Universidade de Lisboa
Faculdade de Farmácia

Multi-approach analysis of the metagenome of a marine sponge containing latrunculin A

Sandra Cristina Godinho Pires da Silva

Mestrado Integrado em Ciências Farmacêuticas

2017

Universidade de Lisboa

Faculdade de Farmácia

Multi-approach analysis of the metagenome of a marine sponge containing latrunculin A

Sandra Cristina Godinho Pires da Silva

Trabalho de campo de Mestrado Integrado em Ciências Farmacêuticas apresentada à Universidade de Lisboa através da Faculdade de Farmácia

Orientador FFUL: Professor Doutor Carlos São-José

2017

The work that led to the production of this document was carried at the Pharmazeutische Biologie Instituit of the Universität Bonn, under the “Indobio” Biodiversity and Health project framework. Indobio consortium is funded by the Federal Ministry of Education and Research (BMBF) with the grants 16GW0117K and 16GW0118.

Supervisor: Prof. Dr. Till F. Schäberle
Institut für Pharmazeutische Biologie der Universität Bonn
Justus-Liebig-Universität Gießen

Co-supervisor: Dr. Katja M. Fisch
Institut für Pharmazeutische Biologie der Universität Bonn
Justus-Liebig-Universität Gießen

Abstract

Marine natural products, such as those encountered in sponges, are a prolific source of drug leads. In cancer treatment, several sponges' secondary metabolites have shown potent cytotoxic activities. One of those is latrunculin A, a macrolide whose biosynthetic production relying on a hybrid pathway involving polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs) is still unknown. Considering the amounting evidence that the sponges' symbionts are the true producers of these compounds, it is reasonable to search for the biosynthetic gene cluster of latrunculin A in the sponges' metagenome. To achieve this goal, a multi-approach analysis was conducted in marine sponges' samples, some of them containing latrunculin A.

A sponge's metagenomic DNA library of 3500 clones was produced. Obtaining high molecular weight metagenomic DNA was a critical point which seems to have conditioned successful vector packaging in λ phage particles. Instead, transformation of the vector into the bacterial host was conducted through electroporation and heat-shock. A preliminary PCR screen of the transformants revealed the presence of PKS genes.

In parallel, PCR screening of metagenomic DNA of diverse sponge samples was conducted. Due to the hybrid PKS/NRPS origin of latrunculin A, focus was directed towards the presence of PKS and NRPS genes. Several PCR-amplified sequences exhibited homology to the ketosynthase (KS) domain of PKSs. Specific KS primers were designed to latter screening of the produced metagenomic library.

Microbiome profiling of three different sponge samples was achieved through 16S rRNA gene sequencing. Even though it was not possible to identify probable latrunculin A' producers, an analysis of the biosynthetic potential of some of the most abundant symbionts was performed. This demonstrated that further investigation of these sponges for the discovery of novel compounds is promising.

Key words: marine sponges, latrunculin A, metagenomic library, biosynthetic gene cluster (BGC), microbiome profiling.

Resumo

Os produtos naturais marinhos, como os encontrados nas esponjas marinhas, são uma prolífica fonte de fármacos. Em particular, no tratamento de cancro, vários metabolitos secundários de esponjas têm mostrado potente atividade citotóxica. Um deles é a latrunculina A, um composto de origem híbrida policetídica e peptídica não ribossomal (PKS/NRPS), cuja via biossintética é ainda desconhecida. Considerando a evidência crescente de que os verdadeiros produtores destes compostos são os simbiontes das esponjas, é lógico procurar o cluster de genes biossintéticos da latrunculina A no metagenoma das esponjas. Para atingir esse objetivo, foi planeada uma abordagem múltipla de análise de amostras de esponjas marinhas, algumas delas contendo latrunculina A.

A partir de uma esponja marinha, foi produzida uma biblioteca de ADN metagenómico contendo 3500 clones. A obtenção de ADN metagenómico de alto tamanho molecular foi um ponto crítico que parece ter condicionado o sucesso do empacotamento do vetor no fago λ. Em vez disso, a transformação do vetor no hospedeiro bacteriano foi efetuada por eletroporação e choque térmico. A presença de genes para sintases de polipéptidos na biblioteca foi confirmada por rastreio inicial dos transformantes por PCR.

Em paralelo, realizou-se o rastreio por PCR do ADN metagenómico de diversas amostras de esponjas. Devido à origem híbrida PKS/NRPS da latrunculina A, o foco foi direcionado para a presença de genes PKS e NRPS. Diversas sequências amplificadas por PCR demonstraram homologia para com o domínio ketosintase (KS) de sintases de policétidos Primers específicos para KS foram desenhados para posterior uso no rastreio da biblioteca metagenómica produzida.

O perfil microbiano de três diferentes amostras de esponjas foi obtido através de sequenciação do gene ARN 16S ribossomal. Apesar de não ter sido possível identificar prováveis produtores de latrunculina A, foi realizada uma análise do potencial biossintético dos simbiontes mais abundantes. Esta demonstrou que parecem promissoras futuras investigações destas esponjas para descoberta de novos compostos.

Palavras-chave: esponjas marinhas, latrunculina A, biblioteca metagenómica, cluster de genes biossintéticos, perfil do microbioma.

Acknowledgments

This work is the end of an incredible journey that will definitely mark my path. From my time in Bonn to the period of writing and revising this dissertation, I have undercome a lot of challenges. More than a more competent and knowledgeable scientist, I became truly passionate about marine sponges and metagenomics. I am sure this is just the beginning of a long road ahead.

First of all, I want to thank Prof. Dr. Till Schäberle for giving me this opportunity and for receiving me so well in his research group. His knowledge and insights combined with his kind guidance were crucial to the success of this work.

I wouldn’t forget to deeply thank Prof. Dr. Carlos São-José, who through the years has continuously guided me in my passion for science. I am grateful for his detailed revision of this thesis and for its availability to support me in this project.

I want to thank Dr. Katja Fisch for her support and technical advice. Without her guidance, I wouldn’t have achieved the results I needed in such a short period of time. For Zerlina goes my deep gratitude for so kindly introducing me to the laboratory and to the work there developed. My appreciation also extends to my laboratory colleagues Stephan, Luis, Isabella, Jamshid, Nils, Cornelia, Riyanti, Max, Henrik, Fayrouz and Sarah, among many others, who made my journey a lot easier. Thank you for every word of knowledge and sympathy.

To the ones that stood by my side in these challenging five years, my deep thanks. Leo, Jorge, Márcio, Maria, Milene e Joana, it wouldn’t have been the same without your friendship. To my friends at home, Sara, Rafael, Susana, Bruna e João I thank for their genuine interest in my work and for every kind word of encouragement.

To my family goes my deep love and gratitude for always believing in me. Without their support, nothing of this would have been possible.

And finally, I want to thank my life companion Filipe, for whom I dedicate this thesis. He has supported me more than it can be described.

Abbreviations

A domain	Adenylation domain
ACP	Acyl carrier protein
antiSMASH	Antibiotics & Secondary Metabolite analysis Shell
AT	Acyltransferase
BAC	Bacterial artificial chromosome
BLAST	Basic local alignment search tool
BGC	Biosynthetic gene cluster
bp	Base pair
ºC	Degree Celsius
C domain	Condensation domain
CHS	Chalcone synthase
cm	Centimetre
CTAB	Cetyl trimethylammonium bromide
DEBS	6-deoxyerythronolide B synthase
DH	Dehydratase
dH_2O	Distilled water
DMSO	Dimethyl sulfoxide
DNA	Deoxyribose nucleic acid
EDTA	N,N,N´,N´-Ethylendiamintetraacetat
ER	Enoylreductase
FDA	Food and Drug Administration
h	Hour
HMW	High molecular weight
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	Kilobase pair
KR	Ketoreductase
KS	Ketosynthase
kV	Kilovolt
LB medium	Luria-Bertani medium
LMP	Low melting point
min	Minute
mg	Milligram
ml	Millilitres
MT	Methyltransferase
NCBI	National Centre for Biotechnology Information
NGS	Next generation sequencing
NRP	Nonribosomal peptide
NRPS	Nonribosomal peptide synthetases
OD	Optical density
PCR	Polymerase chain reaction
PCP	Peptidyl-carrier protein
PDB	Phage dilution buffer
PKS	Polyketide synthase
pH	Potential of hydrogen
RNA	Ribonucleic acid
rpm	Revolutions per minute
rRNA	Ribosomal RNA
RT	Room temperature
sec	Second
TE	Thioesterase
UV	Ultraviolet
V	Volt
WGS	Whole genome shotgun

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF COMPOUNDS' STRUCTURES

1 Introduction

Natural products are small organic molecules produced by living organisms. Although not being considered essential for growth or reproduction, they provide important ecological advantages in the ecosystems where they are found [1]. Since the dawn of humankind, these chemical substances rouse interest due to their interesting biological activities and therefore have been used to treat diseases.

As the statistics show, natural products and their derivatives have been a productive source of leads for drug development: between 1981 and 2010, 34% of all medicines approved by the Food and Drug Administration (FDA, USA) had a natural product origin [2,3]. Examples of natural product-derived drugs which revolutionized medicine are the antibiotics penicillin **(18)**[1], erythromycin **(6)** and tetracycline **(7)**, the antiparasitic avermectin **(12)**, antimalarials like quinine and artemisinin, the cholesterol lowering statins, immunosuppressants as cyclosporine **(21)** and rapamycin **(22)** and anticancer drugs, particularly tubulin-binding agents, like taxol [4].

1.1 Marine natural products

Oceans are incredible diversified reservoirs of life forms. Considering that about 70% of the Earth's surface is covered by water and that life itself first appeared in this environment, more than two billion years ago, it is safe to assume that a huge biotechnological potential lies within these ecosystems. In the last 40 years, several efforts were undertaken to explore these unique organisms and subsequently, their associated compounds [5]. That resulted in several new classes of therapeutic agents inspired by secondary metabolites produced by marine forms with biological activities towards cancer, infectious diseases, inflammation, allergy, and many more [5,6]. In recent years, research on marine natural products has observed a significant increase. By 2010, more than 15,000 marine natural products had been discovered, with 8368 of them only being recorded in the previous decade [7]. Interestingly, nearly 30% of all these compounds were discovered on marine sponges, making these animals the richest source of biologically active marine secondary metabolites [7,8]. In Table 1 are sumarized some of the most significant discoveries of sponge natural compounds.

---

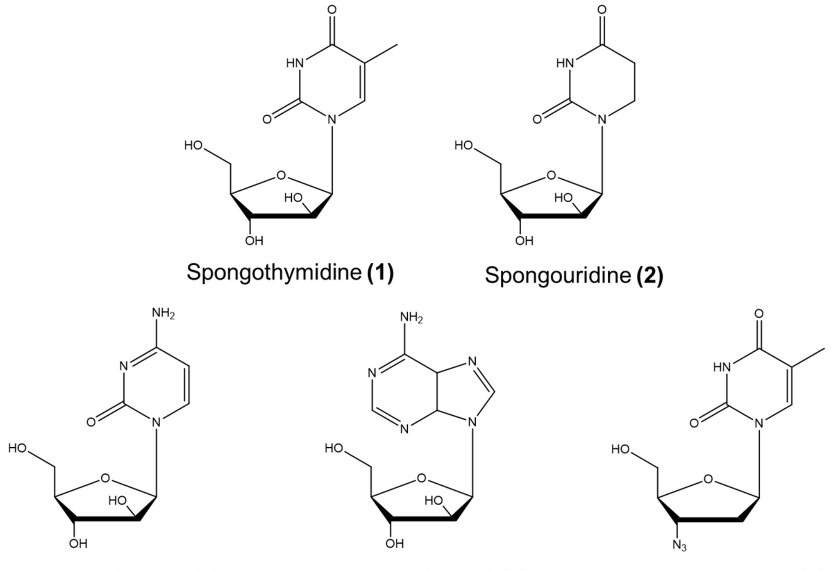
[1] The number denotes for the order in which the molecular structures of the compounds are presented in following figures.

Table 1 Selected natural compounds discovered from sponges [8,9].

Natural product	Sponge(s)	Biosynthetic Class of Agent	Activity	Reference
Hemiasterlin	*Hemiasterella minor, Auletta sp., Siphonochalin sp.*	Tripeptide	Antitumoral	[10,11]
Laulimalide	*Cacospongeia mycofijiensis*	Macrocyclic lactone	Antitumoral	[12]
Spongeothymidine **(1)** Spongeouridine **(2)**	*Cryptotethia crypta*	Nucleoside	Antiviral, antitumoral	[13]
Manzamine A	*Haliclona sp.*	Alkaloid	Anti-malaria, anti-tuberculosis, anti-HIV	[14]
Bengazole, bengamide	*Jaspis sp.*	Polyketide-nonribosomal peptide hybrid	Antifungal	[15,16]
Psymberin **(31)** (Irciniastatin A)	*Psammocinia bulbosa*	Polyketide	Antitumoral	[17]
Debromohymenialdisine (DBH)	*Stylotella aurantium*	Pyrroloazepine	Anti-Alzheimer	[18]
Swinholide A	*Theonella swinhoei*	Macrocyclic lactone	Antitumoral	[19]
Onnamide A **(29)**	*Theonella swinhoei*	Polyketide	Antitumoral	[20]
Topsentin	*Topsentia genitrix,, Spongosorites ruetzleri*	Alkaloid	Inflammation, asthma	[21,22]
Manoalide	*Luffariella variabilis*	Cyclohexane sesterterpenoid	Antiinflammatory, antibiotic	[23]
Halichondrin B	*Halichondria okadai Halichondrida*	Polyether macrolide	Antitumoral	[24]
Latrunculin A **(28)**	*Latrunculia magnifica*	Thiazole macrolide	Antitumoral	[25]
Pateamine A	*Mycale sp., Poecilosclerida*	Thiazole macrolide	Imunnesupressive	[26]
Discodermin	*Discodermia kiiensis, Lithistida*	Cyclic peptide	Antibacterial	[27]

1.1.1 Marine sponges

The discovery of the nucleosides spongothymidine **(1)** and spongouridine **(2)** from the caribbean sponge *Tethya crypta* by Bergmann *et al.* awoke the interest on marine sponges in the early 1950's [6,13,28]. These secondary metabolites provided the basis for the synthesis of cytarabine (Ara-C) **(3)**, the first marine-derived anticancer agent, and the antiviral drugs vidarabine (Ara-A) **(4)** and zidovudine (AZT) **(5)** (Figure 1) [29]. This marked the beginning of an intensive research on sponges, resulting in more than 4851 compounds reported until 2014 [7].



Spongothymidine **(1)** Spongouridine **(2)**

Cytarabine (Ara-C) **(3)** Vidarabine (Ara-A) **(4)** Zidovudine (AZT) **(5)**

Figure 1 Drugs developed from sponge natural compounds. Spongothymidine **(1)** and spongouridine **(2)** inspired the development of an antitumoral compound - Cytarabine (Ara-C) **(3)**, and two antiviral drugs - Vidarabine (Ara-A) **(4)** and Zidovudine (AZT) **(5).**

Being on Earth for more than 580 million years, sponges, from the phylum Porifera, are among the most ancient living Metazoa. More than 8,600 sponge's species are reported but is estimated that this number can round up to 15,000 species [30]. Sponges have been broadly categorized in 4 classes: Calcarea, Demospongiae and Hexactinellida and the recently recognized Homoscleromorpha [31]. These sessile animals are particularly abundant in tropical seawater ecosystems although being also present in temperate regions and in freshwater. Their ecological role in the benthic communities is crucial. Occupying up to 80% of the available substrate, they provide habitat for several species, take part in the biogeochemical fluxes and mediate the consumption and the release of nutrients [30].

Sponges have a simple body structure which lacks tissues or sensory organs, placing cells in direct contact with the surrounding seawater. That is crucial to their feeding strategy as they daily pump large volumes of water (up to 24 m^3 kg^{-1} sponge day^{-1}) [32] to sequester food particles and nutrients, leaving the expelled water essentially sterile. Despite lacking solid physical defences, sponges possess chemically complex defence mechanisms which allows them to live in extremely competitive environments [33].

Sponges house highly diverse, yet specific, symbiotic communities which can constitute more than 35% of the host's biomass [34,35]. The consortia of bacteria, archaea, unicellular algae, fungi, and viruses that reside within the sponge make essential contributions to many aspects of the sponge's physiology and ecology [8]. In fact, the whole community can be treated as one ecological unit, termed the holobiont [30]. Interestingly, many microbial phylotypes seem to live exclusively within sponge hosts and not in the surrounding seawater, making these communities often quite specific [8]. The most prevalent sponge-associated microorganisms are represented in the phyla Proteobacteria, Chloroflexi, Actinobacteria, Acidobacteria, Nitrospirae and in the candidate phylum, Poribacteria which occurs almost exclusively in marine sponges [8,30].

Even though much remains to be explained about the complexity of these associations, our knowledge is rapidly improving. Regarding the ecological roles of symbiotic organisms, one of the primary functions is related with host's defence. The symbionts' production of potent secondary metabolites seems crucial to the thrive of the physically fragile sponges [7,36]. In fact, it is suggested that the invertebrate-derived metabolites are actually produced by their microbial communities. This hypothesis, called the symbiotic hypothesis of marine natural product synthesis, is further supported by the bacterial like structure of many active compounds found in these environments. That has led to a renovated interest on studying these microbial communities for the bioprospecting of new drugs [37].

Nevertheless, the laboratorial cultivation of the bacterial symbionts offers a complex challenge. It is estimated that less than 1% of all bacteria are culturable with conventional techniques, a phenomenon called the Great Plate Count Anomaly [38], and about 70% of all bacterial phyla identified so far contain exclusively uncultured species [39]. Additionally, even when symbionts' cultivation conditions are achieved and microorganisms are cultured, the absence of critical environmental triggers often leads to the non-production of the target compound [37]. This helps to explain the strong focus of the scientific community on genomic technologies in the study of the metabolic potential of the sponges' symbionts [8].

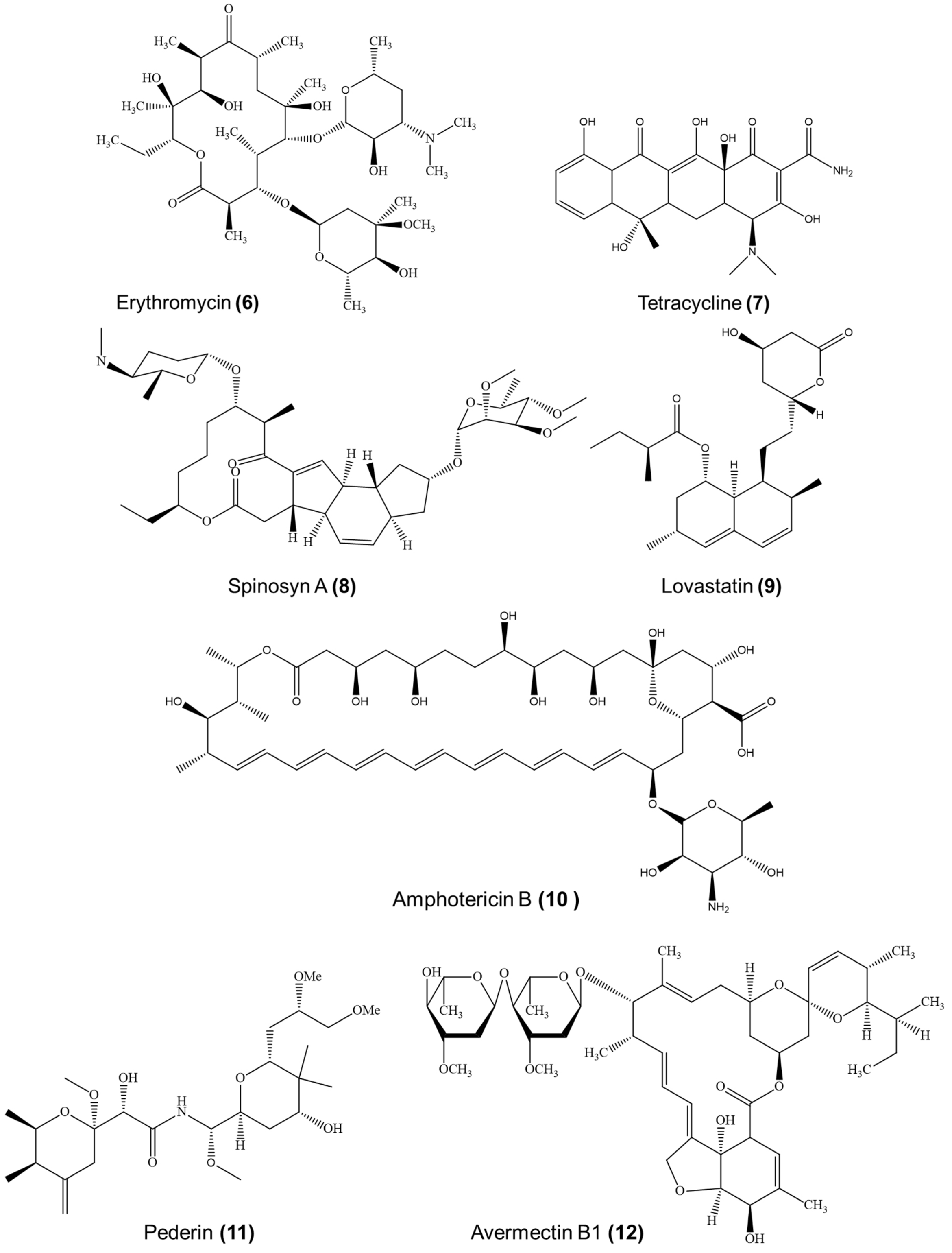
1.2 Microbial secondary metabolites

Sponges, and their associated microflora, produce complex secondary metabolites from various biological precursors. Between the diversified types of synthesised compounds are alkaloids, terpenoids, glycosides, phenols, phenazines, polyketides, fatty acid products and peptides, amino acid analogues, nucleosides, porphyrins, aliphatic cyclic peroxides and sterols [7,8]. However, the most important sponge natural products are the polyketides and nonribosomal peptides (NRP). These compounds are synthesized by large multifunctional polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs) [40]. From a biosynthetic engineering perspective, the building-block-type assembly of these compounds gives a major contribution for their production and commercialization as drugs [1].

