
**University College Cork,**

**National University of Ireland,**

**Department of Microbiology**

# The biotechnological potential of deep sea sponges and their associated microbiome

**Erik Borchert**

**M.Sc.**

**114222423**

**Thesis submitted for the degree of**

**Doctor of Philosophy**

**May 2017**

**Supervisors: Professor Alan D. W. Dobson**

**Professor Fergal O'Gara**

**Head of School: Professor Gerald F. Fitzgerald**

**Declaration of independence**

I hereby declare that this thesis and the work presented within is my own work and has not been submitted for another degree at Univeristy College Cork or elsewhere.

_____

Erik Borchert

*This thesis is dedicated to my parents, my grandmother and my girlfriend.*

# Acknowledgments

I would like to express my gratefulness for my main supervisor Professor Alan Dobson and his crucial role in making this thesis happen. His constant support, challenging mindset and easy to approach attitude have contributed significantly to this work. To the same extent I am very thankful for the support from Doctor Stephen Jackson, who was supervising me on a day to day basis in the laboratory and always willing to review my scientific texts and presentations. Furthermore I would like to thank my second supervisor Professor Fergal O'Gara and Doctor Jonathan Kennedy for their support.

This thesis was conducted as part of the European Union funded Marie-Skłodowska Curie initial training network BluePharmTrain. The BluePharmTrain consortium contributed greatly to this thesis in many different aspects, not only scientifically, but also interpersonally. The semi-annual meetings provided great support, with enlightening conversations, inspiring exchange of ideas, as well as many good memories of excellent companionship and helped to not loose track or feel lost during the duration of the PhD. Therefore I would like to thank everyone involved in this consortium from the different principal investigators, to the Postdocs and the PhD students, for the great time we had.

Furthermore I would like to express my gratitude to my lab and office neighbors Doctor Lekha Margassery, Doctor Lynn Naughton, Jamie Fitzgerald, Eduardo Leão and Beatrice Pulido. All of them have been a great help in making this possible and each and every one of the aforementioned contributed in many different ways. Additionally I would like to thank my collaborators Doctor Ragnar Jóhansson, Doctor Viggó Marteinsson, Doctor Joseph Selvin, Doctor Seghal Kiran, Emilie Dwyer, Sinéad Flynn and Stephen Knobloch for their various contributions to different manuscripts that are part of this thesis.

Finally I would like to thank my whole family for their support throughout the duration of this thesis, the visits and all the food packages were well appreciated. Especially I would like to express my sincere happiness about having my beloved girlfriend Sandra by my side, who was willing to keep up with me for the duration of this thesis nonetheless shared time was extremely scarce. She was always supportive and encouraging, as well as willing to listen to all my complaints and was able to make me smile at any given time.

# Abstract

Shallow water sponges are known to be a prolific source of bioactive compounds and interesting enzymes. In particular shallow water sponges from temperate and warm environments have been investigated in the last couple of decades due to their easy accessibility. Sponges have been shown to harbour dense microbial communities, which were subsequently identified to be the source of most of the isolated bioactive compounds and enzymes. Marine sponges are widespread in our oceans, the biggest interconnected habitat on our whole planet. Sponges can be found not only in shallow water regions but also in the deep sea. The deep sea, comprising approximately anything deeper than 200 m with respect to sea level, makes up an immense area of the oceans, keeping in mind that the mean depth of the oceans is 3800 m. The biodiversity of the deep sea is hard to assess as 95% of the oceans are hypothesized to be unexplored, in this respect it is interesting to note that we have send more people to the moon than to the Marianas Trench, the deepest part of the oceans. Nonetheless already a few deep sea studies have changed our perception from a supposedly very hostile living environment, due to the huge pressure, low temperature and absence of sunlight to a treasure trove of to date largely unexplored marine life, especially with hot spots for living beings and biological diversity like hydrothermal vents, sponge and coral gardens. The marine life in the deep sea has in millions of years adapted to the aforementioned conditions and is therefore believed to be considerably different from other environments, therefore novel or considerably different chemistry particularly with respect to small molecules and novel modes of action for enzymes of industrial interest and antimicrobial compounds can be expected.

The study presented here aims to provide a better understanding of the microbiota of deep sea sponges via applying different next generation sequencing approaches (MiSeq, PacBio, 454 pyrosequencing) as well as standard marine cultivation methods and various enzyme activity assays.

In chapter two the metagenomes of three different deep sea sponge species (*Inflatella pellicula*, *Stelletta normani* and *Poecillastra compressa*) have been investigated for their potential to encode conserved domains of polyketide synthases and non-ribosomal peptide synthetase clusters. These clusters are involved in the production of secondary metabolites that are

beneficial to the sponges as defence mechanism, but could also be used in the pharmacological industry as novel drug leads for example as anticancer or antimicrobial medicines. A huge number of potentially novel adenylation and especially ketosynthase domains were oberverd in the metagenome of the investigated sponge species. Sequence similarities to domains from gene clusters known to be involved in the production of different classes of antibiotics and other bioactive compounds including lipopeptides, glycopeptides, macrolides and hepatotoxins have been identified.

The next chapter studies a common marine microbial isolate that can be retrieved from various marine sources. The *Pseudoalteromonas* spp. isolates described herein have all been isolated from deep sea sponges (*Inflatella pellicula*, *Sericolophus hawaiicus* and *Poecillastra compressa*). The isolates were studied with respect to their biotechnological potential, with a particular focus on their enzymatic activity profiles and their potential for cold adaptation. Furthermore the whole genomes of these isolates and two reference strains were compared with a particular focus on genes potentially involved in symbiosis and secondary metabolism. The isolated *Pseudoalteromonas* spp. were shown to be cold adapted and to express various enzymatic activities, with only one activity being truly cold active. The genome comparison revealed an open pan-genome for all investigated isolates, but no enrichment in symbiosis related genes in the sponge isolates was observed. Nonetheless all the isolates harboured a highly conserved bacteriocin gene cluster with a tetratricopeptide repeat domain, which can be involved in host-association.

Chapter four describes the screening and characterization of a novel cold-active esterase found via a function-based screening of a metagenomic fosmid library of the deep sea sponge *Stelletta normani*. Besides the enzyme defining activity parameters, the esterase was compared to other lipolytic enzymes and *in situ* docking studies were performed. The newly described esterase is part of the type IV hormone sensitive lipase family and is to the best of our knowledge the first truly cold active esterase of this family. The esterase is most active at alkaline pH, mimicking seawater conditions and displays a wide range of halotolerance; coupled with its cold activity this enzyme is potentially desirable for industrial applications in bioremediation and production of biodiesel.

# Contents

## 3. Biotechnological potential of cold adapted *Pseudoalteromonas* spp. isolated from 'deep sea' sponges



## 4. Characterization of a novel cold active deep sea esterase from a metagenomic library from the sponge *Stelletta normani*

# 1. Introduction

## 1.1 Marine Sponges

Sponges are one of the oldest extant metazoans having branched off from other metazoans at least 640 million years ago (Yin *et al.*, 2015). The phylum *Porifera* consist of approximately 8500 species, with over 80% belonging to the main group of *Demospongiae* (common sponges), the three other main orders are *Calcarea* (calcareous sponges), *Hexactinellida* (glass sponges) and *Homoscleromorpha* (encrusting sponges) (Van Soest *et al.*, 2012; Gazave *et al.*, 2012; Morrow and Cárdenas, 2015). Sponges are either effective filter feeders (Bell, 2008; Hentschel *et al.*, 2012) or carnivorous (Vacelet and Boury-Esnault, 1995). The sessile filter-feeder sponges have important functional roles in their ecosystem, usually categorized into three areas, first being impacts on substrate (bioerosion, reef creation, substrate stabilisation), second, bentho-pelagic coupling (carbon, silicon and nitrogen cycling and oxygen depletion) and third, associations with other organisms (sponges as settlement substrate, microhabitat, as releasers of chemicals) (Bell, 2008) (Figure 1). Generally their lifestyles can be divided into two different stages, one being motile as a larvae and then after settling onto a surface becoming sessile as adults (Ayling, 1980).

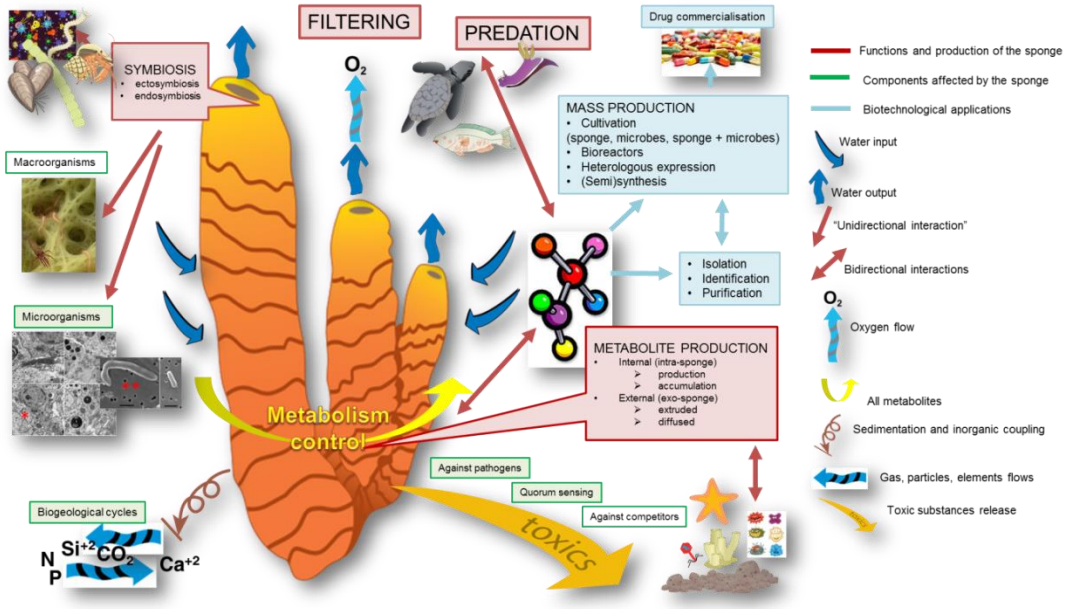



**Figure 1**: Scheme of the roles of filter-feeding sponges in the ecosystem (Steinert *et al.*, 2017).

### 1.1.1 Body plan of sponges

Filter-feeding sponges have a rather simplistic body plan and can be found in various shapes and colours. The outer wall (pinacoderm) is scattered with small pores (ostia), which draw in surrounding seawater for filtering and subsequently ejected from a large opening (osculum) (Hentschel *et al.*, 2012). The directed water flow is generated by the movement of flagella from specialized cells (choanocytes), the water is channelled through aquiferous canals into chambers (choanocyte chamber) where it is filtered and the nearly sterile water is then expelled through the exhalant opening (osculum) (Reiswig, 1971; Wehrl *et al.*, 2007). Sponges show impressive pumping capacities of thousands of litres of water per kilogram of sponge per day (Bell, 2008). Once the water has entered the sponge it is filtered in the choanocyte chambers and particulate matter and microorganisms are taken up by the amoebocyte cells (Figure 2).

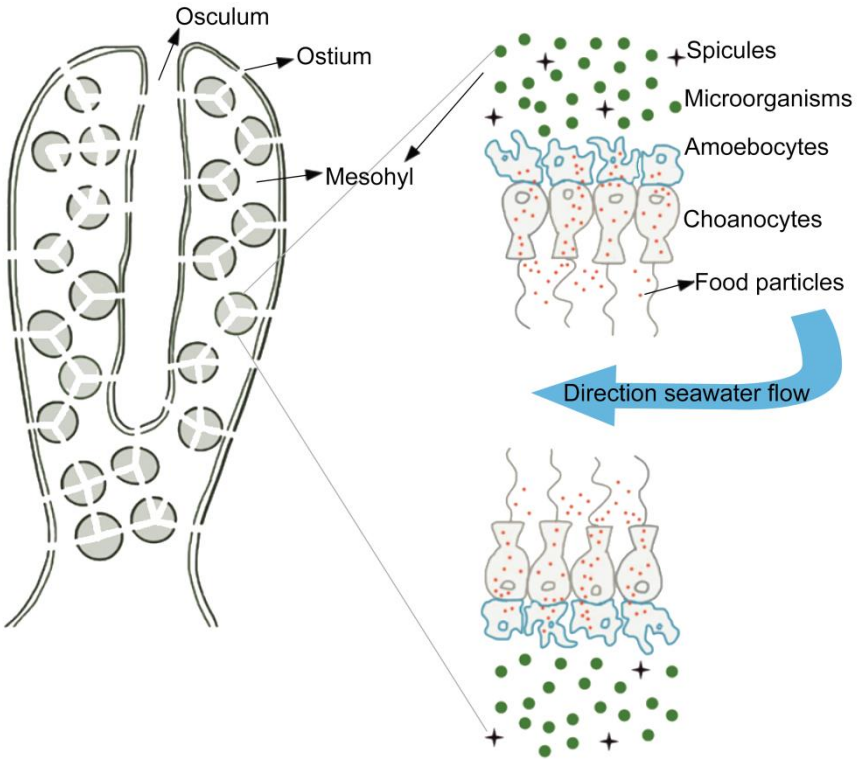



**Figure 2**: Simplified body structure of a filter-feeding sponge on the left, on the right zoom-in on a choanocyte chamber, direction of water flow is indicated (Steinert *et al.*, 2017).

The microorganisms are either phagocytosed or passed on to the mesohyl tissue of the sponge, this mesohyl contains intact microbial cells (potentially symbiotic or

commensalistic), spicules (skeletal framework) and archeaocytes. The spicules are usually made up of silica and provide a skeletal framework for the sponge enabling a proper three-dimensional structure (sponges can range in size from millimetres to meters) and rigidity towards external factors (currents, waves, etc.), as well as internal rigidity as sponges can be soft to fragile, but also rock hard. Spicules are very diverse and are used to identify sponges based on their shape and size (Uriz *et al.*, 2003). In carnivorous sponges specialized spicules (anisochelae) are also used to trap respective food (small crustaceans) (Vacelet and Boury-Esnault, 1995).

### 1.1.2 Microbiome of sponges

Sponges harbour dense microbial communities of up to 38% of their biomass and can even be distinguished by the density of their bacterial population into high microbial abundance (HMA) and low microbial abundance (LMA) sponges. In HMA sponge species the bacterial population can reach a density of $10^8$ – $10^{10}$ bacteria per gram of sponge tissue, whereas in LMA sponge species densities of $10^5$ – $10^6$ bacteria per gram of sponge are observed (Vacelet and Donadey, 1977; Hentschel *et al.*, 2003; Hentschel *et al.*, 2006). The association of microbes and sponges dates back more than 600 million years and is therefore one of the most ancient proven relationships between animals and microorganisms (Taylor *et al.*, 2007; Wilkinson, 1984). The association between sponges and their microbiota seems to be quite unique and selective (especially for HMA sponges), leading to the proposal of specific microbiomes associated with sponges (Kennedy *et al.*, 2014; Moitinho-Silva *et al.*, 2014). In this respect 7500 sponge-derived 16S rRNA gene sequences have been investigated for globally shared sponge-specific clusters, resulting in 173 monophyletic clusters found globally (Simister *et al.*, 2012). Unfortunately, other studies have subsequently found that part of the sponge specific clusters can be found in very low abundances in different marine environments (Taylor *et al.*, 2013; Thomas *et al.*, 2016). Besides bacteria, archaea and eukarya are also associated with sponges (Hentschel *et al.*, 2012). Archaea have been shown to dominate the microbial community of certain deep sea sponge species, making up 70% of the microbial community of the deep sea sponge *Inflatella pellicula* (Jackson *et al.*, 2013). The types of associated microorganisms is extremely diverse and includes 47 prokaryotic phyla, such as *Actinobacteria*, *Chloroflexi*, *Cyanobacteria*, *Proteobacteria* ($\alpha$, $\beta$, $\gamma$, $\delta$) and several

candidate phyla, like *Poribacteria*, *Tectomicrobia* and the Sponge Associated Unclassified Lineage (SAUL) being the most prominent (Hentschel *et al.*, 2012; Wilson *et al.*, 2014). Besides bacteria, fungi are also commonly isolated from sponges and are investigated in respect to their biological activity producing secondary metabolites (Hoeller *et al.*, 2000; Imhoff, 2016). Due to the filter-feeding nature of sponges, fungi are especially enriched in sponges as their numbers in ocean waters are usually quite low in comparison to bacteria. This leads also to the isolation of less common fungal genera like *Beauveria*, *Botryosphaeria*, *Epicoccum*, *Tritirachium* and *Paraphaeosphaeria* from sponges (Hoeller *et al.*, 2000; Indriani, 2007; Paz *et al.*, 2010).

#### 1.1.2.1 Sponge symbionts

Sponge associated microorganisms are quite diverse and identifying them as true sponge symbionts is still a matter of debate and different genomic and metabolic features must be taken into account (Figure 4). Genomic features associated with sponge symbionts include, for example, an overrepresentation of genes containing ankyrin (AR) and tetratricopetide repeats (TPR) (Thomas *et al.*, 2010). AR and TPR mediate protein-protein interactions in eukaryotes and these proteins are involved in different functional processes like transcriptional initiation, cell cycle regulation, cytoskeleton proteins, ion transport and signal transduction (Blatch and Lässle, 1999; Hryniewicz-Jankowska *et al.*, 2002). AR proteins have been shown to mediate the uptake of bacterial cells into amoebal cells, which are functionally analogous to sponge amoebocytes (Reynolds and Thomas, 2016) and therefore play a key role in acquiring symbionts and distinguishing between symbionts and non-symbionts. Another feature of the symbiotic community of a sponge is the ability to nitrify ammonia (Bayer *et al.*, 2008; Thomas *et al.*, 2010) and recently it has been shown that members of the symbiotic community are also involved in detoxification processes by mineralizing ubiquitous environmental toxins like arsenic and barium, which normally accumulate in higher trophic-level organisms (Keren *et al.*, 2017). The candidate phylum *Poribacteria*, which is almost exclusively found in sponges has become a model microorganism for true symbionts in sponges (Fieseler *et al.*, 2004; Siegl *et al.*, 2011). This phylum is widespread across various sponge species (*Demospongiae*) from different oceans

(Lafi *et al.*, 2009). Members of the phylum *Poribacteria* are able to form bacterial microcompartments and are rich in different types of eukaryotic-like protein domains, especially TPR, AR and low-density lipoprotein receptor repeats (Kamke *et al.*, 2014). Besides *Poribacteria*, there is also the filamentous symbiont *Entotheonella* (phylum *Tectomicrobia*), which has been studied in depth for its potential to produce secondary metabolites and has been shown to be responsible for the production of nearly all bioactive compounds derived from the sponge *Theonella swinhoei* (Wilson *et al.*, 2014). Besides its secondary metabolic potential this bacterium also contains eukaryotic-like proteins and is involved in the mineralization of arsenic and barium (Liu *et al.*, 2016; Keren *et al.*, 2017).

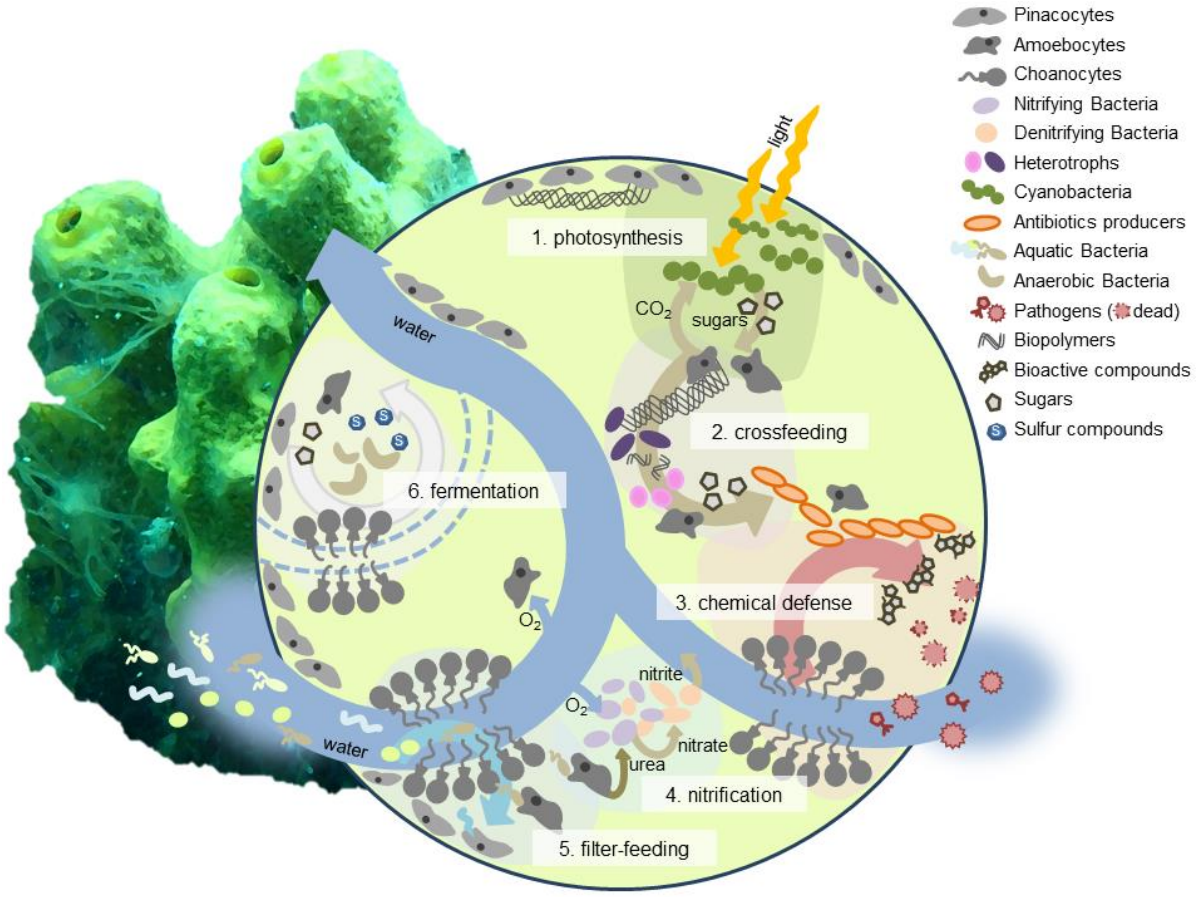



**Figure 4:** Diagram of functions and functionalities provided by sponge associated and symbiotic microorganisms (Steinert *et al.*, 2017).

### 1.1.2.2 Marine natural products from sponges and their microbiota

Nearly 30% of all marine natural products discovered so far originate from sponges and their microbiota, making sponges the richest source of new marine natural products (Mehbub *et al.*, 2014). Due to the sessile filter-feeding nature of sponges they rely solely on stored bioactive natural products (Unson *et al.*, 1994) as a matter of defence when confronted by predators, these metabolites can be either cytotoxic, antibiotic or feeding deterrent (Proksch, 1994; Pawlik *et al.*, 2002). More than 200 new compounds from sponges are reported each year (Laport *et al.*, 2009). The identified natural products have a broad spectrum of biological activities, including antibacterial, anticancer, antifouling, antifungal, anti-inflammatory, antiviral, antiprotozoal, antihelminthic, immunosuppressive, neurosuppressive, neuroprotective and other bioactivities (Sipkema *et al.*, 2005; Blunt *et al.*, 2016). Novel effective drug leads that can potentially be found in sponges are urgently needed to fight evolving infectious microorganisms, as well as fungal and viral diseases (Sagar *et al.*, 2010). Furthermore other diseases like cancers are becoming more and more prevalent in our society and the marine environment continues to be screened as a source of novel anticancer agents, with sponges in particular being a promising source of novel agents (Bhanot *et al.*, 2011). Promising natural products with anticancer activity from sponges are Discodermolide, Hemiasterlins A&B, modified Halichondrin B, KRN-700, Alipkinidine, Fascaphycins, Isohomohalichondrin B, Halichondrin B, Laulimalide/Fijianolide, 5-Methoxyamphimedine and Variolin (Crews *et al.*, 2003). More and more evidence is being gathered and it is already broadly accepted that most of the natural products isolated from marine sponges are actually produced by their microbiota (Piel *et al.*, 2004). Many isolated bioactive compounds display strong structural similarity to complex polyketides and non-ribosomal peptides, which are to date solely known from microorganisms, supporting the hypothesis of a bacterial origin for most of the compounds. Bioactive compounds produced by microbes are normally referred to as secondary metabolites; these are organic compounds that are not directly involved in primary cell functions, such as growth, development or reproduction. These metabolites are produced by specialized gene clusters, including polyketide synthases (PKS, type I, II, III), non-ribosomal peptide synthetases (NRPS), hybrid PKS-NRPS, terpene synthases and clusters involved in the biosynthesis of lantibiotics, nucleosides, bacteriocins, melanins, beta-lactams, phenols, alkaloids and many more. Most

notable in respect to natural product biosynthesis by a symbiont is the filamentous bacterium *Entotheonella*, isolated from the sponge *Theonella swinhoei*, whose genome was sequenced by single cell genomics. For almost all bioactive compounds found in the sponge *Theonella swinhoei*, including polytheonamides, onnamide-type compounds, keramamides, cyclotheonamides and proteusin biosynthetic genes encoding these compounds could be found in the *Entotheonella* genome (Wilson *et al.*, 2014). Nonetheless there are a few exceptions like avarol, stevensine, crambesicidins and some brominated isoxazoline alkaloids that are known to be produced by the sponge itself and not its microbiota (Uriz *et al.*, 1996; Andrade *et al.*, 1999; Turon *et al.*, 2000; Ternon *et al.*, 2016).

## 1.2 The need for novel antimicrobials

Before the 'antibiotic era' infectious diseases such as tuberculosis, syphilis, cholera, smallpox, plague, mumps and many others had been major causes of death among the human population. The discovery of Penicillin by Alexander Fleming in 1928 (Fleming, 2001) and of other antibiotics thereafter lead to a major increases in life expectancy and quality of life. These discoveries together with advances in healthcare (vaccination, antisepsis, public health measures and sanitation) in the early 1950s and in the next decades, in the so called "Golden era of antibiotic discovery"; resulted in infectious diseases stepping down from being the major cause of morbidity in the general population, relative to other diseases which became more prevalent due to lifestyle choices as well as the rise of life expectancy such as cardiovascular diseases, cancer, and stroke.

Soon after the discovery of Penicillin, concerns were raised about the possible development of antibacterial resistance to the antibiotic (even from Alexander Fleming himself) (Levy, 2002). Unfortunately these fears have materialized to date, not only with respect to Penicillin but to a great extent with respect to virtually any antibiotic in medicinal use today (Brown and Wright, 2016). It is hypothesized that we live today in the 'post antibiotic era'(Alanis, 2005), where the number of antibiotic-resistant pathogenic bacteria is rapidly increasing and infectious diseases are again becoming a more common cause of human death. Bacteria have proven to be very capable of adapting to the various antibiotics in current use and they acquire these adaptive traits via a variety of mechanisms involving

mutation, conjugation, transformation and transduction (Table 1). The rapid spread of resistance amongst opportunistic human pathogens to antimicrobials is a huge threat to the healthcare system and future development of the human population in general. Unfortunately we currently seem to be unable to keep pace with the ever evolving drug resistance amongst microbial pathogens. In fact we are even speeding up the process of the evolution of antibiotic resistance ourselves by misusing antimicrobials; particularly in the widespread and quite indiscriminate use of antibiotics in the agricultural/aquaculture areas (Davies and Davies, 2010). To date most of the antimicrobials in use have been isolated from microorganisms from terrestrial, temperate or tropical environments. In order to find novel bioactive compounds with new modes of action it is widely believed that microorganism from different environmental ecosystems (marine ecosystem, shallow water and the deep sea) and ecological niches (hydrothermal vents, saline brines sediments) need to be targeted.

**Table 1**: Antibiotic families, mechanism of action and resistance mechanism (adapted from (Alanis, 2005; Davies and Davies, 2010)

| Antibiotic family | Antibiotic target | Resistance mechanism |
|---|---|---|
| Beta-lactams | Inhibition of cell wall synthesis | Beta-lactamases, efflux, altered target |
| Glycopeptides | Inhibition of cell wall synthesis | Reprogramming peptidoglycan biosynthesis |
| Cyclic Lipopeptides | Inhibition of cell wall synthesis | Altered target |
| Tetracyclines | Inhibition of protein synthesis | Monooxygenation, efflux, altered target |
| Aminoglycosides | Inhibition of protein synthesis | Phosphorylation, acetylation, nucleotidylation, efflux, altered target |
| Streptogramins | Inhibition of protein synthesis | C-O lyase, acetylation, efflux, altered target |
| Oxazolidonones | Inhibition of protein synthesis | Efflux, altered target |
| Macrolides | Inhibition of protein synthesis | Hydrolysis, glycosylation, phosphorylation, efflux, altered target |
| Lincosamindes | Inhibition of protein synthesis | Nucleotidylation, efflux, altered |

| | | |
|---|---|---|
| | | target |
| Fluoroquinolones | Inhibition of DNA synthesis | Acetylation, efflux, altered target |
| Rifamycins | Inhibition of RNA synthesis | ADP-Ribosylation, efflux, altered target |
| Sulfonamides | $C_1$ metabolism | Efflux, altered target |
| Polymyxins | Membrane disorganizing agents | Efflux, altered target |
| Nitroimidazole | others | Altered target |
| Phenicols | Inhibition of protein synthesis | Acetylation, efflux, altered target |
| Pyrimidines | $C_1$ metabolism | Efflux, altered target |

### 1.3 The marine habitat in respect to its biotechnological potential

The sheer size of the oceans and therefore the marine habitat, with all its unique niches is not only interesting with respect to a potential source of novel drug leads, but also for bioprospecting for novel enzymes and biocatalysts (Figure 5). Marine biocatalysts potentially offer novel properties like high salt tolerance, hyperthermostability, barophilicity and cold adaptivity, as well as novel chemical and stereochemical properties (Debashish *et al.*, 2005; Trincone, 2011). The number and variety of enzymes studied from the marine environment is astonishing and includes proteases, peroxidases, chitinases, carbohydrolases (amylases, cellulases, xylanases), agarases, lipases, esterases and many more. The main practical applications for biocatalysts and marine enzymes are found primarily in five domains of industrial applications, those being chemistry, pharmacology, food, cosmetics and agricultural according to related patents from the period of 1973-2007 (Leary *et al.*, 2009).

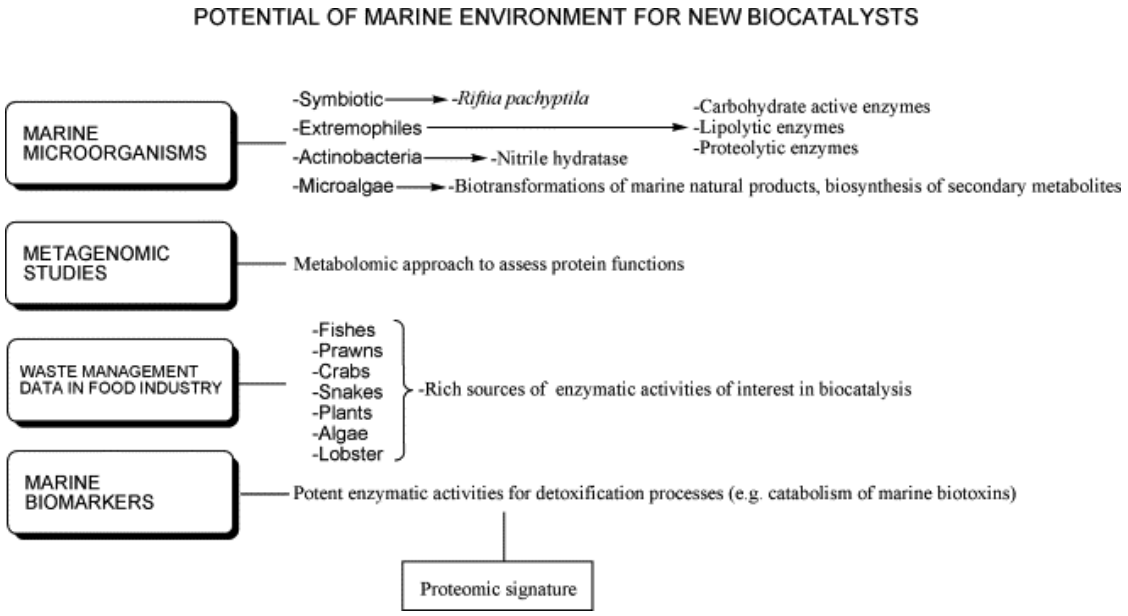



**Figure 5**: Scheme of targets and approaches to exploit the marine biotechnological potential (Trincone, 2010).

Different niches for extremophiles in the marine environment include hydrothermal vents for (hyper-)thermophiles and the deep sea for psychro- and barophiles to name the most prominent (Mesbah and Sarmiento, 2016). Furthermore enzymes classified as acidophilic, alkaliphilic, endolith, metallotolerant, radioresistant and toxitolerant are highly sought after and can be widely found in the marine environment (Trincone, 2011; Selvin *et al.*, 2012; Nigam, 2013; Dumorné *et al.*, 2017) and are also referred to as extremozymes (Elleuche *et al.*, 2014; Hough and Danson, 1999). Marine microorganisms able to grow at high temperatures are for example the archaeon *Pyrococcus furiosus* and the bacterium *Thermotoga maritima*. Both microorganisms have been exploited for their biocatalytic potential and especially an alcohol dehydrogenase from *Pyrococcus furiosus* and a glucoside hydrolase from *Thermotoga maritima* are of potential industrial interest, with both enzymes displaying, besides their thermostability, a high tolerance towards organic solvents (Goyal *et al.*, 2001; Jiang *et al.*, 2004a; Jiang *et al.*, 2004b; Zhu *et al.*, 2006).

The contrast of hyperthermostability and cold adaptivity is also of interest for industrial applications. Cold active biocatalysts can be used as additives in food-, detergent industry and bioremediation processes. Psychrophilic enzymes are advantageous because

they help to reduce energy costs, as well as the risk of microbial contamination and chemical side reactions, additionally they possess high specific activity and low heat stability, allowing for easy inactivation (Cavicchioli *et al.*, 2002; Trincone, 2011; Santiago *et al.*, 2016;). Microorganisms from the marine environment known to encode cold active enzymes are for example *Pseudoalteromonas arctica* and *Shewanella* sp. G5. A cold-active esterase from *Pseudoalteromonas arctica* has been successfully cloned and characterized, and it shows a broad substrate specificity for short chain fatty acid esters and is also capable of hydrolysing medically relevant esters like the anti-inflammatory drugs naproxen, ketoprofen and ibuprofen (Al Khudary *et al.*, 2010). From *Shewanella* sp. G5, which is able to use cellobiose as carbon source, a cold-active β-glucosidase with potential application in the winemaking industry has been reported (Cristobal *et al.*, 2009; Cristóbal *et al.*, 2016).

