Universidad Autónoma de Madrid

Centro de Investigaciones Biológicas Margarita Salas (CSIC)

Programa de Doctorado en Microbiología

Characterization of labrenzin biosynthesis in marine alphaproteobacterium *Labrenzia* sp. PHM005

PhD thesis

Dina Kačar

2021

Acknowledgements

This Thesis is a result of contributions of many people in so many ways. Only looking back at it, is possible to realize how it evolved during time along with so many influences that shaped this work and myself, personally and professionally. I want to express deep gratitude to all and single one person that has been a part of my adventure because I grew with them and I will never forget it.

I want to thank Jose, who was the one who saw the potential in me in the first place and gave me the opportunity to learn as much as I could. The times we had brainstorming and numerous discussions were never dull and would always end up with better ideas. Thanks to him I also learned to question everything and all the time, which made me a slightly better scientist.

Thanks to Bea, who taught me to keep a cool head and relativize when it was needed, but also to prioritize well. I could count on her on a daily basis and she was always ready to help with all my experimental and personal doubts.

This Thesis would not be possible without the collaboration with PharmaMar, who allowed this work to happen. I want to thank to Carmen, Fernando, Ada, Pilar and Elena for your dedication, commitment and high professionalism during these years. Your help was essential for my learning process and scientific accomplishments.

This work also brought us to together with the group of Jörn Piel. I am very thankful for all the insights, suggestions and great teamwork received from Jörn and Stefan.

I would like to thank to my heroes from our lab 344, the place where we spent so many hours together and eventually became a family. Juan and Gerardo, we worked hard and party harder. There is no person who could share my pains better than you. But above all, we laughed so much and had a great time together. Sonia is the most positive person I know. Thank you for reminding me to seek the best in people and situations, and thank you for being a friend. Thank you, Loreine, for showing your support and kindness when needed. Even though for a short time, I won't forget the company and contributions of visiting pals Gabriel, Paula, Maria I...

There were times that I needed some technical assistance and you guys never denied your help. So thank you Juan for leaving what you were doing to help me out, Virginia and Natalia for immense help with HPLC and bioreactor; Manuel for letting me the magical MFD strain that saved my Thesis; Gonzalo for borrowing the right plasmid at the right time and not to forget all the technical service from the Technical Service, you rock. I would like to thank to all the CIB staff for so many pleasant interactions.

I really appreciate all the inputs of Auxi, Eduardo, Manuel and Pedro throughout the years.

Thank you Ana, for managing my displaced bills, fast and efficient as you are, and thanks for giving away your secrets of how to do make perfect competent cells. Your technical assistence was impeccable.

Thank you Unai, David and Elena, for keeping track of my *Lost and Found* stuff, especially concerning pipettes, Petri plates, tubes and similar crime tools. You all contributed to my *Lost and Found* box. Thank you for the patience.

Now, if anything is important to keep our little scientific minds running, it is a good cup of *café con leche*. Thank you Raquél for keeping me well fed. It was highly appreciated. Speaking of coffee, the group Café 2.0 was the best invent of all times. Thank you for all the fun we had, especially Alberto for hilarious impressions of my accent, Maria C for being thoughtful and caring, Sofia for your contagious energy and good humour.

I wish all the jobs would come in a pack with wonderful work mates as you are. I was happy to coincide with you and would always repeat the experience. Thank you (Helena, Ana María, Maria Manoli, Sergio, Fran, Natalia T, Ryan, Oliver, Roberto, Susana, Cris, Irene, ...) for sharing lab days and weekends with me. My experience in CIB would be significantly less fun if it weren't for the *Rubenes*. You were the essence of my adventure and the cause of so many anecdotes I will be telling many years on. I will never be able to repeat those experiences somewhere else. So thank you for all the good times we had, lunch stories we shared, laughs and fun activities. I hope to see you again.

I want to thank all my friends who followed me on this path somewhere between Croatia and Spain and stayed faithful. Thank you Fer for the greatest patience and support during these years and your family who was always there for me.

Finally, I want to thank my family for their unconditional love and support and for letting me be what I wanted to be.

Dina

Index

Abbreviations

aa	amino acid
ABC	ATP binding cassette
ACP	acyl-carrier protein
ADC	antibody-drug conjugates
Ap^R	ampicillin resistant
APS	ammonium persulfate
AT	acyl-transferase
BSA	Bovine Serum Albumin
BVMO	Baeyer-Villiger monooxygenase
Cy	cyclization domain
CYP450	cytochrome P450
cDNA	coding DNA
Cm^R	chloramphenicol resistant
DAD	Diode-Array Detection
DH	dehydratase
DMSO	dimethyl sulfoxide
dNTP	deoxynucleoside triphosphate
ER	enoyl-reductase
FAS	fatty acid synthase
FDA	Food and Drug Administration
Fdr	ferredoxin reductase
Fld	FMN-containing flavodoxin
Gm^R	gentamycin resistant
GFP	Green Fluorescent Protein
HPLC	High Pressure Liquid Chromatography
Km^R	kanamycin resistant
KR	ketoreductase
KS	ketosynthase
LB	Luria-Bertani
MNP	marine natural product
MS	mass spectrometry
MT	methyltransferase
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
nm	nanometer
NMR	nuclear magnetic resonance
NP	natural product
NRP	non-ribosomal peptide
NRPS	non-ribosomal peptide synthetase
OD	optical density
OR	oxidoreductase
ORF	open reading frame
OSMAC	One-Strain-Many-Compounds
PCP	peptidyl carrier protein
PCR	polymerase chain reaction
PKS	polyketide synthase

Abbreviations

PPTase	phospho-pantheteinyl transferase
ROV	Remotely Operated Vehicle
RP	redox partner
RT-PCR	reverse transcription PCR
SAM	S-adenosylmethionine
SDR	short-chain dehydratase/reductase
Sm^R	streptomycin resistant
TAE	Tris-Acetate-EDTA
TE	thioesterase
UV	ultraviolet
wt	wild type

Summary

Summary

The last two decades, natural product discovery and investigation increased tremendously due to the advances in molecular and analytical tools as well as the constant demand for developing new drugs, especially antibiotics and antitumorals. Polyketides are one of the most diverse compounds in nature, and their biological activity is of extreme interest for the pharmaceutical industry. This work studies the biosynthesis of polyketide labrenzin, produced by the marine alphaproteobacterium *Labrenzia* sp. PHM005 isolated from the marine sediments collected at the depth of 18 m near the coast of Kenya. This compound is a new analog of pederin, one of the most toxic molecules in nature, and has raised interest for the development of new antitumor agents.

Today, more than 30 different molecules constitute the pederin-like family of polyketides, containing almost an identical core region and variable polyketide or amino acid termini. Nonetheless, the gene clusters associated to their biosynthesis have mostly been identified in symbiont metagenomes, as in the case of the polyketides pederin, onnamide, psymberin, diaphorin, and nosperin, thus, making their study difficult or impossible. However, *Labrenzia* sp. PHM005 is a free-living producer of labrenzin and can be easily cultivated in the laboratory, allowing for the first time the in-depth characterization of a pederin-biosynthesis gene cluster.

For that purpose, the complete genome of *Labrenzia* sp. PHM005 has been sequenced and the 79 kb *trans*-AT *PKS* gene cluster responsible for the synthesis of labrenzin has been identified. The bioinformatic analyses allowed for the *in silico* characterization of *lab* genes, identification of the PKS/NRPS domains, delimitation of gene cluster and assembly insights through comparative analyses with other homolog clusters. Genetic engineering tools have been developed to facilitate a functional gene analysis, to decipher the function of most of the genes involved in the polyketide biosynthesis, to develop an elicitors screening method via translational fusions, and to generate 10 new labrenzin intermediates and 10 more hybrid analogs, including pederin.

The inactivation of *lab3*, *lab4*, *lab11* and *lab13* genes encoding HMGS, PKS, ECH and PKS/NRPS proteins, respectively, caused the absence of labrenzin and its intermediates in the PHM005 culture extracts demonstrating that these genes are essential for the biosynthesis of labrenzin polyketide backbone. To understand the role of the complete polyketide backbone in the biosynthesis of labrenzin a point mutation in the last catalytic site of thioesterase domain (Lab13)

was performed. Since labrenzin production drastically decreased in this mutant, it was concluded that the polyketide backbone has to be fully synthetized and released from the PKS/NRPS complex for its further hydrolytic processing into labrenzin. In addition, it was demonstrated that the polyketide cleavage happens inside the cells.

Concerning labrenzin production, the knockout of *trans*-AT Lab9 significantly decreased the production of labrenzin and its intermediates supporting its role in proofreading. This mutation also demonstrated that the other *trans*-AT, Lab10, alone is sufficient for polyketide assembly.

Polyketide tailoring reactions have shown to be a useful target for generating *à-la-carte* intermediates. Methyltransferases Lab6 and Lab16 are responsible for *O*-methylations of labrenzin core. Lab6 has a dual function and it sequentially methylates the C10-OH and C17-OH groups. Lab16 methylates the C6-OH and its knockout caused the quirality of the new generated intermediates. Both enzymes are substrate specific and their activity is not redundant. This work also revealed that oxidoreductase Lab5 is involved in the hydroxylation of C7 position, since its elimination caused the accumulation of the intermediates lacking one -OH group. On the other hand, a tandem deletion of cytochrome P450 Lab7 and oxidoreductase Lab8 demonstrated that both enzymes are necessary for the hydroxylation at C10 position.

It was possible to heterologously express methyltransferase PedO, encoded in the *Paederus fuscipes* symbiont metagenome, and to produce for the first time pederin in a cultivable bacterium. It was demonstrated that PedO is less substrate specific, since it catalysed more than one methylation.

Finally, the genomic and transcriptomic analyses carried out with this cultivable bacterium revealed new insights into the biosynthesis of pederin family compounds and set the foundations for the future studies destined to better understanding of the biosynthetic mechanisms of the complex PKS/NRPS systems as well as industrial applications and developments of new drugs.

Introduction

1. Natural products in drug discovery

Nature has always been a great provider of resources for humankind to produce food and materials, build communities and make their overall life easier. Historically, natural plant products were the first to be medically used in treatments of different kinds of diseases, since the ancient civilizations of Mesopotamia, Egypt, China or India (Cragg and Newman, 2013). One of the biggest historical outbreaks in fighting against bacterial infections and malaria were some well-known natural products such as penicillin isolated from *Penicillium*, quinine isolated from the bark of *Cinchona* species or artemisinin isolated from wormwood *Artemisia afra*. Other plants used for the isolation of anticancer compounds were alkaloids paclitaxel (Taxol®) from *Taxus brevifolia* and vinblastine from *Catharanthus roseus*. Whereas in the traditional medicine the total plant extract would be applied for a certain remedy, the modern medicine aims to the application of standardized protocols, safety, controlled dosage and precise molecular targeting in disease treatments. That inevitably requires isolation of the active compound and subsequent pharmacokinetic, pharmacodynamic and other analyses for a proper drug development.

Although natural sources have been widely used in drug discovery throughout the years, there are still some important challenges to overcome. Often, chemical mixtures from the total extracts have synergetic effect. It is not uncommon a single molecule isolated from the natural active extract turns out to be biologically inactive. On the other hand, the exact mechanism of biological activity in some natural drugs often remains unknown. Discovery of natural products may be also impaired as some biodiversity protection laws try to protect species from extinction and environments from human alterations. Another challenge is to improve the screening methods to avoid the rediscovery of previously identified natural compounds, known as dereplication, which is both money and time consuming. In line with that, the advances in computational and analytical methods are extremely important in the determination of new bioactive molecules (Halton and Clark, 2019; Thomford et al., 2018).

Despite the difficulties in the isolation of natural compounds, elucidation of molecular mechanisms, challenges in production and standardization procedures, about 42% of all the FDA approved drugs from 1981 to 2019 were natural-derived (Newman and Cragg, 2020). That including biological drugs: peptides isolated from an organism or biotechnological products; entirely natural products; plant natural products and derived natural products with synthetic

modification. Figure 1 shows a graphical distribution of FDA drugs from some of the categories. Most of the natural products and their derivatives are antibacterial and antiparasitic drugs, followed by anticancer drugs.

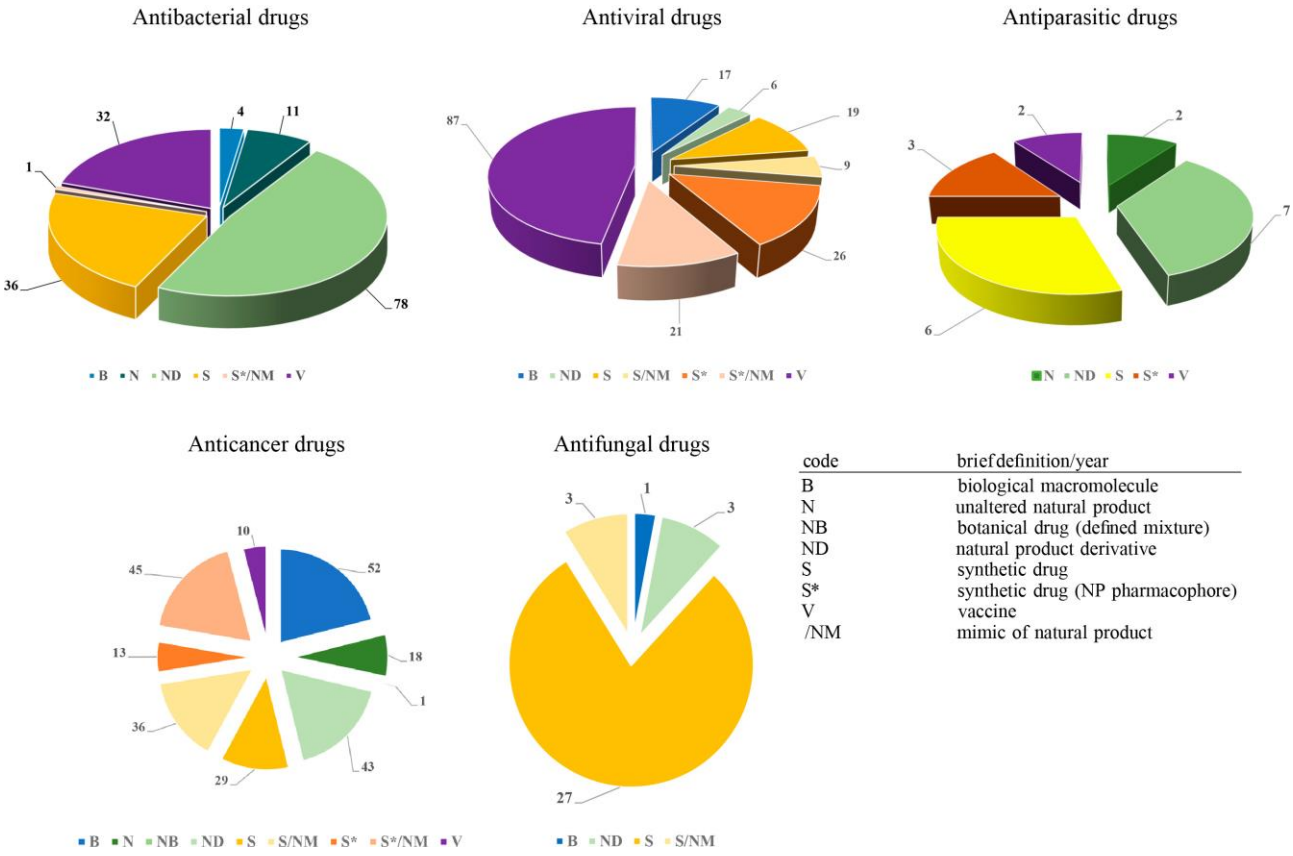


Figure 1. Pie charts representing the number of FDA approved drugs by their source from January 1981 to September 2019 (modified from Newman and Cragg, 2020).

After the Golden Age of antibiotics in the 1950's and 1960's, a peak of the natural drug production was reached in the short period from 1985 to 1988 and since then the discovery of new products has been declining (Newman and Cragg, 2020). Nevertheless, the emergence of new and rapid evolution of existing diseases, as well as the apparition of pathogens multi-resistant to antibiotics, have made both industry and academia to turn back to the natural product discovery considering the advantages that offer natural chemicals like high diversity, biochemical specificity and ability to interact with and bind to biological targets (Martins et al., 2014).

1.1. Marine-derived compounds

Oceans and seas occupy around 71% of the planet, so it is not surprising that the discovery of bioactive compounds from the invertebrates and microorganisms present in marine ecosystems have become extremely popular over last five decades. Particularly knowing that the average ocean depth is 3800 m and only less than 5% and 0.01% of deep sea and deep sea floor has been explored, respectively (Ramirez-Llodra et al., 2010). The inspection of marine environments has excelled in 1970's with the introduction of modern snorkeling and scuba diving and in 1990's with the implementation of remotely operated vehicles (ROVs) that allowed the exploration of larger depths (Dias et al., 2012). Marine invertebrate organisms like sponges, corals, bryozoans and tunicates were the first sources of marine natural products (MNPs).

Green, brown and red marine algae have been broadly studied for their antibacterial and antifungal properties. The dominant compounds isolated from these algae are terpenoids with wide range of biological activities: antimicrobial (*e.g.*, crenuladial, a diterpenoid isolated from the brown alga *Dilophus ligatus*); anticancer (*e.g.*, diterpenoids from *Dictyota dichotoma*) and even pesticide (*e.g.*, halogenated terpenes from the red alga *Plocamium cartilagineum*) (Dias et al., 2012).

Sponges (phylum Porifera) are one of the most primitive animals completely lacking nervous, circulatory and digestive system. They obtain food and other resources by filtering the whole marine environment, so it is not surprising that numerous natural products have been isolated from sponge biomass. Moreover, as sessile organisms, they have developed multiple chemical defense mechanisms against predators. Historically, several products isolated from sponges have generated potent pharmaceutical compounds such as, the antiviral nucleoside Ara-A, the anticancer agent Ara-C, and the macrolide eribulin mesylate approved for the treatment of metastatic breast cancer (Dias et al., 2012; Dyshlovoy and Honecker, 2019). Marine sponges have been the primary source of the MNPs until the last decade. The number of newly isolated compounds started slightly declining, assumingly due to the research shift towards marine microorganisms, often demonstrated to be the real producers of pharmaceutical substances (Taylor et al., 2007).

Species of marine fungi *Aspergillus*, *Penicillium*, *Talaromyces* and *Trichoderma* are also known to produce a wide spectrum of biologically active molecules as radical-scavengers, inhibitors of isocitrate dehydrogenase, anticancer agents and antibiotics against methicillin-resistant strains of *Staphylococcus aureus*, vancomycin-resistant strains of *Enterococcus faecalis*, or *Vibrio* sp. (Nicoletti and Vinale, 2018). Apart from fungi, numerous free-living bacteria and symbionts associated with sponges and other marine invertebrates produce remarkably versatile natural products. Since 2014, the microorganisms are dominating the MNPs discovery whencompared with higher organisms. The biggest producer of natural products among the marine bacteria is actinomycete genus *Streptomyces*, occupying more than 50% of bacterial MNPs followed by *Pseudoalteromonas*, *Nocardiopsis* and *Bacillus* (Carroll et al., 2019). In addition, since 2016, the number of MNPs identified from cyanobacteria increased by more than 70%.

1.1.1. Marine anticancer drugs

It has been 50 years since the first antitumor drug of marine origin, cytarabine (Ara-C, Cytosar-U®), was FDA approved for leukemia treatment and introduced on the markets (Dyshlovoy and Honecker, 2019). The original prototype molecules spongothymidine and spongouridine were isolated 18 years earlier from sponge *Cryptotethia crypta* collected by the cost of Florida (Bergmann and Feeney, 1951). Marine-derived compounds are known to possess extremely cytotoxic activity. The most potent antitumor compound ever known, spongistatin, originated from a sponge, has an IC_{50} of 10^{-12} M in breast cancer cells (Stonik, 2009). Although the MNPs exhibit very strong cytotoxic properties, they are not necessarily the best option for anticancer treatments in humans or animals and often are excluded from further clinical trials due to their adverse side effects. This is mainly the reason why marine anticancer drugs, that managed to get to the clinic, are very limited. Approved marine-derived compounds used in cancer treatment as of August 2020 are shown in Table 1 (Mayer, 2020).

Table 1. Anticancer marine-derived drugs approved by the responsible institutions until 2020.

Compound	Type	Trademark	Marine organism	Disease	Company
Lurbinectedin	**Alkaloid**	**Zepzelca TM (2020)***	Tunicate	Metastatic Small Cell Lung Cancer	**PharmaMar**
Enfortumab Vedotinejfv	ADC (MMAE)	**PADCEVTM (2019)***	Mollusk/ cyanobacterium	Metastatic urothelial cancer	**Astellas Pharma & Seattle Genetics**
Polatuzumab vedotin (DCDS-4501A)	ADC (MMAE)	**Polivy TM (2019)***	Mollusk/ cyanobacterium	Non-Hodgkin lymphoma, Chronic lymphocytic leukemia, Lymphoma, B-Cell lymphoma, Follicular	**Genentech/ Roche**
Plitidepsin**	Depsipeptide	**Aplidin®**	Tunicate	Multiple Myeloma, Leukemia, Lymphoma	**PharmaMar**
Trabectedin (ET-743)	Alkaloid	**Yondelis® (2015) ***	Tunicate	Soft Tissue Sarcoma and Ovarian Cancer	**PharmaMar**
Brentuximab vedotin (SGN-35)	ADC(MMAE)	**Adcetris® (2011)***	Mollusk/ cyanobacterium	Anaplastic large T-cell systemic malignant lymphoma, Hodgkin's disease	**Seattle Genetics**
Eribulin Mesylate (E7389)	Macrolide	**Halaven® (2010)***	Sponge	Metastatic Breast Cancer	**Eisai Inc.**
Cytarabine (Ara-C)	Nucleoside	**Cytosar-U® (1969)***	Sponge	Leukemia	**Pfizer**

***FDA Years Approved ** Australia Dec 2018 Approved.**

According to global marine pharmaceutical clinical pipeline updated by Mayer (2020), there are currently 21 other potential anticancer candidates undergoing phase I, II or III of clinical trials. Besides, more than 10000 marine originated compounds have reported *in vitro* activities and about 2000 showed *in vivo* biological activities (Dyshlovoy and Honecker, 2019). The dynamic field of marine drugs discovery and production is rapidly growing. During the last 10 years, the number of newly discovered MNPs reached more than 1000 per year and the advances in synthesis methods allowed the mass production of active substances necessary for the clinical trials. Currently, there are 4 promising drug candidates undergoing phase III clinical trials in the US: Tetrodotoxin (Halneuron^{TM}, Wex Pharmaceutical Inc.) developed as a painkiller for chronic cancer-related pain; fungal diketopiperazine Plinabulin (NPI-2358, BeyondSpring Pharmaceuticals Inc.) tested in non-small cell lung cancer and brain tumor; Lurbinectedin (PM01183, Zepsyre®, PharmaMar) tested against small cell lung cancer, ovarian and breast cancer; and β-lactone-γ-lactam Marizomib (NPI-0052, Salinosporamide A, Triphase) targeting multiple cancer diseases.

1.2. Prospects in biotechnology

Despite the isolation of more than 11000 new drug compounds from 1990 (Martins et al., 2014), there have been significant challenges in drug production until the date. One of the most important is the difficulty to harvest the compounds from the often rare and dispersed marine producers. Also, the concentrations of the active compounds in marine organisms are extremely low, whereas to obtain the amount necessary for the chemical elucidation of MNPs it is necessary to collect immense quantities of the marine biomass. For example, initially three tons of sponges were required for the isolation of 0.8 mg of spongistatin and another 400 kg of sponges from different locations was used further to obtain 10 mg of spongistatin. Also, for the isolation of 18 g of bryostatin-1, from the bryozoan *Bugula neritina* that grows in colonies on the ship docks and piers, 10000 gallons of bryozoans were needed (Stonik, 2009). In some cases, the strategy used to produce enough amounts of marine active compounds for the preclinical and clinical studies, was to culture the marine invertebrates in marine farms. This strategy was used by the US National Cancer Institute for the isolation of halichondrin B and for the isolation of trabectedin by PharmaMar. Nowadays, the active compounds are produced mainly by chemical synthesis or semi-synthesis through a mixture of chemical and biotechnological processes using natural or genetically modified microorganisms.

More recently, the production of antibody-drug conjugates (ADCs) has become very popular since it increases the target specificity of the drugs, for instance in cancer therapies. ADCs are composed of the naturally derived cytotoxic compound, a lead molecule with the attached antibody designed to target specific proteins on the cancer cell membranes. By the end of 2019, 83% of the marine anticancer drugs undergoing clinical trials were antibody-drug conjugates (Dyshlovoy and Honecker, 2019).

In the last decade, the advances in DNA sequencing have led to the identification of genomes of new bacteria and opened a new window to inspect their secondary metabolites biosynthesis potential. Nonetheless, only about 30% of known bacterial phyla had cultivable representatives which significantly limited the industrial potential for the production of active compounds based on bacterial fermentations (Achtman and Wagner, 2008). However, thanks to advances in next generation sequencing (NGS) technologies, rapidly growing metagenomics projects are deciphering biosynthetic gene clusters of natural products in uncultured bacteria

(Wilson and Piel, 2013). Identification of novel biosynthetic gene clusters alone is not enough, since many of them are downregulated or 'silenced' even in the cultivable microorganisms due to the lack of the relevant activators or inadequate culture conditions. To overcome these challenges, several approaches have been applied during the recent years such as varying the culture conditions by OSMAC (One Strain Many Compounds) approach, co-culturing, addition of chemical elicitors, improving the translation rates by ribosome engineering, manipulation of specific and global regulators, quorum sensing and heterologous expression among others (Zarins-Tutt et al., 2016).

Another important challenge in MNPs identification is to reduce the high rediscovery rate of natural active products. To avoid time consuming and costly re-isolation of already known products, chemical dereplication plays a crucial role (Dinan, 2006). Technological advances in the recent years have made high-throughput screening (HTS) largely implemented in the industry and analytical methods such as high-performance liquid chromatography (HPLC), nuclear magnetic resonance spectroscopy (NMR) or mass spectrometry (MS) are constantly improving (Kind and Fiehn, 2017; Pérez-Victoria et al., 2016). Particularly, the combined robust chromatographic strategies employed since the last 20 years have brought revolution to chemistry of MNPs, alongside with the rapid growth of spectra databases (Pérez-Victoria et al., 2016).

Together with computational biology and analytical chemistry advances, genome mining coupled with genetic engineering, systems biology and synthetic biology techniques will be the key approaches to optimize the production of valuable natural products in industrial microorganisms. In order to achive that it is essential to understand the molecular basis of biosynthetic mechanisms and determine the functional roles of genes involved in the NPs biosynthesis.

2. Polyketides and non-ribosomal peptides

Secondary metabolism compounds produced by symbionts microorganisms are known to give the host certain competitive advantages like defense from predators, competition for resources or interaction with the same or different species. They are not considered essential for the survival of the host since they are not directly associated with growth or reproduction processes. As described in the previous chapter, natural products are secondary metabolites of great industrial interest due to the wide range of useful products derived from them, such as pharmaceutical drugs, enzyme inhibitors, immuno-suppressants, growth promoters, pigments, toxins, pheromones, pesticides, plastics, etc. Secondary metabolites comprise molecules like alkaloids, terpenes, glycosides, aromatic compounds, and different kinds of polymers, among other, but also polyketides and non-ribosomal peptides, which are the subjects of this doctoral thesis.

Polyketides are remarkably diverse and complex natural products distributed across the wide spectrum of ecological niches including water and soil environments, fungi, plants and animals. Their synthesis has been extensively studied, but still leaves a great number of unresolved questions and mechanisms around this field. In line with the work developed in this doctoral thesis, the next paragraphs will introduce the main issues concerning the polyketide and non-ribosomal peptide synthesis.

2.1. Polyketide synthases

Polyketide synthases (PKSs) are multi-enzymatic megasynthases up to 10 mDa MW responsible for the synthesis of numerous polyketides (Lowry et al., 2016). PKSs share high similarity in biosynthetic precursors, architecture and chemistry with animal saturated fatty acid synthases (FAS). Both are giant enzymes consisting of covalently bonded catalytic domains performing similar chemical reactions. Whereas FAS comes and acts as a single copy and iteratively synthetizes long fatty acid molecules of primary metabolism with unique function, many PKSs combines several covalently connected FAS-like units, called modules, in a sequential manner to generate a structurally and functionally diverse group of secondary metabolic products known as polyketides. Recent advances in genome sequencing and genome mining have led to the identification of different types of PKS systems in a wide range of bacterial, fungal, plant or lichens genomes and metagenomes.

2.1.1. Modular organization of PKS

Modules define minimal biosynthetic units of PKS allowing the selection, processing and assembly of the polyketide precursors. Modular PKSs use small acyl-CoA molecules, mostly malonyl-CoA and methylmalonyl-CoA, as building blocks for polyketide assembly. Each acyl-precursor is being incorporated in the growing polyketide backbone resulting in elongation by two carbon atoms. Depending on the type of PKS, modules can be sequentially bound in the large polypeptide or exist separately. One module consists of characteristic biosynthetic domains necessary for the assembly. There are several types of basic modules or domains:

Ketosynthase (KS) is a key enzyme in polyketide biosynthesis because it catalyzes C-C bond between the growing polyketide chain and the incoming acyl building unit. The mechanism of synthesis is a decarboxylative Claisen condensation. Because it is highly specific for the incoming substrates, this fact is exploited for the bioinformatic predictions of the nascent polyketide intermediates (Helfrich et al., 2019).

Acyl carrier protein (ACP) is a non-catalytic domain that acts as a carrier of the growing polyketide chain and the extender unit. It exists in two forms, *i.e.*, an inactive apo form and an active holo form. It requires a prosthetic 4'-phosphopantetheinyl group for its activity, which is transferred by a 4'-phosphopantetheinyl transferase (PPT).

Acyl-transferase (AT) is in charge of selection of specific acyl units and their covalent binding to 4'-phosphopantetheinyl group of the ACP prior to polyketide extension. It has been reported recently that some AT variants can also have proofreading role by hydrolyzing the thioesters with incorrectly inserted extender unit or stalled intermediates (Jensen et al., 2012).

Apart from the basic domains, several optional domains responsible for changing the oxidation state of the β-keto group or additional processing can be present in PKSs. They can be included in the module or acting in *trans* as separated enzymes. Some examples are:

Ketoreductase (KR) stereo-selectively reduces keto-group of the keto-acyl-ACP intermediate resulting in hydroxyl-group. It belongs to a family of short-chain dehydratases/reductases (SDRs), therefore it is NAD(P)H-dependent.

Dehydratase (DH) catalyzes the reaction of water removal in the β-hydroxy-acyl-ACP intermediate, previously generated by KR, and results in α-β double bond formation in *cis* or *trans* configuration.

Enoyl-reductase (ER) subsequently hydrates the double bond formed by DH and generates the enoyl-ACP intermediate from the acyl-ACP.

Methyltransferase (MT) catalyzes the reaction of the addition of a methyl-group to carbon, nitrogen or oxygen atoms of the polyketide residue, differentiating C-MT, N-MT and O-MT, respectively. In general, they possess two subdomains, one for accepting the methyl donor S-adenosylmethionine (SAM) and the other for the methyl acceptor, which is a nascent polyketide.

Thioesterase (TE) is involved in the releasing of the assembled polyketide from the multi-enzyme complex. This step is catalyzed by cyclization of the final product or hydrolysis of the thioester bond between the ACP and the product. There are two types of TE, *i.e.*, type I acts in *cis* and is positioned at the end of the last PKS module, and type II TE is separated *trans* acting enzyme which can act to release the final product, but also to remove unusual or stalled residues during the assembly (Kotowska and Pawlik, 2014).

2.1.2. PKS assembly mechanism

Based on the polyketide synthesis model best studied, *i.e.*, the 6-deoxyerythronolide B synthase (DEBS), which produces macrocyclic core of the antibiotic erythromycin, modular type I PKSs are the reference for the PKS assembly mechanism. A peculiar property of PKSs is the linearity between their modular architecture and the nascent polyketide chain that they synthesize, whereas the polyketide chemical structure reflects the order of the enzymatic domains in the PKS. This is known as the colinearity principle and allows to predict the molecular structures of polyketides as well as to directed useful genetic modifications for the production of new desired products (Hertweck, 2009).

As previously mentioned, the minimal biosynthetic unit of a PKS is composed of essential catalytic domains forming a single module. Therefore, selection of the extended unit and elongation of the growing polyketide chain is performed in cycles by each module. The modular polyketide synthesis cycle includes four steps and is known as turnstile mechanism (Lowry et al., 2016). The closest metaphorical example for this mechanism would be the metal bar gates in a metro station

which allow to pass only one person at a time, because two persons do not fit in one entry. Following the same analogy, each PKS module can receive the incoming intermediate from the upstream module only after passing the current intermediate to the downstream module. The four transformations of this mechanism include entry translocation (I), transacylation (II), elongation (III) and exit translocation (IV) (Figure 2).

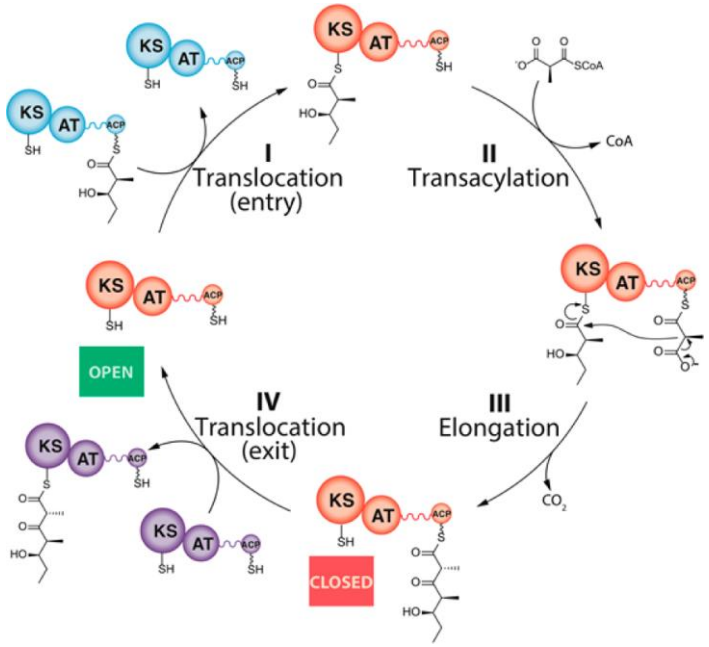


Figure 2. Scheme representing a turnstile mechanism for the PKS assembly. (I) Translocation of the intermediate from the previous module (blue) to the KS domain of the following module (orange); (II) AT-catalyzed transacylation of the extender unit from the methylmalonyl-CoA donor to ACP; (III) Elongation by decarboxylative Claisen condensation; (IV) Translocation of the synthetized intermediate to the ACP of the downstream module (purple) (Lowry *et al.*, 2016).

Entry translocation involves the covalent binding of the polyketide chain from the upstream module to the entry site, whereas the thiol-group of the active cysteine residue of the KS domain. In transacylation step, AT catalyzes the selection and transfer of the acyl-unit from the acyl-CoA donor to the ACP phosphopanteteinyl arm. In the subsequent elongation step, the polyketide intermediate is being elongated via decarboxylative Claisen condensation between the thioester tethered on KS domain and acyl-ACP. Importantly, although the KS domain becomes vacant after the elongation step, it cannot bind new intermediates before the release of the ACP-tethered thioester. Finally, in the last exit translocation step the growing polyketide chain is being translocated from the ACP to the KS of the next module. It is only then, when the KS domain is

ready to receive another intermediate. This way the iterative KS synthesis is prevented and orderly assembly of the polyketide is assured.

Interestingly, although more enzymatic reactions are necessary by FAS to get a full product, as compared to PKSs, when product rates were measured *in vitro*, PKS was more than 100 fold slower than FAS (Smith and Tsai, 2007). It supports the argument that PKS synthesis is highly beneficial for the host.

2.1.3. Classification of PKSs

Polyketide synthases are divided in three canonical groups, depending on the modular organization and their mechanism of biosynthesis, although there are many combinations of basic types known to date. The overview of biosynthesis principles of each PKS type will provide the understanding of the molecular basis for the remarkable diversity of polyketides in the nature.

Modular type I PKSs can be divided in non-iterative multi-catalytic enzymes (see chapter 2.1.2.) and iterative synthases that reuse one module in a cyclic manner, resulting in homogeneous polyketide product composed of the same monomers. Non-iterative PKSs are mainly found in bacteria, like DEBS for the antibiotic erythromycin production (Figure 3A). On the other hand, iterative PKSs are mainly extended in fungi, like producers of lovastatin implied in lowering the blood cholesterol. Depending on the reduction of β-keto-groups, iterative PKSs are further subdivided in non-reducing (NR), partially reducing (PR), or highly reducing (HR) PKSs.

Additionally, type I modular systems are separated in two distinctive groups; *cis*-AT PKSs and *trans*-AT PKSs. The main difference between them is the localization of ATs. In the *cis*-PKS systems AT domains are integrated in the modules of a polyketide synthase (see Figure 3A), whereas in *trans*-AT PKSs ATs are separated enzymes that associate with the PKS *in trans* forming a complex. Another difference is a much higher diversity of *trans*-AT PKS products because their molecular structure usually does not reflect the modular architecture. *Trans*-AT PKSs could have unusual starter units and may including modules with special domains, non-elongating KS^0 domains, modules split on two proteins or possessing additional tailoring enzymes acting *in trans*, apart from the ATs (Helfrich and Piel, 2016).

Type II PKSs are common in prokaryotes, mainly actinomycetes and, unlike type I PKSs, are composed of separate enzymes that make a minimal biosynthetic set, iteratively active KS_α, KS_β

and ACP. The example of this type of synthases is an aromatic polyketide tetracenomycin C (Figure 3B), acting as anthracycline antibiotic.

Type III PKS are the simplest PKSs comprising only iterative ketosynthases and do not use ACPs but instead use directly acyl extender units as substrates. The total size of produced polyketide is defined by the spatial architecture of the catalytic site. The best known type III synthases are plant chalcone and stilbene synthases, although there are examples found in bacteria and fungi. The mechanism of type III PKS is schemed in Figure 3C.

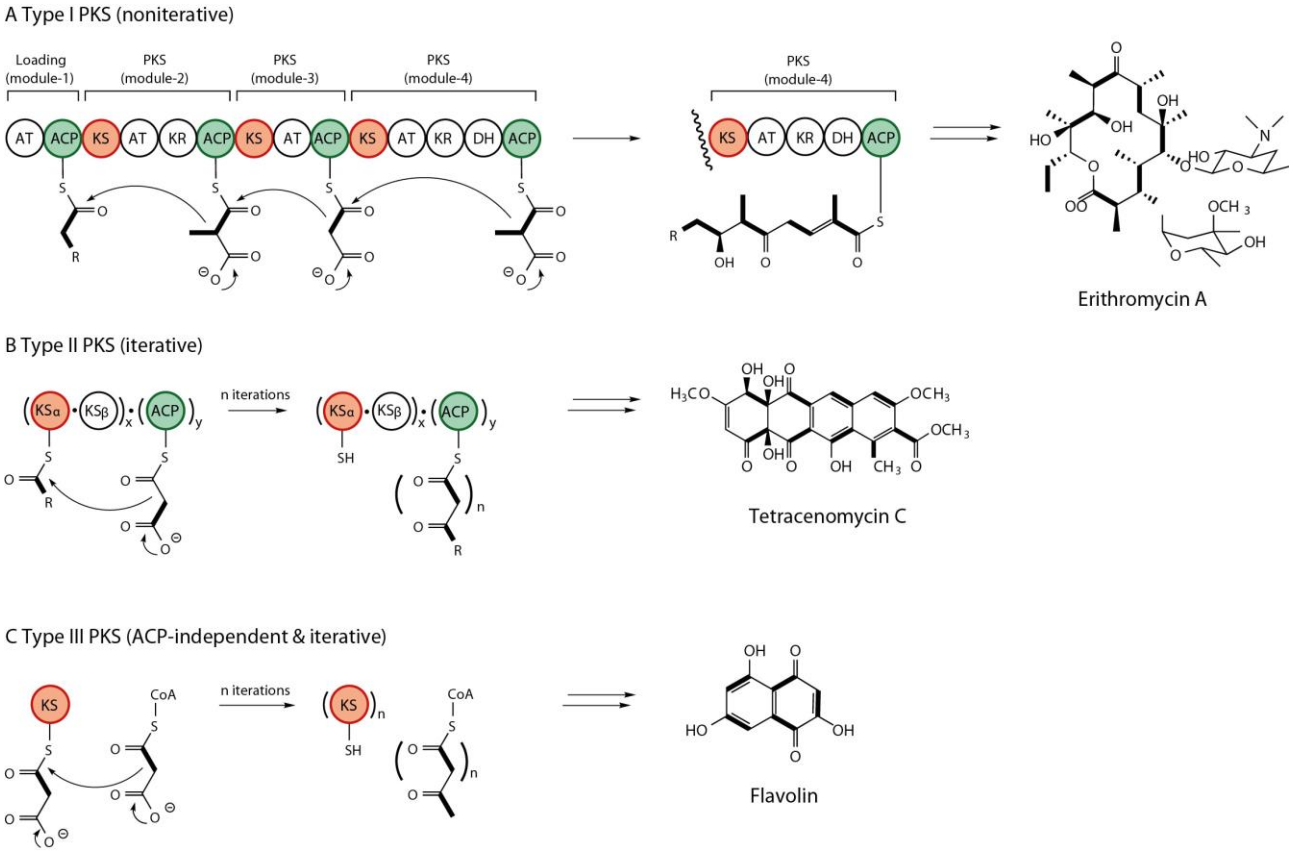


Figure 3. Scheme representing three types of PKS and their corresponding mechanisms of biosynthesis. The final product of each biosynthesis type is shown. A) Mechanism of type I PKS. B) Mechanism of type II PKS. C) Mechanism of type III PKS (modified from Shen, 2003).

2.2. Non-ribosomal peptide synthetases

Ribosomes are main organelles responsible for protein synthesis in all living organisms. Nonetheless, alternatives to produce non-ribosomal peptides with wide varieties of biological activities have important role in the nature. Those molecules are synthetized by non-ribosomal peptide synthetases (NRPSs), macromolecular machineries present in bacteria, cyanobacteria, fungi

and plants. Likewise, PKSs synthetize compounds by generating C-C bonds in an assembly-like fashion, the NRPSs catalyze peptide synthesis by amide bond formation between mostly natural protein-derived amino acids and sometimes acyl units. The integration of different non-protein amino acids, such as halogenated amino acids, aminobenzoic acids, methyl and hydroxy amino acids, phenylglycines and cyclopropyl amino acids (Süssmuth and Mainz, 2017), contributes to the complexity of the chemical structures of NRPs and extends the range of their biological activity.

2.2.1. Modular organization of NRPS

Similar to previously described PKS domain architecture, NRPSs are organized in modules comprising core catalytic domains essential for the peptide bond formation and peptide chain elongation. Although additional modifying domains might be a part of the module, such as methyltransferase (MT) (Jirakkakul et al., 2008), epimerization (E) or cyclization (Cy) domain (Keating et al., 2002), a minimal biosynthetic unit is composed of adenylation domain (A), thiolation domain (T) or otherwise called peptidyl carrier protein (PCP) and condensation domain (C). The last module contains the thioesterase domain (TE) at the C-terminal end which is responsible of releasing the final product by hydrolysis, oligomerization or cyclization (Mootz et al., 2002).

2.2.2. NRPS elongation mechanism

The scheme of biosynthesis mechanism is shown in a Figure 4 (Weber and Marahiel, 2001). The first step is activation of the corresponding amino acid by adenylation domain generating aminoacyl-AMP. Then the substrate is loaded to a thiol group of 4-phosphopanteteyl cofactor of peptidyl carrier protein, using the chemical energy from the phosphoester bond to form a thioester. In the next step the upstream C domain catalyzes the formation of peptide bond between the preceding amino acid bound to a donor position and an activated amino acid delivered to an acceptor position of C domain by PCP. The resulting growing peptide is now loaded to PCP, which allows regenerating of the thiol group of the upstream PCP and translocation to donor position of the downstream C domain to continue the assembly line. The cofactor covalently bound to PCP has a role of swinging arm transferring the intermediates between the domains. This movement is largely random which affects the synthesis speed. The importance of NRPs for the microorganism must be very high, considering that the activity of NRPS is approximately 3 orders of magnitude

slower than the ribosome activity when the protein production rate is compared (Reimer et al., 2018).

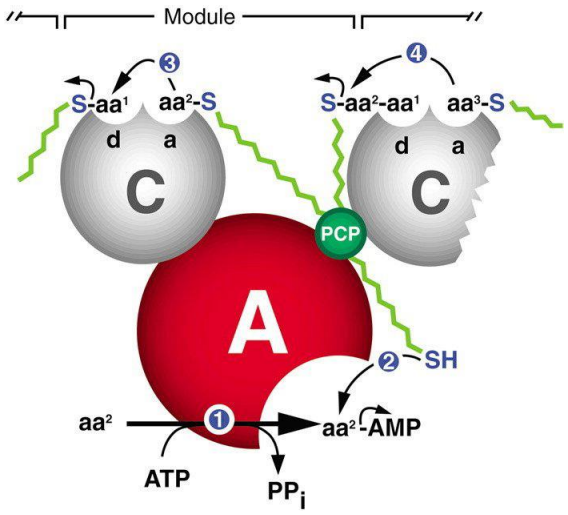


Figure 4. Scheme representing mechanism of peptide bond formation in a modular NRPS biosynthesis: 1) Substrate adenylation by A domain; 2) Loading of PCP-thioester formation; 3) Peptide bond formation by upstream C domain; 4) Peptide bond formation by downstream C domain (modified from Weber *et al.*, 2001).

2.2.3. Types of the NRPSs

Depending on the modular iteration and linearity of biosynthesis, there are three canonical types of NRPSs that synthesize different NRPs (Figure 5).

In linear type A NRPS, each module is responsible for the integration of one amino acid. The same messenger RNA encodes the sequence of the nascent peptide chain on the ribosome during translation, so the resulting chemical structure in NRP assembly can be predicted by the NRPS modular sequence. The number of modules varies, resulting in different NRP length. Some examples of products of linear type A NRPSs are precursors of β-lactam antibiotics like penicillins and cephalosporins (Felnagle et al., 2008), as well as lipopeptide antibiotics like daptomycin (Robbel and Marahiel, 2010) and surfactin (Kraas et al., 2010).

Type B NRPSs are iterative, *i.e.*, all the modules are repeatedly used generating NRPs with repeated moieties, like fungal bassianolide and beauvericin or bacterial gramicidin S and enterobactin (Yu et al., 2017).

Type C NRPSs are responsible for a non-linear NRP biosynthesis employing certain domains inside the module more than once or skipping certain modules. Some modules even miss entire A domains and likely use A domains from neighboring modules (Zhang et al., 2020). Examples of non-linear type C NRPSs products are syringomycin, bleomycin or vibriobactin (Mootz et al., 2002).

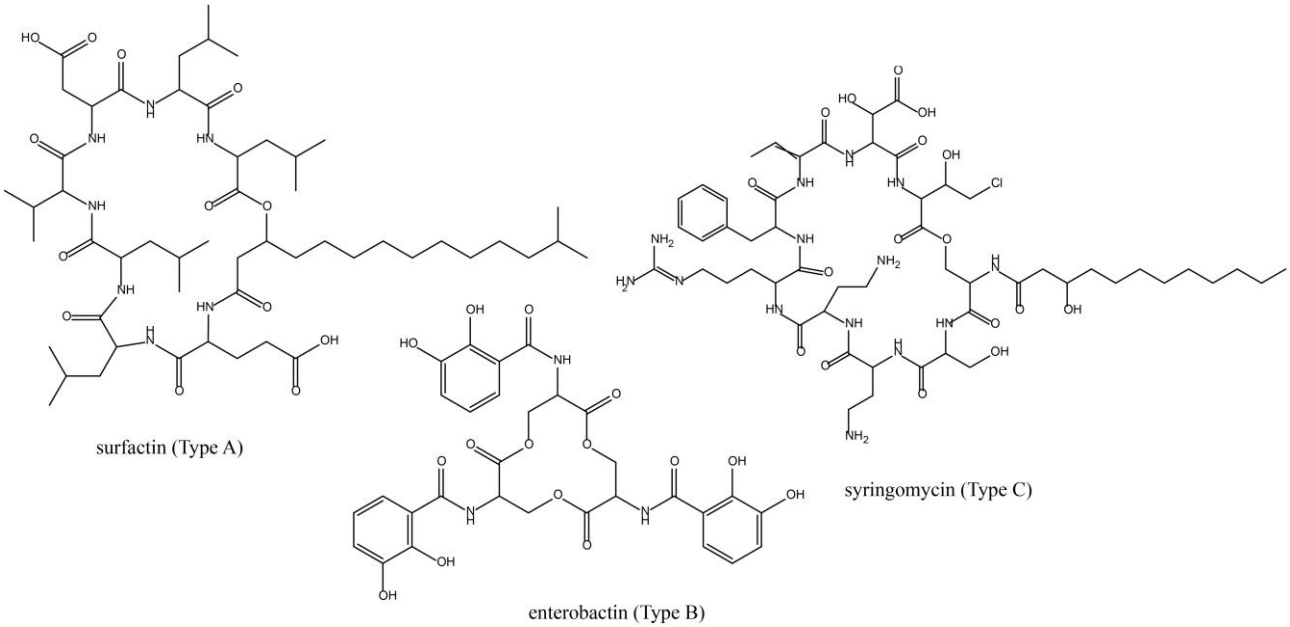


Figure 5. Examples of NRP products produced by type A, B and C NRPSs.

2.3. Hybrid PKS/NRPS

Beside the canonical classification of the NRPSs and PKSs, also there are mixed-type PKSs/NRPSs, both linear and non-linear, which in combination with the possibility of the iteration largely expand the combinatorial power of these multimodular enzyme complexes. In these hybrid systems, PKSs and NRPSs can be tethered in the same multienzymatic molecule or separated in different enzymatic polypeptides. Some examples of hybrid type PKS/NRPS synthetized products are the siderophore yersiniabactin, the tumor promoter microcystin-LR or the immunosuppressant rapamycin (Miyanaga et al., 2018). In marine bacteria *Tistrella* and *Thalassospira* there have been isolated 14 new analogs of immunosuppressant thalassospiramides that differed in structure and size, as a result of non-canonical biosynthesis from the PKSs/NRPSs involved (Ross et al., 2013).

The knowledge gathered through years about the great diversity in modular structure as well as the elucidated mechanisms of PKSs and NRPSs, provided new insights into the impressive biosynthetic versatility of polyketides and an opportunity for the combinatorial biosynthesis in the future. Yet, there are still many functions unknown and questions unanswered. In 2014 was performed a very detailed and extensive comparative analysis between PKS systems obtained from various databases. The survey generated 885 non-redundant PKSs catalog (O'Brien et al., 2014). Among them, only about 14% of the PKS clusters had their natural products identified. The sequencing technology has advanced significantly in recent years, but there are still numerous orphan PKS gene clusters awaiting to be deciphered. Surprisingly, almost 25% of the catalogued gene clusters belonged to *trans*-AT PKS systems, out of which one-half were hybrids of PKS/NRPS. They are most commonly found in microbes who are yet understudied or impossible to cultivate in laboratory conditions. One of the first identified *trans*-AT PKS/NRPS gene clusters was the pederin gene cluster (Piel, 2002). This work provides continuity of Piel's work and broadens the knowledge on pederin-family biosynthetic clusters.

3. Pederin family compounds

Pederin is a potent polyketide toxin known to cause severe dermatitis in humans followed by blisters, skin lesions and necrosis after contact with small rove insects of genus *Paederus*, family *Staphylinidae*, order *Coleoptera* (Figure 6). A collection of 25 million of *Paederus fuscipes* beetles was needed to isolate the crystalized chemical compound for the first time in 1953 (Pavan and Bo, 1953). The exact chemical structure was elucidated later (Quilico et al., 1961). There are more than 600 known species of *Paederus* genus but less than 4% are reported to cause skin lesions. These rove beetles are widespread around the globe, inhabit mostly wet places like riversides and lake banks or irrigated fields. They used to bring the attention after localized epidemic dermatitis outbreaks, which are also common nowadays, usually in subtropical and tropical climates, caused by light attraction of the insects by night, like in the dormitories, hospitals, on riverboats, oil camps or military units (Frank and Kanamitsu, 1987; Heo et al., 2013; Ngatu, 2018; Suwannahitatorn et al., 2014). The insects do not harm by stinging or biting, but only cause the blisters after accidentally, or intentionally, crushing them against the skin. Hours after the contact, the reaction progressively causes rash and itching, blisters and severe skin damage, which depends on the amount of the chemical in contact (Figure 6). It also causes strong ocular reactions like increasing pain, difficulty in opening an eye and conjunctivitis. *Paederus* dermatitis heals by itself over time,

but washing affected skin surface with soup and water prevents infections and reduces the risks formation of lesions.

Figure 6. Images of beetle *P. fuscipes* and the dermatitis skin symptoms it causes (modified from Ngatu (2018)).

Today, there are more than 30 molecules identified that belong to the group of pederin-like molecules that have been isolated from different micro or macro-organisms. Decades after the first pederin isolation, studies have revealed the potent anticancer activities of pederin family compounds and raised the interest of pharmaceutical industry in anticancer drug production. The biosynthesis of the closest pederin analog, labrenzin, is studied and described in this doctoral thesis.

3.1. Pederin characteristics

3.1.1. Toxicity

The review made by Frank and Kanamitsu (1987) describes detailed toxic effects of pederin collected from Japanese literature. Pederin is reported to be at least 15 times more potent than cobra venom, which makes it as one of the most potent animal toxins known. The toxicity was tested on a wide range of vertebrates, however, the evidence that pederin is used as defense against predators remained inconclusive. The pederin effects presented variations in dosage needed to provoke the reaction as well as in responses in different sites of injection in each tested animal. For example, for an ocular reaction, a much higher concentration was needed in chickens than in laboratory mice, rabbits and rats, while frogs and hedgehogs showed no reaction at all. Considering the animals who directly feed on the insects and the lack of the toxin effect, it did not produce expected avoidance of predators. One of the explanations was that pederin was being degraded in animal's stomach since it

was completely destroyed upon treatment with 0.4% HCl (pH 1.8). However, ingestion of pederin could poison horses and cows, and intravenous injections killed rabbits, guinea pigs and mice.

Additionally, it was assayed as a possible antibiotic, but had none or very few effects on bacteria like *Bacillus subtilis* and *Escherichia coli*, whereas in yeast *Saccharomyces cerevisiae* and protozoa *Trichomonas columbae* exhibited strong growth inhibition. *In vitro* studies showed that pederin was active against eukaryotic organisms with 80S ribosomal unit, but not against prokaryotes with 70S ribosome that placed pederin as a ribosome-specific toxin (Tiboni et al., 1968). It was shown that pederin inhibits protein synthesis and mitosis, but the exact molecular target and mode of action was unknown. In an assay with known antibiotics, it was demonstrated that pederin interferes with the steps following the ternary complex formation between mRNA, ribosome and aminoacyl-tRNA (Tiboni et al., 1968).

Pederin was shown to be especially powerful against tumor cell lines *in vitro*. The first reported anticancer activity of pederin inhibited sarcoma 180 tumors in mice (Pavan, 1958). The severe cytological alterations and the blockage of mitosis in normal and HeLa cells were caused by as little as 1 ng/µL of pederin in less than 2 h (Soldati et al., 1966). It was also showed to be very efficient against murine P388 leukemia cells (Narquizian and Kocienski, 2000). These initial studies raised interest of pharmaceutical industry for the further exploration of pederin-family compounds as potent anticancer agents and potential drugs.

3.1.2. First studies on pederin biosynthesis

Pederin was quantified in both male and female *Paederus* species and unexpectedly, female insects contained 10 times more pederin than males (Kellner and Dettner, 1995). Although pederin was present in all parts of an adult beetle, except for the wings, the highest pederin concentration was accumulated in eggs laid by *Paederus* females. However, there were 10% of females unable to biosynthesize pederin, therefore they could not pass it on the eggs. This, what appeared to be female polymorphism phenomena, allowed Kellner and Dettner (1995) to conduct the first study on functional importance of pederin for the species of *Paederus*. It was noticed previously that some species of wolf-spiders pray on *Paederus* offspring (Kurosa, 1958). Therefore, different insect and spider species were collected from the *Paederus* habitat and fed on pederin positive (+) and negative (-) larvae. Insects showed no selection for (+) or (-) larvae, while the wolf-spiders clearly turned away from attacking (+) larvae and showed cleansing ritual afterwards. This finding

supported the role of pederin as a larval chemical defense from specific predators they cannot easily run away from.

The subsequent study wanted to tackle the heredity of the pederin toxin, since its biosynthesis was shown to be matrilineal trait. Nonetheless, the offspring of (+) mothers did not show the expected Mendelian proportions in laboratory grown new generations of *Paederus riparius* (Kellner, 2000). It was the first time Kellner suggested that pederin producer might be a microorganism horizontally transmitted among the beetles. Furthermore, to prove that the real producer was an endosymbiotic bacterium, an antibiotic susceptibility study was performed. Due to the fact that *Paederus* beetles also feed on conspecifics, (-) female offspring was fed with (+) eggs previously treated with antibiotics benzylpenicillin, erythromycin, oxytetracycline or streptomycin (Kellner, 2001). The treatment with antibiotics resulted in loss of pederin content. Although the antibiotic efficacy depended on their spectrum, dosage and duration of exposure, the greater effectiveness of oxytetracycline and streptomycin indicated that the pederin producer was likely a gram-negative bacterium. Next, total DNA was extracted from *Paederus* females to run the 16S rDNA studies for the molecular identification of the symbiont. A phylogenetic three showed close relationships with the genera *Pseudomonas* (Kellner, 2002).

3.2. Pederin-family gene clusters identified from the metagenomes

Symbiotic bacteria have been long known to play an important role in exchanging nutrients with their host and providing defense against predators. However, majority of those symbionts are impossible to grow in laboratory conditions. The rapid development of DNA sequencing technologies and the eagerness to know the real producer of pederin and molecules alike led to several metagenome sequencing projects. It allows obtaining the genetic material of the bacterial symbionts directly from complex samples like soil or higher eukaryote organisms from marine environments to mine for the biosynthetic gene clusters responsible for the production of bioactive molecules (Figure 7). After sample collection and DNA isolation, there are two common natural product identification approaches: 1) creation of a metagenomic library followed by screening by biological activity or PCR; 2) high-throughput sequencing to discover bioactive small molecules and characterize their biosynthetic genes for future applications. Metagenomics opened the door to the identification of numerous secondary metabolites gene clusters of industrial interest, among which are ever growing pederin-group members.

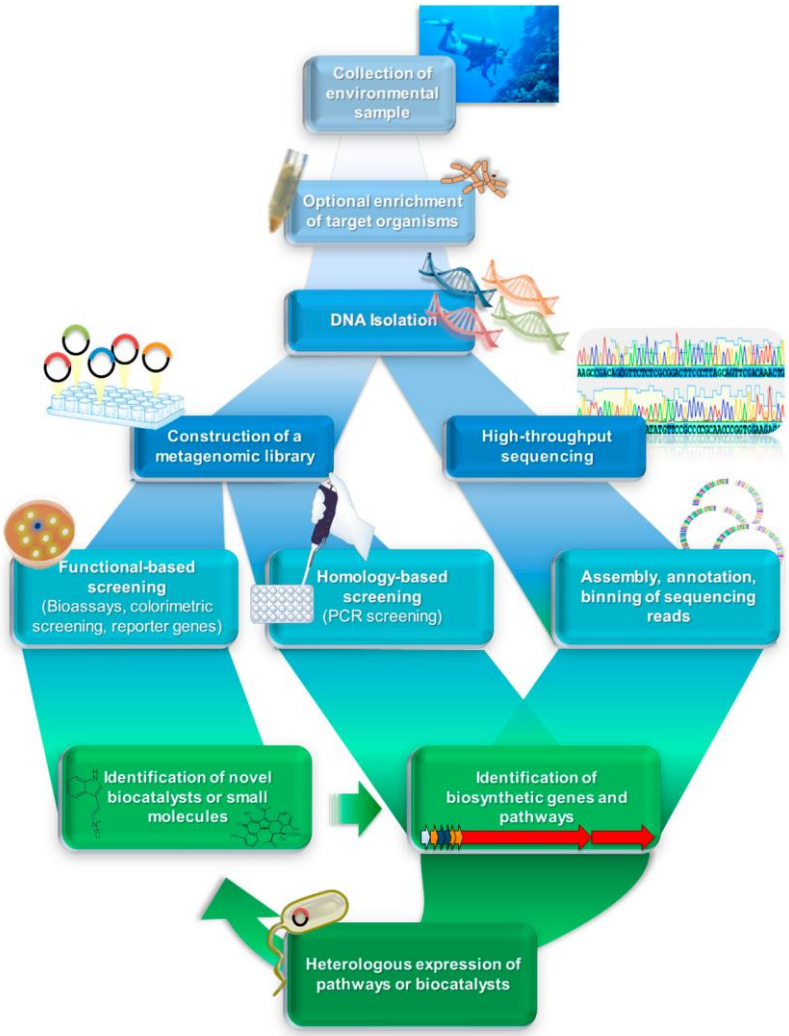


Figure 7. Graphical diagram showing steps of the metagenomic workflow (Wilson and Piel, 2013).

3.2.1. Pederin gene cluster from *Paederus fuscipes*

The first insights into the true producers of pederin were obtained by cloning a fraction of the pederin biosynthesis genes (*ped* cluster) from genomic DNA obtained from *P. fuscipes* beetles (Piel, 2002). The partial *ped* cluster that was initially located on a 54-kb region bordered by transposase pseudogenes encoded a mixed modular PKS/NRPS. The *ped* cluster was distributed among two distinct regions (*pedIJK* and *pedABCDEF*) of the genome from an uncultured bacterial symbiont of the insect that exhibited a similarity to genera *Pseudomonas* (Piel, 2002; Piel et al., 2004a, 2004d). This finding supported the closest relationship to *P. aeruginosa* previously detected in *Paederus sabaeus* (Kellner, 2002). In later studies, the third genomic region was identified to

carry additional *ped* genes (*pedLMNOPQR*) speculated to be involved in the early stages of pederin biosynthesis (Piel et al., 2004c, 2005). Transposase genes, indicating the unusual fragmentation of the secondary metabolites gene cluster, delimited all three *ped* regions.

3.2.2. Pederin-family gene clusters from sponge symbionts

Marine sponges have long been known as a rich source of natural bioactive compounds and a home to innumerous symbiotic organisms. They are one of the oldest multicellular eukaryote organisms and first metazoans inhabiting marine environment (Zumberge et al., 2018). Highly diverse sponge-associated microbes can build up to 40% of their biomass. Moreover, densities of microbial cells can exceed 10^9 per mL of sponge tissue. It is not surprising since 1 kg of sponge approximately can filter up to 24000 L of sea water per day (Taylor et al., 2007). The symbionts mainly contribute to denitrification and anaerobic ammonium oxidation processes, cell signaling, adhesion-related proteins production, removal of toxins from the sponge and transport of nutrients (Webster and Taylor, 2012).

Metagenomic mining projects from marine sponges revealed a great pool of bacterial genes encoding PKSs and NRPSs. Surprisingly, cytotoxic polyketides closely related to pederin have been isolated from several genera of marine sponges distributed all around the planet. To emphasize the patience and scientific dedication invested in isolation of these compounds, in one example 600 fractions combined over 11-year time were needed to chemically elucidate psymberin from *Psammocinia* sp. collected from the coast of Papua New Guinea (Cichewicz et al., 2004). Other pederin-related molecules were: theopederins A-E from *Theonella* sp. collected from Hachijo-jima Island near Tokyo (Fusetani et al., 1992) as well as onnamides A-D from *Theonella* genus (Matsunaga et al., 1992; Sakemi et al., 1988), theopederins K-L isolated from *Discodermia* sp. collected from Honduras (Paul et al., 2002) and mycalamide A from genus *Mycale* obtained from New Zeland (Perry et al., 1988) (Figure 8). All the mentioned compounds, except onnamides, are patented which indicates the importance they have in the pharmaceutical industry regarding their remarkable antitumor properties.