Interestingly, most of the microbial derived natural products are produced by metabolic pathways encoded by chromosomally adjacent genes called biosynthetic gene clusters (BGCs). These BGCs encode enzymes, regulatory proteins and transporters that are necessary to produce, process and export specialized metabolites [41].

1.2.1 Polyketide synthases

Polyketides are the largest family of structurally diverse secondary metabolites, synthesised in both prokaryotic and eukaryotic organisms. In human medicine, they are the source of several biologically active compounds which target both acute and degenerate diseases. As clinically relevant examples of polyketides we can point the antibiotics erythromycin A **(6)** and tetracycline **(7)**, the insecticide spinosyn A **(8)**, the cholesterol-lowering drug lovastatin **(9)**, the antifungal compound amphotericin B **(10)**, the antitumoral agent pederin **(11)** and the anti-helmintic avermectin B1 **(12)** (Figure 2) [42,43].



Erythromycin **(6)**

Tetracycline **(7)**

Spinosyn A **(8)**

Lovastatin **(9)**

Amphotericin B **(10)**

Pederin **(11)**

Avermectin B1 **(12)**

Figure 2 Representative polyketides.

Polyketides are built from simple carboxylic acid monomers through programmed events catalysed by PKSs. These giant multimodular enzymes have a biosynthetic mechanism similar to fatty acid synthases. Each module carries a variable set of catalytic domains and usually adds, by successive rounds of decarboxylative Claisen condensation, one thioesterified acyl extender unit to the growing chain. The mostly used building blocks are the residues of acetate (malonyl-CoA) and propionate (methylmalonyl-CoA), but more complex units can also be used [44].

PKSs are divided into three classes or types, according to the structural organization of their domains. It is also important to acknowledge an additional class of hybrids with NRPSs which will be discussed in Section 1.2.3 (Table 2) [45]. The type I and II PKSs correspond to the previously classified enzymes of fatty acid biosynthesis [43]. Linearly organized catalytic domains within large multifunctional enzymes are referred to as type I, whereas in type II PKS, a complex of discrete monofunctional enzymes can be found. Type II and III PKSs generate aromatic molecules, but use fundamentally different approaches to construct the polycarbonyl chain and channel it into a cyclic fate. Besides the enzyme structures, the mechanism of synthesis is another characteristic to classify PKS. Depending on whether a module is used only once or repeatedly, PKSs can also be classified as modular (non-iterative) or iterative [44].

Table 2 PKS types (adapted from [43,44]).

Type	Extender units	Organisms
Type I	Various	Bacteria and fungi
Type II	Manoyl-CoA	Bacteria
Type III	Manoyl-CoA Acyl-CoA	Mainly plants, some bacteria and fungi
PKS/NRPS hybrids	Manoyl-CoA Amino acids	Bacteria (modular), fungi (iterative)

Even though the structural and organizational diversity of PKSs is tremendous, one has marked the polyketide biosynthesis' research since its discovery: the 6-deoxyerythronolide B synthase (called "DEBS"). This enzyme is responsible for the macrolide erythromycin A **(6)** biosynthesis in the soil bacterium *Saccharopolyspora erythraea* [46]. It was the first PKS to be discovered, in the 1990's, and remains the most intensively studied, serving as the prototypic to which all other modular PKS are compared. DEBS has over 3000 amino acids and is composed of six elongation modules flanked by a loading didomain and a thioesterase [43] (Figure 3). The modules are further organized into three multimodular subunits: DEBS 1, 2 and 3 [42].

The prototypical PKS module, present in the DEBS, is composed by a ketoacyl synthase or ketosynthase (KS), an acyl transferase (AT), and an acyl carrier protein (ACP). In the initial step of the polyketide erythromycin **(6)** synthesis, the carboxy group of the starter unit is chosen by the AT domain and it is covalently attached to a 5'-phosphopantetheinyl SH group of the ACP. Then, the polyketide chain, transferred to the KS domain, is elongated by a Claisen-type thioester condensation, and optionally further modified depending on the domains present on the respective module. A ketoreductase (KR) reduces the β-keto group to an alcohol function, a dehydratase (DH) removes one H_2O molecule, originating an α, β-saturated moiety, and an enoylreductase (ER) converts the olefinic unit in a saturated system [44]. The PKS chain is thus transferred from module to module until its final structure is achieved. Finally, a thiosterase (TE) hydrolyses and releases the completed polyketide chain. In erythromycin **(6)** biosynthesis, the resulting aglycone 6-deoxyerythronolide B **(13)** is further hydroxylated twice, glycosylated twice and methylated on a sugar residue to form the fully active antibiotic compound [42].

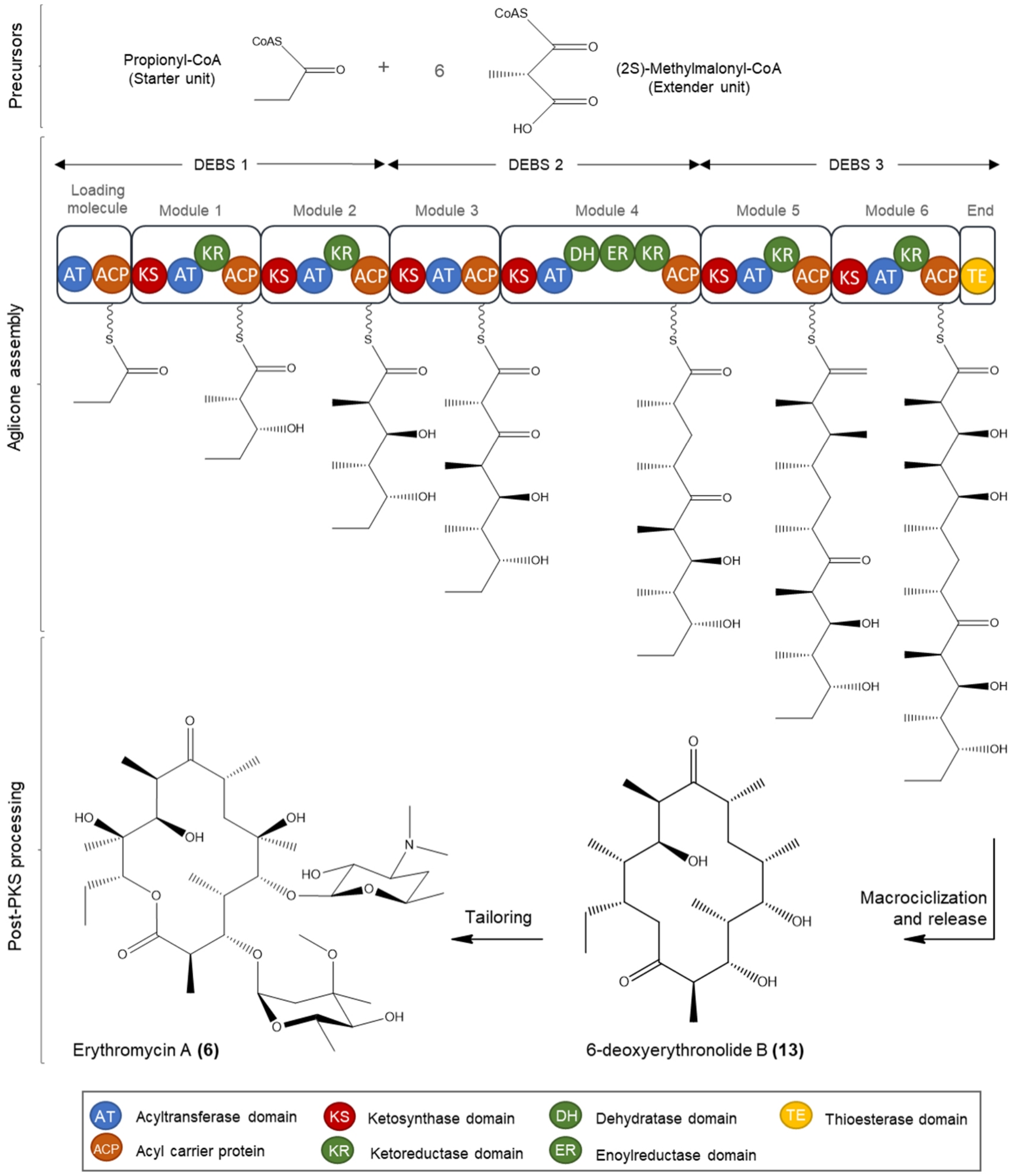
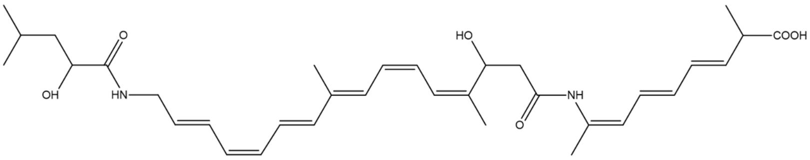


Figure 3 Erythromycin A **(6)** biosynthetic pathway by DEBS. Adapted from [42].

1.1.1.1. Type I PKSs

The already described DEBS is characterized as a *cis*-AT modular type I PKS, the most intensively studied group of PKS enzymes. These enzymes are particularly attractive to drug research due to the characteristic organization of the enzymatic machinery in ‘‘assembly lines’’. As each active site in modular type I PKSs is used only once during polyketide biosynthesis, the order of the PKS modules correlates with the required sequence of biosynthetic transformations [43]. This collinearity rule allows the establishment of good predictions about metabolite structure, considering the enzyme architecture, and the opposite situation is also effective [44]. Secondly, the specific characteristics of these enzymes make them highly prone to drug research, as they are very adaptable and amenable to pathway engineering [45].

Another smaller and architecturally different type I PKS has been recently reported: the *trans*-AT PKS. Differently from *cis*-AT PKSs, in these enzymes the ATs are encoded by discrete genes and the modules lack AT domains. Instead, stand-alone ATs are responsible for loading the extender unit malonyl-CoA to AT-less modules [44]. Altogether, unusual domain sets, appearing redundant modules and novel domain types contribute to the non-linear assembly of these polyketides [43]. This makes metabolite structure prediction a harder approach. The BGC for the antibiotic bacillaene **(14)** (Figure 4)**,** present in several *Bacillus* species, is considered the prototype for this group of type I PKS systems [47].



Bacillaene **(14)**

Figure 4 Bacillaene **(14)**, a polyketide generated by a *trans*-AT PKS system.

1.1.1.2. Type II PKSs

Type II PKSs encompass discrete catalytic functions that are combined into a productive complex. The ‘‘minimal PKS’’ includes the KSα, KSβ, and ACP domains, which perform a defined number of chain-extension cycles to build the final polyketide (Figure 5). While KSβ anchors the growing polyketide chain, KSα catalyses the Claisen type condensation between the growing polyketide and incoming acyl-CoA subunits. The two KS subunits have very similar sequences, being the central difference the absence of the active site cysteine in KSβ. In type II PKSs, additional subunits can include KR, cyclases and aromatases [44].

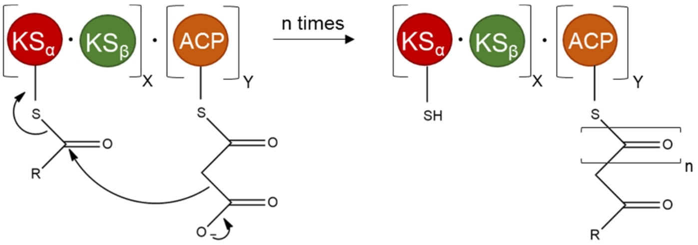
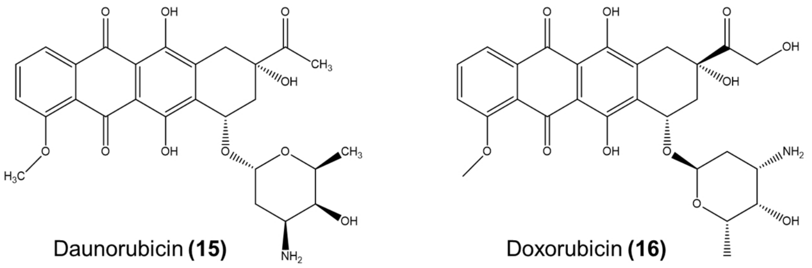


Figure 5 Type II PKS, an iterative system. Adapted from [43].

Typical primary products of type II PKSs are the polyphenols (Figure 6). Well-known examples of this class are the antitumoral agent daunorubicin **(15)** and its derivative doxorubicin **(16)** [47].



Daunorubicin **(15)**

Doxorubicin **(16)**

Figure 6 Daunorubicin **(15)** and its derivative doxorubicin **(16)**.

1.1.1.3. Type III PKSs

The third PKS class links to the multifunctional enzymes of the chalcone synthase (CHS) type, long been found in plants, but more recently have also been encountered in bacteria and fungi [48]. Type III PKSs are composed by a multifunctional active site, responsible for all the steps necessary to the polyketide synthesis. In collaboration with CoA-bound substrates, selects the starter unit, assembles the chain and promotes its folding [44] (Figure 7).

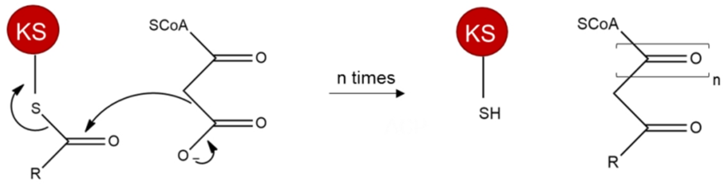
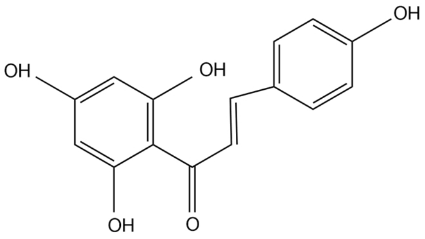


Figure 7 Type III PKS, an iterative and ACP-independent system. Adapted from [43].

The well-known CHS starts from one p-coumaroyl-CoA and three malonyl-CoA units which produces the tetraketide chalcone through subsequent Claisen ester condensation reactions. CHSs provide the starting material for several biologically important phenylpropanoid metabolites [44] (Figure 8).

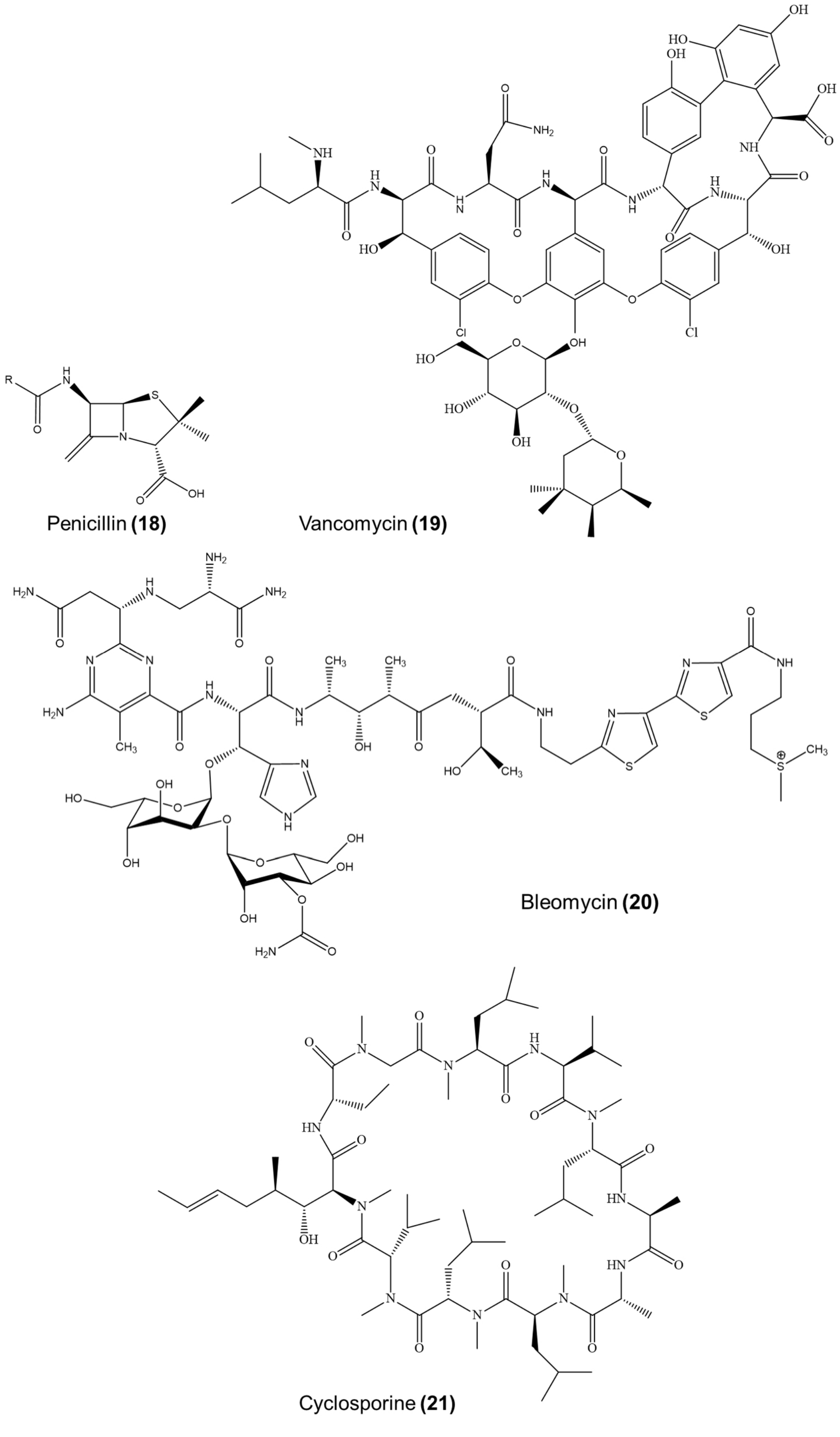


Naringenin Chalcone **(17)**

Figure 8 Naringenin chalcone **(17)**.

1.2.2 Nonribosomal peptide synthetases

NRPS are large, multimodular enzymes that assemble numerous structural and functionally diverse peptides. These enzymes are mostly found in bacteria and fungi. The starting material for the biosynthesis of NRPs are the 20 proteinogenic amino acids and several nonproteinogenic amino and aryl acids [49]. These are sequentially incorporated into the growing peptide chain by specific domains. According to the collinearity rule, the number and the order of the modules defines the number and the order of amino acids in NRPs. More than 20 marketed drugs are NRP, including antibiotics like penicillin **(18)** and vancomycin **(19)**, the antitumoral compound bleomycin **(20)**, and the well-known immunosuppressant cyclosporine **(21)** [37,50] (Figure 9). NRPs are particularly interesting for the development of new therapeutics mostly due to their excellent binding properties, low off-target toxicity, and high stability [51].



Penicillin **(18)**

Vancomycin **(19)**

Bleomycin **(20)**

Cyclosporine **(21)**

Figure 9 NRP-based drugs.

As observed in polyketide biosynthesis, NRPs are usually synthesised in three main stages: building-block assembly, NRPS-mediated peptide assembly and post-NRPS modifications. In the minimal NRPS elongation module, the adenylation (A) domain selects, activates and loads the amino acid. The activated substrate is then transferred to the thiolation domain, also called peptidyl-carrier protein (PCP). Finally, the condensation (C) domain of the downstream module forms an amide bond between the elongated chain and the activated amino acid, coupling to the upstream nascent peptide chain [37,50] (Figure 10).

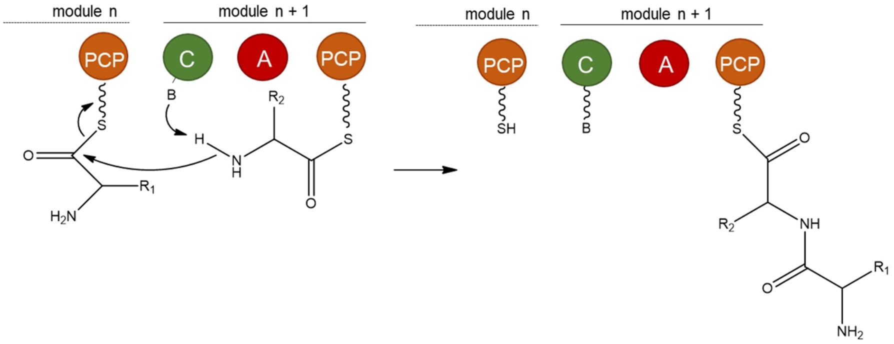


Figure 10 Peptide bond formation in NRPs. Adapted from [49].

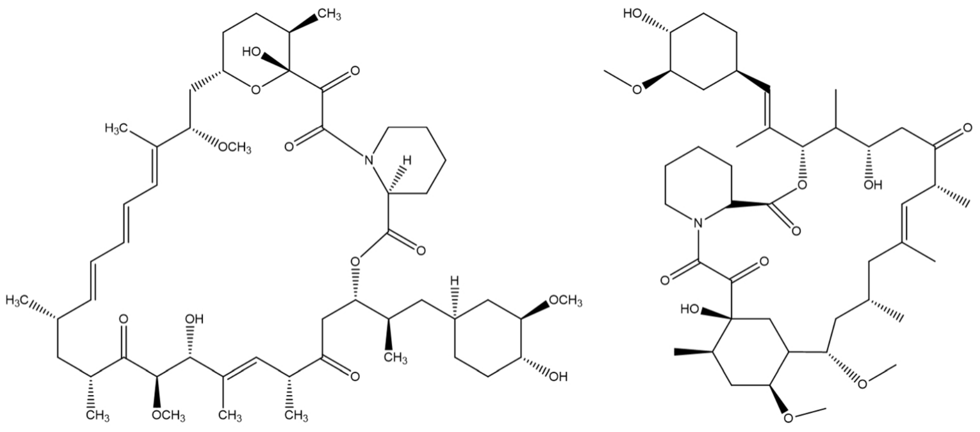
During chain elongation, the growing peptide chain can be further modified. An heterocyclization domain forms a thiazoline or oxazoline ring. These five-membered rings can be further oxidized by oxidation domains into stable thiazole or oxazole heterocylces. An epimerization domain can modify an L-amino acid to an D-amino acid. A methyltransferase can transfer a methyl group from an S-adenosylmethionine to the respective nitrogen or carbon atom. Dual/epimerization can occur, being responsible for both epimerization and condensation reactions [37].