### 1.3.1 Approaches to exploit biotechnological potential

The exploitation of the biotechnological potential of a given environmental sample depended for a long time on the cultivability of the relevant microorganisms. Cultivation approaches have therefore been continuously refined and have become more and more sophisticated, nonetheless approximately 99% of environmental microbes cannot currently be cultured under laboratory conditions (Handelsman, 2004; Singh, 2010). A rise in the use of metagenomics based strategies (a metagenome is defined as all the genomic DNA that is present in a given sample) has aimed to close this gap and besides investigating the biotechnological potential of particular environmental ecosystems has also provided insights into the relationships between microbes, between microbes and their environment and between microbes and their hosts (Streit and Schmitz, 2004; Kennedy *et al.*, 2007; Jackson *et al.*, 2015). Some metagenomic studies initially involve the creation of a metagenomic library, where the environmental DNA is subcloned into suitable vectors, fosmids or bacterial artificial chromosomes and subcloned into a host, with *Escherichia coli* being the most prominent host system employed. The analysis of the generated libraries can be generally distinguished into two types, involving either function-based or sequence-based metagenomic analysis (Venter *et al.*, 2004; Uchiyama and Miyazaki, 2009;) (Figure 6). Function-based metagenomics rely on effective screening methods and are hampered by

insert-size and choice of library host, but the huge advantage is the chance of identifying completely novel functional enzymes (Kennedy *et al.*, 2010; Kennedy *et al.*, 2011). Sequence-based metagenomics are either based on shotgun-sequencing of the partial or whole metagenome (no clone library required) (Schmeisser *et al.*, 2003; Venter *et al.*, 2004; Vieites *et al.*, 2009) or by using PCR with degenerate primers for the gene of interest, colony blotting or radioactive probes (Schloss and Handelsman, 2003; Chen and Murrell, 2010). Major challenges of sequence-based metagenomic approaches lies in the amount of data generated by sequencing and by the low chance of identifying truly novel genes when using degenerate probes.

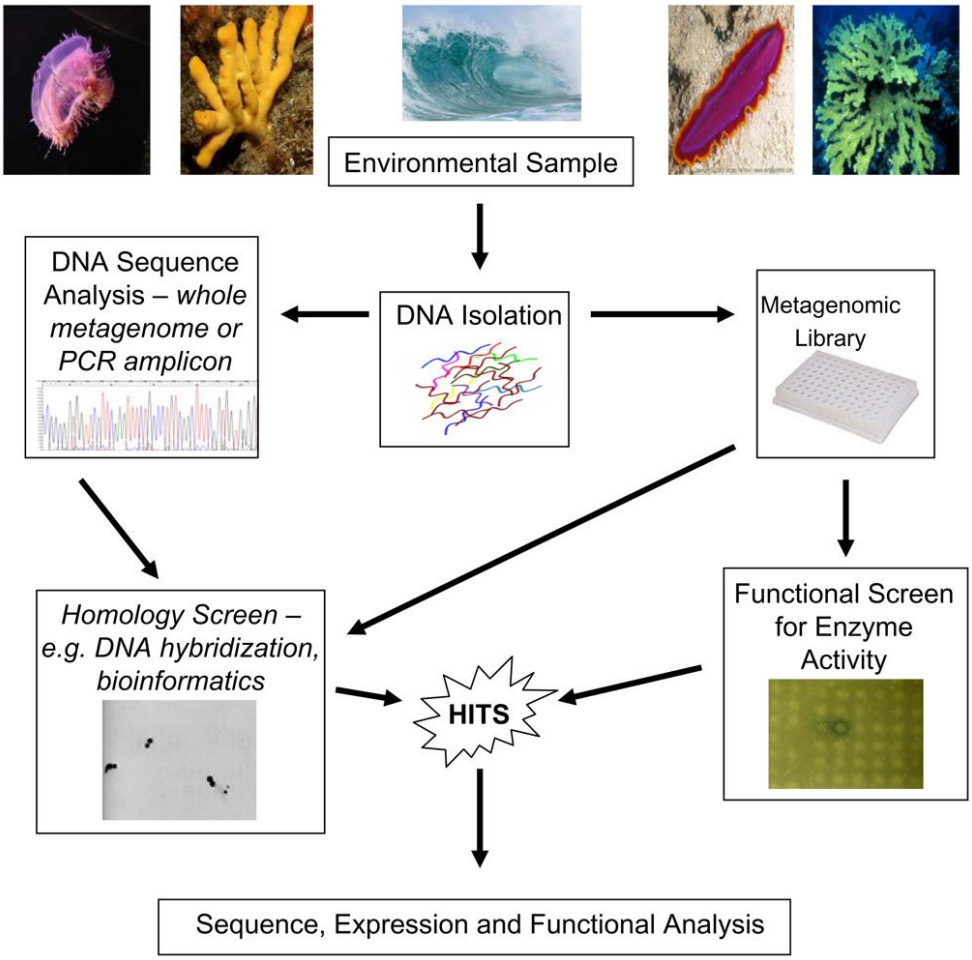



**Figure 6**: Enzyme discovery pipeline for function- and sequence-based metagenomic approaches (Kennedy *et al.*, 2010).

### 1.3.2 Shallow water environment

Shallow water habitats are more easily accessible than the deep sea, but can still require some sophisticated equipment, like self-contained underwater breathing apparatus (SCUBA) or remotely operated vehicles (ROV) to retrieve samples. The marine habitat is acknowledged as a really important resource of novel enzymes, biocatalysts and natural products, but nonetheless the rate of scientific research, and therefore publication and patent outputs is less than from other environments that have been studied (Figure 7)(Trincone, 2010).

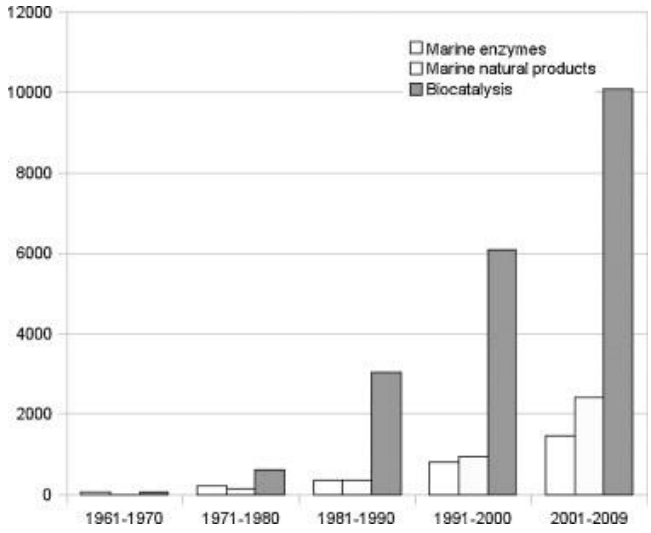



**Figure 7**: Combined number of articles, reviews, patents, etc. containing the concepts of 'marine enzymes', 'marine natural product' and 'biocatalysis' up to the end of 2009 according to MEDLINE and CAPLUS searches (Trincone, 2010).

The range of enzymes described from the marine environments as outlined previously is tremendous and the most common type of biocatalysts are carbohydrate and protein degrading enzymes, as well as lipolytic hydrolases (Lee *et al.*, 2010; Trincone, 2010; Trincone, 2011).

### 1.3.3 The 'deep sea'

The oceans account for 71% of the surface of our planet and is usually divided into coastal and shallow water regions and the 'deep sea', roughly anything deeper than 200 m, with over 50% of the oceans being below 3000 m (mean depths 3800 m), which makes the 'deep sea' the biggest interconnected habitat on the entire planet (Ramirez-Llodra *et al.*, 2010). Due to the harsh conditions prevalent in this ecosystem, most notably high pressure, low temperature and the absence of sunlight it was long believed to be very hostile for living beings. In contrast, expeditions with remotely operated vehicles and manned submarines have shown a wealth of diversity, especially microbial diversity in this extreme environment (Sogin *et al.*, 2006; Jørgensen and Boetius, 2007). In this respect hydrothermal vents within the 'deep sea' have been described in detail as hotspots for living beings and biological diversity (Naganuma, 2000; Zierenberg *et al.*, 2000; Flores *et al.*, 2012; Yang *et al.*, 2013a; Lossouarn *et al.*, 2015;).

An example of an enzyme from the deep sea in industrial use is an $\alpha$-amylase marketed by BASF Enzyme LLC as Fuelzyme®. The enzyme is marketed as having broad temperature and pH operating values and can be used for mash liquefaction in ethanol fuel production. In 2005, Ferrer *et al*. published a metagenomic study of a hypersaline deep-sea anoxic basin, where they identified five esterases and two out of the five were able to function under the harsh conditions of this environment (high salinity, high hydrostatic pressure, anoxia and a sharp chemocline). Furthermore, one of the esterases had a unique adaptive structure:function configuration, enabling it to display high catalytic activity under a wide range of physicochemical conditions (Ferrer *et al.*, 2005).

## 1.4. The pharmaceutical potential of cold environments

Cold environments such as Arctic and Antarctic regions and the deep sea are richly populated by microbes which encounter the same selective pressures and/or even more than their counterparts from moderate or warm environments. Microbes from moderate and warm environments have been extensively studied for their ability to produce antimicrobial compounds, but this resource seems to be exhausted. Keeping in mind the rapid emergence of antimicrobial resistance, we need to look for new sources of antimicrobials. It is widely

believed that environmental conditions shape the chemistry and lifestyle of the native microbial communities, therefore the investigation of constantly cold environments might be advisable. Psychrophiles are an as of yet largely untapped source for novel or considerably different antimicrobial compounds. To date a sustained level of research has not focused on psychrophiles as a source of novel bioactive compounds, but with the ever growing need for new antibiotics due to the aforementioned ongoing threat of antimicrobial resistance; this will undoubtedly change in the future.

### 1.4.1 Bioactive compounds from cold environments

While cold environments have to date been mostly overlooked in the search for new antimicrobials, nonetheless there are several studies focusing on the microbial diversity of cold environments which indicate a high level of diversity within these environments. It is widely believed that high levels of microbial diversity is indicative of high levels of potential antibiotic producing microbes, given that these compounds are likely to be advantageous for the producing organism; particularly in an environment where there is competition for resources. Furthermore given that marine microorganisms have survived under extremes of temperature, salinity, and pressure over many millions of years; then they are likely to have evolved to adapt to these extreme conditions and therefore potentially possess novel biochemistry. Thus due to the environmental differences between cold marine environments and temperate or tropic marine environments coupled with adaptive evolution it can be assumed that the bioactive compounds produced by microorganisms from these cold environments are likely to be quite different from many of the classes of antimicrobials currently in use. To date mostly large-scale and rather unspecific antimicrobial screens of microorganism retrieved from for example alpine sites, benthic mats from Antarctic lakes and sponges from deep-sea and arctic environments have been performed. Therefore more emphasis should be placed on finding new antimicrobials from these sources coupled to a more in depth analysis of the compounds/activities found, because to date only a few studies have concentrated on targeting antimicrobial activity in these environments, nonetheless most of them show promising results.

A comprehensive review on cold-water marine natural products was published in 2007 by Lebar *et al.*, covering most of the compounds identified up until 2005 from cold marine environments (Lebar *et al.*, 2007). In this review natural products from microbes, bacteria, fungi, microalgae, macroalgae, sponges, corals, bryozoans, molluscs, tunicates and echinoderms living in cold marine environments were described. Furthermore Abbas and co-workers subsequently published a review on Arctic and Antarctic sponge secondary metabolites (Abbas *et al.*, 2011). Thus only the more recent advances are mentioned subsequently, but recommend the interested reader to refer to the aforementioned publication if required. Furthermore a general article on marine natural products is typically published on an annual basis (Blunt *et al.*, 2016), but typically no more than 3% of the compounds which are described are retrieved from cold or deep sea sources and even less display antimicrobial activity.

#### 1.4.1.1 Synoxazolidinones

Synoxazolidinones A and B (Figure 8) have been isolated from the ascidian *Synoicum pulmonaria* collected from the Norwegian coast in 2010 (Tadesse *et al.*, 2010). These compounds constitute a novel family of brominated guanidinium oxazolidinones with activities against a range of Gram-positive bacteria, especially against methicillin-resistant *Staphylococcus aureus* (MRSA). Besides the Synoxazolidinones A and B the ascidian also produces Synoxazolidinones C and Pulmonarins, which are also brominated compounds, all of which display some kind of antimicrobial activity especially against micro- and macrofouling organisms in the water column and are therefore of industrial interest (Trepos *et al.*, 2014). However one of the major bottlenecks in the use of bioactive compounds from natural resources for biopharmaceutical application is the quantities produced by the native strains; which are often quite low. Therefore the large-scale production of these compounds would need the harvest of huge amounts of ascidians to fulfil the required demands. Therefore the possibility of total chemical synthesis of Synoxazolidinones as described in Shymanska *et al*. would be required to allow large scale productions, thereby minimizing any potential detrimental environmental impact of large scale harvesting of ascidians (Shymanska *et al.*, 2014).

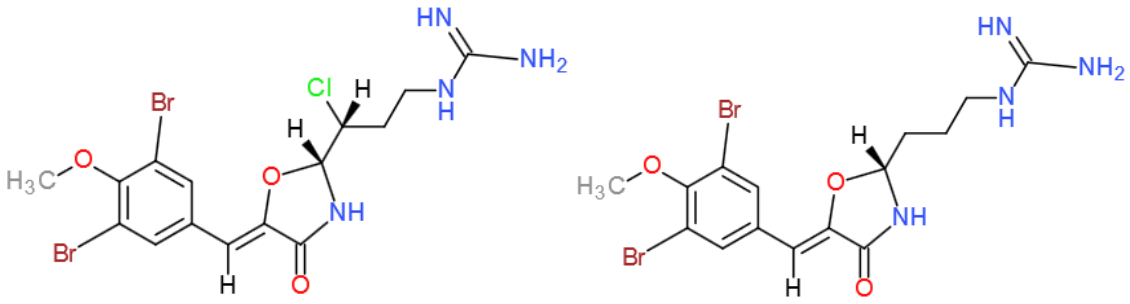



**Figure 8**: Synoxazolidinones A and B (adapted from (Tadesse *et al.*, 2010) and visualized with 2D Sketcher (https://web.chemdoodle.com/demos/sketcher/))

### 1.4.1.2 Microcins

In 2010 a bacteriocin like compound named Serraticin A produced by a psychrophilic microorganism closely related to *Serratia proteamaculans* was isolated from a soil sample from Isla de los Estados, Argentinia and was termed to be the first cold-active compound with antimicrobial activity from *S. proteamaculans*, it was produced at 8°C (Sánchez *et al.*, 2010). The compound showed activity against an *Escherichia coli* and a *Salmonella enterica* strain. The mode of action was proposed to involve either blocking DNA replication or inhibition of the septation process.

### 1.4.1.3 Lantibiotics

Subtilomycin is a class I bacteriocin and was purified from a *Bacillus subtilis* strain isolated from the marine sponge *Haliclona simulans* collected on the west coast of Ireland (Phelan *et al.*, 2013). The peptide shows very good activity against *Clostridium sporogenes*, good activity against *Bacillus cereus*, *Bacillus megaterium*, *Listeria monocytogenes* and *Listeria innocua* and also some activity against *Staphylococcus aureus*, a methicillin resistant *S. aureus* strain and a vancomycin resistant *S. aureus* strain. The bacteriocin also showed strong inhibitory activity against multiple *Candida* species (*C. albicans*, *C. dubliniensis*, *C. lusitaniae* and *C. parapsilosis*). Class I lantibiotics usually interfere with the cell membranes of its target, either by inhibiting membrane biosynthesis or pore forming (McAuliffe *et al.*, 2001).

### 1.4.1.4 Spirotetronate antibiotics

Lobophorins are kijanimicin derivatives classified as medium-sized spirotetronates with a central ring system comprising of 13 carbon atoms (Vieweg *et al.*, 2014). Kijanimicin itself is produced via a modular Type-I polyketide synthase and an operon involved in the attachment and intramolecular cyclization of glycerate units (Zhang *et al.*, 2007). In 2013 Pan and co-workers identified two new groups of compounds namely Lobophorins H and I from a *Streptomyces sp.* isolated from a south China deep sea sediment sample, retrieved from a depth of 2134 m (Pan *et al.*, 2013). Lobophorin H in particular showed potent activity against *Bacillus subtilis* (Figure 9), which was comparable to the activity of ampicillin; unfortunately the mode of action for Lobophorins has not yet been described. Furthermore the Lobophorins seem to be exclusively active against Gram positive bacteria and not against either Gram negative bacteria or fungi, but some of them do display antitumor activities against oral cancer cells (Cruz *et al.*, 2015). Additionally the lobophorins and other kijanimicin derivatives seem to be widespread in nature, lobophorins H and I from deep sea sediment (Pan *et al.*, 2013), lobophorins A and B from a tropical marine bacterium (Jiang *et al.*, 1999) and kijanimicin from the soil actinomycete *Actinomadura kijaniata* (Zhang *et al.*, 2007).

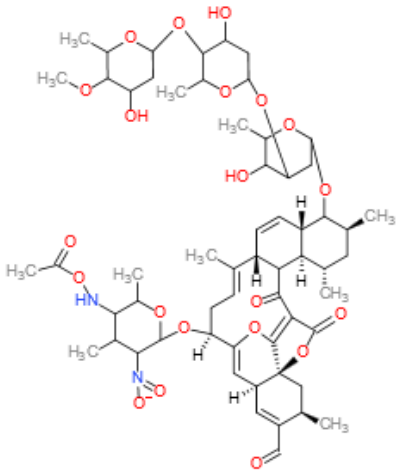


**Figure 9**: Lobophorin H (adapted from (Pan *et al.*, 2013) and visualized with 2D Sketcher)

In 2013 Wang and co-workers screened a large marine-derived library comprising of 4024 bacterial and 533 fungal isolates for growth inhibition of the Bacille Calmette-Guérin an attenuated strain of the bovine tuberculosis bacillus *Mycobacterium bovis* (Wang *et al.*, 2013).

Twenty seven of the screened abstracts (0.6%) showed inhibitory activity. One of the active extracts was from a south China deep sea sediment-derived actinomycete, *Verrucosispora* sp., which was retrieved from 2733 m below sea level. The marine actinomycete *Verrucosispora* sp. is also known to produce other bioactive compounds like proximicins A, B and C and thiocoraline A a cytotoxic thiodepsipeptide (Schneider *et al.*, 2008; Wyche *et al.*, 2011). Structural elucidation of the active fractions revealed the presence of three new abyssomicin polyketides (abyssomicin J, K and L) as well as four known abyssomicins (abyssomicin B, C, D and H), which were formerly also isolated from *Verrucosispora sp.* isolates. The newly isolated abyssomicin J (Figure 10) is a dimeric thioester which in contrast to other members of the abyssomicin family, which are typically monomeric small spirotetronates (central ring system C=11). Abyssomicins are of particular interest as novel antibiotics as they target the *p*-aminobenzoic acid biosynthetic pathway, which is involved in the synthesis of tetrahydrofolate; a pathway unique to multiple microorganisms, but not found in humans (Bister *et al.*, 2004).

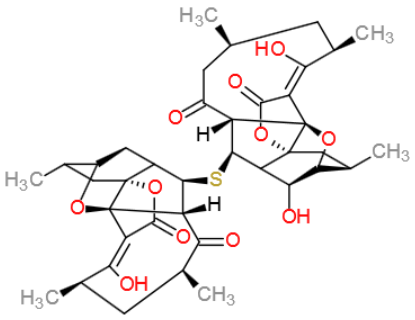



**Figure 10**: Abyssomycin J (adapted from (Wang *et al.*, 2013) and visualized with 2D Sketcher)

### 1.4.2 Other bioactive compounds from cold environments

The actinomycete genus *Serinicoccus* which was firstly discovered in 2004 from a deep sea sediment sample from the Indian ocean retrieved from a depth of 5368 m (Xiao *et al.*, 2011) and which currently contains only three species, all of whom were isolated from marine habitats; were recently reported to produce secondary metabolites and new indole alkaloids (Figure 11) with weak antimicrobial and cytotoxic activities (Yang *et al.*, 2013b).

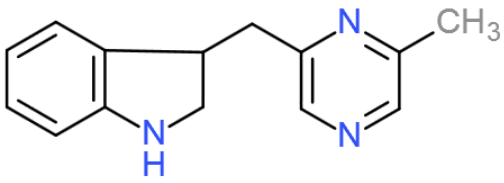



**Figure 11**: New Indole Alkaloid (adapted from (Yang *et al.*, 2013b) and visualized with 2D Sketcher)

Monanchocidins B-E (Figure 12) are unusual polycyclic guanidine alkaloids isolated from the marine sponge *Monanchora pulchra* collected near Urup Island by dredging, which displayed potent antileukemic activities (Makarieva *et al.*, 2011), by inducing apoptosis (Guzii *et al.*, 2010). Monanchocidin A overcomes drug resistance of cancer and tumor cell lines by inducing autophagy and lysosomal membrane permeabilization, making it a promising drug lead (Dyshlovoy *et al.*, 2015). The Monanchocidins are part of the well-known group of pentacyclic guanidine alkaloids with the first representative being Ptilomycin A (Ohizumi *et al.*, 1996). Metabolites of this compound displayed a broad range of biological activities including antifungal, antimicrobial, antimalarial and many other properties.

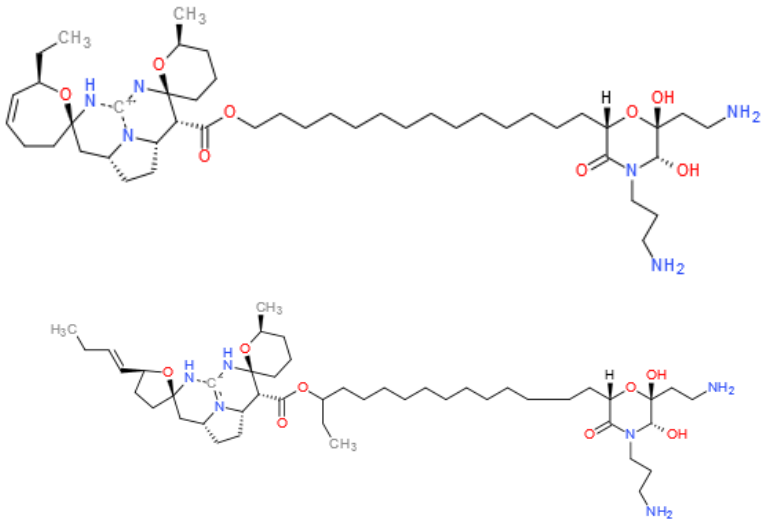



**Figure 12**: Monanchocidins B (top) and E (bottom) (adapted from (Makarieva *et al.*, 2011) and visualized with 2D Sketcher)

New antibacterial compounds namely ent-Eusynstyelamides D, E, and F (Figure 13) were isolated from the arctic bryozoan *Tegella cf. spitzbergensis* and represented the first report of compounds with antimicrobial activity from this organism (Tadesse *et al.*, 2011)

with inhibitory activities against *E. coli*, *S. aureus*, *P. aeruginosa* and *C. glutamicum* strains. These Eusynstyelamides are brominated tryptophan derivatives which were first isolated from the Australian Great Barrier Reef ascidian *Eusynstyela latericius* (Tapiolas *et al.*, 2009), which displayed strong neuronal nitric oxide synthase inhibitory capabilities, which could be of potential use in treating neuropathological disorders, such as stroke, Alzheimer's disease, Parkinson's disease and dementia, which are commonly associated with nitric oxide overproduction (Calabrese *et al.*, 2000).

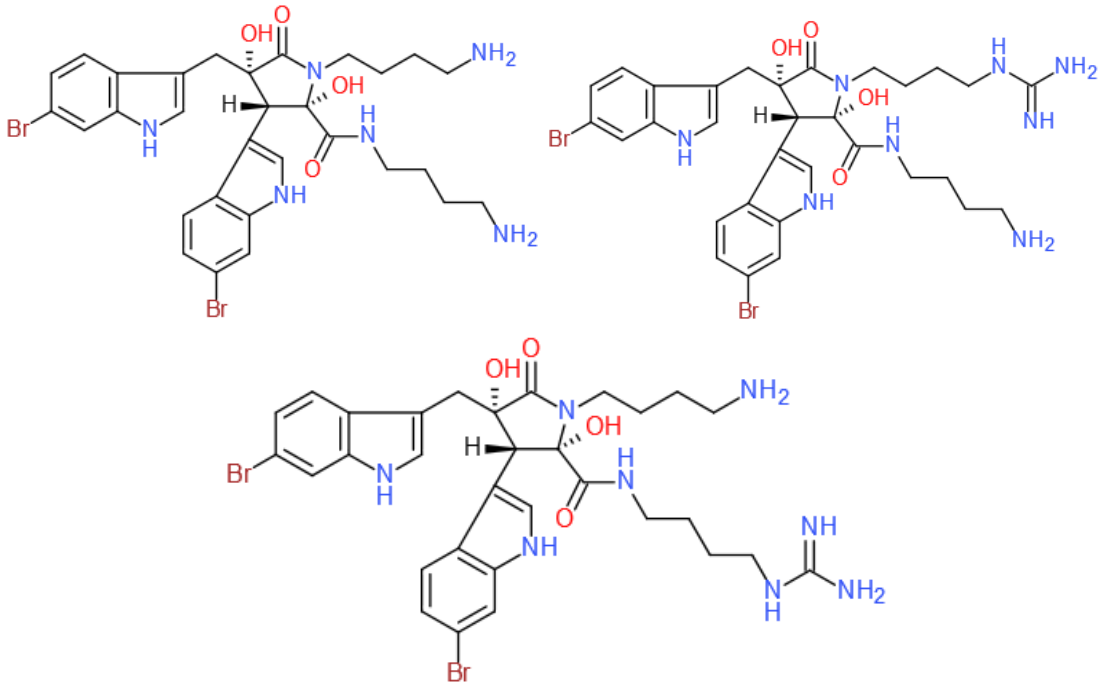



**Figure 13**: Eusynstyelamides D, E and F (clockwise direction) (adapted from (Tadesse *et al.*, 2011) and visualized with 2D Sketcher)

### 1.4.3 Uncharacterized compounds from cold environments

Kim and co-workers isolated bioactive microorganisms from an arctic lichen collected in Spitzbergen (Kim *et al.*, 2014). Lichens are a composite, symbiotic organism comprising of an algae or cyanobacteria and a filamentous fungi. They isolated five bacteria with antibacterial activity, which were closely related to either *Sphingomonas* sp. or *Burkholderia* sp. The isolates were active against Gram-positive (*S. aureus*, *B. subtilis*, *M. luteus*) as well as Gram-negative (*E. coli*, *P. aeruginosa*, *E. cloacae*) indicator strains in disk diffusion tests and minimum inhibitory concentration assays.

Lo Giudice and colleagues studied 580 bacterial isolates retrieved from various Antarctic marine sources, such as seawater, sediment and Antarctic fish intestine collected during four oceanographic campaigns for their antibacterial activities against terrestrial microorganisms (Gram-positive and Gram-negative strains as well as the eukaryotic fungus *Candida albicans*). Twenty two of the isolates showed varying degrees of antibacterial activity against *E. coli*, *Proteus mirabilis*, *Micrococcus luteus* and *B. subtilis*. The active microbial isolates were identified as belonging to two main phylogenetic groups one being Actinobacteria (*Arthrobacter*, *Janibacter*, *Nesterenkomia* and *Rhodococcus* sp.) and the other being γ-Proteobacteria (*Pseudoalteromonas* and *Pseudomonas* sp.) by 16S rRNA gene sequencing analysis. Interestingly the γ-proteobacterial isolates only inhibited the Gram-negative indicator strains *E. coli* and *P. mirabilis*, whereas the activity displayed by the Actinobacteria was more widespread (Lo Giudice *et al.*, 2007).

In another study 132 bacterial isolates retrieved from three Antarctic sponges (*Haliclonissa verrucosa*, *Anoxycalyx joubini* and *Lissodendoryx nobilis*) were screened for antimicrobial activity against more than 70 different *Burkholderia* sp. strains and other indicator strains. *Burkholderia* besides the more common *Pseudomonas aeruginosa* is commonly connected to infections in Cystic fibrosis patients, but due to its resistance to most antibiotics it is very difficult to treat. Most of these isolates exhibited an ability to inhibit the growth of *Burkholderia cepacia* complex bacteria, but not other pathogenic bacteria, which indicates a very specific action against these types of bacteria. The retrieved bacteria belonged mostly to the *Arthrobacter, Pseudoalteromonas, Psychrobacter, Shewanella* and *Roseobacter* genera. The cause of action was believed to be due to the production of an array of volatile organic compounds (VOCs) produced by these isolates rather than by bioactive secondary metabolites, so no evidence was found for polyketide synthase genes and plasmid related sequences involved in the biosynthesis of the VOCs. Interestingly the array of VOCs produced, differed from isolate to isolate and corresponded to the range of observed antimicrobial activities (Papaleo *et al.*, 2012). The aforementioned work lead to the sequencing and comparative analysis of three *Arthrobacter* (Orlandini *et al.*, 2014) and three *Psychrobacter* strains (Fondi *et al.*, 2014) which displayed good antibacterial activity against *Burkholderia cepacia* complex bacteria, but unfortunately none of the studies was able to provide further insights into the cause of antibacterial activity besides excluding known

secondary metabolite gene clusters, suggesting an unknown type of action and/or compound.

A halophilic Antarctic *Nocardioides* sp. retrieved from Antarctic soil has been investigated for the production of enzymes and antimicrobial properties following growth on different carbon sources (Gesheva and Vasileva-Tonkova, 2012). The bacterium displayed differential expression of hydrolytic enzymes and antimicrobial compounds in respect to the available carbon source, which highlights the importance of varying the growth condition in the laboratory to help unlock the 'hidden' potential from environmental isolates. The isolate displayed antimicrobial activity against Gram-positive and Gram-negative bacteria, with the highest activity against *S. aureus* and *Xanthomonas oryzae*. The antimicrobial activity towards *Xanthomonas oryzae* is of special importance here as this bacterium causes bacterial blight in rice, one of the most harmful diseases of rice, therefore with the greatest economic impact. Further analysis suggested that glycolipids and/or lipopeptides could be responsible for the antimicrobial phenotype, depending on the carbon source on which the isolate was cultured.

In summary the marine environment is a promising resource for novel bioactive compounds as well as for novel biocatalysts. Molecules produced by microorganisms from this environment are likely to possess different biochemical characteristics and potential novel mode of actions from those produced by microorganism from terrestrial environments. This is likely due to the inherently different physiochemical characteristics encountered by these marine microorganisms relative to their terrestrial counterparts. Thus these novel bioactive compounds and/or novel biocatalysts are therefore of potential interest for both industrial and medical based applications. Of special interest in this respect is undoubtedly the deep sea, which is one of the least explored environments on our earth and one of the last remaining frontiers awaiting extensive scientific exploitation.

* Parts of this introduction have been used in the book chapter:
Borchert E, Jackson SA, O'Gara F, Dobson ADW: Psychrophiles: From Biodiversity to Biotechnology 2nd Edition 20017, **Chapter 23: Psychrophiles as a source of novel antimicrobials.** Springer Verlag, Berlin Heidelberg

## 1.5 Bibliography


Abbas, S., M. Kelly, J. Bowling, J. Sims, A. Waters, and M. Hamann, 2011, Advancement into the Arctic region for bioactive sponge secondary metabolites: Mar Drugs, v. 9, p. 2423-37.

Al Khudary, R., R. Venkatachalam, M. Katzer, S. Elleuche, and G. Antranikian, 2010, A cold-adapted esterase of a novel marine isolate, *Pseudoalteromonas arctica*: gene cloning, enzyme purification and characterization: Extremophiles, v. 14, p. 273-85.

Alanis, A. J., 2005, Resistance to antibiotics: are we in the post-antibiotic era?: Arch Med Res, v. 36, p. 697-705.

Andrade, P., R. Willoughby, S. Pomponi, and R. Kerr, 1999, Biosynthetic studies of the alkaloid, stevensine, in a cell culture of the marine sponge *Teichaxinella morchella*: Tetrahedron Letters, v. 40, p. 4775-4778.

Ayling, A., 1980, Patterns of sexuality, asexual reproduction and recruitment in some subtidal marine demosponge: The Biological Bulletin, v. 158.

Bayer, K., S. Schmitt, and U. Hentschel, 2008, Physiology, phylogeny and *in situ* evidence for bacterial and archaeal nitrifiers in the marine sponge *Aplysina aerophoba*: Environ Microbiol, v. 10, p. 2942-55.

Bell, J. J., 2008, The functional roles of marine sponges, Estuarine, Coastal and Shelf Science, Elsevier, p. 341-353.

Bhanot, A., R. Sharma, and M. Noolvi, 2011, Natural sources as potential anti-cancer agents: A review: International Journal of Phytomedicine, v. 3, p. 09-26.

Bister, B., D. Bischoff, M. Ströbele, J. Riedlinger, A. Reicke, F. Wolter, A. T. Bull, H. Zähner, H. P. Fiedler, and R. D. Süssmuth, 2004, Abyssomicin C-A polycyclic antibiotic from a marine *Verrucosispora* strain as an inhibitor of the p-aminobenzoic acid/tetrahydrofolate biosynthesis pathway: Angew Chem Int Ed Engl, v. 43, p. 2574-6.

Blatch, G. L., and M. Lässle, 1999, The tetratricopeptide repeat: a structural motif mediating protein-protein interactions: Bioessays, v. 21, p. 932-9.

Blunt, J. W., B. R. Copp, R. A. Keyzers, M. H. Munro, and M. R. Prinsep, 2016, Marine natural products: Nat Prod Rep, v. 33, p. 382-431.

Brown, E. D., and G. D. Wright, 2016, Antibacterial drug discovery in the resistance era: Nature, v. 529, p. 336-43.

Calabrese, V., T. E. Bates, and A. M. Stella, 2000, NO synthase and NO-dependent signal pathways in brain aging and neurodegenerative disorders: the role of oxidant/antioxidant balance: Neurochem Res, v. 25, p. 1315-41.

Cavicchioli, R., K. S. Siddiqui, D. Andrews, and K. R. Sowers, 2002, Low-temperature extremophiles and their applications: Curr Opin Biotechnol, v. 13, p. 253-61.

Chen, Y., and J. C. Murrell, 2010, When metagenomics meets stable-isotope probing: progress and perspectives: Trends Microbiol, v. 18, p. 157-63.

Crews, P., W. Gerwick, F. Schmitz, D. France, K. Bair, A. Wright, and Y. Hallock, 2003, Molecular approaches to discover marine natural product anticancer leads - an update from a drug discovery group collaboration: Pharmaceutical Biology, v. 41, p. 39-52.

Cristobal, H., A. Schmidt, E. Kothe, J. Breccia, and C. Abate, 2009, Characterization of inducible cold-active β-glucosidases from the psychrotolerant bacterium *Shewanella* sp. G5 isolated from a sub-Antarctic ecosystem: Enzyme and Microbial Technology, v. 45, p. 498-506.

Cristóbal, H. A., H. R. Poma, C. M. Abate, and V. B. Rajal, 2016, Quantification of the genetic expression of bgl-A, bgl, and CspA and enzymatic characterization of β-glucosidases from *Shewanella* sp. G5: Mar Biotechnol (NY), v. 18, p. 396-408.

Cruz, P. G., A. M. Fribley, J. R. Miller, M. J. Larsen, P. J. Schultz, R. T. Jacob, G. Tamayo-Castillo, R. J. Kaufman, and D. H. Sherman, 2015, Novel Lobophorins inhibit oral cancer cell growth and induce Atf4- and Chop-dependent cell death in murine fibroblasts: ACS Med Chem Lett, v. 6, p. 877-81.

Davies, J., and D. Davies, 2010, Origins and evolution of antibiotic resistance: Microbiol Mol Biol Rev, v. 74, p. 417-33.