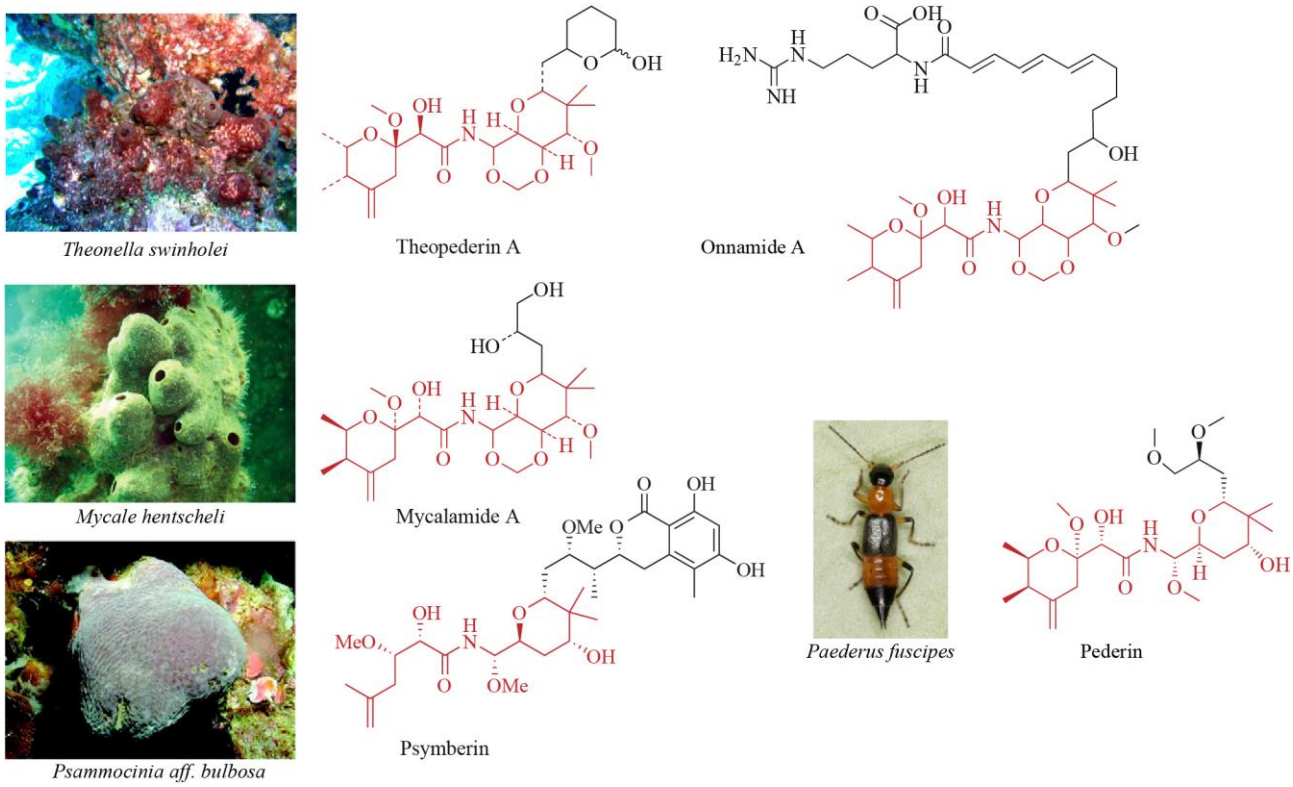


Figure 8. Structurally similar polyketides isolated from *P. fuscipes* beetles and different marine sponges (images taken from Kampa et al., 2013). Molecular core of pederin is colored in red (modified from Newman and Giddings, 2014).

The efforts to identify the gene clusters responsible for the biosynthesis of pederin-family compounds were priority in the following years. Partial onnamide cluster has been identified from the metagenomics library screening from *Theonella* sp. Related genes covering a 36.5 kb region named *onnABCDEFGHI* were highly homologous to *ped* genes identified from *P. fuscipes* metagenome (Piel et al., 2004c). However, one region of the second hybrid PKS/NRPS was missing along with other related tailoring genes. Five years later a complete psymberin gene cluster comprising genes *psyA-N* was found in another metagenomics project using the sponge specimen *Psammocinia aff. bulbosa* (Fisch et al., 2009). Since then, there have been numerous unsuccessfully attempts afterwards to synthetize psymberin chemically (Bielitza and Pietruszka, 2013). Although the complete set of genes necessary for the biosynthesis of this extensively studied antitumor agent was fully available, the heterologous production has not been achieved yet. More recently, using genomic DNA obtained from the sponge *Mycale hentscheli*, a producer of mycalamide A and a great variety of other polyketides, the complete mycalamide gene cluster was revealed (Rust et al., 2020). Whole genome sequencing of sponge microbiome allowed identification of 14 genomes from uncultivable microorganisms, out of which *Candidatus* Entomycale ignis was suggested as a

producer of mycalamide. Four DNA fragments were connected by PCR to generate a *myc* biosynthetic cluster highly similar to previously reported *ped*, *onn* and *psy* gene clusters.

3.2.3. Pederin-family gene clusters from other symbionts

The diverse distribution of pederin-gene clusters across different ecological niches identified from host species divergent one from another have raised great interest in evolutionary relationships of these clusters and deeper understanding of their functional role, along with the defense mechanism. Apart from terrestrial beetles and marine sponges, which had strikingly little specimen and environmental connection, two more pederin members have been discovered from unusual hosts at about the same time. One of them is produced by the lichen-associated symbiont that produces a new pederin analog called nosperin (Kampa et al., 2013). Terrestrial lichen *Peltigera membranacea* was collected in Iceland and whole genome sequencing revealed *trans*-AT PKS gene cluster distributed along 59-kb region that highly resembled pederin type PKS clusters. The core pederin structure is partially shared in nosperin (Figure 9). Phylogenetic analysis of the symbiont producer pointed towards a *Nostoc* sp. cyanobacterium. PCR-based screening also confirmed the presence of *nsp* cluster in other *P. membranacea* collected in Vancouver Island and the mainland. The gene mining showed the presence in several *Nostoc* species isolated from *Peltigera* spp. On the other hand, lichen-no associated cyanobacteria were not positive for *nsp* genes.

Another host-symbiont relationship was described in the case of an Asian citrus pest *Diaphorina citri*. The sequencing of the microbiome of the insect allowed for the assembly of the complete genome of the symbiont bacterium named *Candidatus* Profftella armatura which uncovered the newly identified PKS gene cluster, responsible for the biosynthesis of pederin-family polyketide, diaphorin (Nakabachi et al., 2013). It was the first time the *ped*-like gene cluster was integrated in the whole genome context. Interestingly, the genome size was significantly reduced, since the circular chromosome was only 459,399 bp long, which was considered a trait of a symbiont bacterium in an obligate host-symbiont dependence. The *dip* cluster expanded through two different loci in the genome and was highly similar to *ped* cluster. Even the molecular structure is similar to pederin with the only difference that diaphorin lacks two *O*-methyl groups.

3.3. Pederin-family gene clusters identified in the genomes of free-living microorganisms

The previous studies argued that the *ped*-like gene clusters found in symbiont genomes could provide competitive advantage to their hosts by the means of chemical defense against the predators. It was assumed that the pederin-family compounds were exclusive to symbiont producers. However, a free-living cyanobacterium *Cuspidothrix issatschenkoi* was isolated from moss *Blasia* as a producer of a another pederin-like polyketide cusperin (Kust et al., 2018). The identified *cus* biosynthetic cluster is almost identical to *nsp* cluster, therefore nosperin and cusperin possess very little structural differences (Figure 9). Interestingly, cusperin showed no cytotoxic activity, as opposed to the rest of the pederin-type compounds. Moreover, its producer is a freshwater planktonic cyanobacterium causing toxic water blooms that raised questions about possible roles of pederin-family compounds in the ecological context, other than just defense agents.

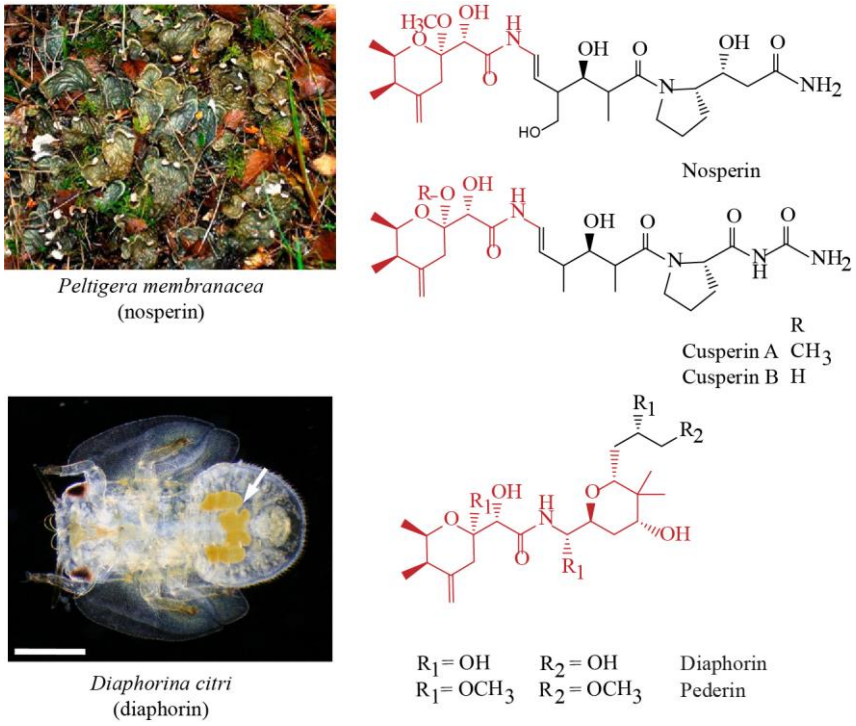


Figure 9. Pederin-family compounds isolated from lichen and insect bacteriome. Image of lichen *P. membranacea* hosting *Nostoc* sp. cyanobacterium, a producer of nosperin (up left, taken from Kampa et al., 2013). Ventral view of a fifth-instar nymph of *D. citri*. The arrow indicates the yellow bacteriome in the abdomen. The scale bar represents 500 µm. (down left, taken from Nakabachi et al., 2013). Respective molecular structures of nosperin, cusperin (image of a producer not shown) and diaphorin are shown on the right.

Introduction

In 2017, an 18-*O*-demethyl pederin (compound 1) has been isolated from the free-living marine alphaproteobacterium *Labrenzia* sp. PHM005. The marine sediment samples were collected back in 2005 at the depth of 18 m near the coast of Kenya. In this doctoral thesis, the compound 1, was renamed as labrenzin. Previous studies on *ped*-like gene clusters did not carried out any functional gene analysis so far, due to inability to either cultivate or genetically modify the producer microorganism. This thesis constitutes one-step forward in understanding the biosynthesis of pederin-type polyketides and opens the door to modify and overproduce these anticancer drugs for industrial and pharmaceutical purposes.

Objectives

The foundation for this doctoral thesis is the recent isolation of labrenzin from the culture of a marine heterotrophic α-Proteobacterium, *Labrenzia* sp. strain PHM005 isolated from marine sediment. This was the first report of the production of this family of natural polyketides by a free-living marine bacterium that could be cultured in the laboratory. Therefore, the focus of this thesis is to characterize for the first time the gene cluster involved in the biosynthesis of the highly cytotoxic polyketide labrenzin. To address properly this aim, the gene characterization is performed through *in silico* analysis and experimental functional studies. Thus, the two main objectives of the thesis and the corresponding specific subobjectives are:

1) *In silico* identification and characterization of labrenzin biosynthetic cluster:

   The aim of this objective will be to localize and annotate the genes for labrenzin biosynthesis in the genome. The specific subobjectives are:

   - Sequencing and analysis of *Labrenzia* sp. PHM005 genome;
   - Delimitation of the labrenzin gene cluster and gene organization;
   - Comparative analysis with homologous gene clusters;

2) Functional analysis of the genes involved in labrenzin biosynthesis:

   The aim of this objective will be to assign specific *in vivo* roles and the activities to the genes responsible for labrenzin biosynthesis. The specific subobjectives are:

   - Exploring transcriptional activity of the genes;
   - Developing genetic engineering tools;
   - Identification of genes that are essential for the polyketide backbone assembly;
   - Identification of tailoring modifications of the core molecule and their assignation to the genes involved;
   - Identification of labrenzin intermediates and generation of new analogs by genetic engineering;

Material and Methods

1. Bacterial strains, plasmids, media and growth conditions

1.1. Microorganisms

Table 2 shows the bacterial strains, their genotype or NCBI accession number and the application in different processes throughout this work.

Table 2 Microorganisms used in this work

Microorganism	**Strain/Reference**	***Genotype/NCBI accession***	**Application**
Escherichia coli	DH10B (Invitrogen)	F⁻. *mcrA, Δ(mrr-hsdRMS-mcrBC), φ80lacZΔM15ΔlacX74, recA1, endA1, araD139, Δ(ara-leu)769, galU, galK, λ⁻, rpsL (*Str^R*), nupG*	Cloning procedures
Escherichia coli	cc118λpir (de Lorenzo and Timmis, 1994)	*Δ(ara-leu), araD, ΔlacX74, galE, galK, phoA, thi1, rpsE, rpoB, argE (Am), recA1, lysogenic (λpir),* Rif^R	Cloning procedures for R6K replication ori plasmids; donor in triparental conjugation
Escherichia coli	MFD*pir* (Ferrières et al., 2010)	MG1655RP4-2-Tc*::[ΔMu1::aac(3)IV-ΔaphA-Δnic35-ΔMu2::zeo] ΔdapA::(erm-pir) ΔrecA*, Apra^R, Zeo^R, Erm^R	Biparental conjugation
Escherichia coli	HB101 (Promega)	F⁻, *thi-1, hsdS20 (rB⁻, mB⁻), supE44, recA13, ara-14, leuB6, proA2, lacY1, galK2, rpsL20* (StrR), *xyl-5, mtl-1*	Helper strain in triparental conjugation
***Labrenzia* sp.**	PHM005 (provided by PharmaMar)	CP041191 and CP041190	labrenzin producer
Stappia indica	PHM037 (provided by PharmaMar)	CP046908	sesbanimide producer
Labrenzia aggregata	PHM038 (provided by PharmaMar)	JABFCZ010000000	sesbanimide producer

1.2. Culture media

Standard overnight *E. coli* cultures were grown aerobically in Luria-Bertani (LB) broth or LB agar at 37 ºC (ref). The medium was supplemented with diamine-pymelic acid (DAP) (1 mM) and antibiotic, when appropriate (see chapter 2 for the concentrations details). Marine bacteria were grown in Marine Broth (MB) Difco 2216 (Sigma-Aldrich) at 30 ºC with shaking at 220 rpm or Marine Agar (MA) Difco 2216 (Sigma-Aldrich), supplemented with antibiotic, when appropriate.

Material and Methods

Unless stated otherwise, all the strains were cultured in falcon tubes or flasks with the medium:air ratio 1:5 (vol:vol).

All the media and the solutions were prepared using MilliQ water followed by the sterilization during 15 min at 121 ºC and 1 atm pressure or by the filtration. The media are described in liquid version; agar plates were obtained by adding 1.5% of bacterial agar, except the already mixed commercial powders, and appropriate concentration of antibiotic, when needed. In addition, DAP (1 mM) was spread on the surface of an agar plate, when required.

Culture medium used to study gene transcription, growth and labrenzin production in *Labrenzia sp.* PHM005 was prepared as follows: 1) modified marine basal media (MBM) (Baumann and Baumann, 1981) with NaCl (20 g/L) and 0.2% glucose as carbon source; 2) modified MBM supplemented with vitamins (MBM + vit) [B12 (50 µg/L), panthotenic acid (50 µg/L), riboflavin (50 µg/L), pyridoxamine (10 µg/L), biotin (20 µg/L), folic acid (20 µg/L), nicotinic acid (50 µg/L), *p*-aminobenzoic acid (50 µg/L), thyamine (50 µg/L)]; (3) MBM supplemented with biotin (20 µg/L) (MBM + B); (4) MBM supplemented with thyamine (50 µg/L) (MBM + T); (5) MBM supplemented with biotin (20 µg/L) and thyamine (50 µg/L) (MBM + B + T). For the growth experiments in other carbon sources, MBM+vit was supplemented with 0.2% citrate, glutarate, benzoate and cyclohexanoate.

Previously described medium MBM+vit containing 2.0% of glucose (M2 medium) was used for enhanced production of labrenzin and other analogs for better peak detection by HPLC/MS.

1.3. Growth conditions for labrenzin production

Labrenzia sp. PHM005 was grown overnight in 50 mL falcon tubes in the MB at 30 ºC with shaking at 200 rpm in the New Brunswick Innova® 44/44R Incubator if not stated differently. The overnight culture was washed in 0.85% NaCl solution and diluted to an optical density (OD_{600}) ≈ 0.1 in 20 ml of fresh medium. For growth monitoring, the OD_{600} was measured by UV-mini 1240 spectrophotometer (Shimadzu) every 3 h during the cultivation period. At the laboratory scale, for the HPLC/MS analysis of the production of labrenzin and other intermediates in wild type (wt) and mutants, the culture was monitorized for 72 h. When necessary, the culture media was

supplemented with appropriate antibiotic (for concentrations see chapter 2) or inducer 3-methyl-benzoate (0.2 mM), when the plasmid expression system required so.

2. Vectors and antibiotics

The cloning vectors and antibiotics used in this work are shown in Table 3. The specific recombinant plasmids constructed in this Thesis are described and named in the corresponding chapters.

Table 3. Plasmids used for cloning in this work

Vector	**Size (kb)**	**Antibiotic resistance**	**Host**	**Replication ori/copy number**	**Reference**
pCR™Blunt II-TOPO®	3.5	Kanamycin	*E. coli*	pUC derived/high	Invitrogen
pSEVA338	4.9	Chloramphenicol	Broad- host	pBBR1/medium	(http://seva.cnb.csic.es) (Silva-Rocha et al., 2013)
pSEVA238	5.1	Kanamycin	Broad- host	pBBR1 /medium	(http://seva.cnb.csic.es)
pSEVA312S	2.2	Chloramphenicol	*E. coli* "pir" strains	R6K /low	(http://seva.cnb.csic.es) (Martínez-García and de Lorenzo, 2011)
pSEVA428S	6.5	Streptomycin	Broad- host	RK2/low	(http://seva.cnb.csic.es)
pSEVA331	2.9	Chloramphenicol	Broad- host	pBBR1/medium	(http://seva.cnb.csic.es)
pSEVA227M_p14g	4.6	Kanamycin	Broad- host	RK2/low	provided by Gonzalo Durante
pSEVA237_p14g	3.9	Kanamycin	Broad- host	RK2/low	provided by Gonzalo Durante

In order to select the clones of interest, in both *E. coli* and *Labrenzia* sp. PHM005, a few antibiotics were employed: chloramphenicol at 34 µg/mL or 5 µg/mL, respectively (Sigma-Aldrich); kanamycin at 50 µg/mL (Roche); and streptomycin at 50 µg/mL (Sigma-Aldrich).

3. DNA isolation

Genomic DNA of *Labrenzia* sp. PHM005 for the routine PCR amplifications and generation of fragments for cloning was isolated with the following procedure: The cell biomass was collected from the agar plate or by centrifuging 10 mL of an overnight culture and harvesting cell pellet. The cells were resuspended in 400 µL of 10 mM Tris-EDTA (pH 8.0) (TE) and incubated at 80 ºC during 20 min. After cooling down at room temperature, 50 µL of lysozyme (10 mg/mL) was

added, mixed gently and incubated 1-12 h at 37 ºC. Afterwards, 75 µL of SDS containing 10% of proteinase K (10 mg/mL) was added and incubated at 65 ºC for 10 min. Subsequently, the 100 µL of 5 M NaCl and 100 µL of CTAB/NaCl, previously warmed up at 65 ºC, was added and incubated at 65 ºC for another 10 min. The addition of 750 µL of chloroform/isoamyl alcohol (24:1) was followed by vortex mixing and centrifuging for 5 min at 14000 x *g*. The aqueous phase was transferred by pipette to an Eppendorf tube and the equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) was added, mixed by vortex and centrifuged for 5 min at 14000 x *g*. The aqueous phase was transferred to a new Eppendorf tube and precipitated by isopropanol. It was further maintained at room temperature minimally for 30 min. Finally, the DNA was centrifuged for 15 min at 14000 x *g* at 4 ºC, washed with 200 µL of 70 % ethanol and centrifuged again for 2 min at 14000 x *g* at 4 ºC. The ethanol was removed carefully by pipetting and the remains were air dried. The precipitated DNA was dissolved in 40-100 µL of TE.

Genomic DNA extractions for genome sequencing purposes of *Labrenzia* sp. PHM005, *Stappia indica* PHM037 and *Labrenzia aggregata* PHM038 were performed using the Blood & Cell Culture DNA Mini Kit (Qiagen) to obtain the DNA with higher purity from 2 mL of an overnight bacterial culture.

Mini-preparation of plasmid DNA was done using the High Pure Plasmid Isolation Kit (Roche Diagnostics) for overnight culture volumes from 2-4 mL.

DNA concentration was measured using NanoPhotometer (Implen).

3.1. Whole genome sequencing

The genome of *Labrenzia* sp. PHM005 was sequenced by a PacBio RSII sequencer and was assembled *de novo* in a single contig. Samples were prepared at PharmaMar and library construction, sequencing, data processing and assembly were provided by Macrogen. The depth of the sequencing was 131.

The genome of strain PHM005 was re-sequenced using an Illumina MiSeq system with a 300 nt pair-end library to determine the presence of plasmids in the strain PHM005 that cannot be detected by the PacBio sequencing method as well as to correct other putative PacBio sequencing

errors. The library was constructed following the manufacturer's recommendations (Nextera DNA Flex, Illumina). The reads were trimmed and assembled into contigs using CLC Genomics Workbench software package (Quiagen). The service of the genome sequencing and assembly was provided at the Complutense University of Madrid, Department of Genomics.

Library construction, sequencing, data processing and assembly of genomes of strains PHM037 and PHM038 were provided by MicrobesNG (https://microbesng.com): genomic DNA libraries were prepared using Nextera XT Library Prep Kit (Illumina, San Diego, USA) following the manufacturer's protocol with the following modifications: 2 ng of DNA instead of 1 ng were used as input, and PCR elongation time was increased to 1 min from 30 s. DNA quantification and library preparation were carried out on a Hamilton Microlab STAR automated liquid handling system. Pooled libraries were quantified using the Kapa Biosystems Library Quantification Kit for Illumina on a Roche light cycler 96 qPCR machine. The genomes were sequenced by Illumina HiSeq technology using a 250 bp paired end protocol (https://microbesng.com). The depth of the sequencing was 30x.

3.2. Sequencing of PCR fragments and plasmid constructions

All cloned DNA fragments were confirmed by Sanger sequencing provided by Secugen S.L. DNA sequencing was done with fluorescently labeled dideoxynucleotide terminators and AmpliTaq FS DNA polymerase (Applied Biosystems Inc., USA) in an ABI Prism 377 automated DNA sequencer (Applied Biosystems Inc., USA).

4. Bioinformatics analysis

Sequence searches by homology was routinely performed using BLAST algorithms (all the variants) in the different databases of the National Center for Biotechnology Information (NCBI), (http://www.ncbi.nlm.nih.gov) and HMMER (http://www.ebi.ac.uk/Tools/hmmer) (Potter et al., 2018). Genome mining for secondary metabolites gene clusters and analysis was performed using online platform antiSMASH 4.0 and updated version antiSMASH 5.0 (Blin et al., 2019). Fully automated Rapid Annotation Subsystem Technology (RAST) Server was used for the genome annotation and gene categorization for sequence analysis before publishing the genome (Aziz et al., 2008). SWISS-MODEL Server (https://swissmodel.expasy.org/) was used for modelling and the

analysis of the thioesterase aa sequence. The sequence alignments, genome and plasmid maps visualizations and cloning simulations were performed using software Geneious versions 10.0.2 and Prime 2020.0.4.

4.1. Genome assembly and annotation of *Labrenzia* sp. PHM005

The genome of strain PHM005 was assembled in one contig by Macrogen. Nevertheless, the PacBio technology had some sequencing errors resulting in incorrect ORFs recognition. *Trans*-AT PKS/NRPS cluster, producer of labrenzin, was manually curated for the mistakes of the structural and functional annotations. The errors were corrected by Sanger sequencing of specific PCR amplicons surrounding the sequencing errors. Additionally, to obtain a high quality genome sequence, the more precise Illumina sequencing technology was applied. The obtained reads were mapped to a PacBio generated contig by Geneious v.10.0.2 and the errors were manually curated.

Illumina sequencing determined the presence of a plasmid, which was assembled in one contig and circularized by overlapping the contig extremities. The estimation of the plasmid copy number per cell was done by dividing the total chromosome and plasmid size by the number of Illumina reads covering them, and by the difference between the resulting numbers. The circular chromosome of strain PHM005 and the plasmid p1BIR were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) and deposited in the NCBI database with the accession numbers CP041191 and CP041190, respectively.

4.2. Whole genome comparison

Whole genome alignment of *Labrenzia* sp. PHM005, *L. alexandrii* DFL-11, *L. aggregata* RMAR6-6, *Labrenzia* sp. CP4 and *Labrenzia* sp. VG12 was performed using the progressive Mauve algorithm with default settings from the Geneious v.10.0.2. In addition, each strain was aligned separately to *Labrenzia* sp. PHM005 to calculate the minimum weight for Locally Collinear Blocks (LCB). Average nucleotide identity (ANI) was calculated using reciprocal best hits (two-way ANI) between two genomic datasets in an online tool developed at Kostas lab (Rodriguez-R and Konstantinidis, 2016).

4.3. PKS/NRPS domain analysis

AntiSMASH v.4.0 and 5.0 was used for a rough secondary metabolites cluster mining and the identification of the individual PKS/NRPS domains. The specificity of NRPS adenylation domains and *trans*-ATs was predicted based on a combination of tools implemented in antiSMASH (Stachelhaus code, NRPS Predictor3, pHMM, SANDPUMA and Minowa). The domains were manually revised for conserved motifs. PKSs and NRPSs domains and their substrate specificities were also analyzed by an online tool TransATor (https://transator.ethz.ch/) that predicts the core structure of a *trans*-AT PKS derived polyketide (Helfrich et al., 2019).

4.4. Delimiting the labrenzin gene cluster

To establish gene cluster boundaries, SEED Viewer read-only browser on the RAST server was used for rapid gene comparison and screening for the common genes between strains. Additionally, the proximal DNA regions from each strain were aligned using MAFFT package in Geneious v.10.0.2 setting the parameters for the scoring matrix BLOSUM62 and the gap open penalty to 1.

4.5. Phylogenetic analysis of the PKS/NRPSs Lab13 and homologs

The phylogenetic analysis of the third PKS/NRPS in the *lab* cluster was performed using a Geneious Tree Builder. The Lab13 closest homologues were collected from the NCBI database by BLASTp search. Only the whole protein sequences were included in the analysis. The sequences were previously aligned by Clustal Omega build in Geneious. The Jukes-Cantor model was used as a genetic distance model, tree building method was Neighbour-Joining. For tree resampling a bootstrap of 1000 replicates was used. Subtrees were separately aligned and the trees were constructed using Neighbour-Joining method in MEGA 7 (Kumar et al., 2016).

5. PCR amplification and purification of DNA fragments

5.1. Design of the oligonucleotides

To design oligonucleotides for different applications (Tables 4-10), the *Labrenzia* sp. PHM005 genome was used as DNA template. For optimal primer design several parameters were

considered like CG content, annealing temperature, the formation of hairpins or dimers and it was done using Primer 3 tool build in Geneious.

Table 4. Oligonucleotides for the PacBio sequencing error correction

Name	Sequence (5'→ 3')	Application and specifications
A F	CTTCCGGATACCGGCCAATC	PCR amplification of ORF4
A R	AAGGTGTCTGGGAGTCCAGT	
B F	AATCTTGGGTCTGATCGCGG	
B R	GCGCACATCGGAAAAGAACA	
C F	AGAGTGATTTCCAACCGCCC	PCR amplification of ORF15
C R	ATTGGACGAGTCTGGGCAAC	
D F	AATTTCATCCAGGGTGCCGT	
D R	ATGCAGTAAACCCCGCCATT	
E F	TTTCCGGGCGAAACTGTCTA	
E R	GTGATCTGGCTTTGGAGGGT	
F F	GCGGTTCCTGCAACAGAATG	
F R	TATGCAGTGATCCGCCAGAC	

Table 5. Oligonucleotides for the RT-PCR

Name	Sequence (5'→ 3')	Application and specifications
4 F	CGGAAGCGCCTGTAATCATG	RT-PCR of *lab4*
4 R	GGGGTTCGGCTGACTATCAT	
5 F	GACGGGTGGATCTTGGCTTT	RT-PCR of *lab5*
5 R	TTATCAGCAACCAGACCGCC	
6 F	CGATCTGGGCCGATATGCTT	RT-PCR of *lab6*
6 R	CCTGGAGCGTTTTGAGAGGT	
7 F	ATCAGATTAACCGGGCGTCC	RT-PCR of *lab7*
7 R	GAGGCGATGCGAAATAAGGC	
8 F	CAAAACTGTGCTGAGCCTGC	RT-PCR of *lab8*
8 R	CGGGTGTTTGCCAAAGTGTC	
9 F	GAGCACCATCAACATAGCGC	RT-PCR of *lab9*
9 R	GGCTGAAACTCTGATTGCCG	
10 F	TCTGAAATGACTGAGGCCGC	RT-PCR of *lab10*
10 R	GGCATTGTATTCGCCCACAC	
HK F	CACCACGACAAAAGAACCCG	RT-PCR of housekeeping gene *rpoD*
HK R	GAGAACCTGGGCTTCGTTGA	

Table 6. Oligonucleotides for the gene replacement constructions

Name	Sequence (5'→ 3')	Application and specifications
3 UP F	CTGCAGCTTCCATCGACAGGGTGATGACATC	Δ*lab3*; *Pst*I underlined
3 UP R	ATCCCGAGCTCGCCTGTTTGATTTCAAAGAGCCGG	Δ*lab3*; *Sac*I underlined
3 DOWN F	ATCCCGAGCTCCTGTTCTGTTGTCACTGACAGGAG	Δ*lab3*; *Sac*I underlined
3 DOWN R	ACTAGTACCTCAGAGCTTCGGTAGGATTTG	Δ*lab3*; *Spe*I underlined
PKS4 UP F	ACTAGTCTAGACGCAGTCGTCCTGATGAGAT	Δ*lab4*; *Xba*I underlined

Name	Sequence (5'→ 3')	Application and specifications
PKS4 UP R	ATCCCGAGCTCATGGGAACGTCCAGTATCGC	Δ*lab4*; *Sac*I underlined
PKS4 DOWN F	ATCCCGAGCTCGGTCGTATTCAACACCCGGT	Δ*lab4*; *Sac*I underlined
PKS4 DOWN R	TACCCAAGCTTAGGATCCTGCAAGAAGCCAC	Δ*lab4*; *Hin*dIII underlined
5-6 UP F	CTGCAGGACCACAATACGCCGTTACAGAAC	Δ*lab5*; Δ*lab5-6*; *Pst*I underlined
5-6 UP R	ATCCCGAGCTCCATTCCACAATCTCCGCTGGATG	Δ*lab5*;Δ*lab5-6*; *Sac*I underlined
5 DOWN F	GAGCTCCAGTACATCCAAAGCAGCCTGAAC	Δ*lab5*; *Sac*I underlined
5 DOWN R	ACTAGTTTCAAGTCCTCCGGGGCTTTATAG	Δ*lab5*; *Spe*I underlined
6 UP F	CTGCAGCAGTACATCCAAAGCAGCCTGAAC	Δ*lab6*; *Pst*I underlined
6 UP R	GAGCTCTTCAAGTCCTCCGGGGCTTTATAG	Δ*lab6*; *Sac*I underlined
5-6 DOWN F	ATCCCGAGCTCAGTACTTCGGATGCCTATCGCTAG	Δ*lab6*;Δ*lab5-6*; *Sac*I underlined
5-6 DOWN R	ACTAGTGGATCGGGCTAAAGTTACTGTTGG	Δ*lab6*;Δ*lab5-6*; *Spe*I underlined
P450-OX UP F	ACTAGTCTAGAACCTCTCAAAACGCTCCAGG	Δ*lab7-8*; *Xba*I underlined
P450-OX UP R	ATCCCGAGCTCCCTGCTTTCTGCCTTTGCTG	Δ*lab7-8; Sac*I underlined
P450-OX DOWN F	ATCCCGAGCTCATCCCCTTTACCATGGCACC	Δ*lab7-8; Sac*I underlined
P450-OX DOWN R	TACCCAAGCTTGGCCTGGACAATTCAGTCGA	Δ*lab7-8*; *Hin*dIII underlined
PKS13 UP2 F	ACTAGTCTAGAGATCCAGCCGGTTCATTTCCAGACC	Δ*lab13*; *Xba*I underlined
PKS13 UP2 R	ATCCCGAGCTCGCAGGCCTATGTCACAGTATCGGAC	Δ*lab13*; *Sac*I underlined
PKS13 DOWN F	ATCCCGAGCTCCGCCGCCATATTTGACCTTG	Δ*lab13*; *Sac*I underlined
PKS13 DOWN R	TACCCAAGCTTCTTCCGACCGTTTTGACACG	Δ*lab13*; *Hin*dIII underlined

Table 7. Oligonucleotides for the screening of *Labrenzia* sp. PHM005 mutants

Name	Sequence (5'→ 3')	Mutant
M1 F	TCTTGGTGGACGAGACCAGT	Δ*lab4*
M1 R	GGCTTCGTGTGGTGAAATGC	
M2 F	ATACTGAATCTCTGGATCAGGCCG	Δ*lab7-8*
M2 R	CATGTTGGGGACACATTCGATCAG	
M5 F	GTCAGTGAGGCCAATTCGC	Δ*lab13*
M5 R	TGAACGCCGGATCGAAGC	
M13 intern F	CTGGTCTGTAAACGTGTTGGGATC	Δ*lab13*
M13 intern R	CGAGGAGTTGGCACGGTATTTAC	
M6 F	GAGACATGATTGTGCATATCCGGC	Δ*lab16*
M6 R	GGCATGATCCGGATAGACGATACTC	
M7 F	CCAGTGCGATGGCAAGAAAGTATC	Δ*lab11-12*
M7 R	TTGATCGATTCTTATCCCCCTGCG	
M8a F	CTTCAGTGATGAGACCTTGGGACAG	Δ*lab9*
M8a R	CAAACCGTCTTCGAAGTTGACCAC	
M9 F	CCACTGAAGTAGTTCCTCCCGAAG	Δ*lab5-6; Δlab5; Δlab6*
M9 R	GGACGAAGAGGAGACACTGACAAA	
5 DOWN R	ACTAGTTTCAAGTCCTCCGGGGCTTTATAG	Δ*lab5*
6 UP F	CTGCAGCAGTACATCCAAAGCAGCCTGAAC	Δ*lab6*
ACP-HMGS F	ACTAGTCTAGACTGCTGTTCTCACCACTCGT	Δ*lab3*
ACP-HMGS R	TACCCAAGCTTAAGCCTATCGGAGCAGAGGA	

Table 8. Oligonucleotide for the overexpression or complementation of genes from *Labrenzia* sp. PHM005, a producer of labrenzin

Name	Sequence (5'→ 3')	Application and specifications
P1 F	ACCATGCCTAGGCC**TTATAA**GTGACTGGCTG CCAGAGCAAAAGAGC	Amplification of *lab4-lab5-lab6* operon region by Gibson assembly; *Psi*I in bold, overlapping region underlined
P1 R	GACTCTAGAGGAT**GTTTAAAC**TTAGCGAGGT TTAGATGCGCCCAGGATCG	Amplification of *lab4-lab5-lab6* operon region by Gibson assembly; *Pme*I in bold, overlapping region underlined
VP F	**GTTTAAAC**ATCCTCTAGAGTCGACCTGCAGG	Amplification of plasmid pSEVA238 by Gibson assembly; *Pme*I in bold, overlapping region underlined
VP R	**TTATAA**GGCCTAGGCATGGTCATGACTC	Amplification of plasmid pSEVA238 by Gibson assembly; *Psi*I in bold, overlapping region underlined
PnPKS4 F	ACTTACCATGCCGCAATACG	Amplification of Pn4 for cloning into pSEVA238_*Psi*I_ *lab4-lab5-lab6_Pme*I
PnPKS4 R	AGCAACACCTCAGAGCTTCG	Amplification of Pn4 for cloning to pSEVA238_*Psi*I_ *lab4-lab5-lab6_Pme*I
MTs F	TGGCTATCTCTAGTAAGGCCTACCC	Verification of cloned methyltransferases by PCR amplification and sequencing
MTs R	CGTAATATCCAGCTGAACGGTCTGG	Verification of cloned methyltransferases by PCR amplification and sequencing
ACP F	ACTAGTCAACGATTAGACTTTGGATGCTGCC	Amplification of gene *lab2* (acyl carrier protein); *Spe*I underlined
ACP R	GAGCTCCCGCTGATTTCAATTGCGCC	Amplification of gene *lab2* (acyl carrier protein); *Sac*I underlined
HMGS F	CGGAATTCCGGATGCAGGGAGAGTGTCACG	Amplification of gene *lab3* (hydroxymethylglutaryl-CoA synthase); *Eco*RI underlined
HMGS R	CGGGATCCCGAAGCCTATCGGAGCAGAGGA	Amplification of gene *lab3* (hydroxymethylglutaryl-CoA synthase); *Bam*HI underlined
OR5 F	GAGCTCACGGGAAGCGGCATGAGC	Amplification of gene *lab5* (oxidoreductase); *Sac*I underlined
OR5 R	ACTAGTGAAAGAAAAATGTGTGCTCGGCCG	Amplification of gene *lab5* (oxidoreductase); *Spe*I underlined
OR5-6 R	ACTAGTCCTACCTAGCGATAGGCATCC	Amplification of genes *lab5* and *lab6* (oxidoreductase and methyltransferase); *Spe*I underlined
FMN-OR F	ACTAGTCCACCAGTCCGTATCTGTTTCCTCC	Amplification of gene *lab8* (oxidoreductase); *Spe*I underlined

Name	Sequence (5'→ 3')	Application and specifications
FMN-OR R	GAGCTCGGGGCTGATACGCAAAACCTGAG	Amplification of gene *lab8* (oxidoreductase); *Sac*I underlined
P450-OR F	ACTAGTGATCCTGGGCGCATCTAAACCTC	Amplification of gene *lab7* (cytochrome P450); *Spe*I underlined
P450-OR R	GAGCTCTCAAACGGCTTCTTATCTGGGAGC	Amplification of gene *lab7* (cytochrome P450); *Sac*I underlined
AT F	ACTAGTTCACCCCATAGTCATCCCGAAAAG	Amplification of gene *lab9* (acyltransferase); *Spe*I underlined
AT R	GAGCTCAAGTCATTCAACGGACTCGCC	Amplification of gene *lab9* (acyltransferase); *Sac*I underlined
AT9 F	GAGCTCTCACCCCATAGTCATCCCGAAAAG	Amplification of genes *lab9* and *lab10* (acyltransferase); *Sac*I underlined
AT10 R	GAGCTCATTCCCTTGGGAACTCTAGAGCCAG	Amplification of genes *lab9* and *lab10* (acyltransferase); *Sac*I underlined
ECH R	GAGCTCGACTAGGGAATAGGGATGTTGTCG	Amplification of gene *lab11* (enoyl-CoA-hydratase); *Sac*I underlined
KS F	ACTAGTCCCTAGTCAGAACTCAAAACCGTGG	Amplification of gene *lab12* (ketosynthase); *Spe*I underlined
KS R	GAGCTCGGTTTCTGGCAAATCTCCTGAACG	Amplification of gene *lab12* (ketosynthase); *Sac*I underlined
11-12 F	ACTAGTCTGGCTCTAGAGTTCCCAAGGGAAT	Amplification of genes *lab11* and *lab12* (enoyl-CoA-hydratase and ketosynthase); *Spe*I underlined
Pn PacI F	TTAATTAAACTTACCATGCCGCAATACG	Amplification of Pn4 for cloning to pSEVA227M_*p14g; Pac*I underlined
Pn AvrII R	CCTAGGAGCAACACCTCAGAGCTTCG	Amplification of Pn4 for cloning to pSEVA227M_*p14g; Pac*I underlined

Table 9. Oligonucleotides used for sequencing the *lab4-lab5-lab6* operon

Name	Sequence (5'→ 3')
16kb R1	GTCCATTCGGGTCGATTGC
16kb R2	CCGCATTATGCCAGGCTGC
16kb R3	CAGGTCTGCCCGTGAAGAGAC
seq1 F	ACAGTGGCAGACCCGGTACTGC
seq2 F	GCCTATGGTGGCTTCTTGC
seq3 F	CAAGAGTGGTTGAAGAAGCAGC
seq4 F	CACGTTCTCTGGAGCAGTGC
seq5 F	ACCCCGACGCCCTTGAAAGG
seq6 F	GGAAATCGAGCGGCACGGAC

Name	Sequence (5'→ 3')
seq7 F	TTGCCCCAAGTGCACCCTTCC
seq8 F	AATCCGCAAGCTGCGGCTGC
seq9 F	AGATGAACTGGCACGGAGTG
seq10 F	GCTATCTTGAGGCGCATGGC
seq11 F	CGTGTCCGATAAGGCTCAGG
seq13 F	GCCCGCTTTACCGATGAGC
seq14 F	GGCACCAGTACGATCCGC
seq15 F	AATCCATGTCGCGATGAGC
seq16 F	TGGTCCGGACGCTCCACAG
seq17 F	GTCTGACGATCCGAAGTGG
seq18 F	CCGCTCTGACTGAAGTGAG
seq19 F	CATGCCTGCGTCAGCGATGC
seq20 F	GAGGCCTTGGACATGAATACCC
seq21 F	GTCACCATCCCGCTACACC
4F	CGGAAGCGCCTGTAATCATG

Table 10. Other oligonucleotides employed in this work

Name	Sequence (5'→ 3')	Application and specifications
F24	CGCCAGGGTTTTCCCAGTCACGAC	Universal oligonucleotides for sequencing inserts in pSEVA plasmids
R24	AGCGGATAACAATTTCACACAGGA	
M13 F	GTAAAACGACGGCCAG	Universal oligonucleotides for sequencing inserts in pCR™Blunt II-TOPO®
M13 R	CAGGAAACAGCTATGAC	

5.2. PCR conditions

5.2.1. PCR amplification of DNA fragments for cloning

In order to obtain blunt end DNA fragments for cloning purposes, the Phusion® High-Fidelity DNA Polymerase (New England Biolabs) was used. The rest of the PCR reaction mix components were added at the concentrations according to the manufacturer protocol (genomic DNA, oligonucleotides, DMSO, dNTPs and water). The final volume of the PCR reaction was 50 µL. The PCR reaction performed in the thermocycler Mastercycler Gradient (Eppendorf) had the following program: initial denaturalization for 2 min at 98 ºC; 30 cycles comprising denaturalization of 10 s at 98 ºC, primer annealing for 15 s at 57-60 ºC (depending on the primer Tm) and extension for 15 s per kb of DNA at 72 ºC; the final extension for 2 min at 72 ºC; and the cooling at 4 ºC.

5.2.2. Long fragment PCR amplification

The high fidelity PrimeSTAR® GXL DNA Polymerase (TaKaRa) was used for a long DNA fragment amplification. Other PCR reaction mix components were added at the concentrations according to the manufacturer protocol (genomic DNA, oligonucleotides, DMSO, dNTPs and water). The final volume of the PCR reaction was 50 µL. The PCR reaction had the following program: initial denaturalization for 5 min at 98 ºC; 30 cycles comprising denaturalization of 10 s at 98 ºC, primer annealing for 15 s at 58 ºC and extension of DNA at 68 ºC (min/kb); the final extension for 15 min at 68 ºC; and the cooling at 4 ºC.

5.2.3. Colony PCR for screening and verification of mutants

NZYTaq II 2× Green Master Mix (NZYTech) was used to perform colony PCR for screening purposes. The oligonucleotides, water and DNA template (single colony picked from a plate) and a Master Mix were added to the PCR reaction tube according to the manufacturer protocol. The total volume was 15 µL. The program for the PCR amplification was: initial denaturalization for 10 min at 95 ºC; 30 cycles comprising denaturalization of 30 s at 94 ºC, primer annealing for 30 s at 55-60 ºC (depending on the primer Tm) and extension for 15 s per kb of DNA at 72 ºC; the final extension for 2 min at 72 ºC; and the cooling at 4 ºC.

5.3. Purification of PCR fragments

The DNA fragments obtained by PCR amplification were subsequently purified using QIAquick® PCR Purification Kit (Quiagen) or purified from the gel using QIAquick® Gel Extraction Kit (Quiagen) according to manufacturer instructions.

6. DNA *in vitro* manipulations

6.1. TOPO cloning

One-step "TOPO® Cloning" strategy was used to clone blunt-end PCR amplification products, obtained previously using high fidelity polymerase, in a pCR™Blunt II-TOPO® vector provided by the Zero Blunt® TOPO® PCR Cloning Kit (Invitrogen). The vector had pUC origin of

replication, thus allowing high copies inside the cell which made it a great tool for storing DNA fragments for further restriction cloning and sequencing.

6.2. Restriction cloning

The enzymatic digestions of plasmid DNA were performed using restriction enzymes and buffers CutSmart or NEB 3.1. from New England BioLabs following the specific manufacturer instructions. When needed, the Shrimp Alkaline Phosphatase (rSAP) (New England BioLabs) was employed following the manufacturer recommendations. The phosphatase dephosphorylated the 5' ends of the DNA thus preventing the empty vector re-ligation. When non-phosphorylated PCR products were used as inserts, T4 Polynucleotide Kinase (New England BioLabs) was used for 5' phosphorylation of DNA for subsequent ligation.

The DNA fragments were resolved by gel electrophoresis in TAE buffer (0.04 M Tris-acetate, 1 mM EDTA, pH 8.0). Likewise, gels were prepared using TAE and 0.7% or 1.5% agarose (Condalab). The samples carried the loading dye (0.25% bromophenol blue, 0.25% xylene cyanol, 15% Ficol type 400). Quick-Load® DNA ladder (New England BioLabs) was used for the fragment size estimation. Electrophoresis was carried out at 100-135 V for 15-30 min in a Mupid®-One electrophoresis system (NIPPON Genetics). Afterwards, the gel was dyed by GelRed® Nucleic Acid Gel Stain (Biotium) for 3-5 min and observed under the UV light (285nm) supplied by the system GelDoc® (BioRad). Subsequent ligation steps required the purification using QIAquick® Gel Extraction Kit (Quiagen) or QIAquick® PCR Purification Kit (Quiagen).

The purified vectors and their respective inserts, previously obtained by PCR or enzymatic digestion of pCR™Blunt II-TOPO® vector carrying the inserts, were ligated using the T4 DNA ligase (New England BioLabs) applying the buffer according to the recommendations of the manufacturer and employing vector:insert molar ratio 1:3. Prior to ligation, the concentration of the vector and the insert was measured spectroscopically by NanoPhotometer (Implen) and/or visualized by gel electrophoresis. The ligation was performed overnight at 16 ºC or 1 h at room temperature. The control mixtures contained water instead of the insert.

6.3. Gibson cloning

Gibson Assembly® Cloning Kit (New England BioLabs) was used to clone long DNA fragments into a desired vector. The procedure was designed to join multiple overlapping DNA fragments, regardless of length, in a single-step reaction. Firstly, the both vector and the long DNA fragment DNA, were amplified by PCR. Following the manufacturer recommendations regarding volumes, the non-purified PCR blunt-end products were mixed together with Gibson Assembly Master Mix (2X) and water. The assembly mixture was incubated at 50 ºC in a Mastercycler Gradient (Eppendorf) for 1 hour. Thermo-competent cells were transformed directly with the assembly mixture upon the completion of the incubation.

7. Plasmid constructions

All the PCR fragments employed in this work were previously cloned in the pCR™Blunt II-TOPO® vector and verified by Sanger sequencing. The plasmid constructions created throughout the course of these studies were analyzed via enzyme digestions or sequencing. Design and visualization of each construction was supported by Geneious software versions.

7.1. Constructions for the genetic replacements

All the plasmids used for here described constructions of site-directed mutations, gene expressions and promotor fusions, as well as specificities of each plasmid are listed in Table 3.

7.1.1. Gene knock-out constructions

Figure 10 shows the scheme for the suicide plasmid constructions necessary for the generation of gene knockout mutants. The upstream and downstream fragments flanking a desired deletion site in a gene of interest were amplified by a PCR (Table 6), purified and ligated using Instant Sticky-end Ligase Master Mix (New England BioLabs) and equal volumes of each fragment following the manufacturer instructions. After incubation at room temperature for 5 min, 1 µL of the ligation mixture was used as a template for the PCR amplification to generate a single UP-DOWN DNA fragment comprising upstream and downstream region connected by the bridging restriction site. The PCR product was purified and cloned in the TOPO vector. The restriction

cloning method was employed to clone the generated UP-DOWN fragment into the suicide vector pSEVA312S.

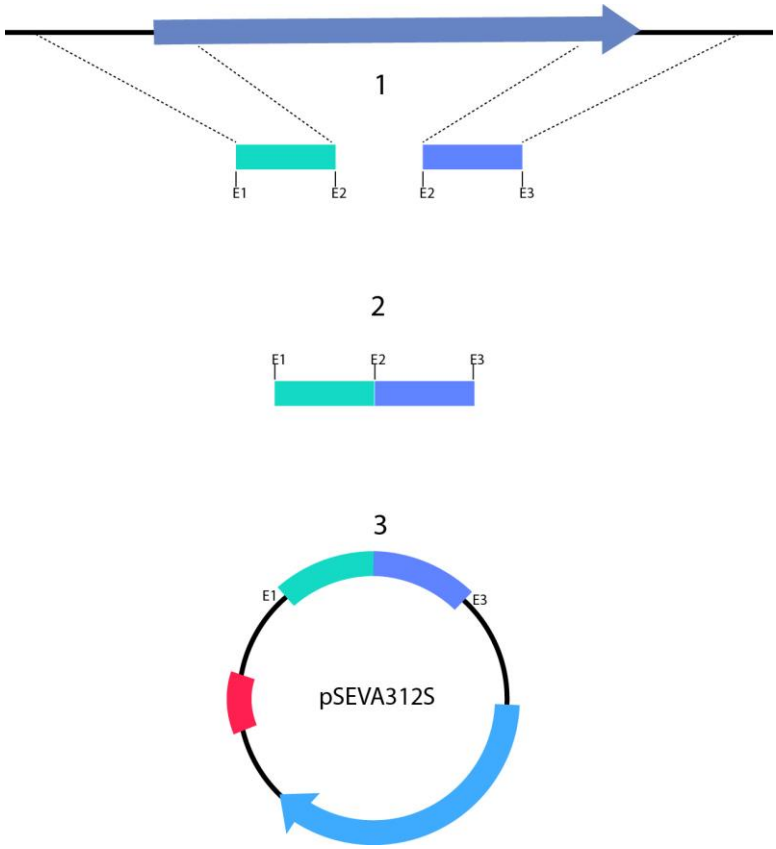


Figure 10. Scheme for the suicide plasmid pSEVA312S construction directed to gene knockouts.

Table 11 shows the upstream and downstream fragment sizes, regions deleted and the restriction enzymes used for each mutant. An exception is a double mutant Δ*lab*6Δ*lab*16 generated by repeating the described gene-knockout protocol for the *lab*16 deletion over already created Δ*lab*6 mutant. Constructions for mutant Δ*lab*2, Δ*lab*9, Δ*lab*16 and Δ*lab*11-12 were designed in Geneious and synthetized by GenScript.

Table 11. Mutant strains generated and their genotipic specifications

Mutant	Upstream (bp)	Downstream (bp)	Deletion (bp)	Enzymes
Δ*lab*3	580	527	563	*Pst*I-*Sac*I-*Spe*I
Δ*lab*4	763	552	763	*Pst*I-*Sac*I-*Spe*I
Δ*lab*5	502	559	481	*Pst*I-*Sac*I-*Spe*I
Δ*lab*6	559	527	787	*Pst*I-*Sac*I-*Spe*I
Δ*lab*5-6	502	527	1826	*Pst*I-*Sac*I-*Spe*I
Δ*lab*7-8	672	692	2477	*Pst*I-*Sac*I-*Spe*I

Mutant	Upstream (bp)	Downstream (bp)	Deletion (bp)	Enzymes
Δ*lab*9	541	512	539	*Pst*I-*Sac*I-*Spe*I
Δ*lab*11-12	556	653	1410	*Pst*I-*Spe*I
Δ*lab*13	578	603	15850	*Pst*I-*Sac*I-*Spe*I
Δ*lab*16	539	559	631	*Sac*II-*Sac*I-*Spe*I

7.1.2. Amino acid serine replacement with an alanine in a conserved thioesterase (TE) motif

In order to make a single point-mutation in a conserved catalytic site in a thioesterase domain at the end of the final PKS/NRPS module in Lab13, a serine from His-Ser-Asn triad was changed to alanine. Region of 819 bp in a thioesterase domain containing an active serine67 in a conserved motif GW**S**SG was modified by changing the nucleotides from the Ser codon in a way it created Ala codon and at the same time generated a *Nhe*I restriction site to facilitate the easy screening for mutants by PCR amplification and enzyme digestion. To do so, a 5'- GCTTGA -3' motif (Ser codon underlined) was changed to a 5'- GCTAGC -3' motif (*Nhe*I restriction site; Ala codon underlined). Digestion with *Nhe*I enzyme would cut the PCR fragment (819 bp) into two pieces (≈400 bp). The design for a modified DNA region flanked by *Pst*I and *Spe*I restriction sites was done using Geneious and synthesized by GenScript.

7.2. Constructions for gene complementation and overexpression

7.2.1. Gibson assembly of *lab4-lab5-lab6* operon

The 15751 bp long region comprising an operon *lab4-lab5-lab6* was cloned in plasmid pSEVA238 under a 3-methyl benzoate inducible promotor P_m, as the promotor is active in both exponential and stationary growth phase. The plasmid is comprised of a medium-copy BBR1 origin of replication, MCS, Pm expression system and kanamycin resistance. Both the long DNA fragment and the plasmid were amplified by PrimeSTAR® GXL DNA Polymerase (TaKaRa) followed by a Gibson cloning without the previous DNA purification steps. The oligonucleotides were designed following the instructions for the Gibson assembly protocol, but also introducing the new enzyme restriction sites for *Psi*I and *Pme*I that do not cut neither the plasmid nor the insert, allowing the use of those sites for additional cloning like changing the promoter or RBS. The resulting plasmid was

pSEVA238_*Psi*I_ *lab4-lab5-lab6_Pme*I (20776 bp) lacking the part of the MCS, including the restriction sites *Eco*RI, *Eco*53kI, *Sac*I, *Acc*65I, *Kpn*I, *Xma*I, *Tsp*MI, *Sma*I and *Bam*HI. The sequence of the operon was verified by Sanger sequencing using oligonucleotides listed in Table 9.

Additionally, the promoter P_m was swapped with the native promoter region of the operon *lab4-lab5-lab6* Pn4 (350 bp) by restriction cloning using restriction sites *Sna*BI and newly created *Psi*I to generate a plasmid pSEVA238_Pn4_*lab4-lab5-lab6_Pme*I.

7.2.2. Methyltransferases overexpression and complementation system

Plasmid pSEVA338 was used as an expression system for three methyltransferases MT6 (Lab6), MT16 (Lab16) and PedO. The artificial operon was designed as indicated in Figure 11, where each gene with the corresponding RBS was flanked by specific restriction sites (blunt cut) in a way that individual genes or the combination of genes could be easily generated by digestion and re-ligation. The operon was synthetized and cloned using *Sac*I and *Spe*I restriction sites by GenScript yielding plasmid pSEVA338_MTs (7699 bp). Other plasmids derived from pSEVA338_MTs were: pSEVA338_lab6 (5854 bp), pSEVA338_lab16 (5705 bp), pSEVA338_PedO (5880 bp), pSEVA338_lab6_lab16 (6691 bp), pSEVA338_PedO_lab6 (6856 bp) and pSEVA338_PedO_lab16 (6736 bp).

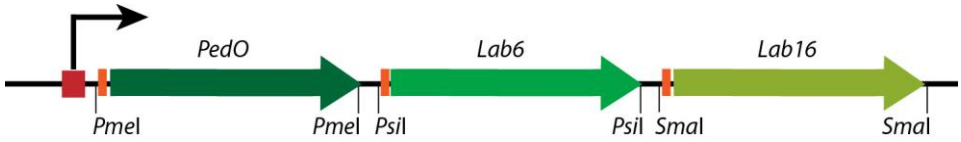


Figure 11. Scheme representing the design of the artificial operon comprising *PedO-Lab6-Lab16* methyltransferases.

7.2.3. Complementation and overexpression of other genes

For the purpose of complementing the created mutations, the respective functional genes were cloned in previously described plasmid pSEVA338 and pSEVA227M_p14g (or pSEVA237_p14g) (Table 3). Plasmid pSEVA338_GFP was previously constructed by cloning a GFP, excised from the pSEVA637 (http://seva.cnb.csic.es) by *Hin*dIII and *Spe*I. Table 12 shows all gene inserts and sizes, plasmids, restriction enzymes used for cloning and the final constructs.

Table 12. Gene expression constructs

Gene	Size (bp)	Cloning plasmid	Enzymes	Construct
HMGS; Lab3	1330	pSEVA338	*Eco*RI-*Bam*HI	pSEVA338_*Lab3*
P450; Lab7	1440	pSEVA338	*Sac*I-*Spe*I	pSEVA237_p14g_*Lab7*
OR; Lab8	1497	pSEVA338	*Sac*I-*Spe*I	pSEVA237_p14g_*Lab8*
P450-OR; Lab7-8	2797	pSEVA338	*Sac*I-*Spe*I	pSEVA237_p14g_*Lab7-8*
AT; Lab9	1089	pSEVA237_p14g	*Sac*I-*Spe*I	pSEVA237_p14g_*Lab9*
ATs; Lab9-10	2382	pSEVA331	*Sac*I	pSEVA331_*Lab9-10*
ECH; Lab11	759	pSEVA338	*Bam*HI-*Spe*I	pSEVA237_p14g_*Lab11*
KS; Lab12	1280	pSEVA338	*Bam*HI-*Spe*I	pSEVA237_p14g_*Lab12*
ECH-KS; Lab11-12	2031	pSEVA338	*Bam*HI-*Spe*I	pSEVA237_p14g_*Lab11-12*

7.3. Fusion of the P_{n4} promoter to a GFP encoding gene in plasmid pSEVA227M_p14g

For the purpose of transcription regulation assays using different elicitors, a strong promoter p14g in plasmid pSEVA227M_p14g was replaced by weak P_{n4} promoter using restriction enzymes *Pac*I and *Avr*II generating a new construct pSEVA227M_Pn4 (4930 bp).

8. Plasmid DNA transformation and construction of mutants

8.1. Preparation and transformation of chemically competent cells

E. coli strains were regularly used for plasmid transformation. To prepare competent cells an overnight pre-inoculum in LB was diluted to approximately 0.05 OD_{600} in a fresh 100 mL LB. The culture was further incubated at 37 ºC and shaking at 220 rpm until it reached 0.48-0.60 OD_{600}. Subsequently, the culture was cooled on ice and centrifuged at 4000 rpm and 4 ºC for 7 min using ultracentrifuge Sorvall Lynx 6000 (Thermo Scientific). The supernatant was discarded and the cells were resuspended carefully in 30 mL of a sterile TBFI solution (100 mM RbCl, 50 mM $MnCl_2$, 30 mM KOAc, 10 mM $CaCl_2$, 15% glycerol, water; pH 5.8) by manually shaking the tubes on ice. The cells were incubated on ice for the next 2 h. Upon the completion of cell incubation, the centrifugation was repeated as before. The supernatant was discarded and the cells were resuspended carefully in 4 mL of a sterile TBFII solution (10 mM MOPS, 10 mM RbCl, 75 mM $CaCl_2$, 15% glycerol, water; pH 6.8) and 200 µL was aliquoted to a sterile 1.5 mL Eppendorf tube on a dry ice and ethanol to allow the instant freezing of the competent cells. The competent cells were stored at -80 ºC.

For plasmid transformation the competent cells were thawed on ice for 10 min and the plasmid DNA or ligation mix and respective controls were added and shortly mixed by pipetting; the cells were incubated on ice for 15-30 min and transferred to a water-bath at 37 ºC for 3 min; afterwards cells were immediately switched to ice for 2 min and recovered in 800 µL of fresh LB. The tubes were incubated at 37 ºC and shaking at 220 rpm for 1 h. Finally, the desired dilutions of transformed cells were spread on Petri plates with LB agar and appropriate antibiotic. The plates were incubated in a stove at 37 ºC.

8.2. Preparation and transformation of electro-competent cells

When required, *E. coli* MFD*pir* strain was used as a donor strain for conjugations and thus was necessary to previously prepare electro-competent cells and transform them with the plasmid DNA. 10 mL of an overnight LB culture was centrifuged in Multispeed Centrifuge CL31R (Thermo Scientific) at 3800 rpm at 4 ºC for 15 min. The supernatant was discarded and the cells were resuspended in 10 mL of sterile water, previously cooled at 4 ºC. The centrifugation was repeated for 10 min. The cells were washed again in 10 mL of sterile water and centrifuged for 8 min. The cells were resuspended for the third time in 1 mL of sterile water and transferred to a sterile 1.5 mL Eppendorf tube. The cells were centrifuged in Multispeed Centrifuge CL31R (Thermo Scientific) at 13000 rpm at 4 ºC for 3 min. Finally, the supernatant was discarded and the cells were resuspended in 100 µL. Then, 3 µL of plasmid was added in a tube with electro-competent cells and gently mixed by pipetting. The whole volume was transferred to a previously cooled 2 mm electroporation cuvette (Cell Projects) and treated by Gene Pulser (BioRad) under these conditions: 25 µF capacitance, 200 Ω resistance, 2.5 kV pulse. Upon the electroporation finalization, the cells were recovered immediately in a 900 µL of fresh LB supplemented with DAP and transferred to a sterile 1.5 Eppendorf tube. Subsequently, the cells were incubated at 37 ºC and shaking at 220 rpm for 1 h. Finally, the desired dilutions of transformed cells were spread on Petri plates with LB agar, DAP and appropriate antibiotic. The plates were incubated at 37 ºC in a stove.

8.3. Gene knockout by homologous recombination in *Labrenzia* sp. PHM005

8.3.1. Homologous recombination using *I-Sce*I endonuclease system

Throughout this work all the deletion mutants were developed based on the *I-Sce*I endonuclease system that induces the DNA repair by homologous recombination (Martínez-García and de Lorenzo, 2011). Nevertheless, the original protocol was optimized to fit the specificities of the strains used.

Figure 12 shows the simplified scheme of an endonuclease *I-Sce*I directed system followed for gene deletion constructions. Gene knockout was performed using two-vector system; (1) the suicide plasmid pSEVA312S carrying chloramphenicol resistance and the upstream and downstream DNA fragments flanking the desired deletion region; (2) the pSEVA428S carrying the *I-Sce*I endonuclease and the streptomycin resistance. The expression of *I-Sce*I endonuclease is controlled by Pm promoter inducible by 3-methylbenzoate. Once it is expressed, *I-Sce*I endonuclease digests the specific site in the suicide plasmid pSEVA312S integrated in the host genome causing double-strand brake and indirectly inducing the homologous recombination repair system. The survival cells manage to repair their genome, but loose the chloramphenicol resistance. The resulting mutant genotype was theoretically 50%. To generate the knockout of the desired gene the first conjugation was needed to integrate the suicide plasmid in the host genome by homologous recombination. During the second conjugation two things occur: the transformation of the plasmid carrying the *I-Sce*I endonuclease; and the expression of the endonuclease independently of the inducer due to the leaky promotor and observed high endonuclease efficiency. The recombination occurs during the second conjugation so that the plated bacteria can be screened for a mutation directly.