The final domain of an NRPS is usually a thioesterase which, through hydrolysis or cyclization, releases the peptide chain from the assembly line. The resulting linear or cyclic peptide can be further modified by post-NRPS reactions [49].

1.2.3 Hybrid PKS/NRPS compounds

Reflecting the inherent similarity between the two megasynthases, mixed PKS/NRPS systems can also occur. PKS/NRPS-derived products can be divided into two main classes, depending on the gene cluster organization. The compound can either be synthesized individually by a NRPS or a PKS and eventually coupled into a hybrid final product or produced by functionally connected PKS/NRPS enzymes, producing a hybrid polyketide-peptide metabolite [37].

Several PKS/NRPS hybrids, such as the antitumoral agents rapamycin **(22)**, known as Sirolimus, and compound FK506, named tacrolimus **(23)**, are clinically valuable natural products [2] (Figure 11).



Rapamycin **(22)** Tacrolimus - FK506 **(23)**

Figure 11 Examples of PKS/NRPS hybrids.

Epothilones A **(24)** and B **(25)**, which were originally detected in 1987 in the myxobacterium *Sorangium cellulosum* [52], have a recognized potent antitumoral activity due to its capacity of interfering in tubulin polymerization. The biosynthetic pathway for the macrolactones epothilones (Figure 12) has been intensively studied [53], providing a good starting point for the prospect of novel PKS/NRPS hybrids.

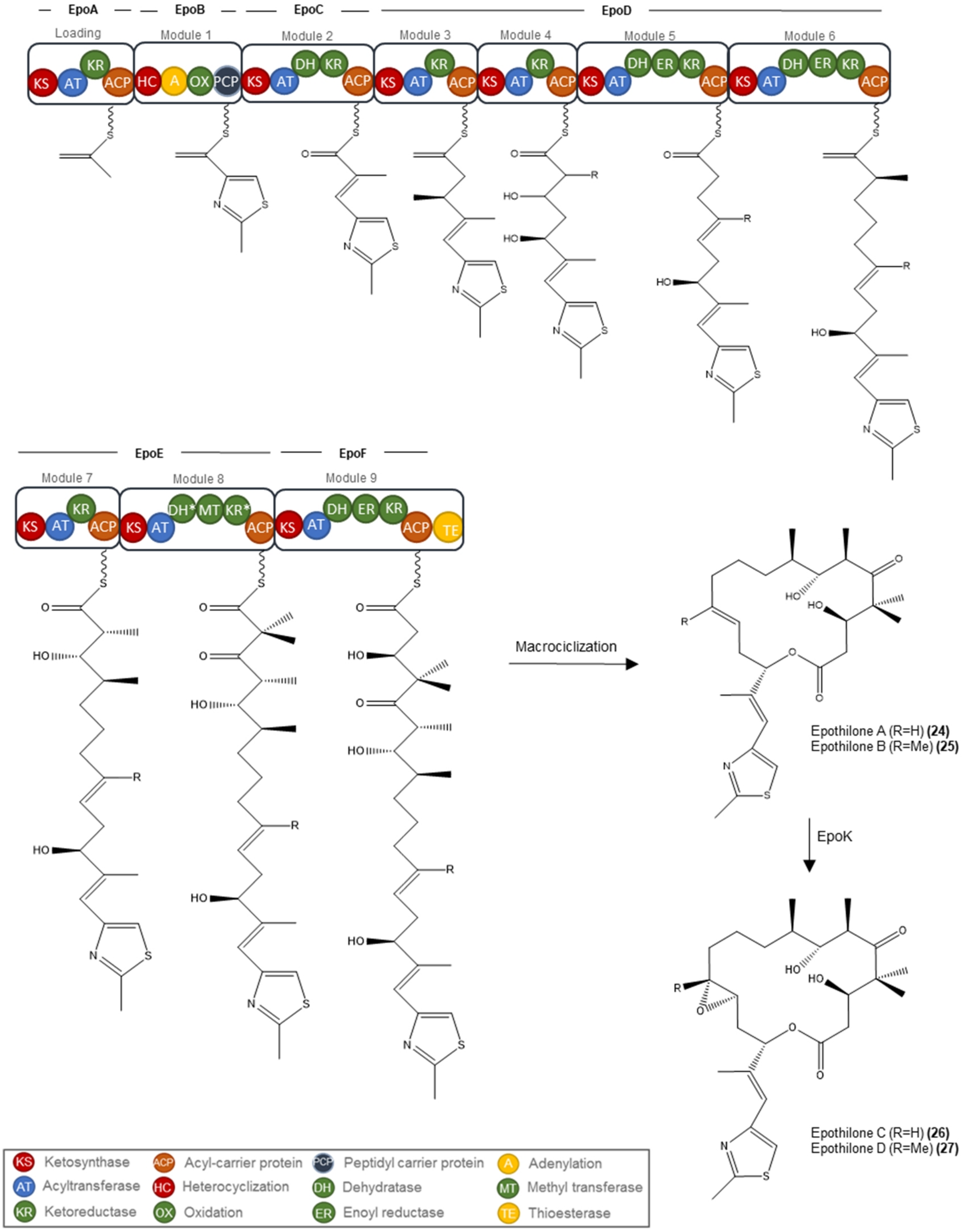


Figure 12 Epothilone biosynthetic pathway. A PKS/NRPS mixed BGC in which one NRPS module (EpoB) is surrounded by PKS modules. Adapted from [53].

1.3 Uncovering the biosynthetic potential within marine sponges

As already described, the biotechnological potential within marine life is almost unlimited. However, the survey of new drug leads has numerous inherent difficulties which have delayed the full exploitation of marine resources. For example, several challenges were found in sample collection and analysis. As the target compounds are normally not abundant in the organisms, considerable amounts of sample for structural elucidation were traditionally necessary. However, high throughput technologies are now allowing to solve structures at a nanomolar scale [4]. Of notice, the powerful mass-spectrometry methods and its associated advanced algorithms used in combination with flow-cell NMR spectrometers [5].

1.3.1 Microbiome analysis

The development of molecular biology technologies revolutionized the approaches on the research of organisms related with interesting natural products. For example, in the late 1980s, the advent of the polymerase chain reaction (PCR) involving the amplification of the 16S ribosomal RNA (rRNA) gene, gave access to unknown unculturable bacteria [54]. The 16S rRNA gene is almost universally present in bacteria. It has nine highly conserved regions which serve as universal primer binding sites for PCR amplification. Simultaneously, the interspecific polymorphisms in its variable regions enable phylogenetic reconstruction, with statistically valid measurements, for bacterial identification and classification [55].

Nowadays, the direct sequencing of 16S rRNA variable regions is the most used technique in the prospect of bacterial phylogeny and taxonomy [56]. It is a powerful tool to visualize the sponge microbiome as also to phylogenetical compare organisms of distinct species, locations and habitats [57-59]. Recently, a global survey was conducted by Torsten *et al.*, providing important conclusions on this matter [35].

Even though this tecnique has these advantages, it only provides a basic idea of the metabolic potential of the sponge microbiome [59]. Other approaches based on the functionality of genes are essential to understand the bigger picture and therefore are briefly presented next.

1.3.2 Genomic approaches

The arrival of the 21st century shed a new light regarding bacterial genomes with the complete sequencing of *Streptomyces coelicolor* strain A3(2) [60]. After more than 50 years of research, only three BGCs were known to exist in the genome of this natural products-producing strain [1]. With the new sequencing approaches, it was discovered that its genome harbours about 30 BCGs. This launched the idea that even well studied strains may hold the potential to synthesize a larger number of compounds than those analytically detected. From this recently found genetic potential, sprouted the concept of natural-product genome mining, which concentrates the efforts on the analysis of the genome [1]. As the cost of sequencing decreased, several bacterial genomes became accessible. In addition, technologies like metagenomics and single cell genomics generate immense data to be analysed [42]. This creates interesting opportunities to expand our knowledge on the secondary metabolism of the sponges' symbionts [1,61].

Next-generation sequencing (NGS) is a DNA sequencing technology first introduced into the market in 2005. In comparison with the most sophisticated capillary sequencers based on the Sanger method, NGS can generate over 100 times more data [62]. It produces shorter (e.g. 100-250) reads with an exceptionally high genome coverage [63], providing a powerful tool to study DNA and RNA samples. Several different platforms are commercially available: 454 Roche, Illumina, Ion Torrent and Helicos, all of them relying on sequencing by synthesis. More recently, new sequencing methods based on single-molecule sequencing are evolving, giving birth to third-generation sequencing technologies like Pacific Biosciences and Oxford Nanopore [64].

1.3.2.1 Genomic mining

Genomic mining is the process of prediction and isolation of natural products based only on the genetic information [65]. The classic approach searches for enzymes involved in the biosynthesis of secondary metabolites, by screening the genome for genes within BGCs [61]. Starting from a BGC is possible to infer the class of the natural compound and, in some case, their structure [41].

However, the predictive accuracy of genome mining is intricately dependent of already made experimental links between the BGCs and the products they encode. This is generally accomplished by using laborious, and often challenging, laboratory experiments such as gene knock-out or heterologous expression [1]. Once a gene function is determined and experimental links are established, the detection of a similar BGC in other strains can be used to predict the production of similar compounds [65]. Nevertheless, from the tens of thousands of compounds already identified by natural product chemistry, the clear majority have not been linked to their respective BGCs [1]. This explains why bioinformatics approaches are still unable of forecast some BGCs [61].

In addition, the inherent complexity of the enzymes encoding for secondary metabolites makes genomic mining a challenging task. Especially large and highly repetitive gene clusters such as NRPS and PKS pathways are rarely fully assembled [65]. In addition, the most promising bacteria on natural-product biosynthesis tend to have large genomes and complex life cycles. Nevertheless, over the past two decades, the BGCs for a broad range of polyketide and nonribosomal peptide natural products have been identified and characterized [45].

1.3.2.2 Metagenomics

Over the past 5 to 10 years, the field of metagenomics has brought significant advances in the study of microbial communities, shedding new light over unknown biosynthetic pathways [66]. Metagenomics (also referred to as environmental and community genomics) can be defined as the application of genomics to uncultured microorganisms [67].

Metagenomic analyses can be based on a known sequence or on the functional genomic potential of an environmental sample. When the sequence of a gene of interest is known, PCR primers or hybridization probes are designed to search for the presence of desired genes. On the other side, a functional metagenomics approach can be employed when sequence data is not known. After identified and isolated, the BGC of a specific compound can be further studied by heterologous expression [9].

Metagenomic techniques focus on the analysis of complex DNA mixtures isolated from entire samples or from enriched preparations. The ultimate goal is to properly assemble individual DNA sequences into genes or genomes [68]. For example, in whole genome shotgun (WGS) metagenomics, all DNA material of an environmental sample is randomly sheared in small fragments, sequenced by NGS and assembled. This primer-independent approach provides a significant amount of unbiased information on genes, functions and organisms and a good estimation of microbial community composition and diversity [69,70].

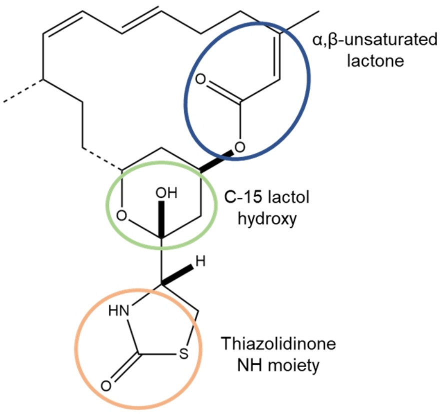
A widely used metagenomic approach is the construction of a metagenomic library. The basic workflow starts with the extraction of the microbial DNA of an environmental sample. After size selected, the collection of DNA fragments is inserted into an appropriate vector which can be, among others, a fosmid (F1 origin-based cosmid vector) or a bacterial artificial chromosome (BAC). The vector-insert constructs are then transformed into a host bacterium, being *Escherichia coli* the most common [67,71].

Several large-insert (40–200 kb) fosmid and BAC clone libraries have been successfully constructed from different sponges. A crucial advantage of these vectors is the capability of including complete gene clusters or operons, simplifying the BGC screening procedure [72]. These successful libraries led to the identification of several new PKS genes, as for example in sponges *Discodermia dissoluta* [73] and *Theonella swinhoei* [74], and new NRPS genes, like those identified in sponges *Haliclona okadai* [75] and *Aplysina aerophoba* [76]. Interestingly, marine metagenomic libraries have also allowed the identification of a wide range of biocatalysts, such as esterases, lipases, and chitinases [9]. However, taking into accont all the efforts, the number of compounds discovered from sponge metagenomic libraries is rather small. The numerous difficulties in recovering high molecular weight (HMW) DNA from sponge symbionts to construct large insert libraries can be pointed out as one of the main reasons for this [75].

1.4 State of the art

Between the myriad of possible applications of marine natural products, one has stood out: its use in the anticancer therapeutic arsenal [77]. Currently, there are four marine inspired commercialized anticancer drugs (two of them from sponges) and several compounds in clinical trials [77]. Sponges are interesting sources of secondary metabolites with cytotoxic activity. In fact, these compounds were found in 10% of all sponge samples studied so far [78]. Nevertheless, the biosynthetic mechanisms of several potential antitumoral compounds are still unknown. One example of this situation is latrunculin A **(28),** a compound firstly described by Kashman and co-workers in the Red Sea sponge *Negombata magnifica* (previously named *Latrunculia magnifica*) [25] whose strong cytotoxicity activity continues to arouse interest (Figure 13).

Latrunculin A **(28)** is a hybrid PKS/NRPS macrolide composed by a 16-membered ring fused to a tetrahydropyran containing a 2-thiazolidinone moiety [79]. It reversibly binds to actin monomers forming 1:1 complexes with G-actin, disrupting its polymerization [80]. In addition, this compound has showed antiviral and antibacterial activities, the capacity of reducing intraocular pressure, inhibition of stress-activated MAP kinase (SAPK) pathway and suppression of the hypoxia-induced factor (HIF-1) activation in breast cancer cells [80]. Even though total synthesis has already been accomplished for latrunculin A **(28)**, the complete biosynthetic mechanism remains unknown.

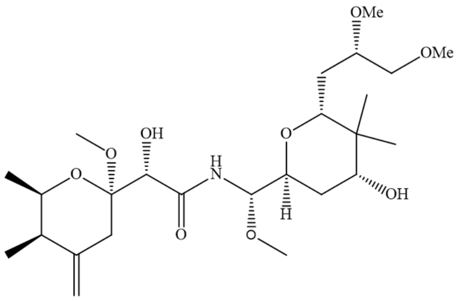


Latrunculin A **(28)**

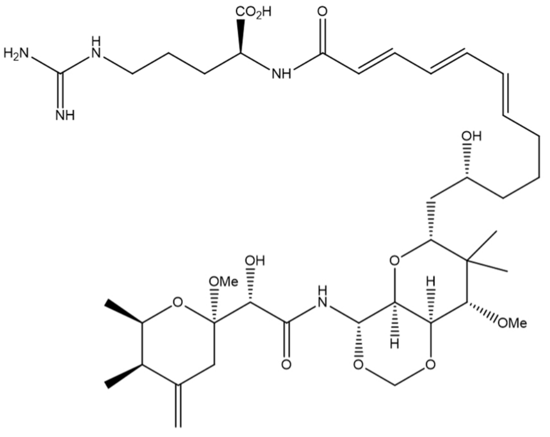
Figure 13 Key pharmacophores of latrunculin A **(28)**.
These are responsible for its actin binding bioactivity [81]**.**

The search for an unknown biosynthetic pathway can embrace several different approaches. In the specific case of sponges' secondary metabolites, the construction of metagenomic libraries with subsequent screening for targeted genes has proved its success. The discovery of several BGCs based on the pederin **(11)** BGC is a good example. Pederin **(11)** is a potent antitumoral agent produced by an uncultured bacterial symbiont of *Paederus fuscipes* rove beetle and its BGC was the first complete PKS/NRPS gene cluster of an uncultured bacterial symbiont to be identified [82]. Later, Piel *et al.* successfully employed a pederin-informed PCR screening to identify PKS gene cluster of onnamide A **(28)**, an interesting compound found on the Japanese sponge *Theonella swinhoei* [74]. As onnamide A **(28)** is very similar in structure to pederin **(11)**, primers for the amplification of PKS gene regions were successfully employed in the identifcation of the candidate gene cluster [9,37].

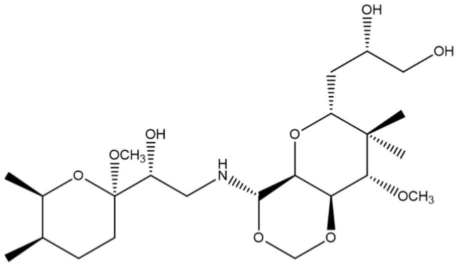
Nevertheless, the use of universal degenerate primers to identify conserved PKS domains has a significant downside: with general primers, too many diversified PKS gene sequences are amplified, difficulting the process of screening for a specific gene cluster. As a solution, Fisch *et al.* applied a nested PCR strategy in which degenerate primers are firstly used to amplify the ketosynthase (KS) domain of the PKS. This is followed by the application of a more specific primer pair. As a result, the BGC from the pederin **(11)** relatives mycalamide **(29)** and psymberin **(30)** were identified from marine sponge metagenomes [9,83] (Figure 14).



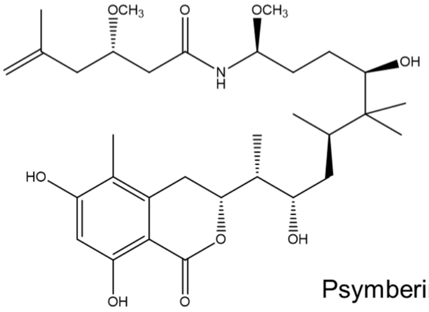
Pederin **(11)**



Onnamide **(29)**



Mycalamide A **(30)**



Psymberin A **(31)**

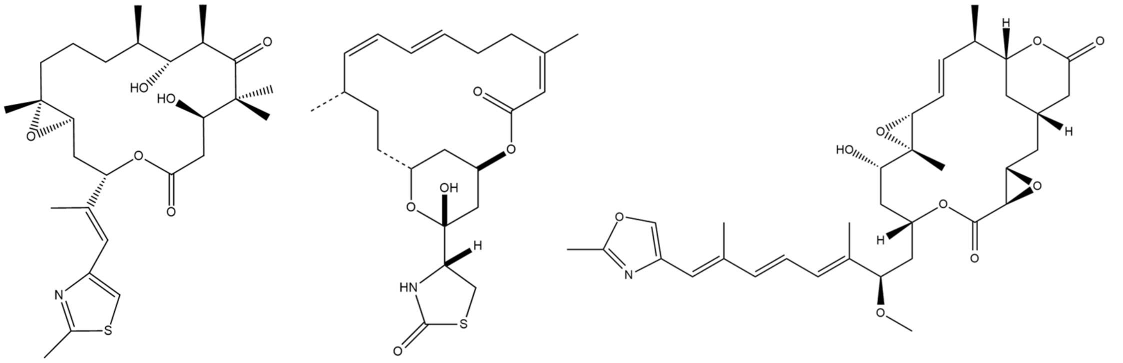
Figure 14 Compounds whose BGC was discovered through a pederin **(11)** informed PCR screening: Onnamide A **(29)** and mycalmide A **(30)** from *T. swinhoei*, and psymberin **(31)**.

1.5 Scope of the work

As exposed in last section, even though latrunculin A **(28)** has several characteristics that picture its potential as a drug lead, its BGC is still unknown. The main goal of this work was to provide new insights on its biosynthetic mechanism. To that end, it was aimed to perform a metagenome analysis of sponges containing latrunculin A **(28)**. The sponge samples were collected in Bunaken National Park (Indonesia), in 2015 and 2016, and are now part of the repertoire of the research group in which this work was developed (professor Till Schäberle lab).

To achieve this goal, three different approaches were pursuit. Firstly, it was intended to produce a representative metagenomic library of a sponge sample. Important insights on the technical application of this method could be obtained, allowing further development and adaptation of the already available knowledge on the specific matter.

Secondly, a structure-based PCR screening of the metagenomic DNA was implemented. Latrunculin A **(28)** shares a carbon skeleton with the cytotoxic secondary metabolite epothilone A **(24)** [79] (Figure 14). In epothilone, the polyketide backbone is synthesized by type I PKS and the thiazole ring is derived from a cysteine incorporated by a NRPS. In addition, latrunculin A **(28)** shares common features with the potent anti-mitotic macrolide rhizoxin **(32)**, particularly the characteristic β-branch of *trans*-AT PKSs. As the biosynthetic pathway of these compounds is well known (epothilone BGC is presented in Figure 14), a gene targeting approach was developed. It was anticipated that PCR screening of PKS and NRPS conserved domains would result in DNA fragments that, after sequencing, could be subjected to bioinformatics analysis.



Epothilone A **(24)** Latrunculin A **(28)** Rhizoxin **(32)**

Figure 15 Chemical structures of epothilone A **(24),** latrunculin A **(28)** and rhizoxin **(32).**

Finally, additional insights on the sponges' microbiome were provided through sponge's microbiome taxonomy profiling with 16S rRNA gene sequencing. The goal was to analyze the possible advantages of this approach on the identification of compounds' producers and clarification of their BGCs.

Altogether, it was envisioned that the combined results of these approaches would provide insights on the biosynthetic production of latrunculin A **(28)**, increasing the potential of its further development as a drug lead and, consequently, of its biotechnological production.

2 Materials and methods

This chapter focuses on the materials and methods used and applied throughout the conducted work. First, a list of the materials is provided which is followed by a detailed description of the carried procedures. Finally, software and databases essential for the bioinformatic analysis of the produced data are presented.

2.1 Materials

2.1.1 Organisms

Sponge samples and bacterial strains used in this work are presented in Tables 3 and 4, respectively.

Table 3 Sponge samples analysed.
L+ and L- state for the presence, or not, of latrunculin A.