Debashish, G., S. Malay, S. Barindra, and M. Joydeep, 2005, Marine enzymes: Adv Biochem Eng Biotechnol, v. 96, p. 189-218.

Dumorné, K., D. C. Córdova, M. Astorga-Eló, and P. Renganathan, 2017, Extremozymes: A potential source for industrial applications: J Microbiol Biotechnol.

Dyshlovoy, S. A., J. Hauschild, K. Amann, K. M. Tabakmakher, S. Venz, R. Walther, A. G. Guzii, T. N. Makarieva, L. K. Shubina, S. N. Fedorov, V. A. Stonik, C. Bokemeyer, S. Balabanov, F. Honecker, and G. von Amsberg, 2015, Marine alkaloid Monanchocidin a overcomes drug resistance by induction of autophagy and lysosomal membrane permeabilization: Oncotarget, v. 6, p. 17328-41.

Elleuche, S., C. Schröder, K. Sahm, and G. Antranikian, 2014, Extremozymes--biocatalysts with unique properties from extremophilic microorganisms: Curr Opin Biotechnol, v. 29, p. 116-23.

Ferrer, M., O. V. Golyshina, T. N. Chernikova, A. N. Khachane, V. A. Martins Dos Santos, M. M. Yakimov, K. N. Timmis, and P. N. Golyshin, 2005, Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin: Chem Biol, v. 12, p. 895-904.

Fieseler, L., M. Horn, M. Wagner, and U. Hentschel, 2004, Discovery of the novel candidate phylum "*Poribacteria*" in marine sponges: Appl Environ Microbiol, v. 70, p. 3724-32.

Fleming, A., 2001, On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*. 1929: Bull World Health Organ, v. 79, p. 780-90.

Flores, G. E., I. D. Wagner, Y. Liu, and A. L. Reysenbach, 2012, Distribution, abundance, and diversity patterns of the thermoacidophilic "deep-sea hydrothermal vent euryarchaeota 2": Front Microbiol, v. 3, p. 47.

Fondi, M., V. Orlandini, E. Perrin, I. Maida, E. Bosi, M. C. Papaleo, L. Michaud, A. Lo Giudice, D. de Pascale, M. L. Tutino, P. Liò, and R. Fani, 2014, Draft genomes of three Antarctic *Psychrobacter* strains producing antimicrobial compounds against *Burkholderia cepacia* complex, opportunistic human pathogens: Mar Genomics, v. 13, p. 37-8.

Gazave, E., P. Lapebie, A. Ereskovsky, J. Vacelet, E. Renard, P. Cardenas, and C. Borchiellini, 2012, No longer Demospongiae: *Homoscleromorpha* formal nomination as a fourth class of *Porifera*.: Hydrobiologica, v. 687, p. 3-10.

Gesheva, V., and E. Vasileva-Tonkova, 2012, Production of enzymes and antimicrobial compounds by halophilic Antarctic *Nocardioides* sp. grown on different carbon sources: World J Microbiol Biotechnol, v. 28, p. 2069-76.

Goyal, K., P. Selvakumar, and K. Hayashi, 2001, Characterization of a thermostable beta-glucosidase (BglB) from *Thermotoga maritima* showing transglycosylation activity: Journal of Molecular Catalysis B: Enzymatic, v. 15, p. 45-53.

Guzii, A. G., T. N. Makarieva, V. A. Denisenko, P. S. Dmitrenok, A. S. Kuzmich, S. A. Dyshlovoy, V. B. Krasokhin, and V. A. Stonik, 2010, Monanchocidin: a new apoptosis-inducing polycyclic guanidine alkaloid from the marine sponge *Monanchora pulchra*: Org Lett, v. 12, p. 4292-5.

Handelsman, J., 2004, Metagenomics: application of genomics to uncultured microorganisms: Microbiol Mol Biol Rev, v. 68, p. 669-85.

Hentschel, U., L. Fieseler, M. Wehrl, C. Gernert, M. Steinert, J. Hacker, and M. Horn, 2003, Microbial diversity of marine sponges: Prog Mol Subcell Biol, v. 37, p. 59-88.

Hentschel, U., J. Piel, S. M. Degnan, and M. W. Taylor, 2012, Genomic insights into the marine sponge microbiome: Nat Rev Microbiol, v. 10, p. 641-54.

Hentschel, U., K. M. Usher, and M. W. Taylor, 2006, Marine sponges as microbial fermenters: FEMS Microbiol Ecol, v. 55, p. 167-77.

Hoeller, U., A. Wright, G. Matthee, G. Konig, S. Draeger, H. Aust, and B. Schulz, 2000, Fungi from marine sponges: diversity, biological activity and secondary metabolites: Mycological Research, v. 104, p. 1354-1365.

Hough, D. W., and M. J. Danson, 1999, Extremozymes: Curr Opin Chem Biol, v. 3, p. 39-46.

Hryniewicz-Jankowska, A., A. Czogalla, E. Bok, and A. F. Sikorsk, 2002, Ankyrins, multifunctional proteins involved in many cellular pathways: Folia Histochem Cytobiol, v. 40, p. 239-49.

Imhoff, J. F., 2016, Natural products from marine fungi--Still an underrepresented resource: Mar Drugs, v. 14, p. 19.

Indriani, I., 2007, Biodiversity of marine-derived fungi and identification of their metabolites, Heinrich-Heine-University Duesselsdorf.

Jackson, S. A., E. Borchert, F. O'Gara, and A. D. Dobson, 2015, Metagenomics for the discovery of novel biosurfactants of environmental interest from marine ecosystems: Curr Opin Biotechnol, v. 33, p. 176-82.

Jackson, S. A., B. Flemer, A. McCann, J. Kennedy, J. P. Morrissey, F. O'Gara, and A. D. Dobson, 2013, Archaea appear to dominate the microbiome of *Inflatella pellicula* deep sea sponges: PLoS One, v. 8, p. e84438.

Jiang, Z., W. Deng, Y. Zhu, L. Li, Y. Sheng, and K. Hayashi, 2004a, The recombinant xylanase B of *Thermotoga maritima* is highly xylan specific and produces exclusively xylobiose from xylans, a unique character for industrial applications: Journal of Molecular Catalysis B: Enzymatic, v. 27, p. 207-213.

Jiang, Z., Y. Zhu, L. Li, X. Yu, I. Kusakabe, M. Kitaoka, and K. Hayashi, 2004b, Transglycosylation reaction of xylanase B from the hyperthermophilic *Thermotoga maritima* with the ability of synthesis of tertiary alkyl beta-D-xylobiosides and xylosides: J Biotechnol, v. 114, p. 125-34.

Jiang, Z. D., P. R. Jensen, and W. Fenical, 1999, Lobophorins A and B, new antiinflammatory macrolides produced by a tropical marine bacterium: Bioorg Med Chem Lett, v. 9, p. 2003-6.

Jørgensen, B. B., and A. Boetius, 2007, Feast and famine--microbial life in the deep-sea bed: Nat Rev Microbiol, v. 5, p. 770-81.

Kamke, J., C. Rinke, P. Schwientek, K. Mavromatis, N. Ivanova, A. Sczyrba, T. Woyke, and U. Hentschel, 2014, The candidate phylum *Poribacteria* by single-cell genomics: new insights into phylogeny, cell-compartmentation, eukaryote-like repeat proteins, and other genomic features: PLoS One, v. 9, p. e87353.

Kennedy, J., B. Flemer, S. A. Jackson, D. P. Lejon, J. P. Morrissey, F. O'Gara, and A. D. Dobson, 2010, Marine metagenomics: new tools for the study and exploitation of marine microbial metabolism: Mar Drugs, v. 8, p. 608-28.

Kennedy, J., B. Flemer, S. A. Jackson, J. P. Morrissey, F. O'Gara, and A. D. Dobson, 2014, Evidence of a putative deep sea specific microbiome in marine sponges: PLoS One, v. 9, p. e91092.

Kennedy, J., J. R. Marchesi, and A. D. Dobson, 2007, Metagenomic approaches to exploit the biotechnological potential of the microbial consortia of marine sponges: Appl Microbiol Biotechnol, v. 75, p. 11-20.

Kennedy, J., N. D. O'Leary, G. S. Kiran, J. P. Morrissey, F. O'Gara, J. Selvin, and A. D. Dobson, 2011, Functional metagenomic strategies for the discovery of novel enzymes

and biosurfactants with biotechnological applications from marine ecosystems: J Appl Microbiol, v. 111, p. 787-99.

Keren, R., B. Mayzel, A. Lavy, I. Polishchuk, D. Levy, S. C. Fakra, B. Pokroy, and M. Ilan, 2017, Sponge-associated bacteria mineralize arsenic and barium on intracellular vesicles: Nat Commun, v. 8, p. 14393.

Kim, M. K., H. Park, and T. J. Oh, 2014, Antibacterial and antioxidant capacity of polar microorganisms isolated from Arctic lichen *Ochrolechia* sp.: Pol J Microbiol, v. 63, p. 317-22.

Lafi, F. F., J. A. Fuerst, L. Fieseler, C. Engels, W. W. Goh, and U. Hentschel, 2009, Widespread distribution of *Poribacteria* in demospongiae: Appl Environ Microbiol, v. 75, p. 5695-9.

Laport, M. S., O. C. Santos, and G. Muricy, 2009, Marine sponges: potential sources of new antimicrobial drugs: Curr Pharm Biotechnol, v. 10, p. 86-105.

Leary, D., M. Vierros, G. Hamon, S. Arico, and C. Monagle, 2009, Marine genetic resources: A review of scientific and commercial interest: Marine Policy, v. 33, p. 183-194.

Lebar, M. D., J. L. Heimbegner, and B. J. Baker, 2007, Cold-water marine natural products: Nat Prod Rep, v. 24, p. 774-97.

Lee, H. S., K. K. Kwon, S. G. Kang, S. S. Cha, S. J. Kim, and J. H. Lee, 2010, Approaches for novel enzyme discovery from marine environments: Curr Opin Biotechnol, v. 21, p. 353-7.

Levy, S. B., 2002, From tragedy the antibiotic era is born, The Antibiotic Paradox: How the misuse of antibiotics destroys their curative powers, v. 2nd ed., Perseus Publishing, p. 1–14.

Liu, F., J. Li, G. Feng, and Z. Li, 2016, New genomic insights into "*Entotheonella*" symbionts in *Theonella swinhoei*: Mixotrophy, anaerobic adaptation, resilience, and interaction: Front Microbiol, v. 7, p. 1333.

Lo Giudice, A., V. Bruni, and L. Michaud, 2007, Characterization of Antarctic psychrotrophic bacteria with antibacterial activities against terrestrial microorganisms: J Basic Microbiol, v. 47, p. 496-505.

Lossouarn, J., S. Dupont, A. Gorlas, C. Mercier, N. Bienvenu, E. Marguet, P. Forterre, and C. Geslin, 2015, An abyssal mobilome: viruses, plasmids and vesicles from deep-sea hydrothermal vents: Res Microbiol, v. 166, p. 742-52.

Makarieva, T. N., K. M. Tabakmaher, A. G. Guzii, V. A. Denisenko, P. S. Dmitrenok, L. K. Shubina, A. S. Kuzmich, H. S. Lee, and V. A. Stonik, 2011, Monanchocidins B-E: polycyclic guanidine alkaloids with potent antileukemic activities from the sponge *Monanchora pulchra*: J Nat Prod, v. 74, p. 1952-8.

McAuliffe, O., R. P. Ross, and C. Hill, 2001, Lantibiotics: structure, biosynthesis and mode of action: FEMS Microbiol Rev, v. 25, p. 285-308.

Mehbub, M. F., J. Lei, C. Franco, and W. Zhang, 2014, Marine sponge derived natural products between 2001 and 2010: trends and opportunities for discovery of bioactives: Mar Drugs, v. 12, p. 4539-77.

Mesbah, N. M., and F. Sarmiento, 2016, Editorial: Enzymes from extreme environments: Front Bioeng Biotechnol, v. 4, p. 24.

Moitinho-Silva, L., K. Bayer, C. V. Cannistraci, E. C. Giles, T. Ryu, L. Seridi, T. Ravasi, and U. Hentschel, 2014, Specificity and transcriptional activity of microbiota associated with low and high microbial abundance sponges from the Red Sea: Mol Ecol, v. 23, p. 1348-63.

Morrow, C., and P. Cárdenas, 2015, Proposal for a revised classification of the Demospongiae (*Porifera*): Front Zool, v. 12, p. 7.

Naganuma, T., 2000, [Microbes on the edge of global biosphere]: Biol Sci Space, v. 14, p. 323-31.

Nigam, P. S., 2013, Microbial enzymes with special characteristics for biotechnological applications: Biomolecules, v. 3, p. 597-611.

Ohizumi, Y., S. Sasaki, T. Kusumi, and I. I. Ohtani, 1996, Ptilomycalin A, a novel Na+, K(+)- or Ca2(+)-ATPase inhibitor, competitively interacts with ATP at its binding site: Eur J Pharmacol, v. 310, p. 95-8.

Orlandini, V., I. Maida, M. Fondi, E. Perrin, M. C. Papaleo, E. Bosi, D. de Pascale, M. L. Tutino, L. Michaud, A. Lo Giudice, and R. Fani, 2014, Genomic analysis of three sponge-associated *Arthrobacter* Antarctic strains, inhibiting the growth of *Burkholderia*

*cepacia* complex bacteria by synthesizing volatile organic compounds: Microbiol Res, v. 169, p. 593-601.

Pan, H. Q., S. Y. Zhang, N. Wang, Z. L. Li, H. M. Hua, J. C. Hu, and S. J. Wang, 2013, New spirotetronate antibiotics, lobophorins H and I, from a South China Sea-derived *Streptomyces* sp. 12A35: Mar Drugs, v. 11, p. 3891-901.

Papaleo, M. C., M. Fondi, I. Maida, E. Perrin, A. Lo Giudice, L. Michaud, S. Mangano, G. Bartolucci, R. Romoli, and R. Fani, 2012, Sponge-associated microbial Antarctic communities exhibiting antimicrobial activity against *Burkholderia cepacia* complex bacteria: Biotechnol Adv, v. 30, p. 272-93.

Pawlik, J. R., G. McFall, and S. Zea, 2002, Does the odor from sponges of the genus *Ircinia* protect them from fish predators?: J Chem Ecol, v. 28, p. 1103-15.

Paz, Z., M. Komon-Zelazowska, I. Druzhinina, M. Aveskamp, A. Shnaiderman, Y. Aluma, S. Carmeli, M. Ilan, and O. Yarden, 2010, Diversity and potential antifungal properties of fungi associated with a Mediterranean sponge: Fungal Diversity, v. 42, p. 17-26.

Phelan, R. W., M. Barret, P. D. Cotter, P. M. O'Connor, R. Chen, J. P. Morrissey, A. D. Dobson, F. O'Gara, and T. M. Barbosa, 2013, Subtilomycin: a new lantibiotic from *Bacillus subtilis* strain MMA7 isolated from the marine sponge *Haliclona simulans*: Mar Drugs, v. 11, p. 1878-98.

Piel, J., D. Hui, G. Wen, D. Butzke, M. Platzer, N. Fusetani, and S. Matsunaga, 2004, Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*: Proc. Natl. Acad. Sci. U.S.A., v. 101, p. 16222-16227.

Proksch, P., 1994, Defensive roles for secondary metabolites from marine sponges and sponge-feeding nudibranchs: Toxicon, v. 32, p. 639-55.

Ramirez-Llodra, E., A. Brandt, R. Danovaro, B. De Mol, E. Escobar, C. German, L. Levin, P. Arbizu, L. Menot, and P. Buhl-Mortensen, 2010, Deep, diverse and definitely different: unique attributes of the world's largest ecosystem: Biogeosciences, v. 7, p. p.2851-2899.

Reiswig, H., 1971, Particle feeding in natural populations of three marine demosponges: The Biological Bulletin, v. 141, p. 568-591.

Reynolds, D., and T. Thomas, 2016, Evolution and function of eukaryotic-like proteins from sponge symbionts: Mol Ecol, v. 25, p. 5242-5253.

Sagar, S., M. Kaur, and K. P. Minneman, 2010, Antiviral lead compounds from marine sponges: Mar Drugs, v. 8, p. 2619-38.

Santiago, M., C. A. Ramírez-Sarmiento, R. A. Zamora, and L. P. Parra, 2016, Discovery, molecular mechanisms, and industrial applications of cold-active enzymes: Front Microbiol, v. 7, p. 1408.

Schloss, P. D., and J. Handelsman, 2003, Biotechnological prospects from metagenomics: Curr Opin Biotechnol, v. 14, p. 303-10.

Schmeisser, C., C. Stöckigt, C. Raasch, J. Wingender, K. N. Timmis, D. F. Wenderoth, H. C. Flemming, H. Liesegang, R. A. Schmitz, K. E. Jaeger, and W. R. Streit, 2003, Metagenome survey of biofilms in drinking-water networks: Appl Environ Microbiol, v. 69, p. 7298-309.

Schneider, K., S. Keller, F. E. Wolter, L. Röglin, W. Beil, O. Seitz, G. Nicholson, C. Bruntner, J. Riedlinger, H. P. Fiedler, and R. D. Süssmuth, 2008, Proximicins A, B, and C-antitumor furan analogues of netropsin from the marine actinomycete *Verrucosispora* induce upregulation of p53 and the cyclin kinase inhibitor p21: Angew Chem Int Ed Engl, v. 47, p. 3258-61.

Selvin, J., J. Kennedy, D. P. Lejon, G. S. Kiran, and A. D. Dobson, 2012, Isolation identification and biochemical characterization of a novel halo-tolerant lipase from the metagenome of the marine sponge *Haliclona simulans*: Microb Cell Fact, v. 11, p. 72.

Shymanska, N. V., I. H. An, and J. G. Pierce, 2014, A rapid synthesis of 4-oxazolidinones: total synthesis of synoxazolidinones A and B: Angew Chem Int Ed Engl, v. 53, p. 5401-4.

Siegl, A., J. Kamke, T. Hochmuth, J. Piel, M. Richter, C. Liang, T. Dandekar, and U. Hentschel, 2011, Single-cell genomics reveals the lifestyle of *Poribacteria*, a candidate phylum symbiotically associated with marine sponges: ISME J, v. 5, p. 61-70.

Simister, R. L., P. Deines, E. S. Botté, N. S. Webster, and M. W. Taylor, 2012, Sponge-specific clusters revisited: a comprehensive phylogeny of sponge-associated microorganisms: Environ Microbiol, v. 14, p. 517-24.

Singh, B. K., 2010, Exploring microbial diversity for biotechnology: the way forward: Trends Biotechnol, v. 28, p. 111-6.

Sipkema, D., M. C. Franssen, R. Osinga, J. Tramper, and R. H. Wijffels, 2005, Marine sponges as pharmacy: Mar Biotechnol (NY), v. 7, p. 142-62.

Sogin, M. L., H. G. Morrison, J. A. Huber, D. Mark Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, and G. J. Herndl, 2006, Microbial diversity in the deep sea and the underexplored "rare biosphere": Proc Natl Acad Sci U S A, v. 103, p. 12115-20.

Steinert, G., C. Huete-Stauffer, N. Aas-Valleriani, E. Borchert, A. Bhushan, A. Campbell, M. Mares, A. Costa, J. Gutleben, S. Knobloch, R. Lee, S. Munroe, D. Naik, E. Peters, E. Stokes, W. Wang, E. Einarsd óttír, and D. Sipkema, 2017, BluePharmTrain - A European Sponge Biotechnology Project, Grand Challenges in Marine Biotechnology: Springer Verlag, Springer, Berlin Heidelberg.

Streit, W. R., and R. A. Schmitz, 2004, Metagenomics--the key to the uncultured microbes: Curr Opin Microbiol, v. 7, p. 492-8.

Sánchez, L. A., M. Hedström, M. A. Delgado, and O. D. Delgado, 2010, Production, purification and characterization of serraticin A, a novel cold-active antimicrobial produced by *Serratia proteamaculans* 136: J Appl Microbiol, v. 109, p. 936-45.

Tadesse, M., M. B. Strøm, J. Svenson, M. Jaspars, B. F. Milne, V. Tørfoss, J. H. Andersen, E. Hansen, K. Stensvåg, and T. Haug, 2010, Synoxazolidinones A and B: novel bioactive alkaloids from the ascidian *Synoicum pulmonaria*: Org Lett, v. 12, p. 4752-5.

Tadesse, M., J. N. Tabudravu, M. Jaspars, M. B. Strøm, E. Hansen, J. H. Andersen, P. E. Kristiansen, and T. Haug, 2011, The antibacterial ent-eusynstyelamide B and eusynstyelamides D, E, and F from the Arctic bryozoan *Tegella cf. spitzbergensis*: J Nat Prod, v. 74, p. 837-41.

Tapiolas, D. M., B. F. Bowden, E. Abou-Mansour, R. H. Willis, J. R. Doyle, A. N. Muirhead, C. Liptrot, L. E. Llewellyn, C. W. Wolff, A. D. Wright, and C. A. Motti, 2009, Eusynstyelamides A, B, and C, nNOS inhibitors, from the ascidian *Eusynstyela latericius*: J Nat Prod, v. 72, p. 1115-20.

Taylor, M. W., R. Radax, D. Steger, and M. Wagner, 2007, Sponge-associated microorganisms: evolution, ecology, and biotechnological potential: Microbiol Mol Biol Rev, v. 71, p. 295-347.

Taylor, M. W., P. Tsai, R. L. Simister, P. Deines, E. Botte, G. Ericson, S. Schmitt, and N. S. Webster, 2013, 'Sponge-specific' bacteria are widespread (but rare) in diverse marine environments: ISME J, v. 7, p. 438-43.

Ternon, E., L. Zarate, S. Chenesseau, J. Croué, R. Dumollard, M. T. Suzuki, and O. P. Thomas, 2016, Spherulization as a process for the exudation of chemical cues by the encrusting sponge *C. crambe*: Sci Rep, v. 6, p. 29474.

Thomas, T., L. Moitinho-Silva, M. Lurgi, J. R. Björk, C. Easson, C. Astudillo-García, J. B. Olson, P. M. Erwin, S. López-Legentil, H. Luter, A. Chaves-Fonnegra, R. Costa, P. J. Schupp, L. Steindler, D. Erpenbeck, J. Gilbert, R. Knight, G. Ackermann, J. Victor Lopez, M. W. Taylor, R. W. Thacker, J. M. Montoya, U. Hentschel, and N. S. Webster, 2016, Diversity, structure and convergent evolution of the global sponge microbiome: Nat Commun, v. 7, p. 11870.

Thomas, T., D. Rusch, M. Z. DeMaere, P. Y. Yung, M. Lewis, A. Halpern, K. B. Heidelberg, S. Egan, P. D. Steinberg, and S. Kjelleberg, 2010, Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis: ISME J, v. 4, p. 1557-67.

Trepos, R., G. Cervin, C. Hellio, H. Pavia, W. Stensen, K. Stensvåg, J. S. Svendsen, T. Haug, and J. Svenson, 2014, Antifouling compounds from the sub-arctic ascidian *Synoicum pulmonaria*: synoxazolidinones A and C, pulmonarins A and B, and synthetic analogues: J Nat Prod, v. 77, p. 2105-13.

Trincone, A., 2010, Potential biocatalysts originating from sea environments: Journal of Molecular Catalysis B: Enzymatic, v. 66, p. 241-256.

Trincone, A., 2011, Marine biocatalysts: enzymatic features and applications: Mar Drugs, v. 9, p. 478-99.

Turon, X., M. Becerro, and M. Uriz, 2000, Distribution of brominated compounds within the sponge *Aplysina aerophoba*: coupling of X-ray microanalysis with cryofixation techniques: Cell and Tissue Research, v. 301, p. 311-322.

Uchiyama, T., and K. Miyazaki, 2009, Functional metagenomics for enzyme discovery: challenges to efficient screening: Curr Opin Biotechnol, v. 20, p. 616-22.

Unson, M., N. Holland, and D. Faulkner, 1994, A brominated secondary metabolite synthesized by the cyanobacterial symbiont of a marine sponge and accumulation of the crystalline metabolite in the sponge tissue: Marine Biology, v. 119, p. 1-11.

Uriz, M., X. Turon, J. Galera, and J. Tur, 1996, New light on the cell location of avarol within the sponge *Dysidea avara* (Dendroceratida): Cell and Tissue Research, v. 285, p. 519-527.

Uriz, M. J., X. Turon, M. A. Becerro, and G. Agell, 2003, Siliceous spicules and skeleton frameworks in sponges: origin, diversity, ultrastructural patterns, and biological functions: Microsc Res Tech, v. 62, p. 279-99.

Vacelet, J., and N. Boury-Esnault, 1995, Carnivorous sponges: Nature, v. 373, p. 333-335.

Vacelet, J., and C. Donadey, 1977, Electron microscope study of the association between some sponges and bacteria: Journal of Experimental Marine Biology and Ecology, v. 30, p. 301-314.

Van Soest, R. W., N. Boury-Esnault, J. Vacelet, M. Dohrmann, D. Erpenbeck, N. J. De Voogd, N. Santodomingo, B. Vanhoorne, M. Kelly, and J. N. Hooper, 2012, Global diversity of sponges (Porifera): PLoS One, v. 7, p. e35105.

Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. H. Rogers, and H. O. Smith, 2004, Environmental genome shotgun sequencing of the Sargasso Sea: Science, v. 304, p. 66-74.

Vieites, J. M., M. E. Guazzaroni, A. Beloqui, P. N. Golyshin, and M. Ferrer, 2009, Metagenomics approaches in systems microbiology: FEMS Microbiol Rev, v. 33, p. 236-55.

Vieweg, L., S. Reichau, R. Schobert, P. F. Leadlay, and R. D. Süssmuth, 2014, Recent advances in the field of bioactive tetronates: Nat Prod Rep, v. 31, p. 1554-84.

Wang, Q., F. Song, X. Xiao, P. Huang, L. Li, A. Monte, W. M. Abdel-Mageed, J. Wang, H. Guo, W. He, F. Xie, H. Dai, M. Liu, C. Chen, H. Xu, A. M. Piggott, X. Liu, R. J. Capon, and L. Zhang, 2013, Abyssomicins from the South China Sea deep-sea sediment *Verrucosispora* sp.: natural thioether Michael addition adducts as antitubercular prodrugs: Angew Chem Int Ed Engl, v. 52, p. 1231-4.

Wehrl, M., M. Steinert, and U. Hentschel, 2007, Bacterial uptake by the marine sponge *Aplysina aerophoba*: Microb Ecol, v. 53, p. 355-65.

Wilkinson, C., 1984, Immunological evidence for the precambrian origin of bacterial symbioses in marine sponges: Proceedings of the Royal Society, v. 220.

Wilson, M. C., T. Mori, C. Rückert, A. R. Uria, M. J. Helf, K. Takada, C. Gernert, U. A. Steffens, N. Heycke, S. Schmitt, C. Rinke, E. J. Helfrich, A. O. Brachmann, C. Gurgui, T. Wakimoto, M. Kracht, M. Crüsemann, U. Hentschel, I. Abe, S. Matsunaga, J. Kalinowski, H. Takeyama, and J. Piel, 2014, An environmental bacterial taxon with a large and distinct metabolic repertoire: Nature, v. 506, p. 58-62.

Wyche, T. P., Y. Hou, D. Braun, H. C. Cohen, M. P. Xiong, and T. S. Bugni, 2011, First natural analogs of the cytotoxic thiodepsipeptide thiocoraline A from a marine *Verrucosispora* sp: J Org Chem, v. 76, p. 6542-7.

Xiao, J., Y. Luo, S. Xie, and J. Xu, 2011, *Serinicoccus profundi* sp. nov., an actinomycete isolated from deep-sea sediment, and emended description of the genus *Serinicoccus*: Int J Syst Evol Microbiol, v. 61, p. 16-9.

Yang, J. S., B. Lu, D. F. Chen, Y. Q. Yu, F. Yang, H. Nagasawa, S. Tsuchida, Y. Fujiwara, and W. J. Yang, 2013a, When did decapods invade hydrothermal vents? Clues from the Western Pacific and Indian Oceans: Mol Biol Evol, v. 30, p. 305-9.

Yang, X. W., G. Y. Zhang, J. X. Ying, B. Yang, X. F. Zhou, A. Steinmetz, Y. H. Liu, and N. Wang, 2013b, Isolation, characterization, and bioactivity evaluation of 3-((6-methylpyrazin-2-yl)methyl)-1H-indole, a new alkaloid from a deep-sea-derived actinomycete *Serinicoccus profundi* sp. nov: Mar Drugs, v. 11, p. 33-9.

Yin, Z., M. Zhu, E. H. Davidson, D. J. Bottjer, F. Zhao, and P. Tafforeau, 2015, Sponge grade body fossil with cellular resolution dating 60 Myr before the Cambrian: Proc Natl Acad Sci U S A, v. 112, p. E1453-60.

Zhang, H., J. A. White-Phillip, C. E. Melançon, H. J. Kwon, W. L. Yu, and H. W. Liu, 2007, Elucidation of the kijanimicin gene cluster: insights into the biosynthesis of spirotetronate antibiotics and nitrosugars: J Am Chem Soc, v. 129, p. 14670-83.

Zhu, D., H. Malik, and L. Hua, 2006, Asymmetric ketone reduction by a hyperthermophilic alcohol dehydrogenase. The substrate specificity, enantioselectivity and tolerance of organic solvents: Tetrahedron: Asymmetry, v. 17, p. 3010-3014.

Zierenberg, R. A., M. W. Adams, and A. J. Arp, 2000, Life in extreme environments: hydrothermal vents: Proc Natl Acad Sci U S A, v. 97, p. 12961-2.

# Chapter 2

## Diversity of natural product biosynthetic genes in the microbiome of the deep sea sponges *Inflatella pellicula*, *Poecillastra compressa* and *Stelletta normani*

**Erik Borchert[1], Stephen A. Jackson[1], Fergal O'Gara[1,2,3], Alan D.W. Dobson[1,4]**

[1] School of Microbiology, University College Cork, National University of Ireland, Cork, Ireland

[2] Biomerit Research Centre, University College Cork, National University of Ireland, Cork, Ireland

[3] School of Biomedical Sciences, Curtin Health Innovation Research Institute, Curtin University, Perth, Australia.

[4] Environmental Research Institute, University College Cork, National University of Ireland, Cork, Ireland

Published in the Journal Frontiers in Microbiology, 2016, 7: 1027

## 2.1 Abstract

Three different deep sea sponge species, *Inflatella pellicula*, *Poecillastra compressa* and *Stelletta normani* comprising of seven individual samples, retrieved from depths of 760 to 2900 m below sea level, were investigated using 454 pyrosequencing for their secondary metabolomic potential targeting adenylation domain and ketosynthase domain sequences. The data obtained suggest a diverse microbial origin of nonribosomal peptide synthetases and polyketide synthase fragments, that in part correlates with their respective microbial community structures that were previously described and reveals an untapped source of potential novelty. The sequences, especially the ketosynthase fragments, display extensive clade formations which are clearly distinct from sequences hosted in public databases, therefore highlighting the potential of the microbiome of these deep sea sponges to produce potentially novel small molecule chemistry. Furthermore sequence similarities to gene clusters known to be involved in the production of many classes of antibiotics and toxins including lipopeptides, glycopeptides, macrolides and hepatotoxins were also identified.

## 2.2 Introduction

Marine sponges (*Porifera*) are important members of marine benthic communities in our oceans, and continue to attract attention due to their remarkably diverse bacterial, archaeal and eukaryotic microbial community structures (Webster and Taylor, 2012), and their importance as a source of novel natural products. Many of the sponge-microbial associates are symbionts involved in nutrient cycling and may also play a role in the sponge's chemical defence mechanisms (Taylor *et al.*, 2007; Bell, 2008; Webster *et al.*, 2010; Hentschel *et al.*, 2012). Sponges are typically sessile filter feeders, filtering large quantities of seawater which contains microbes and viruses that are potentially harmful to the sponge. Thus the ability of members of their microbial communities to produce secondary metabolites with the potential to augment the sponges own chemical defence mechanisms is likely to be advantageous. Sponges are one of the oldest extant metazoans on earth and appear to be obligatorily associated with their bacterial endosymbiotic communities. It is reasonable to expect divergent evolution of ancestral genes among these endosymbionts to the extent that the resulting gene products are likely to be significantly different to those of a

terrestrial origin. This is likely to be particularly true of the endosymbionts of deep-sea sponges which have been exposed to extremes of temperature, salinity and pressure for many millions of years. The adaptation of sponge endosymbionts to these extreme conditions can be expected to also have been facilitated by increased horizontal gene transfer frequencies that are known to be high amongst marine microbial communities, resulting in increases in the genomic flexibility within these bacterial populations (Penn *et al.*, 2009; Sobecky and Hazen, 2009; McDaniel *et al.*, 2010).

Numerous studies have been undertaken to date to investigate the microbial ecology and the biological potential of marine shallow water habitats (Aylward *et al.*, 2015). In marked contrast even though our oceans have a mean depth of 3800 m, with 50% being below 3000 m deep, deep-sea marine environments have only rarely been explored with respect to their potential to genetically encode secondary metabolites of clinical or industrial utility (Ramirez-Llodra *et al.*, 2010). This lack of exploration is most likely due to the technical difficulties and costs associated with sampling at lower depths, with only 5% of the "deep sea" having to date been explored with remote instrumentation (Ramirez-Llodra *et al.*, 2010). Therefore it can be assumed that to date we have only "scratched the surface" of the true biotechnological potential of our oceans, particularly the deep sea.

The identification of novel bioactive compounds and the metabolic potential of microbial communities from various terrestrial or marine habitats have mostly been investigated using a variety of different approaches including direct chemical extraction methods, enhancing cultivability of microorganisms (Sipkema *et al.*, 2011), and testing of isolated microorganisms (Gurgui and Piel, 2010). Novel natural products from the marine environment include, for example, new antimicrobial agents (Jang *et al.*, 2013), novel bioactive compounds (Reen *et al.*, 2015), antifouling agents (Fusetani, 2011) and various enzymes of industrial interest (Satpute *et al.*, 2010; Jackson *et al.*, 2015). However the overall diversity of the secondary metabolite biosynthetic potential present within these environments is difficult to assess given that the majority of bacteria are not readily cultured using currently available microbiological methods (Uria and Piel, 2009).

Polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) gene clusters encode for modular arrangements of different enzymes that are able to extend,

modify, connect and alter a variety of substrates to produce unique compounds with specific enzymatic, chemical or antimicrobial properties (Hertweck, 2009; Khosla, 2009; Helfrich *et al.*, 2014). Each PKS or NRPS gene cluster produces a specific secondary metabolite and the presence of diversity in these gene clusters is indicative of diverse secondary metabolism products. The conserved nature of PKS and NRPS allows the design of degenerate primers to target specific domains which these gene clusters have in common, such as ketosynthase domains in PKS or adenylation and condensation domains in NRPS clusters (Reddy *et al.*, 2012; Woodhouse *et al.*, 2013; Charlop Powers *et al.*, 2014). To assess these clusters and to help overcome the problems associated with culture dependent approaches, efforts have focused on the analysis of community DNA isolated directly from the environment in question, which can provide a means of exploring their secondary metabolic potential (Trindade-Silva *et al*., 2013; Woodhouse *et al*., 2013). Nonetheless, to date, only a few studies have been published which have investigated the secondary metabolic potential of a mixed microbial community using next generation sequencing (NGS) technologies. The resultant sequencing depths have the potential to reveal the entire secondary metabolomic potential of a microbial cohort, something not achievable prior to the advent of NGS. Previous NGS studies targeting secondary metabolism genes have focused on soils (Reddy *et al.*, 2012; Charlop-Powers *et al.*, 2014), and marine sponges (Woodhouse *et al.*, 2013). NGS technologies have to date been primarily used to study microbial abundance via 16S rRNA gene sequencing (Sogin *et al.*, 2006). In contrast, clone libraries, functional metagenomic libraries and comparable techniques have been used to target secondary metabolite gene clusters to estimate the potential of a given microbial community (Rocha-Martin *et al.*, 2014). Reddy *et al.*, 2012, investigated three geographically distinct soil samples and found comparably similar distribution of major bacterial phyla in those soils using 16S rRNA gene analysis, but almost completely distinct sets of secondary metabolite biosynthetic gene sequences. In that study they investigated the presence of specific parts of PKS, NRPS and PKS/NRPS hybrid clusters, namely the ketosynthase domain (KS) of Type I PKS and the adenylation domain (AD) of NRPS clusters. The primers they used were designed to amplify conserved regions of these domains, including the catalytic active site and yielded a PCR product of approximately 795 bp and 760 bp for AD and KS domains respectively

(Ayuso-Sacido and Genilloud, 2005; Schirmer *et al.*, 2005), which correlates with the expected average size of 454 pyrosequencing reads.

We have previously investigated the microbial diversity of the deep sea sponges *Inflatella pellicula*, *Poecillastra compressa* and *Stelletta normani* by 16S rRNA gene pyrosequencing and found that they contained diverse bacteria and archaea, with *I. pellicula* in particular being dominated by archaea (Jackson *et al.*, 2013, Kennedy *et al.*, 2014). Here we investigate the potential for secondary metabolite production of the microbiome of these deep sea sponges to produce novel natural products, utilizing 454 pyrosequencing, targeting PKS and NRPS gene clusters, using the aforementioned Reddy *et al.* PCR primer sets. We report that the microbial communities associated with these deep sea sponges do indeed harbor a wide variety of these genes. The results clearly show relatedness to genes that are involved in the synthesis of known classes of bioactive compounds, for example lipopeptides, glycopeptides, macrolides and hepatotoxins. However, and importantly, there is also a large proportion of comparably different sequences which are only distantly related to domains from known Type I PKS and NRPS sequences.

## 2.3 Materials and methods

### 2.3.1 Sample collection

Sponge samples ($n$ = 7) of the species *Stelletta normani*, *Inflatella pellicula* and *Poecillastra compressa* were collected in Irish territorial waters off the west coast of Ireland using the remotely operated vehicle (R.O.V.) *Holland I* during the Biodiscovery cruises 2010 (2 x *I. pellicula*, 1 x *S. normani* and 1 x *P. compressa*) and 2013 (2 x *S. normani* and 1 x *P. compressa*) aboard the R.V. *Celtic Explorer* (Table 1). After collection the samples were rinsed with sterile artificial seawater (3.33% (*w*/*v*) Instant Ocean, Aquarium Systems) to remove exogenous materials and stored at -80°C until further processing.

**Table 1:** Sample collection data (*samples also used in Jackson *et al.* 2013 and Kennedy *et al.* 2014 to generate 16S rRNA data)

| Sample | ID | latitude | longitude | Depths [m] |
|---|---|---|---|---|
| *I. pellicula** | BD226 | 54.2419 | -12.6938 | 2900 |
| *I. pellicula** | BD92 | 54.0015 | -12.3100 | 748 |
| *S. normani** | BD243 | 54.0015 | -12.3100 | 1350 |
| *S. normani* | BDV1267 | 54.0500 | -12.5333 | 2400 |
| *S. normani* | BDV1379 | 53.9861 | -12.6100 | 760 |
| *P. compressa** | BD130 | 54.0633 | -12.4131 | 1469 |
| *P. compressa* | BDV1346 | 54.0500 | -12.5833 | 1250 |

### 2.3.2 Metagenomic DNA extraction and purification

Frozen sponge tissues of all samples were ground in a sterile mortar with a pestle under liquid nitrogen. The obtained ground tissue was suspended in lysis buffer (100 mM Tris, 100 mM EDTA, 1.5 M NaCl (*w/v*), 1% CTAB (*w/v*), 2% SDS (*w/v*) in a 1:5 ratio and subsequently incubated for two hours at 70°C (Kennedy *et al.*, 2008). Solution was centrifuged until a clear solution was obtained. Afterwards DNA was precipitated using 0.7 volumes Isopropanol for 30 min at room temperature., followed by centrifugation 6000g, 30 min. Supernatant was discarded, pellet was washed with 70% Ethanol, centrifuged again, after supernatant removal air dried and finally resuspended in an appropriate amount of Tris-EDTA buffer (10mM Tris, 1mM EDTA, pH 8.0). The metagenomic DNA was then analyzed by gel electrophoresis, spectrophotometrically quantified (NanoDrop ND-1000) and stored at -20°C until usage.

### 2.3.3 PCR amplicon generation

Primer design was adapted from Reddy *et al.* (Reddy *et al.*, 2012). In short, each primer consists of a 454 sequencing adaptor, a unique 10 bp identifier tag to allow for sequencing of different amplicons/genes in the same region of a 454 plate and degenerate target sequence to either amplify a fragment (approximately 795 bp) of a conserved region in NRPS adenylation domains (AD) or a fragment (approx. 760 bp) of a ketosynthase domain (KS) from type I polyketide synthases (PKS) (see S1 table).

For the amplification of AD gene fragments from seven samples three different PCR conditions were used. The first reaction mixture (50 µl) comprised of 10 ng DNA, 0.5 µM each primer, 200 µM deoxynucleoside triphosphate (dNTP), 1x Q5 reaction buffer (New England Biolabs) and 1 U Q5 Hot start DNA polymerase (New England Biolabs). PCR amplification conditions for mix one were 35 cycles of 98°C for 10 s, 70°C for 30 s, 72°C for 30 s, followed by a final extension at 72°C for 3 min. The second mix contained 1x Phusion Buffer (New England Biolabs), 10 ng DNA, 200 µM dNTPs, 0.5 µM each primer and 1 U Phusion polymerase. PCR amplification from the second reaction mixture comprised 30 cycles of 98°C for 10 s, 68°C for 30 s, 72°C for 30 s and a final extension step at 72°C for 5 min. The third mix included 1x Failsafe Buffer E (Epicentre, FailSafe PCR System) 10 ng DNA, 200 µM dNTPs, 0.5 µM each primer and 2.5 U DreamTaq DNA polymerase (ThermoFisher Scientific). Third mix amplification was as follows: 35 cycles of 95°C for 60 s, 60°C for 60 s, 72°C for 2 min and a final extension at 72°C for 10 min.

For the amplification of KS fragments from five samples only one PCR mix was employed, which is similar to the third mix from the AD amplification, except that buffer E was replaced with buffer F from the FailSafe PCR System. Conditions for amplification were as follows: 35 cycles of 95°C for 40 s, 50°C for 40 s, 72°C for 75 s and a final extension for 5 min at 72°C (Brady *et al.* 2007). All samples were used for AD amplification, but only five for KS amplification (all three *S. normani*, one *I. pellicula* and one *P. compressa* (2010 Cruise) sample.

### 2.3.4 Pyrosequencing and data processing

The amplicons were gel purified and quantified using a spectrophotometer (NanoDrop ND-1000) and a fluorometer (Qubit™ Fluorometer Invitrogen). For library preparation, amplicons generated from all twelve samples were pooled in a single sample to a final concentration of $1.26 \times 10^9$ molecules/µl and pyrosequenced on $1/8^{th}$ of a plate for a 454 GS-FLX+ (Macrogen Inc.) sequencing run. The resulting sequences were quality filtered by removal of low quality (mean quality score below 25), short (less than 150 bp), homopolymer (limit of 6) and ambiguous reads (read contains more than 6 ambiguous

bases) and sorted by sample species using QIIME (Caporaso *et al.*, 2010). MG-RAST (Meyer *et al.*, 2008) was used to dereplicate the quality filtered reads, resulting in deletion of 56.9% of AD and 68.6% of KS sequences respectively. Manually constructed and publicly available reference sequence databases were used to sort/identify the quality filtered sequences using QIIME and NaPDos (e-Value Cutoff of $1e^{-5}$ and minimum match length of 100 aa) (Ziemert *et al.*, 2012). Manually constructed reference databases were established by screening the NCBI database for primer targets and screening known secondary metabolites gene clusters for primer binding sites and by confirming that the adjacent sequences were either ketosynthase or adenylation domains. In this way each reference data set comprised 30 to 40 unique sequences, which were then used in QIIME to pick reference OTUs (pick_open_reference_otus.py) using the UCLUST algorithm (Edgar, 2010) with pre-clustering at 60% identity to the references. The resultant representative OTUs were analyzed using MEGA, iTOL (Letunic and Bork, 2007, 2011) and MG-RAST (Meyer *et al.*, 2008). The NaPDos tool was used to compare the obtained representative KS OTUs to sequences deposited in this database and to calculate phylogenetic trees, later visualized by iTOL. Representative sequences were also checked manually by using the BLAST algorithm against the NCBI database to exclude unwanted sequences, for example fatty acid production affiliated sequences, and to verify the AD and KS domain character of the sequence reads. The data (raw reads) is deposited in the NCBI Sequence Read Archive (SRA) database under the accession number SRP070811.

**2.4 Results:**

The 454 pyrosequencing resulted in 109,079 reads of which 57,993 passed quality filtering and were subsequently analyzed downstream. Of these 57,993 sequences, 2,385 reads account for AD domain sequences and 55,608 reads for KS domain sequences. Dereplication using MG-RAST (Gomez-Alvarez *et al.*, 2009) resulted in 15,865 unique sequences, 1,621 AD reads and 14,244 KS reads respectively. The average length of the remaining sequences after dereplication was 398±205 bp (AD) and 473±168 bp (KS). A breakdown of the numbers of sequences included for further analysis and the representative sequences are provided in Table 2. Chao1 and Shannon diversity estimates were calculated using QIIME with 3% divergence and are listed in Table 3 for each individual sample.

**Table 2:** Breakdown of retrieved sequences after quality control and number of picked reference OTUs

| Species | No. of sequences | Average length | GC content | No. of reads after dereplication | No. of rep. OTUs |
|---|---|---|---|---|---|
| *I. pellicula* AD | 760 | 427 bp | 66.1% | 351 | 35 |
| *P. compressa* AD | 688 | 249 bp | 62.8% | 664 | 14 |
| *S. normani* AD | 937 | 485 bp | 67.8% | 606 | 31 |
| *I. pellicula* KS | 10227 | 505 bp | 53.1% | 3125 | 72 |
| *P. compressa* KS | 8167 | 467 bp | 49.6% | 2514 | 50 |
| *S. normani* KS | 37214 | 485 bp | 57.1% | 8605 | 109 |

**Table 3:** Chao1 and Shannon diversity indices

| Sample | Chao1 | Shannon |
|---|---|---|
| *I. pellicula* A AD | 4.0 | 0.87 |
| *I. pellicula* B AD | 31.0 | 4.89 |
| *P. compressa* A AD | 13.0 | 2.19 |
| *P. compressa* B AD | 1.0 | 0 |
| *S. normani* A AD | 12.33 | 2.95 |
| *S. normani* B AD | 19.3 | 4.16 |
| *S. normani* C AD | 8.0 | 2.88 |
| *I. pellicula* B KS | 76.0 | 5.92 |
| *P. compressa* A KS | 50.0 | 4.98 |
| *S. normani* A KS | 56.12 | 5.08 |
| *S. normani* B KS | 5.0 | 2.19 |
| *S. normani* C KS | 59.13 | 5.70 |

The taxonomic abundances were calculated by MG-RAST after dereplication of the quality filtered reads. The most dominant phylogenetic assignations in the AD sequences comprise of *Proteobacteria*, *Cyanobacteria*, *Firmicutes*, *Actinobacteria*, *Verrucomicrobia* and *Chloroflexi* (Figure 1). *Proteobacteria* account for 49% of the sequences from *I. pellicula*, 53% from *P. compressa* and for 43% from *S. normani* and is therefore the most abundant phylum contributing AD sequences in all three sponge species. A difference in the abundances is observable in the amount of cyanobacterial (0.87%) and *Chloroflexi* (7.82%) affiliated sequences in *P. compressa* in contrast to *I. pellicula* (14.7%, 2.14%) and *S. normani* (18.8%, 2.23%) respectively. The KS sequences are dominated by *Proteobacteria*, *Cyanobacteria*, *Firmicutes*, *Actinobacteria*, *Planctomycetes* and *Verrucomicrobia* (Figure 2). The proteobacterial

KS sequences represent 38.63%, 28.94% and 37.26% of the sequences in *I. pellicula*, *P. compressa* and *S. normani* respectively. Observable differences are notable in the percentile distribution of *Cyanobacteria* (19.15% *I. pellicula* and 21.04% *P. compressa* in contrast to 8.34% in *S. normani*), *Actinobacteria* (8.53% in *I. pellicula* and 10.53% in *S. normani* in contrast to 19.15% in *P. compressa*), *Planctomycetes* (1% in *I. pellicula*, 4.49% in *P. compressa* and 8.18% in *S. normani*), *Firmicutes* (4.65% *I. pellicula* and 4.79% *P. compressa* and 6.92% in *S. normani*) and *Verrucomicrobia* (1.38% *I. pellicula*, 2.44% *P. compressa*, 2% *S. normani*) derived KS sequences.

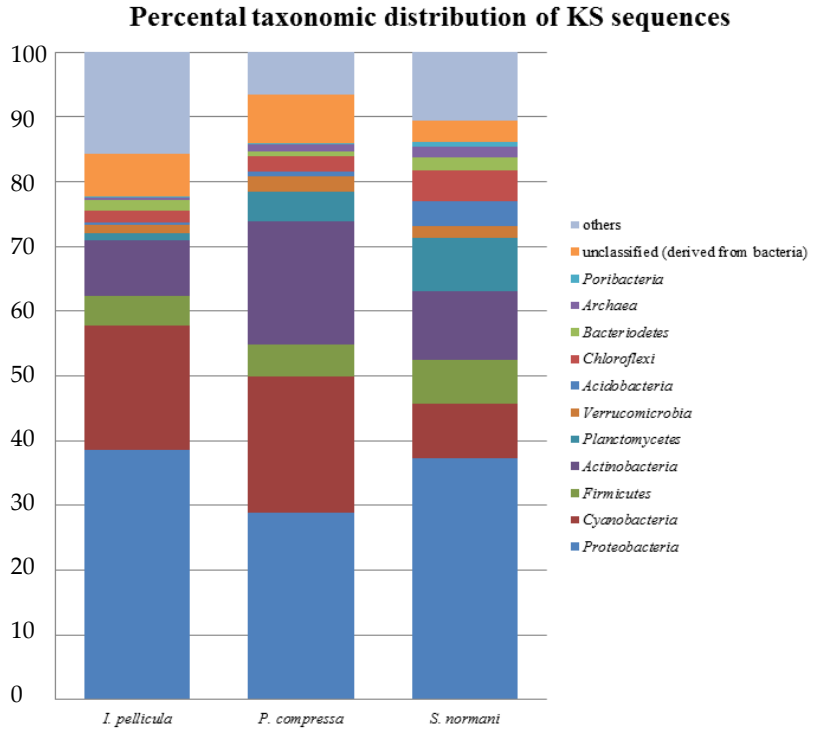



**Figure 1:** Percental distribution of AD sequences. Barchart based on taxonomic identification of raw reads by MG-RAST after dereplication.

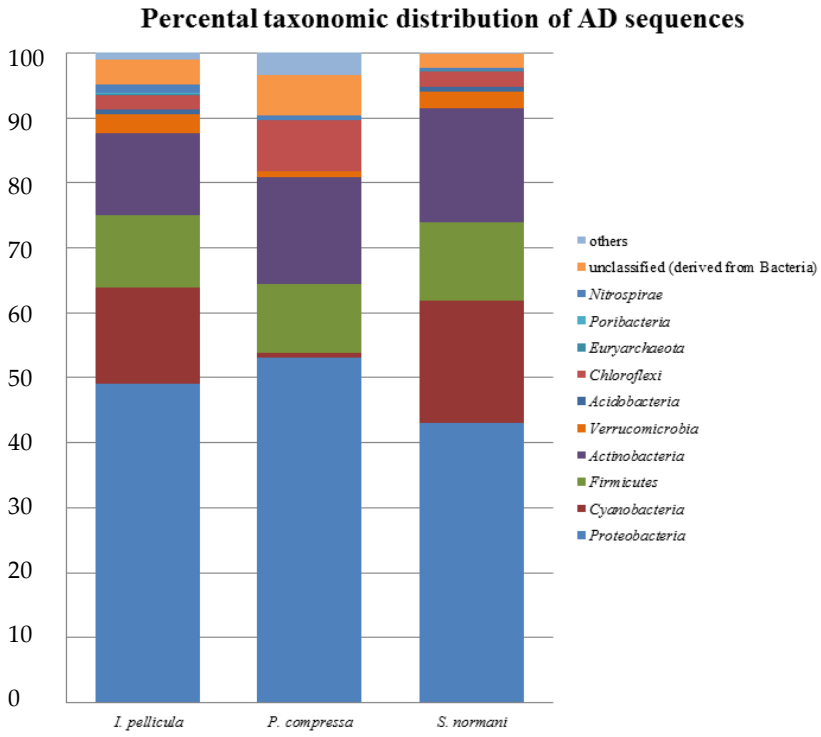



**Figure 2:** Percental distribution of KS sequences. Barchart based on taxonomic identification of raw reads by MG-RAST after dereplication.

### 2.4.1 *Inflatella pellicula*

The sponge samples from *Inflatella pellicula* yielded 351 AD sequences and 3,125 KS sequences after quality filtering and dereplication in QIIME and MG-RAST, resulting in 35 and 72 merged unique representative sequences respectively. Of the 35 AD sequences 18 have a length over 190 bp (up to 697 bp) and were identified as true adenylation domain sequences by BLASTX searches. The predicted taxonomic origin of these sequences is diverse with similarities to genes from species including *Clostridium* sp., *Pseudomonas* sp., *Sorangium cellulosum*, *Microcystis* sp., *Micromonospora* sp., *Streptomyces* sp., *Silvibacterium bohemicum*, *Nostoc* sp., with *Streptomyces* sp., *Microcystis* sp. and *Sorangium cellulosum* being the most prominent origins (level of protein identity ranging from 40 to 60%). As can be seen from Figure 3 (*I. pellicula* tag is colored in red) the obtained reference sequences seem to be distantly related to adenylation domains from macrolides (Epothilone), lipopeptides (Daptomycin) and glycopeptides biosynthetic gene clusters (Vancomycin, Bleomycin, Balhimycin) and dissimilar to streptogramine (Pristinamycin), cyanoginosine (Microcystin), bacteriocin (Enterocin) or depsipeptide (Chondramides) biosynthetic genes (AD domains of compounds in brackets were used to construct the reference data set) (Figure 3).

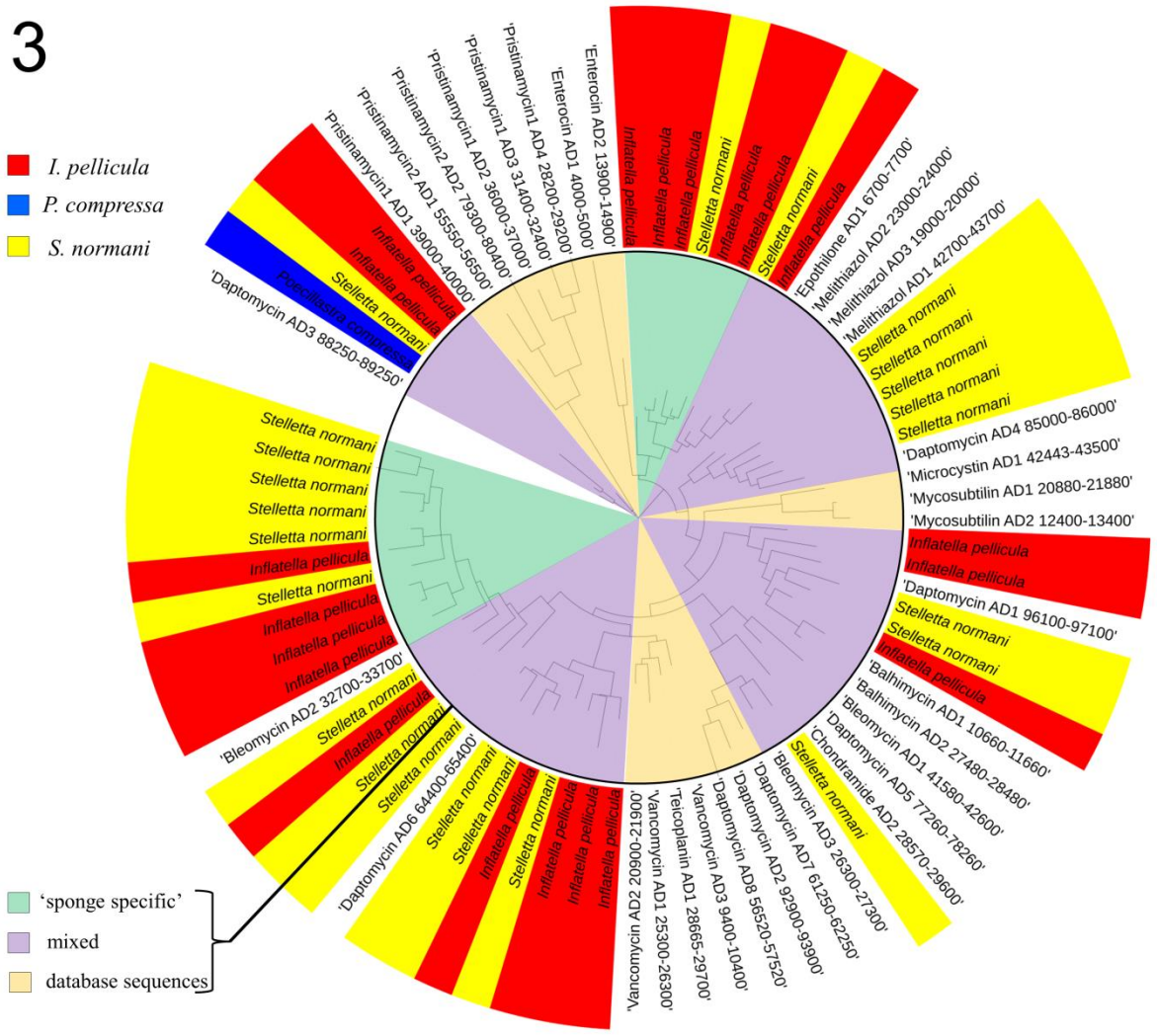
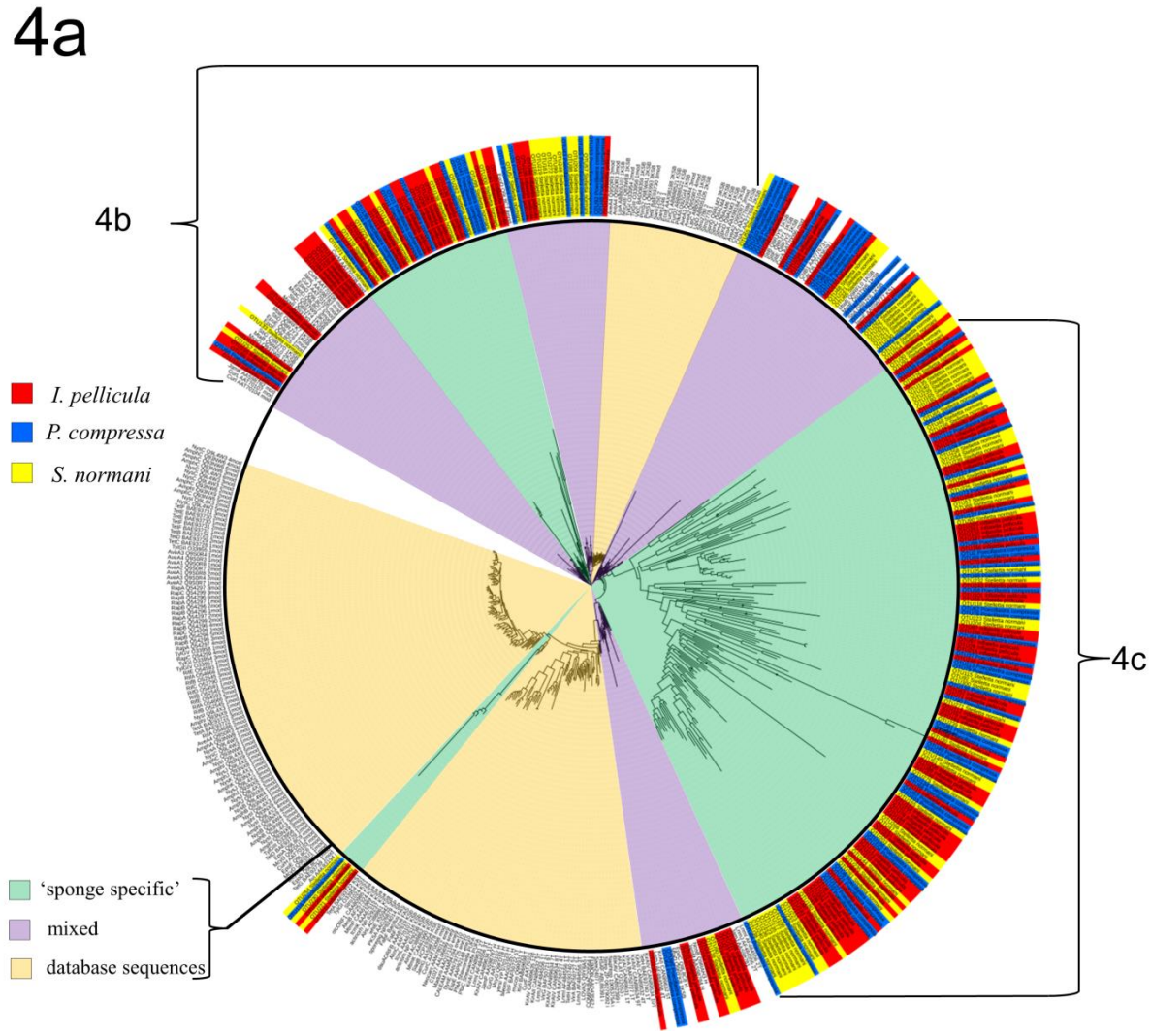



**Figure 3:** Phylogenetic distribution of obtained reference AD sequences compared to a manually constructed reference sequence dataset. The alignment was performed in MEGA software using CLUSTAL W (Thompson *et al.*, 1994) for nucleotide alignment. For phylogenetic tree construction the results were transferred to iTOL software. The sequences from *P. compressa* are blue coloured, from *I. pellicula* red coloured and from *S. normani* yellow coloured. The inner circle highlights the origin of different sequences, pale orange indicates populated only by reference sequences from the manually constructed reference dataset, green clades only comprise of sponge derived sequences and purple clades represent mixed clades.

The KS sequences were not manually checked after reference sequence picking by QIIME, but rather a second quality control step was used by analyzing the sequences with the NaPDos database. This repository consists of 96 different PKS, NRPS and PKS/NRPS hybrid pathways with chemically characterized products. These pathways comprise 648 reference sequences for KS and condensation domains as each pathway may contain several

KS or C domains (see S2 table, for alignment scores). The putative taxonomic origin of the KS domain sequences consists of *Mycobacteria* sp., *Cylindrospermum* sp., *Lyngbya majuscula*, *Sorangium cellulosum*, *Paenibacillus* sp., *Candidatus Endobugula sertula* , *Burkholderia* sp., *Stigmatella aurantica*, *Streptomyces* sp. and many more with uncultured bacteria of marine origin and cyanobacteria being the most prominent (protein identity levels varies from 40 to 70%). Phylogenetic clustering (Figure 4) of the KS domain sequences (*I. pellicula* red) resulted in clade formation (purple sector, 'mixed') with reference KS sequences from known lipopeptide, macrolide biosynthetic genes and a large clade of diverse sequences which were unaffiliated (green sector, 'sponge specific') to a reference sequence.



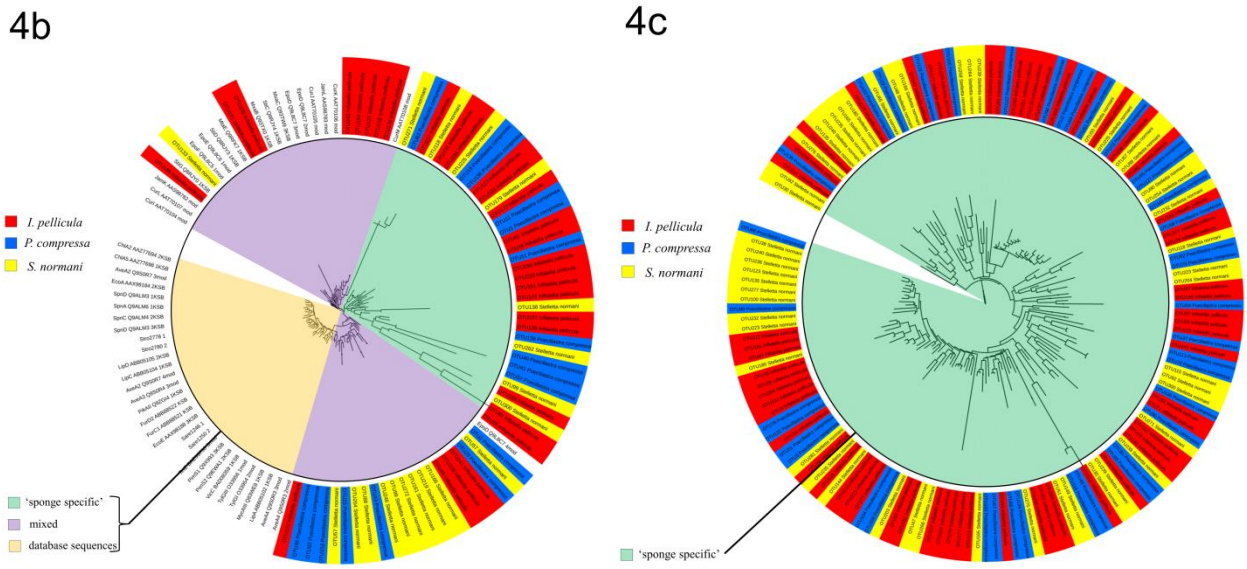



**Figure 4:** Phylogenetic distribution of obtained reference KS sequences compared to reference sequences from NaPDos. The KS domain detection settings were set to minimal length of 100 aa and an e-Value Cutoff of $1e^{-5}$. For phylogenetic tree construction the results were transferred to the iTOL software. The sequences from *P. compressa* are blue coloured, from *I. pellicula* red coloured and from *S. normani* yellow coloured. The inner circle highlights the origin of different sequences, pale orange indicates populated only by reference sequences from the NaPDos database, green clades only comprise of sponge derived sequences ('sponge specific') and purple clades represent mixed clades. 4a shows the phylogenetic tree of all obtained reference KS sequences. 4b is a subtree of figure 4a displaying the three different kinds of observed clade formation. 4c shows a large clade solely made up of obtained reference KS sequences unrelated to references sequences from the NaPDos database.

### 2.4.2 *Poecillastra compressa*

The sponge samples from *Poecillastra compressa* yielded 664 AD sequences and 2,514 KS sequences after dereplication, resulting in 14 and 50 merged unique representative sequences respectively. Of the 14 AD sequences only one was found to be a true adenylation domain with a considerable length (325 bp). The remaining 13 sequences comprised of short reads (120 to 160 bp) with similarities to elongation factors or hypothetical proteins. BLASTX search of the single AD sequence displayed a 51% protein identity to a protein from *Streptomyces* sp. and 39% identity to tyrocidine synthase 3 (tyrocidine is a cyclic decapeptide) from a *Streptomyces* sp.

Forty nine of the 50 KS domain sequences passed the second quality control step. BLASTX was used to investigate the taxonomic origin of these KS sequences, resulting in similarities to previously reported KS sequences from *Lyngbya majuscula*, *Stigmatella aurantica*, *Mycobacterium* sp., *Chondromyces apiculatus*, *Sorangium cellulosum*, *Streptomyces* sp., *and Nannocystis pusilla*. The majority of these KS sequences displayed most similarity to *Cyanobacteria* and to KS sequences from uncultured bacteria of both soil and marine origin. Clustering of the KS sequences (*P. compressa* blue colored tag Figure 4) was performed with the NaPDos reference database and yielded clade formation to KS sequences from known bioactive compounds such as streptogramins, lipopeptides, polyethers, orthosomycin antibiotics and macrolides. Clades were also formed which were clearly distinct from the reference sequences (Figure 4); with protein identity levels ranging from 37% to 75%.

**2.4.3 *Stelletta normani***

The sponge samples from *Stelletta normani* yielded 606 AD sequences and 8605 KS sequences after dereplication. This resulted in 31 and 109 merged unique representative sequences respectively and is therefore the most diverse of the three sample species. Five of the 31 AD domain sequences were discarded due to length restrictions (shorter than 180 bp). A BLASTX search was conducted to look for protein similarities and similarities were predominately found to proteins from *Bacillus* sp., *Stigmatella aurantica*, *Hyella* sp., *Nostoc* sp. and *Microcystis* sp., *Cylindrospermum* sp., *Brevibacillus* sp., *Streptomyces* sp., *Planktothrix* sp., *Nitratireductor* sp. and *Methylobacter* sp. When clustered with known AD domain sequences the obtained sequences (*S. normani* sequences tagged yellow, Figure 3) formed clades with genes that produce lipopeptides , glycopeptides and with sequences derived from the beta-methoxyacrylate inhibitor Melithiazol, with some sequences clustering apart from the reference sequence (Figure 3).

The initial reference sequence picking via QIIME resulted in 109 sequences, of which 55 passed the second quality filter step (NaPDos). BLASTX search of these sequences yielded similarities to proteins from *Sorangium cellulosum*, *Amycolatopsis* sp., *Mycobacterium* sp., *Streptomyces* sp., *Scytonema* sp., *Lyngbya majuscula*, *Clostridium* sp., *Candidatus Thiomargarita nelsonii* and to KS sequences from both uncultured soil and marine bacteria were observed. The KS sequences cluster with biosynthetic genes from lipopeptides, orthosomycin

antibiotic, macrolides and a large cluster of diverse sequences distantly related to KS sequences from gene clusters known to produce Jamaicamides and Melithiazol (Figure 4).

## 2.5 Discussion:

The secondary metabolomic potential of the microbiome of three different deep sea sponge species, *Inflatella pellicula*, *Poecillastra compressa* and *Stelletta normani* was investigated using 454 pyrosequencing; to detect the presence of PKS and NRPS gene cluster associated genes, targeting AD and KS domain sequences (Table 1). The use of a next-generation sequencing approach, circumvents the problems associated with the cultivation of bacteria from these sponges. This study supplements a previous 16S rRNA gene based approach we had employed to study the microbial ecology of these deep sea sponges (Jackson *et al*., 2013; Kennedy *et al*., 2014). Given that NGS analysis of marine sponge metagenomes result in the generation of large data sets (Table 2), it is therefore important that strict quality control is employed so as not to lead to incorrect interpretation of the data. To reflect this the number of raw reads used here has been reduced by approx. 85% in total, 46.8% after quality filtering and 38.5% after dereplication (using default parameters in QIIME and MG-RAST) (Table 2).