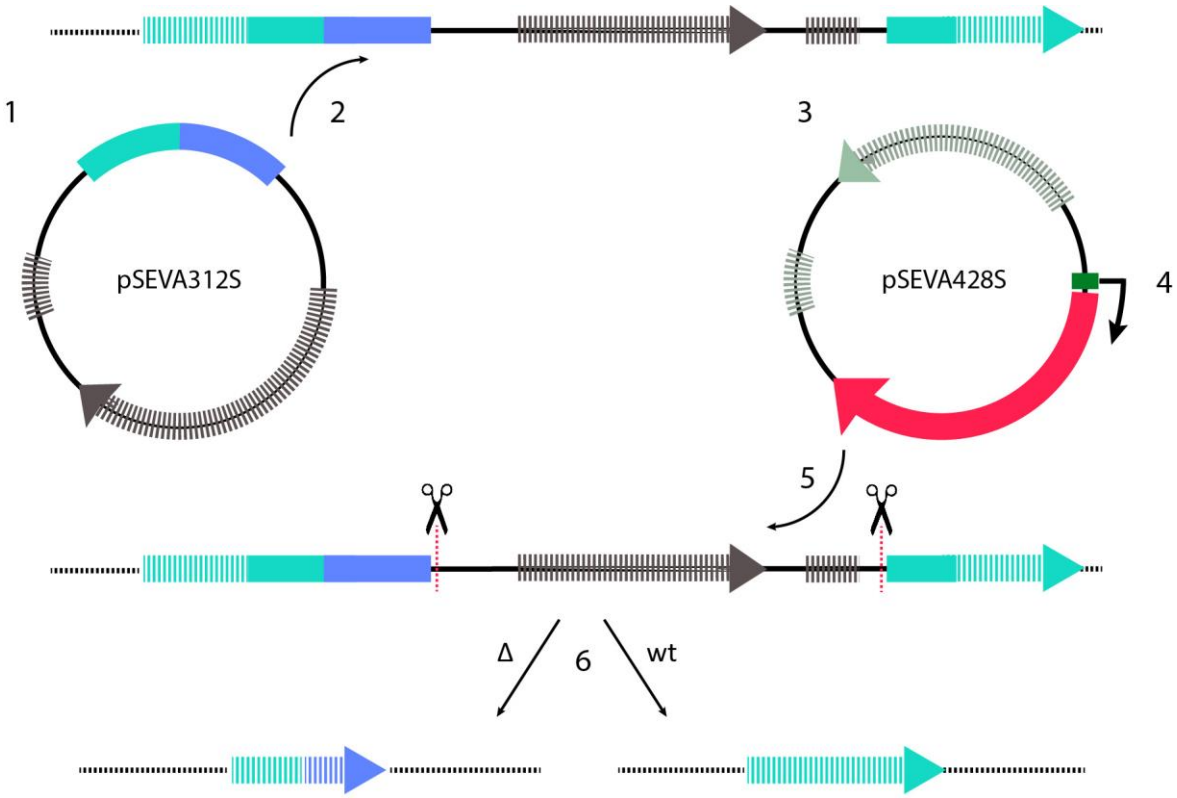


Figure 12. Scheme showing the endonuclease *I*-SceI induced homologous recombination repair for the creation of deletion mutants.

8.3.2. Triparental conjugation

Triparental conjugation was performed using *E. coli* cc118λpir carrying the suicide plasmid as a donor strain, *E. coli* HB101 harboring the plasmid pRK600 as a helper strain and *Labrenzia* sp. PHM005 as a receipt strain. 1 mL of overnight cultures was centrifuged at 13000 rpm for 1 min and the pellet was washed in a 500 µL of sterile 0.85% NaCl solution. The cells were centrifuged again and the pellet was resuspended to a final volume of 200 µL of NaCl solution. 50 µL of each strain were mixed and pipetted to a 0.22 µL filter disc placed on the MA plate. The plate was incubated overnight at 30 ºC. The next day the filter mating discs were collected in a 15 mL tube with 1 mL of a sterile 0.85 % NaCl solution and vortexed thoroughly to detach the cells from the filter. Afterwards, the 100 µL and the rest of the cells were plated on MA plates containing as selective antibiotics chloramphenicol (5 µg/mL) and kanamycin (10 µg/mL). The transconjugants T1 were selected after 5-7 days and verified by colony PCR. Subsequently, pSEVA428S was transformed to

PHM005 T1 by triparental conjugation and selected on the MA plates with selective antibiotics chloramphenicol (5 µg/mL), kanamycin (10 µg/mL) and streptomycin (50 µg/mL).

8.3.3. Biparental conjugation

In order to optimize the conjugation and selection procedure, the biparental conjugation was introduced using as a donor strain *E. coli* MFD*pir*, which is an auxotroph for diamine-salicylic acid (DAP). The conjugation procedure was as described previously, with the following exceptions:

First transformation was selected on MA containing as selective antibiotic chloramphenicol (5 µg/mL) and second recombination was selected on the MA plates with selective antibiotic streptomycin (50 µg/mL). The transconjugants T1 were selected on MA containing as selective antibiotic chloramphenicol (5 µg/mL); then, pSEVA428S was transformed to PHM005 T1 by biparental conjugation during 6 h; the mutation candidates were selected on the MA plates with selective antibiotic streptomycin (50 µg/mL) and verified by colony PCR.

9. RNA extraction and transcriptomics analyses

9.1. RNA extraction and quantitative analysis by RT-PCR

To isolate a total RNA from bacterial cells a High Pure RNA Isolation Kit (Roche) was used with a few modifications of the manufacturer protocol. A volume of 5 mL of a culture in an exponential growth phase was used for the extraction, centrifuged at 3800 rpm at 4 ºC for 10 min. The cell pellet was resuspended in 500 µL of TE buffer. To ensure improved cell lysis, 50 µL of lysozyme (50 mg/mL) (Sigma-Aldrich) was added to the cell suspension and incubated at 37 ºC for 10 min. Additionally, the cell suspension was instantly deep frozen in a dry ice and ethanol followed by a defrost in a water bath at 37 ºC in three series for 1-2 min. The rest of the protocol was followed as instructed by the manufacturer. The purified RNA was further incubated with Ambion® TURBO DNA-free™ DNase Treatment and Removal Reagents (Life Technologies) kit to remove the contaminating DNA from the RNA samples. The RNA concentration was measured using a NanoPhotometer (Implen).

Reverse transcription of 1 µg of DNA-free RNA per sample was done using the Transcriptor First Strand cDNA Synthesis Kit (Roche) to generate cDNA following the manufacturer protocol.

RT- PCR semi-quantitative analyses were carried out to check the mRNA levels. The *rpoD* gene was chosen as the housekeeping gene. Taq-polymerase and 10X Standard Reaction Buffer (Biotools) was used for the PCR amplification. Water, dNTPs and oligonucleotides were added as indicated in a manufacturer protocol. The PCR amplification was done using 1 µL of cDNA sample and program was: initial denaturalization for 5 min at 95 ºC; 25-30 cycles comprising denaturalization of 30 s at 95 ºC, primer annealing for 30 s at 57-60 ºC (depending on the primer Tm) and extension for 1 min per kb of DNA at 72 ºC; the final extension for 2 min at 72 ºC; and the cooling at 4 ºC. Relative mRNA expression was visualized in 1.5% agarose gel by gel electrophoresis. Negative controls contained RNA samples without the reverse transcriptase.

9.2. RNA-seq and 5'Cap RNA-seq analysis

The RNA extraction from *Labrenzia* sp. PHM005 for the transcriptomics analysis was performed at the early exponential phase (12 h) in a modified MBM+vit medium as described previously. Three sample replicates were used for the RNA-seq analysis. The construction of cDNA library, sequencing and bioinformatics analyses were carried out by Vertis Biotechnologie AG. The percentage of the mapped reads to all annotated genes was 78.4%, 80.8%, and 77.0% for each replicate sample.

Cappable-seq is a method (Ettwiller et al., 2016) for the precise identification of bacterial transcription start sites (TSSs) used in this work for TSSs determination in the genome of *Labrenzia* sp. PHM005. The sequencing and the bioinformatics analysis was performed by Vertis Biotechnologie AG using the same RNA samples described above.

10. Fluorescence measurement in GFP producing bacterial cultures

To measure the GFP fluorescence and optical density (OD_{600}), an automatic plate reader Varioskan Flash (Thermo Scientific) was used. The appropriate dilutions of the bacterial culture in MBM+vit were prepared in 0.85 % NaCl solution and 200 µL of each cell suspension was added in three technical replicates to a 96-well plate (Falcon). The excitation wavelength was 485 nm and the emission wavelength was 511 nm. The raw data was automatically collected in an excel sheet. The fluorescence was normalized manually by OD_{600} value. Data analysis and representation was done using Microsoft Excel.

11. SDS-PAGE and protein concentration determination

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was used for qualitative analysis of the protein expression of cloned genes and verification of plasmid expression systems.

11.1. Extraction of intracellular proteins

10 mL of an overnight bacterial culture was centrifuged at 3800 rpm for 10 min at 4 ºC and the supernatant was discarded. The cell pellet was resuspended in 0.85 % NaCl solution and centrifuged in Multispeed Centrifuge CL31R (Thermo Scientific) at 3800 rpm for 10 min at 4 ºC and the supernatant was discarded. The washing procedure was repeated two times. Finally, the cell pellet was resuspended in 1 mL of buffer 20 mM Tris-HCl (pH 8.0) and transferred to a 2 mL Eppendorf tube. Next, the cells were lysed by sonication in a Ultrasons sonicator (Selecta) by applying 5-6 pulses at the maximum intensity for 30 s. The time between pulses, the samples were placed on ice. Finally, the lysed cells were centrifuged in MiniSpin centrifuge (Eppendorf) at 13000 rpm for 1 min. The supernatant was transferred by pipetting to a clean Eppendorf tube and the pellet containing cell wall and insoluble proteins was resuspended in 1 mL of 20 mM Tris-HCl (pH 8.0) for further analysis.

11.2. Determination of soluble protein concentration

The concentration of soluble protein obtained from the intracellular extracts was determined by Bradford (Bradford, 1976) was used for the determination of soluble protein concentration. Bradford reagent at 5x (Bio-Rad) was diluted with water to make 1x Bradford reagent solution. The appropriate cell suspension dilutions were prepared and 20 µL of each sample was mixed with 1 mL Bradford solution. The protein absorption was measured spectrophotometrically at 595 nm with UV-mini 1240 spectrophotometer (Shimadzu). To calculate the protein concentrations, a previously prepared standard curve with bovine seroalbumin was used:

The optimal concentration to load the gel was 15-25 µg.

11.3. SDS-PAGE

To prepare the polyacrylamide gel for electrophoresis, a running gel (12.5%) and a stacking gel were prepared as indicated in a Table 13. First, the running gel was poured in the glass plates of Mini PROTEAN Tetra System (Bio-Rad) and let to polymerize. Then, the stacking gel was poured over the running gel and let to polymerize.

Table 13. Polyacrylamide gel preparation: reagents and volumes

Running gel (12.5%)	**Volume**	**Stacking gel**	**Volume**
Acrylamide 40%	3.13 mL	Acrylamide 40%	0.250 mL
1 M Tris-HCl (pH 8.0)	3.75 mL	1 M Tris-HCl (pH 8.0)	0.125 mL
SDS 20%	50 µL	SDS 20%	12.5 µL
Water	3.01 mL	Water	1.950 mL
APS 10%	50 µL	APS 10%	12.5 µL
TEMED	8 µL	TEMED	2.5 µL
TOTAL	**10 mL**	**TOTAL**	**2.35 mL**

The gel electrophoresis was conducted in a PowerPac Basic power supply (Bio-Rad) and the conditions were: initial 120 V for the stacking gel and 80 V for the running gel. Ladder used was BlueStar Prestained Protein Marker (Nippon Genetics). The gel was stained by Coomassie Brilliant Blue R 250 (Serva) while slowly agitating.

12. Extraction, purification and identification of polyketide compounds

12.1. Polyketide extraction procedure of culture supernatant

Upon the fermentation completion in MBM media, the 20 mL culture was centrifuged in Multispeed Centrifuge CL31R (Thermo Scientific) at 3800 rpm for 15 min at 4 ºC, the supernatant was frozen at -80 ºC and subsequently freeze dried. The lyophilized product was then dissolved in 4 mL of distilled water and equal volume of ethyl acetate. It was further thoroughly homogenized by vortex during 30 s and centrifuged at room temperature. The organic phase was separated by pipetting and the extraction was repeated. All the organic phase was dried by vacuum centrifugation to a solid extract which was further dissolved in 150 µL methanol and filtered for a HPLC/MS analysis.

12.2. Extraction procedure of the intracellular polyketide compounds

An overnight culture of *Labrenzia* sp. PHM005, wt and respective mutants, was diluted to an OD_{600} of 0.1 in a fresh 200 mL of MB medium. For the purpose of an assay of the labrenzin transport inside and outside the cell, 50 ng of purified labrenzin dissolved in methanol was added to a culture medium of non-producing mutant of *Labrenzia* sp. PHM005. After 24-72 h of incubation at 30 ºC and shaking at 220 rpm in New Brunswick Innova® 44/44R Incubator, the cultures were centrifuged using an ultracentrifuge Sorvall Lynx 6000 (Thermo Scientific) (rotor F14-6x250y) at 4000 rpm for 20-30 min at 4 ºC. Afterwards, the cells were washed and resuspended in a 10 mL of water followed by a centrifugation in a Multispeed Centrifuge CL31R (Thermo Scientific) at 3800 rpm for 10 min at 4 ºC. The washing procedure was repeated. The cell pellet was finally resuspended in 10 mL of water and the cells were submitted to high-pressure cell rupture using French Pressure Cell Press (American Instruments Company) in order to break the cells. After centrifucation in Multispeed Centrifuge CL31R (Thermo Scientific) at 3800 rpm for 15 min at 4 ºC, the supernatant was separated for further extraction (see chapter 12.1) and HPLC/MS analysis.

12.3. HPLC/MS chromatography

HPLC-MS was carried out using a mass spectrometry system (Thermo Mod. FinniganTM LXQTM) with on line HPLC Surveyor comprising pump with four separate solvent feeds (Surveyor MS pump Plus), an auto-sampler for multi sample analysis (Surveyor AS Plus), a Photo Diode Array detector (Surveyor PDA Plus) and a mass spectrometer LXQ equipped with linear ion trap. Ionization source used was electrospray ionization (ESI). Separation was performed on a C18 column (ZORBAX Eclipse plus C18, 5 µm, 4.6 × 250 mm, Agilent Technologies, Santa Clara, CA, United States). Two methods were developed throughout this thesis project. Short run method was following: solvent A was 100% water and solvent B was 100% acetonitrile. The flow rate was 1 ml min^{-1} and the gradient (per cent solvent A/B) was t = 2 min, 100% A; t = 5 min, 95% A; t = 25 min, 0% A; t = 27 min, 0% A; t = 30 min, 100% A; t = 35 min, 100% A. Long run method was adjusted for a 500µL min^{-1} flow rate with the following gradient: t = 2 min, 100% A; t = 8 min, 95% A; t = 40 min, 55% A; t = 53 min, 0% A; t = 55 min, 0% A; t = 57 min, 100% A; t = 65 min, 100% A. UV detection was set to 210 nm, 260 nm and 350 nm.

Results

1. Genomic study of *Labrenzia* sp. PHM005

Labrenzia sp. PHM005 was isolated from a marine sediment sample collected at 15 m of depth in the Indian Ocean by the coast of Kenya in 2004. PharmaMar collected the sample, isolated the strain, fermented the culture, identified and purified the novel pederin-family polyketide in the culture supernatant, later named labrenzin. The compound exhibited potent cytotoxic effect against four human cancer cell lines, and the GI_{50} (compound concentration that produces 50% cell growth inhibition, as compared to control cultures) was in the range $(2.0–2.9) \times 10^{-9}$ M (Schleissner et al., 2017). This was the first example reported of a free-living bacterium, which can be cultured in the laboratory, as a producer of a pederin-like compound. This opened the door to the genomic study of the gene cluster responsible for the biosynthesis of labrenzin and better understanding of the molecular mechanisms involved.

1.1. Genome sequencing

To start, the genome of *Labrenzia* sp. PHM005 was sequenced *de novo* by PacBio system and assembled in a single contig (see Materials and Methods 3.1.). Nonetheless, several sequencing errors were detected in the *trans*-AT PKS sequences that resulted in frameshifts that rendered several unexpected stop codons. Sanger sequencing of specific PCR amplicons, surrounding the sequencing errors, as schemed in the Figure 13 corrected the errors. To determine the presence of plasmids in the strain PHM005 that cannot be detected by the PacBio sequencing method as well as to correct other putative sequencing errors, the genome of strain PHM005 was re-sequenced using an Illumina MiSeq system (see Material and Methods 3.1).

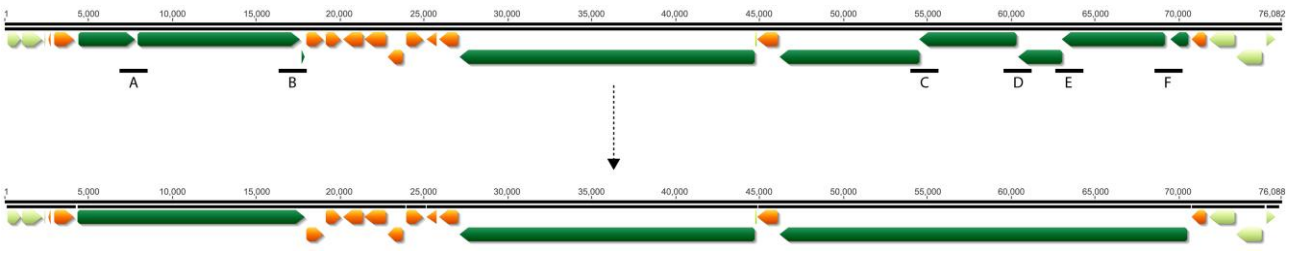


Figure 13. Correction of *lab* gene cluster sequence by Sanger sequencing. PCR-amplified regions for the correction of the ORFs are indicated by letters: A and B for gene *lab4*; C, D, E, F for gene *lab15*.

Results

The genome sequence of *Labrenzia* sp. strain PHM005 revealed two replicons comprising a circular chromosome of 6,167,349 bp (accession CP041191) and a circular plasmid (named p1BIR) of 19,450 bp (accession CP041190) shown in Figure 14. The genome average G + C content is 55%, contains 51 tRNAs, 3 copies of rRNA sequences (5S-23S-16S) and 5,988 coding sequences. Plasmid has a G + C content of 48.5% and contains 17 ORFs most of them coding for putative replication and conjugation related proteins (Table 14). The plasmid encodes two putative Abi family proteins involved in bacteriophage abortive infection resistance (BIR) systems (Garvey et al., 1995). A plasmid copy number of about 60 was calculated taking into account the genome and plasmid sizes and the number of Illumina reads covering both, *i.e.*, 3,820,045 reads and 712,018 reads, respectively (SRA accession SRS5035208). The whole genome of strain PHM005 was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) before the submission to the GenBank database.

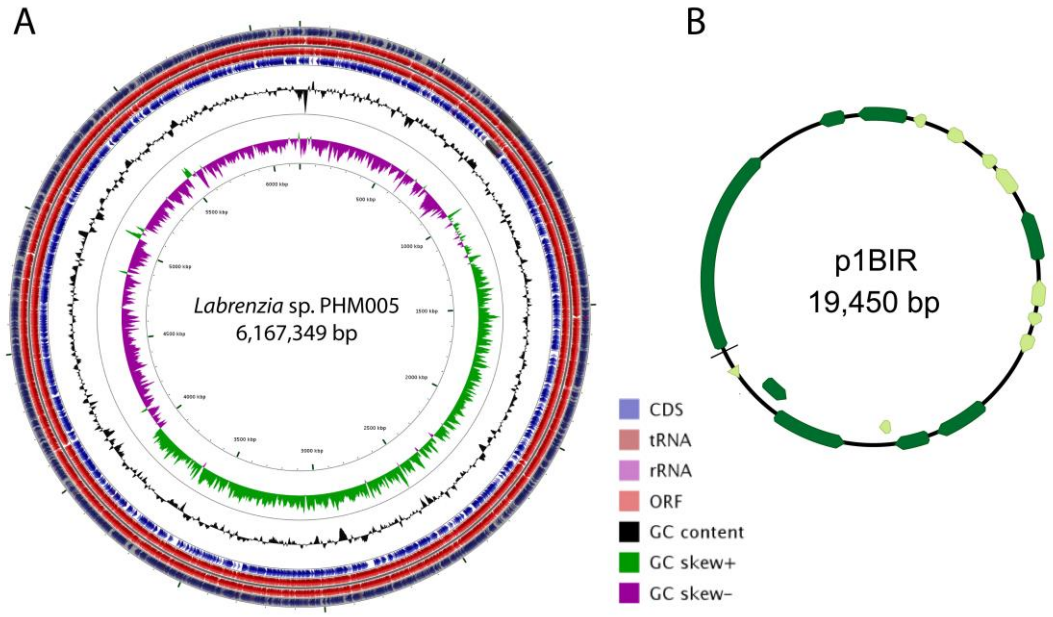


Figure 14. Circular maps of the genome of *Labrenzia* sp. PHM005 and plasmid p1BIR. (A) From outer to inner circles: The first four circles show the coding sequence (CDS), transfer ribonucleic acid (tRNA), ribosomal ribonucleic acid (rRNA), and open reading frame (ORF). The fifth circle represents the GC content (black). The sixth circle demonstrates the GC skew curve (positive GC skew, green; negative GC skew, violet). The genome position scaled in 500 kbp from base 1 is shown on the inner circle. (B) The circular map of plasmid p1BIR (Geneious version 10.0 created by Biomatters. Available from http://www.geneious.com). Dark green arrows represent genes coding for proteins with homology in other species from NCBI database and light green represents hypothetical proteins.

Table 14. Coding sequences of p1BIR and the closest homologs.

Putative protein	BLAST homolog	Origin	Accession number	Cover/identity	Pfam domain
Conjugal transfer coupling protein TraG	type IV secretory system conjugative DNA transfer family protein	*Geminicoccus roseus*	WP_084506409.1	50/39	Type IV secretory system Conjugative DNA transfer
Site-specific recombinases, DNA invertase	recombinase family protein	*Acuticoccus kandeliae*	WP_108663905.1	99/76	Resolvase, N terminal domain; helix-turn-helix domain
Abortive infection bacteriophage resistance protein	Abi family protein	*Sphingomonas melonis*	WP_020493507.1	88/58	no hits
hypothetical protein	hypothetical protein	*Alteromonas* sp. W12	WP_075176678.1	79/46	no hits
hypothetical protein	hypothetical protein	*Sinorhizobium fredii*	WP_037456464.1	93/33	no hits
hypothetical protein	hypothetical protein	*Bradyrhizobium ottawaense*	WP_091977009.1	86/43	no hits
hypothetical protein	hypothetical protein X773_05775	*Mesorhizobium* sp. LSJC285A00	ESW87211.1	87/83	no hits
Abortive infection bacteriophage resistance protein	Abi family protein	*Rhizobiales bacterium*	WP_113396613.1	92/54	Abi-like protein
hypothetical protein	hypothetical protein A4X03_g9487	*Tilletia caries*	OAI97131.1	90/43	no hits
hypothetical protein	alpha/beta fold hydrolase	*Vibrio fluvialis*	WP_044362754.1	83/30	no hits
hypothetical protein	hypothetical protein	*Labrenzia* sp. VG12	WP_094072921.1	94/33	no hits
RepA	plasmid replication initiator RepA	*Bacteroidetes bacterium*	PCJ62743.1	96/68	Replication initiator protein A
ParA	chromosome partitioning protein ParA	*Tropicibacter naphthalenivorans*	WP_058249280.1	97/79	VirC1 protein
hypothetical protein	hypothetical protein	*Acuticoccus yangtzensis*	WP_108676665.1	96/59	no hits
Relaxase/Mobilisation nuclease domain-containing protein	Relaxase/Mobilisation nuclease domain-containing protein	*Mesorhizobium australicum*	SMH26052.1	86/35	Relaxase/Mobilisation nuclease domain
MobC	plasmid mobilization relaxosome protein MobC	*Hyphomonas beringensis*	WP_034797206.1	54/49	no hits
hypothetical protein	hypothetical protein TEF_21935	*Rhizobiales bacterium* NRL2	ANK83160.1	84/45	no hits

1.2. Comparative genomics analysis

Functional comparison of genome sequences by Rapid Annotation Subsystem Technology (RAST) Server (Aziz et al., 2008) revealed *Labrenzia alexandrii* (strain DFL-11) (score 548) as the closest neighbor of strain PHM005, followed by *L. aggregata* (strain IAM 12614) (score 539), *Roseibium* sp. (strain TrichSKD4) (score 328), *Agrobacterium tumefaciens* (strain C58) (score 207) and *Sinorhizobium meliloti* (strain 5A14) (score 198). The genomes of five *Labrenzia* species, *i.e.*, *Labrenzia* sp. (strain PHM005) (accession CP041191), *L. alexandrii* (strain DFL-11) (accession CM011002), *L. aggregata* (strain RMAR6-6) (accession CP019630), *Labrenzia* sp. (strain CP4) (accession CP011927) and *Labrenzia* sp. (strain VG12) (accession CP022529), were analyzed for average nucleotide identity (ANI) to provide intraspecies and interspecies relationships (Table 15). These four strains were chosen because their whole genomes were the only complete genomes of the closest strains uploaded to the NCBI database at the time. Strain PHM005 appears to be most similar to *L. alexandrii* (strain DFL-11) although both genomes do not seem to belong to the same species since the ANI value is below 95% (Goris et al., 2007). Nonetheless, only *L. aggregata* (strain RMAR6-6) and *Labrenzia* sp. (strain CP4) most likely belong to the same species showing an ANI value of 97.76%.

Table 15. Average nucleotide identity (ANI) comparison of *Labrenzia* genomes.

ANI (%)	*Labrenzia* sp.PHM005	*Labrenzia alexandrii* DFL-11	*Labrenzia aggregata* RMAR6-6	*Labrenzia* sp. CP4	*Labrenzia* sp. VG12
Labrenzia sp.PHM005	100				
Labrenzia alexandrii DFL-11	80,23	100			
Labrenzia aggregata RMAR6-6	78,74	78,85	100		
Labrenzia sp. CP4	78,87	78,68	97,76	100	
Labrenzia sp. VG12	79,22	78,81	80,28	80,39	100

Using the five previously described *Labrenzia* species, a whole genome alignment using a progressive Mauve algorithm from the Geneious version 10.0.2 was performed. According to these algorithms, *L. alexandrii* (strain DFL-11) showed the highest local collinearity number, which supports its maximum similarity with PHM005 strain.

Table 16. Four *Labrenzia* genomes aligned to *Labrenzia* sp. PHM005 using progressive Mauve algorithm individually.

***Labrenzia* sp. PHM005 aligned to:**	**minimum weight for Locally Collinear Blocks**
***Labrenzia alexandrii* DFL-11**	164
***Labrenzia aggregata* RMAR-6**	42
***Labrenzia* sp. CP4**	92
***Labrenzia* sp. VG12**	72

2. *In silico* analysis of the *lab* cluster

The circular genome of *Labrenzia* sp. PHM005 was firstly set for a secondary metabolite gene cluster mining using an online platform antiSMASH 4.0. In total, 101 gene clusters have been identified: (*i*) one bacteriocin or other unspecified ribosomally synthesized and post-translationally modified peptide product (RiPP) cluster; (*ii*) 4 putative fatty acid clusters; (*iii*) 4 putative saccharide clusters; (*iv*) 89 putative clusters of unknown type; (*v*) one putative polyhydroxyalkanoate biosynthetic gene cluster; (*vi*) one putative mixed PKS/saccharide cluster; (*vii*) one *trans*-AT PKS/NRPS cluster. The later was studied in detail, as it showed high similarity with pederin gene cluster, as presented in the following sections. Additionally, the rest of the clusters were manually screened for similarity in other microorganisms and we have counted 17 clusters which lie within the order of Rhizobiales sharing from 10% to 32% of cluster genes (Table 17).

Table 17. Homologous gene clusters in other microorganisms according to ClusterBlast by antiSMASH 4.0.

cluster	**genome location**	**Cluster Blast hits**	**NCBI sequence ID**	**%**
t1pks-saccharide	480196 - 549637	*Labrenzia alexandrii* DFL-11	EQ973121	85
		Rhizobium sp. LC145	LBHV01000004	25
putative	668100 - 685508	*Pseudomonas* sp. GM74 PMI34	AKJG01000048	17
		Rhizobium tropici CIAT 899	CP004015	15
saccharide	686985 - 711837	*Rhizobium etli bv. mimosae* str. IE4771	CP006986	14
		Azorhizobium caulinodans ORS 571	AP009384	11
putative	732631 - 741605	*Kiloniella* sp. P1-1	LANI01000017	27
		Rhizobium rhizogenes NBRC 13257	BAYX01000006	22
fatty_acid	935479 - 956702	*Rhodomicrobium vannielii* ATCC 17100	CP002292	20
		Rhizobium sp. Leaf391	LMQG01000042	17
putative	993628 - 1003997	*Rhizobium* sp. Leaf311	LMNZ01000009	12
		Rhizobium rubi NBRC 13261	BBJU01000029	10
putative	2030747 - 2043034	*Burkholderia* sp. Leaf177	LMPF01000021	23

cluster	genome location	Cluster Blast hits	NCBI sequence ID	%
		Bradyrhizobium valentinum strain LmjM3	LLXX01000143	20
putative	2140378 - 2156286	*Bradyrhizobium elkanii* strain UASWS1015	JXOF01000173	13
		Bradyrhizobium sp. YR681 PMI42	AKIY01000259	10
putative	2225472 - 2235766	*Mesorhizobium* sp. LNJC405B00	AYWC01000002	15
		Mesorhizobium sp. Root554	LMGA01000001	12
putative	2359048 - 2367711	*Phaeobacter* sp. CECT 5382	CYSG01000021	34
		Bradyrhizobium sp. DFCI-1	AMFB01000012	21
putative	2379843 - 2396431	*Rhizobium rubi* NBRC 13261	BBJU01000023	17
putative	3516897 - 3522239	*Neorhizobium galegae bv. officinalis*	CCRH01000015	17
fatty_acid	4522154 - 4543389	*Sinorhizobium fredii* HH103	HE616890	25
putative	4543774 - 4553183	*Bradyrhizobium* sp. STM 3843	CAFK01000193	16
putative	4658717 - 4682200	*Mesorhizobium* sp. Root695	LMHO01000015	20
fatty_acid	4699018 - 4719995	*Phaeospirillum fulvum* MGU-K5	AQPH01000012	48
		Mesorhizobium sp. Root102	LMCP01000045	32
putative	4825217 - 4843781	*Haematobacter missouriensis* strain CCUG 52307	JFGS01000018	25
		Bradyrhizobium sp. LTSP849	JYMR01000055	25

2.1. Identification and delimitation of the cluster

The genome of *L. alexandrii* (strain DFL-11) was used to estimate the size and the putative boundaries of the *trans*-AT PKS cluster in the genome of strain PHM005. The alignment of both genomes is shown in Figure 15A. This alignment revealed that the *trans*-AT PKS cluster has been probably inserted within two rDNA regions. The putative cluster limits were further compared with the whole genomes of other three previously mentioned species of *Labrenzia*, *i.e.*, *L. aggregata* (strain RMAR6-6) (accession CP019630), *Labrenzia* sp. (strain CP4) (accession CP011927) and *Labrenzia* sp. (strain VG12) (accession CP022529). Interestingly, the genes located downstream from the gene cluster were found in the same orientation in all species. The similarity and the corresponding gene positions in each species are shown in Figure 15B. On the contrary, the genes located upstream the *trans*-AT PKS cluster were only found in *L. alexandrii* (strain DFL-11) (gene position at genome between 3,628,890 bp and 3,648,023 bp following the same order). Accordingly, to these comparisons, the homologous genes present in at least two genomes were discarded as candidates for the biosynthesis of labrenzin allowing us to fix the putative boundaries.

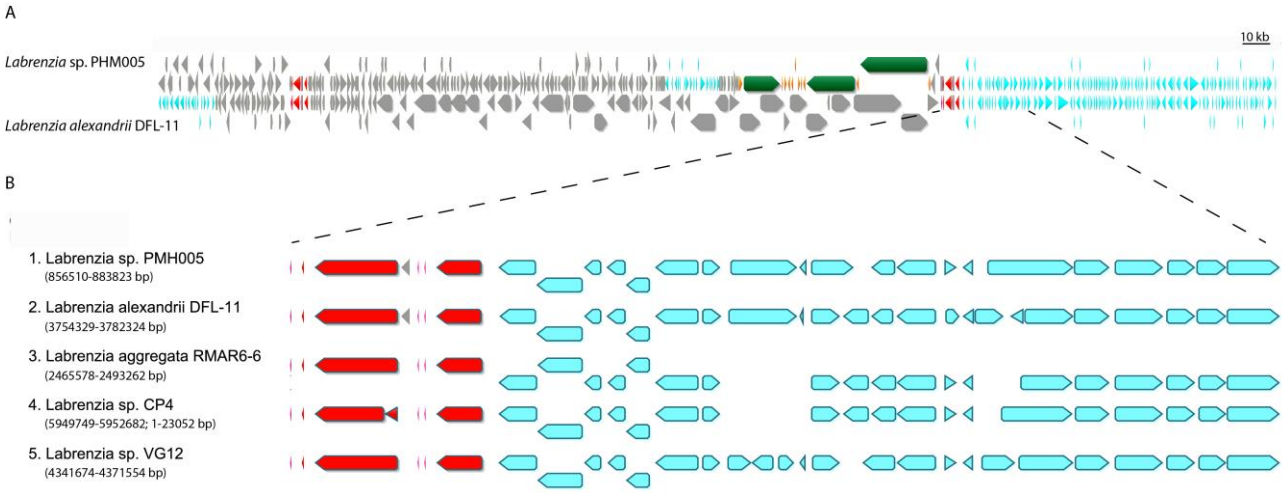


Figure 15. Representation of the lab gene cluster and its bordering genes distributed among the selected *Labrenzia* species genomes. A) MAFFT alignment of *L. alexandrii* DFL-11 and *Labrenzia* sp. PHM005. Coding sequences are represented by arrows: gray, CDS; red, ribosomal DNA; green and orange, *lab* gene cluster; light blue; identical genes. B) MAFFT alignment and localization of downstream bordering genes in *Labrenzia* species.

Finally, the *trans*-AT PKS/NRPS cluster, further named *lab* cluster, consists of a 79-kb region (776,792–855,905 bp) comprising 3 genes encoding multi-domain PKS/NRPS proteins (*lab4*, *lab13*, and *lab15*), 16 auxiliary enzymes with assigned functions and 6 uncharacterized ORFs annotated as hypothetical proteins. The results of the BLAST protein homology analysis indicate the similarity with pederin, onnamide and diaphorin gene clusters Table 18.

Results

Table 18. BLAST homology of the *Labrenzia* sp. PHM005 *trans*-AT PKS gene cluster as of May 2017. ORFs are listed from left to right of the cluster (upstream to downstream).

ORF	Annotation (RAST)	BLAST homolog	Origin	accession number	cover/identity %
HYP1	DUF3089 domain-containing protein	DUF3089 domain-containing protein	*Labrenzia* sp. DG1229	WP_051644561.1	100/65
HYP2	ParA family protein	ParA family protein	*Labrenzia marina*	WP_103225501.1	100/73
HYP3	DUF697 domain-containing protein	DUF697 domain-containing protein	*Labrenzia* sp. Alg231-36	WP_108873020.1	100/81
HYP4	DUF697 domain-containing protein	DUF697 domain-containing protein	*Labrenzia* sp. Alg231-36	WP_108873020.1	100/81
HYP5	prohibitin family protein	prohibitin family protein	*Stappia indica*	WP_067221789.1	98/62
Lab1	ABC transporter substrate-binding protein	ABC transporter substrate-binding protein	*Oceanibaculum indicum*	WP_008945422.1	91/37
HYP6	hypothetical protein	hypothetical protein	*Oceanibaculum indicum*	WP_008945421.1	95/62
Lab2	acyl carrier protein	acyl carrier protein PedN	*Candidatus* Profftella armatura	WP_020915408.1	95/44
Lab3	hydroxymethylglutaryl-CoA synthase family protein	3-hydroxy-3-methylglutaryl-ACP synthase	*Clostridium beijerinckii*	WP_023973754.1	100/65

ORF	Annotation (RAST)	BLAST homolog	Origin	accession number	cover/identity %
Lab3	hydroxymethylglutaryl-CoA synthase family protein	PedP	Symbiont bacterium of *Paederus fuscipes*	AAW33975.1	99/61
Lab4	*trans*-AT polyketide synthase type I	OnnB	Symbiont bacterium of *Theonella swinhoei*	AAV97870.1	79/40
		putative type I polyketide synthase PedI	Symbiont bacterium of *Paederus fuscipes*	AAR19304.1	96/42
Lab5	LLM class flavin-dependent oxidoreductase	luciferase	*Methylocucumis oryzae*	WP_045777999.1	100/64
		flavin-dependent oxygenase	*Nostoc* sp. *'Peltigera membranacea* cyanobiont'	ADA69238.1	100/60
		putative oxygenase PedJ	Symbiont bacterium of *Paederus fuscipes*	AAR19305.1	100/61
		OnnC	Symbiont bacterium of *Theonella swinhoei*	AAV97871.1	100/63
Lab6	methyltransferase domain-containing protein	OnnG	Symbiont bacterium of *Theonella swinhoei*	AAV97875.1	99/51
		OnnD	symbiont bacterium of *Theonella swinhoei*	AAV97872.1	97/46
		hypothetical protein ETSY1_46125 (plasmid)	Candidatus *Entotheonella* sp. TSY1	: ETX03779.1	99/51
		PedO	Symbiont bacterium of *Paederus fuscipes*	AAW33974.1	97/43
		putative methyltransferase PedA	Symbiont bacterium of *Paederus fuscipes*	AAS47557.1	97/47
		SAM-dependent methyltransferase, partial	*Methylocucumis oryzae*	WP_052700174.1	72/58
Lab7	cytochrome P450	cytochrome P450	*Labrenzia alba*	WP_055112014.1	97/54
		cytochrome P450	*Oceanibaculum indicum*	WP_008944496.1	98/54
		cytochrome P450 monooxygenase	*Magnetospirillum gryphiswaldense* MSR-1	CAM75136.1	98/48

ORF	Annotation (RAST)	BLAST homolog	Origin	accession number	cover/identity %
Lab8	FMN-dependent oxidoreductase	FMN-dependent oxidoreductase PedB	Symbiont bacterium of *Paederus fuscipes*	AAS47558.1	98/56
Lab8		putative FMN-dependent oxidoreductase, PedB-like protein	*Candidatus* Profftella armatura	WP_020915402.1	98/52
Lab9	acyltransferase domain-containing protein	acyltransferase domain-containing protein	*Methylobacter tundripaludum*	WP_104425066.1	98/41
		polyketide biosynthesis acyltransferase	*Candidatus* Profftella armatura	WP_020915461.1	92/34
		putative acyltransferase PedC	Symbiont bacterium of *Paederus fuscipes*	AAS47559.1	98/35
Lab10	ACP S-malonyltransferase	malonyl CoA-acyl carrier protein transacylase	*Burkholderia gladioli*	WP_036053944.1	98/51
		malonyl CoA-acyl carrier protein transacylase	*Methylocucumis oryzae*	WP_045778127.1	98/50
		malonyl CoA-acyl carrier protein transacylase	*Candidatus* Entotheonella sp. (ex. *Theonella swinhoei*)	AKQ22695.1	98/52
		putative acyltransferase PedD	Symbiont bacterium of *Paederus fuscipes*	AAS47563.1	95/51
Lab11	enoyl-CoA hydratase/isomerase	enoyl-CoA hydratase	*Oceanibaculum indicum*	WP_008944492.1	94/50
		enoyl-CoA hydratase	*Labrenzia* sp. DG1229	WP_029061353.1	97/43
		PedL	Symbiont bacterium of *Paederus fuscipes*	AAW33971.1	91/43
Lab12	polyketide beta-ketoacyl:ACP synthase	polyketide beta-ketoacyl:ACP synthase	*Oceanibaculum indicum*	WP_008944493.1	98/44
		Polyketide biosynthesis malonyl-ACP decarboxylase PksF	*Labrenzia alba*	CTQ66097.1	97/42
		PedM*	Symbiont bacterium of *Paederus fuscipes*	AAW33972.1	100/27

ORF	Annotation (RAST)	BLAST homolog	Origin	accession number	cover/identity %
Lab13	mixed type I polyketide synthase - peptide synthetase	mixed type I polyketide synthase - peptide synthetase	Symbiont bacterium of *Paederus fuscipes*	AAS47562.1	99/42
		non-ribosomal peptide synthetase	*Magnetovibrio blakemorei*	WP_069956861.1	98/36
		Polyketide synthase PksN	*Labrenzia alba*	CTQ74038.1	98/35
Lab14	monooxygenase	putative FAD-dependent monooxygenase	Symbiont bacterium of *Paederus fuscipes*	AAS47561.1	94/73
		flavin-containing monooxygenase PedG-like protein	*Candidatus* Profftella armatura	WP_020915400.1	98/71
		monooxygenase	*Labrenzia alba*	WP_055112032.1	97/62
Lab15	mixed type I polyketide synthase - peptide synthetase	mixed type I polyketide synthase/nonribosomal peptide synthetase	Symbiont bacterium of *Paederus fuscipes*	AAS47564.1	99/42
		non-ribosomal peptide synthetase	*Methylobacter tundripaludum*	WP_104425077.1	96/46
		OnnI, partial	Symbiont bacterium of *Theonella swinhoei*	AAV97877.1	92/46
Lab16	methyltransferase domain-containing protein	OnnH	Symbiont bacterium of *Theonella swinhoei*	AAV97876.1	99/43
		putative methyltransferase PedE	Symbiont bacterium of *Paederus fuscipes*	AAS47560.1	98/51
		class I SAM-dependent methyltransferase	*Methylocucumis oryzae*	WP_045777863.1	94/57
Lab17	ATP-binding cassette domain-containing protein	hypothetical protein ETSY1_10275	*Candidatus* Entotheonella sp. TSY1	ETX00685.1	99/33
		cyclic peptide transporter	*Oceanibaculum indicum*	WP_008944488.1	98/52

Results

ORF	Annotation (RAST)	BLAST homolog	Origin	accession number	cover/identity %
Lab18	cyclic peptide export ABC transporter	cyclic peptide transporter	*Oceanibaculum indicum*	WP_008944489.1	99/51
		fused multidrug transport subunits of ABC superfamily transporter: permease component/ATP-binding protein	*Magnetospirillum gryphiswaldense*	WP_024079861.1	95/40
		ABC transporter, ATP-binding/permease protein	*Magnetospirillum gryphiswaldense* MSR-1	CAM74928.1	98/51
Lab19	4'-phosphopantetheinyl transferase superfamily protein	4'-phosphopantetheinyl transferase superfamily protein	*Methylobacter tundripaludum*	WP_104424065.1	97/42
		4'-phosphopantetheinyl transferase	*Candidatus* Profftella armatura	WP_020915412.1	82/31

* BLAST Global Alignment

2.2. Hypothesis for the labrenzin biosynthetic pathway

The core labrenzin molecule is constructed following an “assembly line” biosynthesis by the modular PKSs/NRPSs and further modified by the tailoring enzymes as proposed (Figure 16B). The structures of biosynthesis intermediates of labrenzin have been deducted from the specificity of enzymatic domains of each PKS/NRPS module. The domain analysis is based on the antiSMASH 4.0 results and the described conserved motifs of the three PKS/NRPSs (El-Sayed et al., 2003).

According to the modular structure predicted by antiSMASH v.4.0., the PKS (Lab4) from strain PHM005 consists of a “starter” module 0 with a GCN5-related N-acetyltransferase family (GNAT) domain that catalyzes the decarboxylation of malonyl-CoA to generate acetyl-CoA and its translocation to the ACP unit. After that, three elongating β-ketoacyl synthases (KSs) incorporate the three malonyl-CoA building units. Compared to PedI from *P. fuscipes* symbiont (Piel *et al.*, 2004), the first module of Lab4 contains an additional dehydratase (DH) domain, most probably inactive, since it lacks the typical conserved motif HXXXGXXXXP (El-Sayed et al., 2003). In addition, the presence of only one ECH domain in the third module does not appear to affect the structure of the first pyran ring of the labrenzin core structure.

The domain architecture of the PKS/NRPS (Lab15) from strain PHM005 is almost identical to the architecture of PedF enzyme (Piel, 2002). The main singularity observed in Lab15 is the presence of a putative inactive DH domain attached to the first non-elongating KS0, while in PedF there is one putative inactive DH domain in the last module.

According to antiSMASH predictions, the PKS/NRPS (Lab13) from strain PHM005, follows the domain order described in PedH, which is hypothesized by (Piel, 2002) to synthesize the chain containing the terminal arginine residue found in onnamide analogs (Matsunaga et al., 1992). However, there are a few exceptions. Substrate predictions of Lab13 by pHHM and SANDPUMA indicate arginine and alanine, respectively, as candidate amino acids for the chain terminal of the putative onnamide-like molecule; therefore, it cannot be concluded with accuracy the last amino acid incorporation. Lab13 also possesses another putative inactive DH domain in the third module. The last non-elongating KS domain lacks the histidine conserved residue and the ACP domain located downstream lacks the conserved GxDS motif (El-Sayed et al., 2003).

As presented in Figure 16B, it can be suspected that the genes encoding tailoring oxidoreductases and methyltransferases would be responsible for the hydroxylations and methylations of the polyketide core molecule, respectively. Since labrenzin intermediate containing the lateral chain has not been detected in the culture extracts, it can be assumed that some other enzyme, encoded in the *lab* cluster or elsewhere in the genome should hydrolyze the final polyketide product into two products: labrenzin and the cryptic moiety. Following the initial hypothesis set by Piel's research group over the years, it can be assumed that a yet unknown C-C bond cleavage mechanism is probably involved in the labrenzin biosynthesis. One possibility is to accept the premature cleavage of the polyketide intermediate at the end of the PKS/NRPS15, and the other possibility is to assume the hydrolysis of the final polyketide product after the biosynthesis on the PKS complex. The possible candidates for the hydrolytic cleavage of the molecule could be the cytochrome P450 (Lab7) and the monooxygenase (Lab14), although their mechanisms of action remain unknown.

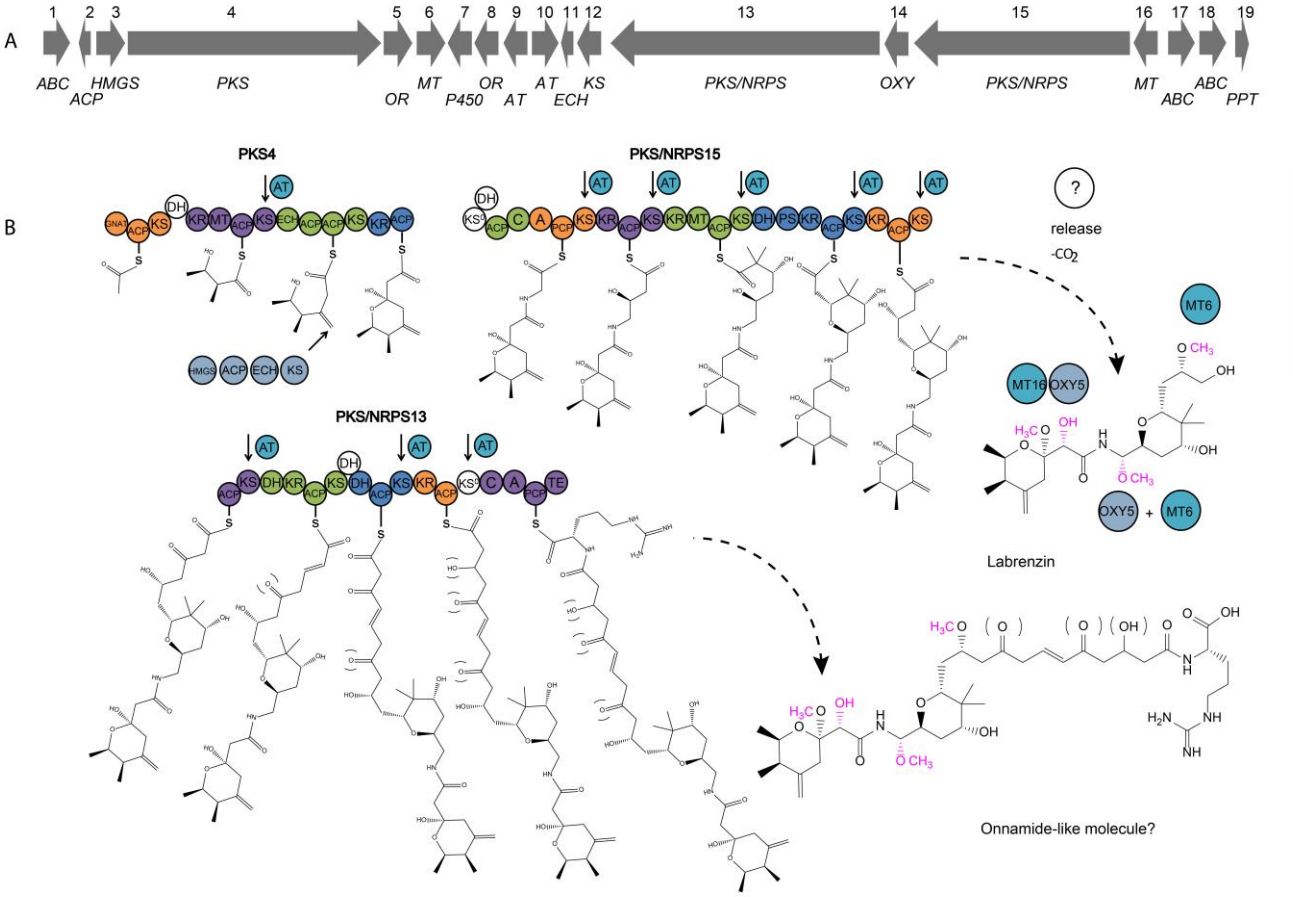


Figure 16. A) Map of the lab cluster comprising the genes for labrenzin production. B) Scheme predicting the pederin/onnamide modular biosynthesis in the free-living Alfaproteobacterium *Labrenzia* sp. PHM005. Specific domains have been identified and analyzed using antiSMASH. Domains (shown in circles) that are part of the same module are represented in the same color, less the putative non-functional domains in white. Module boundaries are defined according to recent improvements (Vander Wood and Keatinge-Clay, 2018). GNAT, GCN5-related N-acetyltransferase domain; ACP, acyl carrier protein domain; KS, ketosynthase domain; KS0, non-elongating ketosynthase domain; KR, ketoreductase domain; MT, methyltransferase domain; ECH, enoyl- CoA hydratase/isomerase domain; A, adenylation domain; AT, acyltransferase; C, condensation domain; DH, dehydratase domain; PS, pyran synthase; PCP, peptidyl carrier protein domain; TE, thioesterase domain; OR, oxidoreductase; MT, methyltransferase. Putative *trans*-AT docking site associated with its respective KS is indicated by arrows. Two pathways are indicated: the synthesis of the labrenzin and the onnamide A questioned to be the complete molecule of the proposed pathway. Putative functional groups are marked inside brackets and putative modifications by tailoring enzymes are colored in pink.

2.3. Comparative analyses with homologous gene clusters

2.3.1. Pederin family gene clusters

Using as a reference the genes of the pederin cluster from the symbiont of *P. fuscipes*, the homologous genes and their protein identities of the corresponding genes of *Labrenzia* sp. PHM005 are visualized in Figure 17. The first observed difference between the two almost identical biosynthetic systems is the gene distribution in their host genomes. In the symbiont bacterium of *P. fuscipes*, the pederin genes are clustered in three genomic islands flanked by IS elements and transposons (Piel et al., 2004d, 2005). However, the pederin biosynthetic genes of strain PHM005 are clustered together, suggesting its horizontal gene transfer, a very common phenomenon in bacteria (Roth and Lawrence, 1996).

The *lab* gene cluster of strain PHM005 comprises three putative ABC transporters, one cytochrome P450 and one 4'-phosphopantetheinyl transferase, which have not been identified in *P. fuscipes* or *T. swinhoei*. On the other hand, there are three proteins in the *P. fuscipes* symbiont, that are absent in *Labrenzia* sp. PHM005, *i.e.*, a putative PedQ esterase, a PedR regulator and a protein of unknown function named PedK in *P. fuscipes* or OnnE and OnnF in *T. swinhoei* (Piel et al., 2004b). The tailoring enzymes in strain PHM005 showed protein identity <60%, with the exception of the hydroxymethylglutaryl-CoA synthase (HMGS) and the oxygenase (OXY) which have 61% and 73% identity, respectively. Among the tailoring enzymes in *lab* cluster, there is a presence of HMGS, stand-alone ACP, KS and ECH, two putative acetyl-transferases (AT) and one oxidoreductase (OR). Another difference regarding the *ped* cluster is the number of methyl-transferases (MTs). *Ped* cluster described in the beetle symbiont comprises three MTs (PedA, PedE and PedO) while *lab* cluster in PHM005 strain has only two.

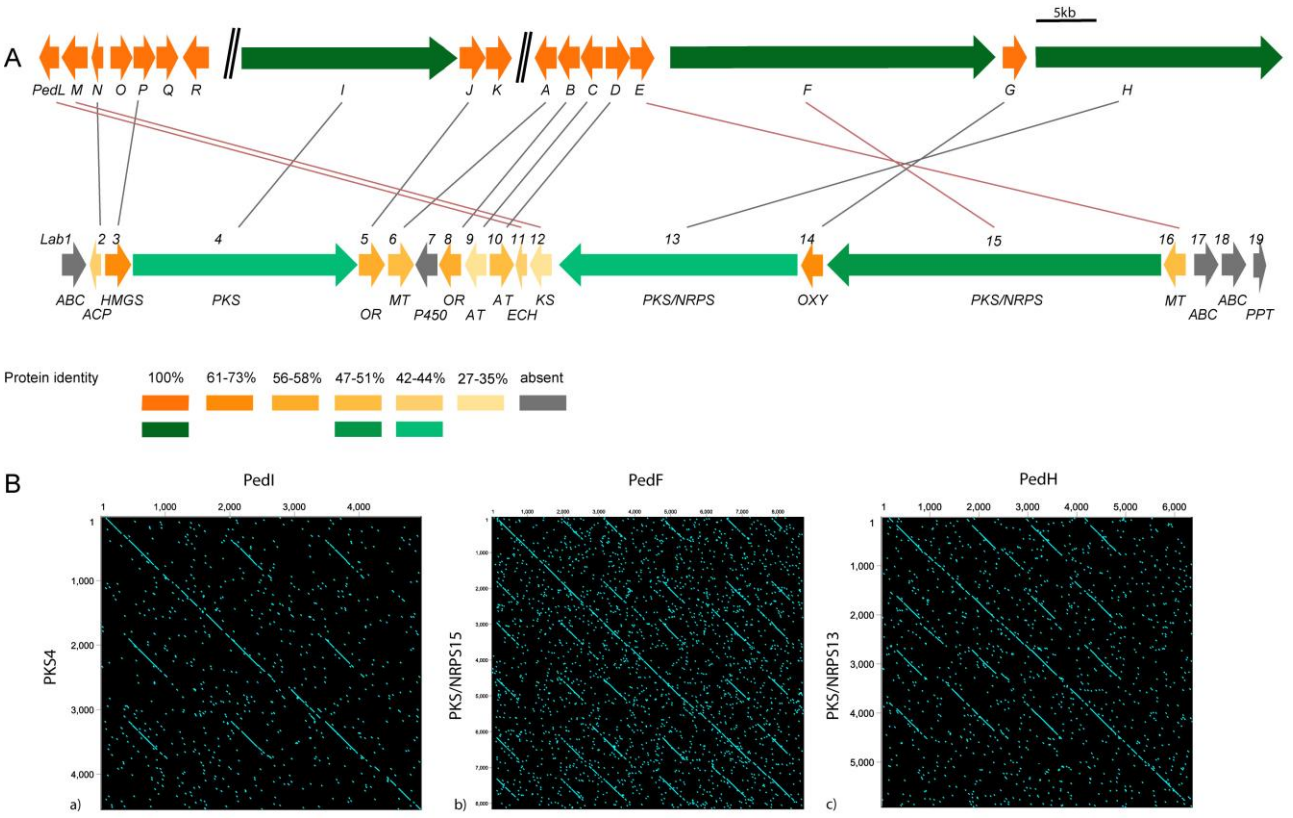


Figure 17. Comparative analysis of a pederin cluster from a symbiont bacterium of *P. fuscipes* and the free-living alphaproteobacterium *Labrenzia* sp. PHM005. Modified scheme of the pederin cluster as described from a: A) bacterium symbiont of *P. fuscipes* (Piel et al., 2004c) (above) and the newly described labrenzin cluster from *Labrenzia* sp. PHM005 (below). Homologous genes between two clusters are indicated by lines. Protein identity is visualized using colored scale that matches the identity percentage of all the genes from the cluster, using as reference the symbiont bacterium of *P. fuscipes* genes. PKS and NRPS genes are shown in green and tailoring enzymes in orange and gray. ABC, transporter; ACP, acyl carrier protein; AT, acyltransferase; ECH, enoyl-CoA hydratase/isomerase domain; HMGS, hydroxymethylglutaryl-CoA synthase; KS, ketosynthase; MT, methyltransferase; OR, oxidoreductase; OXY, monooxygenase; P450, cytochrome hydroxylase; PPT, phosphopantetheinyl transferase. B) Graphical representation of the homology using dotplotting between PKS 4, 13, and 15 and PedI, PedH, and PedF, respectively.

2.3.1.1. PKS and hybrid PKS/NRPS similarities with other homologs

Apart from pederin and onnamide gene clusters mentioned above, other biosynthetic gene clusters of compounds from the pederin family, such as those described for the biosynthesis of diaphorin, psymberin, nosperin and cusperin were inspected for the PKS/NRPS domain architecture using the antiSMASH v.4.0 software (Figure 18A). The starting PKS responsible for the synthesis of the first pyran ring of labrenzin is conserved in all these clusters, except in the psymberin cluster where PKS lacks the second module resulting in the absence of three C atoms. There is also a difference in the number of the tandem ACP domains found in the third module of the starting PKS. In addition, the third module comprises two enoyl-CoA hydratase (ECH1 and ECH2) domains in OnnB, while the rest of the starting PKSs contain a single ECH domain based on antiSMASH analysis. The ECH domains can exist as separate enzymes, as a part of the HMGS-cassette, or as embedded domains in the PKSs, like ECH2 in curacin A (Gu et al., 2006). It is hypothesized that the ECH domains are responsible for the exomethylene group formation in the pederin pyran structure (Piel et al., 2004d). The amino acid sequence of the putative ECH1 domains in all 7 PKSs were aligned using as reference the characterized ECH1 functional domain (CurE) from the curacin A biosynthetic pathway (Gu et al., 2006) (Figure 18B). The OnnB, NspA, CusA, and PedI maintain the conserved amino acid motifs, while DipP, PsyA, and PKS4 have deleted the regions containing the conserved motifs of ECH1, suggesting that they might not be functional and that only one functional ECH domain is enough to generate the corresponding compounds. It is also possible that some other tailoring enzyme participates in that particular biosynthetic step.

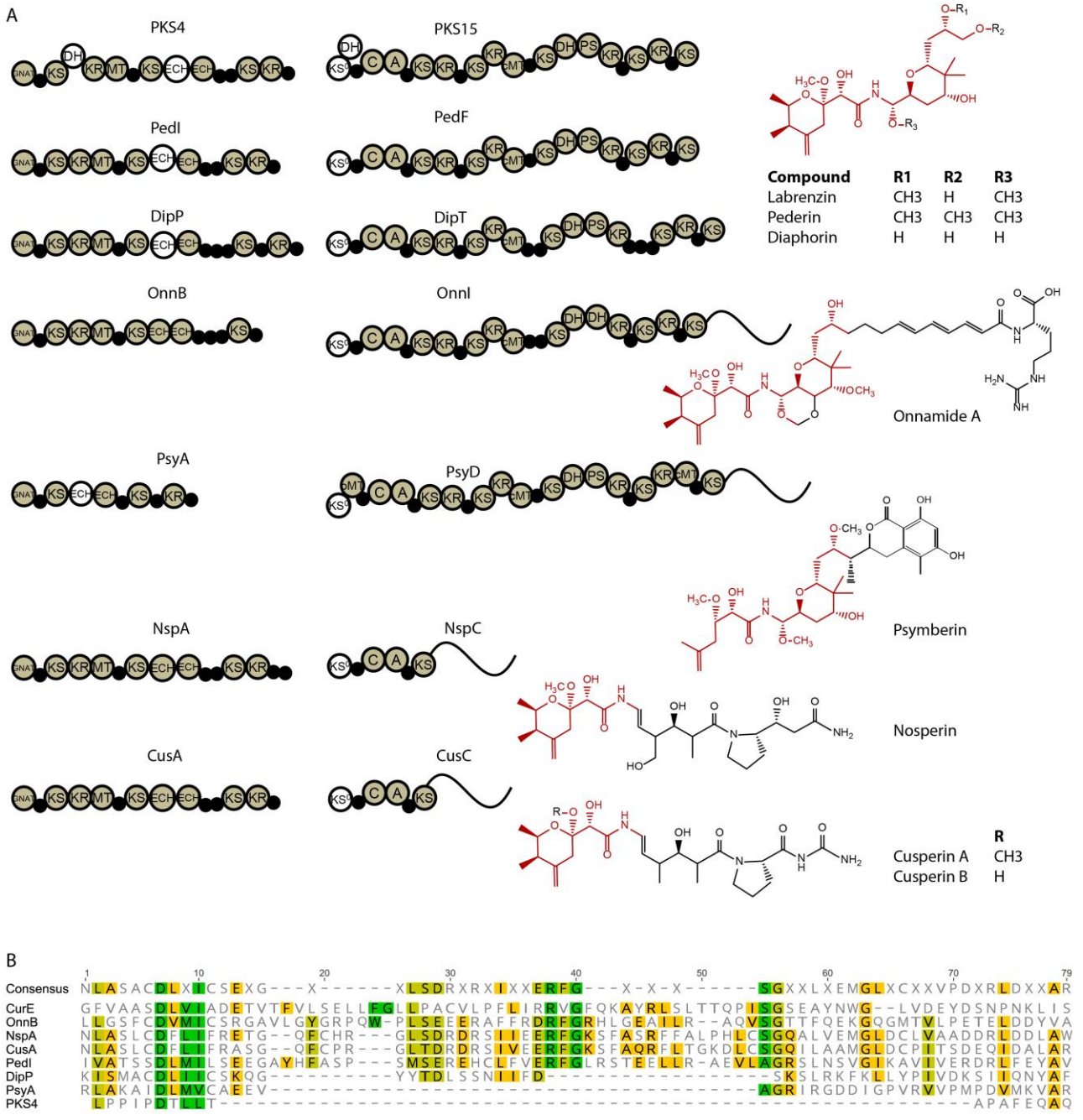


Figure 18. A) Modular scheme and domain comparison of PKS/NRPSs responsible for the biosynthesis of pederin family compounds isolated from organisms of different ecological niches: labrenzin, *Labrenzia* sp. PHM005; pederin, *P. fuscipes* symbiont; diaphorin, *Candidatus* Profftella armatura; onnamide A, *T. swinhoei* symbiont; psymberin, *Psammocinia* sp. symbiont; nosperin, *Nostoc* sp.; cusperin, *Cuspidothrix issatschenkoi* (top to bottom). Common core structure of the respective polyketides is colored red. B) Alignment (aa) of putative ECH1 domains from the first PKS in onnamide A, nosperin, cusperin, pederin, diaphorin, psymberin and labrenzin (top to bottom).

2.3.2. Sesbanimide gene cluster

During this thesis, two additional marine bacterial genomes that contained a *trans*-AT polyketide cluster strikingly similar to *lab* cluster were sequenced. *S. indica* PHM037 was isolated from the sponge *Pandaros acanthifolium* collected in the Caribbean Sea (Martinique Island) in October 2003 and *L. aggregata* PHM038 was isolated from the sponge belonging to the genus of *Plakortis* in Nosy Be (Madagascar) in November 2017. Both strains were sequenced as explained in Material and Methods 3.1. To explore the secondary metabolite gene clusters contained in the genomes of PHM037 and PHM038 strains the online platform of antiSMASH 5.0 was used. Genome mining revealed one almost identical *trans*-AT PKS cluster in both strains. Since the strains were producers of novel glutaramide-type polyketides, denominated sesbanimides, the cluster was named as *sbn* gene cluster.