Sponge sample	Source
Chdi15Bu13S^{L+}	Bunaken, 2015
Chdi15Bu24+26S^{L+}	Bunaken, 2015
Chan15Bu6S^{L+}	Bunaken, 2015
Chdi16Sa3-5S^{L-}	Bunaken, 2016
Chan15Bu13S^{L-}	Bunaken, 2015

Specimens of sponges were previously collected by scuba diving and snorkelling in the coastline of Bunaken National Park, Indonesia, in 2015 and 2016. The sampling actions were performed under the “Indobio” Biodiversity and Health project framework. Immediately after collection, they were preserved in 95% ethanol and kept at -20 ºC. Previous work confirmed the presence of the compound of interest, latrunculin A, in three of these sponges: Chdi15Bu13S^{L+}, Chdi15Bu24+26S^{L+} and Chan15Bu6S^{L+} (Figure 16).

Figure 16 Underwater photographs of the sponge samples.
(a) Chdi15Bu13S^{L+}, **(b)** Chdi15Bu24+26S^{L+}, **(c)** Chan15Bu6S^{L+}.

Most part of this work, however, was carried with sample Chi15Bu13S^{L+} (Figure 17).

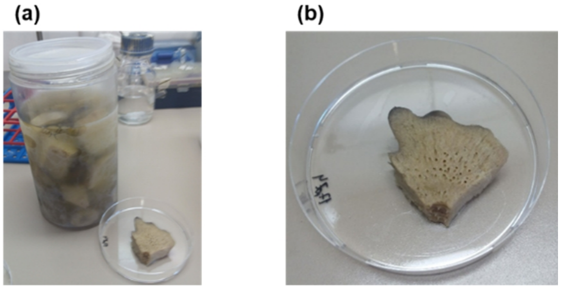


Figure 17 Chdi15Bu13S^{L+} sponge sample. **(a)** Recipient of conservation. **(b)** Example of a cut for metagenomic DNA extraction in a petri dish.

Table 4 Bacterial strains used in this work.

Bacterial host	Genotype	Manufacturer
E. coli α-Select Silver Competent Cells	F-deoR endA1 recA1 relA1 gyrA96 hsdR17(rk-, mk+) supE44 thi-1 phoA Δ(lacZYA argF)U169 Φ80lacZΔM15λ-	Bioline
E. coli EPI-300™-T1^R	F- mcrA D(mrr-hsdRMS-mcrBC) f80dlacZDM15 DlacX74 recA1 endA1 araD139 D(ara, leu)7697 galU galK λ- rpsL nupG trfA tonA dhfr	Epicentre Biotechnologies

2.1.2 Chemicals, solutions and media

Chemical reagents and solutions, antibiotics, buffers and culture media used in this work are listed in Tables 5, 6, 7 and 8, respectively. Culture media were autoclaved at 121 ºC for 20 min.

Table 5 Solvents and reagents used.

Substance	Manufacturer	Substance	Manufacturer
Agar	Roth	NaCl	Merck
Chloroform	Roth	NaOH	Merck
CTAB (Cetrimonium bromide)	Fluka	Nitrogen, liquid	Linde
dATP	Fermentas	peqGOLD Agarose	PEQLAB Biotechnologie
DMSO (Dimethyl sulfoxide)	Roth	peqGOLD Low-Melt Agarose	PEQLAB Biotechnologie
dNTP (Deoxynucleoside triphosphates)	Promega	Phenol	Merck
EDTA (Ethylenediamine-tetraacetic acid)	Sigma-Aldrich	Potassium chloride	Merck
Ethanol 99,8% p.a	Roth	SeaPrep Agarose	Lonza
Ethidium bromide	Roth	Sodium acetate	Merck
Gel Loading Dye	Termo Scientific	Sodium chloride	Merck
Glycerol	Roth	Tryptone	Roth
Glycerin	Merck	Tris-Base	Roth
IPTG (Isopropyl-ß-D-thiogalactopyranoside)	Roth	Tris-HCl	Roth
Isoamyl Alcohol	Roth	Urea	Roth
Isopropanol	Roth	Yeast extract	Fluka

Table 6 Antibiotics for selection of recombinant strains.

Antibiotic	Media concentration	Stock	Solvent	Manufacturer
Carbenicillin	100 µg/ ml	100 x	Water	Fluka
Chloramphenicol	25 µg/ ml	1000 x	Ethanol	Fluka

Table 7 Buffers and solutions.

Buffer	Composition	Store
Phage dilution buffer (PDB)	10 mM Tris–HCl (pH 8.3) 100 mM, NaCl 10 mM $MgCl_2$	RT
Alkaline lysis solution I (P1)	50 mM Tris-HCl (pH 8.0) 10 mM EDTA	4 ºC
Alkaline lysis solution II (P2)	200 mM NaOH 1% SDS	RT
Alkaline lysis solution III (P3)	3 M KOAc (pH 5.5)	RT
Sponge Lysis Buffer	8 M urea 2% sarkosyl 1 M NaCl 50 mM EDTA	RT
TE Buffer	10 mM Tris-HCl (pH 7.5) 1 mM EDTA (pH 8.0)	RT
TAE Buffer 50x	242 g Tris-Base 57.1 ml acetic acid 100 ml 0.5 M EDTA Add water to 1000 ml	RT

Table 8 Bacterial cultivation media.

Media	Composition
Luria-Bertani (LB) medium	10 g tryptone 5 g yeast extract 10 g NaCl Add water to 1000 ml of final volume (pH 7.5)
LB agar	15 g agar in 1 L of LB broth medium
SeaPrep agarose	5 g SeaPrep agarose in 1 L of LB broth medium

2.1.3 Molecular biology materials and enzymes

MaxPlax™ Lambda Packaging Extracts (Epicentre Biotechnologies) were used to package the pCC2FOS™ vector library. Molecular biology enzymes (Table 9) and molecular weight markers (Table 10) were stored and used according to manufacturers' instructions. The list of primers employed in this work is presented in Table 11.

Table 9 Used enzymes.

Enzyme	Manufacturer
Agarase	Fermentas GmbH
GoTaq® G2 Flexi DNA Polymerase	Promega
Fast-Link™ DNA Ligase	Epicentre
T4 DNA Ligase	Promega
T4 DNA Ligase	New-England Biolabs
β-agarase I	New-England Biolabs

Table 10 Molecular weight markers, used in agarose gel electrophoresis to size estimation.

Molecular weight marker	Manufacturer
1 kb DNA Extension Ladder	Invitrogen
Gene Ruler 1 kb Plus DNA Ladder	Thermoscientific
GeneRuler DNA Ladder Mix	Thermoscientific

Table 11 List of primers.

Nr.	Primer' name	Sequence (5'- 3')	Lenght (bp)	Source
-	T7	TAATACGACTCACTATA	17	Promega
-	SP6	TATTTAGGTGACACTATAG	19	
Primers for KS-domain of PKS				
KF0001	Ks1up	MGiGARGCiHWiSMiATGGAYCCiCARCAiMG	32	[83]
KF0002	KSdn1	GGRTCiCCiARiSWiGTiCCiGTiCCRTG	28	
KF0003	Ksup	MGNGARGCNNWNSMNATGGAYCCNCARCANMG	32	[83]
KF0004	KSdn	GGRTCNCCNARNSWNGTNCCNGTNCCRTG	28	
KF0019	nKSfor1	TRATYGCNAAYCGNATYTCNYA	22	Previous work[2]
KF0020	nKSrev1	GTRCCRTGNRYYTCRATRTANCC	23	
KF0033	nKSrev2	GTNCCRTGNGCYTCRATRTA	20	
KF0034	nKS for3	GCNYTNGARGAYGCNGGNTA	20	
KF0040	KSupR	MGNGARGCNNWNSMNATGGAYCCNCGNCANMG	32	
Primers for A-domain of NRPS				
KF0025	A2forA	AARDSNGGNGSNGSNTAYBNCC	19	[73]
KF0026	A8revA	CKRWRNCCNCKNANYTTNACYTG	23	
KF0027	A3Fpae	GGWCDACHGGHMANCCHAARGG	22	[84]
KF0028	A7Rpae	GGCAKCCATYTYGCCARGTCNCCKGT	26	
KF0029	A2forAi	AARDSIGGIGSIGSITAYBICC	19	[73]
KF0030	A8revAi	CKRWRICCICKIAIYTTIACYTG	23	
General Primers for 16S rDNA region				
-	fD1	CCGAATTCGTCGACAACAGAGTTTGATCCTGGCTAG	36	[85]
-	rD1	CCCGGGATCCAAGCTTAAGGAGGTGATCCAGCC	34	

---

[2] Primers developed by Dr. Katja Fisch, available in the research group.

The primers were purchased from Eurofins MWG Operin (Ebersberg, Germany). The oligonucleotides were reconstituted in Tris-HCl buffer, diluted in TAE buffer to a concentration of 100 ρmol/µl and stored at -20 °C.

2.1.4 Molecular biological kits and equipment

Molecular biology kits and equipment employed in this work are listed in Tables 12 and 13, respectively.

Table 12 Molecular biological kits.

Molecular biological kit	Manufacturer
pGEM®-T Vector System I	Promega
CopyControl™ HTP Fosmid Library Production Kit with pCC2FOS™ Vector	Epicentre Biotechnologies
Fast-Link™ DNA Ligation Kit	Epicentre Biotechnologies
Wizard® Genomic DNA Purification Kit	Promega
DNA Clean & Concentrator™	Zymo Research Corporation
Zymoclean™ Large Fragment DNA Recovery Kit	Zymo Research Corporation
QIAGEN Genomic-tip 100/G	Qiagen
Fast Gene™ Gel/PCR Extraction Kit	Nippon Genetics

Table 13 Equipment.

Material	Manufacturer
Autoclave Varioklav®	Labortechnik AG
Biometra T3000 Thermocycler	Biometra GmbH
Centrifuge 5415 D	Eppendorf
Centrifuge tubes (15/50 ml)	TPP AG
Corning® 15ml, 50 ml PP Centrifuge Tubes	Corning Incorporated
BioSpectrometer®	Eppendorf
Eppendorf tubes 0.5, 1.5, 2 ml	Eppendorf
Eppendorf µCuvette® G1.0	Eppendorf
Gel chambers Horizon™ 58 and Horizon™ 11.14	Thermoscientif
Heraeus™ Fresco™ 17 Microcentrifuge	Termoscientific
Heraeus™ Multifuge™ X1R	Termoscientific
Inolab pH meter	WTW
Kodak DC290	Kodak GmbH
Laminar Airflow Clean Bench BSB 4A (Hera Safe, Class II)	Heraeus
Milli-Q® Water System	Millipore
Power Pac™ Basic Power Supply	BioRad
Sub-Cell® Model 192 Cell	BioRad
Thermal Mixer (TMix)	Analytik Jena

2.2 Methods

2.2.1 Metagenomic DNA isolation from sponges

To isolate the metagenomic DNA from sponges, the method proposed by the authors Gurgui and Piel was applied [86]. Starting with the sponge sample conserved in 95% ethanol, at -20 ºC, 1 to 30 g of tissue was cut with a scalpel. Under liquid nitrogen and using a prechilled mortar and pestle, the sponge was ground to a fine powder. In some cases, sterile sea sand was used to enhance the grinding.

To access the differences between the quality of metagenomic DNA collected with or without previous concentration of the bacterial cells, an intermediate step was tried. After grounding the sponge to a fine powder, the sample was separated into different 50 ml falcon tubes with artificial sea water. To pellet the sponge cells and other detritus, a centrifugation (300 rpm, 30 min) was carried. The supernatant was then transferred to new tubes and the bacterial cells were pelleted (10,000 rpm, 30 min). To enhance cell lysis, the bacterial pellet was sometimes frozen.

One ml of sponge lysis buffer (Table 7) was added to each 100 mg of sponge tissue (or 10 mL to each bacterial pellet) and the mixture was incubated in a 60 ºC water bath for 20 min, with gentle mixing every 5 min. The sample was extracted at least twice with phenol/chloroform/isoamyl alcohol (24:25:1) and the phases separated by centrifugation (10,000 rpm, 4 ºC, 5 min). The aqueous phase was recovered and then extracted with the same volume of chloroform. The DNA was then precipitated with two volumes of ice cold 100% ethanol and 1/10 volumes of 3M sodium acetate (pH 7). After half an hour at -20 ºC, the precipitate was centrifuged (10 000 rpm, 4 ºC, 30 min) and the supernatant removed. The pellet was washed twice with 70% ethanol (10,000 rpm, 4 ºC, 15 min), air-dried and resuspended in 500 µl of Tris-HCl buffer (10 mM, pH 8.5).

The resulting extracted sponge's metagenomic DNA was used through all the work. It was the basis for metagenomic library production (Section 2.2.2), the template for the screen and amplification of key biosynthetic gene clusters (Section 2.2.4) as well as the template for 16S rRNA gene sequencing and consequent microbiome profiling.

2.2.2 Construction of a metagenomic library

The methodology for the construction of the metagenomic library was based on the optimization proposed by Gurgui and Piel [86] for the protocol of the commercial kit CopyControl™ Fosmid Library (Epicentre). In addition, several modifications were introduced based on the results obtained during this study. We describe next the general procedure. In Chapters 3 and 4, some alternative steps and their respective results are presented.

2.2.2.1 DNA size selection by LMP agarose gel electrophoresis

The success of the fosmid library construction relies on the use of a correct DNA size as the lambda phages are only capable of packaging DNA from 30 to 45 kb, being 36 kb the optimal size. Bearing this in mind, a gel electrophoresis is performed with low melting point (LMP) agarose to select the desired DNA size from the previous isolated metagenomic DNA. In addition, this step also removes contaminants which could inactivate further enzyme reactions.

A 20 cm long and 15 cm wide electrophoresis gel was set in a Sub-Cell® Model 192 (BioRad) chamber. To load 200-300 µl of metagenomic DNA, a larger well was prepared on a normal electrophoresis gel comb by covering the small wells with tape. This large well was flanked by two smaller ones, which were used for loading the DNA size marker. A gel at 1% of peqGOLD Low Melt Agarose was prepared in 1x TAE buffer and poured in the previously prepared electrophoresis chamber. The gel was then cooled for 1 hour to achieve optimal consistence and afterwards the DNA sample and the DNA size marker were uniformly loaded into the gel. First, the gel was resolved at 50-60 V for 10 min to guarantee the transference of the DNA onto to the gel matrix. Secondly, the voltage was lowered to 30-40 V and the gel electrophoresis was carried out overnight.

After the run was completed, a 2-cm slice of gel was removed with a scalpel from both sides of the gel. Each slice contained the whole marker lane and a small portion of the well where the DNA sample was loaded. To minimize UV light exposure to the sponge metagenomic DNA, only these two pieces were stained in a 1 µg/ml ethidium bromide bath, for 20 min. Afterwards, the stained slices were exposed to UV and the position of the top and the bottom of the 40 kb marker was marked with a scalpel. All the gel slices were realigned, and the desired band of DNA was located and excised from the gel.

The cut-out gel slices were placed in 15 ml falcon tubes and molten at 65 °C for 15 min. The mixture was supplemented with 1 µl of GELase enzyme per 300 µl of melted agarose and with pre-warmed buffer to 1x final concentration. The solution was incubated in a 42 ºC water bath for 2 hours. After this time, the enzyme was inactivated for 10 min in a 70 ºC water bath. A centrifugation step (10,000 rpm, 30 min, RT) was performed to pellet any insoluble oligosaccharides and the supernatant, after being transferred to a new falcon, was precipitated for 15 min at RT with 1/10 volumes of 3 M sodium acetate (pH 7.0) and 2.5 volumes of 100% ethanol. The precipitated DNA was pelleted (10,000 rpm, 30 min, RT) and twice washed with ice-cold 70% ethanol (10,000 rpm, 20 min, 4 ºC). The supernatant was removed, the DNA pellet air-dried and resuspended in 55 µl of Tris-HCl buffer.

2.2.2.2 Blunt-ending of size selected DNA

To clone the high molecular DNA fragments into the pCC2FOS™ vector, it is necessary to end repair the genomic DNA because the vector contain blunt ends. This step culminates in blunt-ended, 5'-phosphorylated DNA. The blunt-ending reaction contained the following components:

8 µl	10x end-repair buffer
8 µl	2.5 mM dNTP mix
8 µl	10 mM ATP
x µl	Size selected metagenomic DNA
4 µl	End-repair mix (T4 DNA polymerase and T4 polynucleotide kinase)
Add sterile water to 80 µl of final volume	

The reaction mixture, after the removal of any air bubble, was incubated for 45 min at RT. To inactivate the reaction in progress, 10 mM of EDTA were added and the mixture was put at 70 ºC for 10 min. The metagenomic DNA was then precipitated with 120 µl of H_2O, 20 µl of 3 M sodium acetate (pH 5.2) and 140 µl of isopropanol, at RT for

30 min. After that period, the mixture was spin in a microcentrifuge (top speed, 30 min, 4 ºC) and the supernatant (about 95%) was carefully removed. 500 µl of ice-cold 70% ethanol were added to the pellet and another centrifugation step was carried out (top speed, 10 min, 4 ºC). After total removal of the supernatant, the pellet was air-dried at RT. Finally, 15 µl of Tris-HCl buffer (pH 8.5) was added and the DNA was resuspended on ice for ≈30 min. An analytical 1% agarose gel was performed to assess the DNA concentration and size.

2.2.2.3 Ligation of end-repaired DNA into the pCC2FOS™ vector

In this part of the protocol, it is optimal to have a 10:1 molar ratio of vector to DNA-insert. Depending on the obtained quality of the blunt-ended metagenomic DNA, a single ligation reaction will produce ≈10^3-10^6 clones. Differently from all the other steps of the metagenomic library construction protocol, the used ligase enzyme was different from the one supplied by Epicentre. The T4 DNA ligase from New England Biolabs has proven to be more efficient and was, therefore, considered a safer choice to assure an optimal ligation.

The following components were added to a PCR tube, in the order listed below, and gently mixed by after each addition. The final reaction mixture was incubated at 16 ºC overnight in a thermocycler.

1 µl	10x T4 DNA ligase buffer
1 µl	pCC2FOS™ fosmid (0.5 µg/ µl)
x µl	Blunt ended insert DNA
x µl	T4 DNA ligase (New England Biolabs)
Add sterile water to 10 µl of final volume	

On the next day, T4 DNA ligase was inactivated at 65 ºC for 10 to 15 min and the packaging reaction was immediately performed.

2.2.2.4 Packaging of fosmid clones

On the day before the packaging reaction, a colony of EPI-300™-T1^R *E. coli* was inoculated, from a fresh LB agar plate, in 5 ml of LB broth supplemented with 10 mM $MgSO_4$. The cell culture was grown overnight (30 ºC, 200 rpm). On the day of the packaging, 5 ml of the ON culture were inoculated in 50 ml of LB broth with 10 mM $MgSO_4$. The cells were grown at 37 ºC (200 rpm) until an OD_{600nm} of 0.8 to 1.0 was reached and stored at 4 ºC until being necessary.

For one packaging reaction, 25 µl of thawed MaxPlax™ Lambda Packaging Extracts (Epicentre Biotechnologies) were transferred to a microcentrifuge tube placed on ice and 10 µl of the ligated fosmid DNA were added. The solution was gently mixed by pipetting (with cut tips) and incubated in a 30°C water bath for 90 min. After this time, additional 25 µl of thawed packaging extract were added and the mixture was incubated at 30 ºC for an additional 90 min period. Finally, phage dilution buffer (PDB) was added to a final volume of 1 ml.

Before plating the library, it was necessary to determine the titer of fosmid clones packaged in the phage particles to estimate their number in the library. Serial dilutions into PDB were prepared (1:10, 1:10^2 and 1:$10^{3)}$ and 10 µl of each dilution was added to 100 µl of previously grown EPI-300™-T1^R *E. coli* cells. After an incubation step at

37 ºC for 20 min, the infected cells were spread onto LB plates supplemented with 12.5 µg/ml chloramphenicol. Subsequently to an ON incubation, the number of colonies per plate as counted and the titer of the phage packaging was determined with the following equation (1):

$$(x\ colonies) \times (dilution\ factor) \times 100 = cfu/mL \tag{1}$$

2.2.2.5 Library plating

The packaging reaction (1 ml) was mixed with 10 ml of previously grown EPI-300™-T1^R *E. coli* cells. Next, a calculated volume of LB-Sea Prep Agarose (5 g/l) was added, to obtain ≈1,000 cfus (colony forming units)/ml, as well as 12.5 µg/ml of chloramphenicol. One-ml aliquots were distributed to 2 ml sterile screw-caps, previously put on wet ice. After one hour, the vials were placed at 37 ºC for 16-18h. Afterwards, the grown colonies were briefly vortex, 50% glycerol was added, and the vials mixed by inversion. The metagenomic library was stored at -80 ºC [87].

2.2.3 General workflow for the identification of key biosynthetic genes

The screening for biosynthetic genes of interest was conducted in metagenomic DNA extracted from sponge samples (Section 2.2.1). First, PCR amplifications targeting KS and A domains were performed (Section 2.2.4). After visualization of the PCR products on an electrophoresis gel (Section 2.2.5), selected DNA fragments were purified (Section 2.2.6) and ligated into the pGEM®-T vector (Section 2.2.7). The ligated vector was transformed into α-Select Silver or EPI-300™-T1^R competent *E. coli* cells through chemotransformation by heat-shock (Section 2.2.8) and electroporation (Section 2.2.9), respectively. Positive recombinants, selected through blue/white selection, were picked and a whole cell PCR was performed. After gel electrophoresis visualization and gel purification of the DNA products, their concentration and purity were checked (Section 2.2.10). Finally, products would be sent to an external service (GATC Biotech AG) for Sanger sequencing. The retrieved sequences were then subjected to a bioinformatics analysis (Section 2.3).