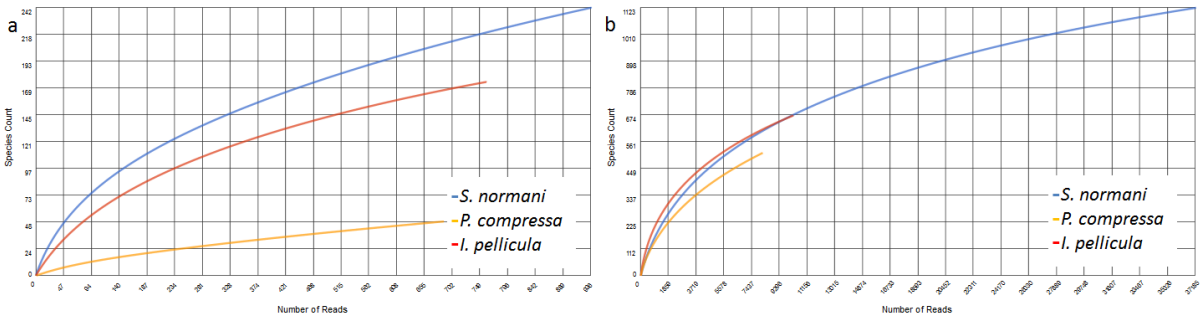



**Figure 5:** Rarefaction curves of the obtained AD (a) and KS (b) sequences. This figure was generated using MG-RAST were the data was compared to the *Non-Redundant Multi-Source Protein Annotation Database* with a minimal identity Cutoff of 60% identity to account for the observed low identity to known sequences, maximal e-Value Cutoff of $1e^{-5}$ and a minimal alignment length Cutoff of 15 aa.

The resulting number of sequences in the final analysis while clearly not representative of the entire biosynthetic potential of the sponge microbial communities are nonetheless significant in that they indicate the presence of PKS and NRPS diversity within these deep sea sponges. Rarefaction curves indicate sufficient coverage, indicated by the plateau of the curve, for only one out of three sample species (Figure 5). The possibility exists, as previously alluded to in Brady *et al.* 2007 that the use of these degenerate primers may lead to the selective amplification of proteobacterial and actinobacterial AD and KS sequences. However, in this instance we feel that possible overrepresentation is likely to be marginal as we have previously reported that *Proteobacteria* and *Actinobacteria* account for a substantial portion of the microbial community of sponges and also of the communities in the deep sea sponges investigated here (Kennedy *et al.* 2014). In that study a 16S rRNA gene sequencing based approach was employed to investigate the microbial communities of four deep sea sponges and of the surrounding seawater, we found that the microbial community of those sponges comprise to a large extent of *Proteobacteria* (especially *γ-Proteobacteria*), *Chloroflexi* (*Stelletta normani*), *Actinobacteria* and *Bacteroidetes*. The predicted taxonomic sources of the KS and AD reads presented here are in the main part, well represented in the aforementioned 16S rRNA gene dataset as well as *Firmicutes* and particularly *Cyanobacteria* (Figure 1 and 2), with the most prominent phyla being *Proteobacteria*, *Actinobacteria* and *Cyanobacteria*. Furthermore *Actinobacteria* or more specifically *Streptomyces* (Chater *et al.*, 2010) and many classes of the diverse phylum of *Proteobacteria* are noted producers of potent secondary metabolites (Gerth *et al.*, 1996; Wenzel and Müller, 2009). Though the proposed phyletic assignments of our KS and AD domain sequences are further validated by the observed similarities between the 16S rRNA gene data and the phylogenetic distribution of the KS and AD sequences, caution is required in the interpretation of these assignments. Putative taxonomic origins of functional genes are not fully reflective of the actual taxonomic source of these genes but are merely indications of sequence identity between a query sequence and its most similar sequence match. Nonetheless the prominent occurrence of *Cyanobacteria* affiliated sequences is puzzling as this bacterial phylum is not present in the 16S rRNA gene datasets and is not expected to be. *Cyanobacteria* rely on photosynthesis for energy generation which does not occur at depths greater than 200m. A possible explanation for this is a high rate of horizontal gene transfer of NRPS and PKS cluster affiliated

sequences, which is known to frequently occur in marine sponge metagenomes. PKS and NRPS gene clusters are also known to be also encoded on 'genomic/pathogenicity islands', that are rich in mobile genetic elements, therefore enhancing their potential transfer frequencies (Fischbach *et al.*, 2008; Ridley *et al.*, 2008; Ziemert *et al.*, 2014).

The phylogenetic trees constructed from the sequences obtained for AD and KS domain fragments clearly sheds further light on the hidden biological potential of microbial populations associated with these deep sea sponges. It is evident that a portion of the sequences, in particular the KS domain sequences form their own diverse clades which are clearly distinct from KS sequences from genes encoding known bioactive compounds (Figure 6). The use of the BLASTX algorithm was particularly illuminating when investigating these KS and AD sequences, given that the comparable long reads achieved with 454 pyrosequencing (up to 700 bp) allowed a more robust analysis to be performed. Many common 'hits' are similar to sequences of marine origin like KS sequences from uncultured bacteria identified from shallow water sponges or to *Mycobacterium marinum*, *Cyanobacteria* and *Streptomyces* sp.. Furthermore it is worth mentioning the occasional appearance of KS domain 'hits' with sequences from *Sorangium cellulosum* a myxobacteria inhabiting soil environments and the producer of Epothilone (Gerth *et al.*, 1996). Other sequences showed similarities to genes from genera which are known to produce Hectochlorin, Jamaicamides, Gulmirecins (Schieferdecker *et al.*, 2014) and Nostophycin (Fewer *et al.*, 2011) amongst others. Hectochlorin was first isolated from the marine cyanobacteria *Lyngbya majuscula* and is a product of a mixed PKS/NRPS pathway and displays potent antifungal and cytotoxic properties (Ramaswamy *et al.*, 2007). The Jamaicamides are lipopeptides which are also of mixed PKS/NRPS origin. They are produced by the marine cyanobacteria *Lyngbya majuscula* and display sodium channel blocking capabilities (Edwards *et al.*, 2004). The origin of the AD domain fragments is also quite diverse with the closest 'hits' being to AD genes from *Brevibacillus*, *Streptomyces*, *Pseudomonas*, *Nostoc* and *Clostridium* species. Actual 'hits' with known bioactive compounds for AD sequences comprise of similarities to the AD domain from the genecluster encoding Simocyclinone, an angucycline antibiotic with topoisomerase inhibitory activity (Flatman *et al.*, 2005) and an AD domain from Microcystin which is an hepatotoxin produced by *Cyanobacteria* (Dawson, 1998). Furthermore, comparatively few AD domain fragments

(compared to KS sequences) were retrieved from the data (2,385 before and 1,621 sequences after quality filtering, Table 2), which may be due to a low abundance of this sequence type in deep sea sponges. The KS and AD domain fragment sequences can be distinguished by either clustering with reference sequences or by forming their own clades, which are only very distantly related to the database sequences used for comparison (Figure 3, 4). This is particularly true in the case of the KS sequences which make up a clade of sequences which are clearly distinct from KS sequences from genes involved in the synthesis of known bioactive compounds (Figure 4). These clades are very diverse, as is evident from the individual branch lengths in the phylogenetic tree (Figure 4a, b, c). Furthermore, the KS and AD sequences show similarities to genes linked to the production of a broad range of antibiotics and toxins of different groups. These include lipopeptides, glycopeptides, macrolides, streptogramins, depsipepdtides, cyanoginosines, bacteriocins and hepatotoxins. Thus while sequence similarity searches and sequence cladograms indicate degrees of similarity with known PKS and NRPS gene fragments, degrees of novelty or divergence are also very obvious (Figure 3, 4). In particular the KS and AD gene fragments which have been identified here form clades which are clearly distinct from those of known antibiotic related gene clusters. This indicates that potential novel biodiversity with respect to marine natural products is likely to be present in these deep sea sponge microbiomes.

In conclusion, this study reveals that PKS and NRPS affiliated domains are prevalent among the genomes of the members of the microbial communities of these deep sea sponges, which may potentially also be from symbiotic members of the community and therefore be sponge-specific. Nonetheless further research needs to be performed to allocate the biological potential identified here to whole gene clusters and possible gene products. The exploitation of this potential may however be difficult to achieve, particularly bearing in mind the difficulties in obtaining samples from these depths and the sample size requirements involved. However given the potential biodiversity that we report here, with respect to natural product biosynthetic genes, such difficulties may be worth overcoming, particularly given the ongoing need for novel bioactive polyketides and nonribosomal peptides.

## 2.6 Bibliography


Aylward, F. O., J. M. Eppley, J. M. Smith, F. P. Chavez, C. A. Scholin, and E. F. DeLong, 2015, Microbial community transcriptional networks are conserved in three domains at ocean basin scales: Proc Natl Acad Sci U S A, v. 112, p. 5443-8.

Ayuso-Sacido, A., and O. Genilloud, 2005, New PCR primers for the screening of NRPS and PKS-I systems in *Actinomycetes*: detection and distribution of these biosynthetic gene sequences in major taxonomic groups: Microb Ecol, v. 49, p. 10-24.

Bell, J. J., 2008, The functional roles of marine sponges, Estuarine, Coastal and Shelf Science, Elsevier, p. 341-353.

Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Peña, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, and R. Knight, 2010, QIIME allows analysis of high-throughput community sequencing data: Nat Methods, v. 7, p. 335-6.

Charlop-Powers, Z., J. G. Owen, B. V. Reddy, M. A. Ternei, and S. F. Brady, 2014, Chemical-biogeographic survey of secondary metabolism in soil: Proc Natl Acad Sci U S A, v. 111, p. 3757-62.

Chater, K. F., S. Biró, K. J. Lee, T. Palmer, and H. Schrempf, 2010, The complex extracellular biology of *Streptomyces*: FEMS Microbiol Rev, v. 34, p. 171-98.

Dawson, R. M., 1998, The toxicology of microcystins: Toxicon, v. 36, p. 953-62.

Edgar, R. C., 2010, Search and clustering orders of magnitude faster than BLAST: Bioinformatics, v. 26, p. 2460-1.

Edwards, D. J., B. L. Marquez, L. M. Nogle, K. McPhail, D. E. Goeger, M. A. Roberts, and W. H. Gerwick, 2004, Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*: Chem Biol, v. 11, p. 817-33.

Fewer, D. P., J. Osterholm, L. Rouhiainen, J. Jokela, M. Wahlsten, and K. Sivonen, 2011, Nostophycin biosynthesis is directed by a hybrid polyketide synthase-nonribosomal

peptide synthetase in the toxic cyanobacterium *Nostoc* sp. strain 152: Appl Environ Microbiol, v. 77, p. 8034-40.

Fischbach, M. A., C. T. Walsh, and J. Clardy, 2008, The evolution of gene collectives: How natural selection drives chemical innovation: Proc Natl Acad Sci U S A, v. 105, p. 4601-8.

Flatman, R. H., A. J. Howells, L. Heide, H. P. Fiedler, and A. Maxwell, 2005, Simocyclinone D8, an inhibitor of DNA gyrase with a novel mode of action: Antimicrob Agents Chemother, v. 49, p. 1093-100.

Fusetani, N., 2011, Antifouling marine natural products: Nat Prod Rep, v. 28, p. 400-10.

Gerth, K., N. Bedorf, G. Höfle, H. Irschik, and H. Reichenbach, 1996, Epothilons A and B: antifungal and cytotoxic compounds from *Sorangium cellulosum* (Myxobacteria). Production, physico-chemical and biological properties: J Antibiot (Tokyo), v. 49, p. 560-3.

Gomez-Alvarez, V., T. K. Teal, and T. M. Schmidt, 2009, Systematic artifacts in metagenomes from complex microbial communities: ISME J, v. 3, p. 1314-7.

Gurgui, C., and J. Piel, 2010, Metagenomic approaches to identify and isolate bioactive natural products from microbiota of marine sponges: Methods Mol Biol, v. 668, p. 247-64.

Helfrich, E. J., S. Reiter, and J. Piel, 2014, Recent advances in genome-based polyketide discovery: Curr Opin Biotechnol, v. 29, p. 107-15.

Hentschel, U., J. Piel, S. M. Degnan, and M. W. Taylor, 2012, Genomic insights into the marine sponge microbiome: Nat Rev Microbiol, v. 10, p. 641-54.

Hertweck, C., 2009, The biosynthetic logic of polyketide diversity: Angew Chem Int Ed Engl, v. 48, p. 4688-716.

Jackson, S. A., E. Borchert, F. O'Gara, and A. D. Dobson, 2015, Metagenomics for the discovery of novel biosurfactants of environmental interest from marine ecosystems: Curr Opin Biotechnol, v. 33, p. 176-82.

Jackson, S. A., B. Flemer, A. McCann, J. Kennedy, J. P. Morrissey, F. O'Gara, and A. D. Dobson, 2013, Archaea appear to dominate the microbiome of *Inflatella pellicula* deep sea sponges: PLoS One, v. 8, p. e84438.

Jang, K. H., S. J. Nam, J. B. Locke, C. A. Kauffman, D. S. Beatty, L. A. Paul, and W. Fenical, 2013, Anthracimycin, a potent anthrax antibiotic from a marine-derived *Actinomycete*: Angew Chem Int Ed Engl, v. 52, p. 7822-4.

Kennedy, J., C. E. Codling, B. V. Jones, A. D. Dobson, and J. R. Marchesi, 2008, Diversity of microbes associated with the marine sponge, *Haliclona simulans*, isolated from Irish waters and identification of polyketide synthase genes from the sponge metagenome: Environ Microbiol, v. 10, p. 1888-902.

Kennedy, J., N. D. O'Leary, G. S. Kiran, J. P. Morrissey, F. O'Gara, J. Selvin, and A. D. Dobson, 2011, Functional metagenomic strategies for the discovery of novel enzymes and biosurfactants with biotechnological applications from marine ecosystems: J Appl Microbiol, v. 111, p. 787-99.

Kennedy, J., Flemer, B., Jackson, S.A., Morrissey, J., O'Gara, F., Dobson, A.D.W. 2014, Evidence of a putative deep sea specific microbiome in marine sponges: PLoS One, 9(3): e91092.

Khosla, C., 2009, Structures and mechanisms of polyketide synthases: J Org Chem, v. 74, p. 6416-20.

Letunic, I., and P. Bork, 2007, Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation: Bioinformatics, v. 23, p. 127-8.

Letunic, I., and P. Bork, 2011, Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy: Nucleic Acids Res, v. 39, p. W475-8.

McDaniel, L. D., E. Young, J. Delaney, F. Ruhnau, K. B. Ritchie, and J. H. Paul, 2010, High frequency of horizontal gene transfer in the oceans: Science, v. 330, p. 50.

Meyer, F., D. Paarmann, M. D'Souza, R. Olson, E. M. Glass, M. Kubal, T. Paczian, A. Rodriguez, R. Stevens, A. Wilke, J. Wilkening, and R. A. Edwards, 2008, The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes: BMC Bioinformatics, v. 9, p. 386.

Penn, K., C. Jenkins, M. Nett, D. W. Udwary, E. A. Gontang, R. P. McGlinchey, B. Foster, A. Lapidus, S. Podell, E. E. Allen, B. S. Moore, and P. R. Jensen, 2009, Genomic islands link secondary metabolism to functional adaptation in marine *Actinobacteria*: ISME J, v. 3, p. 1193-203.

Ramaswamy, A. V., C. M. Sorrels, and W. H. Gerwick, 2007, Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*: J Nat Prod, v. 70, p. 1977-86.

Ramirez-Llodra, E., A. Brandt, R. Danovaro, B. De Mol, E. Escobar, C. German, L. Levin, P. Arbizu, L. Menot, and P. Buhl-Mortensen, 2010, Deep, diverse and definitely different: unique attributes of the world's largest ecosystem: Biogeosciences, v. 7, p. p.2851-2899.

Reddy, B. V., D. Kallifidas, J. H. Kim, Z. Charlop-Powers, Z. Feng, and S. F. Brady, 2012, Natural product biosynthetic gene diversity in geographically distinct soil microbiomes: Appl Environ Microbiol, v. 78, p. 3744-52.

Reen, F. J., J. A. Gutiérrez-Barranquero, A. D. Dobson, C. Adams, and F. O'Gara, 2015, Emerging concepts promising new horizons for marine biodiscovery and synthetic biology: Mar Drugs, v. 13, p. 2924-54.

Ridley, C. P., H. Y. Lee, and C. Khosla, 2008, Evolution of polyketide synthases in bacteria: Proc Natl Acad Sci U S A, v. 105, p. 4595-600.

Rocha-Martin, J., C. Harrington, A. D. Dobson, and F. O'Gara, 2014, Emerging strategies and integrated systems microbiology technologies for biodiscovery of marine bioactive compounds: Mar Drugs, v. 12, p. 3516-59.

Satpute, S. K., I. M. Banat, P. K. Dhakephalkar, A. G. Banpurkar, and B. A. Chopade, 2010, Biosurfactants, bioemulsifiers and exopolysaccharides from marine microorganisms: Biotechnol Adv, v. 28, p. 436-50.

Schieferdecker, S., S. König, C. Weigel, H. M. Dahse, O. Werz, and M. Nett, 2014, Structure and biosynthetic assembly of gulmirecins, macrolide antibiotics from the predatory bacterium *Pyxidicoccus fallax*: Chemistry, v. 20, p. 15933-40.

Schirmer, A., R. Gadkari, C. D. Reeves, F. Ibrahim, E. F. DeLong, and C. R. Hutchinson, 2005, Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissoluta*: Appl Environ Microbiol, v. 71, p. 4840-9.

Sipkema, D., K. Schippers, W. J. Maalcke, Y. Yang, S. Salim, and H. W. Blanch, 2011, Multiple approaches to enhance the cultivability of bacteria associated with the marine sponge *Haliclona* (gellius) sp: Appl Environ Microbiol, v. 77, p. 2130-40.

Sobecky, P. A., and T. H. Hazen, 2009, Horizontal gene transfer and mobile genetic elements in marine systems: Methods Mol Biol, v. 532, p. 435-53.

Sogin, M. L., H. G. Morrison, J. A. Huber, D. Mark Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, and G. J. Herndl, 2006, Microbial diversity in the deep sea and the underexplored "rare biosphere": Proc Natl Acad Sci U S A, v. 103, p. 12115-20.

Taylor, M. W., R. Radax, D. Steger, and M. Wagner, 2007, Sponge-associated microorganisms: evolution, ecology, and biotechnological potential: Microbiol Mol Biol Rev, v. 71, p. 295-347.

Thompson, J. D., D. G. Higgins, and T. J. Gibson, 1994, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice: Nucleic Acids Res, v. 22, p. 4673-80.

Trindade-Silva, A. E., C. P. Rua, B. G. Andrade, A. C. Vicente, G. G. Silva, R. G. Berlinck, and F. L. Thompson, 2013, Polyketide synthase gene diversity within the microbiome of the sponge *Arenosclera brasiliensis*, endemic to the Southern Atlantic Ocean: Appl Environ Microbiol, v. 79, p. 1598-605.

Uria, A., and J. Piel, 2009, Cultivation-independent approaches to investigate the chemistry of marine symbiotic bacteria: Phytochemistry Reviews, v. 8, p. pp 401-414.

Webster, N. S., and M. W. Taylor, 2012, Marine sponges and their microbial symbionts: love and other relationships: Environ Microbiol, v. 14, p. 335-46.

Webster, N. S., M. W. Taylor, F. Behnam, S. Lücker, T. Rattei, S. Whalan, M. Horn, and M. Wagner, 2010, Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts: Environ Microbiol, v. 12, p. 2070-82.

Wenzel, S. C., and R. Müller, 2009, *Myxobacteria*--'microbial factories' for the production of bioactive secondary metabolites: Mol Biosyst, v. 5, p. 567-74.

Whitman, W. B., D. C. Coleman, and W. J. Wiebe, 1998, Prokaryotes: the unseen majority: Proc Natl Acad Sci U S A, v. 95, p. 6578-83.

Woodhouse, J. N., L. Fan, M. V. Brown, T. Thomas, and B. A. Neilan, 2013, Deep sequencing of non-ribosomal peptide synthetases and polyketide synthases from the microbiomes of Australian marine sponges: ISME J, v. 7, p. 1842-51.

Ziemert, N., A. Lechner, M. Wietz, N. Millán-Aguiñaga, K. L. Chavarria, and P. R. Jensen, 2014, Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*: Proc Natl Acad Sci U S A, v. 111, p. E1130-9.

Ziemert, N., S. Podell, K. Penn, J. H. Badger, E. Allen, and P. R. Jensen, 2012, The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity: PLoS One, v. 7, p. e34064.

# Chapter 3

## Biotechnological potential of cold adapted *Pseudoalteromonas* spp. isolated from 'deep sea' sponges

**Erik Borchert[1], Stephen Knobloch[2], Emilie Dwyer[1], Sinéad O'Flynn[1], Stephen A. Jackson[1], Ragnar Jóhannsson[2], Viggó T. Marteinsson[2], Fergal O'Gara[1,3,4] and Alan D.W. Dobson[1]**

[1] School of Microbiology, University College Cork, National University of Ireland, Cork, Ireland

[2] Department of Research and Innovation, Matís ohf., Reykjavik, Iceland

[3] Biomerit Research Centre, University College Cork, National University of Ireland, Cork, Ireland

[4] School of Biomedical Sciences, Curtin Health Innovation Research Institute, Curtin University, Perth, Australia

## 3.1 Abstract:

The marine genus *Pseudoalteromonas* is known for its versatile biotechnological potential with respect to the production of antimicrobials and enzymes of industrial interest. We have sequenced the genomes of three *Pseudoalteromonas* sp. strains isolated from different deep sea sponges on the Illumina MiSeq platform. The isolates have been screened for various industrially important enzymes and comparative genomics has been applied to investigate potential relationships between the isolates and their host organisms, while comparing them to free-living *Pseudoalteromonas* spp. from shallow and deep sea environments. The genomes of the sponge associated *Pseudoalteromonas* strains contained much lower levels of potential eukaryotic-like proteins which are known to be enriched in symbiotic sponge associated microorganisms, than might be expected for true sponge symbionts. While all the *Pseudoalteromonas* shared a large distinct subset of genes, nonetheless the number of unique and accessory genes is quite large and defines the pan-genome as open. Enzymatic screens indicate that a vast array of enzyme activities are expressed by the isolates including β-galactosidase, β-glucosidase and protease activities and further tests identified these activities to be both psychrophilic and mesophilic, as well as favoring alkaline pH conditions.

## 3.2 Introduction

The genus *Pseudoalteromonas* are a subgroup of Gram-negative Gammaproteobacteria with common features including, a requirement for $Na^{2+}$ ions, motility and aerobic and chemoheterotrophic metabolism. The genus was first described by Gauthier and co-workers and separated from the genus *Alteromonas* (Gauthier *et al.*, 1995). The genus can be divided into either pigmented or non-pigmented species, with members of the genus being known to possess the ability to produce a wide array of bioactive compounds. The pigmented species in particular are known to produce a range of antimicrobial and antifouling compounds which display activity against a broad spectrum of organisms and have as a result been widely investigated in the past (Holmström *et al.*, 1996; Egan *et al.*, 2002; Bowman, 2007; Fehér *et al.*, 2010). While the non-pigmented species are typically not antimicrobial

producers, they are however versatile producers of an array of different extracellular enzymes that are of potential biotechnological interest (Cieśliński *et al.*, 2005; Dobretsov *et al.*, 2007; Mo *et al.*, 2009; Oh *et al.*, 2010; Yan *et al.*, 2009). *Pseudoalteromonas* are one of the most frequently isolated bacteria from marine environments (Holmström and Kjelleberg, 1999) and are routinely found in association with various eukaryotic hosts in these environments such as tunicates (Holmström *et al.*, 1998), algae (Egan *et al.*, 2001), sponges (Ivanova *et al.*, 2002), mussels (Ivanova *et al.*, 1998), pufferfish (Simidu *et al.*, 1990) as well as algae and marine plants (Akagawa-Matsushita *et al.*, 1992; Yoshikawa *et al.*, 1997). They have also been isolated as free living in seawater (Bozal *et al.*, 1997), sea ice (Bowman, 1998) and marine sediment (Qin *et al.*, 2011).

The deep oceans as an ecosystem are of growing interest to the scientific community. While the mean depth of the oceans is 3800 m, about 50% is deeper than 3000 m. With only 5% of the 'deep sea' having to date been explored, it is clear that the biotechnological potential of this unique ecosystem has yet to be fully exploited (Borchert *et al.*, 2016; Ramirez-Llodra *et al.*, 2010; Sipkema, 2016). We have previously reported on the microbial biodiversity of deep sea sponges sampled at depths of between 760-2900 m below sea level, indicating that the microbial community structures of these sponges may represent an untapped source of potential microbial biodiversity (Jackson *et al.*, 2013; Kennedy *et al.*, 2014). Bacterial and fungal communities from deep sea sediments also continue to receive attention, not only from an ecological standpoint (Xu *et al.*, 2014), but also due to the ability of microorganisms isolated from this ecosystem to produce novel bioactive molecules (Li *et al.*, 2016; Wu *et al.*, 2016) and enzymes of biotechnological importance (Shao *et al.*, 2015; Yang *et al.*, 2016). Cold-active enzymes are of particular interest as they possess a range of structural features that promote flexibility at the active site, low substrate affinity and high specific activity at low temperatures. These characteristics are important in industrial biocatalysis, not only from an energy savings standpoint, but also due to the fact that reactions at low temperatures prevents undesirable chemical side reactions which can occur at higher temperatures; while also allowing rapid thermal inactivation of these enzymes, due to their thermolabile properties (Cavicchioli *et al.*, 2002; Santiago *et al.*, 2016).

*Pseudoalteromonas* strains have previously been reported to produce a number of cold adapted enzymes including DNA ligase (Georlette *et al.*, 2000), pectate lyase (Truong *et al.*,

2001), β-galactosidase (Cieśliński *et al.*, 2005), subtilase (Yan *et al.*, 2009) and agarase (Oh *et al.*, 2010), with enzyme production in *Pseudoalteromonas haloplanktis* TAC125 and other *Pseudoalteromonas* strains namely sp. ANT506, sp. ANT178, sp. KMM701 and sp. CF6-2 in particular being studied in more detail (Santiago *et al.*, 2016; Yang *et al.*, 2016). With this in mind, this study focused on the isolation and comparative genomics of three non-pigmented *Pseudoalteromonas* spp. isolated at different depths from both marine sponges and sediment; in an effort to assess their biotechnological potential.

The genomes share a large pangenome and have a considerable number of unique gene clusters, but only a small number of genes are associated with potential host interaction in all the investigated genomes, irrespective of whether or not they have been isolated from sponges, from deep sea sediment or ocean water. Furthermore *Pseudoalteromonas* strains EB27, SK18 and SK20, isolated from deep sea sponges do not share a large number of genes that could be attributed to a symbiotic lifestyle. While the strains displayed cold-adapted growth characteristics, they are unlikely to be true psychrophiles. The three strains did however display a number of interesting enzyme activities including β-glucosidase, protease and β-galactosidase activities and displayed properties that are favorable to industrial applications, such as alkaline pH optima and cold-adaptation.

## 3.3 Materials and Methods

### 3.3.1. Sponge collection and isolation of microorganisms

The sponges (*Poecillastra compressa*, *Inflatella pellicula* and *Sericolophus hawaiicus*) used for the isolation of microorganisms have been collected of the west coast of Ireland during the Biodiscovery cruises 2010 and 2013 by the remotely operated vehicle *Holland I* on board the R.V. *Celtic Explorer*. The sponges were rinsed directly after collection with sterile artificial sea water (3.33% (w/v) Instant Ocean, Aquarium Systems) to remove any exogenous material and were subsequently stored at -80°C until further processing. The isolation of microorganisms was performed as follows; small sponge pieces were macerated with a sterile razor blade and serially diluted with artificial seawater and plated onto SYP-SW plates (10g/l starch, 4g/l yeast extract, 2g/l peptone, 33.3g/l artificial sea salt, 1.5% agar). The

plates were inspected daily for colonies and incubated for four weeks at 28°C. All colonies were restreaked until pure cultures were obtained.

### 3.3.2. Enzyme activity plate screenings

The pure cultures were tested for different enzyme activities. All screenings were carried out at 28°C and incubated for three to four days.

Protease screening was carried out using SYP-SW plates supplemented with 2% skim milk (Sigma-Aldrich), a clear halo around the colonies after incubation indicates a possible protease activity. Positive colonies were further tested on SYP-SW plates supplemented with 40 ng/ml X-Gal to differentiate between true protease activity and β-glucosidase/β-galactosidase activity, a blue colour change of the colony would indicate that it is rather the latter activity.

Cellulase activity was tested using SYP-SW plates supplemented with 0.1% Ostazin brilliant red hydroxyethyl-cellulose (OBR-HEC; Slovak Academy of Science, Institute of Chemistry); clear halos around colonies indicate cellulase activity.

Lipase activity was investigated via adding 1% tributyrin (Sigma-Aldrich) to the SYP-SW plates, again a clear halo around the colonies indicates a lipase or esterase activity.

### 3.3.3. Enzyme assays and growth characterization

Native enzyme assays for β-glucosidase and protease activity were carried out with aliquots of overnight cultures from the respective *Pseudoalteromonas* sp. isolates. β-glucosidase assays were carried out by adding 150 µl of overnight culture to 1350 µl 0.1 M potassium phosphate buffer solution (pH 8.5, pH 6.0 and 7.0 gave little to no detectable activity) and 22.5 µl 0.1 M p-nitrophenyl-β-D glucopyranoside as substrate, incubated for 1.5 h at different temperatures and absorbance at 420 nm was measured subsequently. The β-galactosidase assay followed a similar protocol as the β-glucosidase assay, besides that the substrate was substituted with 45 µl 0.05 M p-nitrophenyl-β-D galactopyranoside. For the protease assay 150 µl overnight culture were added to 250 µl 2% azocasein solution (0.1 M Tris-HCl, pH 8.0) followed by incubation for 1 h at different temperatures, then 900 µl 10% trichloroacetic acid were added and incubated for 15 min at room temperature to stop the

reaction and afterwards centrifuged for 10 min at max. speed. 600 µl of the supernatant were combined with 700 µl 1 M NaOH and the absorbance was measured at 440 nm.

Overnight cultures incubated at 28°C and 180 rpm were diluted the next day in 30 ml marine broth (Difco marine broth 2216) to an optical density at 600 nm of 0.05 and subsequently incubated at different temperatures (4°C , 23°C , 28°C , 37°C) in shaking incubators. The growth was monitored hourly by measuring the optical density at 600 nm. The specific growth rate (mu) and the generation/doubling time ($t_{gen}$) was calculated using the formula $mu=(\ln(X_1)-\ln(X_0))/(t_1-t_0)$ and $t_{gen}=(0.693/mu)*60$, with $X_0$ being the optical density at the beginning of the exponential growth phase (approximate $OD_{600}$ of 0.15), $X_1$ a time point within the exponential growth phase ($OD_{600}$ between 0.15 and 1.0) and t the time passed between $X_0$ and $X_1$ in hours.

### 3.3.4. Genomic DNA isolation and sequencing

Genomic DNA isolation was carried out by processing a 10ml overnight culture (SYP-SW medium, 28°C, 180 rpm), after centrifugation the media was removed and 2 ml lysis buffer (2% SDS, 1% CTAB, 100 mM Tris, 100 mM EDTA, 1.5 M NaCl, pH 8.0) were added and incubated in a water bath at 70°C with occasional mixing for two hours. The cell lysate was centrifuged until a clear lysate was obtained, 0.7 volumes of Isopropanol were subsequently added to precipitate the genomic DNA (30 min, room temperature). After centrifugation the supernatant was discarded and the obtained pellet was washed with 70% ethanol, then centrifuged again and after supernatant removal, air dried and finally resuspended in an appropriate amount of Tris-EDTA buffer (10 mM Tris, 1 mM EDTA, pH 8.0). DNA quality was assessed by running 5 µl DNA on an agarose gel resulting in a high molecular weight band. Purity of DNA was measured on a NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). DNA was quantified with a Qubit dsDNA HS assay (Thermo Fisher Scientific) prior to preparing genomic libraries with the Nextera XT DNA Library Prep Kit (Illumina, San Diego, USA) according to the manufacturer's instructions. Final libraries were barcoded with Nextera XT indices, assessed on a Bioanalyzer High Sensitivity DNA chip (Agilent Technologies, Santa Clara, CA, USA) and sequenced together on an Illumina MiSeq platform using paired-end 300 bp chemistry. Raw sequence data was quality trimmed using Trimmomatic version 0.36 (Bolger

*et al.*, 2014) removing short reads and trimming both ends of reads containing low quality bases. Quality trimmed reads were assembled using SPAdes version 3.7.0 (Bankevich *et al.*, 2012) in paired-end mode with default settings. Full length 16S rRNA gene sequences were predicted using RNAmmer (Lagesen *et al.*, 2007). (the genomes are deposited in the NCBI database under the accession numbers MTQB00000000, MTQC00000000, MTQD00000000)

### 3.3.5. Genome analysis and comparison

The draft genomes were annotated using the RAST pipeline (Aziz *et al.*, 2008; Brettin *et al.*, 2015; Overbeek *et al.*, 2014). The genomes of *Pseudoalteromonas haloplanktis* TAC125 (Médigue *et al.*, 2005) and *Pseudoalteromonas* sp. SM9913 (Qin *et al.*, 2011) were used as reference genomes for comparison and annotated again in the same manner as the newly isolated *Pseudoalteromonas* spp. to rule out annotation biases between different software packages. Genome comparison was carried out by using the BPGA pipeline (Chaudhari *et al.*, 2016) and manually screening the genomes for enzymes of industrial interest. The genomes were screened for secondary metabolite gene clusters using antiSMASH (Blin *et al.*, 2013; Medema *et al.*, 2011; Weber *et al.*, 2015).

## 3.4 Results

### 3.4.1. Enzymatic activity profile

The three *Pseudoalteromonas* spp. isolates displayed a range of different enzyme activity profiles (Table 1). The *Poecillastra compressa* isolate EB27 (retrieved from a depth of 1480 m) displayed the greatest range of different activities, with the most prominent being β-glucosidase, protease and cellulase activity. SK20 (*Inflatella pellicula*, 2900 m) exhibited strong β-galactosidase activity and SK18 (*Sericolophus hawaiicus*, 2129 m) displayed high levels of protease activity. All of the isolates displayed some lipolytic activity. EB27 and SK18 also displayed low levels of amylase activity.