The region of 64 kb encodes a modular PKS, a mixed type I PKS/non-ribosomal peptide synthetase (PKS/NRPS), 16 auxiliary enzymes with assigned functions and 5 hypothetical proteins, possibly involved in sesbanimide biosynthesis. Figure 19A shows the *sbn* cluster delimited by the marginal genes shared by PHM037 and PHM038 and named alphabetically *sbnA*-*W*. Interestingly, the cluster comprises 19 genes homologs to the *lab* genes, hypothetical included, which is more than 80% of the *sbn* cluster. The *sbn* ORFs and their homologous protein are shown in Table 19. Based on comparative analyses, computational predictions, and the elucidated chemical structures of sesbanimide intermediates it has been proposed the sesbanimide assembly on a multi-domain modular PKS and PKS/NRPS with the assistance of several *trans*-acting protein components of the enzymatic complex (Figure 19B) based on the gene structure of PHM037.

Table 19. ORFs from the *sbn* cluster and their closest homologues in the RefSeq database.

ORF	Protein/ gene	putativefunction/homologue	Origin	Cover/identity %	Accession number of a protein homologue
SbnA	**AT/AH**	acyltransferase domain-containing protein	*Stappia indica* USBA 352	100/97.82	WP_067221761.1
SbnB	**4'-PPT**	4'-phosphopantetheinyl transferase	*Stappia indica* USBA 352	99/98.17	WP_067221763.1
SbnC	**EST**	metallophosphoesterase	*Stappia indica* USBA 352	100/97.22	WP_067221764.1
SbnD	**MT**	FkbM family methyltransferase	*Stappia indica* USBA 352	100/94.03	WP_083202408.1
SbnE	**P450**	cytochrome P450	*Stappia indica* USBA 352	100/97.61	WP_083206337.1
SbnF	**HMGS**	hydroxymethylglutaryl-CoA synthase	*Stappia* sp. ARW1T	100/98.09	WP_120269012.1

ORF	Protein/ gene	putativefunction/homologue	Origin	Cover/identity %	Accession number of a protein homologue
SbnG	**ACP**	acyl carrier protein	*Stappia indica* USBA 352	100/98.77	WP_067221771.1
SbnH	**KS**	polyketide beta-ketoacyl:ACP synthase	*Stappia* sp. ARW1T	100/91.85	WP_120269013.1
SbnI	**ECH**	enoyl-CoA hydratase/isomerase	*Stappia indica* USBA 352	100/94.94	WP_067339094.1
SbnJ	**AMT**	asparagine synthase (glutamine-hydrolyzing)	*Stappia indica* USBA 352	100/93.87	WP_067221776.1
SbnK	**ACP**	acyl carrier protein	*Stappia* sp. ARW1T	100/96.34	WP_120269016.1
SbnL	**ABC**	cyclic peptide export ABC transporter	*Stappia indica* USBA 352	100/96.33	WP_083202407.1
SbnM	**ABC**	cyclic peptide export ABC transporter	*Stappia indica* USBA 352	100/97.27	WP_097175571.1
SbnN	**AT**	ACP S-malonyltransferase	*Stappia indica* USBA 352	100/94.66	WP_097175572.1
SbnO	**PKS**	SDR family NAD(P)-dependent oxidoreductase	*Stappia* sp. ARW1T	100/88.66	WP_120269020.1
SbnP	**OXY**	NAD(P)-binding domain-containing protein; monooxygenase	*Stappia indica* USBA 352	100/96.50	WP_097175574.1
SbnQ	**PKS/NRPS**	non-ribosomal peptide synthetase	*Stappia indica* USBA 352	100/86.90	WP_067221787.1
SbnR	**HYP1**	prohibitin family protein	*Stappia indica* USBA 352	100/97.92	WP_067221789.1
SbnS	**ABC**	ABC transporter substrate-binding protein	*Stappia indica* USBA 352	100/92.36	WP_067221790.1
SbnT	**HYP2**	hypothetical protein	*Stappia indica* USBA 352	97/94.37	WP_067221792.1
SbnU	**HYP3**	DUF697 domain-containing protein	*Stappia* sp.ARW1T	100/97.17	WP_120269026.1
SbnV	**HYP4**	DUF697 domain-containing protein	*Stappia* sp. ARW1T	99/86.78	WP_147421864.1
SbnW	**HYP5**	ParA family protein	*Stappia indica* USBA 352	100/99.08	WP_097175579.1

In analogy to biochemically characterized glutaramide assembly in gladiofungin molecule (Niehs et al., 2020), the loading module composed of the amidotransferase SbnJ and the ACP SbnK should be responsible for transferring the amino group from amino acid substrate, likely glutamine, to activated malonate. The complete glutaramide starting unit is further generated with the aid of the first two modules in the PKS SbnO. The loading module together with the unusual KS-B-ACP module are present in PKSs of other glutaramides, such as cyclohexamide (Yin et al., 2014), migrastatin, iso-migrastatin, dorrigocin (Lim et al., 2009), 9-methylstreptimidone (Wang et al., 2013), and lactimidomycin (Seo et al., 2014). The remaining portion of the sesbanimide A core structure is assembled on the downstream SbnO modules. The last module contains tandem ECH domains and tandem ACPs as known from other *trans*-AT PKS modules that generate exomethylene branches (Helfrich and Piel, 2016; Piel, 2002). Both ACPs possess a distinctive motif DSxxxxxW identified as important for β-branching (Kosol et al., 2018). The genes *sbnF-I* encoding HMGS, ACP, KS, and ECH homologs, respectively, form a HCS cassette to assist in β-branching.

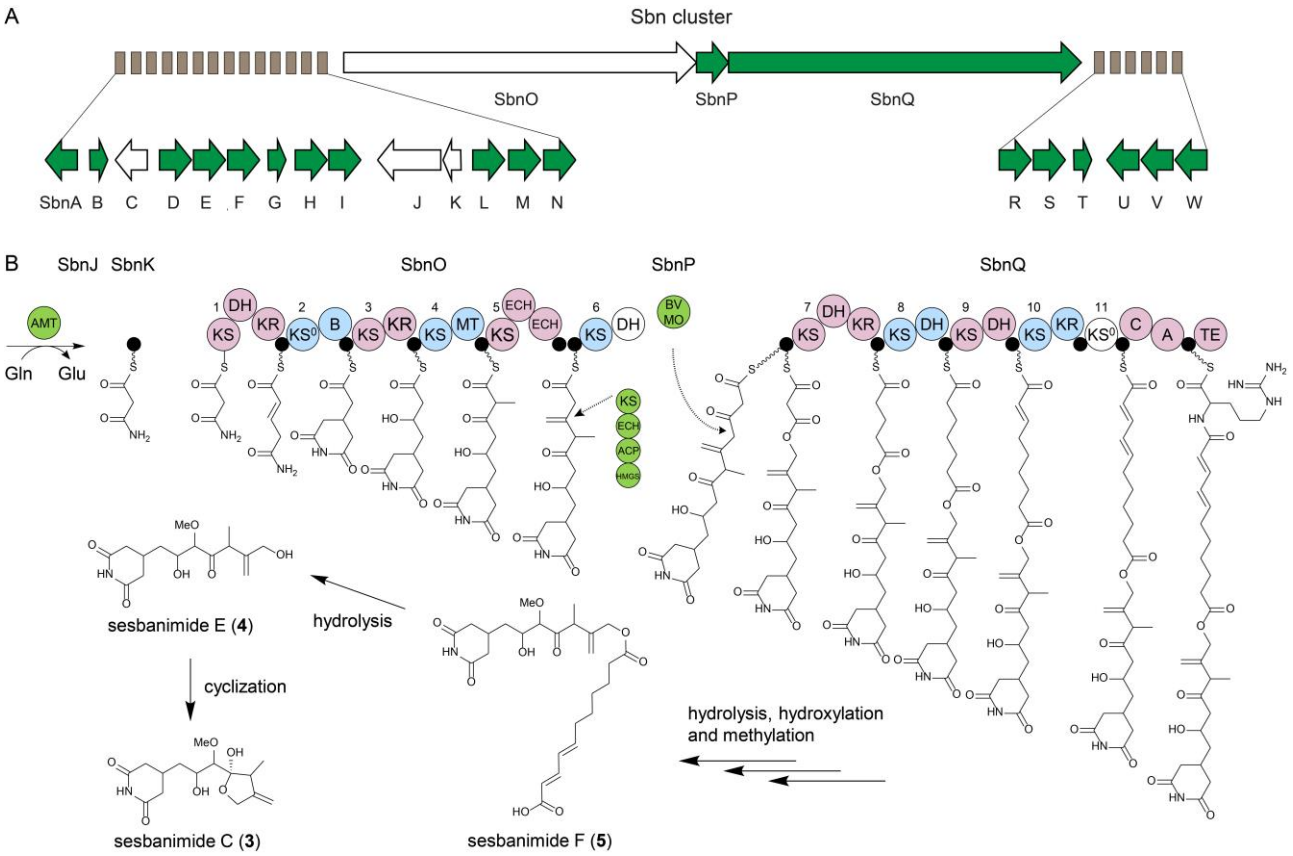


Figure 19. Proposed biosynthetic pathway for the assembly of sesbanimide polyketides. A) Genes comprising the *sbn* gene cluster for sesbanimide production. The homolog genes from the *lab* cluster are colored in green. B) Assembly of the polyketide core structure by PKS and PKS/NRPS multimodular complexes and *trans*-acting components: the loading module comprising amidotranferase (AMT) and associated acyl-carrier protein (ACP); β-branching cassette comprising hydroxymethylglutaryl-CoA synthase (HMGS), stand-alone ACP, ketosynthase (KS), enoyl-CoA hydratase (ECH); Baeyer-Villiger monooxygenase (BVMO). Other PKS domains indicated are ketoreductase (KR), dehydratase (DH), branching domain (B), non-elongating ketosynthase (KS^0), methyltransferase (MT), condensation domain (C), adenylation domain (A), peptidyl carrier protein (PCP), thioesterase domain (TE). *Trans*-acting proteins responsible for β-branching and oxygenation are indicated by discontinued arrows.

An unusual feature of the *sbn* PKS is the additional large protein SbnQ comprising four PKS and one NRPS module, which do not correspond to moieties of previously reported sesbanimides. However, the fortuitous discovery of sesbanimide F carrying an extended lipoid moiety provides a rationale for this feature. The internal ester moiety of this congener is collinear with a split module comprising the SbnOQ interface and the monooxygenase SbnP, a module type recently shown to insert oxygen by Baeyer-Villiger oxidation resulting in ester formation (Meoded et al., 2018). The lipid moiety of sesbanimide F matches well to the subsequent PKS modules, but the compound lacks the final incorporation of an amino acid suggested by the terminal NRPS module of SbnQ, hypothetically an arginine according to NRPS2 Predictor (Röttig et al., 2011).

2.3.2.1. PKS and hybrid PKS/NRPS similarities with other homologs

Surprisingly, a comparative analysis of the modular architecture of SbnO revealed that the N-terminal glutaramide-forming modules, alongside with the loading module, and the C-terminal modules have originated from different gene clusters from different bacterial phyla (Figure 20). The loading module composed of AMT and ACP proteins, denominated SbnJ and SbnK, respectively, and the first three modules of SbnO, are architecturally identical to previously identified glutaramide polyketide clusters of cyclohexamide (Yin et al., 2014), migrastatin, *iso*-migrastatin, dorrigocin (Lim et al., 2009), 9-methylstreptimidone (Wang et al., 2013) and lactimidomycin (Seo et al., 2014) as well as in glutaramide clusters of several actinomycetes and *Burkholderia* spp. genomes (Stulberg et al., 2016). On the other hand, the last two modules of SbnO, comprising distinctive tandem ECH and ACP domains, are absent in the glutaramide clusters from *Streptomyces* and *Burkholderia* genomes. Moreover, protein identity analysis of downstream SbnO domains revealed similarities with the pederin-family gene clusters of labrenzin, pederin (Piel, 2002; Piel et al., 2004d), nosperin (Kampa et al., 2013), and cusperin (Kust et al., 2018), as well as with the clusters of onnamide (Piel et al., 2004c) and oocydin A (Matilla et al., 2012).

Results

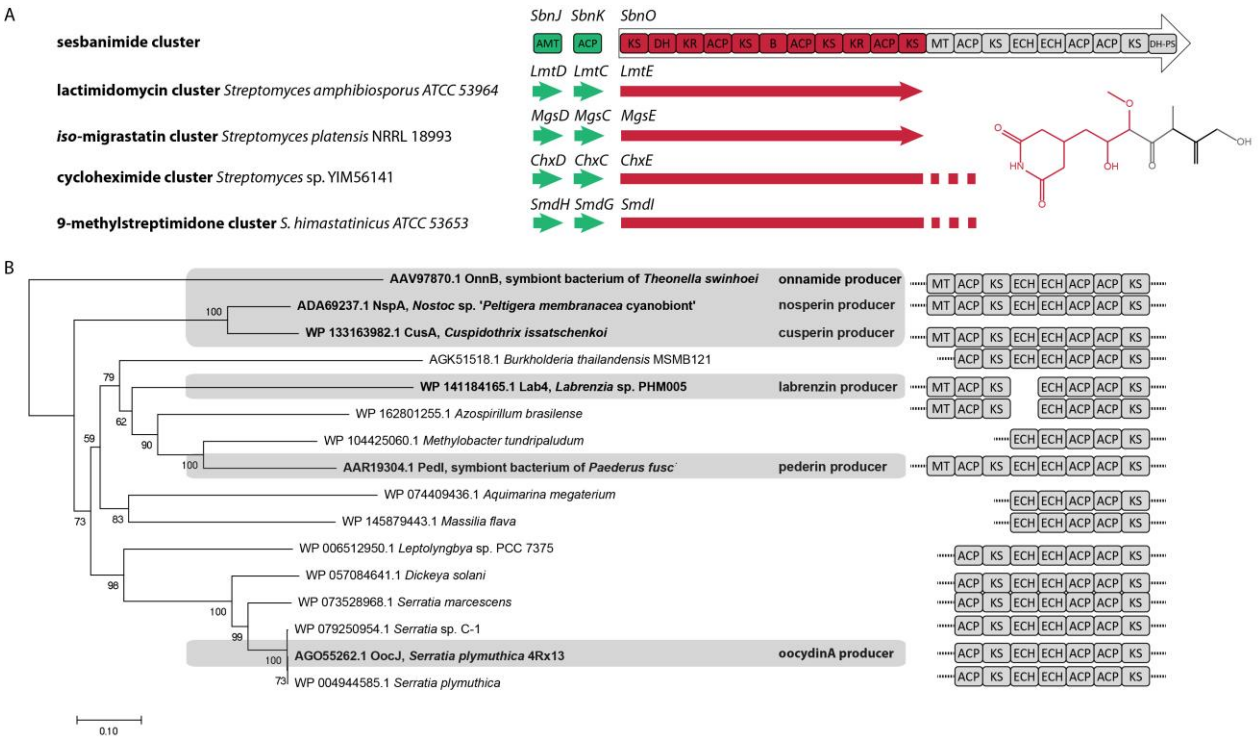


Figure 20. Protein identity and functionality comparisons of SbnO modules. A) Genetic scheme showing AMT, ACP and PKS from *sbn* cluster and characterized homologs in *Streptomyces* species. Sesbanimide C molecular structure (right) is shown in two colors; each color marks the moiety assembled on the respective PKS modules (the same color). B) Neighbor-joining tree based on the partial SbnO protein sequences included in the study (shadowed in grey) and the alignment with other homologous genes. Identified polyketides and their respective producers are indicated in bold. PKS domains (right) in homolog protein sequences are identified using TransATor (http://transator.ethz.ch).The scale shows evolutionary distances. Bootstrap values are indicated for each node.

The across-genomes screening for the SbnQ homologues showed their presence in distant PKS clusters producing structurally different polyketides lacking glutaramide moieties. To investigate the genome distribution of SbnQ, a Neighbor-joining (Saitou and Nei, 1987) a phylogenetic analysis of the closest homologues was performed (Figure 21). Assuming OnnJ as a common ancestor from the onnamide cluster of the symbiont bacterium of *T. swinhoei*, it was poosible to find a clade of pederin-family gene clusters producing different polyketides, *i.e.*, diaphorin in a bacterium symbiont of the plant pathogen *D. citri* (Nakabachi et al., 2013), pederin in a bacterium symbiont of the beetle *P. fuscipes* (Piel et al., 2004b), and labrenzin in *Labrenzia* sp. PHM005, as well as other orphan clusters of *Azospirillum brasilense* and *Methylobacter tundripaludum*. Homologues of SbnQ, like DipO, PedH and Lab13, respectively, are always present

as downstream PKS/NRPS in those clusters, forming hybrid polyketide complexes with divergent upstream PKSs.

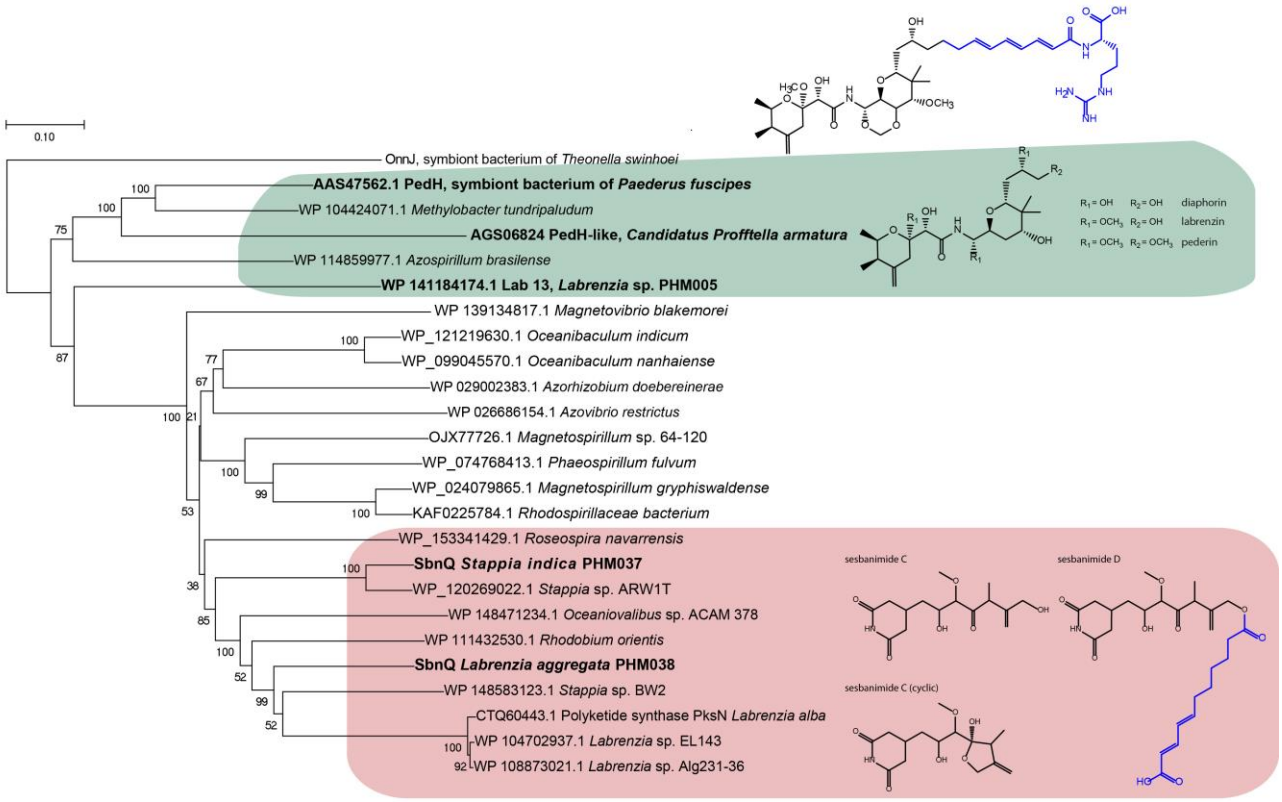


Figure 21. Neighbor-joining tree of SbnQ (PKS/NRPS) gene sequences, responsible for the assembly of the lipoid moiety of sesbanimide D (blue), and homologues. The characterized gene clusters are indicated in bold and their respective polyketide structures are shown (right). The scale shows evolutionary distances. Bootstrap values are indicated for each node.

Moreover, when the modular architecture of SbnQ was compated with Lab13 from the *lab* cluster (Figure 22), it was observed that the order of domains in the corresponding modules is nearly conserved in both PKS/NRPSs and matches the assembly of lipoid moiety of sesbanimide F in *S. indica* PHM037. The sole modular difference detected in Lab13 is the lack of the second KS-DH-ACP module, when compared to SbnQ. Such difference implies that the polyketide moiety in the missing labrenzin analog might be shorter by one malonate building unit. In addition, the final labrenzin product most probably is a similar ester-like polyketide.

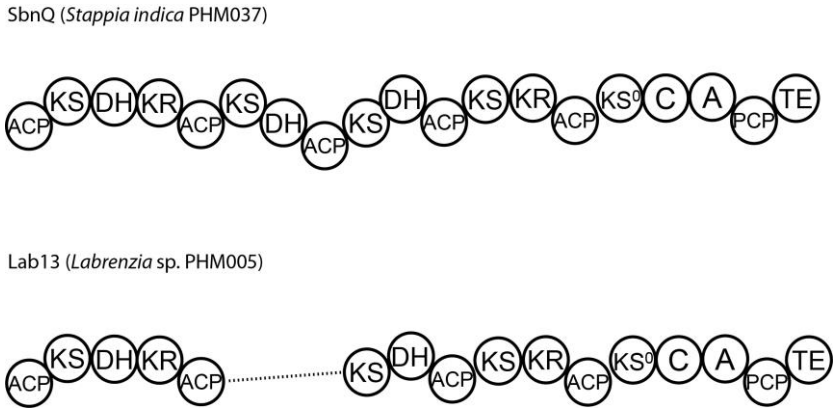


Figure 22. Scheme of the domain organization in Lab13 and SbnQ (PKS/NRPS).

3. Growth and transcriptional analysis

3.1. Growth properties of *Labrenzia* sp. PHM005

In order to get the insights into the growth and physiological characteristics of this marine strain, several growth media were compared. Our results showed that *Labrenzia* sp. PHM005 is able to grow in the rich medium Marine Broth (MB) and in the defined medium Marine Basal Medium (MBM) containing glucose as carbon source and supplemented with vitamins as described in Material and Methods (chapter 1.2.1.) (Figure 23A). Labrenzin was detected by mass spectrometry in both media within the first 24 h of growth (Figure 23B). Although labrenzin was detected in cultures of all tested media, biotin and thiamin are essential for the optimal growth of *Labrenzia* sp. PHM005.

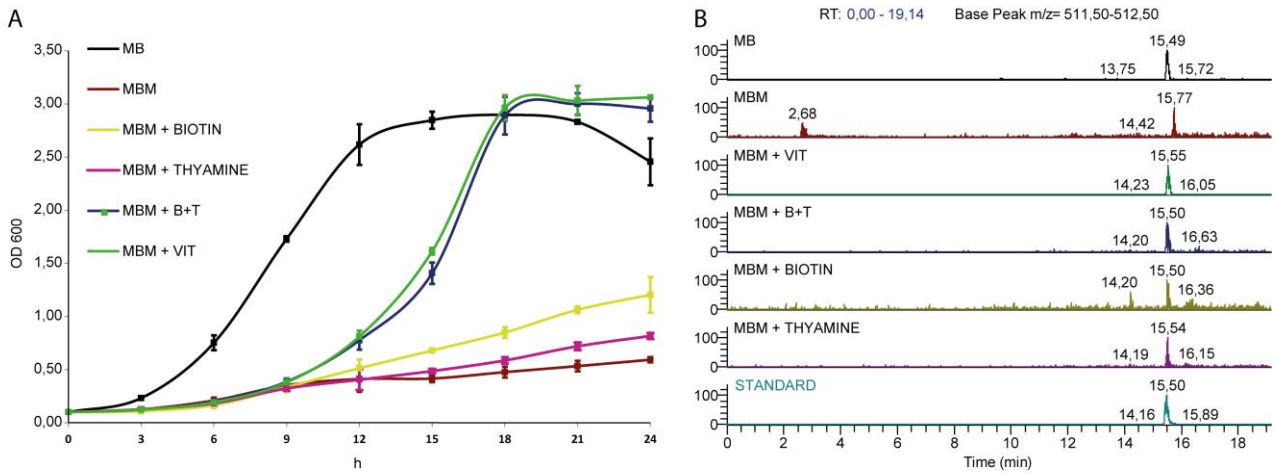


Figure 23. Growth curves of *Labrenzia* sp. PHM005 and extracted ion chromatograms of culture extracts. (A) Strain is grown in Marine Broth (MB) and modified Marine Basal Medium (MBM) supplemented with different vitamin mixtures as indicated in the legend. (B) MS chromatograms showing the labrenzin extracted from different media after 24 h of cultivation (from up to down): MB; MBM; MBM + VIT; MBM + B + T; MBM + Biotin; MBM + Thyamine; standard. Results represent the means of three replicates.

Additionally, it was tested the growth on different sources of carbon, as the sole carbon source, such as citrate, glutarate, benzoate and cyclohexane (Figure 24). The results showed that the strain PHM005 could only grow in citrate and glutarate, although it did not exceed an optical density of 1.

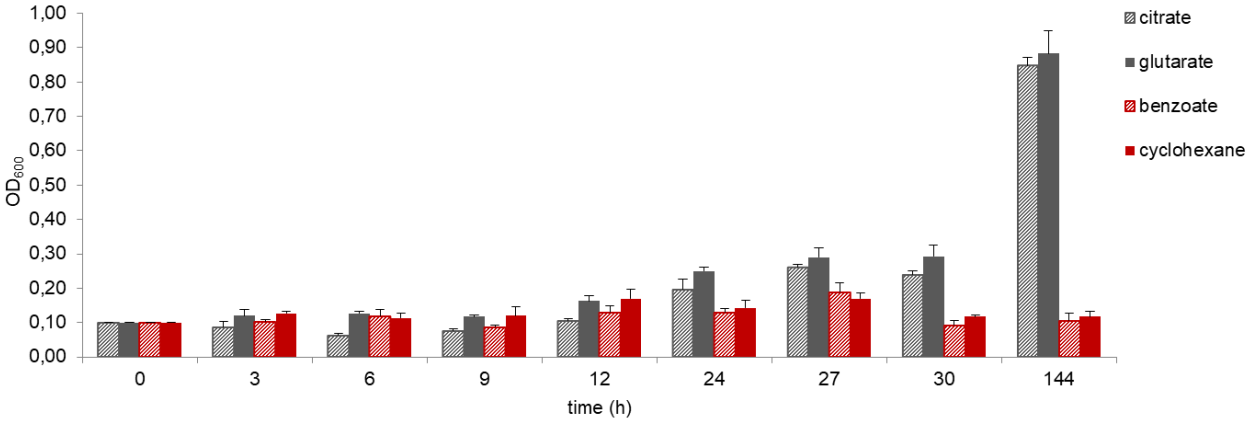


Figure 24. Growth monitoring of *Labrenzia* sp. PHM005 up to 144 h in MBM+VIT medium with different carbon sources. Results represent the means of three replicates.

Results

The growth requirement of NaCl was tested as an important growth factor for marine bacteria, starting from 20 g/L, an amount similar to that contained in commercial MB (19.45 g/L of NaCl). MB was used since this marine bacterium was unable to grow in the conventional LB medium currently used for *E. coli* in the lab. To determine the minimal NaCl concentration the cells were cultivated in the presence of increasing concentration of NaCl. The NaCl effect on the growth of *Labrenzia* sp. PHM005 is shown in Figure 25.

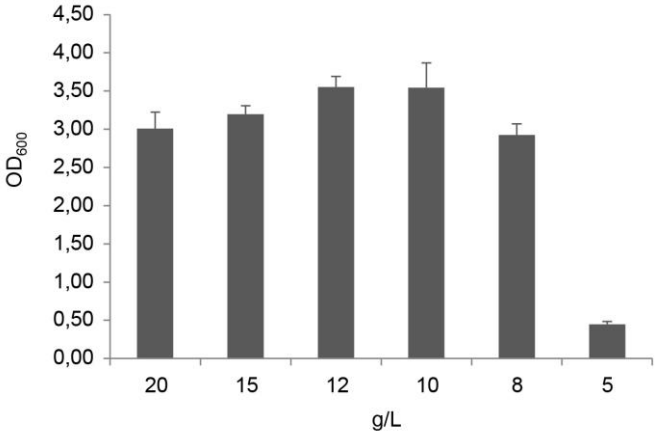


Figure 25. Growth of *Labrenzia* sp. PHM005 culture after 24 h in MBM+VIT medium supplemented with different concentrations of NaCl. Results represent the means of three replicates.

To conlcude, the optimal medium chosen for the successive cultivations of *Labrenzia* sp. PHM005 wt strain and mutant strains was MBM+VIT medium modified to contain 20 g/L of NaCl and 0.2% of glucose as carbon source; and M2 medium as a modified MBM+VIT medium with only difference being 2.0% of glucose.

3.2. Gene transcription analyses

3.2.1. Transcriptomics and identification of transcription initiation sites

The transcriptional activity of *lab* cluster genes was tested on MBM+VIT medium. Figure 26 shows the expression levels of *lab* cluster genes and the housekeeping genes in the early exponential phase. All *lab* cluster genes are expressed suggesting that these genes are activated during the exponential phase of growth and none of them remain silent. Interestingly, some of the hypothetical proteins were upregulated under this growing condition, such as Hyp4, Hyp5 and Hyp6. DUF697 domain-containing protein (Hyp4) is a small protein of 196 aa with uncharacterized

function. Pfam analysis revealed three transmembrane motifs (45-71 aa; 92-113 aa; 119-140aa) and IUPred2A web interface appointed to two protein disorder motifs at N and C-terminal side of the protein. Disordered segments characterize flexible conformations whose properties depend on environmental changes, as redox potential, or binding of other proteins (Mészáros et al., 2018). Hyp5 is a 289 aa prohibitin-family protein with SPFH/Band-7 family domain and a transmembrane motif (20-41 aa) according to Pfam database. This protein is phylogenetically related to stomatin-like integral membrane domain proteins (Tavernarakis et al., 1999). Another upregulated hypothetical peptide, Hyp6, is the smallest of only 65 aa and characterized with two transmembrane motifs (6-24 aa; 31-52 aa). In line with the domain analysis, these upregulated proteins might be involved in regulation of *lab* cluster or the polyketide transport.

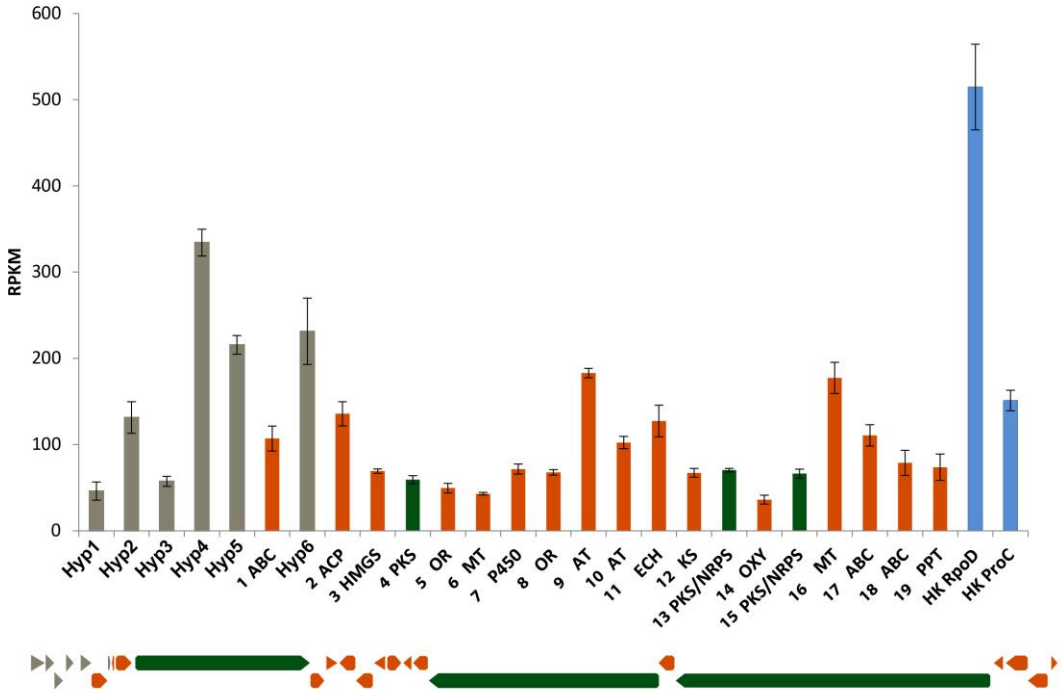


Figure 26. *Lab* cluster gene expression levels in early exponential phase cultured in MBM+VIT medium. The order of the bars follows the order of the genes (from left to right). The results represent the mean of three replicates. Blue bars represent housekeeping genes RpoD and ProC.

In addition to transcriptomics analysis, the cappable-seq method was employed for the identification of transcription initiation sites (TSS), or +1 sites, with the aim to localize the promotor regions within the *lab* cluster and confirm which genes were organized in operons. Table 20 shows the exact genome locations of the TSSs, strand, an approximate TATA box sequence

predicted by online tool BPROM (www.softberry.com) and manually inspected RBS. Figure 27 shows the scheme of the *lab* genes' transcription. Based on the TSS and genes localizations, the possible operons of the lab cluster would comprise: 1) Hyp2-Hyp3-Hyp4; 2) Hyp5-ABC1; 3) Hyp6-HMGS; 4) OR5-MT6; 5) AT9-OR8-P450; 6) PKS13-KS12; 7) PKS15-OXY14; 8) ABC18-ABC17.

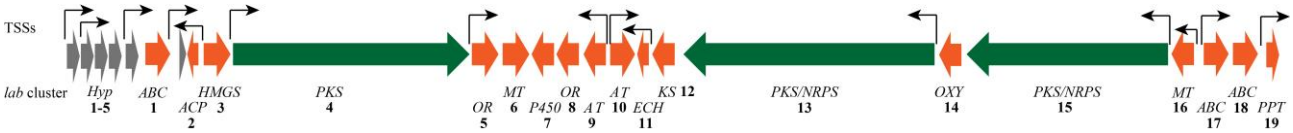


Figure 27. Scheme representing the lab cluster and TSSs. Direction of the transcription is shown by the direction of the arrow.

Table 20. Cappable-seq identification of transcription initiation sites (TSSs) and bioinformatics predictions of RNA polymerase binding sequences and ribosome binding sequences (RBS).

Genes	TSS	Strand	-10 box (score)	-35 box (score)	RBS (position)
Hyp1	776,746	+	ATGTTTAAT (44)	TTGCCT (56)	AGAACGGG (-4)
Hyp2/	777,937	+	TTGTACTAT (57)	TTGCTC (31)	**GAGAA**TCAGC**AGAG**CC**AAGAA** (-4) /
Hyp3					**AAAA**C**GGG**CGCT**GAG** (-4)
Hyp4					AAAACAGGAGA (-6)
Hyp5/	780,616	+	TGTTATCCT (65)	AAGCAA (12)	GGACG (-4) /
ABC1					GGCGAAG (-1)
ACP2	783,366	-	TGCTTCAAT (34)	CTGCCG (20)	**AAGAA**CAACCCC**AGGGA**C**GCAA**CA (-3)
Hyp6/	782,594	+	AGACAAACT (34)	TCGAAA (23)	AAAGGAGA (-8)
HMGS3					AGGGAGAGTG (-7)
PKS4	784,686	+	GGATATCAG (23)	TTGCCG (55)	**AGAG**C**AAAAGAG**CC**GGG** (-10)
OR5-	798,420	+	ATCGATCAT (22)	ATGACG (30)	AAAAAGCACGGGAAGCGG (-2) / AGCAGGAAA
MT6					(-7)
AT9/	804,336	-	CCTTACAGT (35)	TTCAAA (40)	GGAC (-8) /
OR8/					GAGGCGG (-5) /
P4507					**GGGAG**C**GCAA** (-5)
AT10	804,396	+	CGCTATCAT (66)	GTGAGT (5)	**AGG**TTC**GAG** (-1)
ECH11	806,354	-	ATTTAATCT (50)	GTGCCT (23)	AGGG (-1)
PKS13/	825,420	-	CAGGATGAT (24)	TTGTCG (47)	**GAAAG**CCC**AGG** (-1) ; GGGAGATG (-5) ;
KS12					**GAAG**CC**GAG** (-4) ***** / AACAACAG (-4)
PKS15/	851,097	-	GGCCAGGAT (26)	TTGCTC (31)	GGAAAG (-7) /
OXY14					**GGGAGGG**CGTT**GAGG** (-2)
MT16	852,103	-	TGTTAGGAT (60)	TTGCCA (61)	AAGGAAAAGA (-4)
ABC18/	855,618	-	/	/	GGGGG (-5)

Genes	TSS	Strand	-10 box (score)	-35 box (score)	RBS (position)
ABC17			AGCTATAAA (45)	ATGATG (22)	**GAAGGAAGGGA**CAT**GAAAA** (-1)
PPT19			CGGTACTAT (60)	TTCTAA (27)	GGAT (-8)
***alternative RBSs**					

3.2.2. Activity of the P_n (*lab4*) promoter in the presence of chemical elicitors

Among different approaches for inducing the silent gene cluster, sub-inhibitory concentrations of antibiotics showed positive effect on various metabolic phenomena, such as transcription activation, activation of the quorum sensing regulators, biofilm production or activation of the biosynthesis of the cryptic natural products (Zarins-Tutt et al., 2016). During this work it was established a screening method in *Labrenzia* sp. PHM005 for the effectors of the promoter activity. For the proof of concept, the promoter P_{14g} from pSEVA227M_p14g (kind gift from Gonzalo Durante) was replaced with a native promoter (P_{n4}) from the *lab* cluster (*lab4*), situated between *lab3* and lab4 gene (see Figure 27), to construct a pSEVA227M_Pn4.

Plasmid pSEVA227M_p14g has a few convenient updates for the screening process (Figure 28). The strong promoter P_{14g} is a derivative of P_{Lac} promoter engineered to be unaffected by LacI represor. AG-rich bicistronic RBS was designed to ensure improved translation of GFP. Finally, the monomeric super folder Green Fluorescent Protein (msfgfp) was especially engineered to be able to assemble correctly in adverse conditions. All three factors contribute to emitting stronger fluorescence.

Figure 28. Scheme showing DNA extraction from the pSEVA227M_p14g zooming the promoter $P_{14g,}$ bicistronic linker and partial sequence of the GFP (left to right).

As previously shown (Figure 26), the transcription level of *lab4* gene encoding the first PKS responsible for the intiation of labrenzin biosynthesis was relatively low under cultivation in MBM+VIT (0.2% glucose) medium. The fluorescence of the GFP expressed from P_n and P_{14g} promoters (Figure 28) was even visible at the white light from the pellets of transforrmed *Labrenzia* sp. PHM005 (Figure 29A) and the fluorescence was largely emphasized under purple light (Figure

29B). Purple light, as a part of visible light, does not have damaging effect on the cells, therefore is a great tool for manual screening directly from the plates or the culture pellets.

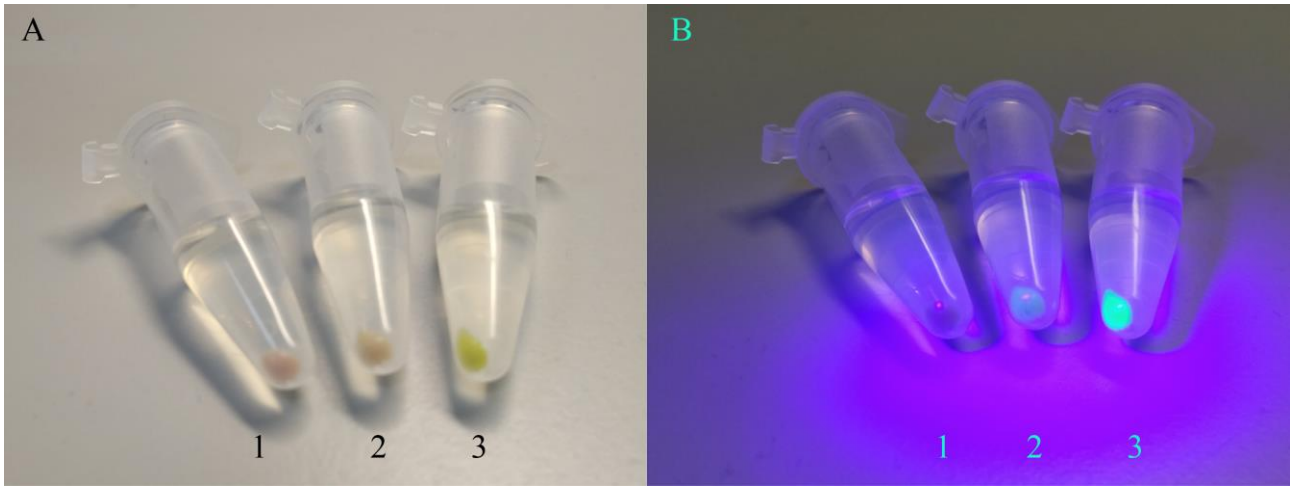


Figure 29. GFP expression in *Labrenzia* sp. PHM005 (pSEVA227M_Pn4) (under P_n promoter) (2) and *Labrenzia* sp. PHM005 (pSEVA227M_p14g) (under P_{14g} promoter) (3). *Labrenzia* sp. PHM005 (1) was used as a control. A) white light; B) blue light.

Some toxic compounds added in sub-inhibitory concentrations do not inhibit the growth of bacteria but might increase the levels of gene expression as an indirect consequence of intervening in the other metabolic pathways (Goh et al., 2002). As a proof of concept, different antibiotics such as trimethoprim, piperacillin, tylosin tartrate, and cephaloridine, were used. The cultivation of *Labrenzia* sp. PHM005 (pSEVA227M_Pn4) strain in 10 mL of M2 medium was monitored every 24 h during 4 days. The expression difference between the different conditions faded as the bacterium entered the stationary growth phase somewhere between 48 h and 72 h of cultivation (Figure 30A). All compounds tested significantly increased the transcriptional activity of the P_{n4} promoter ($p<0.05$) at 24 h of cultivation, *i.e.*, in the exponential growth phase, with trimethoprim eliciting the best significant effect ($p<0.005$) (Figure 30B).

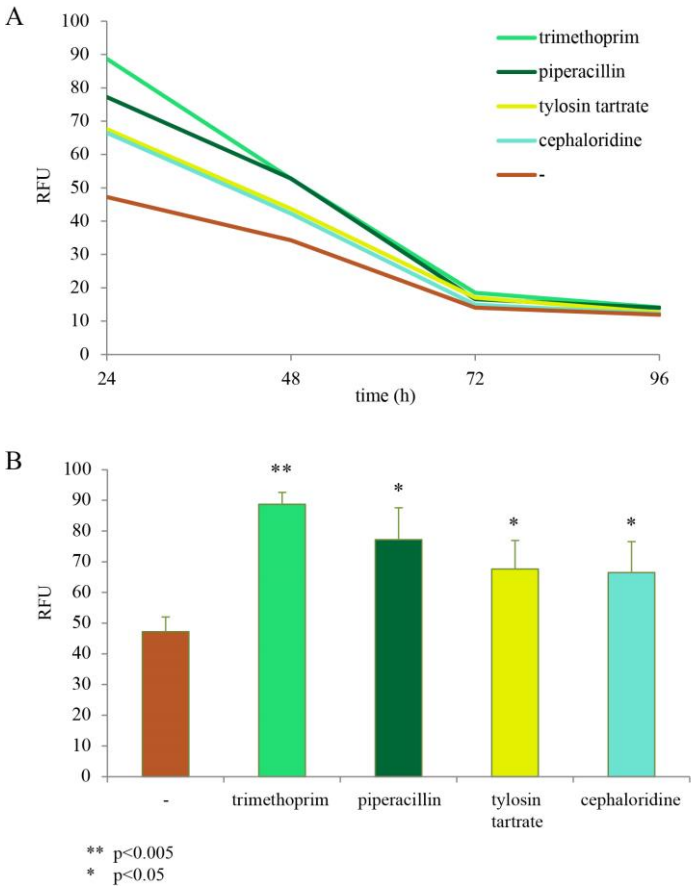


Figure 30. The transcriptional activity of the P_{n4} promoter expressed as relative fluorescence in the presence of trimethoprim, piperacillin, tylosin tartrate and cephaloridine during the cultivation of *Labrenzia* sp. PHM005 (pSEVA227M_Pn4) in M2 medium. A) GFP expression during 96 h of cultivation. B) GFP expression at 24 h. Significant data is considered for *p<0.05 and **p<0.005. Data represents the means of three replicates.

These results not only show the modifications of the transcriptional activity of P_{n4} promoter in the presence of the elicitors but also reveal that this promoter has the highest activity during the exponential growth phase.

The next step was to measure the labrenzin production in the wt upon the addition of the elicitors in order to investigate the correlation between the gene transcription and the polyketide production. Figure 31 shows that the production of labrenzin was not significantly increased in the presence of the elicitors as compared to the control.

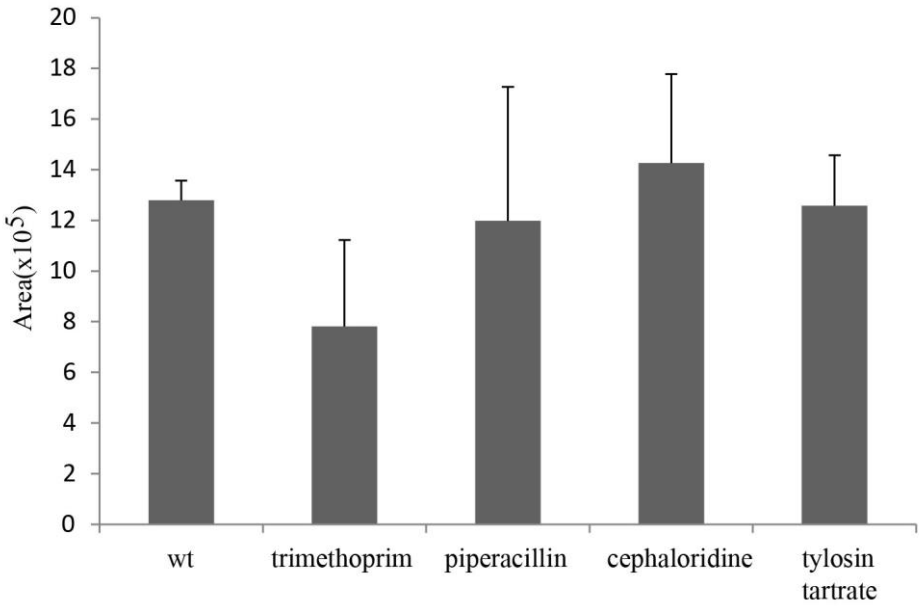


Figure 31. Production of labrenzin in *Labrenzia* sp. PHM005 after 72 h of cultivation in M2 medium expressed as the peak area obtained from MS. The means of three replicates are shown.

4. Functional genetic study of the *lab* cluster

A putative gene cluster responsible for the synthesis of labrenzin, the *lab* cluster, has been identified and characterized *in silico* (see Results chapter 2), showing that it encodes a *trans*-AT mixed type PKS/NRPS biosynthetic pathway that is almost identical to gene clusters responsible for the synthesis of pederin-family compounds. The cluster consists of a 79-kb region comprising three genes encoding multidomain hybrid polyketide synthase/non-ribosomal peptide synthetase (PKS/NRPS) proteins (Lab4, Lab13, and Lab15), and 16 auxiliary enzymes (Table 18). Since this is the first time a pederin gene cluster is identified in a cultivable bacterium, one of the main objectives of this doctoral thesis was to functionally characterize the *lab* cluster genes.

4.1. Assembly of the polyketide backbone

Firstly, it can be important to understand the polyketide backbone synthesis and to determine the key biosynthetic enzymes. Moreover, it was interesting to find the final product of the polyketide biosynthesis and identify the polyketide processing enzyme.

4.1.1. Functional assignation to *lab* gene cluster

To investigate the association between the described gene cluster and the production of labrenzin a gene knockout was perfomed by homologous recombination where the sequence comprising the putative promoter region of the gene encoding PKS Lab4 and the region encoding the GNAT domain of the first module of Lab4 was deleted to generate a non-producing mutant Δ*lab4*. This strain can be also used as a negative control strain for future experiments.

The analyses of the culture extracts of the wt and mutant strain are shown in Figure 32A–C. As expected, the mutant strain was unable to produce labrenzin. The gene expression analysis performed with the mutant strain suggested that the deletion produces a polar effect on the expression of the genes encoding the putative oxidoreductase (Lab5) and the methyltransferase (Lab6) that are located downstream of the *lab4* gene encoding a PKS, suggesting that the three genes are transcribed as an operon (Figure 32D).

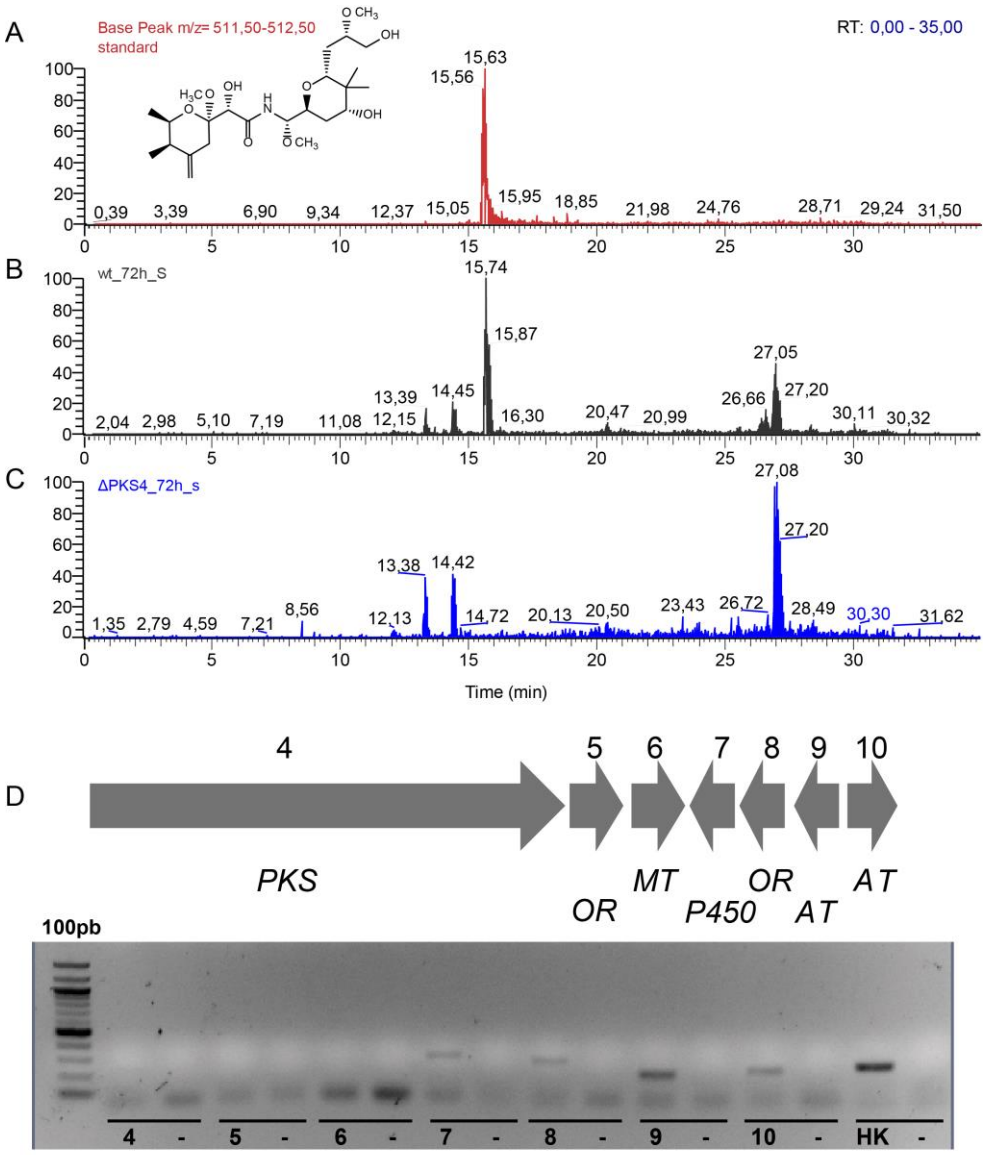


Figure 32. Extracted ion chromatograms of the supernatant extracts obtained after 72 h of cultivation of *Labrenzia* sp. PHM005 in MBM+VIT medium: A) labrenzin standard; B) wt; C) Δ*lab4* mutant. D) RT-PCR analysis of the mRNA expression of *lab* gene cluster in the Δ*lab4* mutant cultured in MBM+VIT medium. Numbers represent genes from the cluster. Each control lane (-) contains a sample lacking the reverse transcriptase.

In the early attempts to clone the entire *lab* cluster for its heterologous expression in other model microorganism a 16 kb fragment of the cluster comprising operon *lab4-lab5-lab6* has been amplified by PCR and cloned into pSEVA338 plasmid (see Material and Methods section 7.2.1). The resulting plasmid was named pSEVA338_*lab4-lab5-lab6*. This 16-kb construct can be used to complement the non-producing Δ*lab4* mutant strain and evaluate the expression and production of such a complex protein in a plasmid. For that purpose, an additional modification of the plasmid was made by switching the original P_m promotor of pSEVA338 plasmid for the native *lab4* P_{n4} promoter. The resulting plasmid was named pSEVA338_*Pn_lab4-lab5-lab6*.

The experiment was set up considering two different physical conditions, *i.e.*, temperature and shaking. The three PHM005 strains used in this experiment were: 1) wt; 2) Δ*lab4* (pSEVA338_*Pn_lab4-lab5-lab6*); 3) Δ*lab4* (pSEVA338_*lab4-lab5-lab6*). Strains were cultured in MBM+VIT medium as described in Material and Methods 1.3. The strain Δ*lab4* (pSEVA338_*lab4-lab5-lab6*) was cultivated with or without the addition of inducer 3-methyl-benzoate (0.2 mM) at the exponential phase.

Figure 33 shows the relative abundance of labrenzin in different conditions. The results indicate that the production of labrenzin is somewhat better in the strains where *lab4-lab5-lab6* operon is expressed from the native P_{n4} promoter. Nonetheless, the expression in the complemented strains did not reach the labrenzin levels produced in the wt strain. The results observed under higher temperature and shacking showed a very low labrenzin production in all complemented strains when compared to the production of wt. It is worth to mention that the labrenzin production by wt was higher under higher temperature and shacking (1.15E5 peak intensity) compared to the condition of lower temperature and shacking (6.39E4 peak intensity).

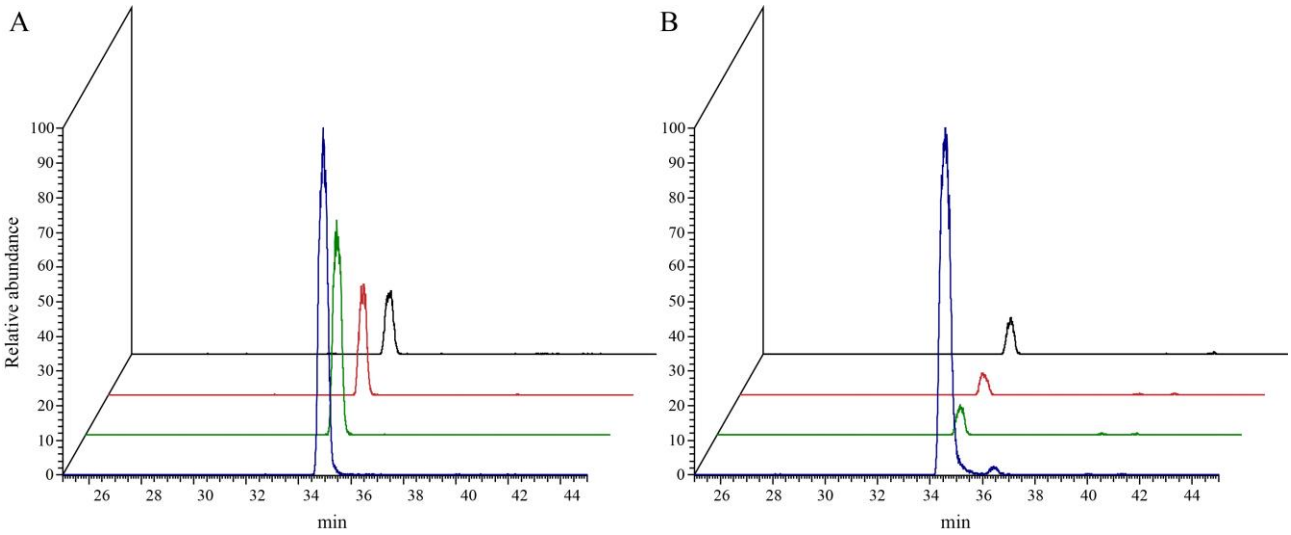


Figure 33. Production of labrenzin expressed as a relative abundance of MS extracted m/z = 512 ion peak intensity in different conditions. A) 72 h cultivation of strains in MBM+VIT with rotation set to 2.5 (New Brunswick Scientific Co. Gyrotory Water Bath Shaker Model G76) and 19 °C. B) 72 h cultivation of strains in MBM+VIT with shaking at 200 rpm (New Brunswick Innova® 44/44R) at 30 °C. Labrenzin production by: wt *Labrenzia* sp. PHM005 (blue), mutant Δ*lab4* (pSEVA338_*Pn_lab4-lab5-lab6*) (green), mutant Δ*lab4* (pSEVA338_*Pm_lab4-lab5-lab6*) not induced (red), mutant Δ*lab4* (pSEVA338_*lab4-lab5-lab6*) induced (black).

After sequencing and analyzing the the cloned region, one-point mutation was observed in PKS Lab4 coding gene resulting in the change of a threonine (codon ACC) by an isoleucine (codon ATC), but as this residue is not part of any conserved domain of the PKS Lab4, it should not be considered significant.

4.1.2. *Trans*-acyltransferases

The functional PKS type I modules consist of catalytic domains that are essential for the assembly and elongation of the polyketide, like KS, ACP and AT (see Introduction chapter 2.1.1). In *trans*-AT PKS, though, ATs are individual enzymes that interact with the modular ACP domains by selecting and loading the appropriate acyl-unit *in trans*. The *lab* cluster consists of two *trans*-ATs whose function was addressed through mutagenesis experiments.

4.1.2.1. *Trans*-AT/hydrolase Lab9 mutation

Lab9 is a *trans*-AT enzyme annotated as an acyltransferase domain-containing protein. Its homology with PedC suggested that it might play the same role as PedC as a broad substrate-editing enzyme in labrenzin biosynthesis. On the other hand, Lab10, an ACP-S-malonyltransferase, was hypothetized to be a malonate-unit loader. In order to test that hypotesis in labrenzin biosynthesis, the *lab9* gene was inactived and complemented *in trans*.

The *lab9* gene was deleted by homologous recombination generating the mutant strain Δ*lab9* to test whether the coding *trans*-AT was essential for the labrenzin biosynthesis. The labrenzin production was reduced about one order of magnitude in the mutant strain (Figure 34A). Surprisingly, the complementation *in trans* of the mutant strain Δ*lab9* with plasmid pSEVA227M_*lab9* did not restore the labrenzin levels of wt (Figure 34B).

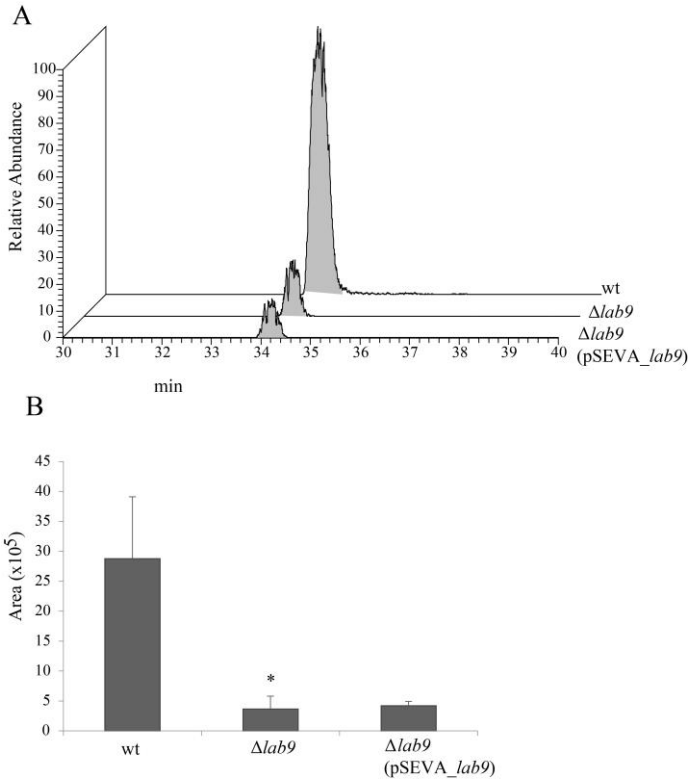


Figure 34. Labrenzin production in PHM005 wt, and mutant strains Δlab9 and Δlab9 (pSEVA227M_lab9). A) MS ion extraction (m/z =512) showing the relative abundance of labrenzin produced by the strains after 72 h of cultivation in M2 medium. B) Abundance of labrenzin expressed as peak area obtained from the MS spectra. Results represent means of three replicates. * Significant data are considered for $p<0.05$.

4.1.2.2. Simultaneous overexpression of Lab9 and Lab10 in the wt strain

In order to improve labrenzin yields, both ATs (Lab9 and Lab10) from the *lab* cluster were cloned in a promoter-less middle-copy plasmid pSEVA331 yielding plasmid pSEVA331_AT9-10 that was subsequently transferred to the wt strain. The region (2382 bp) comprising the *lab9* and *lab10*, that are transcribed in opposed directions, containing their respective promoter regions was amplified by PCR and cloned in plasmid pSEVA331 to create plasmid pSEVA331_*lab9_lab10*. The culture supernatants were collected after 72 h of cultivation in M2 medium followed by the extraction and HPLC/MS analysis for the production of the polyketides. The results are shown in the (Figure 35). The recombinant strain wt (pSEVA331_*lab9-10*) did not show any significant increase in the production of labrenzin **1** and labrenzin analogs **2** and **3**, as identified previously by PharmaMar (personal communication) when compared to the wt, but was somewhat higher when compared to the recombinant wt (pSEVA331) strain habouring the empty plasmid.

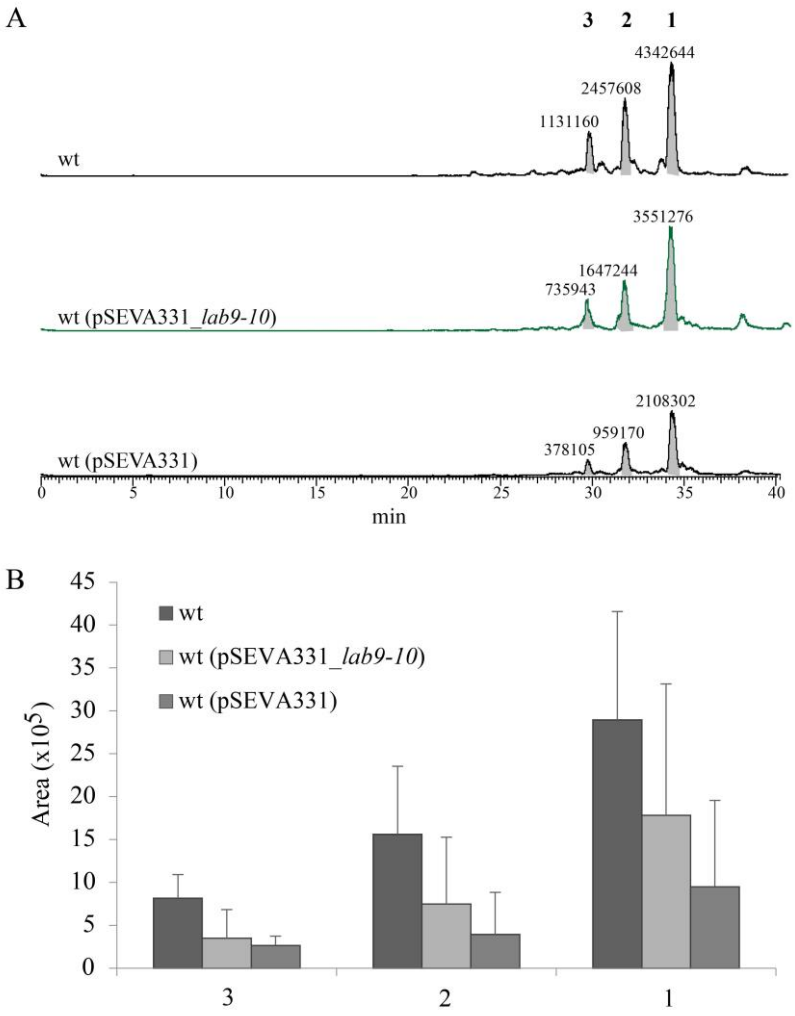


Figure 35. Production of labrenzin analogs in PHM005 strains wt, wt (pSEVA331_*lab9-10*) and wt (pSEVA331). A) MS chromatograms showing extracted ions in the range of m/z = 468-512 and intermediates 1 (m/z = 468), 2 (m/z = 498) and 3 (m/z = 512) of labrenzin biosynthesis obtained from the culture extracts after 72 h of cultivation in the M2 medium. The area of each peak is indicated above. B) Abundance of labrenzin analogs expressed as peak area obtained from the MS spectra. Results represent means of three replicates.