2.2.4 Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) is the method of choice for rapid amplification of nucleic acid fragments, being therefore used for the screening of genes of interest. The used primers were described above (Section 2.1.3); for details see the Results section. In terms of reaction conditions, the applied method was adapted from Katja *et al.* [83]. A typical PCR reaction mixture was constituted by the following components:

10 µl	5x Green GoTaq® Reaction Buffer
2.5 µl	10x $MgCl_2$-solution (25 mM)
1 µl	DMSO
1 µl	Primer 1 (100 µM)
1 µl	Primer 2 (100 µM)
1 µl	dNTPs (10 mM)
0.25 µl	GoTaq® DNA Polymerase
2.5 µl	DNA template
Add sterile water to 50 µl of final volume	

When degenerate primers were applied, 2.5 µl of each primer were used. In terms of PCR conditions, the following program was set in the thermocycler:

1. Initial denaturation	94°C	2 min	
2. Denaturation	94°C	30 s	
3. Annealing	50°C	30 s	34 ×
4. Elongation	72°C	60 s	
5. Final elongation	72°C	10 min	
6. Cooling	4°C	hold	

The elongation time was estimated from the length of the target region (about 1 min/kb). Steps 2, 3 and 4 were repeated 34 times.

2.2.4.1 Optimization of PCR Conditions

In some cases, optimization of PCR conditions was necessary. For example, gradient PCRs, with a temperature range of ± 4 ºC from a set value, were performed. These had the purpose of guarantee the specificity of primer hybridization by using the optimal annealing temperature. $MgCl_2$ concentration was also adjusted through testing the effect of different quantities on *Taq* polymerase activity. Due to the DMSO capacity of avoiding the formation of secondary structures in GC-rich templates, this substance was sometimes added to the PCR reaction mixture. PCR reactions were simultaneously performed with undiluted and diluted metagenomic DNA to reduce the possibility of PCR inhibition by samples' components.

2.2.4.2 Whole-cell PCR

This strategy was used to perform PCRs directly from bacterial colonies instead of using isolated DNA. In blue/white selection, after inoculation of pCC2FOS^{TM}-DNA vector transformants in LB/carbenicillin/IPTG/X-Gal plates, positive (white) colonies were picked and directly put in previously prepared PCR mixtures. The initial denaturation step of the PCR was prolonged to 10 min to ensure the complete release of the bacterial DNA, by cell wall and membrane disruption. The rest of the reaction was performed as already described (Section 2.2.4).

2.2.5 Agarose gel electrophoresis

Gel electrophoresis is a standard method to separate, identify and purify DNA accordingly to its size. In this study, gel electrophoresis was carried out essentially as described by Sambrook and Russell [88], to analyse PCR products and extracted genomic DNA. Standard gels were prepared by melting 1% agarose in 1x TAE buffer. However, when DNA fragments of <500 bp were expected, an 3% agarose electrophoresis gel was run for better visualization. Except for the PCR products which were directly loaded on the gel, other samples were previously mixed with 6x loading dye. Molecular weight markers were also applied to the gel to enable DNA size estimation. Small gels were usually run 40 min at 100 V. To reduce possible laboratory contamination with ethidium bromide, gels were only stained after the electrophoretic run. After 3-5 min in ethidium bromide bath, the background staining was removed by passing the gel in a water bath. The detection of DNA bands was performed by exposing the gel to ultraviolet light. Gels were documented by Kodak DC 290 Zoom Digital Camera System

2.2.6 DNA recovery from agarose gels

After electrophoresis gel visualization, DNA bands of interest were extracted from the agarose gel and purified using one of the following commercial extraction kits: DNA Clean & Concentrator™ (Zymo Research), Zymoclean™ Large Fragment DNA Recovery Kit (Zymo Research) and Fast Gene™ Gel/PCR Extraction Kit (Nippon Genetics). These kits were used accordingly with manufacturer's instructions.

2.2.7 Ligation of PCR products

The pGEM®-T Vector System I (Promega) was used to clone PCR products. This kit allows a complete ligation in just one hour, at room temperature. For optimal efficiency, the ligation reaction was normally performed at 16 ºC, ON. A ratio of 3:1 (insert:vector) is optimal. Therefore, the ideal amount of DNA insert was calculated using equation (2):

$$\frac{50\,ng\ vector * kb\ insert}{3.0\,kb\ vector} * \frac{3}{1} = ng\ of\ insert \tag{2}$$

The following ligation mixture was performed to standard ligations:

5 µl	2X rapid ligation buffer for T4 DNA ligase
1 µl	pGEM®-T Vector (50 ng)
x µl	PCR product
1 µl	T4 DNA ligase
Ad sterile water to 10 µl of final volume	

This vector has several advantages, one of them is the suitability for Sanger sequencing using the general primers T7 and SP6.

2.2.8 Chemotransformation of competent cells by heat-shock

For each ligation (10 µl), two tubes of α-Select Silver Competent Cells (Bioline) were thawed on ice. A 2-µl sample of the ligation was immediately pipetted to one tube and the remaining 8 µl to the second tube. After 20 minutes on ice, a heat shock of 42 ºC for 90 seconds was performed stimulate entry of the ligations products. Then, 500 µl of LB broth were added and the cells incubated for 1 hour at 37 ºC, for plasmid expression. In the case of transformation of DNA cloned into pGEM®-T vector, 50 µl of the transformants were plated in a LB/carbenicillin/IPTG/X-Gal plate and the remaining volume on a second plate. After an overnight incubation at 30 ºC, blue/white selection would be performed to select for the recombinant plasmids. When the used vector did not contain this selection mechanism, the transformants were selected in LB plates with the appropriate antibiotic.

2.2.9 Preparation of electrocompetent cells and electroporation

For the electroporation of ligation products into host cells, freshly made electrocompetent cells were used. Both the induction of competence and the electroporation steps were adapted from the method proposed by Tu *et al.* that relies on the increase of transformation efficiency by working at room temperature [89]. For the preparation of electrocompetent cells, a 1-ml overnight culture was diluted in 100 ml of LB medium. The culture was grown at 37 ºC for approximately two hours until an OD_{600nm}= 0.5-0.6 was reached. At this point, a centrifugation step (10,000 rpm, 5 min, RT) was done to pellet all bacteria. The cells were then washed with dH_20, centrifuged and the supernatant was removed. This washing step was repeated at least three times to guarantee a reduced ionic strength of the medium. The cells were resuspended in 300 µl of dH_2O and 300 ng of DNA were added to the prepared cells. The DNA-cell mixture was then transferred into 1 mm-gap cuvette and an electroshock of a voltage between 1.2 kV and 2.5 kV was applied. LB medium (500 µl) was immediately added and the cells recovered after 1-hour incubation at 37 ºC. At the end, the cell suspension was spread onto LB plates containing the appropriate antibiotics.

2.2.10 Determination of DNA concentration and purity

A UV/VIS spectrophotometer (Biophotometer, Eppendorf) was used to determine nucleic acid concentration and to access the quality of DNA solutions. At a wavelength of 260 nm (A_{260}) the absorption of 1.0 corresponds to a concentration of 50 µg/ml of double stranded DNA. As aromatic amino acids of proteins show significant absorption at of 280 nm (A_{280}), the ratio A_{260}/A_{280} is used to determine the purity of the sample. A sample containing 100% of nucleic acids would have a A_{260}/A_{280} ratio of 2.0. To ensure DNA quality, values between 1.8 and 2 are desired. Depending on the methodology involved, the concentration and quality of DNA was also estimated through agarose gel electrophoresis by comparison with an appropriate DNA ladder.

2.3 Software and databases

Several bioinformatics tools were used during this work. Sequencing results and corresponding chromatograms were visualized with the software Snapgene Viewer® (GSL Biotech). Through it, a rapid analysis was performed to remove the nucleotides corresponding to unspecific peaks. Then, VecScreen, an online tool provided by the National Centre for Biotechnology Information (NCBI), was used to remove the vector backbone from the nucleotide sequence. The treated sequences were then submitted to Basic Local Alignment Search Tool (BLAST), also from NCBI, for homology searches.

When primer designing was necessary, Clone Manager® 9 was the software employed. For the deduction of amino acid sequences from the nucleotide sequences was used ExPASy, an online tool of the SIB Bioinformatics Resource Portal, [90]. In addition, some genomes of interest were analysed for BGC identification with the online platform Antibiotics & Secondary Metabolite analysis Shell (antiSMASH) [91].

2.3.1 16S rRNA gene sequencing

Part of this work consisted on the analysis of sponges' bacterial composition by 16S rRNA gene sequencing. Extracted metagenomic DNA (Section 2.2.1) was sent to CeMET GmbH (https://www.cemet.de/en/), which performed all the steps from sample quality control to the report of all bacteria found on the sample. The used NGS technology was Illumina MiSeq and the reads were compared with NCBI database. Taxonomic binning was performed using the lowest common ancestor (LCA) algorithm.

The sequencing results were analysed with the free software MEGAN V6.8.13 (Community Edition) [92]. This tool allowed taxonomic profile comparison between DNA reads of different sponge samples.

3 Results

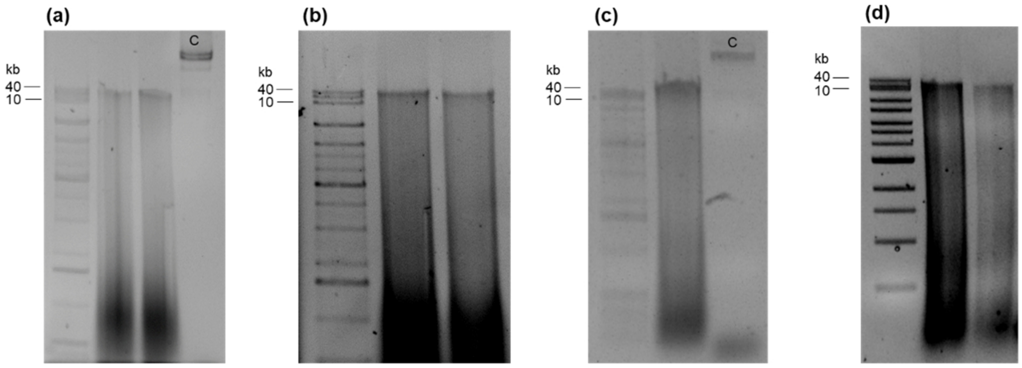
Throughout this work, a total of five sponge samples were analysed. The number of steps successfully performed with each one depended on the workflow and on the obtained results. Table 14 is a summary of all the procedures applied to each sponge sample. Particular emphasis was given to sample Chi15Bu13S $^{L+}$, one of the three samples known to have the compound latrunculin A.

Table 14 Proceedings carried out with each sponge sample during this work.

Sponge sample	Steps successfully achieved					
	Metagenomic DNA isolation	DNA size selection	Metagenomic library construction	PCR screening		16S rRNA sequencing
				KS genes	16S rRNA gene	
Chdi15Bu13S $^{L+}$	✓	✓	✓	✓		✓
Chdi15Bu24+26S^{L+}	✓	✓		✓	✓	
Chan15Bu6S $^{L+}$	✓	✓		✓	✓	✓
Chdi16Sa3-5S $^{L-}$	✓			✓		
Chan15Bu13S $^{L-}$	✓					✓

3.1 Metagenomic library construction

The approaches followed in this work empolyed sponge’s metagenomic DNA. Therefore, the extraction of metagenomic DNA was the first step performed in the analysis of each sponge sample. In sponge sample Chdi15Bu13S^{L+}, for which a metagenomic library was planned, special efforts were made to obtain HMW (>30 kb) DNA. Due to the difficulties in achieving this goal, the procedure was repeated several times. This led to the modification of method’s details in order to achieve optimization for the sample’s specific characteristics. The amount of starting material, the concentration of bacterial cells and the use of sterile sea sand are examples of optimized conditions. Figure 18 presents electropheresis images of several extraction attempts. Metagenomic DNA of panel (c) was selected for the subsequent steps of the protocol.



Starting material	30 g	5 g	8 g	1 g
Applied DNA	1 µl	1 µl	1,5 µl	1 µl
Cell concentration	Yes	Yes	No	No
Sterile sea sand	No	Yes	Yes	Yes
CTAB	No	No	No	No

Figure 18 Different attempts of metagenomic DNA extraction from Chdi15Bu13S^{L+}. Electrophoresis gel images are shown. (C: control, unspecific DNA at ≈50 ng/µl).

In parallel with the application of the described method for metagenomic DNA extraction (Section 2.2.1), the commercial kit Qiagen genomic tip 100/G was tried. Nevertheless, this did not solve the problem in obtaining HMW DNA (Figure 19). In addition, the quality of the extracted DNA proved to be lower due to the considerable amount of fragmented DNA (smear in the gel).

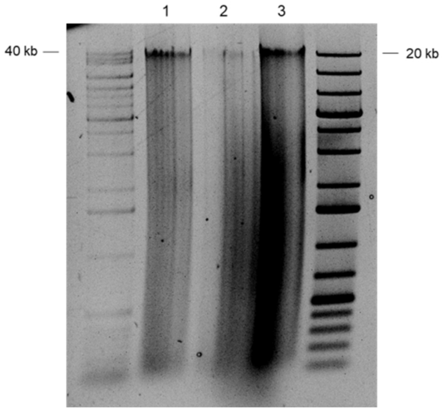


Figure 19 Metagenomic DNA extraction from Chdi15Bu13S^{L+} with Qiagen Genomic Tip. Starting material: 2 g. Electrophoresis gel image is shown.

After the extraction of the metagenomic DNA, the subsequent parts of the protocol were carried with minor adjustments. Figure 20 exhibits a sequence of electrophoresis gels run at the end of each step. To avoid DNA fragmentation, extra care was taken when handling metagenomic DNA. The use of cut tips, low centrifugation rates and the avoidance of repeated pipetting of the DNA, are examples of measures taken. Nevertheless, a decrease in DNA size from each step to the next is observed.

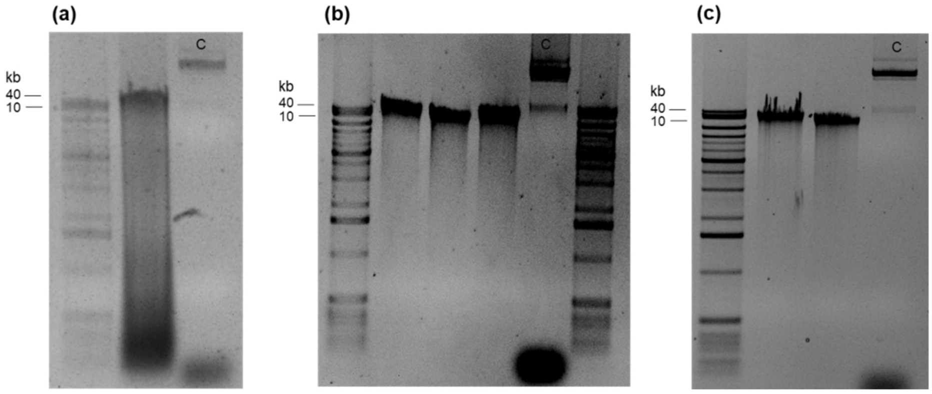
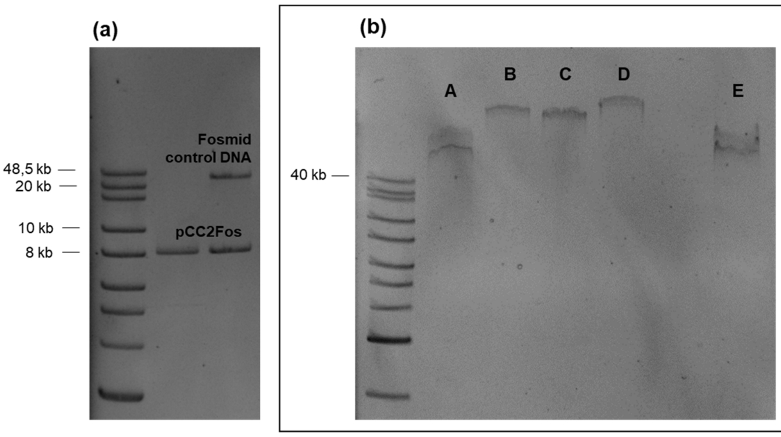


Figure 20 Electrophoresis gels of metagenomic DNA after different steps of Chdi15Bu13S^{L+} library construction. **(a)** After metagenomic DNA extraction. **(b)** After purification and size selection by running on a LMP agarose electrophoresis gel. **(c)** After blunt-ending reaction.

The ligation of the size-selected and blunt-ended metagenomic DNA into the pCC2FOS™ vector was another critical point. Several tries were performed until the reaction's product run in an agarose electrophoresis gel was compatible with the ligation step. Figure 21 (b) presents the last, and only, successful group of ligation reactions. Even though a 10:1 molar ratio (vector:DNA) is described as optimal [86], multiple combinations were simultaneously tried to improve the chances of achieving a positive outcome. Surprisingly, all the 5 combinations had bands compatible with a ligation product. Nevertheless, as they were all slightly different from each other, it was not possible to determine which ones had the optimal ligation.



(c)

Ligation	DNA (µL)	Vector (µL)	Proportion (DNA:Vector)
A	5	0,5	10:1
B	2,5	0,5	5:1
C	3	1	3:1
D	3	0,5	6:1
E	5	1	5:1

[Vector pCC2FOS] 0,5 µg/µl
[Insert DNA] 0,05 µg/µl

Figure 21 Ligation of fosmid control DNA and metagenomic sponge DNA into pCC2FOS™ vector. Reactions were done at 16 ºC for 16 h. **(a)** Agarose electrophoresis gel of the negative control of the ligation (Fosmid control DNA (42 kb) and pCC2FOS™ (8181 bp)). **(b)** Agarose electrophoresis gel of the successful ligation of blunt-ended DNA from Chdi15Bu13S^{L+} in pCC2FOS™ vector. **(c)** Scheme of the ligation reactions presented in (b).

The next step in the metagenomic production workflow is the packaging of the ligated vector into the lambda phage (Section 2.2.4). Two attempts were performed with ligation A and B. The putative phages of the two packaging reactions were used to infect EPI-300™-T1^R *E. coli* cells, with subsequent plating on selection agar plates. However, both attempts failed (no colonies at all for ligation B and two colonies in 1:10 dilution plate for ligation A).

To confirm if the method was being correctly applied, a control DNA was used in the same way as the metagenomic sponge DNA (from the ligation reaction until the inoculation on agar plates). This positive control gave a titer of 1,4x10^7 cfu/ml (Figure 22). Considering the data described in the protocol for CopyControl^{TM} Fosmid Library, which states that a titer of >1x10^7 cfu/ml should be expected, the procedure was being properly conducted.

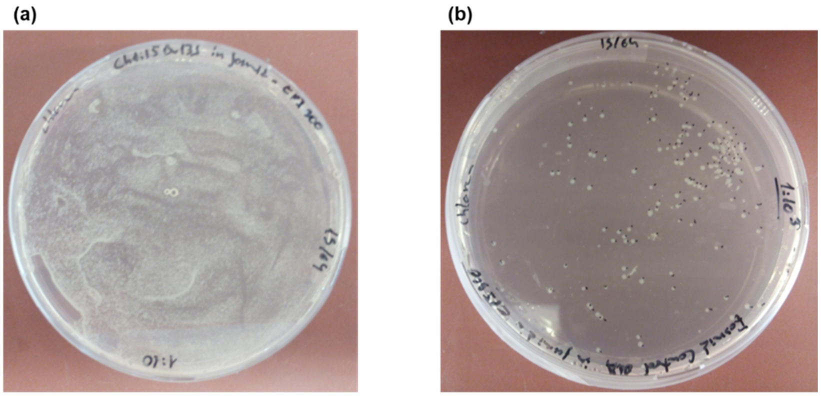


Figure 22 Positive control: EPI-300™-T1^R *E. coli* cells after transduction of pCC2FOS™-fosmid control DNA, packaged in λ-phage.
The cells, plated in LB/chloramphenicol plates, gave a titer of 1,4x10^7 cfu/ml.
(a) 1:10 dilution. **(b)** 1:10^3 dilution.

After the absence of results from the application of the conventional method (lambda packaging) a novel approach was engaged. This consisted on an adaption of the traditional cloning methods for the transformation of bacterial competent cells: electroporation and chemotransformation. These two methods were performed essentially as explained in Sections 2.2.8 and 2.2.9 with ligations D and E (Figure 21). The 10-μl Ligation D was divided in 2- and 8-μl samples that were used to separately transform α-silver *E.coli* cells, through heat-shock. Ligation E, also with 10 μl, was divided in 5 different samples (0.5, 1, 2, 3 and 3.5 μl), which were used in 5 electroporations of EPI-300™-T1^R *E. coli* cells. After the transformation procedure, each mixture was supplemented with 500 μl LB and incubated for one hour to allow cell recovery. After that, 200 μl of each culture were plated in LB/chloramphenicol and incubated ON. On average, the number of colonies resulting from the 7 transformations was around 200 cfu/plate. The transformants resulting from chemotransformation were called α2 and α8, whereas those arising from electroporation were designated E0.5, E1, E2, E3 and E3.5.

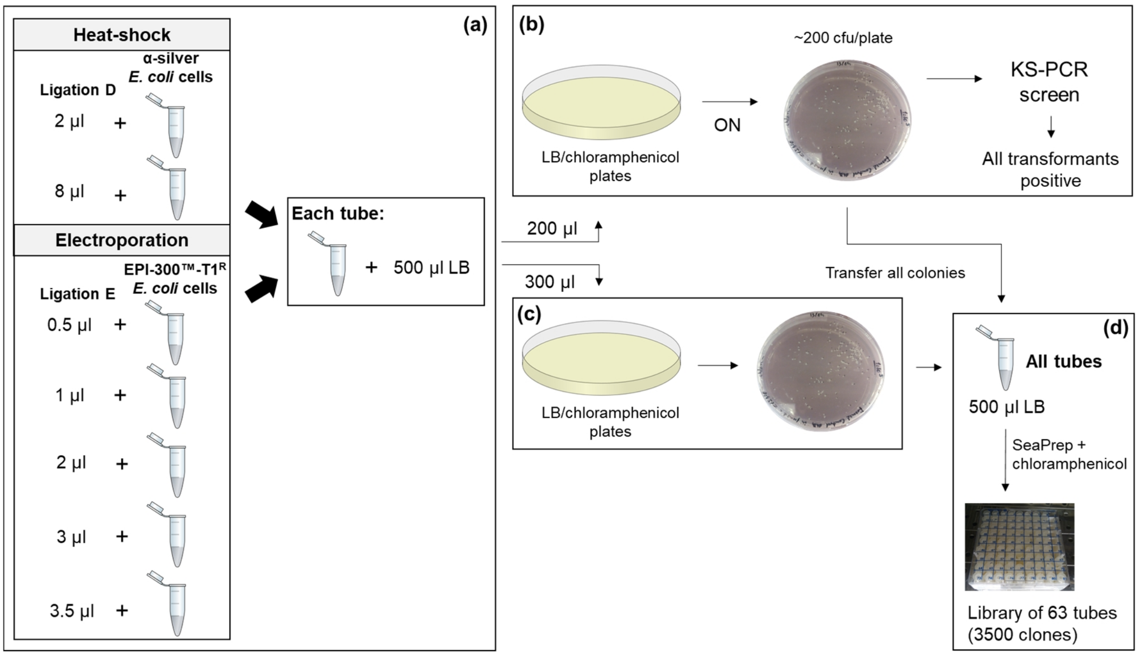


Figure 23 Workflow diagram of all the steps from transformation of host cells with pCC2FOS™-metagenomic DNA to the production of the metagenomic library. **(a)** Transformation of ligation D into α-silver cells by heat shock and electroporation of ligation E in EPI-300™-T1^R cells. **(b)** Transformants were confirmed as having KS genes **(c)** Plating of the rest of the transformed cells **(d)** All the colonies on the LB/chloramphenicol plates were used for the production of the library.