**Table 1**: Enzyme active profile of the *Pseudoalteromonas* sp. Isolates based on plate screenings. Activity is depicted as 'X' and intensity (size of halo) is indicated by the number of 'Xs', with 'X' low activity and 'XXX' describing highest activity. (Glc = β-glucosidase, Gal = β-galactosidase)

| Isolate ID | Sponge | Depth [m] | Cellulase | Lipase | Protease | β-Glc/Gal |
|---|---|---|---|---|---|---|
| EB27 | *Poecillastra compressa* | 1480 | XX | X | XXX | XXX (Glc) |
| SK18 | *Sericolophus hawaiicus* | 2129 | - | X | XXX | - |
| SK20 | *Inflatella pellicula* | 2900 | - | X | - | XXX (Gal) |

Based on the strong protease and β-glucosidase activity in EB27 and the protease and β-galactosidase activity in SK18 these activities were further characterized. Native assays were performed under different temperatures and pH conditions; to provide an insight into the general biochemical characteristics of the enzymes produced by these deep sea microorganisms. Both β-glucosidase and protease activity was markedly affected by pH, with optimal activity being observed at pH 8.5, with a complete loss of activity being observed at a lower (pH 6.0) or only slight activity at neutral pH (data not shown). All assays were subsequently conducted at pH 8.5. The temperature dependency of the enzymes was assessed, with β-glucosidase activity in EB27 being observed over a wide temperature range. Maximum activity was observed at 23°C, with lower levels being observed at 37°C and 4°C respectively (Figure 1). β-glucosidase activity was also observed over a wide temperature range in both SK18 and SK20 albeit that the levels of activity were lower.

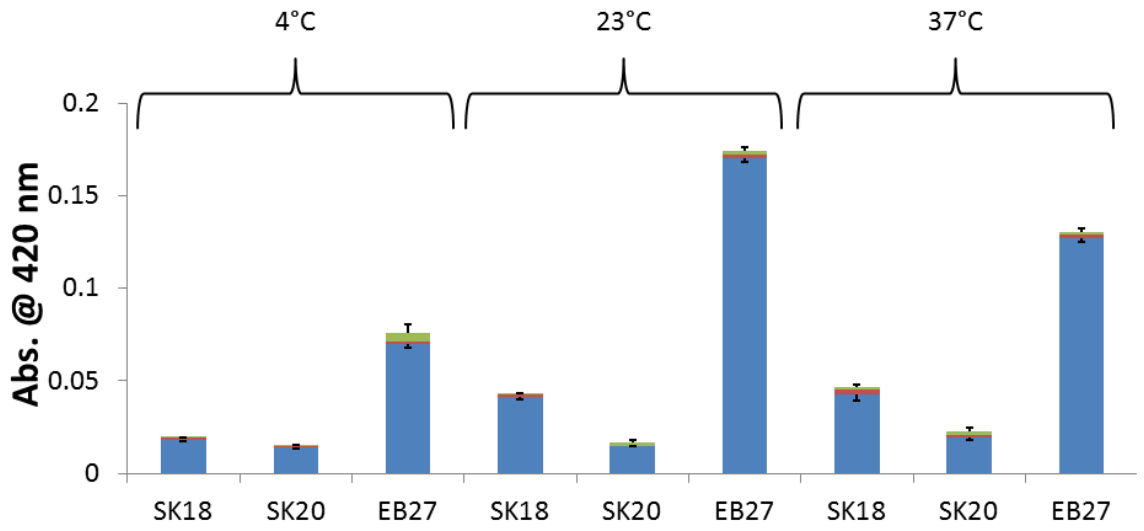



**Figure 1**: β-glucosidase activity of all isolates in native activity assays at different temperatures (*n*=3, pH 8.5). (Barchart with integrated box-whisker plots, green colored $Q_3$ and red colored $Q_1$)

Protease activity was also assessed in both EB27 and SK18, with activity in EB27 in particular being observed over a wide temperature range from 4°C to 55°C. Protease activity in EB27 was highest at 37°C with good levels of activity still being observed at both 45°C and 55°C. The highest level of protease activity in SK18 was also observed at 37°C, but as with EB27, good levels of activity were also observed at higher temperatures (Figure 2).

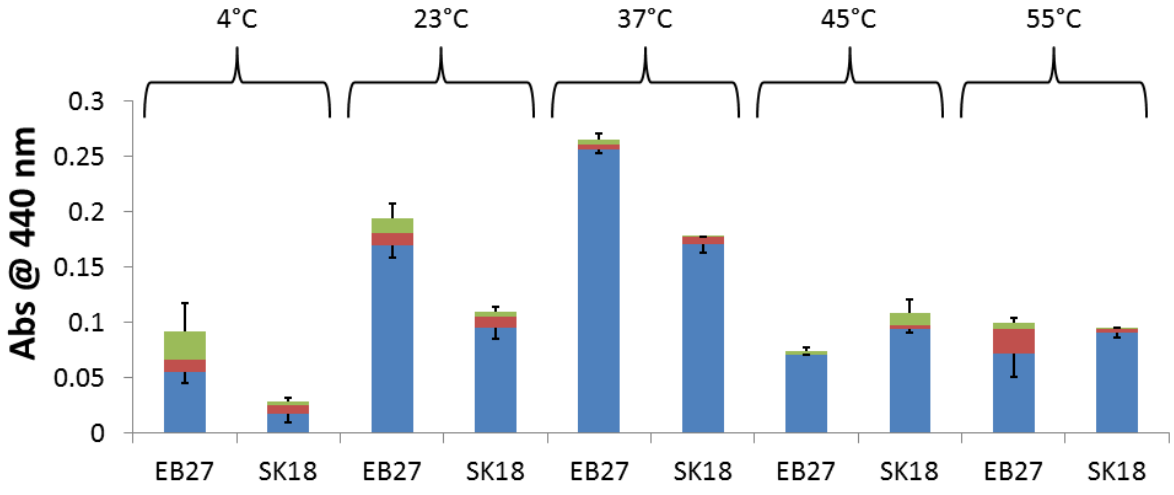



**Figure 2**: Protease activity assay of the isolates EB27 and SK18 at different temperatures (*n*=3, pH 8.0). (Barchart with integrated box-whisker plots, green colored $Q_3$ and red colored $Q_1$)

β-galactosidase assays were performed at a range of different temperatures ranging from 4°C, to 55°C for EB27 and SK20 (Figure 3). The highest activity was observed at 45°C for both strains, but with SK20 displaying higher activity at temperatures above 28°C. Thus interestingly despite the fact that these enzymes are produced by *Pseudoalteromonas* strains which were isolated from depths ranging from 1480 and 2900 metres, where temperatures are typically on average around 2°C; β-glucosidase, β-galactosidase and protease activity in these strains are not all cold adapted.

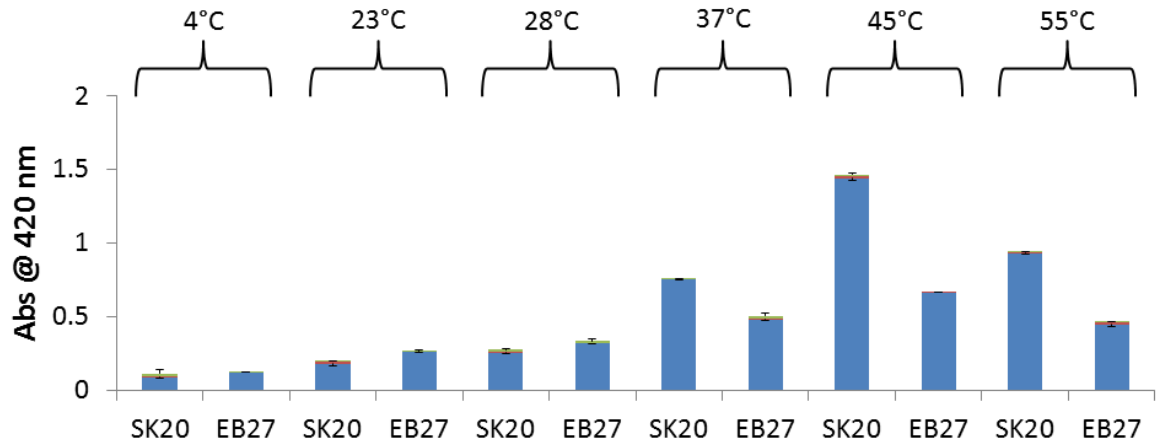



**Figure 3:** β-galactosidase assay of the isolates EB27 and SK18 at different temperatures (*n*=3, pH 8.5). ((Barchart with integrated box-whisker plots, green colored $Q_3$ and red colored $Q_1$)

To assess the general temperature-growth profile of the three isolates we performed growth experiments at different temperatures (4°C, 23°C, 28°C and 37°C) and calculated the growth rate and doubling time (Table 2). The specific growth rate ranged from 0.28 for SK18 (doubling time 159.36 min) to 0.5 for EB27 (doubling time 88.3 min) at 4°C to the optimal for EB27 at 23°C 1.03 (doubling time 40.6 min) and for SK18 and SK20 at 28°C being 2.08 and 1.41 respectively (doubling times of 20 and 29.5 min). The growth rate declined slowly for SK18 and SK20 at 37°C, 1.46 and 0.9 and for EB27 being 0.82.

**Table 2**: Growth characteristics (specific growth rate and generation time) of the *Pseudoalteromonas* sponge isolates at different temperatures (*n*=3), including standard errors.

| ID | 4°C | 23°C | 28°C | 37°C |
|---|---|---|---|---|
| | mu; $t_{gen}$ [min] | mu; $t_{gen}$ [min] | mu; $t_{gen}$[min] | mu; $t_{gen}$[min] |
| EB27 | 0.54±0.15 ; 88.3±20.01 | 1.03±0.05 ; 40.6±2.1 | 0.98±0.04 ; 42.46±1.84 | 0.82±0.09 ; 51.39±6.75 |
| SK18 | 0.28±0.06 ; 159.36±12.75 | 1.76±0.095 ; 23.8±1.24 | 2.08±0.013 ; 20±0.13 | 1.46±0.16 ; 29.04±2.9 |
| SK20 | 0.29±0.02 ; 144.92±31.2 | 0.99±0.08 ; 42.36±3.56 | 1.41±0.01 ; 29.5±0.24 | 0.9±0.14 ; 48.66±7.29 |

We decided to sequence the genomes of these three *Pseudoalteromonas* strains in an attempt to gain a better understanding of their biotechnological potential based on these preliminary extracellular enzyme profiles, together with the fact that other *Pseudoalteromonas* strains such as *Pseudoalteromonas haloplanktis* strains TAC125, TAE79, Sp22b and AS-11 have all been shown to produce a large number of biotechnologically important biocatalysts (Pulicherla KK and KRS, 2013). In addition representatives of the genus *Pseudoalteromonas* have also been shown to produce a broad array of bioactive molecules such as antibiotics, antitumor agents and toxins/antitoxins (Bosi *et al.*, 2017; Holmström and Kjelleberg, 1999; Isnansetyo and Kamei, 2003; Sannino *et al.*, 2017; Xie *et al.*, 2012).

### 3.4.2 Genome sequencing and assembly

The three *Pseudoalteromonas* genomes were sequenced on the MiSeq platform and the coverage obtained ranged from 196x to 230x. The number of identified coding DNA sequences (CDS) ranged from 3582 to 4012, with EB27 having the largest genome of 4.56 Mb and 4012 CDS and SK18 the smallest genome with 3.98 Mb and 3582 CDS (Table 3). The sequencing results fall within the size range of known *Pseudoalteromonas* spp. genomes (*Pseudoalteromonas haloplanktis* TAC125 3.85 Mb (Médigue *et al.*, 2005) to *Pseudoalteromonas atlantica* T6c 5.1 Mb (Grigoriev *et al.*, 2012; Nordberg *et al.*, 2014)), and do not appear to display any unusual patterns which may relate to their host sponge origin, like genome size or differing GC content.

**Table 3**: Genome sequencing statistics and genome features of reference strains. (CDS, coding DNA sequences, N50 weighted median length of the sequences making up 50% of genome size)

| ID | Genome size [Mb] | GC content | N50 [kb] | Contigs | CDS | No. of RNAs | Coverage |
|---|---|---|---|---|---|---|---|
| TAC125 | 3.85 | 40.1% | n/a | n/a | 3473 | 134 | n/a |
| SM9913 | 4.04 | 40.3% | n/a | n/a | 3699 | 87 | n/a |
| EB27 | 4.56 | 39.1% | 216.9 | 114 | 4012 | 136 | 196x |
| SK18 | 3.98 | 40.2% | 156.5 | 115 | 3582 | 110 | 213x |
| SK20 | 4.15 | 40.3% | 98.5 | 213 | 3811 | 139 | 230x |

### 3.4.3 Genome comparison

The BPGA pan-genome pipeline was used to compare the whole genome sequences (Figure 4)(Chaudhari *et al.*, 2016). *Pseudoalteromonas haloplanktis* TAC125 was used as a reference strain representing a shallow water isolate and *Pseudoalteromonas* sp. SM9913 was used as a deep sea reference strain as it had been retrieved from a deep sea sediment sample (1855 m) (Médigue *et al.*, 2005; Qin *et al.*, 2011). A phylogenetic comparison of the 16S rRNA gene of the isolates, reference strains and a number of relevant type strains defined our isolates as true *Pseudoalteromonas* spp. (Figure 5). The 16S rRNA gene from the strains used for the whole genome comparison in this study were identified from the respective genomes by RNAmmer (Lagesen *et al.*, 2007). A pan-genome analysis, based on the comparison of all translated protein sequences; was then performed. The number of translated protein sequences present ranged from 3422 for *Pseudoalteromonas haloplanktis* TAC125 to 3941 for *Pseudoalteromonas* sp. EB27, with 2482 of these proteins being orthologs; making up 72.5% of the smallest genome and 62.9% of the largest genome. The number of unique proteins or paralogs is quite large ranging from 308 to 809. The sponge isolates share only 10 protein clusters not found in the free living reference strains TAC125 and SM9913. These clusters include genes potentially encoding cation efflux proteins, integrases, recombinases and proteins potentially involved in multidrug resistance and which may play a role in helping the *Pseudoalteromonas* strains adapt to life inside the sponge and in helping them cope with other microorganisms inhabiting the sponge. For example recombinases and integrases are known to mediate horizontal gene transfer, which is believed to play a key role in the

genomic evolution of symbionts (Fan *et al.*, 2012). When comparing the individual isolates to the reference strains, EB27 shared 183 clusters with TAC125 and only 28 with SM9913, while SK18 shared 14 clusters with TAC125 and 75 clusters with SM9913. In addition SK20 shared 60 clusters with TAC125 and 120 clusters with SM9913.

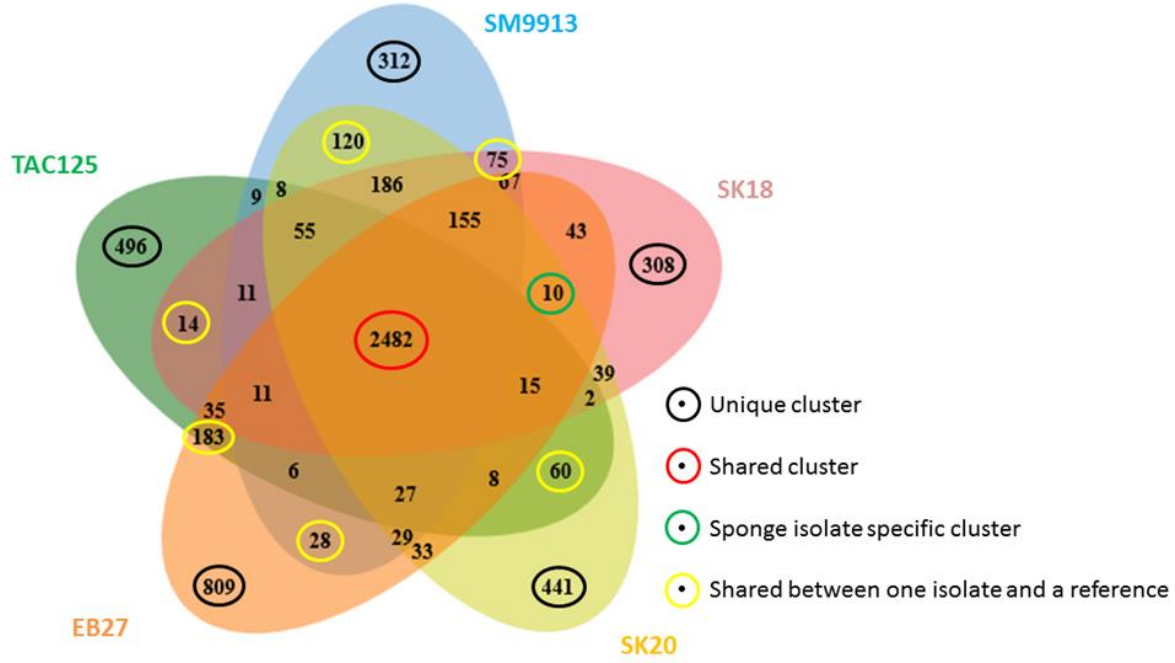



**Figure 4**: Whole genome comparison of translated non redundant protein clusters from all three isolates and the two reference genomes (generated with (Bardou *et al.*, 2014)). Green coloured is *Pseudoalteromonas haloplanktis* TAC125, blue coloured *Pseudoalteromonas* sp. SM9913, light red coloured is SK18, yellow coloured is SK20 and orange coloured is EB27.

The distribution of the protein clusters of orthologous groups (COG) affiliated with biological functions can be seen in Figure 6 (generated with (Chaudhari *et al.*, 2016)). The unique genes as mentioned earlier make up in total approximately 8.5% to 20% per genome. The potential function of these genes appears to be widespread and affiliated with many different cellular functions such as signal transduction mechanisms, cell wall, membrane and envelope biogenesis, recombination and repair and many with only general or unknown function, so that no obvious pattern is evident. The accessory genes appear to be affiliated with several functions such as signal transduction mechanisms, defense mechanism and proteins with either an as yet unknown function or only a general prediction, but again no obvious link with a specific function. The core genome mainly

contributes towards cell cycle control, cell division, chromosome partitioning, translation, ribosomal structure, biogenesis and nucleotide transport and metabolism (Figure 6). According to the COG distribution the KEGG distribution of the translated genomes can be found in the supplementary file 1.

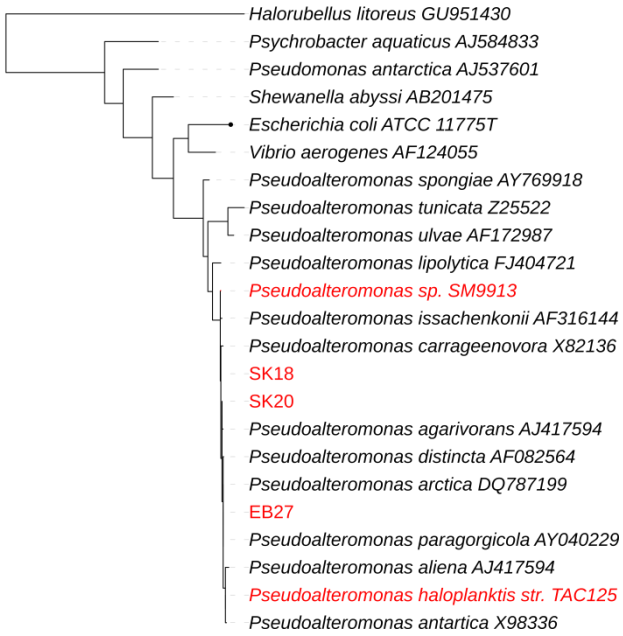



**Figure 5**: Phylogenetic comparison of the isolates investigated and reference strains (marked in red are the isolates that are used for this study. Maximum likelihood bootstrap consensus tree from 1000 repli cates, calculated with MEGA6.0 (Tamura *et al.*, 2013) and visualized with iTOL (Letunic and Bork, 2007, 2011)).

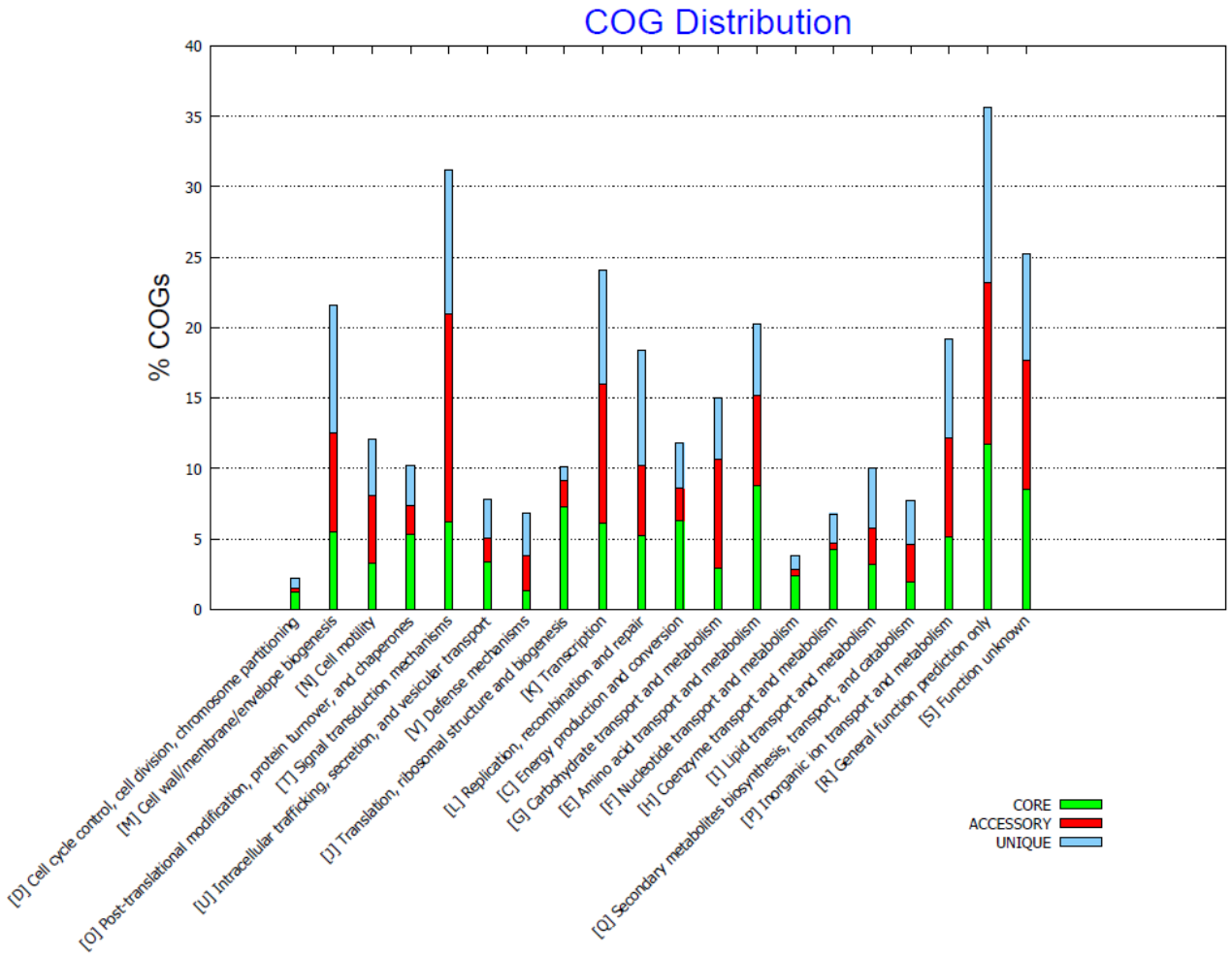



**Figure 6**: Cluster of orthologous groups (COG) distribution of the core, accessory and unique genes of the five investigated *Pseudoalteromonas* genomes (generated with (Chaudhari *et al.*, 2016)).

The pan-genome analysis revealed an open pan-genome for the five *Pseudoalteromonas* isolates investigated here. Therefore the number of dispensable or accessory genes is orders of magnitude larger than the size of the core genome and increases with the number of additional genomes (Figure 7), as defined by (Medini *et al.*, 2005). For the five genomes the pan genome contains 6077 genes and the core genome is made up of 2482 genes. This is in line with recent findings for other non-pigmented *Pseudoalteromonas* spp. (Bosi *et al.*, 2017).

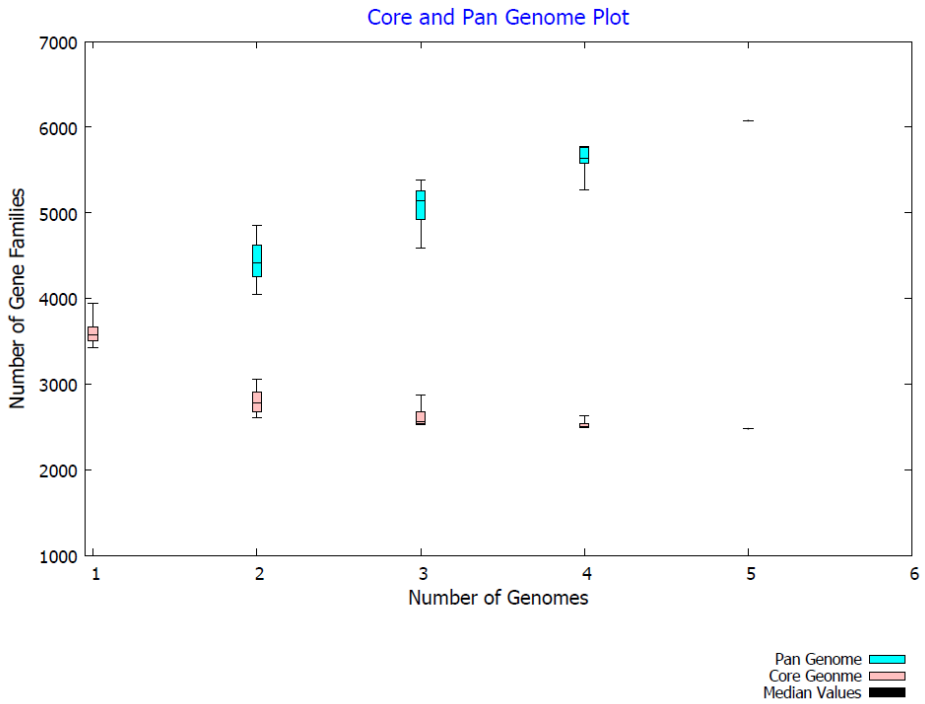


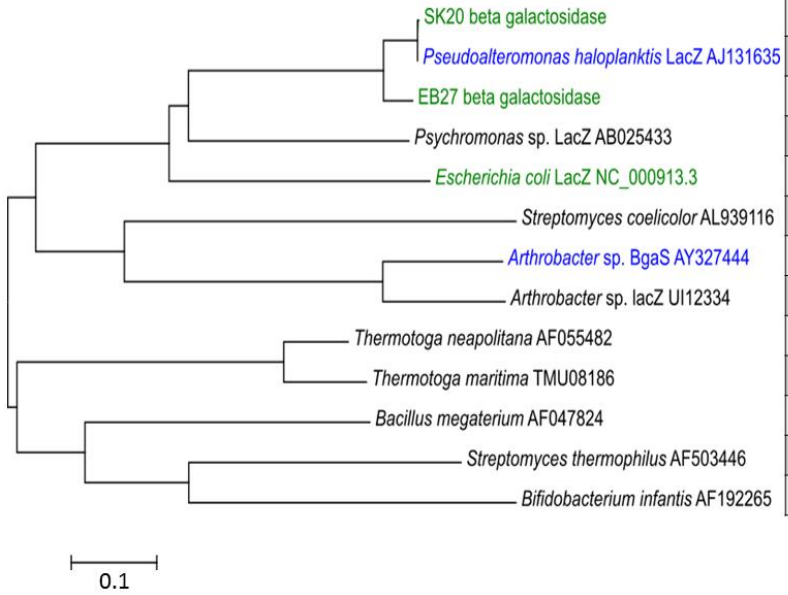

**Figure 7**: Core vs. pan-genome size plot generated with BPGA (Chaudhari *et al.*, 2016).

The genome sequences were then manually screened for genes encoding enzymes of potential industrial interest and were found to be quite rich in potential lipases/esterases and proteases and to contain a relatively small number of potential β-galactosidase, β-glucosidase and cellulase genes (Table 4). The number of potential genes encoding these enzyme activities does not however reflect the phenotypes seen in the plate screening assays (Table 1). For example the presence of β-glucosidase genes does not necessarily lead to a positive phenotypic assay for this enzyme activity, all investigated *Pseudoalteromonas* strains contain at least two β-glucosidase genes, except TAC125, but only EB27 displays this enzyme active in the plate screenings. However the increased number of potential β-glucosidase and cellulase encoding genes in the genome of EB27 may account for the positive screening results in the plate assays for these enzyme activities.

**Table 4**: Abundance of genes encoding for enzymes of potential industrial interest

| ID | Lipase/Est. | β-galactosidase | Protease | β-glucosidase | Cellulase |
|---|---|---|---|---|---|
| TAC125 | 49 | 0 | 35 | 0 | 2 |
| SM9913 | 67 | 0 | 42 | 2 | 3 |
| EB27 | 69 | 1 | 48 | 4 | 5 |
| SK18 | 63 | 0 | 39 | 2 | 3 |
| SK20 | 56 | 1 | 40 | 2 | 3 |

Given the good levels of β -galactosidase activities we observed at a range of different temperatures in both EB27 and SK20 (Figure 3) and given that only one β-galactosidases gene was present in each genome then we reasoned that these genes are likely to be responsible for the observed activity. We phylogenetically compared the β-galactosidases from EB27 and SK20 with other β-galactosidases from different bacterial strains, including cold active enzymes from *Pseudoalteromonas haloplanktis* TAE79 (Hoyoux *et al.*, 2001) and *Arthrobacter* sp. (Coker *et al.*, 2003) (Figure 8). Interestingly the β-galactosidase genes from our two isolates SK20 and EB27 are closely related to a *lacZ* gene from *Pseudoalteromonas haloplanktis* TAE79 which was found to be cold-adapted, protein sequence alignments indicate protein identity levels of 99% for SK20 and 92% for EB27.




| Protein identity | |
|---|---|
| SK20 | EB27 |
| - | 91% |
| 99% | 92% |
| 91% | - |
| 53% | 55% |
| 51% | 51% |
| 32% | 31% |
| 32% | 31% |
| 31% | 31% |
| 37% | 37% |
| 37% | 37% |
| 36% | 38% |
| 31% | 32% |
| 31% | 32% |

**Figure 8**: Phylogenetic tree of different family 2 β-galactosidase genes, levels of protein identity [%] to the β-galactosidase genes from SK20 and EB27 are given in the table. Mesophilic β-galactosidase genes are colored in green, true cold-adapted β-galactosidase genes are colored in blue and the genes written in black have not been investigated for optimal temperature to date. The tree was generated with the maximum-likelihood method and 500 bootstrap replicate.

Deferred antagonism based antimicrobial assays were also performed in an effort to determine whether the three *Pseudoalteromonas* strains displayed any bioactivity against clinically relevant pathogens. The isolates were grown on both low and rich nutrient media and then overlaid with a number of clinically relevant test strains such as *Escherichia coli* 12210, *Staphylococcus aureus* NCD0 949, *Bacillus subtilis* 1E32, *Pseudomonas aeruginosa* PA-O1, *Acinetobacter johnsonii* WH00185, *Enterobacter faecium* NCIMB 11508, *Klebsiella pneumonia* NCIMB 13218 and *Enterobacter aerogenes* NCIMB 10102 in soft LB-agar. No bioactivity was observed, despite the fact that all three *Pseudoalteromonas* genomes contained at least one potential bacteriocin gene cluster (EB27 contained two bacteriocin gene clusters), with SK18 and EB27 also containing potential arylpolyene and siderophore encoding gene cluster (Table 5).

**Table 5**: Abundance of secondary metabolite gene clusters

| ID | Bacteriocin | Arylpolyene | Siderophore |
|---|---|---|---|
| TAC125 | 1 | 1 | - |
| SM9913 | 1 | - | 1 |
| EB27 | 2 | 1 | - |
| SK18 | 1 | 1 | 1 |
| SK20 | 1 | - | - |

In addition when TAC125 and SM9913 were subsequently analysed, one potential bacteriocin gene cluster was found to be present and highly conserved between the different isolates and the reference genomes (Figure 9). This gene cluster consists of 13 different genes with an average total size of 10.8 kb, except in SK20 which only consists of seven genes with a total of 6.1 kb (Figure 9). In addition to the conserved bacteriocin gene cluster that can be found in all isolates, EB27 has a second small bacteriocin gene cluster spanning 10 kb, which is considerably different from the other clusters and is not a reduced form of the conserved cluster as in SK20.

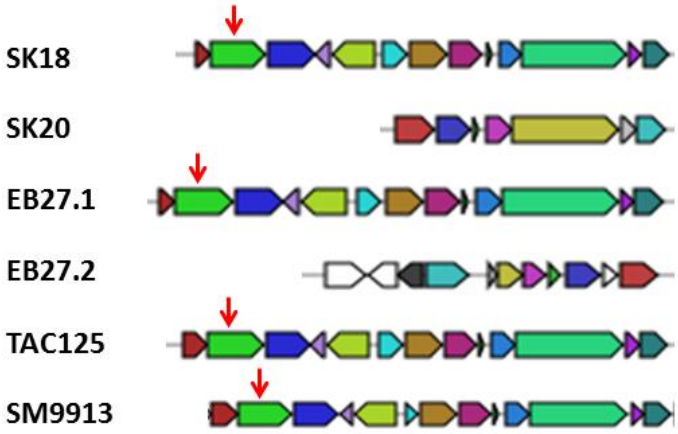



**Figure 9**: Organization of the Bacteriocin gene clusters found in the investigated genomes (adapted from antiSMASH (Blin *et al.*, 2013; Medema *et al.*, 2011; Weber *et al.*, 2015)). Red coloured genes are biosynthetic genes, blue coloured transport-related genes, green coloured regulatory genes and grey coloured other/unidentified genes. The red arrow points add the gene containing a tetratricopeptide repeat.

Given that these *Pseudoalteromonas* spp. had been isolated from different sources such as sponges (EB27, SK18, SK20), deep sea sediment (SM9913), as well as from open Antarctic seawater, we decided to investigate the presence of potential eukaryotic-like proteins such as ankyrin-repeats (ANK) and tetratricopeptide repeats domain-encoding proteins (TRP) which are known to be enriched in symbiotic sponge associated microorganisms (Reynolds and Thomas, 2016) (Table 6). The genomes of all isolates contain a small number of genes with ankyrin and tetratricopeptide repeats, which are present at much lower levels than might be expected from a true sponge symbiont such as *Poribacteria* sp. which contain at least 23 genes with tetratricopeptides repeats in its genome (Siegl *et al.*, 2011).

**Table 6**: Abundance of genes suggested being involved in a symbiotic relationship

| ID | Ankyrin repeats | Tetratricopeptide repeats | Nitrite reductase | Proteases | Sulfatases | Peptidases |
|---|---|---|---|---|---|---|
| TAC125 | 2 | 2 | 1 | 35 | 0 | 58 |
| SM9913 | 1 | 2 | 0 | 42 | 1 | 63 |
| EB27 | 2 | 2 | 3 | 48 | 0 | 65 |
| SK18 | 1 | 2 | 0 | 39 | 1 | 63 |
| SK20 | 1 | 2 | 0 | 40 | 1 | 58 |

## 3.5 Discussion

*Pseudoalteromonas* spp. are known to be multitalented with respect to the production of enzymes of industrial interest; with for example agarases (Oh *et al.*, 2010), galactosidases (Cieśliński *et al.*, 2005), proteases (Lee *et al.*, 2002), subtilases (Yan *et al.*, 2009) and phospholipases (Mo et al., 2009) from this genus being described. Furthermore some isolates are also able to produce acidic exopolysaccharides involved in biofilm formation (Bartlett *et al.*, 1988) as well as antimicrobial compounds (Longeon *et al.*, 2004; Zhang and Enomoto, 2011). In general the *Pseudoalteromonas* genus can be divided into pigmented and non-pigmented species, with the first producing mostly antimicrobial and antifouling compounds and the latter being more versatile in the production of different enzymes (Bowman, 2007). The *Pseudoalteromonas* spp. isolates described herein are naturally non-

pigmented and are therefore no exception to the aforementioned general classification as they produce a variety of different enzymes, but display no antimicrobial activity under the assay conditions tested, but interestingly they do possess potential bacteriocin and siderophore gene clusters in their genomes (Table 5, Figure 9).