4.1.3. Cryptic polyketide moiety synthesis

The *lab* cluster contains one PKS, Lab4, and two hybrid PKS/NRPSs, Lab13 and Lab15. While Lab 4 and Lab15 are responsible for the biosynthesis of the labrenzin structural core, the polyketide moiety expected to be synthetized by Lab13, has remained cryptic until now. Interestingly, the Lab13 homologs are conserved in numerous orphan gene clusters distributed across bacterial phylogeny (Figure 21). Part of the corresponding polyketide moiety was identified

in sesbanimide F produced by *S. indica* PHM037 during this doctoral thesis and according to the conserved domain architecture between Lab13 and SbnQ (Figure 22), it should be identified some labrenzin analog intermediates in *Labrenzia* sp. PHM005 cultures similar to onnamide or similar to sesbanimide F.

4.1.3.1. Identification of labrenzin in the cytosol

Since it is only possible to predict the molecular structure of the unidentified labrenzin intermediate (onnamide- or sesbanimide F-like intermediates) by computation, without knowing the exact molecular mass and fragmentation pattern, its ion extraction in total MS scans is unreliable. Assuming that the final molecule (onamide-like) is completely synthetized by PKS, released and later processed by hydrolysis into at least two molecules, labrenzin and a residual moiety, several possible scenarios can be proposed.

One scenario is that the whole molecule was fragmented in two parts during its transportation outside the cells as a result of transport cleavage. The residual moiety could remain bound to the cell membrane, which can explain the absence of the residual moiety in the culture extracts. In that scenario, labrenzin would only be detectable in the culture supernatant, but not in the internal cytosol.

A second scenario can be that whole molecule is processed inside the cells and thus, labrenzin should be detectable in the cytosol.

In order to test these hypotheses, the previously constructed non-producing mutant Δ*lab4* was used as a control together with wt strain to analyze the presence of labrenzin in the cytosol by HPLC/MS. A purified labrenzin was also added to the Δ*lab4* cultures to check if labrenzin can be uptaken by the cells. The extraction of intracellular metabolome was performed after cell rupture by applying high-pressure as described in Materials and Methods 12.1. Figure 36 shows that labrenzin can be detected in the wt cytosol indicating that the large polyketide product is likely hydrolyzed intracellularly. The absence of labrenzin in the cytosol of Δ*lab4* when it was added to the culture medium suggests that labrenzin cannot be uptaked by the cells. Thus, the possibility that the detected cytosolic labrenzin could be the result of the previous secretion and a futher uptake was discarded.

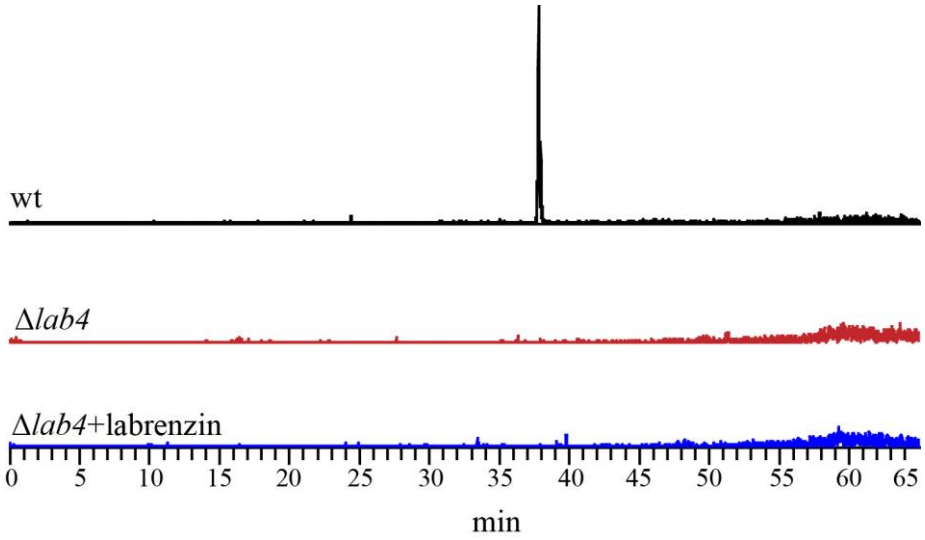


Figure 36. MS ion extraction m/z = 512 of different intracellular extracts obtained after 72 h of cultivation in MBM+VIT medium of wt, Δ*lab4* mutant and Δ*lab4* mutant supplemented with pure labrenzin (Δ*lab4*+labrenzin).

4.1.3.2. Functional analysis of *lab13* gene

It has been hypothesized that the complete onamide-like product has to be synthetized and released from the enzymatic complex, to be further processed into labrenzin and the other residual moiety. However, another possibility could be that labrenzin could be released from the enzymatic complex before the whole molecule is synthesized as the result of a translocation fail of the intermediate between Lab15 and Lab13 PKS/NRPs units. If this was the case perhaps, the final Lab13 PKS/NRPs unit could not be necessary to produce labrenzin. Even more, the absence of Lab13 unit could favor such fail and further increase the production of labrenzin.

In order to test this hypothesis, a mutant strain Δ*lab13* was constructed carrying a complete deletion of *lab13* gene. The culture supernatants were collected after 72 h of cultivation in M2 medium followed by the extraction and HPLC/MS analysis for the production of the polyketides. Figure 37 shows that Δ*lab13* mutant lacking the third PKS/NRPS unit does not produce labrenzin nor the other two labrenzin analogs identified previously by PharmaMar (personal communication). This result demonstrate that Lab13 is required for the synthesis of labrenzin, but in addition it suggests that most probaly the onnamide-like polyketide has to be synthetized prior its processing into labrenzin.

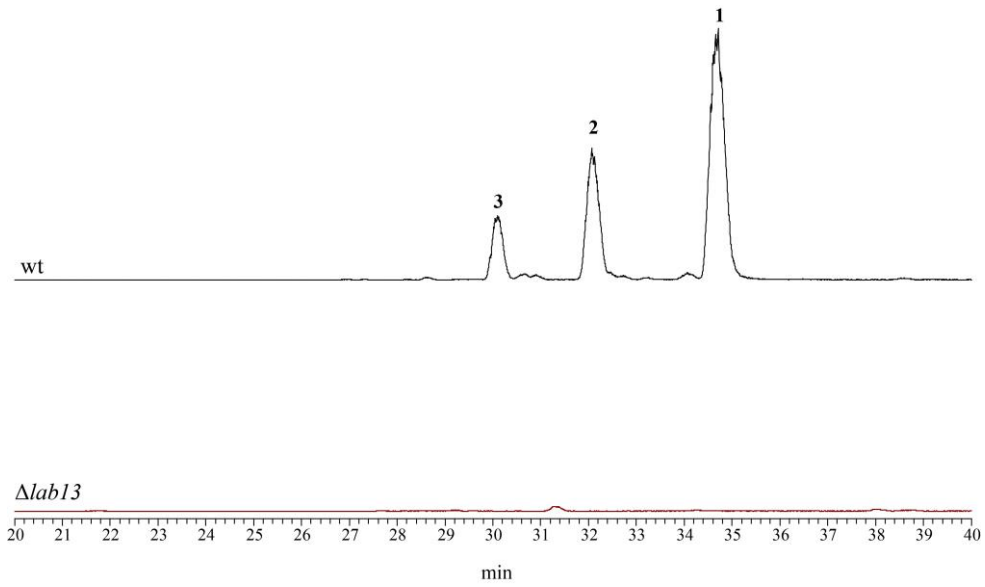


Figure 37. MS chromatograms showing extracted ions in the range of m/z = 468-512 with intermediates 1 (m/z = 468), 2 (m/z = 498) and 3 (m/z = 512) of labrenzin biosynthesis obtained from the wt and Δ*lab13* mutant culture extracts after 72 h of cultivation in M2 medium.

4.1.3.3. Cryptic labrenzin intermediate screening

Labrenzin absorption at any of the tested wavelengths (210 nm, 215 nm, 254 nm, 260 nm, 300 nm and 350 nm) is very low. This is the reason why labrenzin and other intermediates can be only monitored using mass spectrometry. On the other hand, molecules with higher content of double bonds like onnamide-like polyketides (Matsunaga et al., 1992) generally present acceptable adsorption levels at wavelengths of 260-350 nm. In order to find onnamide-like molecules the UV chromatograms of wt and Δ*lab13* culture extracts were compared (Figure 38 and Figure 40).

A new different peak was detected at 260 nm at 41 min in the wt extract, as compared to Δ*lab13* extract, indicating a more hydrophobic molecule than the known labrenzin intermediates (Figure 39). Interestingly, the peak was not visible in the extract of Δ*lab13* mutant strongly indicating it could be related to labrenzin biosynthesis. However, there is no difference between the strains analyzing the spectra at 210 nm and 350 nm (Figure 38 and Figure 40, respectively).

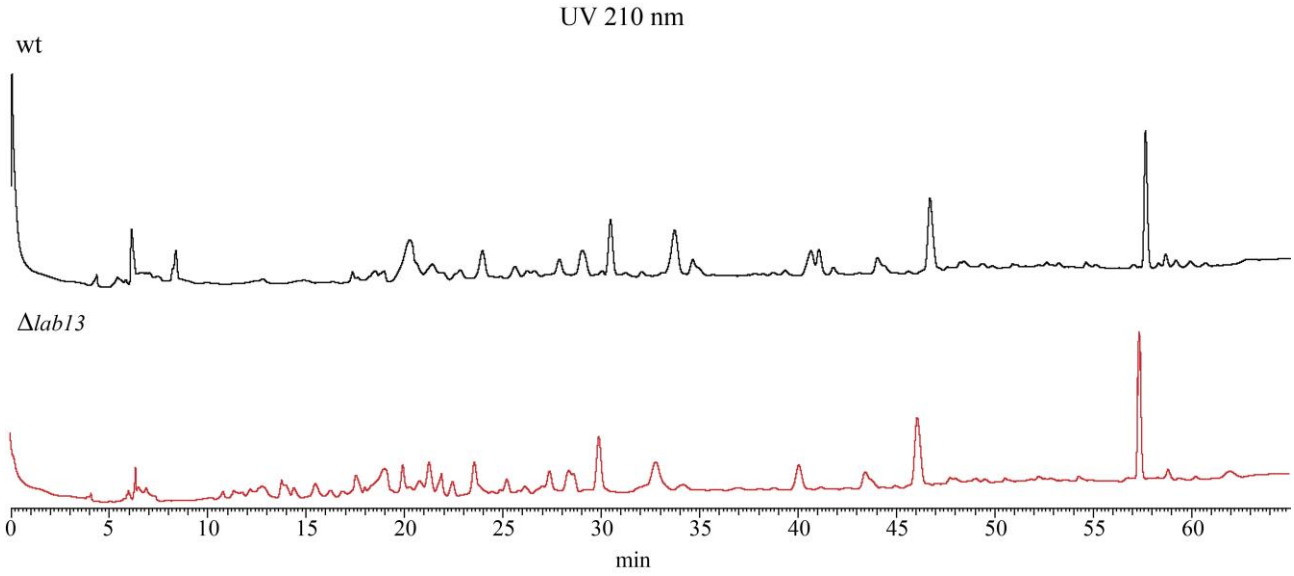


Figure 38. UV chromatograms at 210 nm of wt and Δ*lab13* mutant culture extracts obtained after 72 h of cultivation in M2 medium.

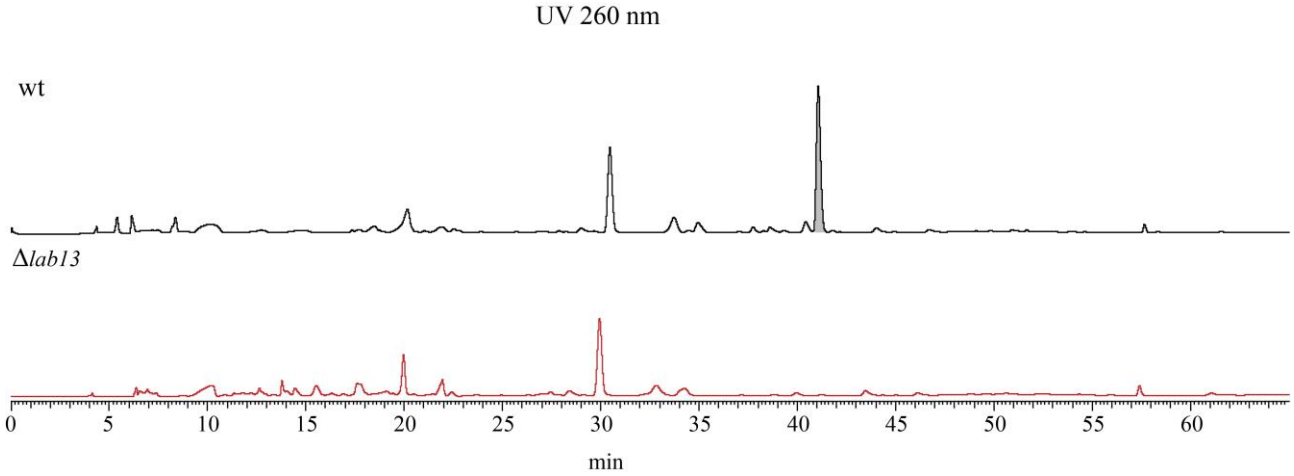


Figure 39. UV chromatograms at 260 nm of wt and Δ*lab13* mutant culture extracts obtained after 72 h of cultivation in M2 medium.

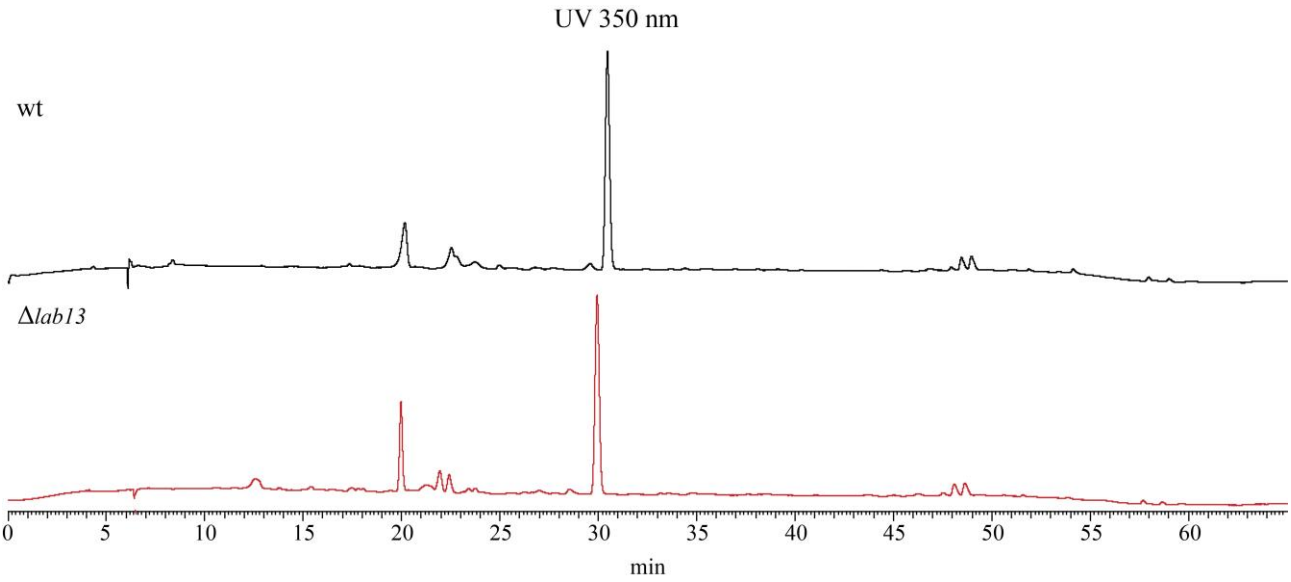


Figure 40. UV chromatograms at 350 nm of wt and Δ*lab13* mutant culture extracts obtained after 72 h of cultivation in M2 medium.

MS spectrum showed in Figure 41 revealed a peak at the same retention time with a dominant ion m/z = 840 and a less intensive ion m/z = 389. The chemical structure of the unknown molecule that would correspond to a molecular mass (in addition with Na^+) of 840, is proposed although there may be other possibilities including a longer side chain or even more double bond content.

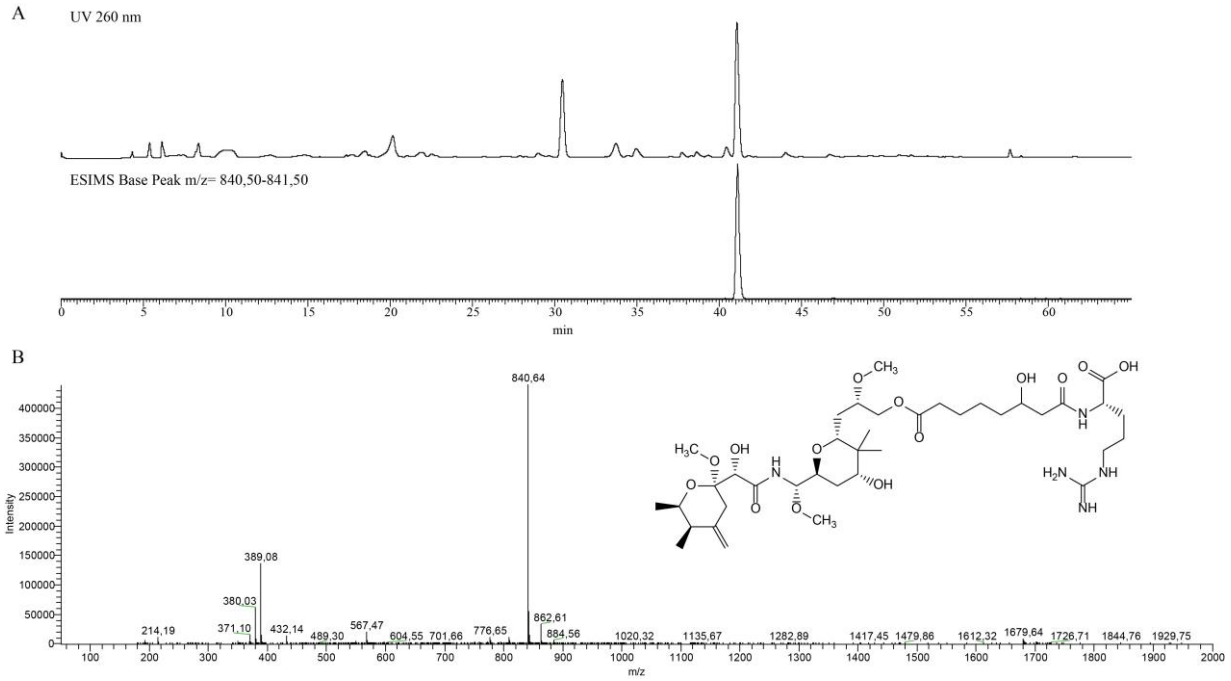


Figure 41. HPLC/MS analysis of the wt culture extracts obtained after 72 h of cultivation in M2 medium. A) UV chromatogram at 260 nm and MS extraction of ion m/z = 840. B) Ion fragmentation of the extracted peak and the possible chemical structure of the molecule.

4.1.3.4. Functional analysis of the thioesterase domain (TE)

Although previous results indicated the third PKS/NRPS Lab13 was essential for the production of labrenzin, bearing in mind that almost the complete gene was eliminated (≈16 kb), it casted doubt on a correct enzymatic complex assembly due to the absence of the last unit. This is, the Lab13 deletion might had prevented or altered the correct formation of the partial multienzymatic complex bewteen the PKS Lab4 and PKS/NRPS Lab15 units.

To ensure the correct conformation of the entire multi-enzymatic complex, the activity of the thioesterase (TE), the last domain of the last module of Lab13, responsible for the polyketide release, was specifically eliminated by performing a minimal genetic modification. The serine of the catalytic center composed by the Ser-Asp-His triad, well known in the esterase family proteins (Schneider and Marahiel, 1998), was changed by a less reactive alanine resulting in the construction of the *lab13*TES>A mutant. The Ser to Ala change is known to to disrupt the catalytic activity in this family of proteins while preserving their three-dimensional conformation (Pazirandeh et al.,

1991). Using the SWISS-MODEL Server (https://swissmodel.expasy.org/) the model of the Lab13 TE domain was obtained and identified the Ser67-Asp94-His217 triad (Figure 42).

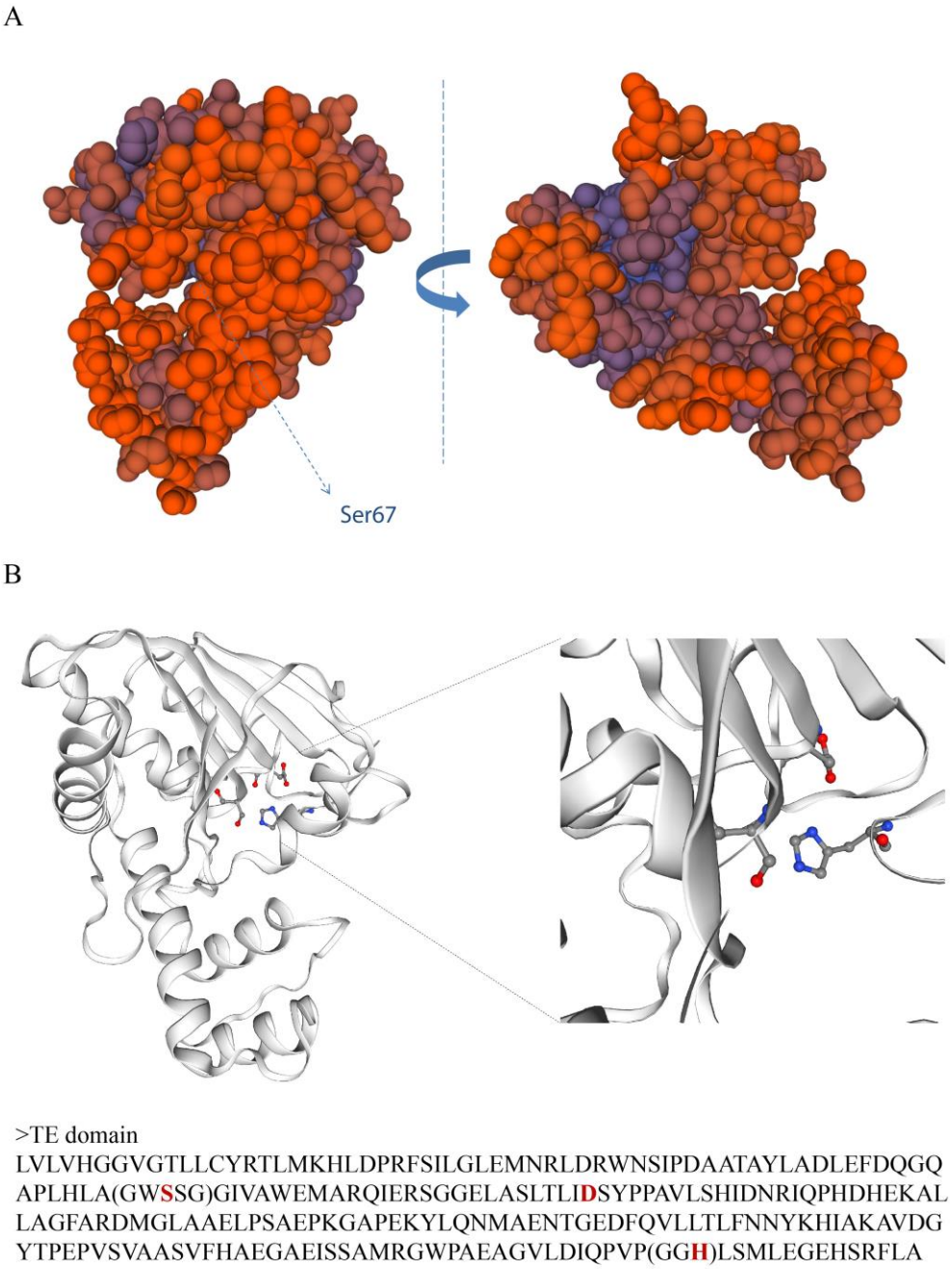


Figure 42. SWISS-MODEL of the Lab13 TE three-dimensional conformation. A) Spacefill model of TE with QMEAN color scheme showing the catalytic pocket from the side-left and right and indicating the active serine site. B) Cartoon model of TE in white color and zoom in of the amino acid triad Ser-Asp-His. The complete amino acid sequence with conserved catalytic motifs is shown below.

The construction of the *lab13*TES>A mutant would allow to determine if labrenzin is produced only after the final onnamide-like product is relased from the multi-enzymatic complex or labrenzin can be produced before this release. The culture supernatants were collected after 72 h of cultivation in M2 medium followed by the extraction and HPLC/MS analysis for the production of the polyketides. The HPLC/MS analysis of the *lab13*TES>A culture extracts, as opposed to wt, showed a drastic reduction in production of labrenzin and the analog intermediates most probably due to the strong reduction of TE activity (Figure 43). Interestingly, although the labrenzin and other intermediates were present in the extracts, the 840 ion was not detectable in the *lab13*TES>A culture extract (not showed in the figure). On the other hand, the presence of different peaks indicates the possibility of the premature release of stalled intermediates. All these results suggest that the synthesis and release of the complete onnamide-like molecule is critical for the maximum yield of labrenzin and that the ester hydrolysis happens after the release from the PKS/NRPS complex.

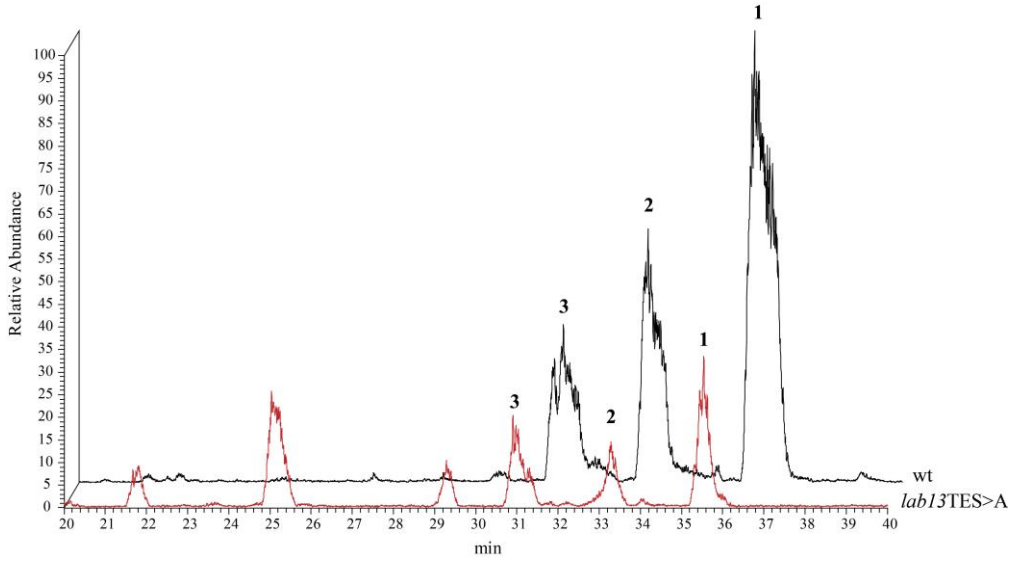


Figure 43. MS chromatograms showing extracted ions in the range of m/z = 468-512 with intermediates 1 (m/z = 512), 2 (m/z = 498) and 3 (m/z = 468) of labrenzin biosynthesis obtained from the wt and *lab13*TES>A culture extracts after 72 h of cultivation in M2 medium. The abundance of the *lab13*TES>A peaks are shown relative to wt peaks (100%).

4.2. Analysis of the polyketide modifications by tailoring enzymes

Tailoring in polyketide biosynthesis consists in the additional reactions catalyzed by independent enzymes that modify the core polyketide intermediate synthetized by PKS and therefore contributes to the structural and functional versatility of the natural products. Different tailoring processes can be observed, but according to the modifying enzymes, they can be classified in two general categories. On-line tailoring enzymes are *trans*-acting enzymes that engage with the multi-enzymatic PKS/NRPS complex and perform modifications during the core intermediate synthesis, like β-branching enzymes in *trans*-AT PKSs (Calderone, 2008) or amino acid heterocyclizations in NRPSs (Sundaram and Hertweck, 2016), while the other tailoring enzymes allow the post-PKS modifications such as hydroxylations and methylations (Han et al., 2018; Mohammad et al., 2018). In line with the objectives of this thesis, several functional genetic analyses were perfomed to determine the chemical role of the tailoring enzymes and the essentiality for the synthesis of the products. Additionally, a wide spectrum of new labrenzin analogues have been prepared as potential new candidates for future drug developments.

4.2.1. HCS cassette β-branching enzymes

The most common feature of *trans*-AT PKS systems is the presence of single AT enzymes that incorporate α-carbon branches, such as methylmalonyl, into a growing polyketide chain. However, *trans*-AT PKS clusters often encode other freestanding enzymes that can introduce unusual β-branches. Those enzymes form the so-called β-branching cassette, usually composed of stand-alone hydroxymethylglutaryl-CoA synthase (HMGS), non-elongating ketosynthase (KS^0), acyl carrier peptide (ACP) and one or two enoyl-CoA hydratases (ECH). The first evidence of the β-branching mechanism in a polyketide biosynthesis was demonstrated biochemically in the bacillaene cluster (Calderone et al., 2006). Since then, there were numerous reports of examples of HCS cassettes, contributing to the polyketide diversity like in myxovirescin (methoxymethyl and ethyl branches), pederin/onnamide (exomethylene group), bryostatin (acrylic ester), curacin (cyclopropyl group), mupirocin (methyl), leinamycin (1,3-dioxo-1,2-dithiolane moiety) and jamaicamide (vinyl chloride) (Calderone, 2008; Kusebauch et al., 2009; Walker et al., 2020).

The key role in these processes holds HMGS since it is in charge of performing a complex aldol addition and the substrate selection. To explore the functional role of HMGS, it was generated a deletion mutant Δ*lab3* expecting to produce a different labrenzin intermediate. Surprisingly, the *Δlab3* strain did not produce labrenzin suggesting an essential role of HMGS in labrenzin assembly, and a clear interaction with the PKS domains (Figure 44).

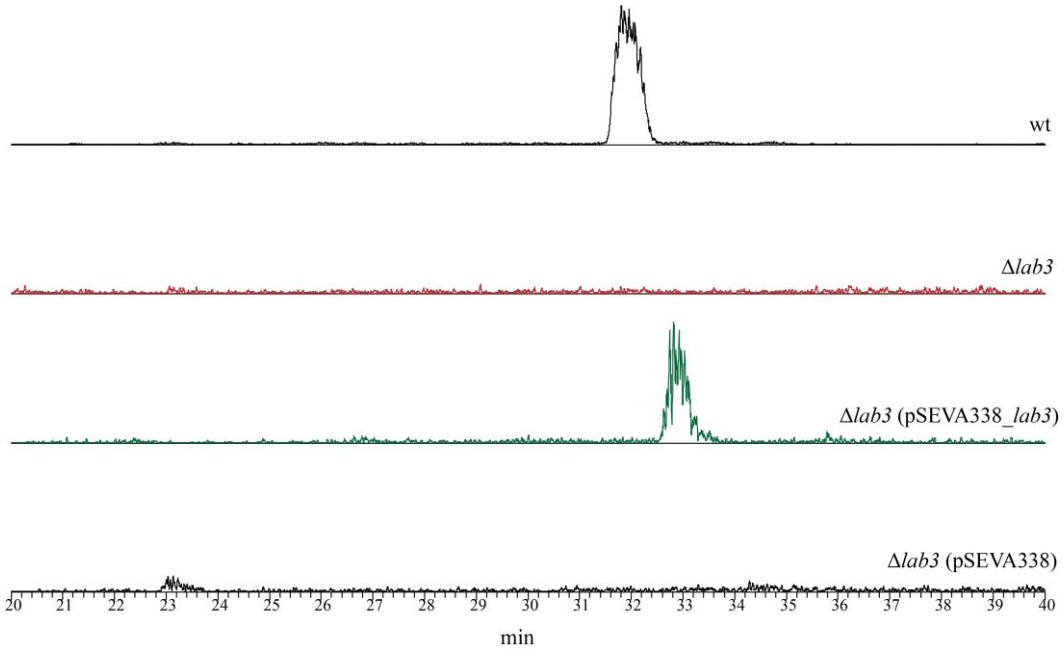


Figure 44. MS chromatograms show extracted ion m/z = 512 (labrenzin) of the supernatant extracts obtained after 72 h of Δ*lab3* cultures cultivation in M2 medium.

Next, a double deletion mutant Δ*lab11*-Δ*lab12* with both ECH (*lab11*) and KS (*lab12*) inactivated was constructed to investigate their implication in labrezin β-branching. The engineered strain was unable to synthetize labrenzin (Figure 45). The deletion was complemented by co-expressing both enzymes and by expressing each one separately. The labrenzin production was recovered after complementation with both functional enzymes, as expected, and with Lab11 alone. A complementation only with Lab12 was unsuccessful. This experiment suggests a crucial role of *lab11* gene in labrenzin assembly, but not for *lab12*. These results indicated the possible redundancy of the ketosynthase. Alternatively, it might be involved in downstream processing of the whole polyketide intermediate that remains to be identified.

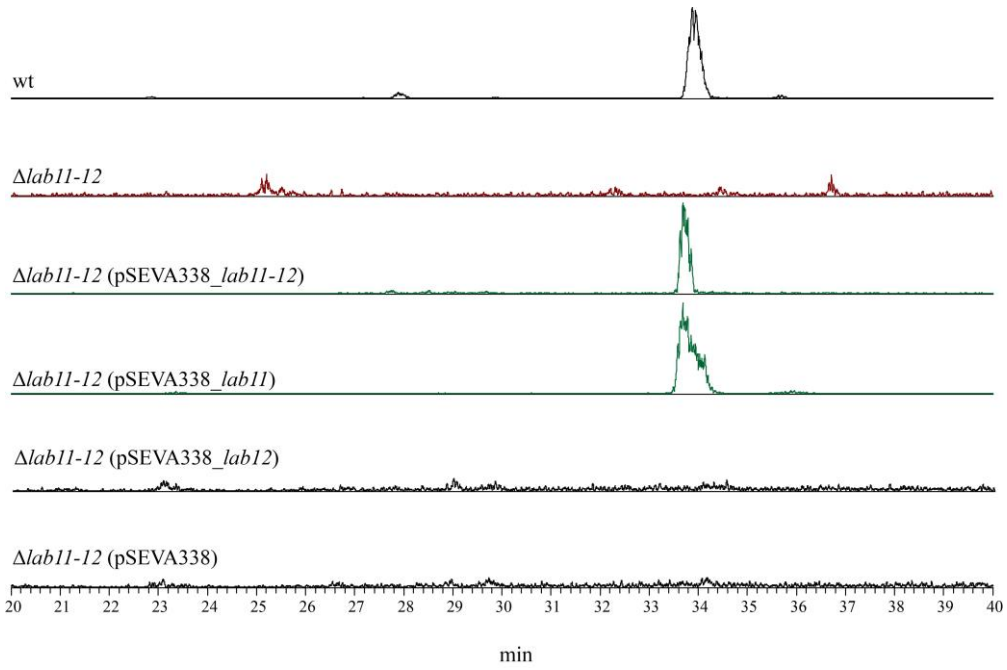


Figure 45. MS chromatograms showing extracted ion m/z = 512 (labrenzin) of the supernatant extracts obtained after 72 h of culturing the indicated strains in M2 medium.

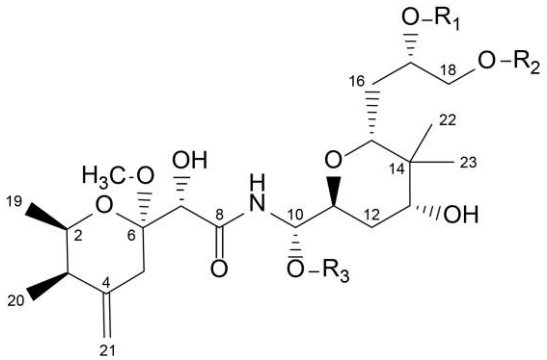
According to the results obtained from Δ*lab3* and Δ*lab11*-Δ*lab12* mutants, it can be concluded that at least HMGS and ECH play essential roles in the early assembly of labrenzin.

4.2.2. Methylations

The early studies on pederin gene cluster proposed a position-specific methylation for each methyltransferase encoded in the pederin (*ped*) gene cluster (Piel, 2002; Zimmermann et al., 2009). *Ped* cluster comprises three MTs (PedA, PedE and PedO) while *lab* cluster has only two MTs (Lab6 and Lab16). In comparison with pederin that has four methoxy groups, labrenzin has only three methoxy groups, lacking the 18-*O*-methylation (Figure 46). Since the pederin structure has four *O*-methylations and the gene cluster only contained three MTs, it was assumed that one of them might have a double function. *In vitro* experimental analysis established that PedO was responsible for the methylation of the C18-OH group in pederin and it was hypothesized that PedA or PedE could have a dual function (Zimmermann et al., 2009). A protein homology comparison between MTs from labrenzin, pederin and diaphorin showed that Lab16 shares 51% identity with PedE, while Lab6 shares 47% identity to PedA. Diaphorin, on the other hand, has only one *O*-methyl group that is probably generated by DipM (Figure 46) because diaphorin gene cluster lacks both PedA and PedO

orthologs (Nakabachi et al., 2013). In this sense, the identity between DipM and PedE is 60%. According to these sequence comparisons, Lab6 and Lab16 seem to be the functional analogs of PedA and PedE, respectively.

It is important to notice that the wt PHM005 strain mainly accumulates two methylated intermediates in the supernatant, *i.e.*, labrenzin (1) with two methylation on C6 and C17 and an analog (2) missing the *O*-methyl group on C17, as identified by PharmaMar (Figure 46). Curiosly, the analog with only one methyl group was barely detectable in the total MS scans. Therefore, it can be assumed that the methylations are produced sequentially, firstly and very quickly on both C6 and C10 and after that, the last methylation occurs on C17. However, this assumption does not inform about which one was the first methylation and which is the enzyme reposible for a dual methylation since several combinations are possible (see below).



Compound	R1	R2	R3
Labrenzin	CH3	H	CH3
Pederin	CH3	CH3	CH3
Diaphorin	H	H	H

Figure 46. Chemical structure of pederin family polyketides.

By overexpressing each one or both MTs in the wt PHM005 strain it would be possible to not only get an indication of the specific functions of each MT, but also test if it would be possible to improve the methylation of the final product and thus increase the labrenzin production. One would expect to notice a reduction in the 17-*O*-demethylated analog (**2**) when the specific methylase for C17 was overproduced. The recombinant culture supernatants were collected after 72 h of cultivation in M2 medium followed by the extraction and HPLC/MS analysis for the production of the polyketides. Figure 47 shows the overexpression of *lab6* and *lab16* genes

encoding the MTs in the PHM005 wt strain. Surprinsingly, when compared to wt, the ratio of 1 and 2 analogs did not change significantly when *lab6* and/or *lab16* genes were overexpressed. These experiments might suggest that the MT activities in wt strain are naturally saturated, but it cannot be discarded that the overexpressed MTs were not functional.

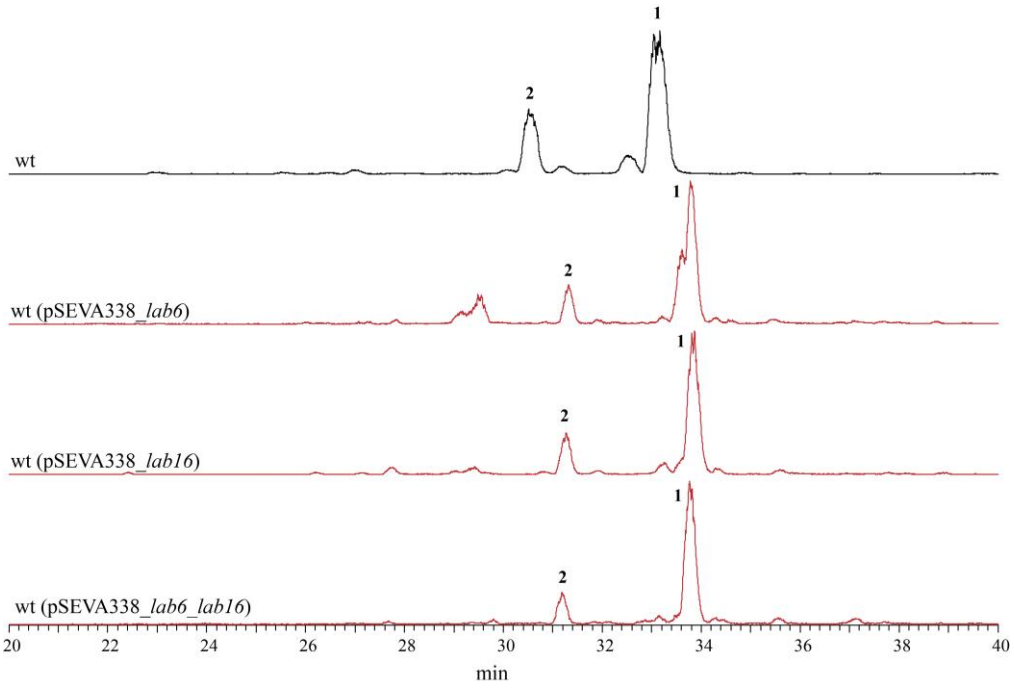


Figure 47. Overexpression of MTs encoded by *lab6* and *lab16* genes in the PHM005 wt strain. MS chromatograms showing extracted ions m/z = 512 (1) (labrenzin) and m/z = 498 (2) (17-*O*-demethyl labrenzin analog) of the supernatant extracts obtained after 72 h of culture in M2 medium.

In order to clarify these results, the deletion of each one or both MTs should provide a better information about their functions. In addition, by creating the mutants lacking the MTs it will be possible to test the functionality of the *lab6* and *lab16* genes by complementation. Moreover, it would be possible to create new demethylated labrenzin analogs, that will allow also to confirm if other MTs encoded in the genome could provide a methylation activity.

4.2.2.1. Function of Lab6 MT

To determine the role of Lab6, the deletion mutant Δ*lab6* was generated. The HPLC/MS analysis of the culture extracts of wt and Δ*lab6* mutant revealed the accumulation of a new labrenzin intermediate (**4**) (Figure 48). The fragmentation pattern (m/z = 484; 394; 412) of this compound is characteristic of two previously identified compounds m/z = 498 and m/z = 512 (m/z = 498; 394; 412 and m/z = 512; 408; 426, respectively) and the molecular mass corresponds to labrenzin missing two $–CH_3$ groups. In addition, the retention time of the intermediate is lower than the other two intermediates previously identified, (**1**) and (**2**), indicating that the new one is a more polar compound (see Material and Methods section 12.3), which perfectly aligns with a demethylated compound, *i.e.*, having free –OH residues. Interestinlgy, this new intermediate has been already described as diaphorin, a natural product isolated from the symbiont of citrus pest *D. citri* (Nakabachi et al., 2013). Remarkably, this is the first time that diaphorin has been produced in a free-living bacterium using genetic engineering techniques.

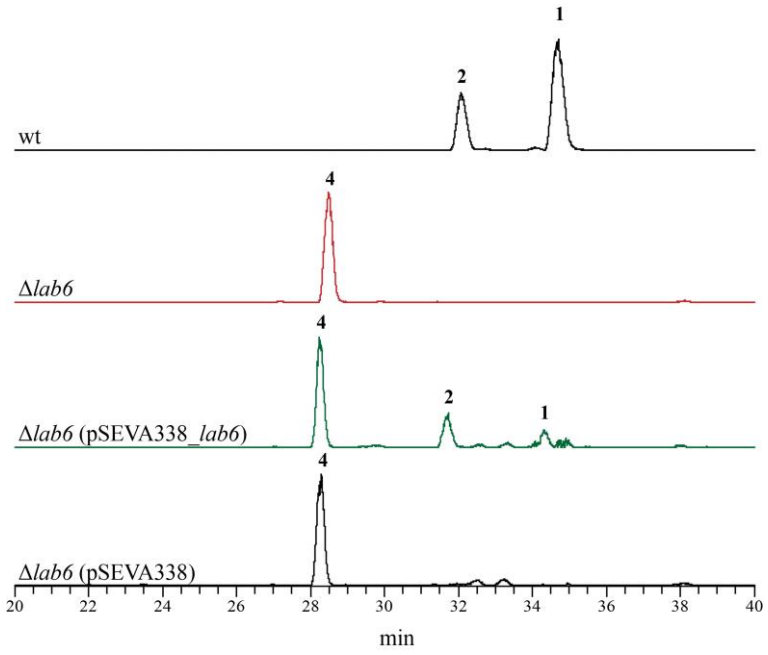


Figure 48. MS chromatograms of the supernatant extracts obtained after 72 h of Δ*lab6* mutant and wt cultures in M2 medium showing extracted ions in the range m/z = 484-512. Intermediates 1 (m/z = 512), 2 (m/z = 498) and 4 (m/z = 484) are indicated.

The *lab6* mutation was complemented with plasmid pSEVA338_lab6 encoding a copy of Lab6. Figure 47 shows that both current labrenzin intermediates (1 and 2) are only partially recovered after complementation, suggesting that Lab6 is not fully functional when expressed *in trans* and explain the poor results observed in Figure 46.

According to these results, it can be proposed that Lab 6 plays a dual function by methylating both C10 and C17 positions (Figure 49). Moreover, taking into account that the intermediate 2 is demethylated at C17 position, it can be proposed that methylation at C17 is produced only after methylation of C10. In addition, since a diaphorin peak is not currently abundant in wt strain, it can be assumed that methylation at C10 is very efficient.

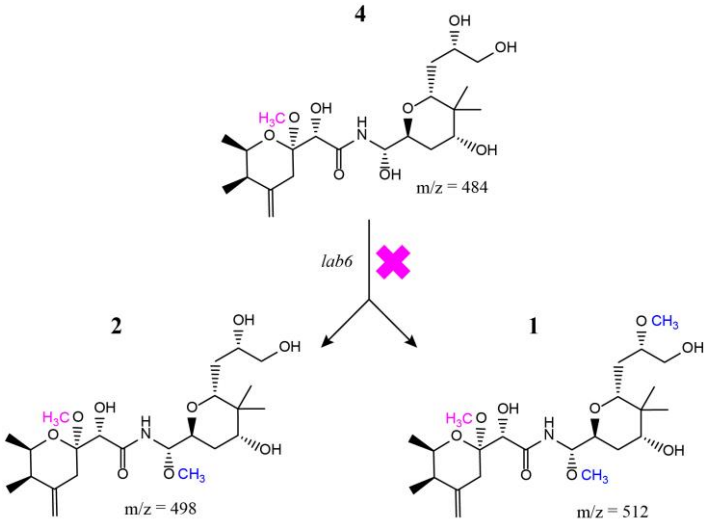


Figure 49. Proposed enzymatic reactions catalyzed by Lab6 MT based on the products found in Δ*lab6* mutant strain.

4.2.2.2. Function of Lab16 MT

Once ascribed a dual function for Lab6, it could be deduced the role of Lab16 as the Mt responsible of C6 methylation. Nevertheless, a deleted mutant strain Δ*lab16* was created to confirm this hypothesis. In addition, this will allow to determine if methylations occur strictly sequentially, *i.e.*, if the methylations of C10 and C17 positions occur only after the first methylation takes place at C6 position or all the methylations are sequentially independent.

Remarkably, Figure 49 shows that the Δ*lab16* mutant strain was able to accumulate several new compounds methylated by Lab6, which were undetectable in the wt strain. Interestingly, chromatograms showed an unexpected chiral separation of peaks after HPLC/MS analysis (Figure 50). Enantiomers **5a** and **5b** have a molecular mass m/z = 484 that corresponds to methylation at C10 position conserving the typical fragmentation for methylated compounds (m/z = 484; 394; 412). Isomers **6a** and **6b** have a molecular mass m/z = 498 with the characteristic ion fragmentation pattern (m/z = 498, 408; 426) for a labrenzin analog missing one methyl group at C6. Enantiomer **6a** overlaps with enantiomer **5b**, which can complicate the isolation of both compounds in case they will be required for physiological activity assays.

The complementation of the Δ*lab16* mutant strain with the pSEVA338-lab16 plasmid carrying a copy of MT Lab16 was only partial. As occured with Lab6 MT the expression of Lab16 MT *in trans* does not seem to be efficient.

The chemical reactions catalyzed by Lab6 in the Δ*lab16* mutant strain and the intermediates produced are represented in Figure 51.

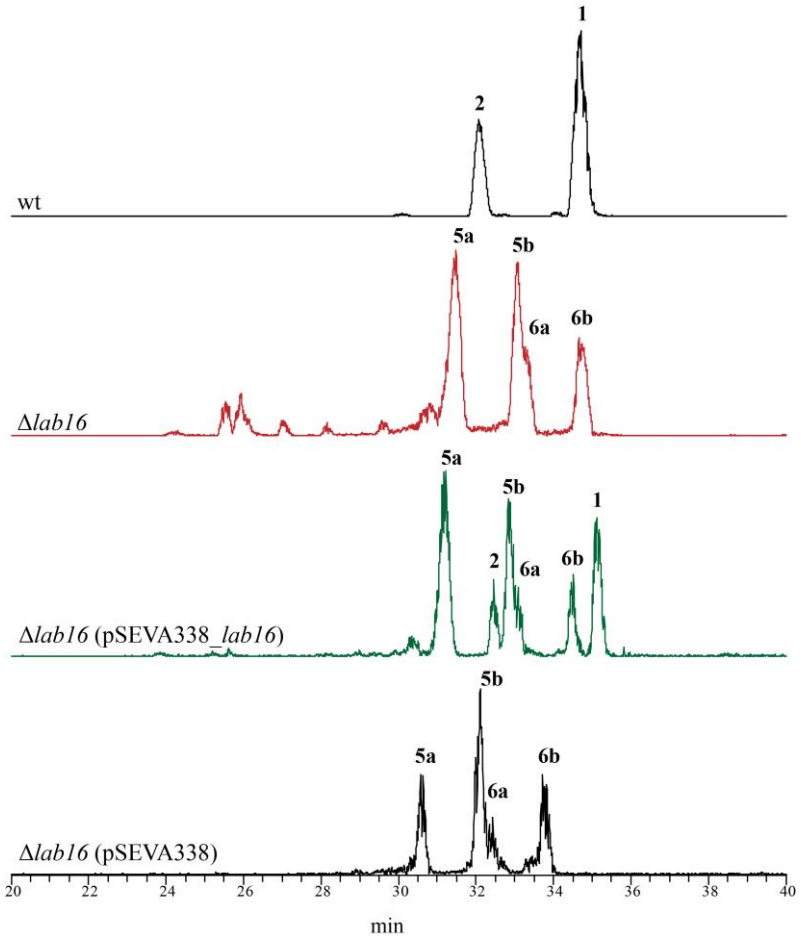


Figure 50. MS chromatograms of the supernatant extracts obtained after 72 h of mutant Δ*lab16* and wt culture cultivation in M2 medium showing extracted ions in the range m/z = 484-512. Intermediates 1 (m/z = 512) and 2 (m/z = 498) produced by wt are shown. New intermediates 5a and 5b (m/z = 484), 6a and 6b (m/z = 498) are quiral isomers.

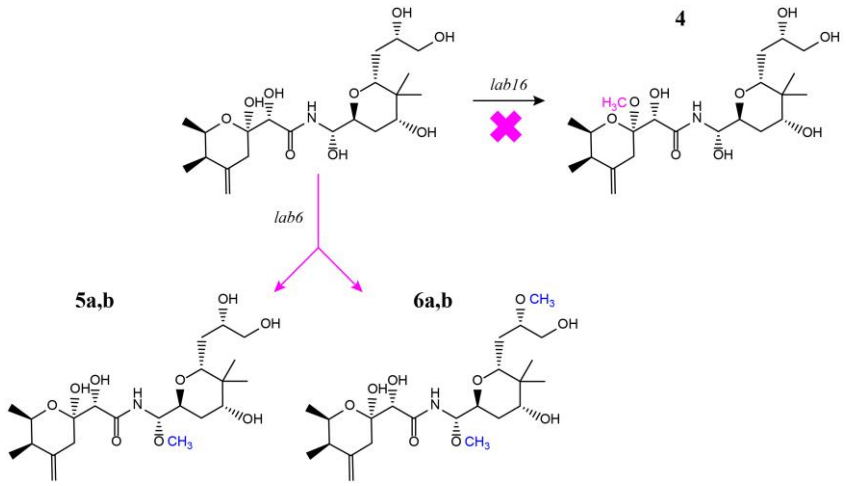


Figure 51. Enzymatic reactions catalyzed by Lab6 MT in Δ*lab16* mutant strain.

These experiments suggest that the methylations at C10 and C17 positions can be produced by Lab6 independently of the existance of a previous methylation at C6 position by Lab16. Nevertheless, these results do not exclude a preferred methylation order that could be ascribed to the different catalytic parameters for the different intermediate substrates according to their dissimilar accomadation in the catalytic centers of both MTs.

4.2.2.3. Construction of a double Δ*lab6*Δ*lab16* mutant

To generate a completely demethylated labrenzin analog and to expand the list of possible drug candidates for the future applications, double Δ*lab6*Δ*lab16* mutant strain, where both MT coding genes have been deleted, was contructed. Figure 52 shows that this mutant is able to produce a completely demethylated labrenzin analog **7** with m/z = 470 following the familiar fragmentation pattern (m/z = 470; 394; 412) of labrenzin intermediates. As expeted, since this new compound is more polar compared to the other methylated derivatives, it is detected at a lower retention time. Interestingly, the relative abundance of the demethylated labrenzin analog, expressed as the ion intensity, was one order of magnitude below the intensity of labrenzin, produced by wt (1.98E4 vs. 2.59E5).

The recovery of the methylated analogs by complementation of Δ*lab6*Δ*lab16* mutant strain with plasmid pSEVA330_lab6_lab16 carrying a copy of both MTs was only partial.

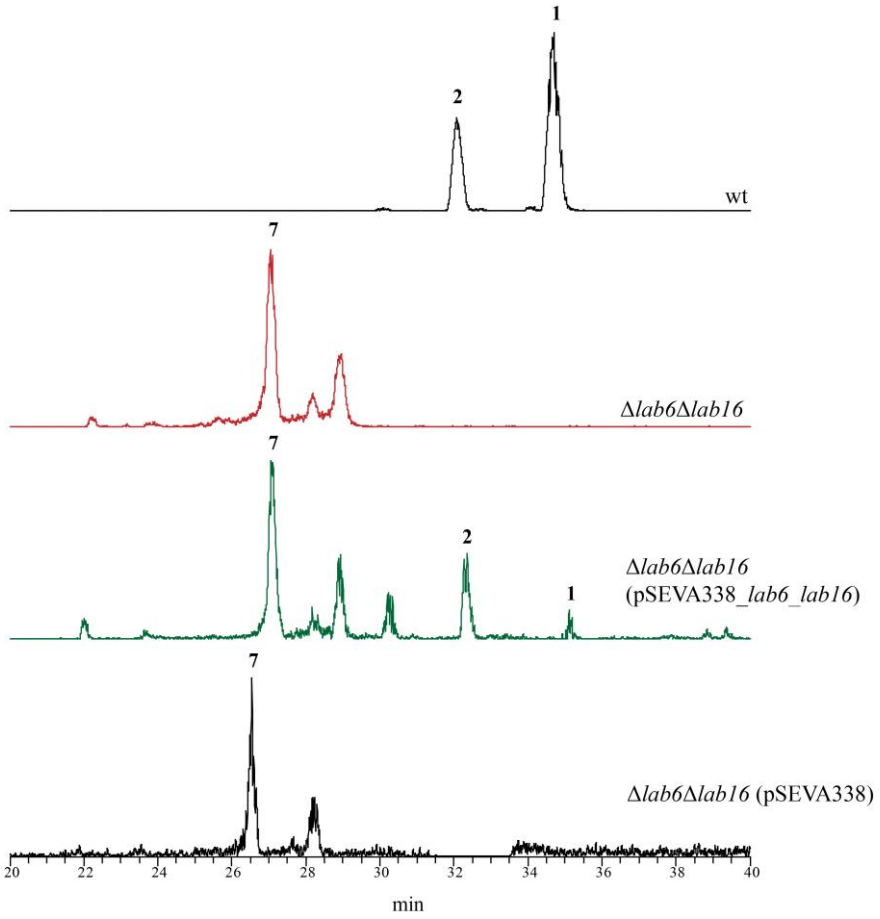


Figure 52. MS chromatograms of the supernatant extracts obtained after 72 h of mutant Δ*lab6*Δ*lab16* and wt culture cultivation in M2 medium showing extracted ions in the range m/z = 470-512. Intermediates 1 (m/z = 512), 2 (m/z = 498) and 7 (m/z = 470) are indicated.

The molecular structure of the intermediate derived from the disrupted tailoring reactions is represented in Figure 53.

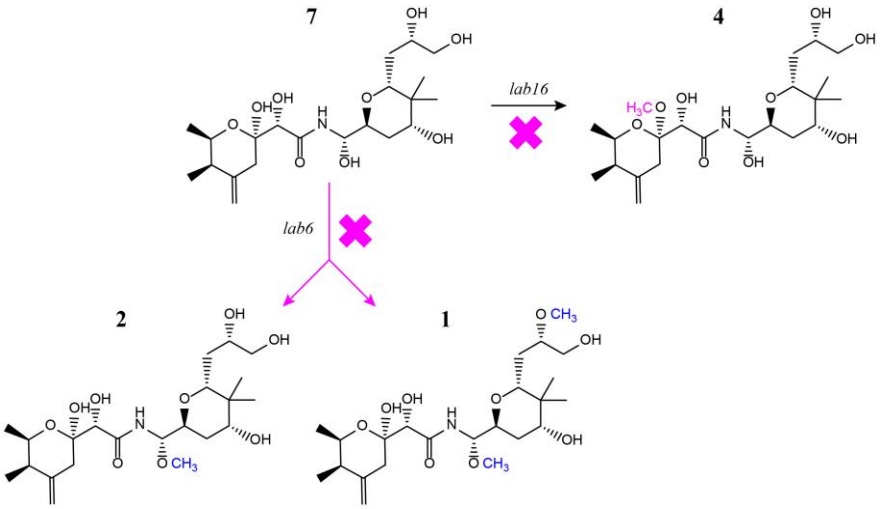


Figure 53. Scheme showing the disrupted enzymatic reactions due to of *lab6* and *lab16* gene inactivation in Δ*lab6*Δ*lab16* mutant strain.

4.2.3. Tailoring hydroxylations

Other tailoring reactions involved in labrenzin biosynthesis, beside methylations, are hydroxylations. In the *lab* cluster there are two putative FMN-dependent oxidoreductases (ORs), named Lab5 and Lab8, and a putative cytochrome P450, named Lab7, that are likely to be involved in the hydroxylations of C7 and C10 (Figure 46). The hydroxylation of C6 is generated by the PKS unit. A possible hypothesis was that at least one of the oxidoreductases would be responsible for at least one of the tailoring modifications. Also, the cytochrome P450 (Lab7) and OR (Lab8) form an operon and their ORFs overlap suggesting a functional dependence.

In order to determine the specific function of each putative hydroxylating genes, several knockout and complementation experiments were performed.

4.2.3.1. Tandem knockout of Lab7 and Lab8

To investigate the functional role of cytochrome P450 Lab 7 and OR Lab8, both genes were deleted, the mutant were screened for the accumulating intermediates and the mutations were complemented by expressing both genes together or independently. Figure 54 shows that when compared to wt, the Δ*lab7-lab8* mutant strain accumulates two intermediates with molecular masses 454 (**8**) and 468 (**3**), while labrenzin (**1**) and its 17-*O*-demethylated analog (**2**) were absent (Figure 54A). Considering the molecular masses of the newly accumulated compounds, it can be deducted the loss of two -CH_3 groups (-28) and one oxygen (-16) for the m/z = 468 compound and one additional loss of one -CH_3 group (-14) for the m/z = 454 compound which would correspond to the molecular structures showed in Figure 54B. Besides, the chemical structure of the 468 peak was previously identified and characterized by NMR by PharmaMar in the wt cultures which additionally supports the MS results obtained from the Δ*lab7-lab8* culture extracts (personal communication, data not shown).

Figure 54C shows the results of the complementation experiments of Δ*lab7*-Δ*lab8* mutant strain with pSEVA338_lab7-lab8, pSEVA338_lab8 and pSEVA338_lab7 plasmids. Interestingly, labrenzin production was observed only when the mutant was complemented by pSEVA338_lab7-lab8. When the *lab7* and *lab8* genes were used to complement the mutant separately, labrenzin was not detected in the culture extracts, suggesting that both enzymes might be required for the generation of the C10-OH group.

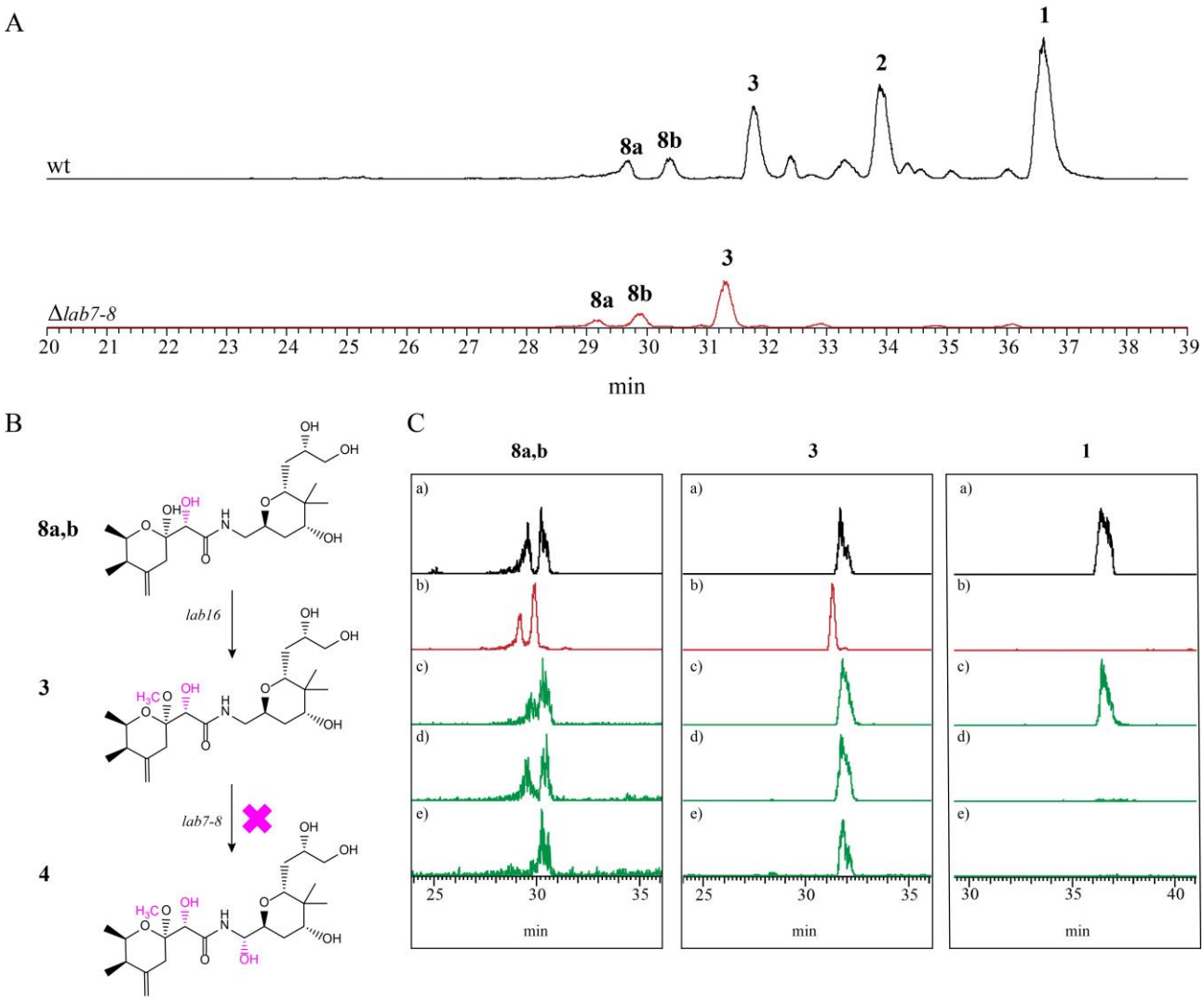


Figure 54. Functional role of Lab7 and Lab8. A) MS chromatograms of extracts obtained after 72 h of Δ*lab7*-Δ*lab8* mutant and wt cultures in M2 medium showing extracted ions in the range m/z = 454-512. The black and white dots indicate peaks m/z = 498 and m/z = 512, respectively. Accumulated polyketide intermediates in the Δ*lab7*-Δ*lab8* mutant are marked with numbers. B) The scheme showing the interrupted pathway of tailoring enzymatic reactions. C) The gene complementation experiment: a) wt used as a positive control; b) Δ*lab7-lab8* used as a negative control; c) Δ*lab7*-Δ*lab8* (pSEVA338_*lab7-lab8*); d) Δ*lab7*-Δ*lab8* (pSEVA338_*lab8*); e) Δ*lab7*-Δ*lab8* (pSEVA338_*lab7*).