Before proceeding to the construction of the library, a KS PCR was carried to verify if the fosmid's inserts contained sequences of interest. To achieve this goal, whole-cell PCR was conducted in colonies picked from the seven transformation plates. Primers listed in Table 11 for KS domain of PKS were used. Even though it was not possible to obtain PCR products for all the four primer pairs, a clear band, compatible with the use of the KS primers KF0019 and KF0020, was obtained for each group of transformed cells, thus confirming the presence of KS sequences (Figure 24).

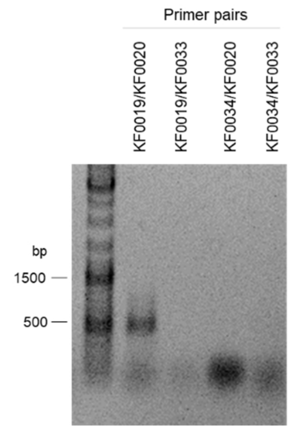


Figure 24 Example of PCR product of the KS-screening of the transformants.
A ≈500 bp band is visible on first lane, corresponding to product of primer pair KF0019/KF0020. The following lanes, with no amplification products, correspond to PCR reactions with primer pairs KF0019/KF0033, KF0034/KF0020 and KF0034/KF0033. Agarose electrophoresis gel is shown.

After confirming the presence of KS sequences in the transformants we proceeded with the production of the metagenomic library. For that, the remaining volume (300 µl) of the transformation mixtures described above was plated in LB/chloramphenicol plates. The total colonies from each transformation, from both the two plating events, were collected to tubes containing 500 µl of LB. To these were added 10 ml of LB/Sea Prep agarose and chloramphenicol to a final concentration of 12.5 µg/ml. The content of each tube was divided in nine 2-ml screw-caps. After overnight incubation, each tube was vortexed and an aliquot of each was removed to produce a pool. A 500-µl volume of 50% glycerol was added to each tube of the library, which was finally stored at -80 ºC. In the end, the metagenomic library was composed of 63 vials and 7 pools (Figure 25). A total of 3500 colonies were collected to produce this library.

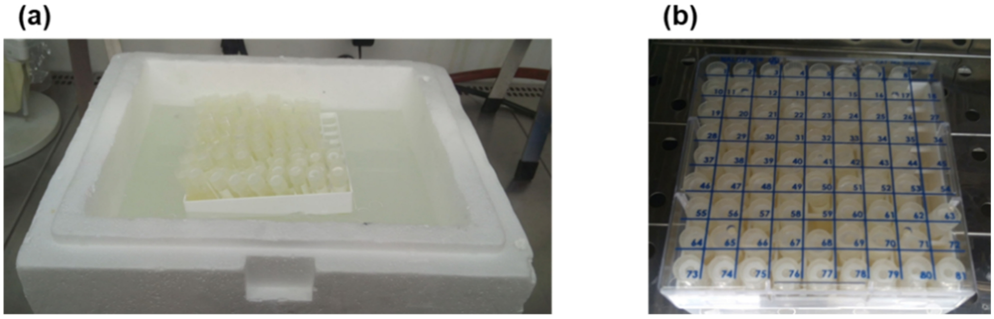


Figure 25 Chdi15Bu13S^{L+} metagenomic library in 3D culture.
(a) 2-ml vials on wet-ice after aliquoted with the transformants in LB/Sea Prep Agarose supplemented with chloramphenicol **(b)** Incubation of the cultures at 37 ºC, in the conservation box.

3.2 Screening for biosynthetic gene clusters (BGCs)

A metagenomic survey of key biosynthetic genes in different sponge samples was performed. This consisted on the use of primers based on KS and A domains since their sequences are the most conserved among PKS and NRPS domains, respectively. After PCR amplification of target regions, PCR products were cloned. Some of the clones were sequenced and analysed by the blastx program.

3.2.1 Polyketide synthases

The identification of PKS genes in the sponge' metagenomic DNA was performed through a nested PCR approach as described by Katja *et al.* [83]. First, general KS primers (KF0001/KF0002 and KF0003/KF0004) originated 700 bp fragments. With these, a second PCR was conducted, resulting in sequences with 480 bp (primers pair KF0019/KF0020 and KF0019/KF0033) and 600 bp (primers pair KF0034/KF0020 and KF0034/KF0033). In Figure 26 it is possible to observe these bands, which have a slightly higher molecular size than the abovementioned due to the presence of vector sequences.

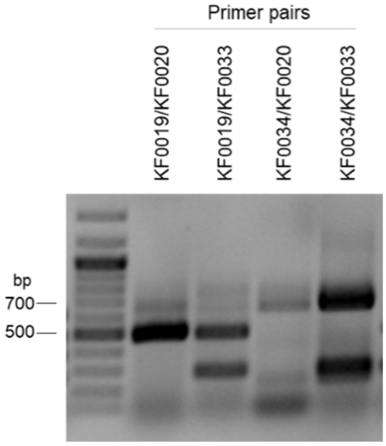


Figure 26 KS PCR screening of Chdi15Bu13S^{L+} metagenomic DNA. Electrophoresis gel image is shown.

This procedure, which was applied in four sponge samples, retrieved a total of 30 different PCR products. After sequenced and analysed in blastx, 16 of them were identified as homologous to PKS genes (Table 15). Some sequences were identified in more than one fragment, providing a perspective in terms of their possible abundance in the metagenomic DNA. Of notice, sequence B.1920.5, with two more identical fragments (B.1920.4 and B.1933.4), and C.2.1933 with one equivalent sequence (C.2.3420).

Table 15 PCR-based KS-screening.

Sponge	PCR product	Size [bp]	Homologue	Identity [%]	Query cover [%]	E value	GeneBank Accession Nr.
Chdi15Bu13S^{L+}	A.3334.3	501	Mixed type I polyketide synthase /nonribosomal peptide [*Candidatus Profftella armatura*]	72	100	5,00E-71	AGS06824.1
	A.1920.2	588	Polyketide synthase [*Streptomyces cattleya*]	66	99	1,00E-55	WP_014144300.1
	A.1920.5	495	Ketosynthase [*Bacillus sp.*]	72	95	1,00E-80	SAJ35041.1
	A.1933.A	348	Polyketide synthase [*Leeia oryzae*]	39	71	4,00E-06	WP_033375960.1
	A.1933.E	331	Polyketide synthase [*Streptomyces himastatinicus*]	47	79	4,00E-09	WP_039937000.1
	A.1933.I	341	Polyketide synthase [*Cylindrospermum stagnale*]	57	38	1,00E-06	WP_015211419.1
	A.2-T7	535	Modular polyketide synthase [uncultured bacterium]	57	83	5,00E-21	AAW84195.1
	A.8.1-T7	529	Ketosynthase [uncultured bacterium]	82	87	7,00E-74	ABK01376.1
Chdi15Bu24+26S^{L+}	B.1920.5	527	Modular polyketide synthase [uncultured bacterium]	98	69	5,00E-27	AAW84195.1
	B.3420.3	687	Ketosynthase [uncultured bacterium]	90	91	5E-116	ABK01341.1
	B.1933.2	277	Ketosynthase, partial [bacterium symbiont of *Plakortis halichondrioides*]	80	94	6,00E-45	KM389572.1
	B.3420.5	626	Modular polyketide synthase ketoacyl synthase domain [symbiont bacterium of *Theonella swinhoei*]	86	98	7E-109	AAQ91806.1
Chan15Bu6S^{L+}	C.2.1933	736	Polyketide synthase, partial [[*Oscillatoria*] sp. PCC 6506]	65	99	1E-102	ACJ46057.1
	C.3.1933	653	Ketosynthase [uncultured bacterium]	95	83	3,00E-61	ABK01328.1
	C.5.1920	670	Polyketide synthase [[*Scytonema hofmanni*] UTEX B 1581]	67	98	4,00E-89	WP_063628697.1
	C.1920.5	700	Modular polyketide synthase [uncultured bacterium]	75	59	7,00E-25	AAW84195.1
Chdi16Sa3-5S^{L-}	D.1920.1.2	483	Acyl transferase domain-containing protein [*Luteibacter sp*. UNCMF366Tsu5.1]	53	90	1,00E-24	SFW64528.1
	D.3334.8	357	Ketosynthase [uncultured bacterium]	58	83	2,00E-28	AGK63336.1

Regarding the sponge from which the metagenomic library was constructed (Chdi15Bu13S^{L+}), six different KS sequences were retrieved with this PCR-based approach. The organisms from which the homologues belonged were from the following phyla: Proteobacteria (two sequences), Actinobacteria (two), Cyanobacteria (one) and Firmicutes (one).

After acknowledging the presence of different PKS gene fragments in the Chdi15Bu13S^{L+} hologenome, next we planned to screen for the whole gene clusters from which each sequence belongs. Using the first six sequences listed in Table 15 as a starting point, six primer pairs were designed with Clone Manager software (Table 16). These will specifically amplify the obtained KS sequences from the clones of the metagenomic library. Consequently, it will be possible to identify the fosmids in which they are included. Hopefully, through the complete sequence of this fosmids it will be possible to the identify the gene clusters from which each sequence belongs. From this point, we are closer to know complete BGCs within the metagenome.

Table 16 Designed primers for BGC screening of the metagenomic library.

Primer' name	Sequence (5'- 3')
A.3334.3-R	CCGTCTTCACTAAGCATTCC
A.3334.3-F	TTTATCGCTACCGCCTCTTC
A.1920.2-F	CGGCGTGTTCTTCTAGCTTG
A.1920.2-R	ACTGCTTCCGCGAATCACC
A.1920.5-F	TAGCGAGCCGATAGAAACC
A.1920.5-R	TACAGCGTCCATCCTTAGC
A.1933.A-F	CACGGATAGCGCAGGAAAG
A.1933.A-R	ATGGGCAGAGGAGGTGTAG
A.1933.E-F	GATTGTGCCGTGAGTTTC
A.1933.E-R	GCGTGTTGTTCTAGCTTG
A.1933.I-F	CACGGGTTGTTCGTCTACCTTG
A.1933.I-R	ATGGGCAGAGGAGGTGTAG

3.2.2 Nonribosomal peptide synthases

A PCR strategy with three different primer pairs for the A-domain (Table 11) was applied in the survey of the sponges' metagenome for NRPS genes. With this approach, a total of 10 different PCR were obtained and sequenced. Unfortunately, none of the 10 sequences revealed significant homology with NRPS genes.

3.3 Sponges' microbial composition

In this section, the microbiome of different sponge samples was analysed. First, a preliminary PCR screening of the sponges' metagenomic DNA was executed. In addition, 16S rRNA gene sequencing was performed in two latrunculin A producing sponges and in one sponge in which this compound was not detected.

3.3.1 16S PCR screening

PCRs with universal 16S primers (Table 11) were performed in metagenomic DNA of sponges Chdi15Bu24+26S^{L+} and Chan15Bu6S^{L+}. A total of six PCR fragments were Sanger sequenced with both forward and reverse primers, retrieving a total of 12 sequences, 10 of them with homologies with sequences in the NCBI database (Table 17).

Table 17 16S ribosomal RNA gene, partial sequences.

Sponge	Fragment	Size [bp]	Closest homologue	Identity [%]	Query cover [%]	E value	GeneBank Acession Nr.	Phylogeny
Chdi15Bu24+26S^{L+}	B.1.16S	1221	*Sphingomonas paucimobilis* strain TY4HX	18	89	5,00E-73	KX055885.1	Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; *Sphingomonas*
	B.16S.8_SP6	1269	Uncultured bacterium clone BA102-C34-seq	90	99	0,00E+00	JX280261.1	Bacteria; environmental samples
	B.16S.8_T7	752	Uncultured marine bacterium clone A7	74	82	6E-125	JX988637.1	Bacteria; environmental samples
Chan15Bu6S^{L+}	C.1.16S_SP6	261	Uncultured delta proteobacterium clone BZ40D8f_f09	93	93	6,00E-116	KF286212.1	Bacteria; Proteobacteria; Deltaproteobacteria; environmental samples
	C.1.16S_T7	352	Uncultured bacterium clone musa167	91	91	2,00E-122	JN023404.1	Bacteria; environmental samples
	C.5.16S_SP6	448	Uncultured bacterium clone TO10919_C25	88	97	0,00E+00	JX206630.1	Bacteria; environmental samples
	C.5.16S_T7	279	Uncultured bacterium clone TO10922_C19	52	88	2,00E-40	JX206654.1	Bacteria; environmental samples
	C.16S.8	751	Uncultured bacterium clone GBc107	56	95	0,00E+00	JQ612241.1	Bacteria; environmental samples
	C.16S.10_SP6	1342	Uncultured bacterium clone Mfav_O03	87	96	0,00E+00	GU118628.1	Bacteria; environmental samples
	C.16S.10_T7	1330	Uncultured bacterium partial	88	95	0,00E+00	HE985097.1	Bacteria; environmental samples

3.3.2 16S rRNA gene sequencing through NGS

Metagenomic DNA samples from Chdi15Bu13S^{L+} and Chan15Bu6S^{L+} (latrunculin A containing sponges) and Chan15Bu13S^{L-} were analysed through 16S rRNA gene sequencing, performed by CeMeT GmbH with NGS Illumina MiSeq system. The lowest ancestor algorithm was used to compare the sequence data with the NCBI taxonomy database. A total of 29894 assigned reads were retrieved for Chdi15Bu13S^{L+}, 21537 for Chan15Bu13S^{L-} and 17133 for Chan15Bu6S^{L+}. MEGAN software was used to graphically exhibit the data (Figure 27).

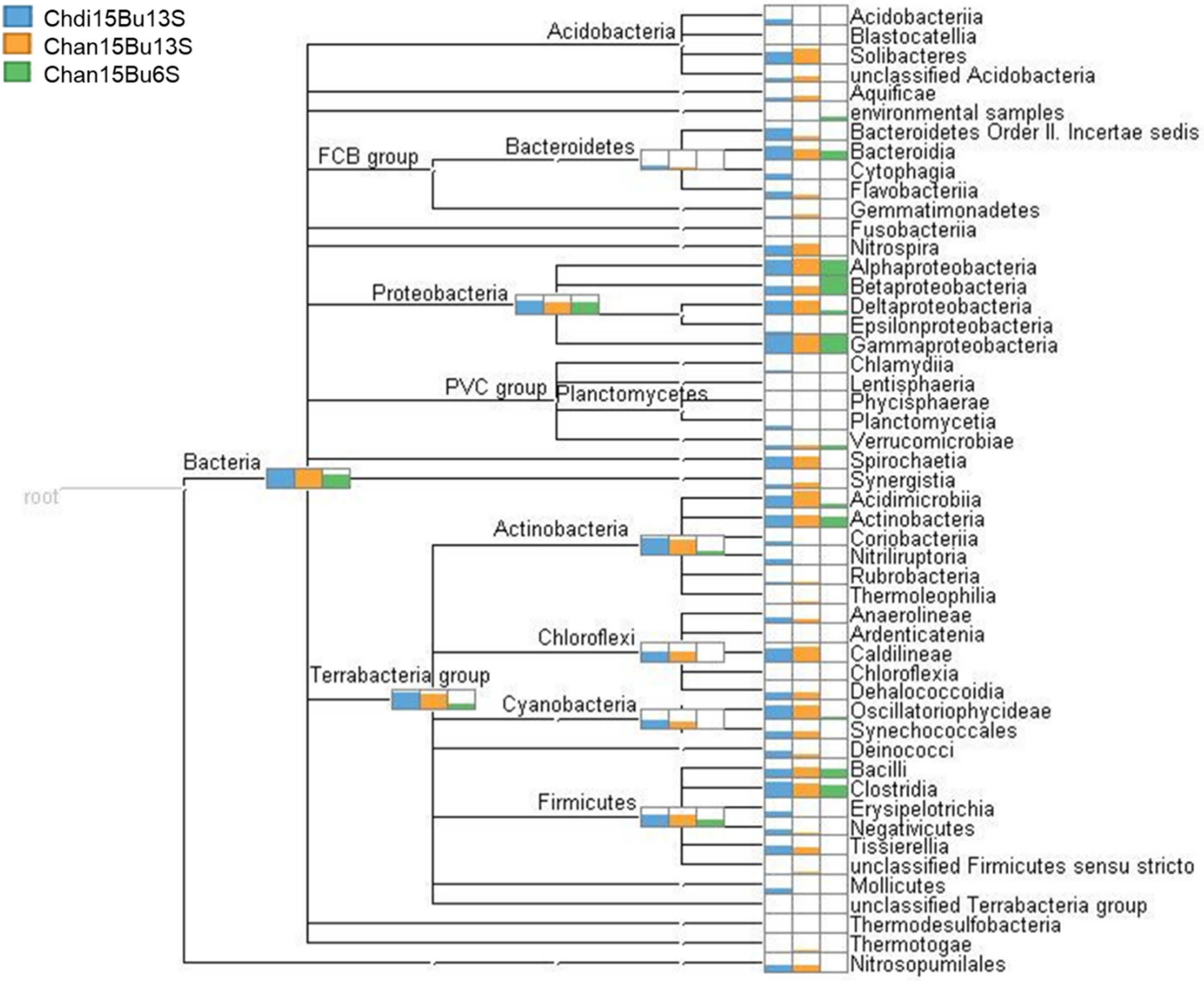


Figure 27 Cladogram of the 16S rRNA taxonomic profiling of Chdi15Bu13S^{L+} (blue), Chan15Bu13S^{L-} (orange) and Chan15Bu6S^{L+} (green) collapsed at 'Class' level. The results are presented in log scale (total of 51,515 assigned reads, normalized to 17,133 per sample).

The higher phylum diversity was found in Chdi15Bu13S^{L+}, with 19 in total. In this sponge, the three most abundant phyla were Proteobacteria (21.69%), Actinobacteria (8.46%) and Firmicutes (6.43%). In Chan15Bu6S^{L+}, Proteobacteria (80.34%), Firmicutes (1.42%) and Actinobacteria (0.63%) were dominant. In Chan15Bu13S^{L-}, they were the Proteobacteria (28.94%), Actinobacteria (10.15%) and Chloroflexi (5.29%) phylum (Figure 27).

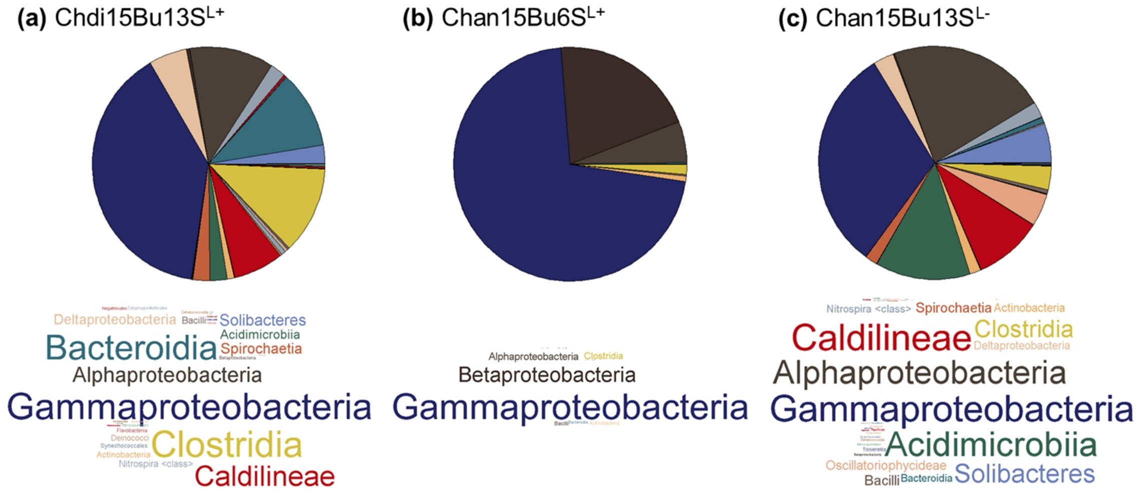


Figure 28 Taxonomy profile, at 'Class' level.
(a) Chdi15Bu13S^{L+}, **(b)** Chan15Bu6S^{L+}, **(c)** Chan15Bu13S^{L-}.

As graphically displayed in Figure 28, at class level, Chdi15Bu13S^{L+} and Chan15Bu13S^{L-} have a similar taxonomical profile. The similarity between them is further demonstrated in Table 18, in which are presented the most abundant bacterial species in each sponge. They both possess *Candidatus Solibacter usitatus* as the most predominant bacteria and *Planktothrix suspense* as the second. Differently from the highly diversified microbiome of Chdi15Bu13S^{L+} and Chan15Bu13S^{L-}, in the sponge Chan15Bu6S^{L+} only four main classes are found. Almost 75% of its reads are assigned to the class Gammaproteobacteria.

Table 18 The first three most predominant bacterial species in each sponge. Relative abundances are shown.