Having isolated a number of *Pseudoalteromonas* strains from deep sea sponges we decided to employ a number of approaches including plate screening, whole genome sequencing and comparative genomics in an attempt to identify genes encoding enzymes with potentially biotechnologically relevant properties. The isolates were found to be cold adapted rather than true psychrophiles according to the observed growth rates and doubling times at various temperatures (Table 2). The isolates grew best at 23°C and 28°C with doubling times ranging from 20 to 40 minutes, compared to a generation time of 20 minutes at 37°C for the mesophilic *E. coli* K-12 strain MG1655 (Sezonov *et al.*, 2007), which shows that our isolates are able to achieve similar growth rates to *E. coli* already at room temperatures, supporting the feasibility to use *Pseudoalteromonas* as an expression system for cold adapted enzymes as demonstrated by Papa *et al.*, 2007. The isolates displayed a number of different enzyme activities including, β-glucosidase, β-galactosidase, protease, and lipase activities (Table 1). We further characterized the β-glucosidase, β-galactosidase and protease activities to investigate possible cold adaptation due to the deep sea origin of the isolates (Figure 1, 2, 3). The protease and β-galactosidase activities were found to be in the mesophilic range, and to be slightly alkaline-active in nature, being optimally active at temperatures between 37°C to 45°C and at a pH of 8.5 (Figure 2, 3). Proteases have an important role in industrial biotechnology, particularly in the detergent, food and pharmaceutical areas. They also find utility as antifouling compounds, with proteases from a deep sea sediment isolated *Pseudoalteromonas* sp. inhibiting larval attachment of the bryozoan *Bugula neritina* (Dobretsov *et al.*, 2007).

A β-galactosidase gene was identified in both SK20 and EB27 and following construction of a phylogenetic tree with relevant closely related β-galactosidases (Figure 8), were found to be closely related to the *lacZ* gene from *Pseudoalteromonas haloplanktis* TAE79 (92% to 99% protein identity). However in contrast to the *P. haloplanktis* LacZ protein which has been reported to be cold active (Hoyoux *et al.*, 2001), the β-galactosidase activity

encoded in the genomes of the SK20 and EB27 isolates appears to be mesophilic in nature, with an optimal temperature of 45°C.

Good levels of β-glucosidase activity were observed in EB27, at temperatures ranging from 4°C to 37°C, with optimal activity at 23°C at a pH of 8.5 (Figure 1). While β-glucosidases are typically involved in important processes in bacteria such as degradation of cellulose and other carbohydrates for nutrient uptake, there is an increased interest in their use in the conversion of lignocellulosic biomass into reducing sugars for ethanol production. While at least two other types of enzymes are also required for the complete degradation of cellulose, namely the endoglucanases and cellobiohydrolases; β-glucosidases are mostly attributed as being the rate limiting enzyme in these processes (Sørensen *et al.*, 2013). They also find industrial applications in wine making where they play a key role in the enzyme mediated release of aromatic compounds from glycosidic precursors present in fruit juices, musts and fermenting products. They are also used in flavour enhancement to improve the organoleptic properties of citrus fruit juices to reduce bitterness (Singh *et al.*, 2016). While the majority of β-glucosidases currently in use are mostly fungal in origin, bacterial derived enzymes are receiving increased, recent interest particularly for biofuel production applications (Singh *et al.*, 2016). Furthermore enzymes from *Pseudoalteromonas* have proven useful in the hydrolysis carbohydrates from algal biomass under alkaline conditions, which is uncommon for terrestrial β-glucosidases and could be used for biofuel production from marine sources ( Matsumoto *et al.*, 2003; M. *et al.*, 2014; Singh *et al.*, 2016).

As mentioned earlier, members of the genus *Pseudoalteromonas* are routinely isolated from a variety of different marcoorganisms. While they have also been isolated from sea water and sediment, they are usually found in association with macroorganisms (Bowman, 2007; Offret *et al.*, 2016). With this in mind we investigated the genomes of our three deep sea sponge associated *Pseudoalteromonas* strains, together with two free living isolates for the presence of potential symbiosis genes, such as genes mediating microbe-host interactions (genes containing eukaryotic-like domains, like ankyrin and tetratricopeptide repeats) or those that may be beneficial in the acquisition or production of nutrients such as proteases, sulfatases or peptidases for the host or the symbiont (Siegl *et al.*, 2011; Kamke *et al.*, 2012). While the genomes of the sponge associated and free living *Pseudoalteromonas* sp. isolates were rich in proteases and sulfatases, they lacked large numbers of genes encoding potential

ankyrin and tetratricopeptide repeats (Table 6). Interestingly one of the genes which did contain a tetratricopeptide repeat is part of the conserved bacteriocin gene cluster that is found in all isolates (Figure 9), which provides some limited evidence of potential microbe-host interactions. In this respect it is known that some bacteriocins are involved in mediating microbe-host interactions via biofilm formation on which a host can settle (Shikuma *et al.*, 2014). However the lack of appreciable numbers of genes containing eukaryotic-like domains amongst the genomes of the *Pseudoalteromonas* sponge isolates appears to suggest that they may not form a true symbiotic relationship with their host. These *Pseudoalteromonas* isolates may however be indirectly beneficial to the sponge by breaking down polysaccharides or other nutrient containing materials and thereby making these available to both themselves and to the sponge; by functioning as either a commensal or transiently associated microbe.

The pan genome of the investigated isolates is open and each genome contains 8.5% to 20% unique genes and the core genome comprises of 2482 genes (Figure 5), whereas the whole pan genome comprises of 6077 genes. The possible functions of the translated genes have been investigated, by assigning them to clusters of orthologous function. The distribution of the unique, accessory and core genes is widespread across different biological functions and no obvious pattern is evident, besides core functions that seem to be conserved in all genomes such as cell cycle control, cell division, chromosome partitioning, translation, ribosomal structure, biogenesis and nucleotide transport and metabolism (Figure 6).

Thus in conclusion following a comparative genomic analysis of non-pigmented sponge associated *Pseudoalteromonas* sp. isolated from different depths and free-living *Pseudoalteromonas* sp. we have demonstrated that these strains share a large open pan-genome and possess a considerable number of unique genes which is in line with results of other genome comparison of non-pigmented *Pseudoalteromonas* spp. (Bosi *et al.*, 2017). We were unable to obtain definitive evidence based on these genome comparisons that non-pigmented *Pseudoalteromonas* spp. form true symbiotic relationships with deep sea sponges. While these non-pigmented strains do not appear to produce antimicrobial compounds; they do however produce a wide variety of different degradative enzymes, such as proteases, lipases, β-glucosidases and β-galactosidases. These enzymes appear to possess

specific industrially important characteristics such as cold-adaptation and activity in the alkaline pH range and are therefore likely to be of interest to different industrial applications. Heterologous expression of these genes in suitable host systems such as *Escherichia coli* may prove useful in their future characterization and in providing sufficient quantities for laboratory scale application studies.

*Stephen A. Jackson, Ragnar Jóhansson, Viggó T. Marteinsson and Alan D.W. Dobson conceived and designed the study; Erik Borchert primarily performed the experiments and analyzed the data, Stephen Knobloch performed the genome sequencing, Emilie Dwyer and Sinéad O'Flynn carried out the enzyme assays.

### 3.6. Bibliography


Akagawa-Matsushita, M, M. M, K. Y, and Y. K, 1992, *Alteromonas atlantica* sp. nov and *Alteromonas carrageenovora* sp. nov., bacteria that decompose algal polysaccharides. Int J Syst Evol Microbiol, v. 42, p. 621-627.

Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, and O. Zagnitko, 2008, The RAST Server: rapid annotations using subsystems technology: BMC Genomics, v. 9, p. 75.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, and P. A. Pevzner, 2012, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing: J Comput Biol, v. 19, p. 455-77.

Bardou, P., J. Mariette, F. Escudié, C. Djemiel, and C. Klopp, 2014, jvenn: an interactive Venn diagram viewer: BMC Bioinformatics, v. 15, p. 293.

Bartlett, D. H., M. E. Wright, and M. Silverman, 1988, Variable expression of extracellular polysaccharide in the marine bacterium *Pseudomonas atlantica* is controlled by genome rearrangement: Proc Natl Acad Sci U S A, v. 85, p. 3923-7.

Blin, K., M. H. Medema, D. Kazempour, M. A. Fischbach, R. Breitling, E. Takano, and T. Weber, 2013, antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers: Nucleic Acids Res, v. 41, p. W204-12.

Bolger, A. M., M. Lohse, and B. Usadel, 2014, Trimmomatic: a flexible trimmer for Illumina sequence data: Bioinformatics, v. 30, p. 2114-20.

Borchert, E., S. A. Jackson, F. O'Gara, and A. D. Dobson, 2016, Diversity of natural product biosynthetic genes in the microbiome of the deep sea sponges *Inflatella pellicula*, *Poecillastra compressa*, and *Stelletta normani*: Front Microbiol, v. 7, p. 1027.

Bosi, E., M. Fondi, V. Orlandini, E. Perrin, I. Maida, D. de Pascale, M. L. Tutino, E. Parrilli, A. Lo Giudice, A. Filloux, and R. Fani, 2017, The pangenome of (Antarctic) *Pseudoalteromonas* bacteria: evolutionary and functional insights: BMC Genomics, v. 18, p. 93.

Bowman, J. P., 1998, *Pseudoalteromonas prydzensis* sp. nov., a psychrotrophic, halotolerant bacterium form Antarctic sea ice: Int J Syst Bacteriol, v. 48 Pt 3, p. 1037-41.

Bowman, J. P., 2007, Bioactive compound synthetic capacity and ecological significance of marine bacterial genus *Pseudoalteromonas*: Mar Drugs, v. 5, p. 220-41.

Bozal, N., E. Tudela, R. Rosselló-Mora, J. Lalucat, and J. Guinea, 1997, *Pseudoalteromonas antarctica* sp. nov., isolated from an Antarctic coastal environment: Int J Syst Bacteriol, v. 47, p. 345-51.

Brettin, T., J. J. Davis, T. Disz, R. A. Edwards, S. Gerdes, G. J. Olsen, R. Olson, R. Overbeek, B. Parrello, G. D. Pusch, M. Shukla, J. A. Thomason, R. Stevens, V. Vonstein, A. R. Wattam, and F. Xia, 2015, RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes: Sci Rep, v. 5, p. 8365.

Cavicchioli, R., K. S. Siddiqui, D. Andrews, and K. R. Sowers, 2002, Low-temperature extremophiles and their applications: Curr Opin Biotechnol, v. 13, p. 253-61.

Chaudhari, N. M., V. K. Gupta, and C. Dutta, 2016, BPGA- an ultra-fast pan-genome analysis pipeline: Sci Rep, v. 6, p. 24373.

Cieśliński, H., J. Kur, A. Białkowska, I. Baran, K. Makowski, and M. Turkiewicz, 2005, Cloning, expression, and purification of a recombinant cold-adapted beta-galactosidase from antarctic bacterium *Pseudoalteromonas* sp. 22b: Protein Expr Purif, v. 39, p. 27-34.

Coker, J. A., P. P. Sheridan, J. Loveland-Curtze, K. R. Gutshall, A. J. Auman, and J. E. Brenchley, 2003, Biochemical characterization of a beta-galactosidase with a low temperature optimum obtained from an Antarctic *Arthrobacter* isolate: J Bacteriol, v. 185, p. 5473-82.

Dobretsov, S., H. Xiong, Y. Xu, L. A. Levin, and P. Y. Qian, 2007, Novel antifoulants: inhibition of larval attachment by proteases: Mar Biotechnol (NY), v. 9, p. 388-97.

Egan, S., C. Holmström, and S. Kjelleberg, 2001, *Pseudoalteromonas ulvae* sp. nov., a bacterium with antifouling activities isolated from the surface of a marine alga: Int J Syst Evol Microbiol, v. 51, p. 1499-504.

Egan, S., S. James, C. Holmström, and S. Kjelleberg, 2002, Correlation between pigmentation and antifouling compounds produced by *Pseudoalteromonas tunicata*: Environ Microbiol, v. 4, p. 433-42.

Fan, L., D. Reynolds, M. Liu, M. Stark, S. Kjelleberg, N. S. Webster, and T. Thomas, 2012, Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts: Proc Natl Acad Sci U S A, v. 109, p. E1878-87.

Fehér, D., R. Barlow, J. McAtee, and T. K. Hemscheidt, 2010, Highly brominated antimicrobial metabolites from a marine *Pseudoalteromonas* sp: J Nat Prod, v. 73, p. 1963-6.

Gauthier, G., M. Gauthier, and R. Christen, 1995, Phylogenetic analysis of the genera *Alteromonas*, *Shewanella*, and *Moritella* using genes coding for small-subunit rRNA sequences and division of the genus *Alteromonas* into two genera, *Alteromonas* (emended) and *Pseudoalteromonas* gen. nov., and proposal of twelve new species combinations: Int J Syst Bacteriol, v. 45, p. 755-61.

Georlette, D., Z. O. Jónsson, F. Van Petegem, J. Chessa, J. Van Beeumen, U. Hübscher, and C. Gerday, 2000, A DNA ligase from the psychrophile *Pseudoalteromonas haloplanktis* gives insights into the adaptation of proteins to low temperatures: Eur J Biochem, v. 267, p. 3502-12.

Grigoriev, I. V., H. Nordberg, I. Shabalov, A. Aerts, M. Cantor, D. Goodstein, A. Kuo, S. Minovitsky, R. Nikitin, R. A. Ohm, R. Otillar, A. Poliakov, I. Ratnere, R. Riley, T. Smirnova, D. Rokhsar, and I. Dubchak, 2012, The genome portal of the Department of Energy Joint Genome Institute: Nucleic Acids Res, v. 40, p. D26-32.

Holmström, C., S. James, S. Egan, and S. Kjelleberg, 1996, Inhibition of common fouling organisms by marine bacterial isolates with special reference to the role of pigmented bacteria: Biofouling, v. 10, p. 251-9.

Holmström, C., S. James, B. A. Neilan, D. C. White, and S. Kjelleberg, 1998, *Pseudoalteromonas tunicata* sp. nov., a bacterium that produces antifouling agents: Int J Syst Bacteriol, v. 48 Pt 4, p. 1205-12.

Holmström, C., and S. Kjelleberg, 1999, Marine *Pseudoalteromonas* species are associated with higher organisms and produce biologically active extracellular agents: FEMS Microbiol Ecol, v. 30, p. 285-293.

Hoyoux, A., I. Jennes, P. Dubois, S. Genicot, F. Dubail, J. M. François, E. Baise, G. Feller, and C. Gerday, 2001, Cold-adapted beta-galactosidase from the Antarctic psychrophile *Pseudoalteromonas haloplanktis*: Appl Environ Microbiol, v. 67, p. 1529-35.

Isnansetyo, A., and Y. Kamei, 2003, MC21-A, a bactericidal antibiotic produced by a new marine bacterium, *Pseudoalteromonas phenolica* sp. nov. O-BC30(T), against methicillin-resistant *Staphylococcus aureus*: Antimicrob Agents Chemother, v. 47, p. 480-8.

Ivanova, E. P., E. A. Kiprianova, V. V. Mikhailov, G. F. Levanova, A. D. Garagulya, N. M. Gorshkova, M. V. Vysotskii, D. V. Nicolau, N. Yumoto, T. Taguchi, and S. Yoshikawa, 1998, Phenotypic diversity of *Pseudoalteromonas citrea* from different marine habitats and emendation of the description: Int J Syst Bacteriol, v. 48 Pt 1, p. 247-56.

Ivanova, E. P., L. S. Shevchenko, T. Sawabe, A. M. Lysenko, V. I. Svetashev, N. M. Gorshkova, M. Satomi, R. Christen, and V. V. Mikhailov, 2002, *Pseudoalteromonas maricaloris* sp. nov., isolated from an Australian sponge, and reclassification of [*Pseudoalteromonas aurantia*] NCIMB 2033 as *Pseudoalteromonas flavipulchra* sp. nov: Int J Syst Evol Microbiol, v. 52, p. 263-71.

Jackson, S. A., B. Flemer, A. McCann, J. Kennedy, J. P. Morrissey, F. O'Gara, and A. D. Dobson, 2013, Archaea appear to dominate the microbiome of *Inflatella pellicula* deep sea sponges: PLoS One, v. 8, p. e84438.

Kamke, J., K. Bayer, T. Woyke, and U. Hentschel, 2012, Exploring symbioses by single-cell genomics: Biol Bull, v. 223, p. 30-43.

Kennedy, J., B. Flemer, S. A. Jackson, J. P. Morrissey, F. O'Gara, and A. D. Dobson, 2014, Evidence of a putative deep sea specific microbiome in marine sponges: PLoS One, v. 9, p. e91092.

Lagesen, K., P. Hallin, E. A. Rødland, H. H. Staerfeldt, T. Rognes, and D. W. Ussery, 2007, RNAmmer: consistent and rapid annotation of ribosomal RNA genes: Nucleic Acids Res, v. 35, p. 3100-8.

Lee, S. O., J. Kato, K. Nakashima, A. Kuroda, T. Ikeda, N. Takiguchi, and H. Ohtake, 2002, Cloning and characterization of extracellular metal protease gene of the algicidal

marine bacterium *Pseudoalteromonas* sp. strain A28: Biosci Biotechnol Biochem, v. 66, p. 1366-9.

Letunic, I., and P. Bork, 2007, Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation: Bioinformatics, v. 23, p. 127-8.

Letunic, I., and P. Bork, 2011, Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy: Nucleic Acids Res, v. 39, p. W475-8.

Li, X. D., X. M. Li, X. Li, G. M. Xu, Y. Liu, and B. G. Wang, 2016, Aspewentins D-H, 20-Nor-isopimarane derivatives from the deep sea sediment-derived fungus *Aspergillus wentii* SD-310: J Nat Prod, v. 79, p. 1347-53.

Longeon, A., J. Peduzzi, M. Barthélemy, S. Corre, J. L. Nicolas, and M. Guyot, 2004, Purification and partial identification of novel antimicrobial protein from marine bacterium *Pseudoalteromonas* species strain X153: Mar Biotechnol (NY), v. 6, p. 633-41.

M. M. El-Naggar, U. M. Abdul-Rouf, H. A. H. Ibrahim, and W. M. M. El-Sayed, 2014, Saccharification of *Ulva lactuca* via *Pseudoalteromonas piscicida* for biofuel production: Journal of Energy and Natural Resources, v. 3, p. 77-84.

Matsumoto, M., H. Yokouchi, N. Suzuki, H. Ohata, and T. Matsunaga, 2003, Saccharification of marine microalgae using marine bacteria for ethanol production: Appl Biochem Biotechnol, v. 105 -108, p. 247-54.

Medema, M. H., K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski, M. A. Fischbach, T. Weber, E. Takano, and R. Breitling, 2011, antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences: Nucleic Acids Res, v. 39, p. W339-46.

Medini, D., C. Donati, H. Tettelin, V. Masignani, and R. Rappuoli, 2005, The microbial pan-genome: Curr Opin Genet Dev, v. 15, p. 589-94.

Mo, S., J. H. Kim, and K. W. Cho, 2009, A novel extracellular phospholipase C purified from a marine bacterium, *Pseudoalteromonas* sp. J937: Biotechnol Lett, v. 31, p. 89-94.

Médigue, C., E. Krin, G. Pascal, V. Barbe, A. Bernsel, P. N. Bertin, F. Cheung, S. Cruveiller, S. D'Amico, A. Duilio, G. Fang, G. Feller, C. Ho, S. Mangenot, G. Marino, J. Nilsson, E. Parrilli, E. P. Rocha, Z. Rouy, A. Sekowska, M. L. Tutino, D. Vallenet, G. von Heijne, and A. Danchin, 2005, Coping with cold: the genome of the versatile marine

Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125: Genome Res, v. 15, p. 1325-35.

Nordberg, H., M. Cantor, S. Dusheyko, S. Hua, A. Poliakov, I. Shabalov, T. Smirnova, I. V. Grigoriev, and I. Dubchak, 2014, The genome portal of the Department of Energy Joint Genome Institute: 2014 updates: Nucleic Acids Res, v. 42, p. D26-31.

Offret, C., F. Desriac, P. Le Chevalier, J. Mounier, C. Jégou, and Y. Fleury, 2016, Spotlight on Antimicrobial Metabolites from the Marine Bacteria *Pseudoalteromonas*: Chemodiversity and Ecological Significance: Mar Drugs, v. 14.

Oh, C., C. Nikapitiya, Y. Lee, I. Whang, S. J. Kim, D. H. Kang, and J. Lee, 2010, Cloning, purification and biochemical characterization of beta agarase from the marine bacterium *Pseudoalteromonas* sp. AG4: J Ind Microbiol Biotechnol, v. 37, p. 483-94.

Overbeek, R., R. Olson, G. D. Pusch, G. J. Olsen, J. J. Davis, T. Disz, R. A. Edwards, S. Gerdes, B. Parrello, M. Shukla, V. Vonstein, A. R. Wattam, F. Xia, and R. Stevens, 2014, The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST): Nucleic Acids Res, v. 42, p. D206-14.

Papa, R., V. Rippa, G. Sannia, G. Marino, and A. Duilio, 2007, An effective cold inducible expression system developed in *Pseudoalteromonas haloplanktis* TAC125: J Biotechnol, v. 127, p. 199-210.

Pulicherla KK, and S. R. KRS, 2013, Marine biocatalysts and their stability: molecular approach. In: Marine Enzymes for Biocatalysis- Sources, Biocatalytic Characteristics and Bioprocesses of Marine Enzymes., *in* A. Trincone, ed., Marine Enzymes for Biocatalysis - Sources, biocatalytic characteristics and bioprocesses of marine enzymes, v. 38, Woodhead Publishing Series in Biomedicine, p. 109-130.

Qin, Q. L., Y. Li, Y. J. Zhang, Z. M. Zhou, W. X. Zhang, X. L. Chen, X. Y. Zhang, B. C. Zhou, L. Wang, and Y. Z. Zhang, 2011, Comparative genomics reveals a deep-sea sediment-adapted life style of *Pseudoalteromonas* sp. SM9913: ISME J, v. 5, p. 274-84.

Ramirez-Llodra, E., A. Brandt, R. Danovaro, B. De Mol, E. Escobar, C. German, L. Levin, P. Arbizu, L. Menot, and P. Buhl-Mortensen, 2010, Deep, diverse and definitely different: unique attributes of the world's largest ecosystem: Biogeosciences, v. 7, p. p.2851-2899.

Reynolds, D., and T. Thomas, 2016, Evolution and function of eukaryotic-like proteins from sponge symbionts: Mol Ecol, v. 25, p. 5242-5253.

Sannino, F., E. Parrilli, G. A. Apuzzo, D. de Pascale, P. Tedesco, I. Maida, E. Perrin, M. Fondi, R. Fani, G. Marino, and M. L. Tutino, 2017, *Pseudoalteromonas haloplanktis* produces methylamine, a volatile compound active against *Burkholderia cepacia* complex strains: N Biotechnol, v. 35, p. 13-18.

Santiago, M., C. A. Ramírez-Sarmiento, R. A. Zamora, and L. P. Parra, 2016, Discovery, molecular mechanisms, and industrial applications of cold-active enzymes: Front Microbiol, v. 7, p. 1408.

Sezonov, G., D. Joseleau-Petit, and R. D'Ari, 2007, *Escherichia coli* physiology in Luria-Bertani broth: J Bacteriol, v. 189, p. 8746-9.

Shao, X., L. Y. Ran, C. Liu, X. L. Chen, X. Y. Zhang, Q. L. Qin, B. C. Zhou, and Y. Z. Zhang, 2015, Culture condition optimization and pilot scale production of the M12 metalloprotease myroilysin produced by the deep-sea bacterium *Myroides profundi* D25: Molecules, v. 20, p. 11891-901.

Shikuma, N. J., M. Pilhofer, G. L. Weiss, M. G. Hadfield, G. J. Jensen, and D. K. Newman, 2014, Marine tubeworm metamorphosis induced by arrays of bacterial phage tail-like structures: Science, v. 343, p. 529-33.

Siegl, A., J. Kamke, T. Hochmuth, J. Piel, M. Richter, C. Liang, T. Dandekar, and U. Hentschel, 2011, Single-cell genomics reveals the lifestyle of *Poribacteria*, a candidate phylum symbiotically associated with marine sponges: ISME J, v. 5, p. 61-70.

Simidu, U., K. Kita-Tsukamoto, T. Yasumoto, and M. Yotsu, 1990, Taxonomy of four marine bacterial strains that produce tetrodotoxin: Int J Syst Bacteriol, v. 40, p. 331-6.

Singh, G., A. K. Verma, and V. Kumar, 2016, Catalytic properties, functional attributes and industrial applications of β-glucosidases: 3 Biotech, v. 6.

Sipkema, D., 2016, Marine biotechnology: diving deeper for drugs: Microb Biotechnol.

Sørensen, A., M. Lübeck, P. S. Lübeck, and B. K. Ahring, 2013, Fungal Beta-glucosidases: a bottleneck in industrial use of lignocellulosic materials: Biomolecules, v. 3, p. 612-31.

Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar, 2013, MEGA6: Molecular Evolutionary Genetics Analysis version 6.0: Mol Biol Evol, v. 30, p. 2725-9.

Truong, L. V., H. Tuyen, E. Helmke, L. T. Binh, and T. Schweder, 2001, Cloning of two pectate lyase genes from the marine Antarctic bacterium *Pseudoalteromonas haloplanktis* strain ANT/505 and characterization of the enzymes: Extremophiles, v. 5, p. 35-44.

Weber, T., K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Bruccoleri, S. Y. Lee, M. A. Fischbach, R. Müller, W. Wohlleben, R. Breitling, E. Takano, and M. H. Medema, 2015, antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters: Nucleic Acids Res, v. 43, p. W237-43.

Wu, B., J. Wiese, R. Schmaljohann, and J. F. Imhoff, 2016, Biscogniauxone, a new isopyrrolonaphthoquinone compound from the fungus *Biscogniauxia mediterranea* isolated from deep-sea sediments: Mar Drugs, v. 14.

Xie, B. B., Y. L. Shu, Q. L. Qin, J. C. Rong, X. Y. Zhang, X. L. Chen, B. C. Zhou, and Y. Z. Zhang, 2012, Genome sequence of the cycloprodigiosin-producing bacterial strain *Pseudoalteromonas rubra* ATCC 29570(T): J Bacteriol, v. 194, p. 1637-8.

Xu, W., K. L. Pang, and Z. H. Luo, 2014, High fungal diversity and abundance recovered in the deep-sea sediments of the Pacific Ocean: Microb Ecol, v. 68, p. 688-98.

Yan, B. Q., X. L. Chen, X. Y. Hou, H. He, B. C. Zhou, and Y. Z. Zhang, 2009, Molecular analysis of the gene encoding a cold-adapted halophilic subtilase from deep-sea psychrotolerant bacterium *Pseudoalteromonas* sp. SM9913: cloning, expression, characterization and function analysis of the C-terminal PPC domains: Extremophiles, v. 13, p. 725-33.

Yang, J., Y. Yu, B. L. Tang, S. Zhong, M. Shi, B. B. Xie, X. Y. Zhang, B. C. Zhou, Y. Z. Zhang, and X. L. Chen, 2016, Pilot-scale production and thermostability improvement of the M23 protease pseudoalterin from the deep sea bacterium *Pseudoalteromonas* sp. CF6-2: Molecules, v. 21.

Yoshikawa, K., T. Takadera, K. Adachi, M. Nishijima, and H. Sano, 1997, Korormicin, a novel antibiotic specifically active against marine gram-negative bacteria, produced by a marine bacterium: J Antibiot (Tokyo), v. 50, p. 949-53.

Zhang, X., and K. Enomoto, 2011, Characterization of a gene cluster and its putative promoter region for violacein biosynthesis in *Pseudoalteromonas* sp. 520P1: Appl Microbiol Biotechnol, v. 90, p. 1963-71.

# Chapter 4

## Characterization of a novel cold active deep sea esterase from a metagenomic library from the sponge *Stelletta normani*

**Erik Borchert[1], Joseph Selvin[2], Seghal G. Kiran[3], Stephen A. Jackson[1], Fergal O'Gara[1,4,5], Alan D.W. Dobson[1,6]**

[1]School of Microbiology, University College Cork, National University of Ireland, Cork, Ireland
[2]Department of Microbiology, School of Life Sciences, Pondicherry University, Pondicherry, India
[3]Department of Food Science and Technology, Pondicherry University, Pondicherry, India
[4]Biomerit Research Centre, University College Cork, National University College Cork, Cork, Ireland
[5]School of Biomedical Sciences, Curtin Health Innovation Research Institute, Curtin University, Perth, Australia
[6]Environmental Research Institute, University College Cork, National University of Ireland, Cork, Ireland

## 4.1 Abstract

Esterases catalyze the hydrolysis of ester bonds in fatty acid esters with short-chain acyl groups. These type of enzymes have numerous industrial applications, particularly in the food, detergent and paper industries; as well as in the production of biodiesel and environmental applications for the degradation of lipid wastes and in bioremediation. Due to the widespread applications of lipolytic enzymes, there continues to be an interest in novel esterases with new properties. Marine ecosystem has long been acknowledged as a significant reservoir of microbial biodiversity and in particular of bacterial enzymes with desirable characteristics for industrial use, such as for example cold adaptation and activity in the alkaline pH range. Given that the vast majority of microorganisms from marine environments are not as yet culturable using standard laboratory conditions, we applied a functional metagenomic approach to exploit the enzymatic potential of one particular marine ecosystem, the microbiome of the deep sea sponge *Stelletta normani*. Screening of a metagenomic library from this sponge resulted in the identification of a number of lipolytic active clones. One of these encoded a highly, cold-active esterase 7N9, and the recombinant esterase was subsequently heterologously expressed in *Escherichia coli*. The esterase was classified as type IV lipolytic enzyme, belonging to the GDSAG subfamily of hormone sensitive lipases. Furthermore the recombinant 7N9 esterase was biochemically characterized and *in silico* docking studies have been performed. The enzyme is most active at alkaline pH (8.0) and displays salt tolerance over a wide range of concentrations. The docking studies supplement the biochemical characterization and confirming its activity towards short-chain fatty acids while as well highlighting the specificity towards certain inhibitors, furthermore the structural difference to a closely related mesophilic esterase is elaborated.

## 4.2 Introduction:

Metagenomic based approaches have proven extremely useful as a means of discovering microbial enzymes with entirely new biochemical properties, thereby exploiting the microbial diversity of a variety of different environmental ecosystems (Kennedy *et al.*,

2011). These approaches are typically employed to help overcome the problem of cultivating bacteria from various different environments, where typically only 0.001% to 1% of bacterial isolates can be recovered and grown under standard laboratory conditions (Bernard *et al.*, 2000). Bioprospecting for novel enzymes with interesting biotechnological applications using metagenomics based approaches particularly from extreme environments such as acidic, cold, hot and hypersaline environments has proven to be particularly successful (Mirete *et al.*, 2016). Nevertheless it is clear that relative to the number of metagenomic sampling sites that have been reported to date that up until now we have largely under sampled many of these with respect to enzyme discovery (Ferrer *et al.*, 2016). Thus a large part of the microbial biodiversity present in the earth's biosphere has yet to be explored or exploited for novel enzymes (Alcaide *et al.*, 2015b).

The impetus to explore novel environments for new industrial enzymes comes from the need to meet the ongoing global demand for these enzymes which in 2014 was estimated to have a value of around $4.2 billion, and which is expected to reach nearly $6.2 billion by 2020 (Singh *et al.*, 2016). The deep oceans are one part of the earth's biosphere which has to date received little attention. With mean depths of 3800 m and 50% of the oceans being deeper than 3000 m, the 'deep sea' constitutes not only a potential large resource from a microbial biodiversity perspective, but also a very unique environment; with temperatures ranging from 2-3$^{o}$C and a salinity of about 3.5% together with hundreds of bars of hydrostatic pressure (Wirsen and Molyneaux, 1999). Thus microbial communities which have adapted to these extremes of temperature, salinity, pressure and low levels of light are likely to possess novel biochemistry; and have enzymes that may be uniquely suited to many industrial processes (Alcaide *et al.*, 2015a). In addition seawater samples are an extremely rich source of potential biocatalytic biodiversity when one considers that with bacteria capable of achieving densities of up to $10^6$ per milliliter of seawater (Azam, 1998), and assuming that there are approximately 3000 genes in a single genome and that 40% of these genes have catalytic activity then there may be as many as $3 \times 10^9$ genes mediating up to $1{\cdot}2 \times 10^9$ putative reactions in a milliliter of seawater (Dinsdale *et al.*, 2008; Vieites *et al.*, 2009). Thus although the deep sea is likely to be a rich source of microbial biocatalytic biodiversity, very few studies have to date attempted to access or exploit this biodiversity;

most likely due to both the technical difficulties and costs associated with sampling at lower depths.

Lipolytic enzymes can be classified into eight different families and numerous subfamilies (Arpigny and Jaeger, 1999). The overall three-dimensional structure of all lipases and esterases is defined by a characteristic $\alpha/\beta$-hydrolase fold (Ollis *et al.*, 1992), with 'true lipases', members of family I; also having a characteristic lid and possessing characteristic interfacial activation properties (Arpigny and Jaeger, 1999). Furthermore lipolytic enzymes can be categorized as either lipases (triacylglycerol hydrolases, EC 3.1.1.3) or esterases (EC 3.1.1.1) corresponding to their specific hydrolytic activity, where lipases hydrolyze long-chain acyl groups to fatty acids and acylglycerols and esterases hydrolyze ester bonds of fatty acid esters with short-chain acyl groups (Verger, 1997). The industrial applications of lipolytic enzymes are wide ranging and include applications in the detergent industry, biodiesel production, food industry, pulp and paper industry, fats and oils production via transesterification, as well as environmental applications for the degradation of lipid wastes (Panda and Gowrishankar, 2005; Jegannathan and Nielsen, 2013; Sharma and Kanwar, 2014; Sasso *et al.*, 2016; Ramnath *et al.*, 2017; Rao *et al.*, 2017). Lipolytic enzymes from *Burkholderia* are for example interesting in biodiesel production, as they can be used for transesterification of waste oils with short chain alcohols in the presence of high levels of methanol (Sasso *et al.*, 2016). Furthermore lipolytic enzymes can be used for bioremediation of environmental hazards (oil spills), which is important in conjunction with the exploitation of new and remote sources of oils, especially in the cold environments (Yang *et al.*, 2009).