4.2.3.2. Knockout of Lab5

Once it has been established the tailoring role of the cytochrome P450 Lab7 and OR Lab8, it appears more clear the possible role of the OR Lab5, that could be most likely responsible for the hydroxylation at C7 position. The *lab5* gene was deleted not only to confirm the tailoring role of OR Lab5, but also to test if the hydroxylation reactions happen in a specific order, *i.e.*, if they are substrate specific. The culture extracts were screened for the new intermediates by HPLC/MS.

Figure 55A shows new methylated intermediates without the C7-OH group, as expected, to confirm the function of Lab5. Three different peaks where observed in the MS spectra. The associated chemical structure are shown in the Figure 55B. Compound **9** with molecular mass 452 corresponds to labrenzin (m/z = 512) without two -CH_3 groups (-14) and two oxygens (-16). Two possible enantiomers **10a** and **10b** with molecular masses 482 correspond to a molecule **9** with the addition of one -CH_3 group (+14) and oxygen (+16), *i.e.*, 452 + 30 = 482. A fully methylated compound **11**, on the other hand, presented with the m/z = 496, fits the compound **10** with the additional methyl group (482+14 = 496).

Figure 55B shows the proposed alternative tailoring pathway once the *lab5* gene was inactivated and the C7-hydroxylation is blocked. The Δ*lab5* mutant strain accumulated compounds **9**, **10** and **11**, but corresponding intermediate between **9** and **10** or **11** was not detected, probably due to the high efficiency of MT Lab6 under these conditions.

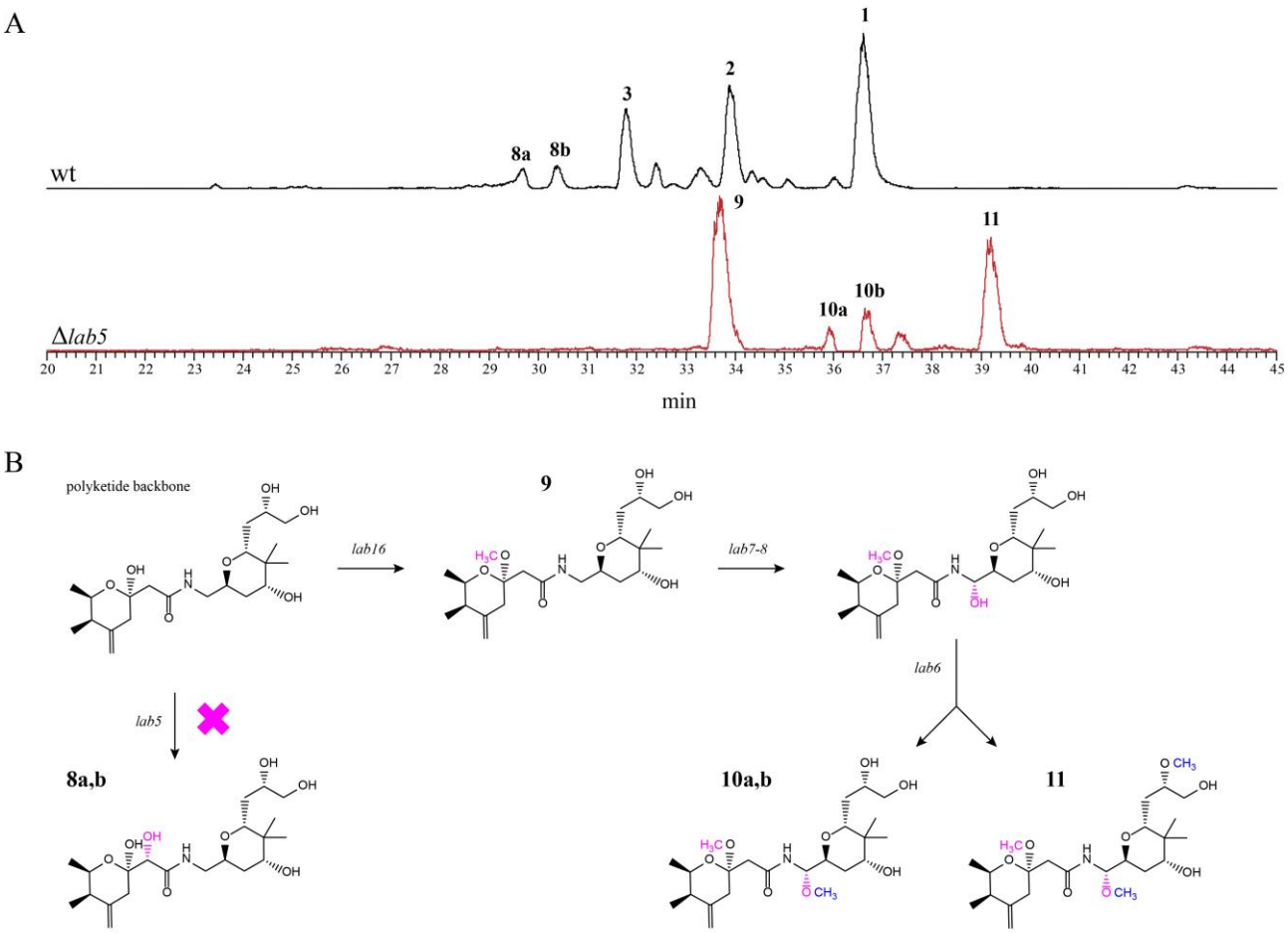


Figure 55. Functional role of OR Lab5. A) MS chromatograms of the supernatant extracts obtained after 72 h of Δ*lab5* mutant and wt cultures in M2 medium showing extracted ions in the range m/z = 452-512. Accumulated polyketide intermediates in are marked with numbers: 1 (m/z = 512), 2 (m/z = 498), 3 (m/z = 468), 8a,b (m/z = 454), 9 (m/z = 452), 10a,b (m/z = 482), 11 (m/z = 496). B) Scheme showing the interrupted hydroxylation reaction and an alternative pathway for tailoring enzymatic reactions.

4.2.3.3. Tandem knockout of Lab5-Lab6

A further step was to eliminate both the OR Lab5 and the MT Lab6, which are organized in an operon, to generate a Δ*lab5-lab6* mutant strain. The idea was to expand the repertoire of labrenzin analogs for future applications. Figure 56 shows the MS analysis of the mutant and wt culture extracts. As expected, the compound **9** (m/z = 452) was accumulated together with a small amount of a new intermediate **12** with molecular mass 468. Since in the mutant the methylation by

Lab6 was blocked, compounds **10** and **11** were not detected (Figure 56A). The updated alternative tailoring pathway is shown in Figure 56B.

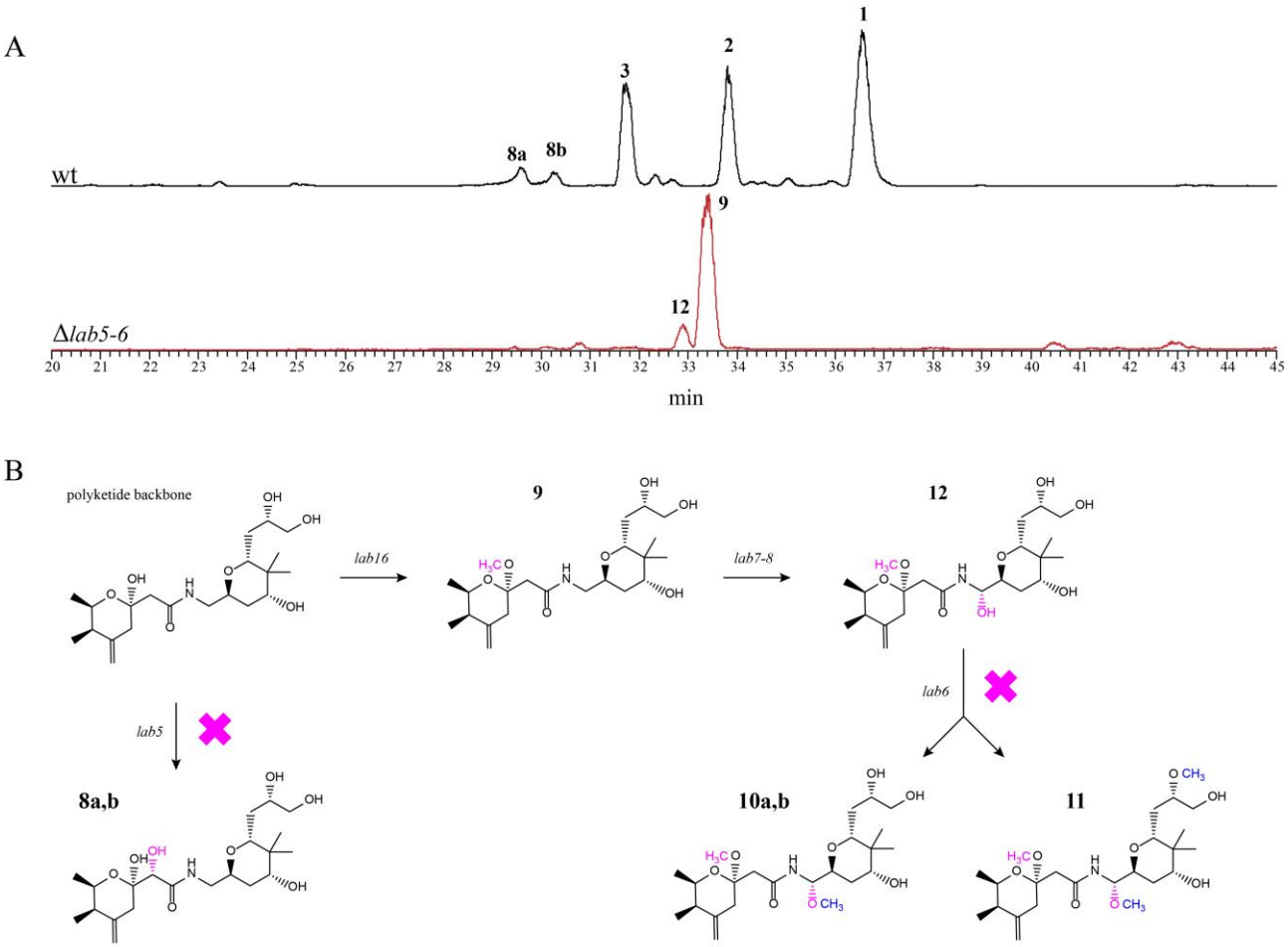


Figure 56. Generation of new labrenzin intermediates by deleting the OR Lab5 and MT Lab6. A) MS chromatograms of the supernatant extracts obtained after 72 h of Δ*lab5*-Δ*lab6* mutant and wt cultures in M2 medium showing extracted ions in the range m/z = 452-512. The molecular masses of each labrenzin intermediate in wt are shown above the respective peak. Accumulated polyketide intermediates in Δ*lab5*-Δ*lab6* mutant are marked with numbers. B) Scheme showing the interrupted hydroxylation and methylation reactions in the tailoring pathway.

4.2.4. Tailoring by heterologous expression PedO

4.2.4.1. Production of pederin in PHM005 wt strain

The difference between pederin and labrenzin is the methylation of C18-OH, that is absent in labrenzin. This was also predictable assuming that the *lab* cluster contains only two MTs, Lab6 and Lab16, characterized in the previous sections, when compared to *ped* cluster, which contains three MTs. The biochemical *in vitro* experiments conducted with mycalamide as substrate, a pederin family compound with free C18-OH group, demonstrated that PedO is the only MT of *ped* cluster capable of methylating at C18 position (Zimmermann et al., 2009).

In order to synthetize pederin for the first time in *Labrenzia* sp. PHM005, a synthetic *pedO* gene was heterologously expressed in this strain. The culture extracts were further analyzed by HPLC/MS. Figure 57A shows that a new intermediate, more hydrophobic than the previous analogs **1** and **2** observed in wt strain. The MS spectrum revealed a new peak **13** at RT = 40.4 min and with ion fragmentation m/z = 526; 440; 422 matching with an additional methyl group, suggesting that it is pederin (Figure 57B).

It is worth to mention that this result allows proposing that the hydrolysis of the onnamide-like polyketide that should originate labrenzin must take place inside the cells. The methylation of C18 of labrenzin by PedO can only take place inside the cells. Thus, if the OH-C18 group is not generated inside the cells it will be imposible to obtain pederin. This observation cannot be deduced from the activities of the other tailoring genes since in those cases it cannot be discarded that the tailoring reactions could be perfomed during the synthesis of the polyketide or even previously to the release of labrenzin.

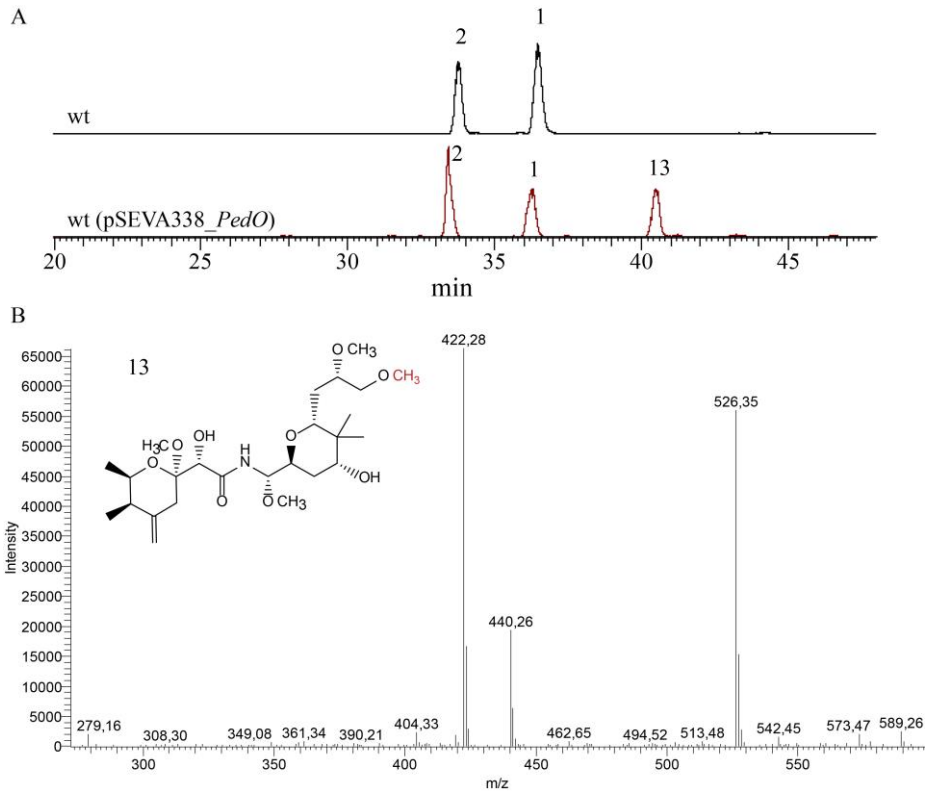


Figure 57. Pederin biosynthesis in *Labrenzia* sp. PHM005 wt (pSEVA338_PedO). A) MS chromatograms of the supernatant extracts obtained after 72 h of cultivation of wt cultures in M2 medium presenting extracted ions in the range m/z = 498-526. Intermediates 1 (m/z = 512), 2 (m/z = 498) and 13 (m/z = 13) are indicated. B) MS ion fragmentation of 13 (pederin).

Although pederin was produced in *Labrenzia* sp. PHM005 wt (pSEVA338_PedO), it was not the most abundant intermediate in the extract. In order to optimize the expression of *pedO* and to obtain sufficient amount of pederin for its purification and structure elucidation, two new genetic approaches were carried out.

Firstly, the relative pederin amount was increased by co-expressing *pedO* with *lab6* and *lab16* genes. Figure 58 shows MS chromatograms of culture extracts of strains harnessing different methyltransferase expression combinations. 17-*O*-demethylated labrenzin (**2**) was the the most abundant peak when *pedO* was expressed alone forming an operon with *lab16* and when the three MT genes were co-expressed. On the other hand, *pedO* expressed together with *lab6*, appeared to be the optimal combination for pederin production under these conditions.

The second approach was to engineer the expression plasmid pSEVA338_*pedO_lab6* with the constitutive strong P_{14g} promoter, replacing the inducible P_m promoter. Nevertheless, the promoter switch did not seem to change the peak ratio as observed in the MS chromatograms.

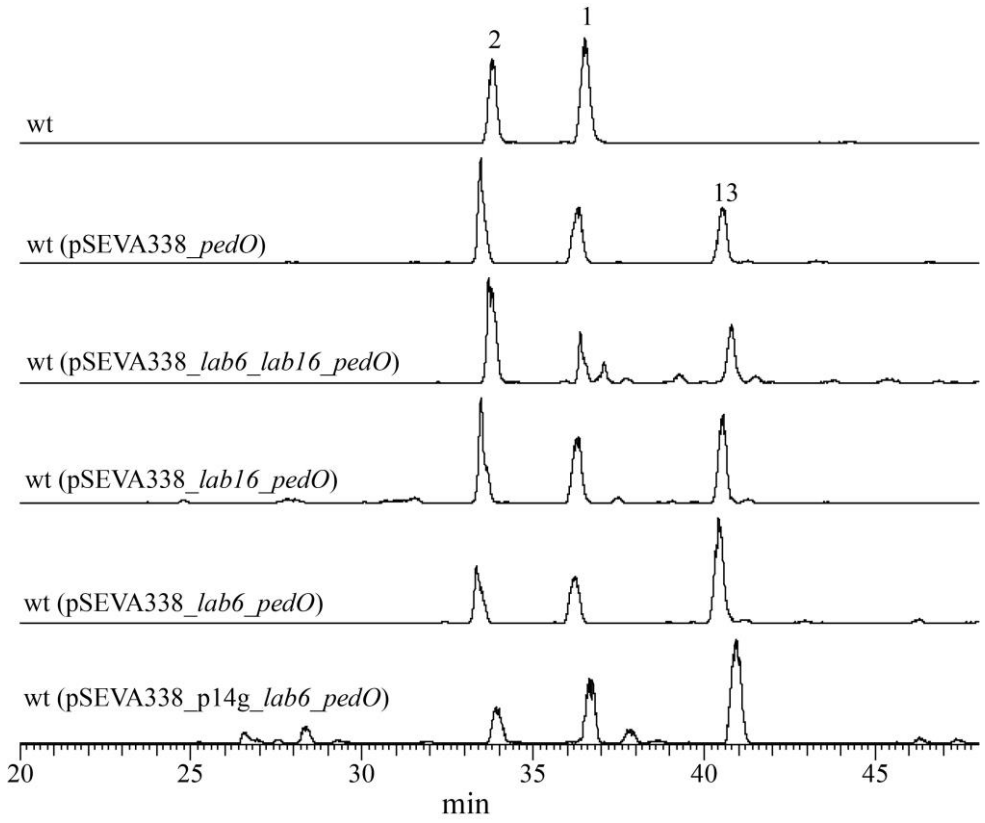


Figure 58. MS chromatograms of the culture extracts obtained after 72 h of cultivation of wt recombinant strains transformed with different plasmids in MBM+VIT medium. Extracted ions range is m/z = 498-526. Intermediates 1 (m/z = 512), 2 (m/z = 498) and 13 (m/z = 526) are indicated.

However, when carbon rich media is used for the cultivation (M2 or 2MLA medium used in PharmaMar), the intermediates' ratio does not maintain. Moreover, pederin is about one order of magnitude less abundant than labrenzin in those cultures (data not shown).

4.2.4.2. Production of novel pederin analogs in PHM005 mutant strains

In addition to heterologously synthetized pederin in *Labrenzia* sp. PHM005 (pSEVA338_PedO), the possibility of *à-la-carte* production of novel pederin analogs for the future applications was explored taking into account that all previously generated mutants accumulated differently tailored molecules, based on hydroxylation and methylation variations. It is assumed that

Results

PedO can catalyze the methylation on other compounds with the available C18-OH group as well, independently of the tailoring of the substrate molecule.

In order to generate new analogs, the pSEVA338_PedO plasmid was expressed in the Δ*lab6,* Δ*lab16* and Δ*lab6*Δ*lab16* mutant strains.

The extract of Δ*lab6* (pSEVA338_PedO) strain showed several new peaks in MS chromatograms (Figure 59). A part from the intermediate **4** lacking two methoxy groups (m/z = 484) with RT = 29.4-30.0 min (see also Figure 48 and Figure 49), the strain also accumulated **3**, 10-demethoxy-labrenzin (m/z = 468) at RT = 28.0-28.6 min. This compound is also accumulated in the Δ*lab7*-Δ*lab8* culture extracts (Figure 54). Therefore, it could be expected to detect the C18-OH methylated molecular counterparts. Summing one additional methyl group the expected products were detected in the total MS scan, as intermediates **14** (m/z = 498, 408; 426) and **16** (m/z = 482; 428), respectively. Unexpectedly, the peaks **15a**, **15b** and **17** were detected in sufficient amount whose molecular mass and fragmentation pattern fit the previously identified labrenzin analogs. Intermediates **15a** and **15b** were isomers with ion fragmentation m/z = 468; 428 characteristic for the compounds **8a** and **8b** (m/z = 454; 414) produced by the Δ*lab7*-Δ*lab8* mutant strain (Figure 54). Interestingly, demethylated counterpart did not accumulate in the culture extracts, suggesting that PedO might have higher affinity towards m/z = 454 intermediate as a substrate than Lab16. Under ideal conditions, compound **15** was expected to be completely transformed into compound **16**, but according to the results presented in Figure 59, Lab16 methylation does not seem to be as efficient.

The most unexpected result is the presence of the compound **17**, whose molecular mass corresponds to two additional methylations (m/z = 496; 442), produced by methylation of compound **16**. This result suggests that PedO might have a dual function that is able to replace the function of Lab6 and methylate another available -OH group. In addition, it can be observed that compound **14** is much less abundant than compounds **15**, **16** or **17**, suggesting that PedO prefers substrates lacking the C10-OH group under these conditions. To confirm the exact molecular structure of **4** it will be necessary to purify and identify it by NMR.

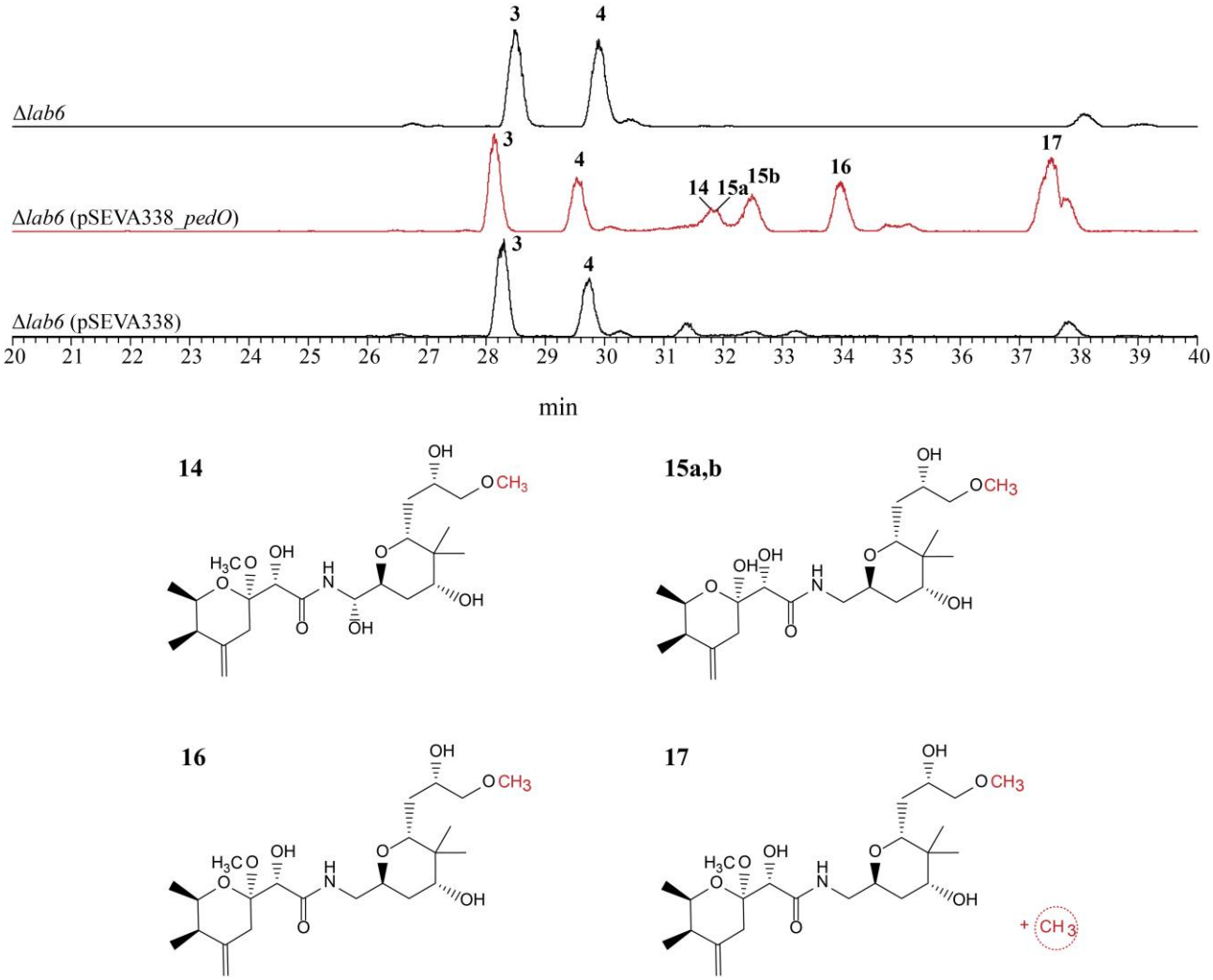


Figure 59. MS chromatograms of the culture extracts obtained after 72 h of cultivation of Δ*lab6* (pSEVA338_PedO) strain in M2 medium. Extracted ions range is m/z = 468-498. Intermediates 3 (m/z = 468), 4 (m/z = 484), 14 (m/z = 498), 15a,b (m/z = 468), 16 (m/z = 482), 17 (m/z = 498) are indicated. Molecular structures of hybrid intermediates are shown below.

On the other hand, four new methylated analogs were detected in the extract of Δ*lab16* (pSEVA338_PedO) strain when it was compared to a control extract (Figure 60). As presented above, the Δ*lab16* mutant generates two isomers lacking the C6-OH methylation: isomers **5a** and **5b** at RT = 30.6 and 31.8, and isomers **6a** and **6b** at RT = 32.4 and 33.8 (see also Figure 50 and Figure 51). Besides these methylated intermediates, the mutant accumulated substantial amount of the two isomers **8a** and **8b** missing the C10-OH group with RT = 27.2-28.8 min (see also the Figure 54). The Δ*lab16* (pSEVA338_PedO) strain showed their molecular counterparts **18a** and **18b** (m/z = 468; 428), **19a** and **19b** (m/z = 498; 408; 426), and **20a** and **20b** (m/z = 512) carrying an

additional methyl group. The unexpected result was again the appearance of the double methylated intermediate **21** with a fragmentation pattern m/z = 482; 442, typical for C10-dehydroxylated intermediate **8** of m/z = 454; 414 (Figure 54). Interestingly, the MS showed only one peak and it can be assumed either that compound **21** is produced from the quiral molecules **8a** and **8b** to a non-quiral intermediate or that PedO methylation favors one configuration over the other. The substantial residual amounts of 1a and 1b indicate that the second methylation is less efficient under these conditions.

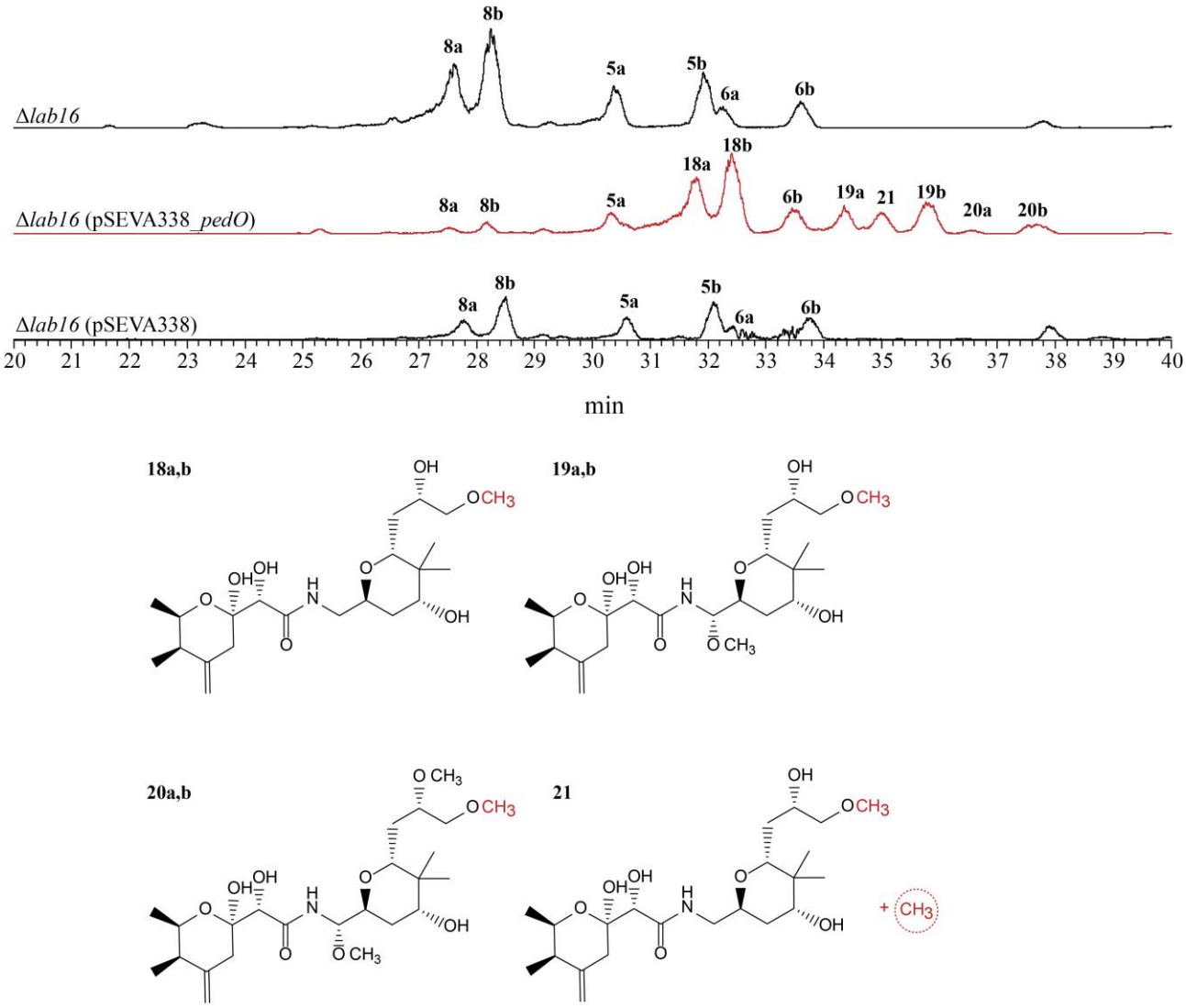


Figure 60. MS chromatograms of the culture extract obtained after 72 h of cultivation of Δ*lab16* (pSEVA338_PedO) in M2 medium. Extracted ions range is m/z = 468-512. Intermediates 5 (m/z = 484), 6 (m/z = 498), 8 (m/z = 454), 18 (m/z = 468), 19 (m/z = 498), 20 (m/z = 512), 21 (m/z = 482) are indicated. Molecular structures of hybrid intermediates are shown below.

Finally, an intermediate **22** with only one methylation at C18-OH position was produced in the Δ*lab6*Δ*lab16* (pSEVA338_PedO) strain (Figure 61). Unfortunately, the abundances of **22** (m/z = 484; 426) and its precursor **7** in the Δ*lab6*Δ*lab16* mutant (m/z = 470; RT = 26.5) were barely detectable in the MS spectra. The strain also accumulated intermediates **18a**, **18b** and **21** as Δ*lab16* mutant from the previous section.

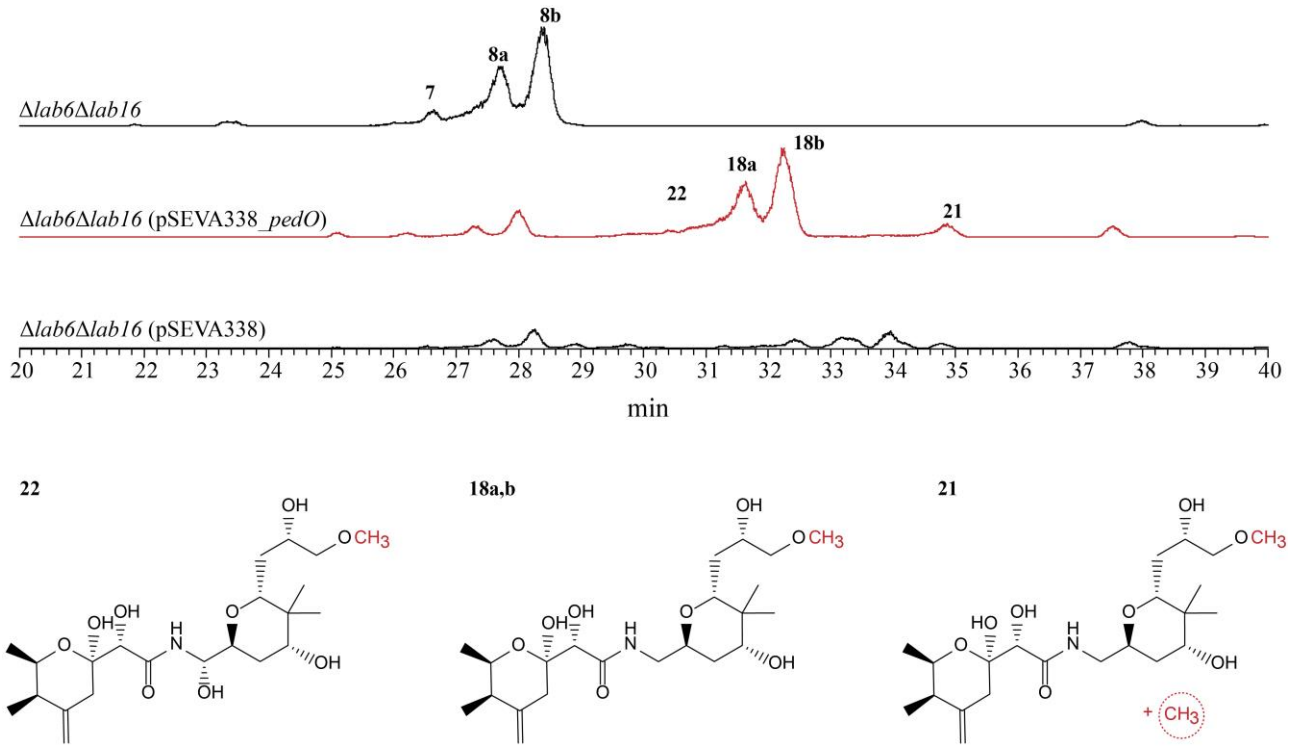


Figure 61. MS chromatograms of the culture extract obtained after 72 h of cultivation of Δ*lab6*Δ*lab16* (pSEVA338_PedO) strain in M2 medium. Extracted ions range is m/z = 468-484. Intermediates 7 (m/z = 470), 8 (m/z = 454), 18 (m/z = 468), 21 (m/z = 482), 22 (m/z = 484) are indicated. Molecular structures of hybrid intermediates are shown.

Discussion

1. *In silico* characterization of the *lab* cluster

The potent and selective bioactivities of pederin family compounds have triggered their pharmaceutical interest during the last years mainly as drug candidates for cancer therapy (Wan et al., 2011), however, the production of pederin analogs had been ascribed to non-cultivable endosymbiont bacteria. Piel and collaborators hypothesized that pederin-like gene clusters first belonged to a free-living bacterium and later have been horizontally transferred among symbionts where they preferentially stayed with the premise that these compounds were more advantageous to the host organism than to the producer itself (Piel et al., 2004c).

Until the recent discovery of a pederin analog produced by the free-living bacterium *Labrenzia* sp. PHM005 (Schleissner et al., 2017), pederin-related compounds were only obtained after extraction from insects or marine organisms (Helfrich and Piel, 2016), and by chemical synthesis through a multi-step process due to its complex structure (Wu et al., 2011). The structure of pederin and onnamide biosynthetic clusters have been reconstructed using partial sequences obtained from metagenomics analyses. Therefore, until now it has not been possible to fix precisely the boundaries of these clusters as well as to carry out functional or regulatory analyses. The discovery of the labrenzin produced by the cultivable strain *Labrenzia* sp. PHM005 has opened a new technical window not only to understand its biosynthesis and regulation, but also to apply biotechnological tools for industrial purposes and for developing large-scale production of these anti-cancer agents.

Based on *in silico* genomic approaches the *lab* cluster was identified and described, a *trans*-AT PKS/NRPS cluster that is responsible for the synthesis of labrenzin, a pederin analog in *Labrenzia* sp. PHM005. This allowed carrying out the comparative analyses between homologue gene clusters for the biosynthesis of psymberin, diaphorin, nosperin and cusperin to understand better the correlation between PKS domains and the chemical structure of growing polyketide. This *lab* cluster consists of three main PKSs/NPRSs genes that might be responsible not only for the synthesis of labrenzin but also for the synthesis of an onnamide-like analog. Interestingly, the partially published sequence of the homologous onnamide gene cluster (*onn* cluster) found in the metagenome of the sponge *T. swinhoei* contains all the PKS genes in a single region (Piel et al., 2004c) like in the PHM005 strain. It has been postulated that the diverse onnamide and pederin

analogs produced by the sponge might be generated by the same *onn* cluster by using alternative synthetic mechanisms (Piel et al., 2004c).

Concerning to onnamide-like molecules, the assembly of sesbanimide analogs in *S. indica* PHM037, novel types of glutaramide-family polyketides was proposed in this work (Figure 19). More than 80% of the *sbn* biosynthetic genes are present in *lab* cluster, although those clusters are responsible for the production of structurally completely different molecules. Moreover, the structure of sesbanimide F revealed the cryptic polyketide moiety co-linear with the SbnQ domain organization. Strikingly, the modular architecture of downstream mixed type PKS/NRPS, SbnQ, shows a high similarity to PedH in pederin and Lab13 in labrenzin gene clusters. This insight provides a clue about the missing polyketide moiety in labrenzin biosynthesis. The unexpected presence of SbnQ homologs in unrelated polyketide gene clusters across phylogenetically distant bacteria raises intriguing questions about the evolutionary relationship between glutaramide-like and pederin-like pathways, as well as the functionality of their synthetic products.

2. Bases of labrenzin production

2.1. *Trans*-AT with proofreading role in labrenzin assembly

In *trans*-AT PKS an external acyltransferase enzyme associates and dissociates with the PKS complex during the acyl transfer process from the activated acyl-CoA to the growing polyketide chain. The *lab* cluster encodes two *trans*-ATs, named Lab9 and Lab10. Interestingly, only a couple of PKS clusters have been described to contain more than one *trans*-AT, *i.e.*, the PedC and PedD *trans*-ATs from the pederin cluster (Piel, 2002), and the BaeC and BaeD *trans*-ATs from the bacillaene gene cluster (Chen et al., 2006). *Trans*-AT Lab9 is homolog to PedC with 35% of amino acid identity and Lab10 is homolog to PedD with 51% of aa identity. *In vitro* experiments with PedC and PedD demonstrated their similar yet distinctive functional roles in pederin biosynthesis (Jensen et al., 2012). While PedD was responsible for loading malonyl units to ACPs, PedC showed proofreading activity hydrolyzing a whole variety of short-chain acyl-intermediates, suggesting that it is responsible for editing the incorrectly inserted acyl-units and thus, for polyketide yield improvement (Jensen et al., 2012).

In this work, it was showed the important role of Lab9 in the efficiency of the polyketide assembly, as its inactivation led to dramatic drop in labrenzin production (Figure 34). This result also supports the previous *in vitro* observation of the hydrolytic activity on its homologue PedC and the speculated the editing function (Jensen et al., 2012). Similar editing role can be find in stand-alone type II TEs that hydrolyze residues attached to modular ACPs and can remove stalled intermediates, release the final product or control the substrate selection. In addition, alike Lab9 knockout, there are several examples of TE type II removal causing significant decrease in the polyketide production up to 85-90% (tylosin), 40-60% (rifamycin B), 90% (candicidin), 84% (surfactin) (Kotowska and Pawlik, 2014). Unalike Lab9, these examples have reached the production in the wild type strain after its complementation. A possible explanation may be a Lab9 misfolding and/or aggregation because of non-optimized overexpression or because the selection of an inadequate expression plasmid.

Lab9 knockout also indicated that the *trans*-AT Lab10 alone is sufficient for the malonyl-unit loading to the PKS and therefore the labrenzin biosynthesis. Overexpression of both *trans*-ATs, using the native promoters in a multicopy plasmid, did not lead to increased polyketide production under the conditions tested. *Trans*-ATs activity might also depend on the substrate availability and assembly efficiency on the global turnover of the PKS system, which may depend on many factors. In conclusion, the bottlenecks in polyketide biosynthesis might be based on something else and the increased production requires fine-tuning of several different variables.

2.2. Transcriptional up-regulation

Another approach to increase the labrenzin production was tested by using chemical elicitors to induce the transcription levels of *lab* genes. It has been shown that the sub-inhibitory concentrations of antibiotics have pleiotropic effect on multiple promoters across the genome, as they recruit the global transcription regulators that may activate the transcription of silent gene clusters (Goh et al., 2002; Zarins-Tutt et al., 2016). The main downside of this approach is the inability to determine the mechanism of gene activation since there are many factors involved and it is challenging to track down the right enhancer.

Discussion

It was showed that all the chemical substances used affected positively the promoter P_{n4} activity, especially using trimethoprim and piperacillin. They were also successfully applied in the first high-throughput elicitor screening resulted in de-silencing of two cryptic gene clusters in *Burkholderia thailandensis* and inducing at least five metabolic pathways (Seyedsayamdost, 2014). Although the *lab4* transcription was upregulated, the production of labrenzin did not significantly increase as compared to wild type under the culture conditions tested. The obvious limitation of the method is the specific targeting of only one gene or one operon at a time. While in some cases the whole gene cluster is organized in one operon, like in didemnin B gene cluster from *Tistrella mobilis* (Xu et al., 2012), *lab* gene cluster has at least 15 identified TSS with the respective promoter regions (Figure 27).

Alternatively, these results suggest that the gene transcription itself may not be the bottleneck in labrenzin production. Other factors should be taken into account in order to improve the labrenzin yields, such as engineering strong RBS sequences to improve the translation or ensuring excess of substrates and necessary cofactors. It is also possible that the presence of the chemicals interferes with other pathways or changes intracellular physical conditions that negatively affects the overall polyketide biosynthesis.

In this thesis, a new tool to improve elicitors screening using GFP as a reporter in *Labrenzia* sp. PHM005 has been developed. Various high-throughput screening methods based on translational fusions of LacZ, XylE or GFP as reporters have been developed until now (Tomm et al., 2019). In this work, the monomeric super folder GFP was used as reporter gene and its production was enhanced using a bicistronic design element resulting in the green fluorescence visible to the naked eye. This property allows to screening many colonies at once on a simple petri dish and may be applied in random mutagenesis screenings in future studies.

3. Essential enzymes for labrenzin biosynthesis

In this work, several gene deletion mutant strains have been produced. The mutations have completely abolished the production of labrenzin, showing that the deleted genes play an essential role in labrenzin biosynthesis. This is the case of *lab3*, *lab4*, *lab11* and *lab13* genes encoding the HMGS, PKS, ECH and PKS/NRPS enzymes, respectively.

3.1. Operon *lab4-lab5-lab6*

The first and most straightforward mutation was performed by the inactivation of *lab4* gene encoding the first PKS generating Δ*lab4* strain. This mutation was useful for demonstrating that the *lab* gene cluster was certainly the labrenzin biosynthetic cluster. In fact, this was the first time that a genetic engineering tool was applied to study a pederin-family gene cluster, since the *ped* gene cluster identified in 2002 could not be manipulated. At the same time, the mutagenesis of *lab4* was designed to disrupt also the P_{n4} promoter and ensure the generation of a non-producing labrenzin strain, to be used as a negative control in future studies. Not surprisingly, the mutation caused a polar effect on the expression of the following two genes, *i.e.*, *lab5* and *lab6*, demonstrating that these genes formed an operon.

Restoring the labrenzin production in the Δ*lab4* (pSEVA338_*lab4_lab5_lab6*) mutant strain by complementation seemed a good opportunity to test the possibility of producing a large gene transcript using a multi-copy plasmid. PKSs are large multi-enzymatic proteins whose overproduction could be rather challenging. As expeted, labrenzin was detected in the completented strain but its production was below the level of wild type. The attempts to increase the production by changing the temperature and shaking were not sucsessful. Several reasons might explain the low production yield in the complemented strain such as protein aggregation caused by overexpression of the protein. Moreover, multi-copy plasmid replication and unregulated protein expression might act as a metabolic burden (Wu et al., 2016). Nevertheless, the reasons should be studied more carefully.

3.2. β-branching enzymes

There have been described many types of β-branching enzymes assisting the polyketide biosynthesis and contributing to their functional and chemical variety (Walker et al., 2020). HMGS cassette (or HCS cassettes) that introduces alkyl side-chains into β-position usually comprise individual HMGS, ACP, ECH and KS. However, there is only limited evidence of their function (Buchholz et al., 2010; Calderone et al., 2006; Haines et al., 2013; Simunovic and Müller, 2007). Among the tailoring enzymes in *lab* cluster, there are free-standing ACP, HMGS, ECH and KS as encoded by *lab2*, *lab3*, *lab11* and *lab12* genes, respectively.

Discussion

In this work, the essentiality of HMGS and ECH was shown, but suprisingly the knockout of the stand-alone KS^0, that it expressed in the same operon with ECH, did not have an effect on the labrenzin production. One possible explanation for this finding is an enzyme redundancy, *i.e.*, other enzyme from elsewhere in the genome could complement the decarboxylation function of KS^0.

On the other hand, the HCS cassette could interact *in trans* with the modular KS, more probable due to the high specificity of the *lab* genes and the fact that HCS cassette already interacts with the PKS modules. In agreement with this possibility, the third module of Lab4 PKS comprises an ECH domain and tandem ACP domains, conserving motif flags for β-branching specificity (Haines et al., 2013), immediately followed by a KS domain that could assist the HCS cassette.

Then, it was surprising that the cluster could need a redundant KS^0. On reason can be that the additional KS could improve the turnover of the branching enzymatic complex and use it to regulate the assembly rate. Another reason could be that stand-alone KS^0 is engaged in the downstream polyketide tailoring.

According to the results obtained from the knockout and complementation studies on HMGS, ECH, KS and backed up with data from other studies, it can be assumed that at least HMGS and ECH from HCS cassete are crucial for labrenzin biosynthesis, probably interacting with the Lab4 PKS module 3 to form the *exo*-methylene branch in labrenzin. Without this *in trans* activity the biosynthesis of the polyketide is impaired. The same outcome was reported in myxovirescin biosynthesis where the HMGS TaC knockout completely abolished the production (Simunovic and Müller, 2007) or in the mupirocin production after deleting HCS cassete genes (Hothersall et al., 2007). Considering this result, it would be interesting to manipulate these genes in order to increase the turnover of PKS and to improve the efficiency of the assembly flow in future works.

ACP encoded by *lab2* gene remains to be studied, although it likely forms a complex with HMGS as described (Walker et al., 2020).

In conclusion, the proposed biochemical mechanism of *exo*-methylene β-branching in labrenzin biosynthesis would include the association of HMGS-ACP complex with the interaction "pocket" in the tandem ACPs in the module 3 of PKS Lab4, as demonstrated by modelling and

mutagenesis (Haines et al., 2013), followed by successive dehydration and decarboxylation steps catalyzed by modular ECH and *in trans* acting ECH, as it was demonstrated in curacin A biosynthesis (Gu et al., 2006).

4. Assembly of the large cryptic labrezin intermediate

4.1. Comparative analysis with sesbanimide biosynthesis

Chemical elucidation of the intermediates of sesbanimide biosynthetic pathway and sequencing of the genome of two marine producers opened the window to understand better the assembly patterns of sesbanimide and other similar polyketides. Strikingly, *sbn* cluster shared 80% of genes with *lab* cluster. The most interesting result of these analyses was the determination of sesbanimide F structure, revealing for the first time a clue to anticipate the structure of the large cryptic polyketide fragment (Figure 19) that should be synthezided by SbnQ, the Lab13 homolog. The structure of lipoid moiety of sesbanimide F reflects the architecture of SbnQ almost perfectly, with the exception of the terminal amino acid that is missing. Thus, sesbanimide F structure provides some clues for the whole polyketide assembly of *lab* cluster.

An initial hypothesis was that the monooxygenase inserted between the two PKSs in *ped* cluster, PedG, would have a hydrolyzing function (Piel, 2002). Interestingly, its SbnP homologue, also located between two hybrid PKSs, could be a Baeyer-Villiger monooxygenases (BVMOs). These enzymes convert ketones and cyclic ketones into their corresponding esters and lactones, respectively (Leisch et al., 2011). The first demonstration of its role in a polyketide biosynthesis was an *in vitro* biochemical assay of individually expressed *cis*-monooxygenase from the FR901464 pathway (Tang et al., 2013). Later, the expression of OocK, a BVMO from oocydin biosynthesis, showed that the enzyme could also produce an oxygen insertion into synthetic substrates. Thus, it was speculated that it is involved in the ester-like polyketide oocydin B synthesis, followed by ester hydrolysis into oocydin A (Meoded et al., 2018). Taking into account these observations, it has been suggested that SbnP is responsible for the oxygen insertion into the growing polyketide chain of sesbanimides that generates the long-chain intermediate sesbanimide F. Moroever, taking into account the high similarity of SbnQ and Lab13 and the BVMO homology with Lab14, it can be expected that labrenzin biosysnthesis uses the same mechanism.

Discussion

To explain the structure of the lipid moiety of sesbanimide F, it is worth to mention that the two first modules of SbnQ (also found in Lab13) should contain an enoyl reductase domain, but they are not present in the protein. This means that this function should be carried out by a *trans*-enoyl reductase. In addition, taking into account the non-canonical distribution of the three dehydratase and two ketoreductase domains in the four PKS modules of SbnQ, these domains should act *trans*-modularly coordinated to generate the double bounds present in sesbanimide F.

The presence of different sesbanimide congeners suggests that the final polyketide product might be processed into a longer intermediate sesbanimide F and shorter sesbanimides A, C, D and E, most probably by hydrolytic esterases or amidases encoded by additional tailoring genes. If amide cleavage was very efficient, it would explain why the fully elongated congener is not detected in the cultures.

4.2. Mutations in downstream PKS/NRPS (Lab13)

Lab13 is a hybrid PKS/NRPS involved in the downstream biosynthesis of the cryptic polyketide moiety that was first associated to onnamide-type molecules (Piel, 2002). Thus, it would be expected to detect in the labrenzin extracts at least a sesbanimide F-like molecule, as identified from the *S. indica* PHM037 culture extracts. One major aim of this thesis was the identification of the large onnamide-like labrenzin intermediate or at least a sebanimide F-like intermediate that can explain the labrenzin origin. The dilemma was to determine weather it was necessary to synthetize the whole polyketide intermediate in order to hydrolyze it later. Perhaps, labrenzin was produced in another way, for instance by a *trans*-acting enzyme, encoded within the cluster or elsewhere in the genome, that could release labrenzin from the PKS complex just at the moment it is synthetized. In this way, labrenzin could be produced before the onnamide-like polyketide could be synthesized. Should that be the case, there must be compelling reasons to justify the conservation of such a hughe *lab13* gene. One reson could be that Lab13 is required to maintain the structure of the whole enzymatic complex. However, it could not explain why all key catalytic sites are still conserved in Lab13. If the only reason to maintain Lab13 was to support the complex structure, one should expect that the enzymatic activity of Lab13 might have disappeared by evolution. On the other hand, taking into account that the number of domains in DipO (Nakabachi et al., 2013) and SbnQ,

two homologues of Lab13, are not identical, the idea that Lab13 migh have lost a domain and it could not be functional cannot be discarded.

The absence of labrenzin in the Δ*lab13* mutant strain confirmed that the Lab13 protein is required for the synthesis of labrenzin. However, this result does not mean that a whole polyketide has to be synthetized in order to produce labrenzin or that Lab13 is functional since, as mentioned above, Lab13 could be only needed to keep the structure of the enzymatic complex.

To address the question of the enzymatic role in the polyketide elongation by Lab13, a point mutation was perfomed in the last thioesterase domain. This domain is required to release the polyketide from the enzymtaic complex. In this way, the 3D structure of Lab13 is not altered and it is possible to confirm is Lab13 is only required to keep the structure of the complex or it is fully functional. To this aim, the serine of the catalytic site was replaced by alanine in order to strongly reduce or even to abolish the thioesterase activity (Witkowski et al., 1994). The *lab13*TES>A mutant only produces a residual amount of labrenzin (Figure 43) suggesting that, as expected, the thiosterase activity was strongly reduced (Pazirandeh et al., 1991), that Lab13 is functional and that the synthesis of the whole onnamide-like polyketide should be a true intermediate in the sysnthesi of labrenzin. Therefore, the residual amount of labrenzin detected in the *lab13*TES>A mutant provides a more solid clue about the role of Lab13, suggesting that is it not only required as a structural support. In this case, since the mutation only impairs the release of the onnamide-like polyketide, it suggests that labrenzin is generated by hydrolysis of the onnamide-like polyketide once it was released from the complex. If labrenzin could be released before the large polyketide could be synthesized the *lab13*TES>A mutant should be able to produce labrenzin in high amounts as the wild type.

On the other hand, the Δ*lab13* mutant allowed to compair the UV and MS chromatographs of crude extracts with those obtained from the wild type strain extracts. In this way, it can be observed that in extract from the wt strain there was a significantly abundant ion $[M+Na]^+$ m/z = 840 that was not present in the extracts of Δ*lab13* mutant. The molecular mass of this compound would correspond to the onnamide-like molecule containing the chain with the terminal amino acid arginine. Although this compound was not purified and the chemical structure was not elucidated yet, there are one additional indication that is in the agreement with the expected chemical profile of

the molecule. The compound strongly absorbs under UV light at 260 nm, that is compatible with the existence of double bonds in the molecule as expected (Matsunaga et al., 1992; Paul et al., 2002).

5. Post-PKS tailoring modifications of labrenzin

5.1. *O*-Methylation

Since the first discovery of the *ped* cluster in the beetle *P. fuscipes* metagenome almost two decades ago (Piel, 2002), there have been identified five more homologous gene clusters in the genomes of symbionts responsible for the biosynthesis of structurally highly similar compounds such as, psymberin (Fisch et al., 2009), diaphorin (Nakabachi et al., 2013), nosperin (Kampa et al., 2013), cusperin (Kust et al., 2018) and mycalamide A (Rust et al., 2020). In spite of the fact that *O*-methylation plays an important role in the diversity of these compounds, there was only one experimental study to study the *O*-methylation in pederin-family polyketides (Zimmermann et al., 2009). The *in vitro* biochemical assays reported that the methyltransferase PedO methylated the available 18C-OH group in mycalamide A.

In this thesis, it has been confirmed the role of PedO since its expression allowed to generate pederin in the wild type strain PHM005 providing the labrenzin with an additional methylation on C18-OH. On the other hand, PedO also allowed generating two new dimethylated biosynthetic analogs in the *Δlab6* and *Δlab16* mutant strains suggesting a double function for this methylase. Interestingly, the double function of PedO was only noticeable upon the intermediates lacking the C10-OH group suggesting that PedO can only perform the second methylation if the C10 position remains non hydroxylated. A similar finding was observed when PedO was unable to methylate the C18-OH group of the mycalamide A intermediate when the neighboring -OH was methylated, *i.e.* C18 *O*-methylation appeared only possible when the C17-OH was non methylated (Zimmermann et al., 2009).

The Δ*lab6* and Δ*lab16* mutant strains expressing PedO showed dominant MS peaks that corresponded to methylated hybrids of C10-dehydroxylated intermediates also indicating a possibility of substrate preference. In this work, PedO was expressed in the MTs mutants but it would be interesting to express it in mutants lacking the hydroxylation functions and test the preference trend. While it is difficult to deduce the correct methylation order of the tailoring

intermediates based only on the MS results, they did indicate some level of sequentiality in the decoration process including both hydroxylations and methylations.

Another example suggesting the substrate preference of MTs would be the poor pederin accumulation in the culture extracts obtained from the wt strain with PedO expressed in the conditions of high labrenzin production established by PharmaMar. If the premise that the C18 *O*-methylation only occurs when the neighboring C17 OH remains demethylated is true (Zimmermann et al., 2009), it may explain the low abundance of pederin compared to labrenzin in the culture extract. On the contrary, in the MBM+vit medium the same expression system yielded dominant pederin versus labrenzin intermediate. One possibility for controlled ratio of labrenzin and pederin could be the fine-tuning of the Lab6 and PedO expression.

As already mentioned, pederin has four *O*-methylations, but only three MTs are present in the *ped* cluster. Knowing that PedO is responsible for the C18 *O*-methylation, it was considered that one of the other MTs, PedA or PedE, would have a double function (Zimmermann et al., 2009). Analogously to pederin, labrenzin is decorated by three *O*-methylations but *lab* cluster only has two MTs, Lab6 and Lab16, homologues of PedA and PedE, respectively (Figure 17). The functional genetic analysis conducted in this thesis revealed the dual function of Lab6, by methylating C10-OH and C17-OH groups. The knockout of *lab6* gene, also demonstrated the role of *lab16* gene. Nonetheless, the *lab16* deletion confirmed the roles of MTs and the elimination of both MTs excluded the possibility of methylation redundancy. In addition, the results obtained by MS analysis showed that the methylation do not occur strictly sequentially. MS analysis of the mutant strains involved in the reactions of hydroxylation showed that the methylations occur even when C7-OH or C10-OH groups are missing, indicating that these MTs accept different substrates. Unexpectedly, the C6-*O*-methylation performed by Lab16 stabilizes the chirality of the intermediates. To obtain a more precise information about the substrate preferences, additional biochemical assays should be performed.

Finally, it was interesting to dicuss why *lab* and *ped* clusters differ in one MT, *i.e.*, the PedO MT. According to the theory of the Selfish Operon model (Roth and Lawrence, 1996), horizontal transfer drives the genes to evolve by forming clusters with the aim that the functionally related genes disseminate together. Interestingly, diaphorin gene cluster, identified from a reduced genome

of obligate symbiont and divided between two loci, contains only one MT. One possibility is that this polyketide gene cluster was gaining new functions as it was being transfered horizontally across different bacterial taxa, and that eventually led to the evolution of *ped* and *lab* gene clusters. It is obvious that the PedO function "competes" with the unprocessed complete product, which is expected to be similar to ester-like polyketide sesbanimide F produced by *S. indica* PHM037. Therefore, it is possible that the presence of PedO depends on the efficiency of the hydrolysis of the polyketide ester, considering it leaves the C18-OH group free for the methylation. As the hydrolysis was becoming more efficient, due to the development of the host metabolism, intracellular redox state or the presence of another esterase-like enzyme in the host genome, it was more likely for the cluster to acquire another function. Nervertheless, other options can be envisioned to explain the presence of different MTs in the perderin-like clusters.

5.2. Hydroxylation

The functional genetic analyses carried out upon oxidoreductases and cytochrome P450 (CYP450) from the *lab* cluster revealed some interesting and unexpected insights into the hydroxylation of the labrenzin and its analogs. CYP450s are universally spread enzymes responsible for the most versatile reactions in nature in terms of substrate range and transformations diversity (Munro et al., 2013). So far, numerous CYP450s involved in polyketide biosynthesis have been characterized, catalyzing diverse reactions such as desaturation (Reddick et al., 2007), hydroxylation (Mo et al., 2019; Stassi et al., 1993) and epoxidation (Ogura et al., 2004). Until 2017 around 200 CYP450s engaged in 14 different transformations, mostly hydroxylations, had been characterized only in *Streptomyces* (Rudolf et al., 2017).

The *lab* cluster contains genes encoding putative oxidoreductases Lab5, Lab8 and CYP450 (Lab7) that were considered good candidates for the tailoring hydroxylation reactions. In this work, the indispensable role of CYP450 and the associated reductase Lab8 in the C10-OH group formation was demonstrated. Initially, Lab7 or Lab8 would be involved in downstream decoration functions and Lab5 could have a dual function, as it was showed for MT Lab6. Instead, the knockout of both genes and the following complementation experiments revealed the involvement of Lab7 and Lab8 in C10 hydroxylation. One indication that those enzymes would act in a codependent way, is their genetic organization in an operon with overlapping ORFs. On the other

hand, most of the CYP450s interact with diverse redox partners (RPs) and accordingly distribute among up to 10 different classes of cytochrome systems (Li et al., 2020). Since Lab8 is annotated as a FMN-dependent oxidoreductase, this hydroxylating system could belong to Class III, which is a three-component system consisting on a ferredoxin reductase, FMN-containing flavodoxin (Fld) and P450.

Lab5, on the other hand, is annotated as LLM class flavin-dependent oxidoreductase and according to Pfam, the closest homolog is luciferase-like monooxygenase (E-value 7.8e-41). Moreover, the experimental analysis demonstrated it is responsible for the introduction of the C7-OH group. The CYP450 is not involved in this reaction because the *lab7-lab8* knockout did not affect C7-hydroxylation. This indicated that Lab5 belong to a different class of reductases than Lab8 and most likely it is sufficient to generate the hydroxylation. A different mechanism of action is also envisioned by the particular atomic environment and electron distribution in a direct proximity of C7 and C10. Another possibility is that some other CYP450 from the *Labrenzia* sp. PHM005 genome might assist this reaction, although, as demonstrated throughout this work, labrenzin related genes are highly specific and show a great level of clustering organization, through operons and overlapping ORFs.

Recollecting the results of all ten intermediates accumulated by different mutants generated during this work, it can be deducted the polyketide tailoring pathway, as it would occur in the wild type PHM005 and the alternative pathways that would lead to the accumulation of certain intermediates in the mutant strains (Figure 62). Lab5 knockout also showed that Lab6 is substrate-tolerant because it is able to methylate different labrenzin analogs, as it revealed new methylated intermediates lacking C7-OH group. Tandem knockout of Lab5 and Lab6 confirmed the established tailoring pathway and expanded the repertoire of labrenzin analogs.

Interestingly, the MS analysis of the peaks abundance in crude extracts of Δ*lab5* and Δ*lab5*-Δ*lab6* mutants, as well as in the extracts of Δ*lab6* and Δ*lab6*-Δ*lab16* mutants indicated a bottleneck in the tailoring flow that is the hydroxylation of the carbon C10 by the hydroxylase Lab7-Lab8.