Chdi15Bu13S^{L+}		**Chan15Bu6S^{L+}**		**Chan15Bu13S^{L-}**	
Candidatus Solibacter usitatus	1,17%	*Xylella fastidiosa*	1,63%	*Candidatus Solibacter usitatus*	2,69%
Planktothrix suspensa	1,06%	*Sinobacterium caligoides*	0,68%	*Planktothrix suspensa*	2,06%
Marichromatium purpuratum	0,86%	*Carboxydocella sporoproducens*	0,58%	*Vibrio navarrensis*	0,84%

Three bacterial species were only identified in latrunculin A containing samples: *Coprococcus eutactus*, *Coprobacter fastidiosus* and *Sporobacter termitidis*. The available genome of the closest strains was analysed in the antiSMASH platform. However, no potentially interesting BGCs were retrieved.

4 Discussion

4.1 Metagenomic library construction

The construction of a sponge metagenomic library poses a complex challenge. Even though several groups [74,93] have successfully tackled this issue, the process itself continues to be very laborious and time-consuming [94]. The main refered challenge is the difficulty in recovering high molecular weight (HMW) DNA. One of the main reasons is the fact that every starting sample is different from the ones previously used because of the spatial-temporal heterogeneity of the sponge and associated organisms. Factors like polysaccharide content, the amount of sponge-associated bacteria and the sample's state of conservation make optimization mandatory for each particular case.

It is important to account for the impact on DNA quality of sample collection and preservation, especially when the goal is to produce a metagenomic library. In this case, the sponge samples were collected by SCUBA diving and preserved in 95% ethanol. In an ideal scenario, right after sponge gathering, the bacterial DNA would have been immediately extracted and the metagenomic library prepared. However, this analysis was only performed aproximately six months after collection. It is hypothesized that this could, in part, justify the difficulties on getting HMW DNA.

4.1.1 Metagenomic DNA isolation

As already stated, the main challenge when constructing a metagenomic library is the extraction of HMW DNA (> 30 kb), with good yield and integrity. Considering that each step contributes with more DNA fragmentation, it is crucial to start with HMW DNA [95]. As described in Section 3.1, this was a critical point of this work which required optimization. Thereby, comments on the impact of protocol's conditions are here exposed.

For one library production, a considerable amount of DNA must be initially collected, as the rate of DNA loss in each step is high. Theoretically, from 1 g of sponge tissue it is possible to collect enough DNA for one library production [86]. Nevertheless, due to the difficulty on recovering sufficiently concentrated DNA, higher quantities of initial sample were tried (1 to 30 g). The best results were obtained with quantities ranging from 3 to 8 g and, from this point, higher amounts of sample were not proportional to a procedure improvement. As the complete bacterial cell lysis is a key factor in the DNA isolation, it can be presumed that higher cell densities decrease the efficiency of the sponge lysis buffer. Other strategies were attempted to maximize cell lysis: the freezing of the bacterial pellet and the usage of sterile sea sand on the grinding process. The latter culminated in a more concentrated DNA. That was expected since the fragmentation of the sponge structures helps to free the symbionts, resulting in a better effect of the lysis buffer.

Since the bacterial DNA is the object of the study, it is hypothesized that the separation of the symbionts' genomic material from the sponge DNA could improve subsequent enzymatic reactions and facilitate heterologous expression [93]. Even though PCR primers will only bind to the specific sequences of bacterial DNA, host genomic material interferes with the quality of the process by allowing more unspecific binding.

Nevertheless, bacterial cell concentration through centrifugation had a negative impact on DNA size and, because of this, this approach was dropped.

Gurundi *et al.* proposed the addition of a CTAB step for the removal of the polysaccharide content, usually high in sponge samples [86]. Its presence can lead to incomplete enzymatic reactions, in the following phases of the protocol. However, this step was ignored due to coprecipitation of CTAB with DNA, and its subsequent loss. Additionally, it is presumed that Chdi15Bu13S^{L+} has low polysaccharide content.

Finally, the Qiagen genomic tip kit failure in extracting good quality metagenomic DNA might be caused by the non-specificity of this kind of commercial kits, which are designed to work on a roll of different samples.

4.1.2 pCC2FOS^{TM}-metagenomic DNA vector transformation into host cells

pCC2FOS^{TM} is a *cos*-based vector which allows packaging by lambda phage for subsequent transduction of *E. coli* cells. The correct packaging of the vector into the lambda phage particles can only be confirmed through its transduction to suitable bacterial cells, with the subsequent colony growth in LB/chloramphenicol plates. When no positive results are obtained, it is difficult to assess the cause. For this reason, a control procedure was carried with Fosmid Control DNA (Epicentre), which gave a positive result. This confirms that the method was being properly performed. Hence, the most probable cause for failure in alpha packaging and, consequently, in cell transformation is the reduced size of the metagenomic DNA. This constrains its packaging into lambda phage heads [95]. Ultimately, it can lead to the production of vector-less, non-infectious phage particles.

Using an unconventional approach, based on transformation by heat-shock and electroporation, it was possible to deliver vector pCC2Fos-metagenomic DNA to *E. coli* cells. Metagenomic libraries prepared with the same experimental setup but with properly-functioning lambda packaging are described as being constituted by 80,000 clones [82] or 410,000 clone [83]. Compared with these libraries, this approach originated significantly less clones (≈3500 clones). This can pose an impediment in the future screening of the library for the BGC of interest [72] as the size of a polyketide gene cluster can vary from 10 to 100 kb [96]. Very often, these are separated in distinct fosmids being necessary to identify and sequence several clones [37]. Therefore, a larger metagenomic library is required to improve the probabilities of succeeding in a future screening. Nevertheless, it was decided to proceed with these transformants because the restart of the whole procedure would be time-consuming.

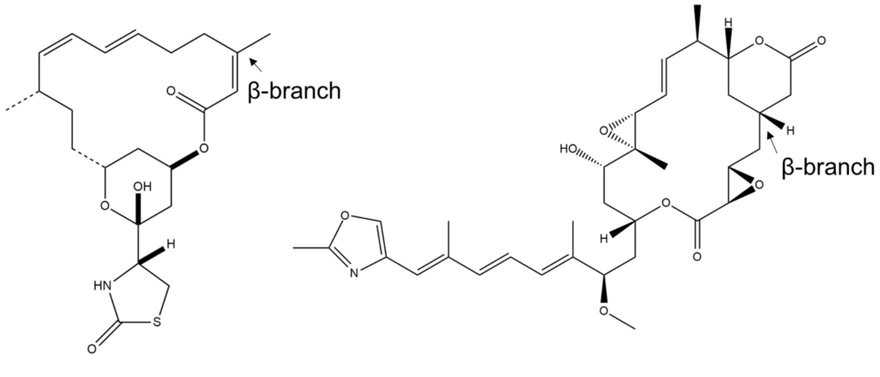
Future screening of the produced metagenomic library will start with primers (Table 11) designed from KS sequences already found in Chi15Bu13S^{L+} sponge (Section 3.2.1). This process can guide to the identification of potential latrunculin A biosynthetic genes as well as to BGCs of known or yet-unidentified polyketide products.

4.2 Screening for biosynthetic gene clusters (BGCs)

While constructing the metagenomic library, it was intended to evaluate the biosynthetic potential of the sponges' samples in study. That was performed using well-described primers for the KS domain of PKSs and for the A domain of NRPSs. Even though it was not possible to obtain NRPS sequences, this PCR strategy confirmed the presence of polyketide synthases' genes in all the sponges screened. In addition, the knowledge of the sequences within sponge Chdi15Bu13S^{L+} allowed the design of specific primers (Table 16) for the later screening of the constructed metagenomic library. Its future use can be specially interesting to uncover the full BGCs from which some KS sequences belong.

It is important to acknowledge the bias associated with PCR-based methods. Even though this approach provides interesting insights into sponge's microbiome as well into some of the BGCs present in the environmental DNA, it only gives a diminish perspective of the sample's full metabolic potential [96].

A structure-based approach was used to target specific features of the compound latrunculin A. Polyketides synthetized by *trans*-AT PKSs normally contain a characteristic carbon branch at the β position, as observed in latrunculin A **(28)** and rhizoxin **(32)** (Figure 29). The primers used in the KS nested PCR are specific for the acyl transferase domain. For example, primer KF0034 targets the EDAGY motif which is conserved in 84% of all *trans*-AT PKSs [83]. In the Supplementary material chapter, where is presented the translation to amino acids of the nucleotide sequences, it is possible to identify this motif, among others.



Latrunculin A **(28)**

Rhizoxin **(32)**

Figure 29 Position of β-branch on the molecular structures of latrunculin A **(28)** and rhizoxin **(32)**.

The homologue of the fragment A.3334.3, with 72% identity, is an interesting sequence for a deeper analysis. *Candidatus Profetella Armatura* is a Betaproteobacteria with a characteristic small genome, 15% of which encoding for a pederin-like polyketide toxin named diaphorine [97]. The indicated identity shows that the sequence retrieved from Chdi15Bu13S^{L+} metagenome probably belongs to a similar, but not the same, gene cluster. Together with the knowledge that diverse pederin-related compounds show potent bioactivities, these are enough points of interest to pursuit the study of this sequence.

Considering than only two sequences were found more than once in the sequenced bacterial clones, a relative high diversity of PKSs seems to exist in this sponge's metagenome. It can be estimated that further KS-PCR screenings would result in more different sequences. This could be an interesting approach for future work.

4.2.1 NRPS-based PCR screening

Due to the hybrid origin of latrunculin A, a different approach, based on PCR screening for the highly conserved A-domain, was attempted. The lack of homologies of the obtained PCR products with NRPS genes can be analysed from separate perspectives. By solely considering these results, it could be stated that NRPS genes are not abundant in the sponges' metagenome and therefore are not easily depicted through the applied PCR strategy. Nevertheless, there is no guarantee that the optimal PCR conditions were achieved for the used primers. If inadequate, this can inhibit the reaction and be responsible for the verified absence of PCR products. Therefore, no significant conclusions can be retrieved from these results.

4.3 Sponges' microbial composition

As already stated in Section 1.3.1, the characterization of the host microbiome is a crucial step towards the full comprehension of these intricately connected systems. In this section, sponges' symbionts will be analysed. Nevertheless, as these approaches are based on metagenomic DNA extracted from sponges' portions, it is necessary to access the associated and inherent bias. Even though it was tried to include representative segments of the sponge (longitudinal cuts, exemplified in Figure 17 (b)), there is no guarantee that it contains representatives of all sponge's microorganisms. Only through more sampling would be possible to discard this bias.

In terms of the analysis of general 16S screening, reduced conclusions could be draw. All the PCR fragments, apart from B.1.16S, were homologous to uncultured bacteria. This is an expectable result because the number of environmental rRNA gene sequences in the databases is considerably higher than the number of sequences from cultivated microorganisms [98].

This was only a preliminary attempt to characterize the sponges' microbiome as it is an inexpensive and fast approach to retrieve data. The following method, 16S rRNA gene NGS sequencing, is the most used technique to phylogenetically characterize a metagenome due to its capacity of revealing most of the microorganisms present on a sample.

4.3.1 16S rRNA gene sequencing through NGS

The main goal of this approach was to verify if it existed a major difference in the microbiome composition of two sponges containing latrunculin A when compared with a sponge lacking the compound. This could provide hints on which bacterial species are more probable to be the compound's producer. However, a similar taxonomic profile was found between Chdi15Bu13S^{L+} and Chan15Bu13S^{L-}, with Chan15Bu6S^{L+} showing a significantly different microbiome. Therefore, from this analysis it is not possible to form strong conclusions on potential latrunculin A producers. Nevertheless, this data is an excellent opportunity to better understand the biotechnological potential of these sponges' symbionts and therefore some points are discussed next.

Gammaproteobacteria was the most predominant class of bacteria in all tree sponge samples. Belonging to the phylum Proteobacteria, this class is known for the synthesis of diversified secondary metabolites and for its wide presence in sponges [30]. Interestingly, in recent years the number of interesting natural products identified from marine proteobacteria has significantly increased. However, it is stated that the full potential of this phylum, particularly in terms of PKS and NRPS genes, is far from being deeply explored [99].

The phylum Actinobacteria is also very predominant in these sponges' microbiomes. This is a highly studied phylum, being its members the source of half of the bioactive secondary metabolites discovered so far [32]. In terms of marine actinomycetes, recent studies have shown significant differences on the produced secondary metabolites, when compared with the ones produced by terrestrial actinomycetes. The strong bioactivities of some discovered compounds shed light to the unexploited potential of this group of marine bacteria [100].

The similarity between Chdi15Bu13S^{L+} and Chan15Bu13S^{L-} is well evident at species level, both having the *Candidatus Solibacter usitatus* as the most predominant bacterial species (Table 17). This is an Acidobacteria of the class Solibacteres with a significant abundance in these two sponges. Interestingly, recent studies have revealed the widespread presence of PKS genes across the Acidobacteria phylum, the most abundant on the planet. However, even though these polyketides are speculated to possess significant antibacterial activities, their identification and characterization are still absent [101]. It is expectable that in the following years more compounds from this phylum will be discovered.

In Chdi15Bu6S^{L-}, the higher number of reads at species level corresponded to *Xylella fastidiosa* (332 reads). This Gammaproteobacteria species is described as a plant pathogen that colonizes the xylem [102]. Therefore, its presence in a sponge microbiome is highly unexpected and this sequencing result must be evaluated. Although 16S rRNA based techniques are highly used for microbiome profiling, they have some associated limitations that can lead to incorrect species identification. 16S rRNA gene sequencing is known to provide genus identification in over 90% of the cases but only 65 to 83% of the times can define bacterial species. These difficulties can be explained by the low numbers of sequences deposited in the databases, the presence of novel taxa and the mistaken classification of species with similar 16S rRNA sequences [103]. In addition, the short length of obtained reads and sequencing errors can contribute to incomplete, or even incorrect, microbiome profiling [54]. This could have been the case of the identification of *Xylella fastidiosa* in this sponge. Second, it can be speculated that contamination of the sample with algae symbionts could have occurred.

The three organisms only identified on latrunculin containing sponges are anaerobic bacteria. At the light of the long scientific believe that anaerobic bacteria are not capable of producing the natural products of interest, these would be considered not worthy of a deeper investigation. However, in recent years, several gene clusters for secondary metabolites like polyketides and NRP have been found in genomes of anaerobic bacteria. Hertweck *et al.* states that considering the limited energy available for anaerobes, the produced compounds must have an important role on their survival [104]. Therefore, it is estimated that, in next years, along with the increase on the number of anaerobic bacteria sequenced, more BGCs for compounds with interesting bioactivities will be found.

An antiSMASH analysis was conducted in the closest available genomes of the organisms presented in Table 20. However, no meaningful results were retrieved. It is important to acknowledge that it is only possible to analyse available genomes of each species and not the specific strains present in the sponge sample. Therefore, significant differences on the BGCs content can be found. For example, while a given sponge's symbiont can possess a BGC, its closely-related strain present in databases may not.

5 Conclusions

5.1 Summary of contributions

The main goal of this work was to provide insights on the biosynthetic pathway of latrunculin A, a potent cytotoxic natural compound found in sponges. To achieve this, a multi-approach analysis of sponge samples containing this compound was planned. This consisted on the production of a metagenomic library of a sponge containing latrunculin A, the screening of the metagenomic DNA of several sponge samples for conserved domains of PKS and NRPS genes and on the bacterial taxonomical profiling of three different sponge samples.

A metagenomic library was produced through transformation of a fosmid vector carrying metagenomic DNA into host cells, following an alternative approach (electroporation and heat-shock). This approach produced significantly less clones than the established method (vector packaging in phage λ particles followed by transduction into host cells). Therefore, it is concluded that despite being possible to transform large vectors into *E. coli* cells through electroporation and heat-shock, it does not produce a sufficiently high number of clones to probabilistically contain the BGC of interest.

Additionally, the KS-PCR screening of the metagenomic DNA confirmed the presence of diverse PKS genes. This approach, combined with taxonomical profiling through 16S rRNA gene sequencing, enlightened the biosynthetic potential of the sponges' symbionts. Comments on the main characteristics of the most predominant bacteria are provided.

5.2 Directions for future work

The work here developed provides a starting point to the search of the biosynthetic gene cluster of latrunculin A through a metagenomic approach. The produced metagenomic library shall be screened for conserved biosynthetic genes. Particularly, KS primers designed from sequences present in the metagenome shall be used.

The described difficulties on producing a large metagenomic DNA library mandate a consideration of different approaches and, therefore, several guidelines are hereafter provided:

- Use of pulse field gel electrophoresis on the extraction of HMW metagenomic DNA for library production. This method can tackle the limitations of standard gel electrophoresis in separating large DNA molecules, thus enhancing the chances of recovering higher molecular weight DNA, with better quality and yield.
- In future expeditions of the group, the metagenomic library should be constructed as soon as possible after sponges' samples collection. In addition, it could be useful to analyse other sample conservation methods, rather than ethanol, to maximize DNA preservation.

- Whole genome sequencing for assembling of the information into contigs. This would provide enough information to verify if parts of the putative latrunculin BGC are encoded in the sponge metagenome.

- Construction of a second metagenomic library with the knowledge gathered in this work for these specific samples. As referred by several research groups, metagenomic library production is most part of the times a try-out process to gather enough information on what the necessary procedure's optimizations for the samples' specific characteristics are.

References

1. Jensen PR. Natural Products and the Gene Cluster Revolution. Trends in Microbiology. 2016;24(12):968–77.
2. Harvey AL, Edrada-Ebel RA, Quinn RJ. The re-emergence of natural products for drug discovery in the genomics era. Nature Reviews Drug Discovery. 2015;14(2):111–29.
3. Newman DJ, Cragg GM. Natural Products as Sources of New Drugs over the 30 Years. Journal of Natural Products. 2012;75(3):311–35.
4. Li JW-H, Vederas JC. Drug Discovery and Natural Products: End of an Era or an Endless Frontier? Science. 2009;325(5937):161–5. A
5. Gerwick WH, Moore BS. Lessons from the past and charting the future of marine natural products drug discovery and chemical biology. Chemistry and Biology. 2012;19:85–98.
6. Newman DJ, Cragg GM. Marine natural products and related compounds in clinical and advanced preclinical trials. Journal of Natural Products. 2004;67(8):1216–38.
7. Mehbub MF, Lei J, Franco C, Zhang W. Marine sponge derived natural products between 2001 and 2010: Trends and opportunities for discovery of bioactives. Marine Drugs. 2014;12(8):4539–77.
8. Hentschel U, Piel J, Degnan SM, Taylor MW. Genomic insights into the marine sponge microbiome. Nature Reviews Microbiology. 2012;10(9):641–54.
9. Kim S-K, editor. Handbook of Marine Biotechnology. Vol. 46, Springer. Springer; 2015. 1512 p.
10. Benayahu TR, Kashman Y. Hemiasterlin and geodiamolide TA; two new cytotoxic peptides from the marine sponge *Hemiasterella minor* (Kirkpatrick). Tetrahedron Letters. 1994;35(25):4453–6.
11. Gamble WR, Durso NA, Fuller RW, Westergaard CK, Johnson TR, Sackett DL, *et al*. Cytotoxic and tubulin-interactive hemiasterlins from *Auletta* sp. and Siphonochalina spp. sponges. Bioorganic & Medicinal Chemistry. 1999 Aug 1;7(8):1611–5.
12. Mooberry SL, Tien G, Hernandez AH, Plubrukarn A, Davidson BS. Laulimalide and isolaulimalide, new paclitaxel-like microtubule-stabilizing agents. Cancer Research. 1999;59(3):653–60.
13. Bergmann W, Feeney RJ. The Isolation of a new Thymine Pentoside from Sponges. Journal of the American Chemical Society. 1950;72(6):2809–10.
14. Sakai R, Kohmoto S, Higa T, Jefford CW, Bernardinelli G. Manzamine B and C, two novel alkaloids from the sponge *Haliclona* sp. Tetrahedron Letters. 1987;28(45):5493–6.
15. Quinoa E, Adamczeski M, Crews P, Bakus GJ. Bengamides, heterocyclic anthelmintics from a Jaspidae marine sponge. The Journal of Organic Chemistry. 1986 Nov 1;51(23):4494–7. 6
16. Groweiss A, Newcomer JJ, O'Keefe BR, Blackman A, Boyd MR. Cytotoxic Metabolites from an Australian Collection of the Sponge Jaspis Species. Journal of Natural Products. 1999;62(12):1691–3.
17. Cichewicz RH, Valeriote FA, Crews P. Psymberin, A Potent Sponge-Derived Cytotoxin from Psammocinia Distantly Related to the Pederin Family. Organic Letters. 2004;6(12):1951–4.
18. Saleem RSZ, Lansdell TA, Tepe JJ. Synthesis and evaluation of debromohymenialdisine-derived Chk2 inhibitors. Bioorganic & Medicinal Chemistry. 2012;20(4):1475–81.
19. Bubb MR, Spector I, Bershadsky AD, Korn ED. Swinholide A is a microfilament disrupting marine toxin that stabilizes actin dimers and severs actin filaments. Vol. 270, Journal of Biological Chemistry. 1995.p.3463–6.

20. Lee K-H, Nishimura S, Matsunaga S, Fusetani N, Horinouchi S, Yoshida M. Inhibition of protein synthesis and activation of stress-activated protein kinases by onnamide A and theopederin B, antitumor marine natural products. Cancer Science. 2005;96(6):357–64.

21. Bartik K, Braekman J-C, Daloze D, Stoller C. Topsentins, new toxic bis-indole alkaloids from the marine sponge *Topsentia genitrix*. Can J Chem. 1987;65:2118–21.

22. Shin J, Seo Y, Cho KW, Rho JR, Sim CJ. New bis(indole) alkaloids of the topsentin class from the sponge *Spongosorites genitrix*. Journal of Natural Products. 1999;62(4):647–9.

23. de Silva ED, Scheuer PJ. Three new sesterterpenoid antibiotics from the marine sponge *Luffariella variabilis* (polejaff). Tetrahedron Letters. 1981;22(33):3147–50.

24. Bai R, Paull KD, Herald CL, Malspeis L, Pettit GR, Hamel E. Halichondrin B and homohalichondrin B, marine natural products binding in the vinca domain of tubulin: Discovery of tubulin-based mechanism of action by analysis of differential cytotoxicity data. Journal of Biological Chemistry. 1991;266(24):15882–9.