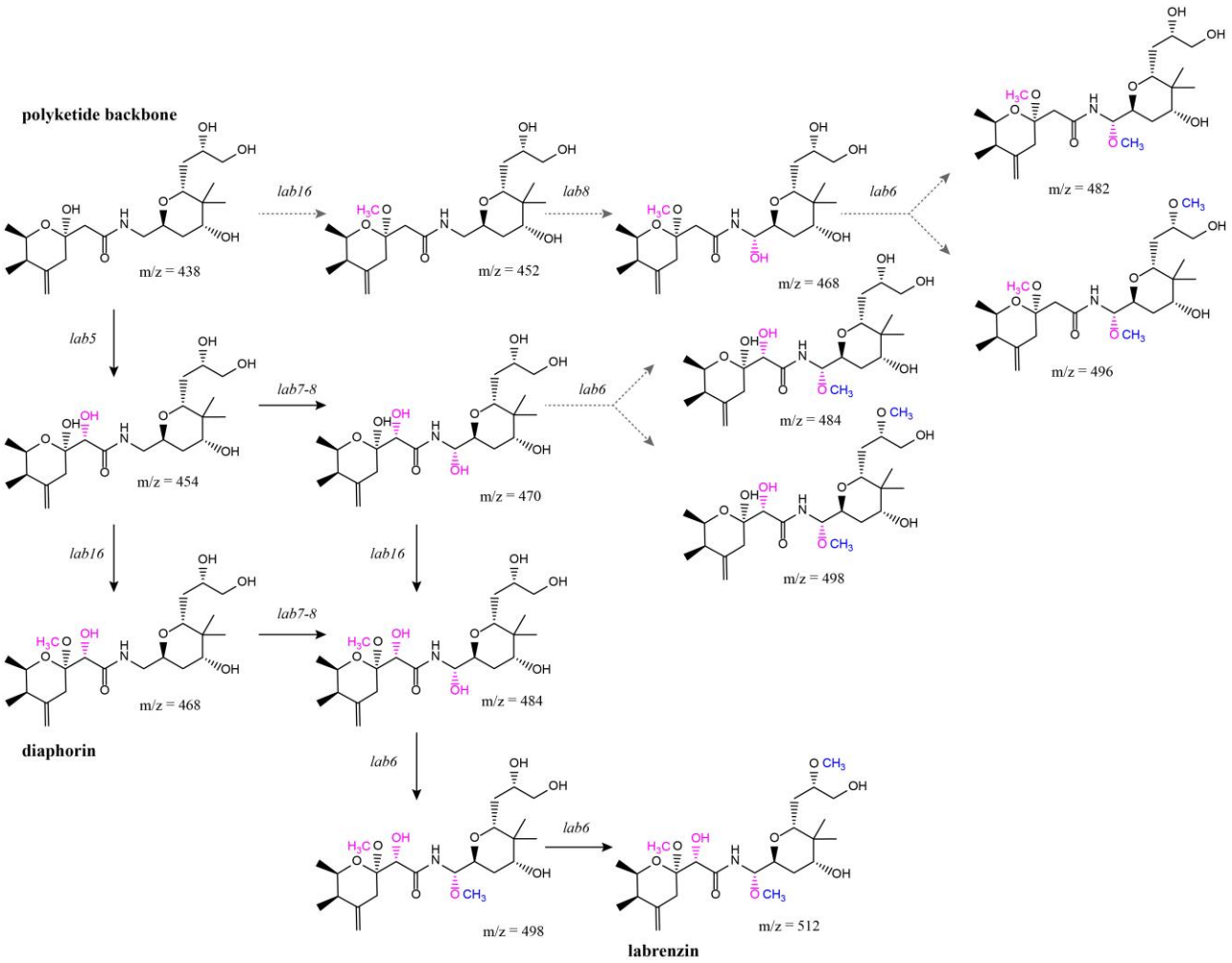


Figure 62. Pathway of labrenzin tailoring reactions in *Labrenzia* sp. PHM005. Continuous line represents the reactions flow in the wild type strain and the dashed line the secondary reactions in mutant strains.

6. Deciphering the labrenzin assembly line and future prospects

Finally, based on the results obtained in this thesis, it can be propose a correction in the initial biosynthesis hypothesis and present the current pathway of the full labrenzin assembly by PKS/NRPS complex including tailoring modifications (Figure 62). The significant change to the initial scheme is due to the oxygen insertion by Baeyer-Villiger monooxygenase (OXY or Lab14) when the polyketide elongation continues by the Lab13 PKS/NRPS. This enzyme should produce an ester-polyketide that is analog to sesbanimide F. OXY is most likely essential for the proper polyketide backbone elongation rather than producing the polyketide hydrolysis, as previously hypothesized (Piel, 2010). Once synthetized, the final polyketide is being released from the

complex by the TE domain. The labrenzin is produced by an unknown hydrolytic mechanism in one or two consecutive steps.

The HMGS and ECH are crucial for the labrenzin assembly and they are likely involved in the early biosynthesis stage by catalyzing β-branching of *exo*-methylene residue. According to the results of this work, the free-standing KS^0 is redundant activity in β-branching or could be involved in downstream processing of the molecule.

Tailoring enzymes involved in labrenzin decoration are MTs (Lab6 and Lab16) and oxidoreductases (Lab5, Lab7 and Lab8). Lab16 methylates the C6-OH and Lab6 has a dual function by successively methylating both C10-OH and C17-OH group. Additionaly, in some clusters there would be a third methyltransferase PedO, catalyzing the C18-OH methylation (not shown in the figure). While the C6-OH group is introduced by modular modifications, C7-OH group is generated by individual Lab5 oxidoreductase. The Lab8 and Lab7 are both needed for the hydroxylation of the C10 carbon and not Lab5 as initially postulated.

Discussion

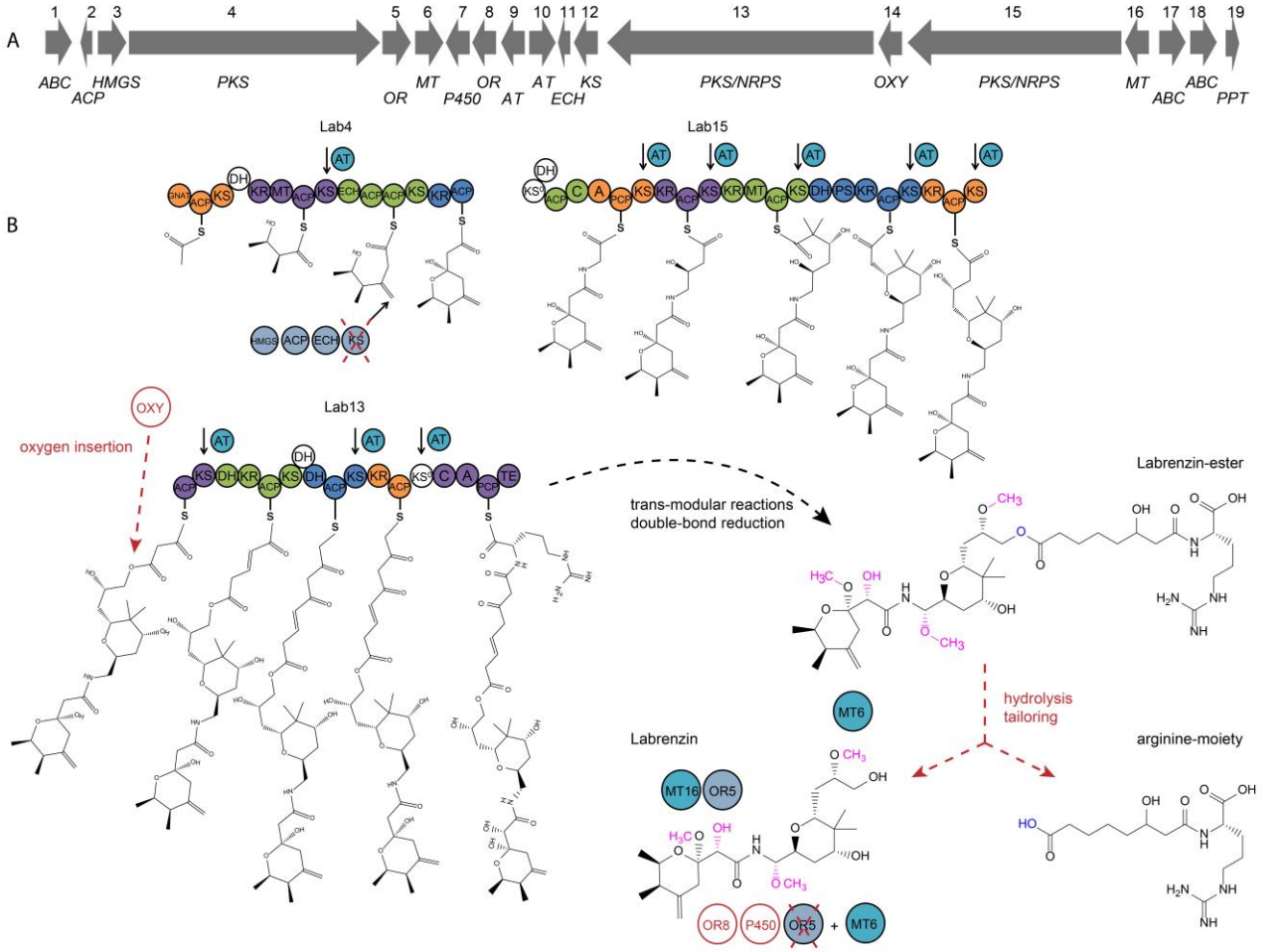


Figure 62. A) Map of the lab cluster comprising genes for labrenzin production. B) Corrected scheme predicting the labrenzin modular biosynthesis in a free-living Alfaproteobacterium *Labrenzia* sp. PHM005. Specific domains have been identified and analyzed using antiSMASH. Domains (shown in circles) that are part of the same module are represented in the same color, less the putative non-functional domains in white. Module boundaries are defined according to recent improvements (Vander Wood and Keatinge-Clay, 2018). GNAT, GCN5-related N-acetyltransferase domain; ACP, acyl carrier protein domain; KS, ketosynthase domain; KS0, non-elongating ketosynthase domain; KR, ketoreductase domain; MT, methyltransferase domain; ECH, enoyl- CoA hydratase/isomerase domain; A, adenylation domain; AT, acyltransferase; C, condensation domain; DH, dehydratase domain; PS, pyran synthase; PCP, peptidyl carrier protein domain; TE, thioesterase domain; OR, oxidoreductase; MT, methyltransferase. Putative *trans*-AT docking site associated with its respective KS is indicated by arrows. Modifications by tailoring enzymes are colored in pink. Specific corrections of the initial biosynthesis hypothesis are indicated in red.

To conclude, the biosynthesis of secondary metabolites is influenced by a wide variety of environmental and physiological signals, presumably reflecting the range of conditions that trigger their production in nature. The genetic manipulation of *Labrenzia* sp. PHM005 has opened the possibility of studying the pathway at a molecular level, not only to unravel the biosynthetic mechanisms for the sysnthesis of pederin-like compounds, but also to generate new pederin family analogs for pharmaceutical purposes. There are remarkable challenges ahead in understanding the regulatory cascades that link environmental and developmental signals to pleiotropic and ultimately pathway-specific regulatory genes for secondary metabolism. The availability of the entire genome sequence of *Labrenzia* sp. PHM005 and the possibility to develop transcriptome analyses and site-directed mutagenesis will definitively provide some clues to fully exploit the industrial potential of this strain.

Conclusions

1. The genome of the marine alphaproteobacterium *Labrenzia* sp. PHM005, a producer of polyketide labrenzin, consists of a circular chromosome (6.1 Mb) and a plasmid p1BIR (19 kb). Its closest related strain is *L. alexandrii* (strain DFL-11).

2. The production of labrenzin is associated with a *trans*-AT PKS gene cluster, denominated *lab* cluster, expanding over 79-kb genome region and comprising three genes encoding PKSs, 16 auxiliary genes and 6 ORFs with still unassigned function.

3. Homologs of *lab* gene cluster are widespread across phylogenetically distant bacterial phyla. Comparative genomics between clusters have contributed to decipher polyketide assembly pathways.

4. A new screening method to find effectors for labrencin production was created by testing molecules that significantly increase the transcription from the P_n promoter of *lab4* gene.

5. The *lab3*, *lab4*, *lab11* and *lab13* genes encoding respectively HMGS, PKS, ECH and PKS/NRPS are essential for the biosynthesis of labrenzin polyketide backbone.

6. The serine mutation of thioesterase catalytic site of Lab13 drastically decreases labrenzin production suggesting that the polyketide backbone has to be fully sinthetized and released from the PKS/NRPS complex for the further hydrolitic processing into labrenzin in the bacterial cytosol.

7. Although the knockout of *trans*-AT Lab9 significantly decreases the yields of labrenzin and its intermediates, *trans*-AT Lab10 alone is sufficient for the polyketide assembly.

8. Methyltransferases Lab6 and Lab16 are responsible for the tailoring reactions of *O*-methylation of the labrenzin molecular core. Lab6 has double function and sequentially methylates the C10-OH and C17-OH. Lab16 methylates the C6-OH and its knockout causes the quirality of the generated intermediates. Both MTs are substrate specific and their activity is not redundant.

Conclusions

9. Oxidoreductase Lab5 is involved in the hydroxylation at C7, whereas Lab7 and Lab8 are necessary for the hydroxylation at C10.

10. PedO encoded in the *P. fuscipes* symbiont can catalyse the methylation of C18-OH of labrenzin generating for the first time pederin in a cultivable bacterium.

11. Knockout experiments in the *Labrenzia* sp. PHM005 have allowed to generate ten different tailored labrenzin analogs.

Bibliography

Achtman, M., and Wagner, M. (2008). Microbial diversity and the genetic nature of microbial species. *Nat. Rev. Microbiol.* 6, 431–440. doi:10.1038/nrmicro1872.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9, 75. doi:10.1186/1471-2164-9-75.

Baumann, P., and Baumann, L. (1981). *The Prokaryotes: A Handbook on Habitats, Isolation and Identification of Bacteria*. , eds. M. P. Starr, H. Stolp, H. G. Trüper, A. Balows, and H. G. Schlegel Springer.

Bergmann, W., and Feeney, R. J. (1951). Contributions to the study of marine products. XXXII. The nuclesides of sponges. I. 1. *J. Org. Chem.* 16, 981–987. doi:10.1021/jo01146a023.

Bielitza, M., and Pietruszka, J. (2013). The psymberin story - Biological properties and approaches towards total and analogue syntheses. *Angew. Chemie - Int. Ed.* 52, 10960–10985. doi:10.1002/anie.201301259.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., et al. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* 47, W81–W87. doi:10.1093/nar/gkz310.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254.

Buchholz, T. J., Rath, C. M., Lopanik, N. B., Gardner, N. P., Håkansson, K., and Sherman, D. H. (2010). Polyketide β-branching in bryostatin biosynthesis: Identification of surrogate acetyl-ACP donors for BryR, an HMG-ACP synthase. *Chem. Biol.* 17, 1092–1100. doi:10.1016/j.chembiol.2010.08.008.

Calderone, C. T. (2008). Isoprenoid-like alkylations in polyketide biosynthesis. *Nat. Prod. Rep.* 25, 845–853. doi:10.1039/b807243d.

Calderone, C. T., Kowtoniuk, W. E., Kelleher, N. L., Walsh, C. T., and Dorrestein, P. C. (2006).

Convergence of isoprene and polyketide biosynthetic machinery: isoprenyl-S-carrier proteins in the pksX pathway of *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A.* 103, 8977–8982. doi:10.1073/pnas.0603148103.

Carroll, A. R., Copp, B. R., Davis, R. A., Keyzers, R. A., and Prinsep, M. R. (2019). Marine natural products. *Nat. Prod. Rep.* 36, 122–173. doi:10.1039/c8np00092a.

Chen, X. H., Vater, J., Piel, J., Franke, P., Scholz, R., Schneider, K., et al. (2006). Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42. *J. Bacteriol.* 188, 4024–4036. doi:10.1128/JB.00052-06.

Cichewicz, R. H., Valeriote, F. A., and Crews, P. (2004). Psymberin, a potent sponge-derived cytotoxin from *Psammocinia* distantly related to the pederin family. *Org. Lett.* 6, 1951–1954. doi:10.1021/ol049503q.

Cragg, G. M., and Newman, D. J. (2013). Natural products: a continuing source of novel drug leads. *Biochim. Biophys. Acta - Gen. Subj.* 1830, 3670–3695. doi:10.1016/j.bbagen.2013.02.008.

de Lorenzo, V., and Timmis, K. N. (1994). Analysis and construction of stable phenotypes in gram-negative bacteria with Tn5- and Tn10-derived minitransposons. *Methods Enzymol.* 235, 386–405. doi:10.1016/0076-6879(94)35157-0.

Dias, D. A., Urban, S., and Roessner, U. (2012). A historical overview of natural products in drug discovery. *Metabolites* 2, 303–336. doi:10.3390/metabo2020303.

Dinan, L. (2006). "Dereplication and Partial Identification of Compounds", in *Natural Products Isolation* (Totowa, NJ: Humana Press), 297–321. doi:10.1385/1-59259-955-9:297.

Dyshlovoy, S. A., and Honecker, F. (2019). Marine compounds and cancer: the first two decades of XXI century. *Mar. Drugs* 18, 20. doi:10.3390/md18010020.

El-Sayed, A. K., Hothersall, J., Cooper, S. M., Stephens, E., Simpson, T. J., and Thomas, C. M. (2003). Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586. *Chem. Biol.* 10, 419–430. doi:10.1016/S1074-5521(03)00091-7.

Ettwiller, L., Buswell, J., Yigit, E., and Schildkraut, I. (2016). A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. *BMC Genomics* 17, 1–14. doi:10.1186/s12864-016-2539-z.

Felnagle, E. A., Jackson, E. E., Chan, Y. A., Podevels, A. M., Berti, A. D., McMahon, M. D., et al. (2008). Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Mol. Pharm.* 5, 191–211. doi:10.1021/mp700137g.

Ferrières, L., Hémery, G., Nham, T., Guérout, A. M., Mazel, D., Beloin, C., et al. (2010). Silent mischief: bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *J. Bacteriol.* 192, 6418–6427. doi:10.1128/JB.00621-10.

Fisch, K. M., Gurgui, C., Heycke, N., van der Sar, S. a., Anderson, S. a., Webb, V. L., et al. (2009). Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. *Nat. Chem. Biol.* 5, 494–501. doi:10.1038/nchembio.176.

Frank, J. H., and Kanamitsu, K. (1987). *Paederus*, sensu lato (Coleoptera: Staphylinidae): natural history and medical importance. *J. Med. Entomol.* 24, 155–191.

Fusetani, N., Sugawara, T., and Matsunaga, S. (1992). Theopederins A-E, potent antitumor metabolites from a marine sponge, *Theonella* sp. *J. Org. Chem.* 57, 3828–3832. doi:10.1021/jo00040a021.

Garvey, P., Fitzgerald, G. F., and Hill, C. (1995). Cloning and DNA sequence analysis of two abortive infection phage resistance determinants from the lactococcal plasmid pNP40. *Appl. Environ. Microbiol.* 61, 4321–4328.

Goh, E. B., Yim, G., Tsui, W., McClure, J. A., Surette, M. G., and Davies, J. (2002). Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics. *Proc. Natl. Acad. Sci. U. S. A.* 99, 17025–17030. doi:10.1073/pnas.252607699.

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M.

(2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi:10.1099/ijs.0.64483-0.

Gu, L., Jia, J., Liu, H., Håkansson, K., Gerwick, W. H., and Sherman, D. H. (2006). Metabolic coupling of dehydration and decarboxylation in the curacin A pathway: functional identification of a mechanistically diverse enzyme pair. *J. Am. Chem. Soc.* 128, 9014–9015. doi:10.1021/ja0626382.

Haines, A. S., Dong, X., Song, Z., Farmer, R., Williams, C., Hothersall, J., et al. (2013). A conserved motif flags acyl carrier proteins for β-branching in polyketide synthesis. *Nat. Chem. Biol.* 9, 685–692. doi:10.1038/nchembio.1342.

Halton, B., and Clark, R. (2019). The International Symposium on Macrocyclic and Supramolecular Chemistry. in *Chemistry in New Zealand*.

Han, J. W., Ng, B. G., Sohng, J. K., Yoon, Y. J., Choi, G. J., and Kim, B. S. (2018). Functional characterization of O-methyltransferases used to catalyse site-specific methylation in the post-tailoring steps of pradimicin biosynthesis. *J. Appl. Microbiol.* 124, 144–154. doi:10.1111/jam.13619.

Piel, J. (2010). Biosynthesis of polyketides by *trans*-AT polyketide synthases. *Nat. Prod. Rep.* 27, 996–1047. doi: 10.1039/b816430b.

Helfrich, E. J. N., and Piel, J. (2016). Biosynthesis of polyketides by *trans*-AT polyketide synthases. *Nat. Prod. Rep.* 33, 231–316. doi:10.1039/c5np00125k.

Helfrich, E. J. N., Ueoka, R., Dolev, A., Rust, M., Meoded, R. A., Bhushan, A., et al. (2019). Automated structure prediction of *trans*-acyltransferase polyketide synthase products. *Nat. Chem. Biol.* 15, 813–821. doi:10.1038/s41589-019-0313-7.

Heo, C. C., Latif, B., Hafiz, W. M., and Zhou, H. Z. (2013). Dermatitis caused by *Paederus fuscipes* Curtis, 1840 (Coleoptera: Staphilinidae) in student hostels in Selangor, Malaysia. *Southeast Asian J. Trop. Med. Public Health* 44, 197.

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

Hothersall, J., Wu, J., Rahman, A. S., Shields, J. A., Haddock, J., Johnson, N., et al. (2007). Mutational analysis reveals that all tailoring region genes are required for production of polyketide antibiotic mupirocin by *Pseudomonas fluorescens*: Pseudomonic acid B biosynthesis precedes pseudomonic acid A. *J. Biol. Chem.* 282, 15451–15461. doi:10.1074/jbc.M701490200.

Jensen, K., Niederkrüger, H., Zimmermann, K., Vagstad, A. L., Moldenhauer, J., Brendel, N., et al. (2012). Polyketide proofreading by an acyltransferase-like enzyme. *Chem. Biol.* 19, 329–339. doi:10.1016/j.chembiol.2012.01.005.

Jirakkakul, J., Punya, J., Pongpattanakitshote, S., Paungmoung, P., Vorapreeda, N., Tachaleat, A., et al. (2008). Identification of the nonribosomal peptide synthetase gene responsible for Bassianolide synthesis in wood-decaying fungus *Xylaria* sp. BCC1067. *Microbiology* 154, 995–1006. doi:10.1099/mic.0.2007/013995-0.

Kampa, A., Gagunashvili, A. N., Gulder, T. A. M., Morinaka, B. I., Daolio, C., Godejohann, M., et al. (2013). Metagenomic natural product discovery in lichen provides evidence for a family of biosynthetic pathways in diverse symbioses. *Proc. Natl. Acad. Sci.* 110, E3129–E3137. doi:10.1073/pnas.1305867110.

Keating, T. A., Marshall, C. G., Walsh, C. T., and Keating, A. E. (2002). The structure of vibh represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.* 9, 522–526. doi:10.1038/nsb810.

Kellner, R. L. L. (2000). Possible genetic basis of pederin polymorphism in rove beetles (*Paederus riparius*). *J. Hered.* 91, 158–162. doi:10.1093/jhered/91.2.158.

Kellner, R. L. L. (2001). Suppression of pederin biosynthesis through antibiotic elimination of endosymbionts in *Paederus sabaeus*. *J. Insect Physiol.* 47, 475–483. doi:10.1016/S0022-1910(00)00140-2.

Bibliography

Kellner, R. L. L. (2002). Molecular identification of an endosymbiotic bacterium associated with pederin biosynthesis in *Paederus sabaeus* (Coleoptera: Staphylinidae). *Insect Biochem. Mol. Biol.* 32, 389–395. doi:10.1016/S0965-1748(01)00115-1.

Kellner, R. L. L., and Dettner, K. (1995). Allocation of pederin during lifetime of *Paederus* rove beetles (Coleoptera: Staphylinidae): Evidence for polymorphism of hemolymph toxin. *J. Chem. Ecol.* 21, 1719–1733. doi:10.1007/BF02033672.

Kind, T., and Fiehn, O. (2017). Strategies for dereplication of natural compounds using high-resolution tandem mass spectrometry. *Phytochem. Lett.* 21, 313–319. doi:10.1016/j.phytol.2016.11.006.

Kosol, S., Jenner, M., Lewandowski, J. R., and Challis, G. L. (2018). Protein–protein interactions in *trans* -AT polyketide synthases. *Nat. Prod. Rep.*, 1097–1109. doi:10.1039/C8NP00066B.

Kotowska, M., and Pawlik, K. (2014). Roles of type II thioesterases and their application for secondary metabolite yield improvement. *Appl. Microbiol. Biotechnol.* 98, 7735–7746. doi:10.1007/s00253-014-5952-8.

Kraas, F. I., Helmetag, V., Wittmann, M., Strieker, M., and Marahiel, M. A. (2010). Functional dissection of surfactin synthetase initiation module reveals insights into the mechanism of lipoinitiation. *Chem. Biol.* 17, 872–880. doi:10.1016/j.chembiol.2010.06.015.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi:10.1093/molbev/msw054.

Kurosa, K. (1958). Studies on poisonous beetles. III. Studies on the life history of *Paederus fuscipes* Curtis (Staphylinidae). *Japan. J. Sanit. Zool* 9, 245–276.

Kusebauch, B., Busch, B., Scherlach, K., Roth, M., and Hertweck, C. (2009). Polyketide-chain branching by an enzymatic Michael addition. *Angew. Chemie - Int. Ed.* 48, 5001–5004. doi:10.1002/anie.200900277.

Kust, A., Mareš, J., Jokela, J., Urajová, P., Hájek, J., Saurav, K., et al. (2018). Discovery of a

pederin family compound in a nonsymbiotic bloom-forming cyanobacterium. *ACS Chem. Biol.* 13, 1123–1129. doi:10.1021/acschembio.7b01048.

Leisch, H., Morley, K., and Lau, P. C. K. (2011). Baeyer-villiger monooxygenases: More than just green chemistry. *Chem. Rev.* 111, 4165–4222. doi:10.1021/cr1003437.

Li, S., Du, L., and Bernhardt, R. (2020). Redox partners: function modulators of bacterial P450 enzymes. *Trends Microbiol.* 28, 445–454. doi:10.1016/j.tim.2020.02.012.

Lim, S. K., Ju, J., Zazopoulos, E., Jiang, H., Seo, J. W., Chen, Y., et al. (2009). Iso-migrastatin, migrastatin, and dorrigocin production in *Streptomyces platensis* NRRL 18993 is governed by a single biosynthetic machinery featuring an acyltransferase-less type I polyketide synthase. *J. Biol. Chem.* 284, 29746–29756. doi:10.1074/jbc.M109.046805.

Lowry, B., Li, X., Robbins, T., Cane, D. E., and Khosla, C. (2016). A turnstile mechanism for the controlled growth of biosynthetic intermediates on assembly line polyketide synthases. *ACS Cent. Sci.* 2, 14–20. doi:10.1021/acscentsci.5b00321.

Martínez-García, E., and de Lorenzo, V. (2011). Engineering multiple genomic deletions in Gram-negative bacteria: analysis of the multi-resistant antibiotic profile of *Pseudomonas putida* KT2440. *Environ. Microbiol.* 13, 2702–2716. doi:10.1111/j.1462-2920.2011.02538.x.

Martins, A., Vieira, H., Gaspar, H., and Santos, S. (2014). Marketed marine natural products in the pharmaceutical and cosmeceutical industries: tips for success. *Mar. Drugs* 12, 1066–1101. doi:10.3390/md12021066.

Matilla, M. A., Stöckmann, H., Leeper, F. J., and Salmond, G. P. C. (2012). Bacterial biosynthetic gene clusters encoding the anti-cancer haterumalide class of molecules: Biogenesis of the broad spectrum antifungal and anti-oomycete compound, oocydin A. *J. Biol. Chem.* 287, 39125–39138. doi:10.1074/jbc.M112.401026.

Matsunaga, S., Fusetani, N., and Nakao, Y. (1992). Eight new cytotoxic metabolites closely related to onnamide A from two marine sponges of the genus *Theonella*. *Tetrahedron* 48, 8369–8376. doi:10.1016/S0040-4020(01)86585-6.

Bibliography

Mayer, A. (2020). Marine Pharmaceutical: the Clinical Pipeline. Available online: https://www.midwestern.edu/departments/marinepharmacology/clinical-pipeline.xml (accessed on August 2020). Available at: https://www.midwestern.edu/departments/marinepharmacology/clinical-pipeline.xml.

Meoded, R. A., Ueoka, R., Helfrich, E. J. N., Jensen, K., Magnus, N., Piechulla, B., et al. (2018). A polyketide synthase component for oxygen insertion into polyketide backbones. *Angew. Chemie - Int. Ed.* 57, 11644–11648. doi:10.1002/anie.201805363.

Mészáros, B., Erdös, G., and Dosztányi, Z. (2018). IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* 46, W329–W337. doi:10.1093/nar/gky384.

Miyanaga, A., Kudo, F., and Eguchi, T. (2018). Protein-protein interactions in polyketide synthase-nonribosomal peptide synthetase hybrid assembly lines. *Nat. Prod. Rep.* 35, 1185–1209. doi:10.1039/c8np00022k.

Mo, X., Gui, C., and Yang, S. (2019). Cytochrome P450 oxidase SlgO1 catalyzes the biotransformation of tirandamycin C to a new tirandamycin derivative. *3 Biotech* 9, 1–9. doi:10.1007/s13205-019-1611-1.

Mohammad, H. H., Connolly, J. A., Song, Z., Hothersall, J., Race, P. R., Willis, C. L., et al. (2018). Fine tuning of antibiotic activity by a tailoring hydroxylase in a *trans*-AT polyketide synthase pathway. *ChemBioChem* 19, 836–841. doi:10.1002/cbic.201800036.

Mootz, H. D., Schwarzer, D., and Marahiel, M. A. (2002). Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *ChemBioChem* 3, 490. doi:10.1002/1439-7633(20020603)3:6<490::AID-CBIC490>3.0.CO;2-N.

Munro, A. W., Girvan, H. M., Mason, A. E., Dunford, A. J., and McLean, K. J. (2013). What makes a P450 tick? *Trends Biochem. Sci.* 38, 140–150. doi:10.1016/j.tibs.2012.11.006.

Nakabachi, A., Ueoka, R., Oshima, K., Teta, R., Mangoni, A., Gurgui, M., et al. (2013). Defensive bacteriome symbiont with a drastically reduced genome. *Curr. Biol.* 23, 1478–1484.

doi:10.1016/j.cub.2013.06.027.

Narquizian, R., and Kocienski, P. J. (2000). The pederin family of antitumor agents: structures, synthesis and biological activity. *Ernst Schering Res. Found. Workshop*, 25–56. doi:10.1007/978-3-662-04042-3_2.

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

Newman, D. J., and Giddings, L. A. (2014). Natural products as leads to antitumor drugs. *Phytochem. Rev.* 13, 123–137. doi:10.1007/s11101-013-9292-6.

Ngatu, N. (2018). "Paederus Dermatitis: Environmental Risk Factors, Clinical Features, and Management", in *Occupational and Environmental Skin Disorders*, eds. N. Ngatu and M. Ikeda (Singapore: Springer), 151–157. doi:https://doi.org/10.1007/978-981-10-8758-5_14.

Nicoletti, R., and Vinale, F. (2018). Bioactive compounds from marine-derived *Aspergillus*, *Penicillium*, *Talaromyces* and *Trichoderma* species. *Mar. Drugs* 16, 15–17. doi:10.3390/md16110408.

Niehs, S. P., Kumpfmüller, J., Dose, B., Little, R. F., Ishida, K., Flórez, L. V., et al. (2020). Insect-associated bacteria assemble the antifungal butenolide gladiofungin by non- canonical polyketide chain termination. *Angew. Chemie Int. Ed.*, 1–6. doi:10.1002/anie.202005711.

O'Brien, R. V., Davis, R. W., Khosla, C., and Hillenmeyer, M. E. (2014). Computational identification and analysis of orphan assembly-line polyketide synthases. *J. Antibiot. (Tokyo).* 67, 89–97. doi:10.1038/ja.2013.125.

Ogura, H., Nishida, C. R., Hoch, U. R., Perera, R., Dawson, J. H., and Ortiz De Montellano, P. R. (2004). EpoK, a cytochrome P450 involved in biosynthesis of the anticancer agents epothilones A and B. Substrate-mediated rescue of a P450 enzyme. *Biochemistry* 43, 14712–14721. doi:10.1021/bi048980d.

Paul, G. K., Gunasekera, S. P., Longley, R. E., and Pomponi, S. A. (2002). Theopederins K and L. highly potent cytotoxic metabolites from a marine sponge *Discodermia* species. *J. Nat. Prod.* 65, 59–61. doi:10.1021/np0103766.

Pavan, M. (1958). Biochemical aspects of insect poisons. in *Proc. 4th Internat. Congr. Biochemistry*, 15–36.

Pavan, M., and Bo, G. (1953). Pederin, toxic principle obtained in the crystalline state from the beetle *Paederus fuscipes.* Curt. *Physiol Comp Oecol* 3, 307–312.

Pazirandeh, M., Chirala, S. S., and Wakil, S. J. (1991). Site-directed mutagenesis studies on the recombinant thioesterase domain of chicken fatty acid synthase expressed in *Escherichia coli. J. Biol. Chem.* 266, 20946–20952.

Pérez-Victoria, I., Martín, J., and Reyes, F. (2016). Combined LC/UV/MS and NMR strategies for the dereplication of marine natural products. *Planta Med.* 82, 857–871. doi:10.1055/s-0042-101763.

Perry, N. B., Blunt, J. W., Munro, M. H. G., and Pannell, L. K. (1988). Mycalamide A, an antiviral compound from a New Zealand sponge of the genus *Mycale. J. Am. Chem. Soc.* 110, 4850–4851. doi:10.1021/ja00222a067.

Piel, J. (2002). A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. *Proc. Natl. Acad. Sci. U. S. A.* 99, 14002–14007. doi:10.1073/pnas.222481399.

Piel, J., Butzke, D., Fusetani, N., Hui, D., Platzer, M., Wen, G., et al. (2005). Exploring the chemistry of uncultivated bacterial symbionts: antitumor polyketides of the pederin family. *J. Nat. Prod.* 68, 472–479. doi:10.1021/np049612d.

Piel, J., Höfer, I., and Hui, D. (2004a). Evidence for a symbiosis island involved in horizontal acquisition of pederin biosynthetic capabilities by the bacterial symbiont of *Paederus fuscipes* beetles. *J. Bacteriol.* 186, 1280–1286. doi:10.1128/JB.186.5.1280–1286.2004.

Piel, J., Hui, D., Fusetani, N., and Matsunaga, S. (2004b). Targeting modular polyketide synthases with iteratively acting acyltransferases from metagenomes of uncultured bacterial consortia. *Environ. Microbiol.* 6, 921–927. doi:10.1111/j.1462-2920.2004.00531.x.

Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N., et al. (2004c). Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 16222–16227. doi:10.1073/pnas.0405976101.

Piel, J., Wen, G., Platzer, M., and Hui, D. (2004d). Unprecedented diversity of catalytic domains in the first four modules of the putative pederin polyketide synthase. *ChemBioChem* 5, 93–98. doi:10.1002/cbic.200300782.

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., and Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Res.* 46, W200–W204. doi:10.1093/nar/gky448.

Quilico, A., Cardani, C., Ghiringhelli, D., and Pavan, M. (1961). Pederina e pseudopederina. *Chim Ind.* 43, 1434–1436.

Ramirez-Llodra, E., Brandt, A., Danovaro, R., De Mol, B., Escobar, E., German, C. R., et al. (2010). Deep, diverse and definitely different: Unique attributes of the world's largest ecosystem. *Biogeosciences* 7, 2851–2899. doi:10.5194/bg-7-2851-2010.

Reddick, J. J., Antolak, S. A., and Raner, G. M. (2007). PksS from *Bacillus subtilis* is a cytochrome P450 involved in bacillaene metabolism. *Biochem. Biophys. Res. Commun.* 358, 363–367. doi:10.1016/j.bbrc.2007.04.151.

Reimer, J. M., Haque, A. S., Tarry, M. J., and Schmeing, T. M. (2018). Piecing together nonribosomal peptide synthesis. *Curr. Opin. Struct. Biol.* 49, 104–113. doi:10.1016/j.sbi.2018.01.011.

Robbel, L., and Marahiel, M. A. (2010). Daptomycin, a bacterial lipopeptide synthesized by a nonribosomal machinery. *J. Biol. Chem.* 285, 27501–27508. doi:10.1074/jbc.R110.128181.

Rodriguez-R, L. M., and Konstantinidis, K. T. (2016). The enveomics collection: a toolbox for

specialized analyses of microbial genomes and metagenomes. *PeerJ Prepr.* 4, e1900v1. doi:10.7287/peerj.preprints.1900v1.

Ross, A. C., Xu, Y., Lu, L., Kersten, R. D., Shao, Z., Al-Suwailem, A. M., et al. (2013). Biosynthetic multitasking facilitates thalassospiramide structural diversity in marine bacteria. *J. Am. Chem. Soc.* 135, 1155–1162. doi:10.1021/ja3119674.

Roth, J. R., and Lawrence, J. G. (1996). Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* 143, 1843–1860. doi:10.1261/rna.130506.

Röttig, M., Medema, M. H., Blin, K., Weber, T., Rausch, C., and Kohlbacher, O. (2011). NRPSpredictor2 - a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* 39, 362–367. doi:10.1093/nar/gkr323.

Rudolf, J. D., Chang, C.-Y., Ma, M., and Shen, B. (2017). Cytochromes P450 for natural product biosynthesis in *Streptomyces*: sequence, structure, and function. *Nat. Prod. Rep.* 34, 1141–1172. doi:10.1039/C7NP00034K.

Rust, M., Helfrich, E. J. N., Freeman, M. F., Nanudorn, P., Field, C. M., Rückert, C., et al. (2020). A multiproducer microbiome generates chemical diversity in the marine sponge *Mycale hentscheli*. *Proc. Natl. Acad. Sci. U. S. A.* 117, 9508–9518. doi:10.1073/pnas.1919245117.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425. doi:10.1093/oxfordjournals.molbev.a040454.

Sakemi, S., Ichiba, T., Kohmoto, S., Saucy, G., and Higa, T. (1988). Isolation and structure elucidation of onnamide A, a new bioactive metabolite of a marine sponge, *Theonella* sp. *J. Am. Chem. Soc.* 110, 4851–4853. doi:10.1021/ja00222a068.

Schleissner, C., Cañedo, L. M., Rodríguez, P., Crespo, C., Zúñiga, P., Peñalver, A., et al. (2017). Bacterial production of a pederin analogue by a free-living marine alphaproteobacterium. *J. Nat. Prod.* 80, 2170–2173. doi:10.1021/acs.jnatprod.7b00408.

Schneider, A., and Marahiel, M. A. (1998). Genetic evidence for a role of thioesterase domains,

integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in *Bacillus subtilis*. *Arch. Microbiol.* 169, 404–410. doi:10.1007/s002030050590.

Seo, J. W., Ma, M., Kwong, T., Ju, J., Lim, S. K., Jiang, H., et al. (2014). Comparative characterization of the lactimidomycin and iso-migrastatin biosynthetic machineries revealing unusual features for acyltransferase-less type I polyketide synthases and providing an opportunity to engineer new analogues. *Biochemistry* 53, 7854–7865. doi:10.1021/bi501396v.

Seyedsayamdost, M. R. (2014). High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *Proc. Natl. Acad. Sci. U. S. A.* 111, 7266–7271. doi:10.1073/pnas.1400019111.

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

Silva-Rocha, R., Martínez-García, E., Calles, B., Chavarría, M., Arce-Rodríguez, A., de las Heras, A., et al. (2013). The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res.* 41, D666–D675. doi:10.1093/nar/gks1119.

Simunovic, V., and Müller, R. (2007). 3-Hydroxy-3-methylglutaryl-CoA-like synthases direct the formation of methyl and ethyl side groups in the biosynthesis of the antibiotic myxovirescin A. *ChemBioChem* 8, 497–500. doi:10.1002/cbic.200700017.

Smith, S., and Tsai, S.-C. (2007). The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.* 24, 1041. doi:10.1039/b603600g.

Soldati, M., Fioretti, A., and Ghione, M. (1966). Cytotoxicity of pederin and some of its derivatives on cultured mammalian cells. *Experientia* 22, 176–178. doi:10.1007/BF01897720.

Stassi, D., Donadio, S., Staver, M. J., and Katz, L. (1993). Identification of a *Saccharopolyspora erythraea* gene required for the final hydroxylation step in erythromycin biosynthesis. *J. Bacteriol.* 175, 182–189. doi:10.1128/jb.175.1.182-189.1993.

Bibliography

Stonik, V. A. (2009). Marine natural products: a way to new drugs. *Acta Naturae* 1, 15–25. doi:10.32607/20758251-2009-1-2-15-25.

Stulberg, E. R., Lozano, G. L., Morin, J. B., Park, H., Baraban, E. G., Mlot, C., et al. (2016). Genomic and secondary metabolite analyses of *Streptomyces* sp. 2AW provide insight into the evolution of the cycloheximide pathway. *Front. Microbiol.* 7, 1–12. doi:10.3389/fmicb.2016.00573.

Sundaram, S., and Hertweck, C. (2016). On-line enzymatic tailoring of polyketides and peptides in thiotemplate systems. *Curr. Opin. Chem. Biol.* 31, 82–94. doi:10.1016/j.cbpa.2016.01.012.

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

Suwannahitatorn, P., Jatapai, A., and Rangsin, R. (2014). An outbreak of *Paederus* dermatitis in Thai military personnel. *J. Med. Assoc. Thail.* 97, S96–S100.

Tang, M.-C., He, H.-Y., Zhang, F., and Tang, G.-L. (2013). Baeyer–Villiger oxidation of acyl carrier protein-tethered thioester to acyl carrier protein-linked thiocarbonate catalyzed by a monooxygenase domain in FR901464 biosynthesis. *ACS Catal.* 3, 444–447. doi:10.1021/cs300819e.

Tavernarakis, N., Driscoll, M., and Kyrpides, N. C. (1999). The SPFH domain: Implicated in regulating targeted protein turnover in stomatins and other membrane-associated proteins. *Trends Biochem. Sci.* 24, 425–427. doi:10.1016/S0968-0004(99)01467-X.

Taylor, M. W., Radax, R., Steger, D., and Wagner, M. (2007). Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol. Mol. Biol. Rev.* 71, 295–347. doi:10.1128/MMBR.00040-06.

Thomford, N. E., Senthebane, D. A., Rowe, A., Munro, D., Seele, P., Maroyi, A., et al. (2018). Natural products for drug discovery in the 21st century: innovations for novel drug discovery. *Int. J. Mol. Sci.* 19. doi:10.3390/ijms19061578.

Tiboni, O., Parisi, B., and Ciferri, O. (1968). The mode of action of pederin, a drug inhibiting protein synthesis in eucaryotic organisms. *G. Bot. Ital.* 102, 337–345. doi:10.1080/11263506809426470.

Tomm, H. A., Ucciferri, L., and Ross, A. C. (2019). Advances in microbial culturing conditions to activate silent biosynthetic gene clusters for novel metabolite production. *J. Ind. Microbiol. Biotechnol.* 46, 1381–1400. doi:10.1007/s10295-019-02198-y.

Vander Wood, D. A., and Keatinge-Clay, A. T. (2018). The modules of *trans*-acyltransferase assembly lines redefined with a central acyl carrier protein. *Proteins Struct. Funct. Bioinforma.* 86, 664–675. doi:10.1002/prot.25493.

Walker, P. D., Weir, A. N. M., Willis, C. L., and Crump, M. P. (2020). Polyketide β-branching: diversity, mechanism and selectivity. *Nat. Prod. Rep.* doi:10.1039/d0np00045k.

Wan, S., Wu, F., Rech, J. C., Green, M. E., Balachandran, R., Horne, W. S., et al. (2011). Total synthesis and biological evaluation of pederin, psymberin, and highly potent analogs. *J. Am. Chem. Soc.* 133, 16668–16679. doi:10.1021/ja207331m.

Wang, B., Song, Y., Luo, M., Chen, Q., Ma, J., Huang, H., et al. (2013). Biosynthesis of 9-methylstreptimidone involves a new decarboxylative step for polyketide terminal diene formation. *Org. Lett.* 15, 1278–1281. doi:10.1021/ol400224n.

Weber, T., and Marahiel, M. A. (2001). Exploring the domain structure of modular nonribosomal peptide synthetases. *Structure* 9, R3. doi:10.1016/S0969-2126(00)00560-8.

Webster, N. S., and Taylor, M. W. (2012). Marine sponges and their microbial symbionts: love and other relationships. *Environ. Microbiol.* 14, 335–346. doi:10.1111/j.1462-2920.2011.02460.x.

Wilson, M. C., and Piel, J. (2013). Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chem. Biol.* 20, 636–647. doi:10.1016/j.chembiol.2013.04.011.

Witkowski, A., Witkowska, H. E., and Smith, S. (1994). Reengineering the specificity of a serine

active-site enzyme. Two active- site mutations convert a hydrolase to a transferase. *J. Biol. Chem.* 269, 379–383.

Wu, F., Green, M. E., and Floreancig, P. E. (2011). Total synthesis of pederin and analogues. *Angew. Chemie - Int. Ed.* 50, 1131–1134. doi:10.1002/anie.201006438.

Wu, G., Yan, Q., Jones, J. A., Tang, Y. J., Fong, S. S., and Koffas, M. A. G. (2016). Metabolic burden: cornerstones in synthetic biology and metabolic engineering applications. *Trends Biotechnol.* 34, 652–664. doi:10.1016/j.tibtech.2016.02.010.

Xu, Y., Kersten, R. D., Nam, S. J., Lu, L., Al-Suwailem, A. M., Zheng, H., et al. (2012). Bacterial biosynthesis and maturation of the didemnin anti-cancer agents. *J. Am. Chem. Soc.* 134, 8625–8632. doi:10.1021/ja301735a.

Yin, M., Yan, Y., Lohman, J. R., Huang, S.-X., Ma, M., Zhao, G.-R., et al. (2014). Cycloheximide and actiphenol production in *Streptomyces* sp. YIM56141 governed by single biosynthetic machinery featuring an acyltransferase-less type I polyketide synthase. *Org. Lett.* 16, 3072–3075. doi:10.1021/ol501179w.

Yu, D., Xu, F., Zhang, S., and Zhan, J. (2017). Decoding and reprogramming fungal iterative nonribosomal peptide synthetases. *Nat. Commun.* 8, 1–11. doi:10.1038/ncomms15349.

Zarins-Tutt, J. S., Barberi, T. T., Gao, H., Mearns-Spragg, A., Zhang, L., Newman, D. J., et al. (2016). Prospecting for new bacterial metabolites: a glossary of approaches for inducing, activating and upregulating the biosynthesis of bacterial cryptic or silent natural products. *Nat. Prod. Rep.* 33, 54–72. doi:10.1039/c5np00111k.

Zhang, J. J., Tang, X., Huan, T., Ross, A. C., and Moore, B. S. (2020). Pass-back chain extension expands multimodular assembly line biosynthesis. *Nat. Chem. Biol.* 16, 42–49. doi:10.1038/s41589-019-0385-4.

Zimmermann, K., Engeser, M., Blunt, J. W., Munro, M. H. G., and Piel, J. (2009). Pederin-type pathways of uncultivated bacterial symbionts: analysis of O-methyltransferases and generation of a biosynthetic hybrid. *J. Am. Chem. Soc.* 131, 2780–2781. doi:10.1021/ja808889k.

Zumberge, J. A., Love, G. D., Cárdenas, P., Sperling, E. A., Gunasekera, S., Rohrssen, M., et al. (2018). Demosponge steroid biomarker 26-methylstigmastane provides evidence for Neoproterozoic animals. *Nat. Ecol. Evol.* 2, 1709–1714. doi:10.1038/s41559-018-0676-2.

Annexes

1. Annex I: DNA sequences used for manipulations

>promoter region Pn (*lab4*)

TTACCATGCCGCAATACGAGGCATTGCTTGAAGGTTGCAAGGCTGTTCCCTTCGGCACG
CGCAACCACCAACCAGATCTTGATCAGGTTCCGGACATGAAATCCTGCATTGCCGATC
AAAGCGCCCAGCTCGGATATCAGCGGCTCTTCCTGAAAGAAATCAAAAACTTCCATCG
CGAATACGATGTACTTTGAGTTGTGTTGTCTCCTCTGCTCCGATAGGCTTACCCAAGGA
TACTTTTAAGAGCGCTTGTCTGCGATACTGGACGTTCCCATCGCAGCAGGCGATGTGCG
AGGGAAAATGCCATTCACGCATTTCGGCAAATCCTACCGAAGCTCTGAGGTGTTGCT

*TATA BOX; TSS indicated in red

>MTs operon construction

GAGCTCGTTTAAACAAGAGGTAATATCATGTCTATATCCGCACCTGATTTTGTAGGCT
TGCTAAAACAGCGTCTGCTGGATAATATTTCGACACAGATAGACCATGTTTTTAATGTG
CCTGCCGATATACAAAAGATGCTGGGCGGCAGTGAGAAAAATCTGTCATTGGAAGAGA
GCGCGAAGATATTCGATGCCGGCATTGTGGAACTGATGCGGAAACCAGAAGAGGATTC
TTCTAAAGAAGGAGGCTCATCTTTCATTAGTCATGTAAATAACCATTACGATCATGTAT
TTTATGACGATAATGGAATGACGGGAGTGTTATTCGGTGATACAGACTATCGGAATCA
CGGCTACTGGGATAGAAATACCGTTTCTCAAGATCAAGCATGCCGGCAATTGCAAGAA
ATACTGCTTGACTTCATCCCAGTGAAAACAGGAAAAATTTTAGACGTCGCGTGTGGAA
TGGGGGCTTCTACTCGTCATCTGCTTAAATACTATCCAGCCGAGAATATATGGGCTATC
AATATTTCGGATAAACAGATCGATACAGCGCGCCGAAATGCACCAGGTTGCCATGTGC
AGGTCATGGATGCCACGAACATGAGCTTCGCTGATGAAGCATTTGAAAATATTCTGTG
CATAGAAGCTGCCTTCCATTTTAACACGCGGCGCAAGTTCTTAGAGGAGGCTTTACGTA
TCCTCAAACCTGGTGGAAGGGTGGTGTTGTCGGATTTTATCTTCAGCTCACCTGAACGA
CTGGAGCAAAATAATATACTTCCTGGCCCAGTCAATCATCTAGCGTCAATTGAGGAAT
ACGCTCAACTTTTGAATGACGTGGGTTTCAGCGATTTCACTATTCAAGACGTAAGTGAT
GAGGTTTGGGGCGGACAGTTCCTGAATGGTACGAGCAAGCTGCATAAAGCATTTTACG
ATGGTAAGCTTGATATCGTTCGCATGACTGAAGCACTATGGTGTTATTACTATATAAAT
GCTTCTACCGAAAAGAGCCTTTTCGTATCTGCACAAAAA**TGA*GTTTAAACTTATAA**A

GCAGGAAATCCAATATGGCTAGCGAACTCAAGGATCTGCGACAGCGGTTGGTTGACCG
GCTTTCGGCTACGGTAGAGCAGAAGATTTCGTCAATCGGATACGTGCCCGAAGATTTG
GTCCGCATTGCGGGCTCCGGCGTGCCAGCAGAACCCAGTCATGATGAAGTCTATAAAG
CCCCGGAGGACTTGAAAGAGGCCATCAACGAACACTACGATTTCTCGTTTTATGCTCGC
GAGACGATCTGGGCCGATATGCTTGCTGGCACGCATTTTCGAAATATTGGCTATTGGGA
TGCAAATACTGAATCTCTGGATCAGGCCGGCCGCAATTTGCAGGATCAACTCCTGGCA
CTATTGCCTCAAAAAACCGGACGGATCCTTGACGTAGCCTGCGGGATGGGCGCCTCTA
CAAAACGGCTTCTGGACACTTACCGGCCCGAAGATGTGTGGGCCATCAACATCTCTGC
CAAACAAATCGAAACCACCTCTCAAAACGCTCCAGGCTGCAATGCACAAGTCATGAGC
GCAACGGAGATGACTTTTGAAGACAATTTTTTTGATGCTGTCGAATGCATCGAAGCCGC
TTTTCATTTCGACACGCGGCGCAAGTTTCTGGAAGACACCCTGCGCATTCTGAAGCCGG
GAGGCCGCTTGGTCATGTCCGATGTTCTGATGACTTCAGGGGCTCGGCTGGAGCAATAT
CCGGTGTTCCCCAACCCGGAAAACCACATTGCCACCATCGAAGATTACAAGTCTGTCTT
GGAAGAAATCGGATACGAAAACATCACAATATCTGATGAGCGGAACAATATTTGGAA
ATCGCATTTCATGGCCACAACCAACCGGATTCACGAAGGATTTCTAGCACGGAAGTAT
AATATCGTTGAGGTCACAGACATGATCTGGACGTATTACGAGTTGGATGCAATTACCG
GCCCTTGCCCGATCCTGGGCGCATCTAAACCTCGCTAA*TTATAACCCGGGAAGGAAA
AGACCTATGTCCCGCAGTACTTTGGAGACTACCGGCGCTAGCAATGATACGGTCGAAG
ATCATTATGACTCACCGGCTCTGCGACTGGGACCGATCCTGTTTGATGAACACTTACAC
TGGGGTTATTGGGACGAAGACAGTCGGGATGCAAGTTTCGGTGCCGCAGCGGAAGCCA
TGTGTCATCGAATGATCGATCGGACCGAAATTGGTCCCGGTGAACGGTTTGTCGATCTG
GGTTGTGGTATTGGCCATCCCGCCTTGAAACTTGCTCAAGCTCGAAGTTGTCATGTCAC
CGGAGTGACGATTAGTGGCTACCAGCATCGTATTGCAGGCGAGAAAGCGGCACAGGC
AGGGTTCTCCGACAGATTGGATTTTTTACAAGCCGATGCCCGCAGCGTCCCATTGCCAG
ACAAGAGTTTTGATGGCGGCTGGTTTTTTGAATCGATTTTTCACATGGGTCATGCAGAA
GCTTTGGGCGAAGCCGCCCGGCTTTTGAAACCCGGCGCAGGCCTTGTCCTGACCGACCT
GCCAACTCTGCCTCACACGACACCTGAATTCATGGACTTTGTCCATGAGCATATCCATT
CGGTCTTTGTTCCAGAAGATCGCTATCCCGCCTTGATGGCAGATGCCGGATTTGAACTT
CTGAACATTGAGGACATTTCCGAAAACGTCATGCCATGGCTGGAAACCAAGTTGCGTG
AAGCCGTGCAGGAAAAGTGGAGTGACGTGGTCCGGCTCATGGGAGATCAGGCAGAAA

AAGCCGTCGACAACTGGTACTATCTATTTGAATATATGGCAGAAAACCTCGGATACAC
CATGATTACGGCCCGCCGATTG**TAA*CCCGGGACTAGT**

2. Annex II: Preliminary and confirmation analyses

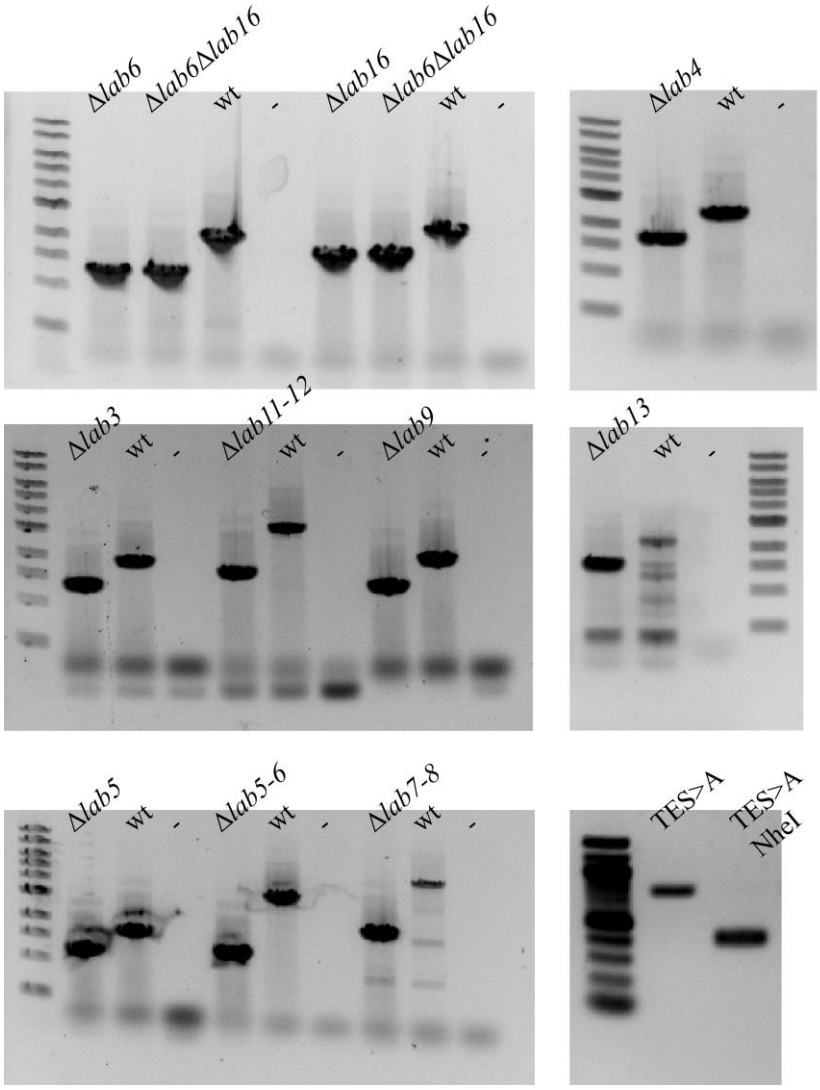


Figure 63. Deletion mutants constructed as described in Material and Methods 7.1.1 and verified by PCR. Quick-Load® DNA ladder (New England BioLabs) of 1 kb was used for the fragment size estimation.

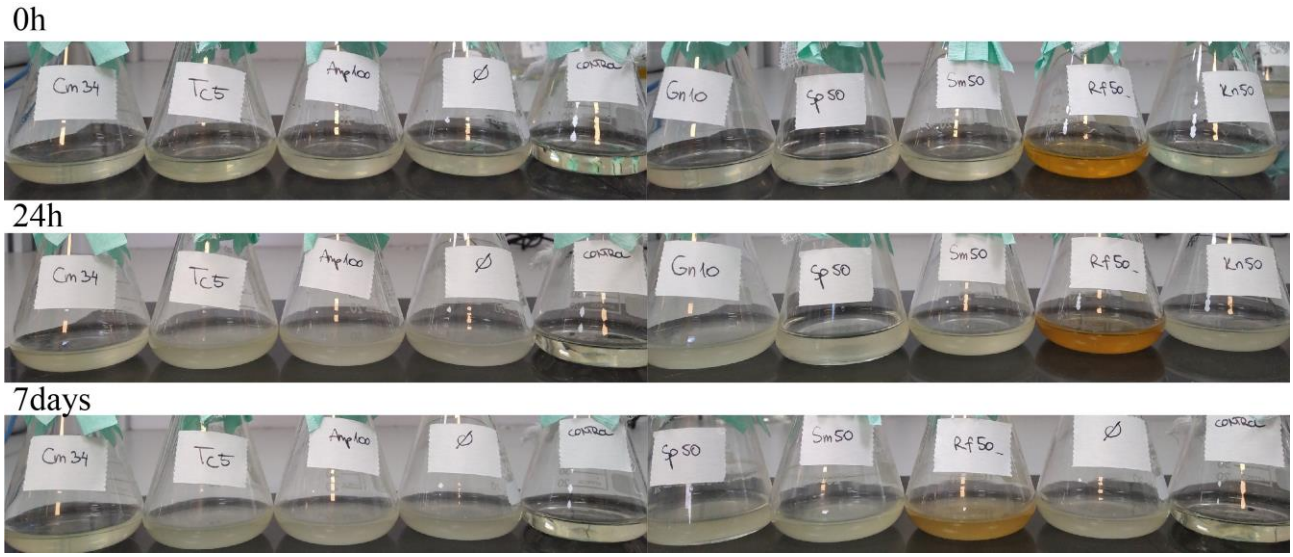


Figure 64. Antibiotic sensitivity of *Labrenzia* sp. PHM005 in MB monitored during 7 days. The antibiotics used were chloramphenicol, tetracycline, ampicillin, gentamycin, spectinomycin, streptomycin, rifampicin and kanamycin. The concentrations of antibiotics are indicated in µg/mL.

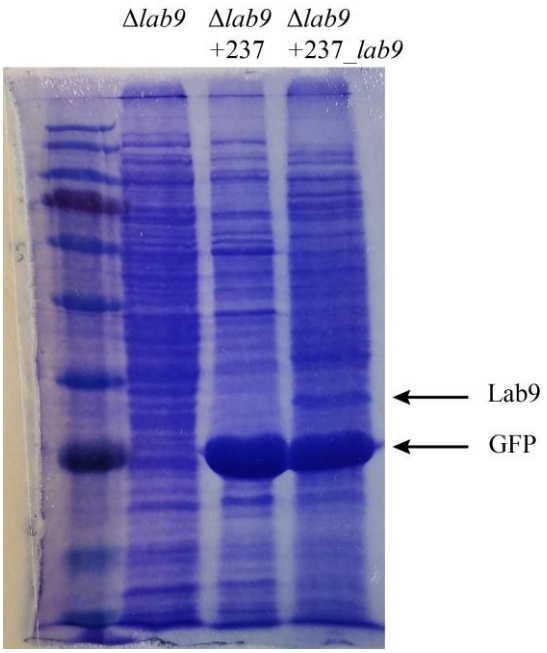


Figure 65. SDS-PAGE of soluble proteins extracted from the cytosol of Δ*lab9*, Δ*lab9*+pSEVA237 and Δ*lab9*+pSEVA237_*lab9*.

3. Annex III: Ion fragmentation of the compounds by ESIMS

Methylations

Specific function of methyltransferase Lab6

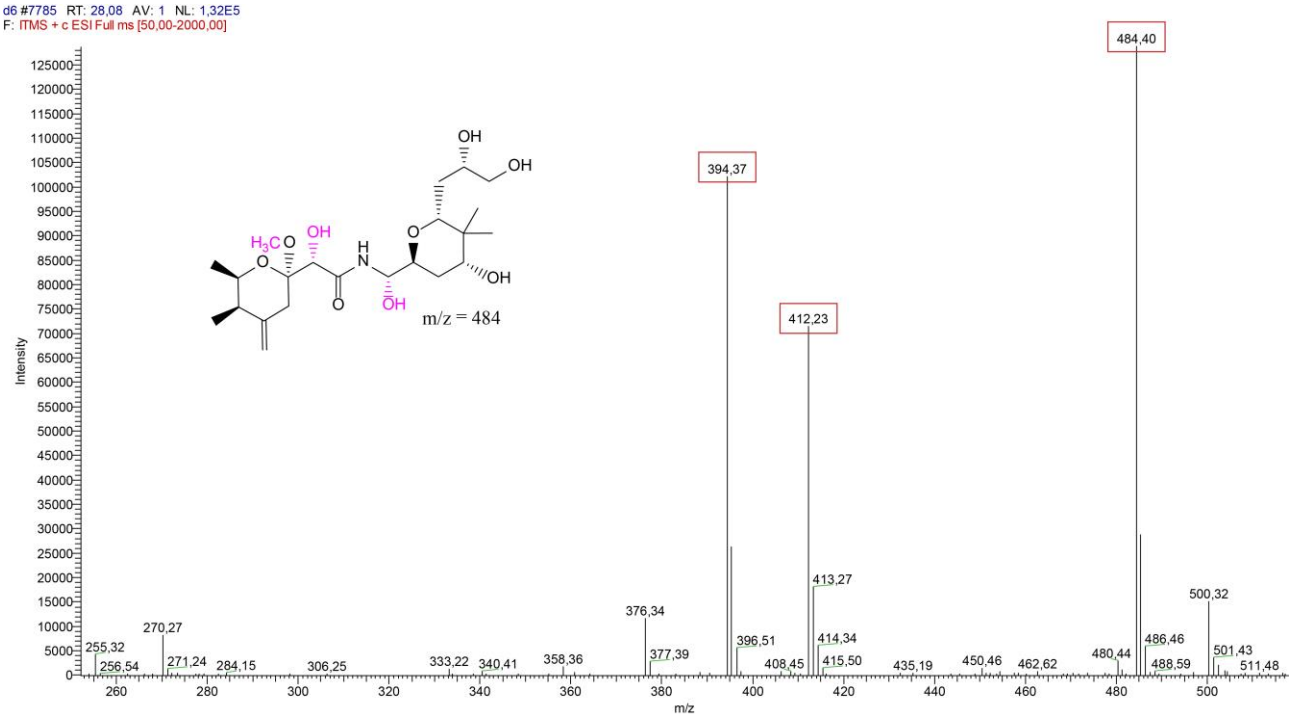


Figure 66. ESIMS fragmentation pattern of compound 4 in Δ*lab6* mutant and the proposed molecular structure. Ion masses include sodium ion.

Annexes

Specific function of methyltransferase Lab16

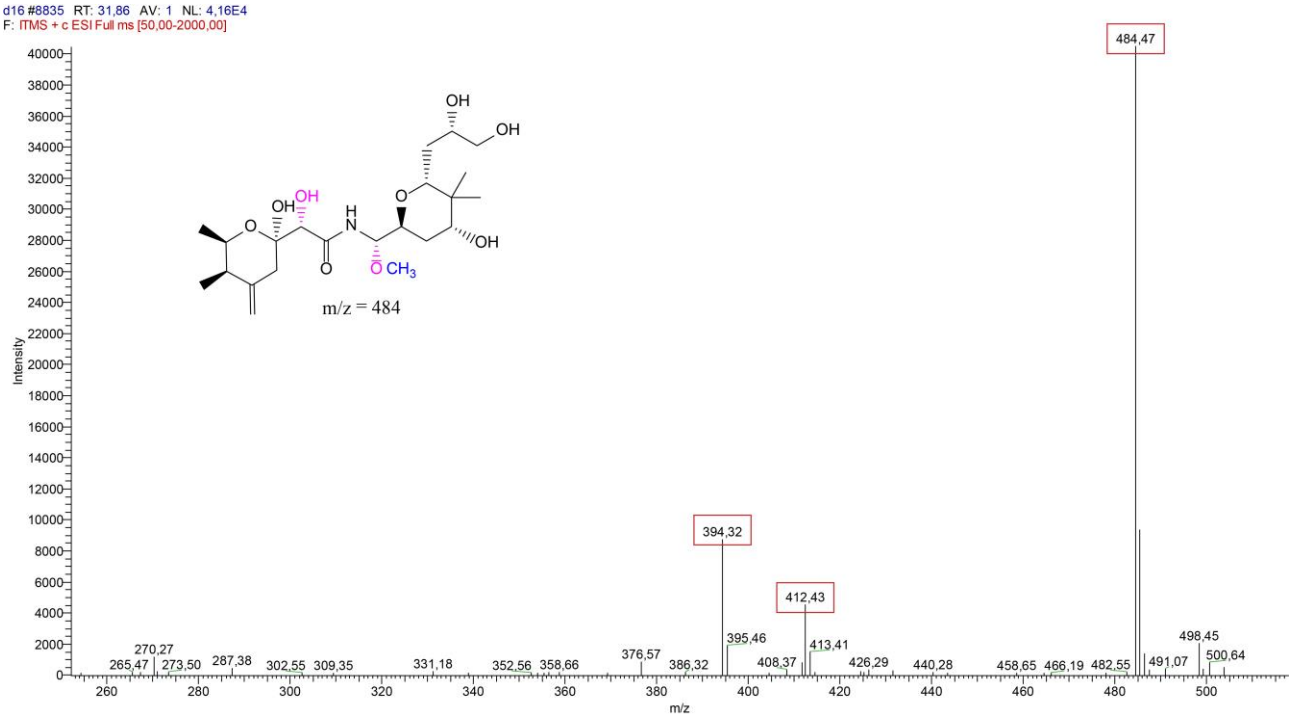


Figure 67. ESIMS fragmentation pattern of compounds 5a and 5b in Δ*lab16* mutant and the proposed molecular structure. Ion masses include sodium ion.

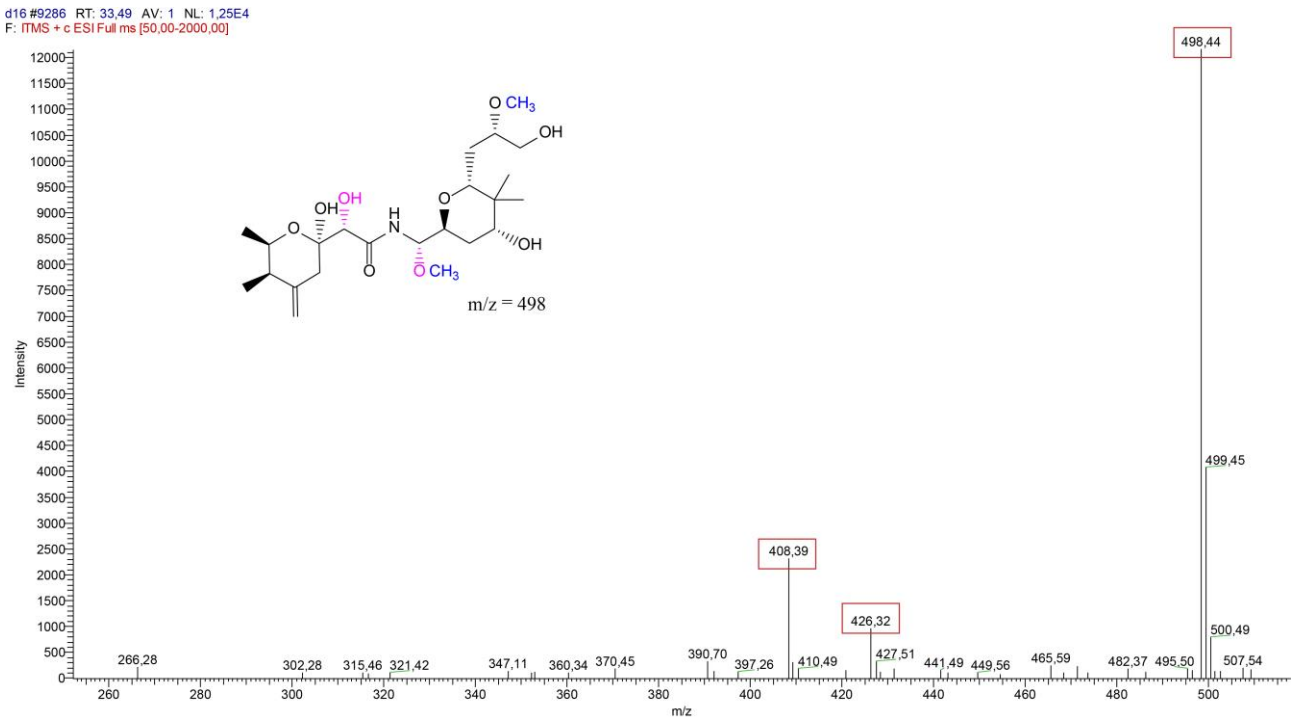


Figure 68. ESIMS fragmentation pattern of compounds 6a and 6b in Δ*lab16* mutant and the proposed molecular structure. Ion masses include sodium ion.

Annexes

Generation of de-methylated labrenzin producer

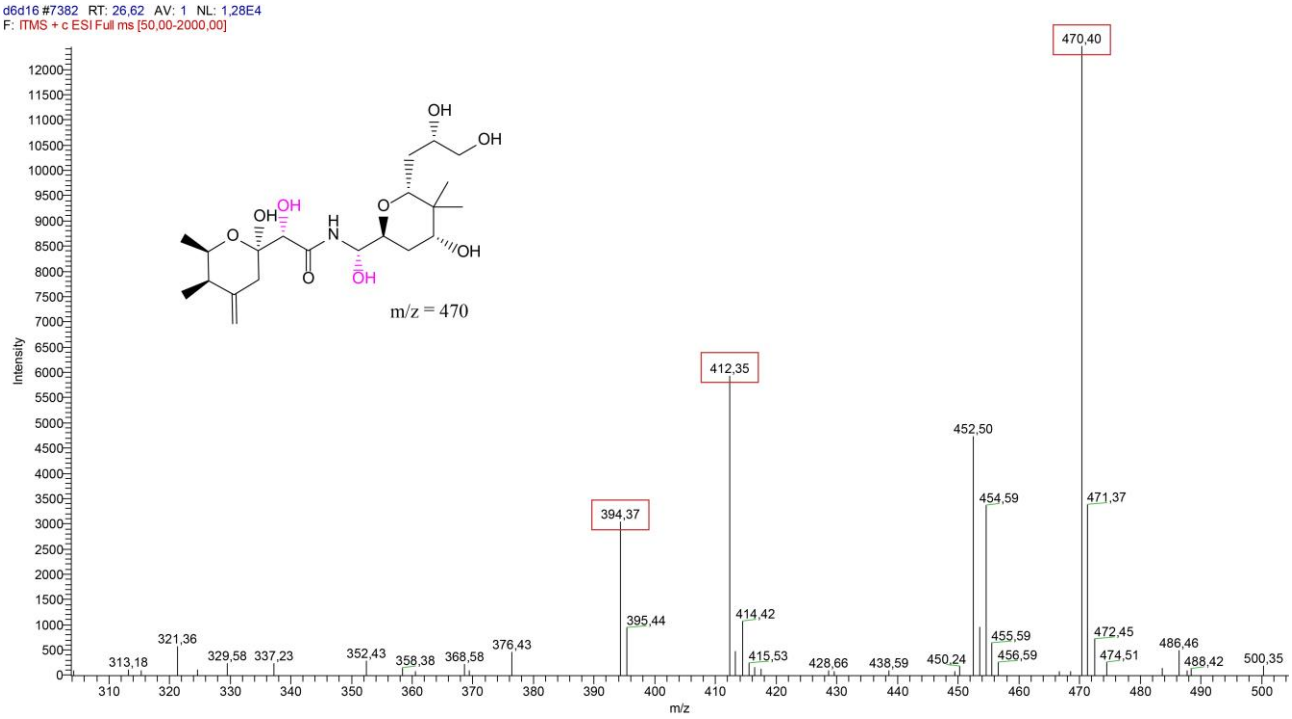


Figure 69. ESIMS fragmentation pattern of compound 7 in Δ*lab6*Δ*lab16* mutant and the proposed molecular structure. Ion masses include sodium ion.

Hydroxylations

Tanadem knockout of Lab7 and Lab8

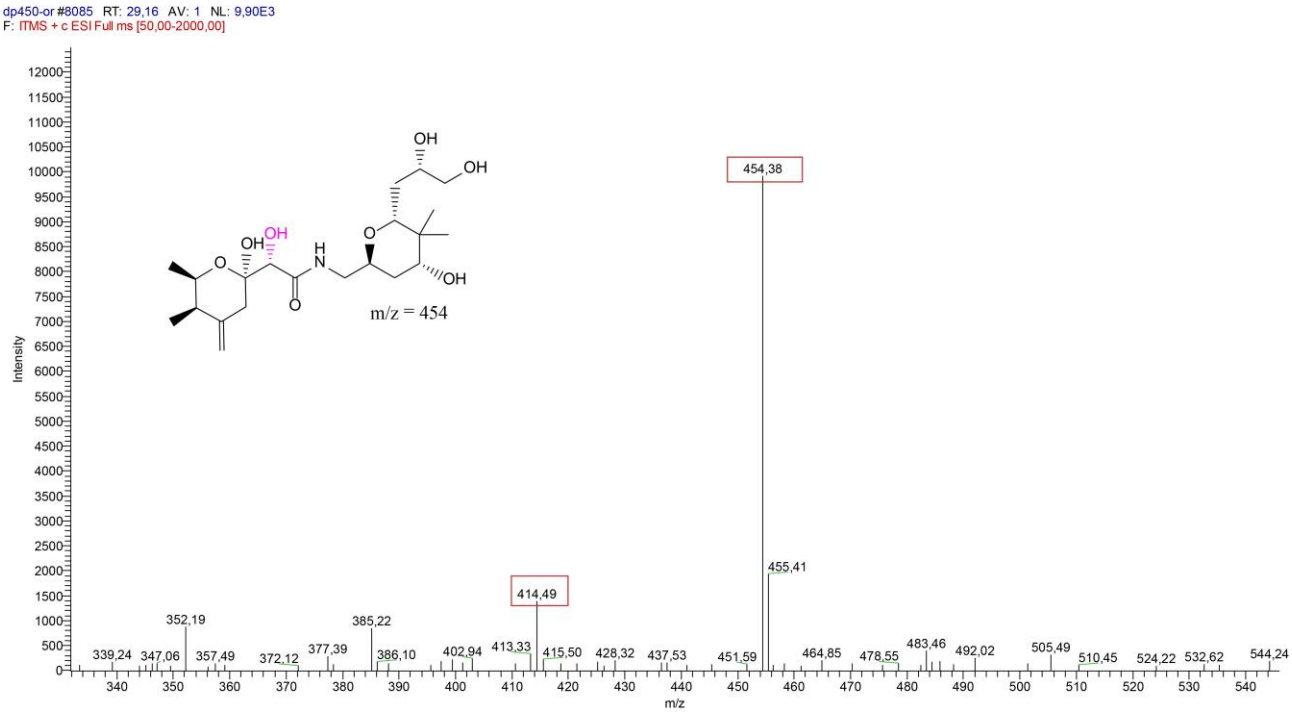


Figure 70. ESIMS fragmentation pattern of compounds 8a and 8b in Δ*lab7*Δ*lab8* mutant and the proposed molecular structure. Ion masses include sodium ion.

Annexes

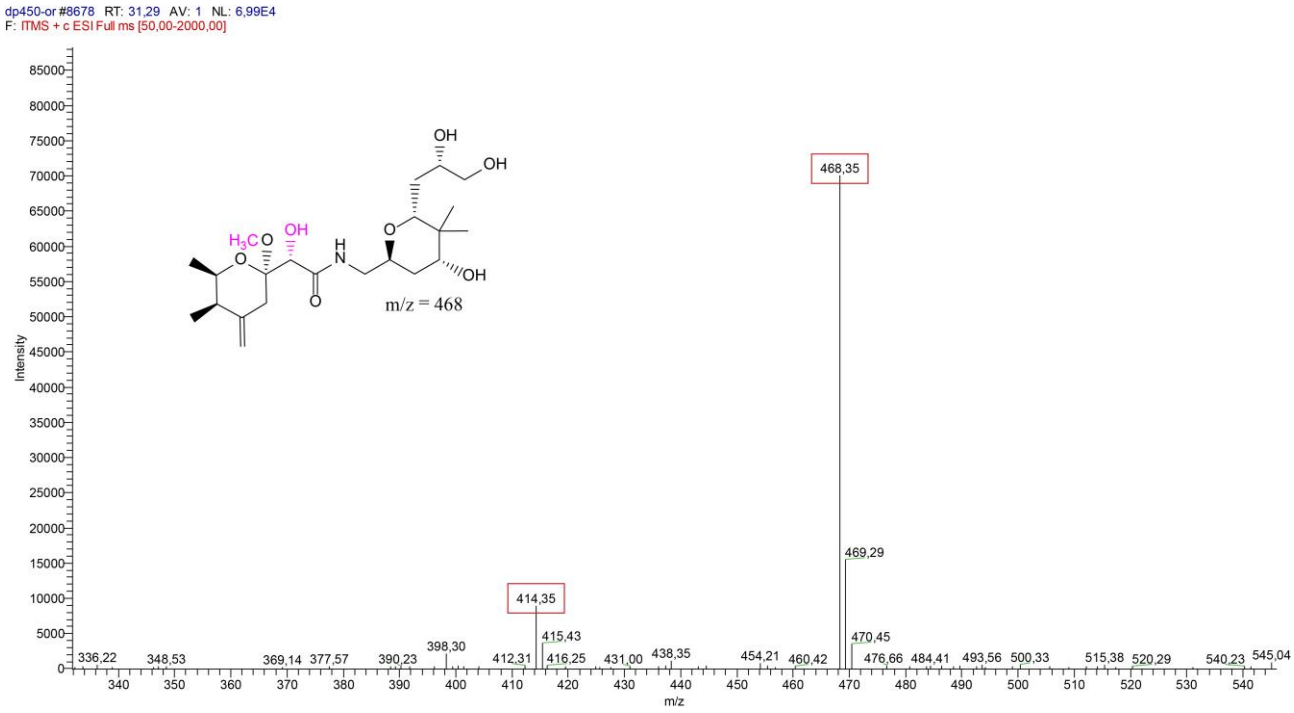


Figure 71. ESIMS fragmentation pattern of compound 3 in Δ*lab7*Δ*lab8* mutant and the proposed molecular structure. Ion masses include sodium ion.

Knockout of Lab5

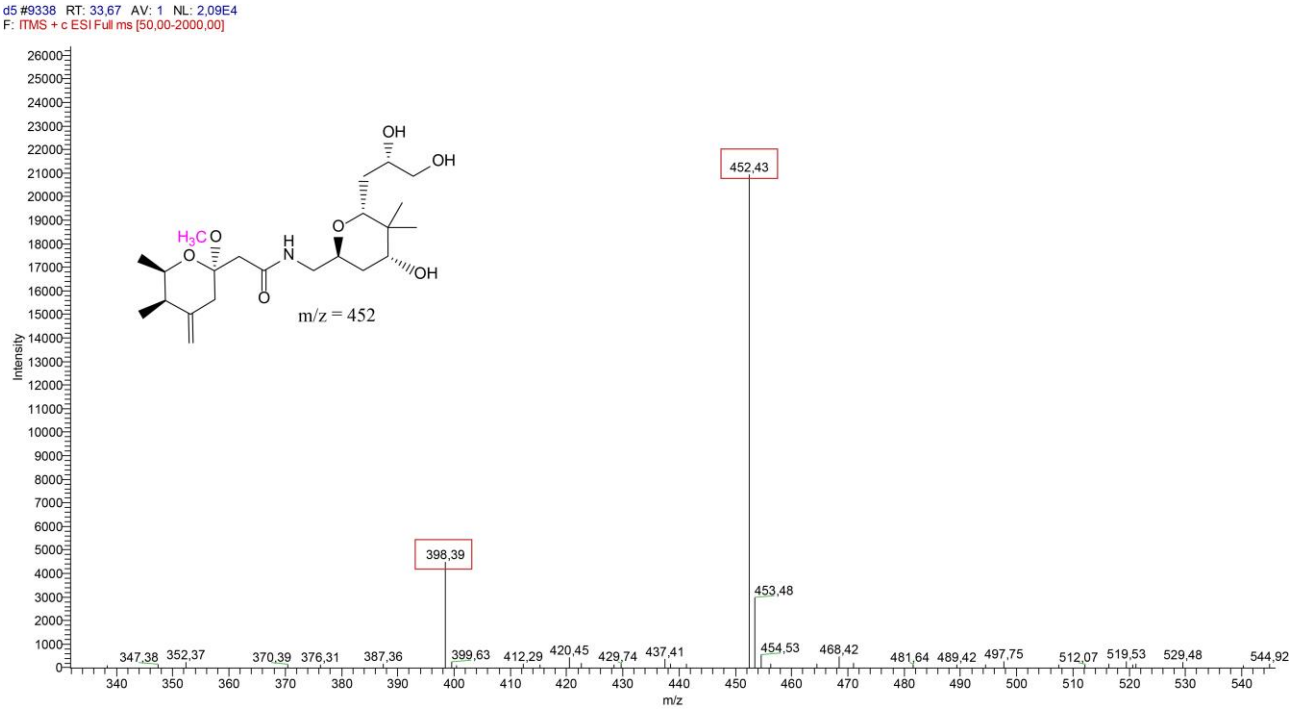


Figure 72. ESIMS fragmentation pattern of compound 9 in Δ*lab5* mutant and the proposed molecular structure. Ion masses include sodium ion.

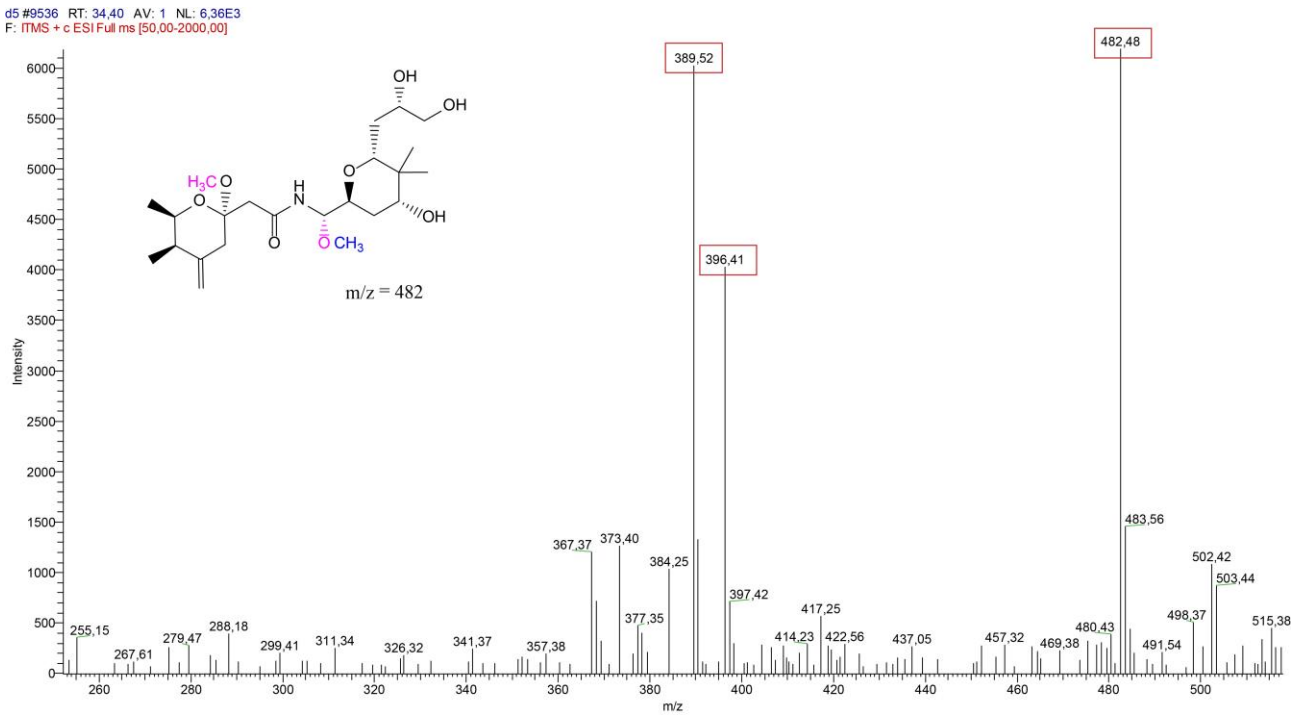


Figure 73. ESIMS fragmentation pattern of compounds 10a and 10b in Δ*lab5* mutant and the proposed molecular structure. Ion masses include sodium ion.

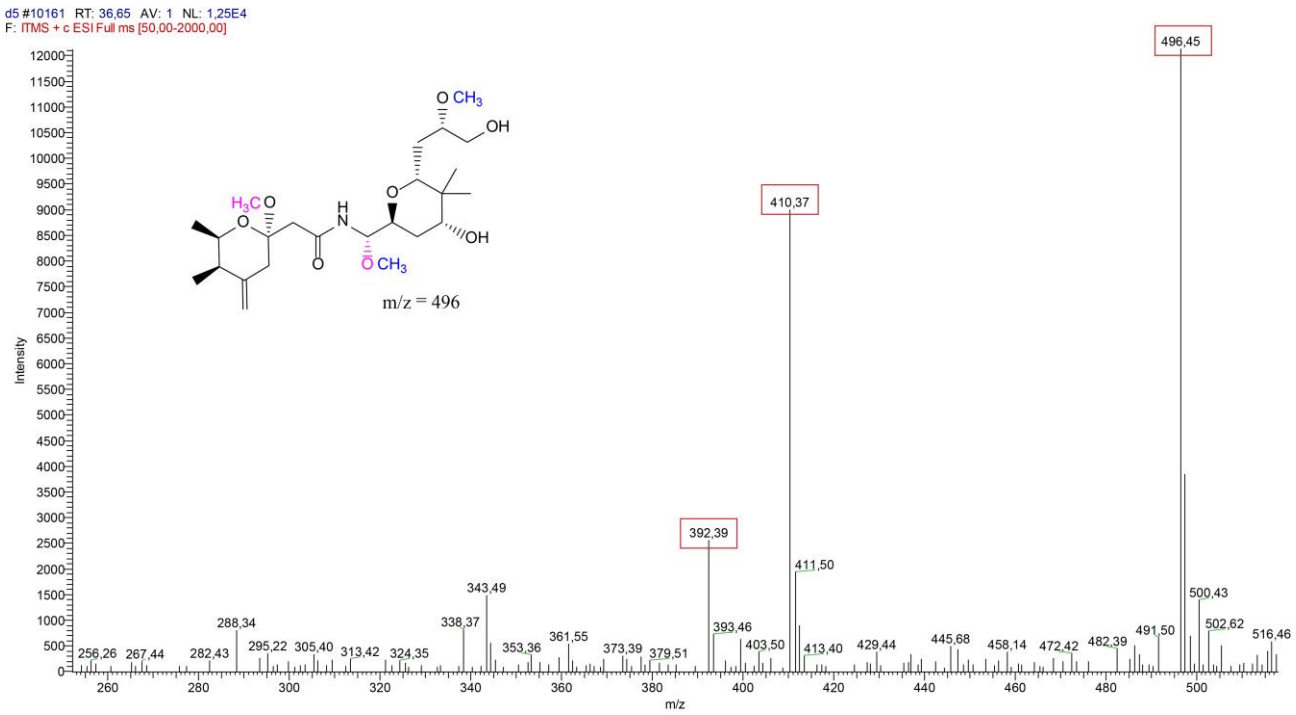


Figure 74. ESIMS fragmentation pattern of compound 11 in Δ*lab5* mutant and the proposed molecular structure. Ion masses include sodium ion.

Tanadem knockout of Lab5-6

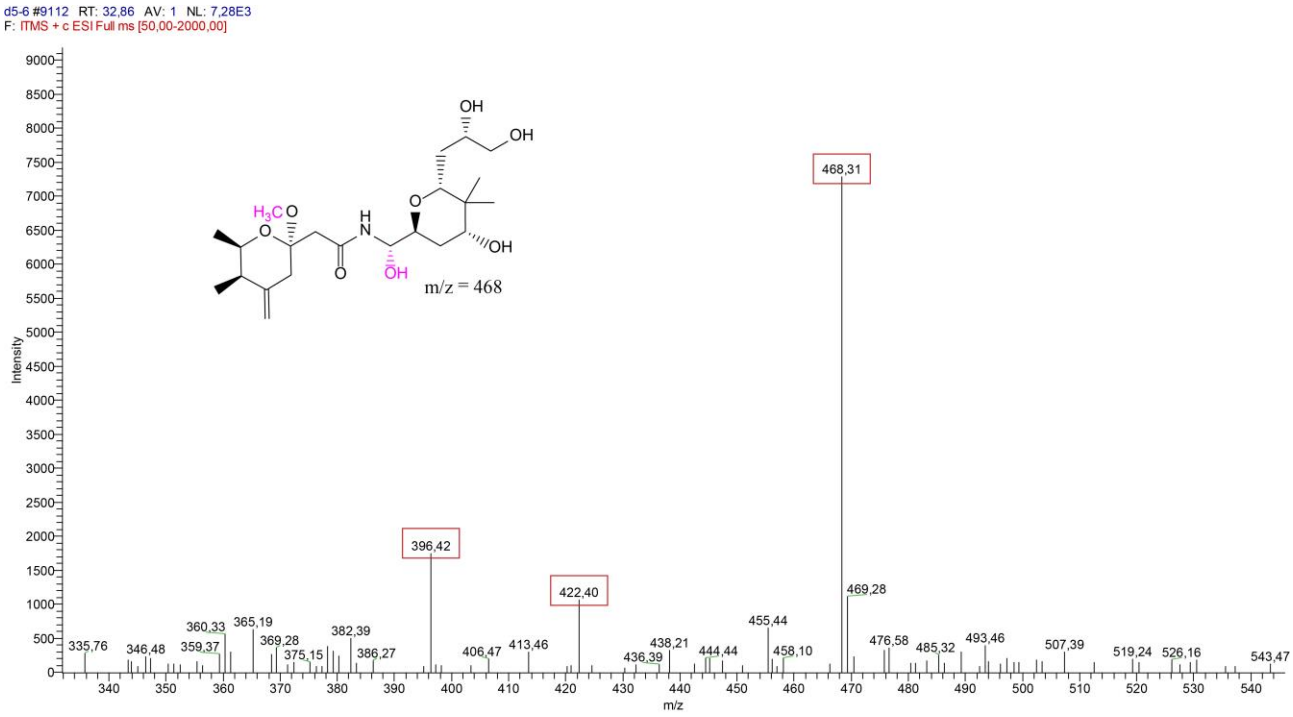


Figure 75. ESIMS fragmentation pattern of compound 12 in Δ*lab5*-Δ*lab6* mutant and the proposed molecular structure. Ion masses include sodium ion.

Tailoring by heterologous expression of methyltransferase PedO

Methyltransferase Lab6 mutant complemented with methyltransferase PedO

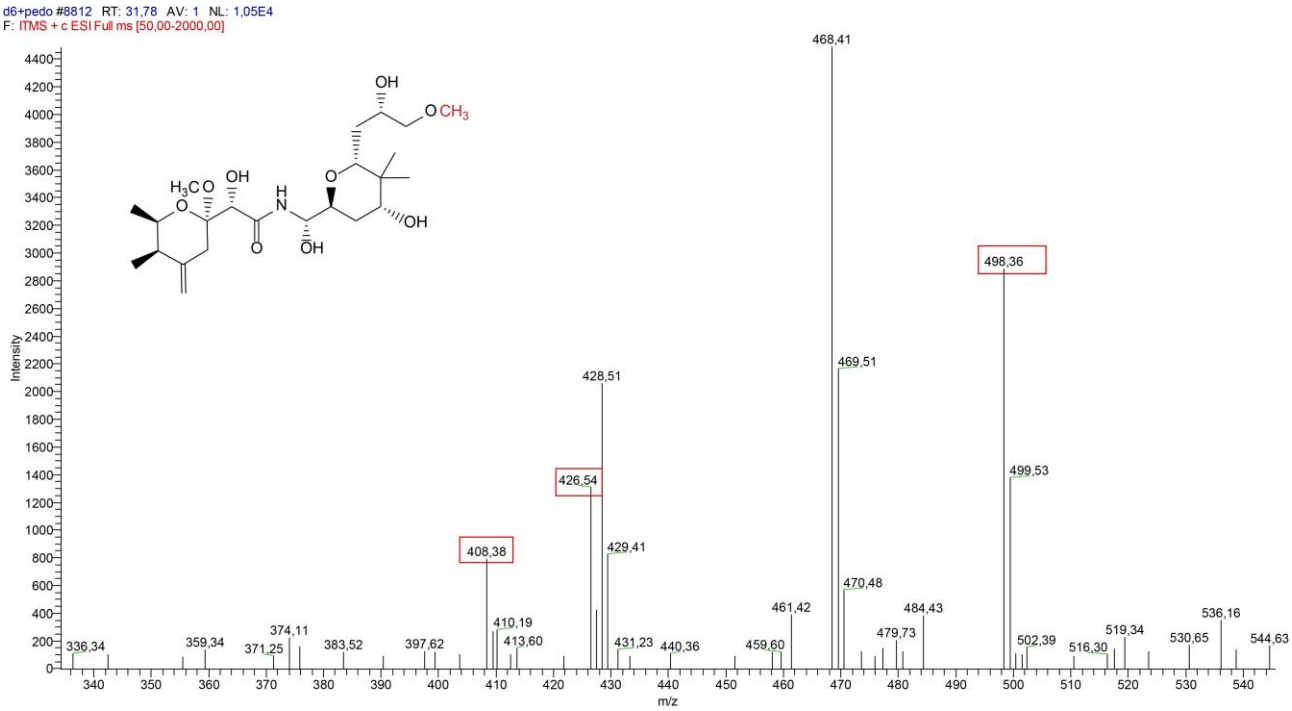


Figure 76. ESIMS fragmentation patern of heterologously produced compound 14 in Δ*lab6* mutant and the proposed molecular structure. Ion masses include sodium ion.

Annexes

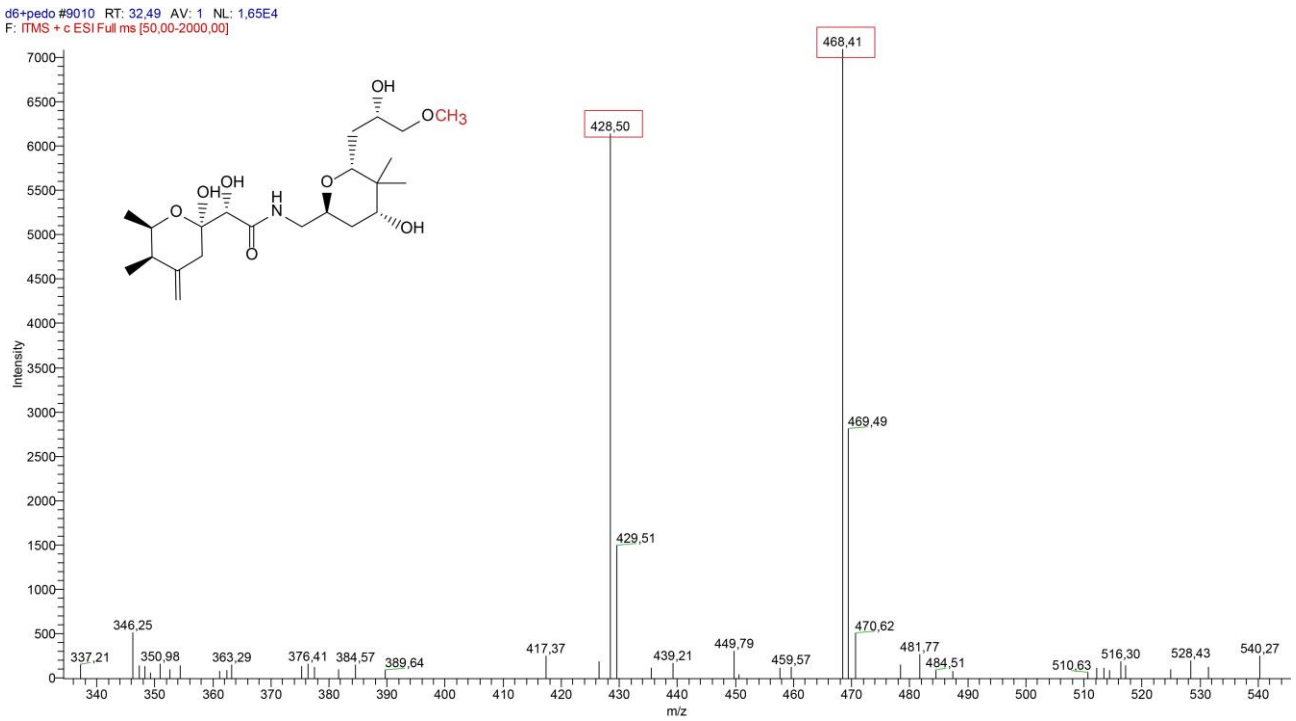


Figure 77. ESIMS fragmentation patern of heterologously produced compound 15 in Δ*lab6* mutant and the proposed molecular structure. Ion masses include sodium ion.

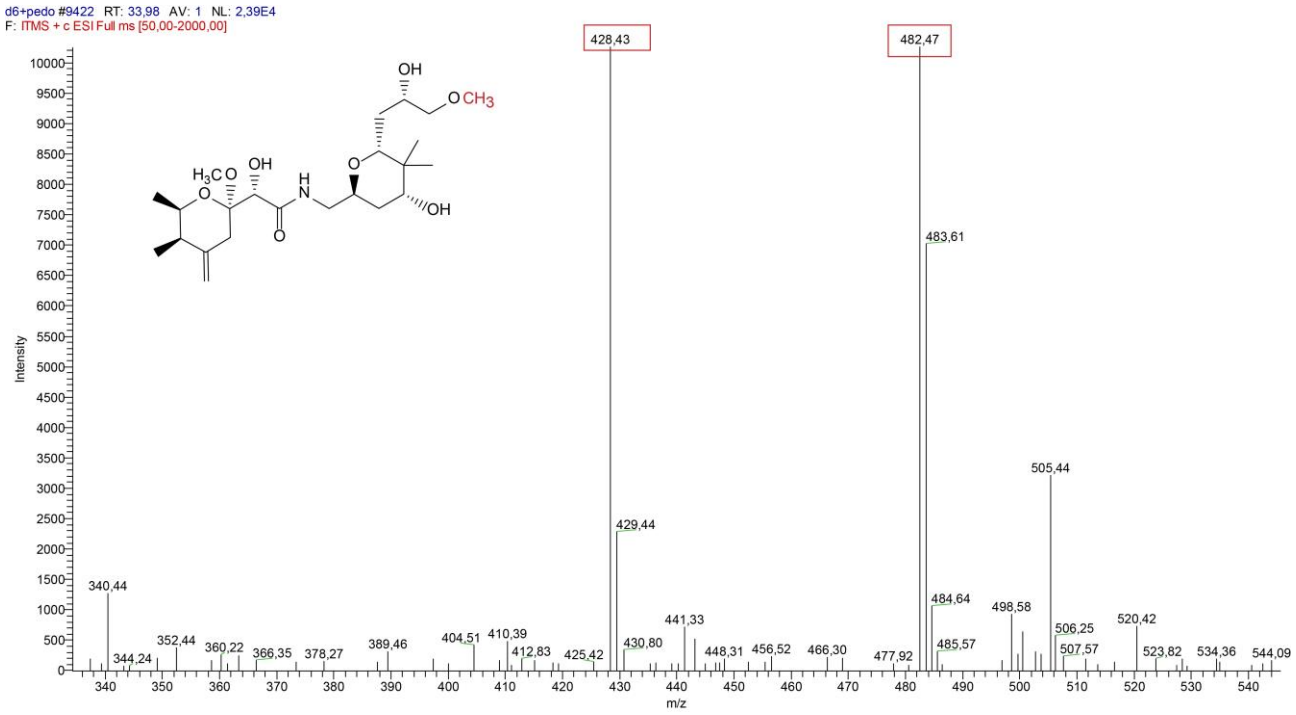


Figure 78. ESIMS fragmentation patern of heterologously produced compound 16 in Δ*lab6* mutant and the proposed molecular structure. Ion masses include sodium ion.

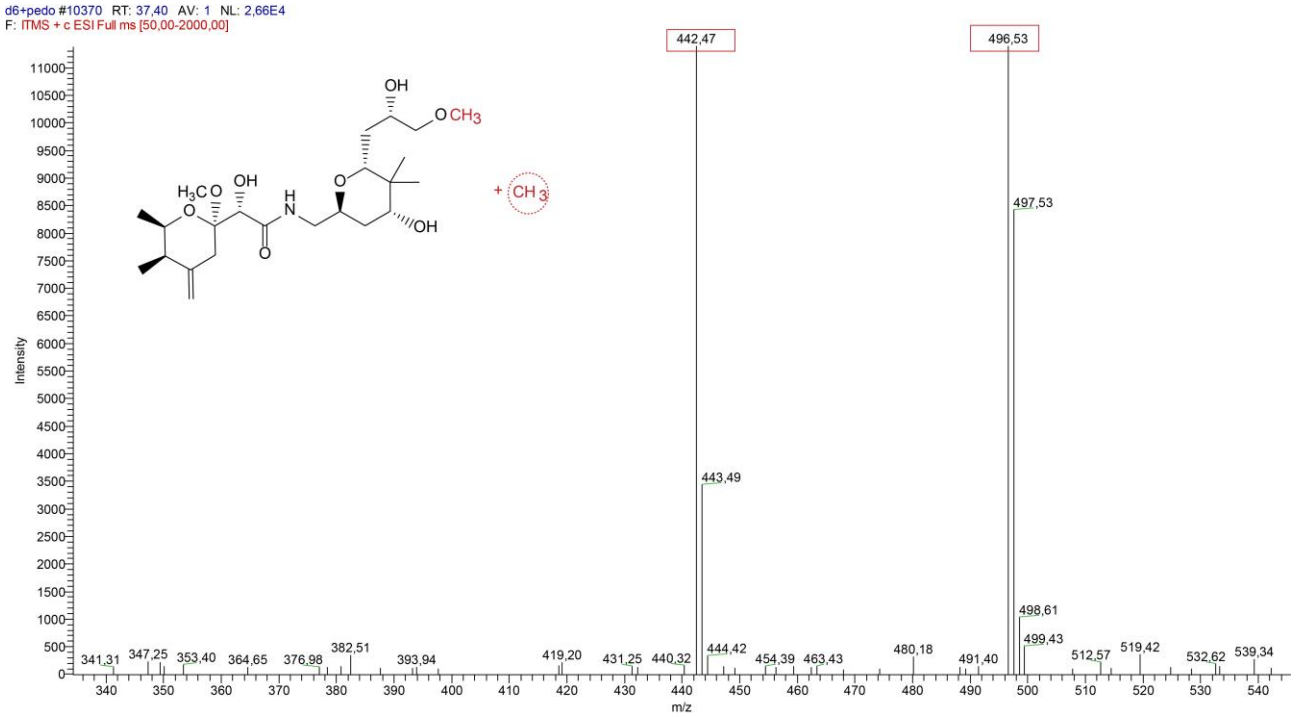


Figure 79. ESIMS fragmentation patern of heterologously produced compound 17 in Δ*lab6* mutant and the proposed molecular structure. Ion masses include sodium ion.

Methyltransferase Lab6 mutant complemented with methyltransferase PedO

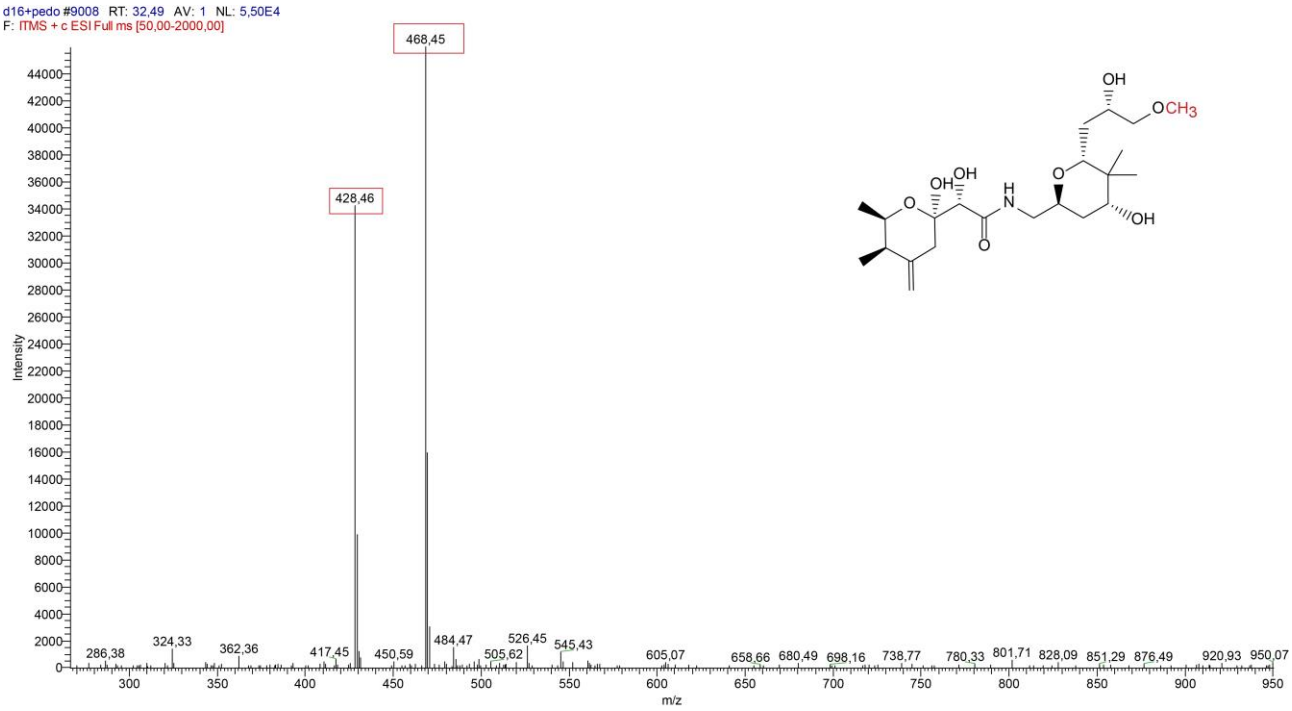


Figure 80. ESIMS fragmentation patern of heterologously produced compound 18 in Δ*lab16* mutant and the proposed molecular structure. Ion masses include sodium ion.

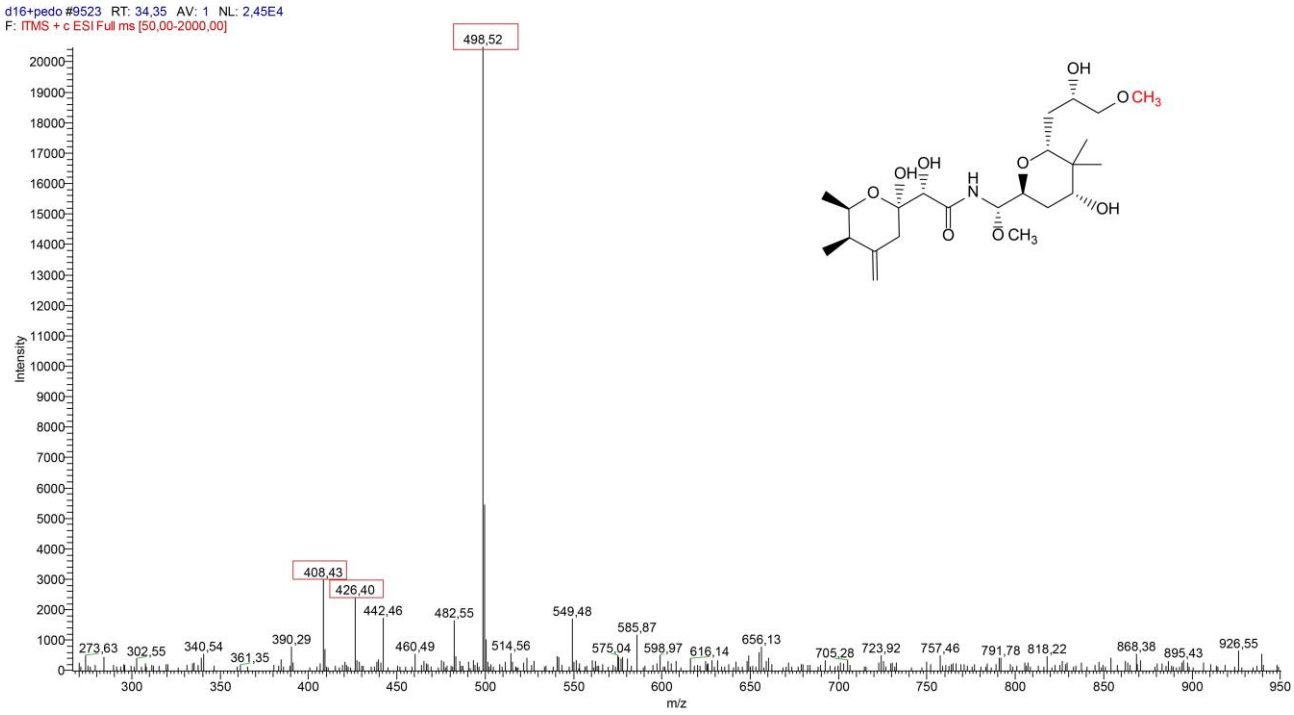


Figure 81. ESIMS fragmentation patern of heterologously produced compound 19 in Δ*lab16* mutant and the proposed molecular structure. Ion masses include sodium ion.

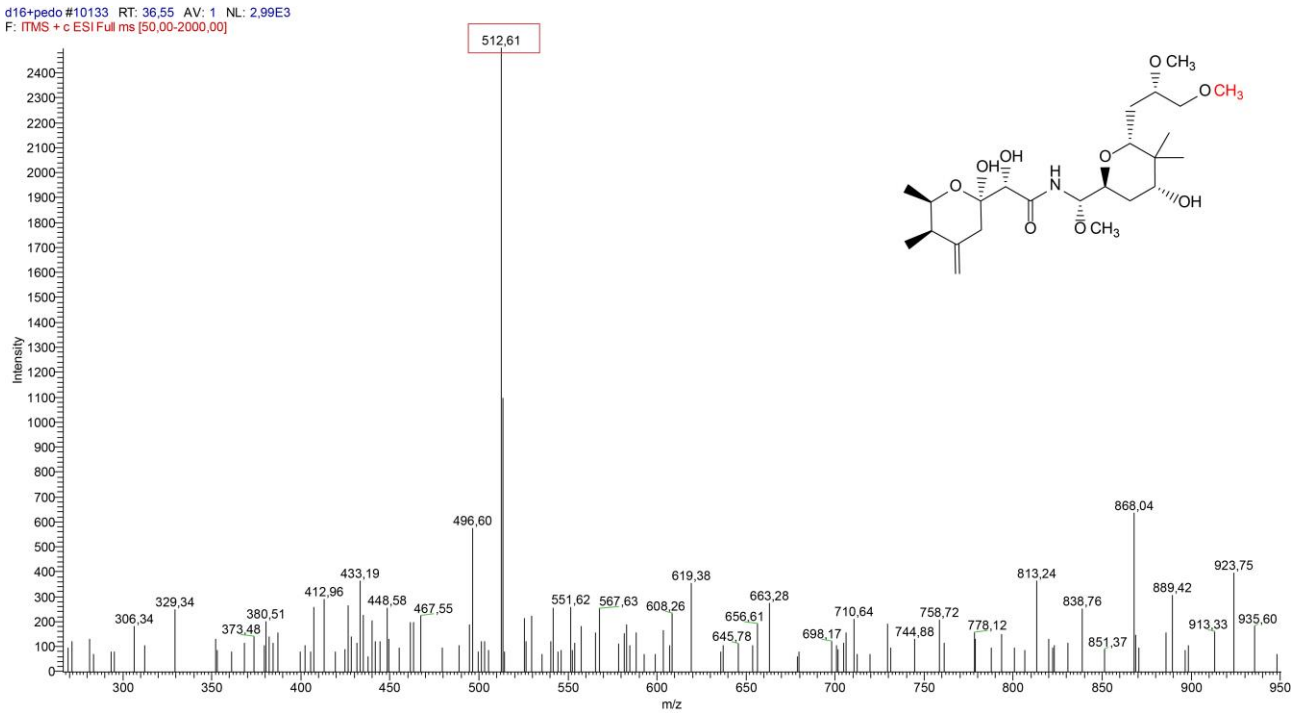


Figure 82. ESIMS fragmentation patern of heterologously produced compound 20 in Δ*lab16* mutant and the proposed molecular structure. Ion masses include sodium ion.

Annexes

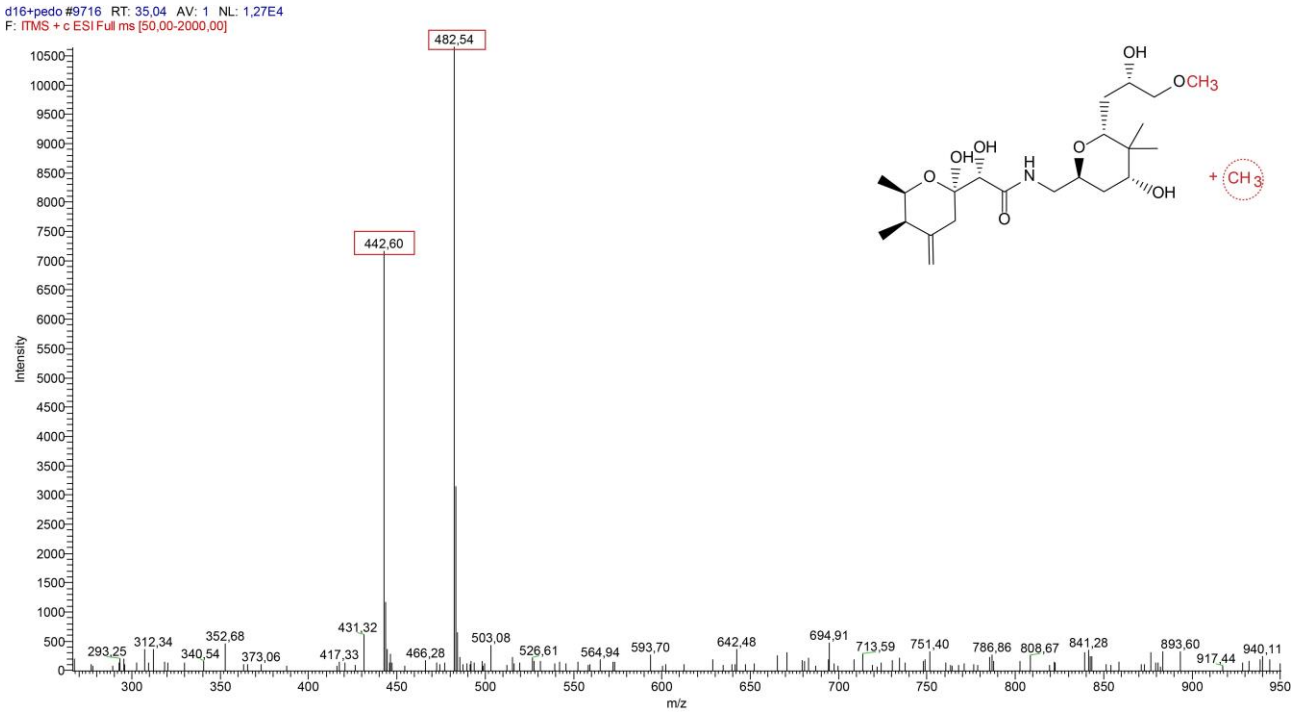


Figure 83. ESIMS fragmentation patern of heterologously produced compound 21 in Δ*lab16* mutant and the proposed molecular structure. Ion masses include sodium ion.

Double methyltransferase mutant complemented with methyltransferase PedO

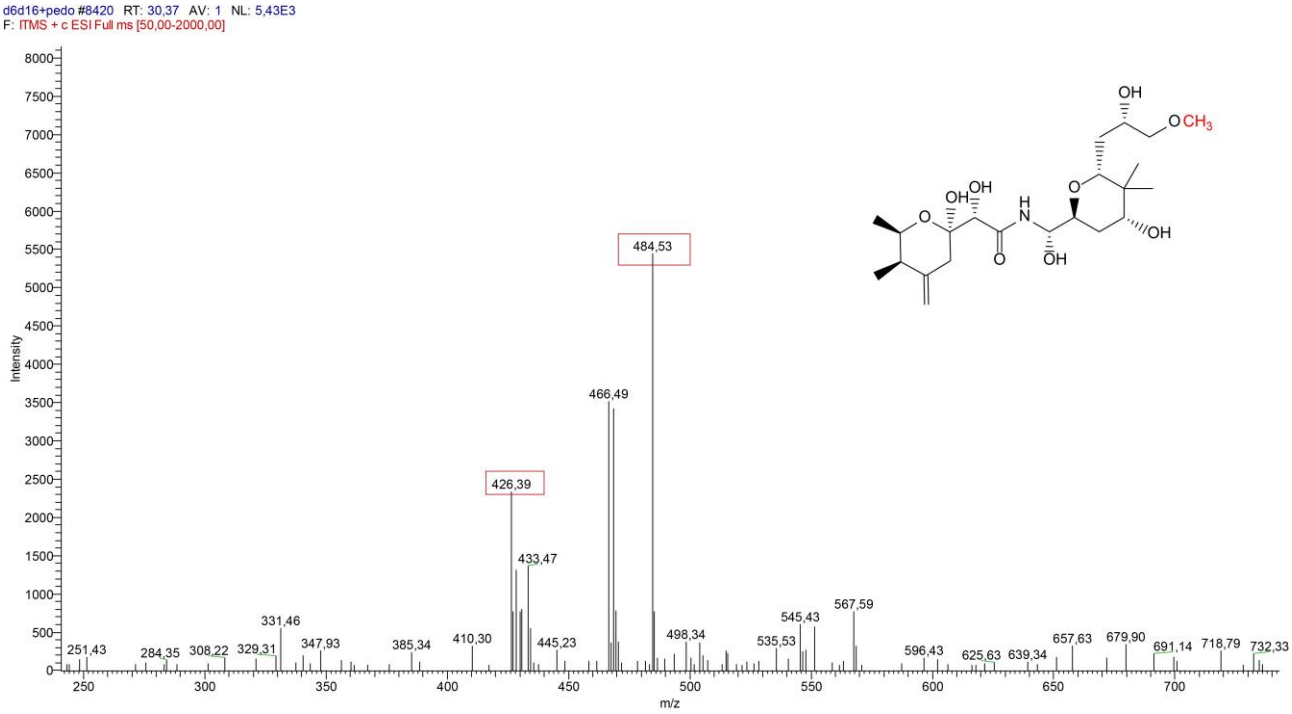


Figure 84. ESIMS fragmentation patern of heterologously produced compound 22 in Δ*lab6*Δ*lab16* mutant and the proposed molecular structure. Ion masses include sodium ion.

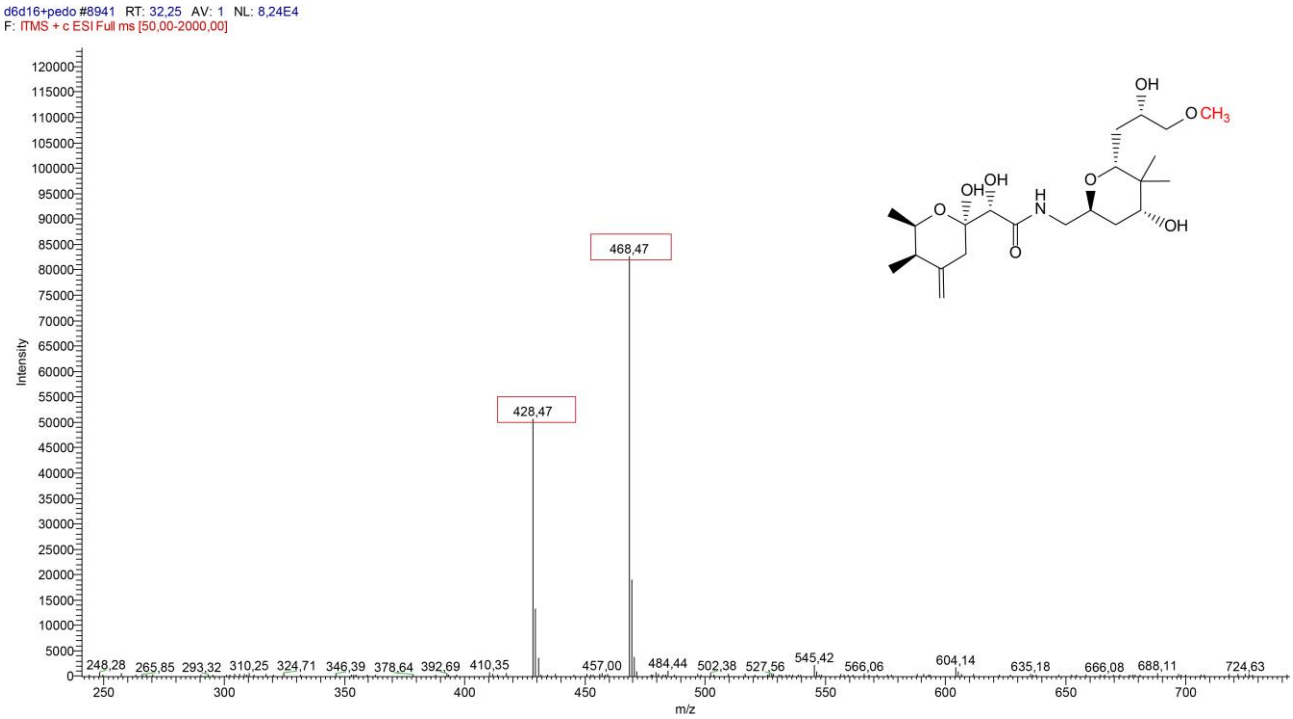


Figure 85. ESIMS fragmentation patern of heterologously produced compound 18 in Δ*lab6*Δ*lab16* mutant and the proposed molecular structure. Ion masses include sodium ion.

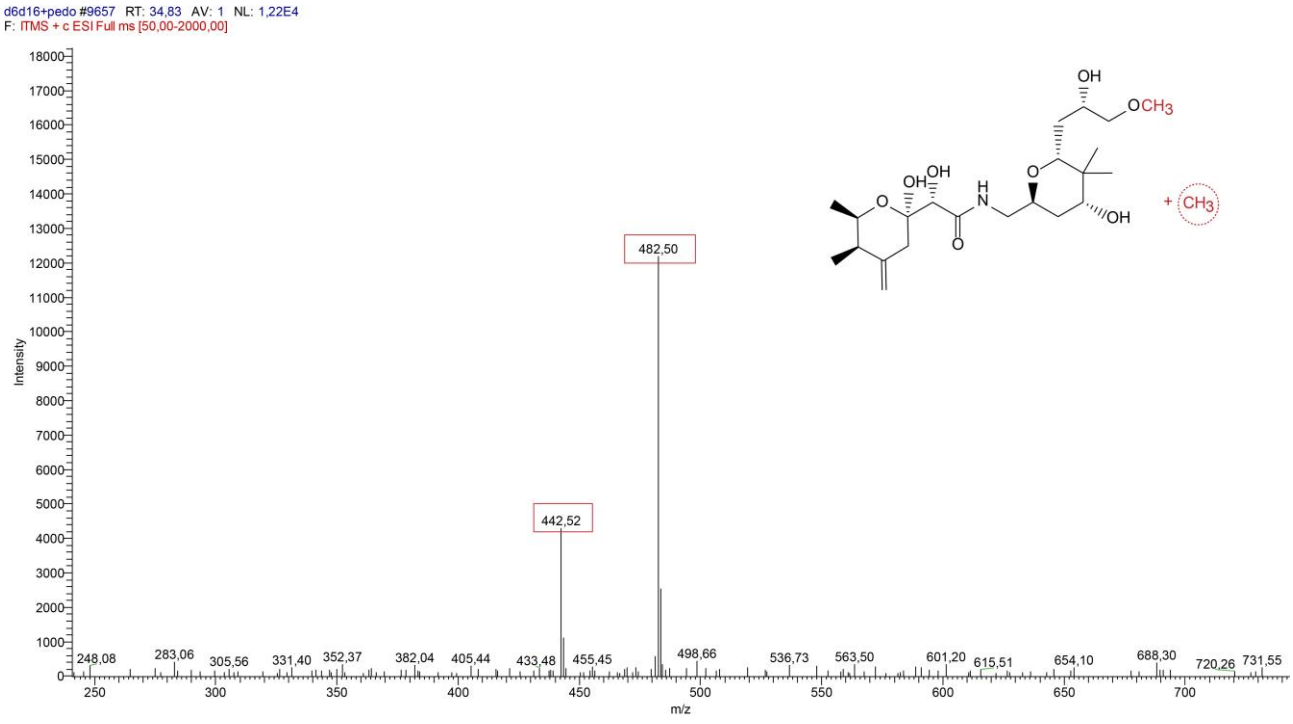


Figure 86. ESIMS fragmentation patern of heterologously produced compound 21 in Δ*lab6*Δ*lab16* mutant and the proposed molecular structure. Ion masses include sodium ion.

4. Annex IV: Publications

Kačar, D., Schleissner, C., Cañedo, L.M., Rodríguez, P., de la Calle, F., Galán, B. and García, J.L., (2019) Genome of *Labrenzia* sp. PHM005 reveals a complete and active *trans*-AT PKS gene cluster for the biosynthesis of labrenzin. *Front. Microbiol.* **10**: 2561-2561.

Kačar, D., Cañedo, L.M., Rodríguez, P., González, E., Galán, B., Schleissner, C., Leopold-Messer, S., Piel, J., Cuevas, C., de la Calle, F., and García, J.L. (2020) Identification of *trans*-AT polyketide clusters in two marine bacteria reveals cryptic similarities between distinct symbiosis factors. *bioRxiv* 2020.09.18.303172.