25. Kashman Y, Groweiss A, Shmueli U. Latrunculin, a new 2-thiazolidinone macrolide from the marine sponge *Latrunculia magnifica*. Tetrahedron Letters. 1980;21(37):3629–32.

26. Northcote PT, Blunt JW, Munro MHG. Pateamine: a potent cytotoxin from the New Zealand Marine sponge, *Mycale* sp. Tetrahedron Letters. 1991;32(44):6411–4.

27. Matsunaga S, Fusetani N, Konosu S. Amino acid composition of Discodermin A, an antimicrobial peptide, from the marine Sponge D*iscodermia kiiensis*. Journal of Natural Products. 1985;48(2):236–41.

28. Bergmann W, Feeney RJ. The nucleosides of Sponges. The Journal of Organic Chemistry. 1951 Jun 1;16(6):981–7.

29. Sipkema D, Franssen MCR, Osinga R, Tramper J, Wijffels RH. Marine sponges as pharmacy. Marine Biotechnology. 2005;7(3):142–62.

30. Webster NS, Thomas T. The Sponge Hologenome. mBio. 2016;7(2):1–14.

31. Morrow C, Cárdenas P. Proposal for a revised classification of the Demospongiae (Porifera). Frontiers in Zoology. 2015;12(1):7.

32. Kennedy J, Marchesi JR, Dobson ADW. Metagenomic approaches to exploit the biotechnological potential of the microbial consortia of marine sponges. Applied Microbiology and Biotechnology. 2007;75(1):11–20.

33. Schmidt EW. From chemical structure to environmental biosynthetic pathways: Navigating marine invertebrate-bacteria associations. Trends in Biotechnology. 2005;23(9):437–40.

34. Hentschel U, Hopke J, Horn M, Anja B, Wagner M, Hacker J, *et al*. Molecular Evidence for a Uniform Microbial Community in Sponges from Different Oceans. Applied and Environmental Microbiology. 2002;68(9):4431–40.

35. Thomas T, Moitinho-Silva L, Lurgi M, Björk JR, Easson C, Astudillo-García C, *et al*. Diversity, structure and convergent evolution of the global sponge microbiome. Nature Communications. 2016;7(May):11870.

36. Lackner G, Peters EE, Helfrich EJN, Piel J. Insights into the lifestyle of uncultured bacterial natural product factories associated with marine sponges. Proceedings of the National Academy of Sciences of the United States of America. 2017;114(3):E347–56.

37. Amoutzias GD, Chaliotis A, Mossialos D. Discovery Strategies of bioactive compounds synthesized by nonribosomal peptide synthetases and type-I polyketide synthases derived from marine microbiomes. Marine Drugs. 2016;14(4).

38. Challinor VL, Bode HB. Bioactive natural products from novel microbial sources. Annals of the New York Academy of Sciences. 2015;1354(1):82–97.

39. Wilson MC, Piel J. Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. Chemistry and Biology. 2013;20(5):636–47.

40. Dunlap WC, Battershill CN, Liptrot CH, Cobb RE, Bourne DG, Jaspars M, *et al*. Biomedicinals from the phytosymbionts of marine invertebrates: A molecular approach. Methods. 2007;42(4):358–76.

41. Medema MH, Fischbach MA. Computational approaches to natural product discovery. Nature Chemical Biology. 2015;11(9):639–48.

42. Weissman KJ, Leadlay PF. Combinatorial biosynthesis of reduced polyketides. Nature Reviews Microbiology. 2005;3(12):925–36.

43. Weissman KJ. Chapter 1 Introduction to Polyketide Biosynthesis. In: Methods in Enzymology. 2009. p. 3–16.

44. Hertweck C. The biosynthetic logic of polyketide diversity. Angewandte Chemie - International Edition. 2009;48(26):4688–716.

45. Chen H, Du L. Iterative Polyketide Biosynthesis by Modular Polyketide Synthases in Bacteria. 2016;33(4):395–401.

46. Cortes J, Haydock SF, Roberts GA, Bevitt DJ, Leadlay PF. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea. Vol. 348, Nature. 1990. p. 176–8.

47. Gulder T, Michael F, Piel J. The Catalytic Diversity of Multimodular Polyketide Synthases: Natural Product Biosynthesis Beyond Textbook Assembly Rules Tobias. Springer. 2011;11(1):13–35.

48. Moore BS, Hopke JN. Discovery of a New Bacterial Polyketide Biosynthetic Pathway. ChemBioChem. 2001;2(1):35–8.

49. Fischbach MA, Walsh CT. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic machinery, and mechanisms. Chemical Reviews. 2006;106(8):3468–96.

50. Süssmuth RD, Mainz A, Roderich DS, Mainz A, Süssmuth RD, Mainz A. Nonribosomal Peptide Synthesis - Principles and Prospects Reviews. AngewChemInt Ed. 2017;3770–822.

51. Agrawal S, Adholeya A, Deshmukh SK. The pharmacological potential of non-ribosomal peptides from marine sponge and tunicates. Frontiers in Pharmacology. 2016;7.

52. Gerth K, Bedorf N, Höfle G, Irschik H, Reichenbach H. Epothilons A and B: antifungal and cytotoxic compounds from *Sorangium cellulosum* (Myxobacteria). Production, physico-chemical and biological properties. The Journal of antibiotics. 1996;49(6):560–3.

53. Molnár I, Schupp T, Ono M, Zirkle R, Milnamow M, Nowak-Thompson B, *et al*. The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from Sorangium cellulosum So ce90. Chemistry & Biology. 2000;7(2):97–109.

54. Poretsky R, Rodriguez-R LM, Luo C, Tsementzi D, Konstantinidis KT. Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. PLoS ONE. 2014;9(4).

55. Rajendhran J, Gunasekaran P. Microbial phylogeny and diversity: Small subunit ribosomal RNA sequence analysis and beyond. Microbiological Research. 2011;166(2):99–110.

56. Yang B, Wang Y, Qian P-Y. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. BMC Bioinformatics. 2016;17(1):135.

57. Hardoim CCP, Esteves AIS, Pires FR, Gonçalves JMS, Cox CJ, Xavier JR, *et al*. Phylogenetically and Spatially Close Marine Sponges Harbour Divergent Bacterial Communities. PLoS ONE. 2012;7(12).

58. Thomas T, Rusch D, DeMaere MZ, Yung PY, Lewis M, Halpern A, *et al*. Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. The ISME journal. 2010;4(12):1557–67.

59. Taylor MW, Radax R, Steger D, Wagner M. Sponge-Associated Microorganisms: Evolution, Ecology, and Biotechnological Potential. Microbiology and Molecular Biology Reviews. 2007;71(2):295–347.

60. Bentley SD, Chater KF, Cerdeno-Tarraga A-M, Challis GL, Thomson NR, James KD, *et al*. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature. 2002 May;417(6885):141–7.

61. Fisch KM, Schäberle TF. Toolbox for Antibiotics Discovery from Microorganisms. Archiv der Pharmazie. 2016;349:1–9.

62. Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. Journal of Applied Genetics. 2011;52(4):413–35.

63. Howe A, Chain PSG. Challenges and opportunities in understanding microbial communities with metagenome assembly. Frontiers in Microbiology. 2015;6(JUL):10–3.

64. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics. 2016;17(6):333–51.

65. Ziemert N, Alanjary M, Weber T. The evolution of genome mining in microbes – a review. Nat Prod Rep. 2016;33(8):988–1005.

66. Thomas T, Gilbert JA, Meyer F. Metagenomics - a guide from sampling to data analysis. Microbial Informatics and Experimentation. 2012;2(1):3.

67. Handelsman J. Metagenomics: Application of Genomics to Uncultured Microorganisms. Microbiology and Molecular Biology Reviews. 2005;68(4):669–85.

68. Ghurye JS, Cepeda-Espinoza V, Pop M, Cepeda-Espinoza V, Pop M. Metagenomic assembly: Overview, challenges and applications. Yale Journal of Biology and Medicine. 2016;89(3):353–62.

69. Ma J, Prince A, Aagaard KM. Use of whole genome shotgun metagenomics: A practical guide for the microbiome-minded physician scientist. Seminars in Reproductive Medicine. 2014;32(1):5–13.

70. Ranjan R, Rani A, Metwally A, Mcgee HS, Perkins DL. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. Biochem Biophys Res Commun. 2016;469(4):967–977.

71. Singh J, Behal A, Singla N, Joshi A, Birbian N, Singh S, *et al*. Metagenomics: Concept, methodology, ecological inference and recent advances. Biotechnology Journal. 2009;4(4):480–94.

72. Dunlap WC, Jaspars M, Hranueli D, Battershill CN, Peric-concha N, Zucko J, *et al*. New Methods for Medicinal Chemistry – Universal Gene Cloning and Expression Systems for Production of Marine Bioactive Metabolites. Current Medicinal Chemistry. 2006;697–710.

73. Schirmer A, Gadkari R, Reeves CD, Ibrahim F, Delong EF, Richard C, *et al*. Metagenomic Analysis Reveals Diverse Polyketide Synthase Gene Clusters in Microorganisms Associated with the Marine Sponge Discodermia dissoluta. Applied and environmental microbiology. 2005;71(8):4840–9.

74. Piel J, Hui D, Wen G, Butzke D, Platzer M, Fusetani N, *et al*. Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(46):16222–7.

75. Ouyang Y, Dai S, Xie L, Kumar MSR, Sun W, Sun H, *et al*. Isolation of high molecular weight DNA from marine sponge bacteria for BAC library construction. Marine Biotechnology. 2010;12(3):318–25.

76. Pimentel-Elardo SM, Grozdanov L, Proksch S, Hentschel U. Diversity of nonribosomal peptide synthetase genes in the microbial metagenomes of marine sponges. Marine Drugs. 2012;10(6):1192–202.

77. Gomes NGMM, Dasari R, Chandra S, Kiss R, Kornienko A. Marine invertebrate metabolites with anticancer activities: Solutions to the ‘supply problem’. Marine Drugs. 2016;14(5).

78. Zhang H, Zhao Z, Wang H. Cytotoxic natural products from marine sponge-derived microorganisms. Marine Drugs. 2017;15(3).

79. Amagata T, Johnson TA, Cichewicz RH, Tenney K, Susan L, Media J, *et al*. Interrogating the Bioactive Pharmacophore of the Latrunculin Chemotype by Investigating the Metabolites of Two Taxonomically Unrelated Sponges. J Med Chem. 2008;51(22):7234–42.

80. Walker M, Kublin JG, Zunt JR. Semisynthetic Latrunculin Derivatives as Inhibitors of Metastatic Breast Cancer: Biological Evaluations, Preliminary Structure- Activity Relationship, and Molecular Modeling Studies. ChemMedChem. 2009;42(1):115–25.

81. El Sayed KA, Youssef DTA, Marchetti D. Bioactive natural and semisynthetic latrunculins. Journal of Natural Products. 2006;69(2):219–23.

82. Piel J. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus beetles*. Proc Natl Acad Sci U S A. 2002;99(22):14002–7.

83. Fisch KM, Gurgui C, Heycke N, van der Sar S a., Anderson S a., Webb VL, *et al*. Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. Nature Chemical Biology. 2009;5(7):494–501.

84. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S Ribosomal DNA Amplification for Phylogenetic Study. American Society of Microbiology. 1991;173(2):697–703.

85. Wu XC, Qian CD, Fang HH, Wen YP, Zhou JY, Zhan ZJ, *et al*. Paenimacrolidin, a novel macrolide antibiotic from *Paenibacillus* sp. F6-B70 active against methicillin-resistant Staphylococcus aureus. Microbial Biotechnology. 2011;4(4):491–502.

86. Gurgui C, Piel J. Metagenomic Approaches to Identify and Isolate Bioactive Natural Products from Microbiota of Marine Sponges. In: Metagenomics Methods and Protocols. Springer Protocols; 2010.

87. Hrvatin S, Piel J. Rapid isolation of rare clones from highly complex DNA libraries by PCR analysis of liquid gel pools. Journal of Microbiological Methods. 2007;68(2):434–6.

88. Sambrook J, Russel, D W. Molecular Cloning: A Laboratory Manual. Vol. 3, Cold Spring Harboc Laboratory Press. 2000. 999 p.

89. Tu Q, Yin J, Fu J, Herrmann J, Li Y, Yin Y, *et al*. Room temperature electrocompetent bacterial cells improve DNA transformation and recombineering efficiency. Scientific reports. 2016;6(24648).

90. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, De Castro E, *et al*. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Research. 2012;40(W1):597–603.

91. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, *et al*. AntiSMASH 3.0 - A comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Research. 2015;43:237–43.

92. Huson DH, Auch A, Qi J, Schuster SC. MEGAN analysis of metagenome data. Gennome Res. 2007;17:377–86.

93. Abe T, Sahin FP, Akiyama K, Naito T, Kishigami M, Miyamoto K, *et al*. Construction of a Metagenomic Library for the Marine Sponge *Halichondria okadai*. Bioscience, Biotechnology, and Biochemistry. 2012;76(4):633–9.

94. Lam KN, Cheng J, Engel K, Neufeld JD, Charles TC. Current and future resources for functional metagenomics. Frontiers in Microbiology. 2015;6(OCT):1–8.

95. Parks R, Graham F. A helper-dependent system for adenovirus vector production helps define a lower limit for efficient DNA packaging. Journal of Virology. 1997;71(4):3293–8.

96. Donsbach K, Rück-Braun K. Combinatorial biosynthesis of polyketides. Organic Synthesis Highlights IV. 2008;3(December):343–9.

97. Piel J, Hui D, Wen G, Butzke D, Platzer M, Fusetani N, *et al*. Defensive bacteriome symbiont with a drastically reduced genome. Applied and Environmental Microbiology. 2013;20(1):636–47.

98. Frank A, Gl O, Yilmaz P, Gerken J, Beccati A, Ciuprina A, *et al*. 25 years of serving the community with ribosomal RNA gene reference databases and tools. Journal of Biotechnology. 2017;(June):0–1.

99. Timmermans ML, Paudel YP, Ross AC. Investigating the Biosynthesis of Natural Products from Marine Proteobacteria: A Survey of Molecules and Strategies. Marine Drugs. 2017;15(8):235.

100. Lam KS. Discovery of novel metabolites from marine actinomycetes. Current Opinion in Microbiology. 2006;9(3):245–51.

101. Kielak AM, Barreto CC, Kowalchuk GA, van Veen JA, Kuramae EE. The ecology of Acidobacteria: Moving beyond genes and genomes. Frontiers in Microbiology. 2016;7(May):1–16.

102. Baldi P, La Porta N. Xylella fastidiosa: Host Range and Advance in Molecular Identification Techniques. Frontiers in Plant Science. 2017;8(June).

103. Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. Journal of Clinical Microbiology. 2007;45(9):2761–4.

104. Letzel A-C, Pidot SJ, Hertweck C. A genomic approach to the cryptic secondary metabolome of the anaerobic world. Nat Prod Rep. 2013;30(3):392–428.

Supplementary Material

A1. Conserved motifs targeted by KS-PCR primers

Primer	Target motif
KF000**1**/KF000**3**	MDPQQR
KF000**2**/KF000**4**	HGTGTSLGDP
KF00**19**	IANRISY
KF00**20**	GYIEVHGT
KF00**33**	YIEAHGT
KF00**34**	EDAGY

A2. Translated protein sequences of KS-PCR products

>A.3334.3
ALEDAGYSPSSLWGSKTALFIATASSGYADQVTKYTDDTDGYSSTGTVPSIGPNRISYYL
NLHGPSEPIETACSSSLVAIHRGVSAIRQGEAEQAIVGGINLIVSPETHVSFTKAGMLSED
GRCKTFSSQANGYVRGEGVGMLYLKPLDSARRDKDTIHGVILGSAE

>A.1920.2
SAVGFNVASELVNTACSSSLVAIHRAINSLHNRESDVAVAGGVSMMLTPKTMLDASQLS
MLSPEGRCKTLDRDANGFVKGEGIGAVVLKPLVQALADGDAVYGVIRGSSVNHGGRAN
SLTAPNPKAQAALVAQAHRLAGFAPETIGYIEAHGT

>A.1920.5
PNRMSFFLNIHGPSEPIETACSSSLVAIHHAILAMNSNGCEMAIAGGVHAIVTPVAHISFSK
AKALAKDGRCKTFSPWADGFAIAEGVGMLILKRLSHAERSGDHIYGIIRGTAVNHGGRAN
SPTAPNPVAQADLLKAAYINAGVDPRTVGYIETHGT

>A.1933.A
MIANRISYLFDFHGPSAAVDTDSAGKNSGAHGAVSRLTRLERTAGLFSTPPLPMTSIFKM
LARGAHMMETETKCSPFIAPGNLHVFCSCVHSCVFRIHGHASTEKKQVALNPLRM

>A.1933.E
SDVAGPGGVSMLWRDTTMLDACHKSMLSPEGRCKTLARESRVLVEGGVAGEVVEKPL
VQASRIYRNSRHNRIX

>A.1933.I
CAVGFDIASELVNTACSSSLVAIHRAINTLHNRESDVAVLGGESMSGGRNPSWMHASTS
CFLQKVVARRLTGCGGCYDRGNRRQGRRTTRGAIIHWIYRNPRHNRSTPAAMAAA

>B.1920.4
FANRISYTFNLTGPSVTIDTACSSSLTAVDTACRELAAETVDAAFAGGANAILLPESYIEFS
RASMLSVSGRCHAFDENADGFVRAEGGALVLLKRLSDALADGDRIHATILATSTNQDGH
TASLMTPSVNSQKAMMHNALQRGEVAPKEIGYIETHGT

>B.1933.2
MRNRQSTTSATMVHSRTPYDSPDSVTTRFRLAQTLQNNDSATFASYIAVCRSIKTLTSR
RTV

>B.1933.4
IANRISYTFNLTGPSVTIDTACSSSLTAVDTACRELAAETVDAAFAGGANAILLPESYIEFS
RASMLSVSGRCHAFDENADGFVRAEGGALVLLKRLSDALADGDRIHATILATSTNQDGH
TASLMTPSVNSQKAMMHNALQRGEVAPKEIGYIEAHGT

```
>B.3420.3
CLGRRGLDPDRLKGSRTGVYTGISNDEYRMLVVDSSKPPEAAGSLYALSGTNLNGTSG
RVSFVLGLMGPAKAVDAACASAMVAVDDAVADLQQGKADLAIAGGVQAILNGRIYELRA
EAMMLSPDGQCKTFDASANGYVRGEGCGVVVLKRLSEAEADGDRIWTVIRGSAVNNG
GTSVGLTVPHTPALVQVMEAALSDAGVVPSEVGYIEIHGTNPAAMATGAVNSAHPX

>B.3420.5
XLDPDRLKGSRTGVYTGISNDEYRMLVVDSSKPPEAAGSLYALSGTNLNGTSGRISFVL
GLVGPAKAVDAACASAMVAVDDAVADLQQGKADLAIAGGVQAILNGRIYELRAEAMMLS
PDGQCKTFDASANGYVRGEGCGVVVLKRLSEAEADGDQIWTVIRGSAVNNGGTSVGLT
VPHTPALVQVMEAALSDAGVVPSEVGYIEAHGT

>C.2.1933
EAQAMDPQQRFFLECAWEALEDAGHLPDACNGRIGLYAGMAISTYLFHLISTGYLDLNN
FIHDSSDLLRVLIGNSKDQLATRASYKLNLTGPSISVQTACSTSLVAVTLACQGLLDY
STOP
DMALAGGVSIRLPQTVGYHYEEDGILSRDGHCRPFDVNAQGTVFGNGLGIVVLKRLEDA
LADYDHIYAVIKGFAVNNDGKEKVGYTAPSVNGQARVVAEALAMADVPPESISYVEAHG
TGTTLGDP

>C.2.3420
EAEDMDPQQRFFLECAWEALEDAGHLPDACNGRIGLYAGMAISTYLFHLISTGYLDLNN
FIHDSSDLLRVLIGNSKDQLATRASYKLNLTGPSISVQTACSTSLVAVTLACQGLLDYQCD
MALAGGVSIRLPQTVGYHYEEDGILSRDGHCRPFDVNAQGTVFGNGLGIVVLKRLEDAL
ADYDHIYAVIKGFAVNNDGKEKVGYTAPSVNGQARVIAEALAMADVPPESISYVEAHGTG
TSLGDP

>C.3.1933
GRGVDPDRLKGSRTGVYTGISNDEYRMLVVDSSKPPEAAGSLYAISGTNLNGTSGRVSF
VLGLMGPAKAVDAACASAMVAVDDAVADLQQGKADLAIAGGVQAILNGRIYELRAEAMM
LSPDGQCKTFDARRTDIAGRRFWGGCPQEVERRGGGWRLRYGSSSGAMAEATVGSA
SVGGSRTPPRWCR

>C.5.1920
RGIALEDAGYLPDACNGRIGLYAGMTISTYLFHLISTGYLDLNNFIHDSSDLLRVLIGNSKD
QLATRASYKLNLTGPSISVQTACSTSLVAVTLACQGLLDYQCDMALAGGVSIRLPQTVGY
HYEEDGILSRDGHCRPFDVNAQGTVFGNGLGIVVLKRLEDALADYDHIYAVIKGFAVNND
GKEKVGYTAPSVNGQARVIAEALAMADVPPESISYIEAHGT

>C.1920.5
IANRISHTFNLTGPSVTIDTACSSSLTAVDTACRKLAAETVDAAFAGGANAILLPESYIEFS
RASMLSVSGRCHAFDENADGFVRAEGGALVLLKRLSDALADGDRIHATILATSTNQDGH
TASLMTPSVNSQKAMMHNALQRGEVAPKEIGYIETHGT

>D.1920-1.2
HGKHKNTFGATFMVADVFKCRTFDERADGYVGGEGYGAIFIKPLAMAEVDGNQIYAIVK
ATAENHVGKATMLTAP

>D.3334-8
XLRLMTQGIIGGVYHRQSAPVFIRLRQGYYIWYFNGDELHSLELMGGSPSILASLISYYLN
LQGPAVATDTPCSSSLVPIHQACHSIQSRESEMVIAGGVYVMTTAQMYIMTL
```