We have previously studied the microbial biodiversity of a number of deep sea sponges sampled at depths between 760-2900 m below sea level, and the sponge species *Stelletta normani* in particular (Kennedy *et al.*, 2014). *S. normani* appears to possess a very diverse microbial community, comparable to high microbial abundance sponges from shallow water habitats (Jackson *et al.*, 2013; Kennedy *et al.*, 2014). Furthermore the microbial community structures of deep sea sponges appear to possess a huge potential secondary metabolite biodiversity (Borchert *et al.*, 2016). With this in mind we set out to assess the biocatalytic potential of the metagenome of the deep sea sponge *S. normani* using a functional metagenomic based approach. The *S. normani* metagenomic fosmid library was

found to express a large number of lipolytic activities, from which we subsequently characterized a cold-active esterase from the hormone sensitive lipase family IV. Cold-active enzymes possess unique biochemical properties that are of particular interest for industrial biocatalysis. These include low substrate affinity, thermolability and high specific activity at low temperatures, which can together help achieve saving in energy costs and in reducing undesirable chemical side reactions, as well as allowing rapid thermal inactivation (Cavicchioli *et al.*, 2002; Santiago *et al.*, 2016). Other 'cold-active' lipolytic enzymes from family IV have previously been described, but these usually possess higher optimal temperatures (35-50°C) (Fu *et al.*, 2011; Hårdeman and Sjöling, 2007), whereas the here described esterase has a high activity at 4-40°C, identifying it at truly 'cold-active'. In addition this work also broadens the description of members of the lipolytic enzyme family IV, as thermophilic and mesophilic enzymes of this family have to date been already described (Rhee *et al.*, 2005).

## 4.3 Materials and Methods:

### 4.3.1 Sponge sampling and metagenomic library preparation

The sponge *Stelletta normani* was sampled in Irish territorial waters off the west coast of Ireland (Latitude 53.9861, Longitude -12.6100) from a depth of 760m with the help of the remotely operated vehicle (ROV) *Holland I* abroad the R.V. *Celtic Explorer* during a Biodiscovery cruise in 2013. The sponge sample was rinsed with sterile artificial seawater [3.33% (w/v) Instant Ocean, Aquarium Systems] and stored at -80°C until further processing.

The total metagenomic DNA was extracted as described in (Kennedy *et al.*, 2008). In brief, the sponge tissue was ground under liquid nitrogen using a sterile pestle and mortar. The obtained sample was suspended in lysis buffer [100 mM Tris, 100 mM EDTA, 1.5 M NaCl (w/v), 1% CTAB (w/v), 2% SDS (w/v)] in a 1:5 ratio and then incubated for 2 h at 70°C. This solution was then centrifuged until a clear solution was obtained, which was subsequently used to precipitate the dissolved metagenomic DNA with 0.7 volumes of isopropanol for 30 minutes at room temperature. The precipitation mixture was centrifuged

at 6000 x *g* for 30 minutes, followed by a washing step with 70% ethanol and the obtained DNA pellet was air dried, before resuspending in a suitable amount of Tris-EDTA buffer (10 mM Tris, 1 mM EDTA, pH 8.0).

The metagenomic DNA was size-separated using pulse field-gel-electrophoresis and the size fraction of ~40 kb was cloned into the fosmid Copy Control pCCERI (derivative of pCC1FOS™) vector for metagenomic library construction. The insert harbouring fosmids were packaged into lambda phages and used to transfect *E. coli* TransforMax™ EPI300™ cells. In total a library of approximately 14,000 clones was generated from the obtained metagenomic DNA and this library was plated onto agar Q-Tray plates containing 1% tributyrin. Positive clones were selected for sequencing on the PacBio platform (Beckman Coulter Genomics).

### 4.3.2 Fosmid sequencing, assembly and annotation

The lipase harbouring fosmid was sequenced on the PacBio platform, by Beckman Coulter Genomics; it was assembled manually from the quality filtered and preassembled reads according to overlapping regions. The assembled fosmid was annotated using the BASys online pipeline (Van Domselaar *et al.*, 2005) and all gene annotations were confirmed by BLAST searches (BLASTX) against the NCBI non redundant protein sequence database.

### 4.3.3 Cloning, expression and purification

The full length esterase was amplified using primers incorporating enzyme restriction sites and allowing in-frame cloning into the pBAD/mycHIS-A overexpression vector (Invitrogen). Forward primer f7N9 (TATATA<u>CC*ATGG*</u>CTAGTCCTGAGCTCGATACGG) incorporates an *Nco*I restriction site (underlined) at the start codon (italics) of the gene. Reverse primer r7N9 (ATATAT<u>AAGCTT</u>GCCAGTGTGCTTTTTAATGAACTCC) incorporates a *Hind*III restriction site replacing the stop codon of the esterase gene. The amplified PCR product and pBAD/mycHis-A were digested with *Nco*I and *Hind*III and subsequently an overnight

ligation was carried out at 4°C at a 10:1 ratio insert to plasmid. Two µl of the ligation reaction were added to 50 µl TOP10 chemically competent cells (ThermoScientific) and the transformation was carried out according to the manufacturer's guidelines. The transformation mixture was plated in different amounts on LB plates containing 50 µg/ml ampicillin, 0.2% arabinose and 1% tributyrin.

Pre inoculum was prepared by inoculating a loop full of culture (*E.coli* TOP10 bearing pBAD/mycHIS-A vector with esterase insert) in 3ml of LB broth supplemented with 50 µg/ml ampicillin and incubated at 37°C in a shaking incubator overnight. Next day 10 ml of LB broth was inoculated with 100 µl of the pre inoculum and 50 µg/ml of ampicillin following incubation at 37°C in a shaking incubator until it reached the mid log phase of growth. Then sterile 0.2% arabinose was added to the culture for dose dependent induction and the culture was then further incubated overnight under the same conditions. The next day protein expression was confirmed by performing SDS-PAGE.

The esterase enzyme was purified from the overnight culture using the Ni-NTA spin column obtained from Qiagen. Enzyme purification steps were followed as described in the Ni-NTA spin kit hand book (Under Native condition). 5 ml of overnight culture was used and centrifuged at 4000 x *g* for 15min in 4°C. The pellet was resuspended in 630 µl of lysis buffer (NPI-10) (50 mM $NaH_2PO_4$, 300 mM NaCl, 10 mM imidazole, pH 8.0) and 70 µl of lysozyme (10mg/ml) was added and kept on ice for 30 min. After this the lysate was centrifuged at 12000 x *g* for 30 min at 4°C and the supernatant was collected. The Ni-NTA column was equilibrated with 600 µl of NPI-10 buffer, centrifuged at 890 x *g* for 2 min at 4°C. 600 µl of the supernatant from the previous step was loaded onto the pre-equilibrated Ni-NTA spin column and then centrifuged at 270 x *g* for 5 min at 4°C, the flow through was collected. The column was washed twice with 600 µl of NPI-20 (50 mM $NaH_2PO_4$, 300 mM NaCl, 20 mM imidazole, pH 8.0) buffer by centrifuging the column at 270 x *g* for 2min at 4°C. Protein was eluted in two fractions by adding 300 µl of NPI-500 (50 mM $NaH_2PO_4$, 300 mM NaCl, 500 mM imidazole, pH 8.0) buffer twice to the column and centrifuged at 890 x g for 2 min at 4°C. The eluted fractions were then checked on SDS-PAGE.

### 4.3.4 Biochemical characterization of recombinant esterase

The pH stability of the esterase enzyme was evaluated at different pHs ranging from 5-10. Cell free supernatant solution and pH buffer were added together in a 1:1 ratio and incubated at 37°C for 1 h. The temperature stability of the esterase was analysed by keeping the cell free supernatant at different temperatures (4°C, 20°C, 25°C, 30°C, 37°C and 40°C) for 1 h. The halotolerance was assessed at sodium chloride concentration ranging from 1% to 24%. The enzyme activity was then tested colorimetrically. For this substrate solutions were prepared comprising of solutions A and B, A comprised of 40 mg p-nitrophenyl palmitate dissolved in 12 ml of isopropanol and B comprising 0.1 g of gum arabic, 0.2 g Sodium deoxycholate, 500 µl Triton X-100 dissolved in 90 ml 50 mM Tris-HCl buffer pH 8. Solutions A and B were mixed in a 1:20 ratio. For each assay 100 µl substrate solution, 50 µl Glycine-NaOH buffer and 10 µl enzyme sample were mixed and pipetted into a microtiter plate, incubated for 45 min at 37°C, and then the absorbance at 410 nm was measured and plotted against a p-nitrophenyl standard curve (Mobarak-Qamsari *et al.*, 2011).

### 4.3.5 Effect of metal ions on enzyme activity

The effect of different metal ions ($Ag^{+}$, $Cu^{2+}$, $K^{+}$, $Co^{2+}$, $Mg^{2+}$ and $Ba^{2+}$, as well as the heavy metal ions $Hg^{+}$ and $Pb^{2+}$) on the enzyme activity was tested by adding different concentrations (1, 2, 3, 4 and 5 mM) of the metal ions to the cell free supernatant following incubation for 1 h at room temperature and subsequently measuring the esterase activity colorimetrically.

### 4.3.6 Docking *in silico* analysis

The esterase sequence obtained from the PacBio whole fosmid sequencing was subjected to BLAST searches at NCBI and a query coverage of 99% of the sequence with 66% identity to the bacterial hormone sensitive lipase E40 (PDB ID: 4xvc) was obtained. The E40 crystal structure was used as a template for homology modelling using Modeller 9.10 (Webb and Sali, 2016) and five models were generated. All the models were stereochemically

optimized by Ramachandran plot and one model was selected for further docking studies (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php). Quality control of the obtained model was performed by using ERRAT (Colovos and Yeates, 1993) and VERITY-3D (Bowie *et al.*, 1991; Lüthy *et al.*, 1992) from the SAVeS 4.0 software (https://services.mbi.ucla.edu/SAVES/). The known inhibitors and substrates were docked against the esterase model *in silico* by using Ligprep and Glide (Friesner *et al.*, 2004) from the Maestro Schrödinger software package (Maestro, 2016).

#### 4.3.7 Enzyme kinetics

Different concentrations of various substrates, pNPP (p-nitrophenyl palmitate), pNPM (p-nitrophenyl myristate), pNPL (p-nitrophenyl laurate), pNPC (p-nitrophenyl caprate), pNPB (p-nitrophenyl butyrate) and pNPA (p-nitrophenyl acetate) from 0.1 mM to 2.0 mM were added to the column purified enzyme sample**.** Based on these values from microplate-readings at 410 nm, $V_{max}$ and $K_m$ values were calculated and Michaelis-Menten plots were generated.

### 4.4. Results

#### 4.4.1 Metagenomic library construction and screening for esterase clones

A metagenomic library was constructed from the marine sponge *Stelletta normani*. The sponge had been collected by an ROV from a depth of 760m. Metagenomic DNA was extracted and size selected for ~40 kb DNA fragments following pulse-field and subsequently concentrated using an Amicon centrifugal concentrator. The library which was constructed using the fosmid vector pCCERI (Selvin *et al.*, 2012) contained approximately 14,000 clones which were screened for lipase activity (Figure 1A). High throughput plate screening using 1% tributyrin resulted in the initial identification of 31 positive clones (data not shown). From amongst the 20

most highly active clones, the 7N9 fosmid was chosen as it displayed the highest level of activity and it was subsequently sequenced using the PacBio system.

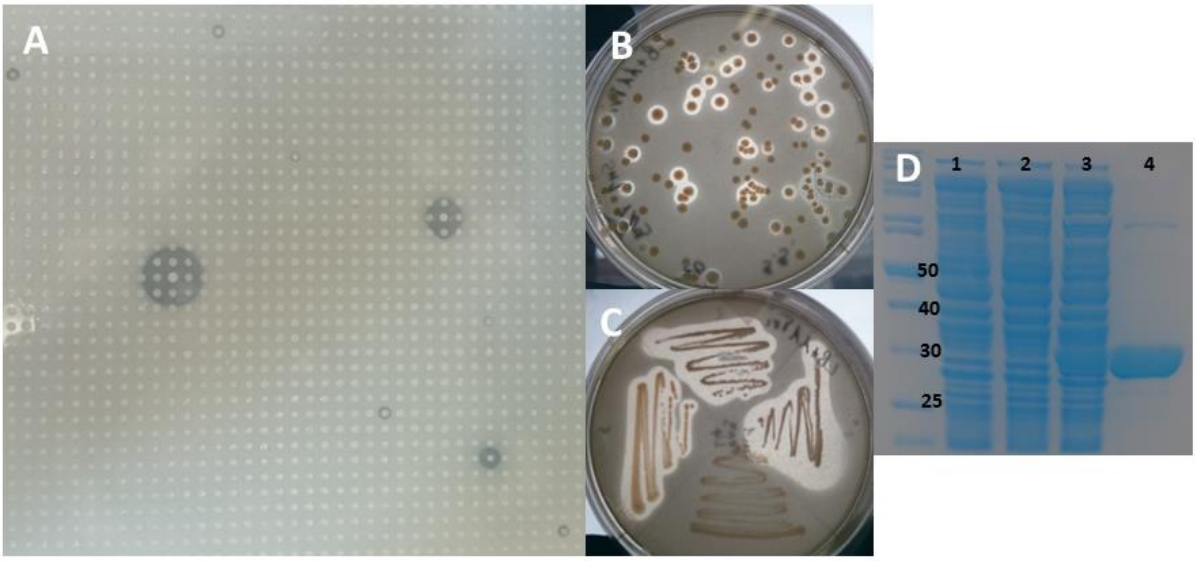



**Figure 1:** Metagenomic library, cloning and purification of 7N9 esterase. A) Metagenomic library of *Stelletta normani* plated onto 1% tributyrin agar, B) Lipase activity of cloned *Escherichia coli* clones containing 7N9 esterase harboring pBad plasmid C) Restreak of active clones. D) SDS-PAGE analysis of the expression and purification of 7N9 esterase, first lane marker, lane 1 induced (0.02% arabinose) *E. coli* culture with empty pBAD expression vector, lane 2 uninduced *E. coli* culture with pBAD harbouring 7N9, lane 3 induced (0.02% araboinose) *E. coli* with pBAD harbouring 7N9, lane 4 partial purification using Ni-NTA resin of 7N9 esterase.

### 4.4.2 Fosmid sequencing and esterase identification

The sequenced fosmid comprised of 41,407 bp and contained 65 coding DNA sequences of which 31 were annotated by BASys (Table 1). A contig (contig 30,107 to 30,997, bah, Table 1) was identified as containing an ORF encoding a gene with putative esterase function (Figure 2). The putative esterase ORF, named 7N9, was found to comprise 296 amino acids, with a GC content of 60.5% and was annotated as an acetyl-hydrolase. BLASTX comparison subsequently classified the esterase as part of the alpha/beta hydrolase family. The enzyme showed highest homology (66%) to an esterase of the bacterial hormone sensitive lipase family (E40), which was itself isolated from marine sediment (Li *et al.*, 2015a). The E40 esterase is part of the GDSAG motif subfamily within the lipase family IV, phylogenetic

comparisons (Figure 3) and multiple sequence alignments (Figure 4) indicate that the 7N9 esterase is part of the same subfamily of lipolytic enzymes as it also contains the characteristic GDSAG motif (hormone sensitive family, Hsl). Furthermore the esterase also contains the highly conserved His-Gly-Gly (HGG) motif, which together with the GDSAG motif involved in the oxyanion hole formation (Mohamed *et al.*, 2013; Ramnath *et al.*, 2017).

**Table 1:** Record of the annotated genes on the fosmid.

| Start | End | Gene | COG | Protein Function |
|---|---|---|---|---|
| 2756 | 1041 | mfd | COG1197 | Transcription-repair-coupling factor |
| 3147 | 2800 | mfd | COG1197 | Transcription-repair-coupling factor |
| 4933 | 5304 | proB | COG0263 | Glutamate 5-kinase |
| 5291 | 5863 | proB | COG0263 | Glutamate 5-kinase |
| 5752 | 6057 | proB | COG0263 | Glutamate 5-kinase |
| 6180 | 6515 | proA | COG0014 | Gamma-glutamyl phosphate reductase |
| 6718 | 7389 | proA | COG0014 | Gamma-glutamyl phosphate reductase |
| 8918 | 8535 | dxs | COG3958 | Putative transketolase C-terminal section |
| 8532 | 7969 | dxs | COG3958 | Putative transketolase C-terminal section |
| 9412 | 8918 | tktB | COG3959 | Putative transketolase N-terminal section |
| 9849 | 9409 | tktA | COG3959 | Putative transketolase N-terminal section |
| 12922 | 13854 | ydcC | - | Uncharacterized protein in dhlA 3'region |
| 16056 | 16493 | repE | - | Replication initiation protein |
| 17273 | 18265 | sopA | COG1192 | Protein sopA |
| 18118 | 18447 | sopA | - | Protein sopA |
| 18447 | 19196 | sopB | COG1475 | Protein sopB |
| 24901 | 24095 | aacC4 | COG2746 | Aminoglycoside N(3')-acetyltransferase IV |
| 25302 | 25640 | traJ | - | Protein traJ |
| 26297 | 25602 | cat | - | Chloramphenicol acetyltransferase |
| 26313 | 26903 | resD | - | Resolvase |
| 27262 | 26915 | betA | COG2303 | Choline dehydrogenase |
| 27470 | 27988 | baiF | COG1804 | Bile acid-CoA hydrolase |
| 27989 | 28825 | baiF | COG1804 | Bile acid-CoA hydrolase |
| 28911 | 29732 | mutM | COG0266 | Formamidopyrimidine-DNA glycosylase |
| 30107 | 30997 | bah | COG0657 | Acetyl-hydrolase |

| | | | | |
|---|---|---|---|---|
| 32466 | 32627 | mutB | COG1884 | Methylmalonyl-CoA mutase large subunit |
| 32678 | 33205 | mutB | COG1884 | Methylmalonyl-CoA mutase large subunit |
| 34009 | 34956 | acoA | COG1071 | Acetoin:2,6-dichlorophenolindophenol oxidoreductase subunit alpha |
| 35222 | 35545 | pdhB | COG0022 | Pyruvate dehydrogenase E1 component subunit beta |
| 35536 | 36117 | acoB | COG0022 | Acetoin:2,6-dichlorophenolindophenol oxidoreductase subunit beta |
| 37965 | 37342 | ysgA | COG0412 | Putative carboxymethylenebutenolidase |

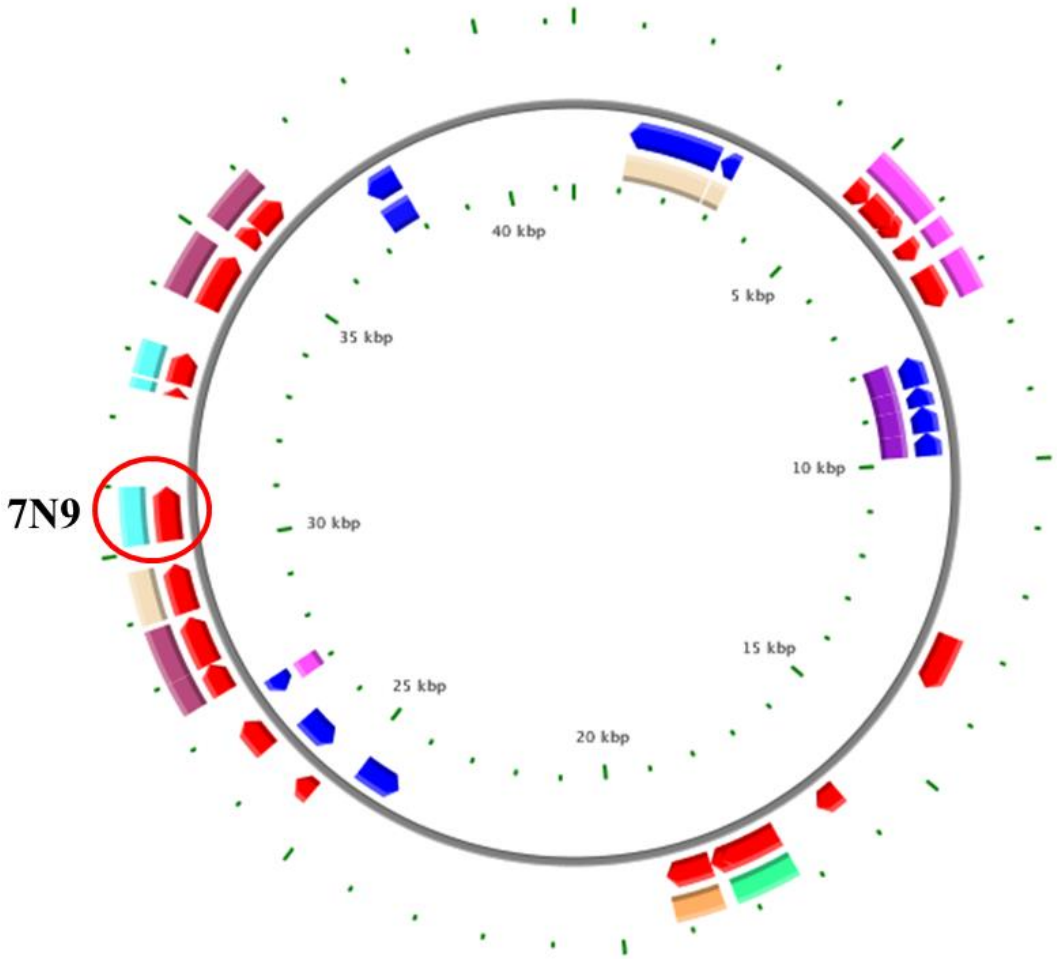



**Figure 2:** Annotated map of the sequenced fosmid bearing the cold active esterase (map generated by BASys (Van Domselaar et al., 2005)). The fosmid backbone starts at 16,5 kbp and ends at approximately 26,5 kbp. The esterase encoding gene is encircled in red.

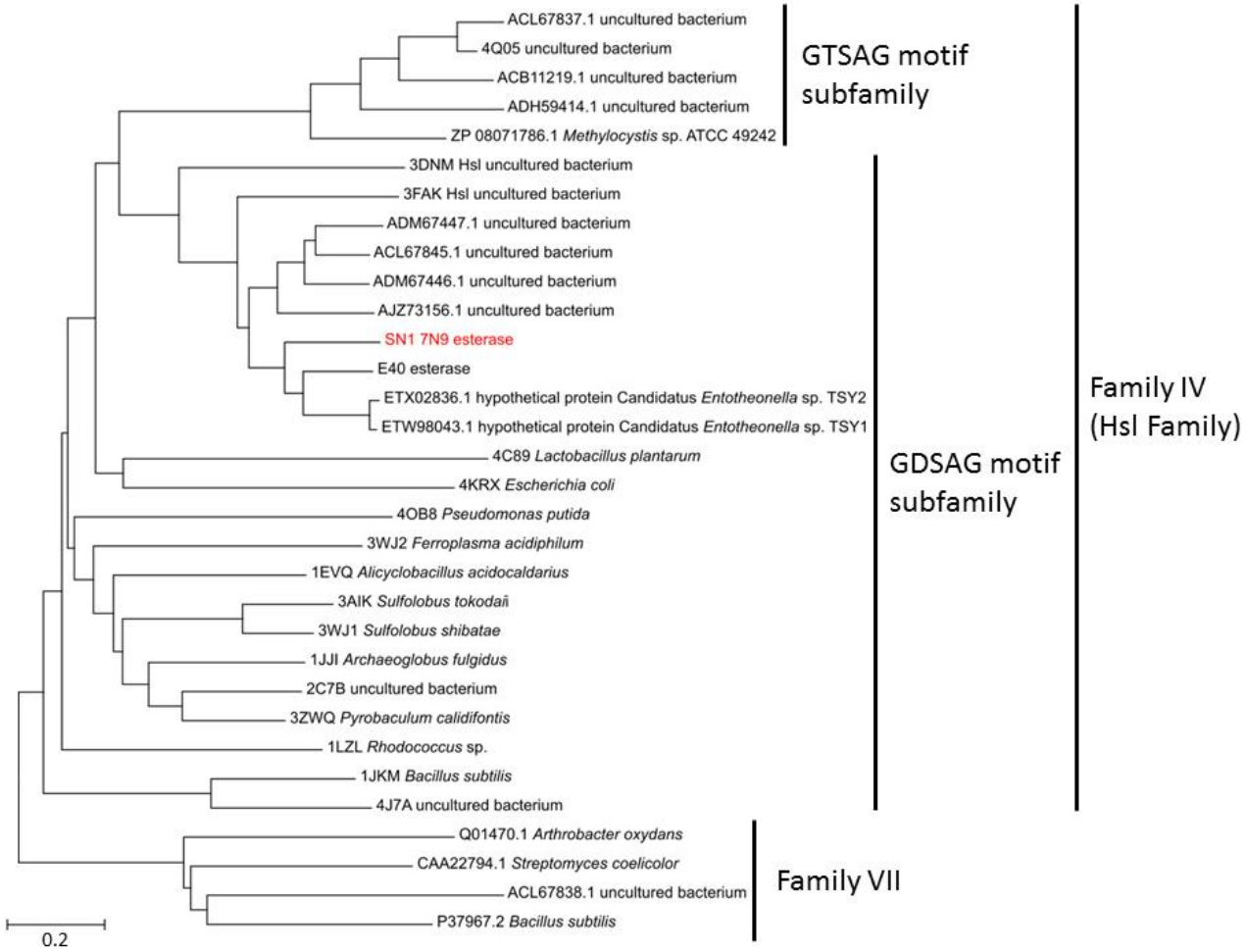



**Figure 3:** Phylogenetic comparison of the cold active esterase and other representative sequences of different lipase families. The phylogenetic tree was built by the neighbor joining method and bootstrap analysis with 500 replicates was conducted, reference sequences from lipase family VII are used as outgroups.

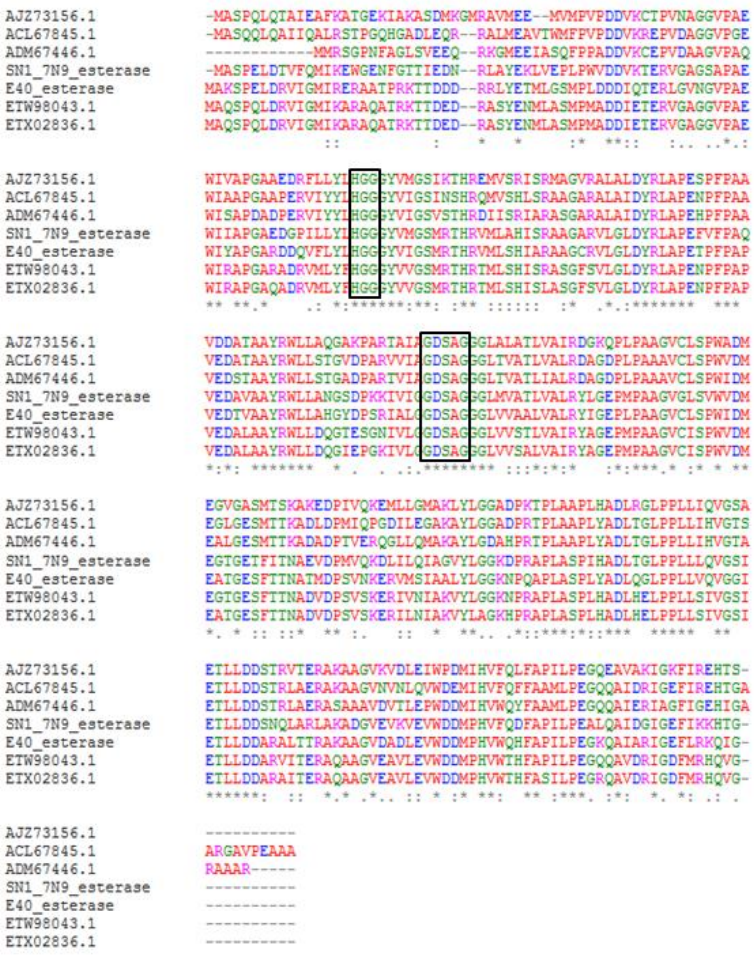



**Figure 4:** Multiple sequence alignment of most closely related esterase sequences. The conserved GDSAG (GXSXG) and HGG motifs are shown in the black boxes (Alignment was produced with Clustal Omega (Li *et al.*, 2015b; McWilliam *et al.*, 2013; Sievers *et al.*, 2011) and MEGA6 (Tamura *et al.*, 2013)). Red colored are small hydrophobic and aromatic amino acids, blue are acidic amino acids, magenta are basic amino acids and green are hydroxyl, sulfhydryl, amine amino acids and glycine. (*) indicate a fully conserved residue, (:) indicate a group of strongly similar residues and (.) indicates conservation of a group of weakly similar residues.

### 4.4.3 Cloning, expression and purification of recombinant 7N9 esterase.

The 7N9 esterase gene was PCR amplified, cloned into the pBAD/mycHIS-A vector, transformed into TOP10 *E. coli* cells and transformants were tested for esterase activity on 1% tributyrin plates (Figure 1B, 1C). Purification of the protein was performed by using the His-tag and a Ni-NTA resin column approach, with the tag being fused to the protein while transforming it into to the expression vector. The ORF encoding the esterase resulted in a

protein of a calculated mass of 31.7 kDa with a theoretical pI of 4.59 and approximately 34 kDa including the fused His-tag and myc-epitope.

### 4.4.4 Docking studies of different substrates and inhibitors

The model of the E40 esterase was used as a template to generate a 3D model of esterase 7N9 and stereo chemical optimization was performed using Ramachandran plotting. When comparing to the template (E40; Pdb id: 4xvc) there is a slight variation in our models CAP and catalytic domain. The template (Pdb id: 4xvc) contains a CAP domain at Met1–Ile45 and a catalytic domain at Gln46–Gly297. Residues Gly76 and Gly77 within the conserved HGG motif comprise the oxyanion hole that is involved in substrate binding for HSL esterases. The catalytic triad composed of residues Ser145, Glu239, and His269 is below the oxyanion hole. In contrast in our model the CAP domain is located at Met1–Lys45 and the catalytic domain at Thr46–Gly296 and the catalytic triad is composing of the residues Ser144, Glu238, and His268 located below the oxyanion hole (Figure 5). The model was subsequently used to dock different substrates and inhibitors (Supplementary file 1.1). Docking scores indicate a high specificity for the substrate pNPA (Figure 6) and the inhibitor Phenylmethansulfonic acid (Table 2). Phenylmethansulfonic acid is also able to covalently bind to the nucleophilic Ser145 of E40. In the supplementary file 1.2 and 1.3 3D binding models of the esterase and the different substrates and inhibitors can be found.

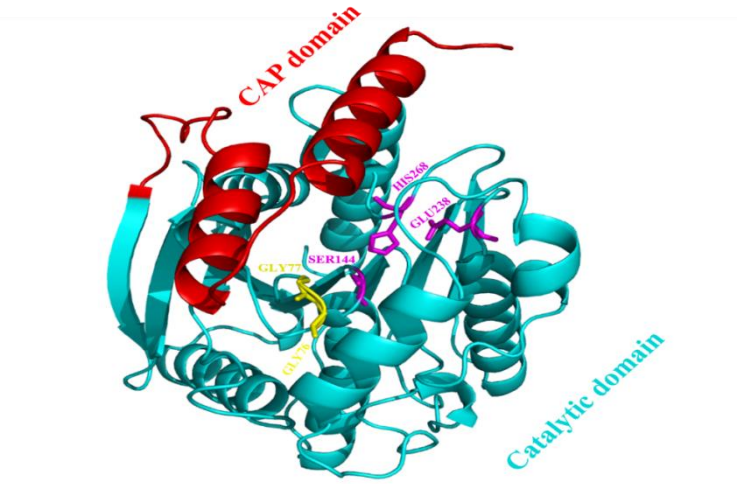



**Figure 5:** Domain architecture of the 7N9 esterase. Highlighted in red is the CAP domain, magenta colored is the catalytic triad Ser144, Glu238 and His268, yellow is the oxyanion hole comprising of residues Gly76 and Gly77 and light blue is the catalytic domain.

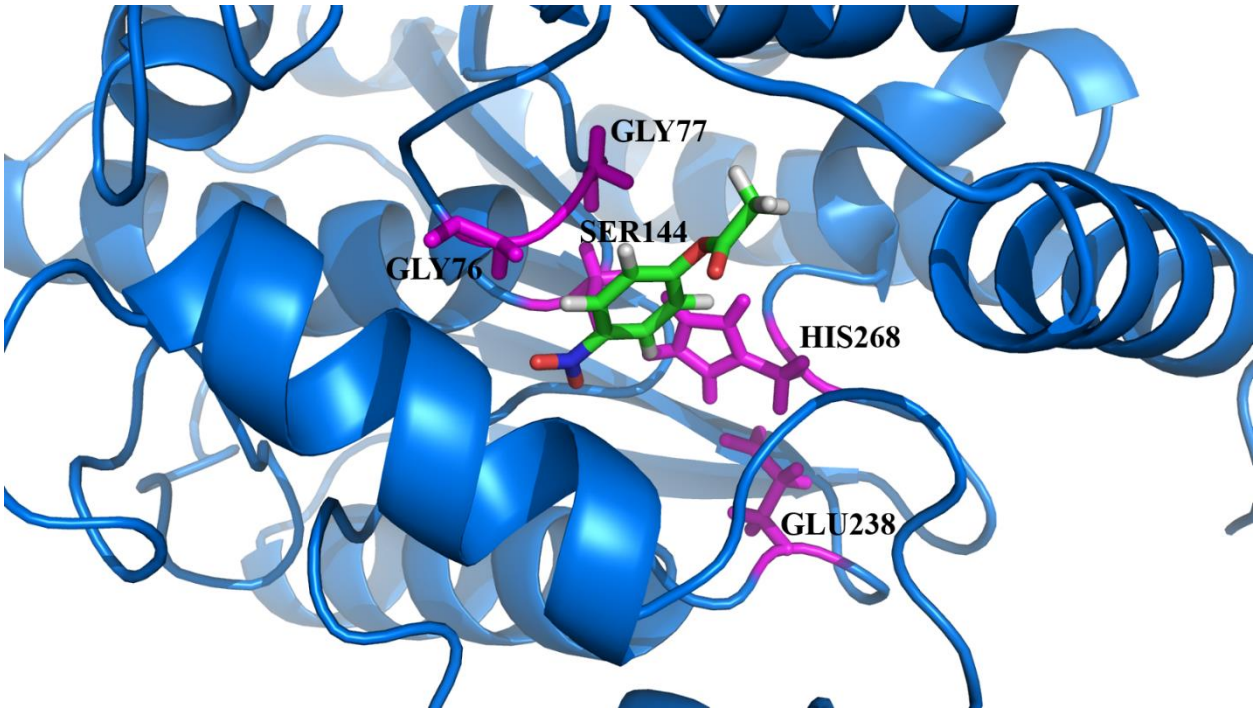



**Figure 6:** 3D docking Model of the most preferred substrate, 4-nitrophenol acetate. The catalytic site residues of 7N9 are highlighted in magenta and the substrate is placed in the centre.

**Table 2:** Docking scores of different substrates and inhibitors with the esterase model.

| Name | Docking Score | Name | Docking Score |
|---|---|---|---|
| 4-methylumbelliferone | -6.092 | Phenylmethanesulfonic acid | -5.998 |
| 4-Nitrophenyl acetate | -5.961 | 5-Carbamoyl-2H-1,2,3-triazole-4-diazonium | -4.676 |
| Tributyrin | -5.379 | Isoxazole | -4.380 |
| 4-Nitrophenyl phosphate | -4.873 | Oleic acid | -1.926 |
| Triacetin | -4.361 | Triacsin C | -1.664 |
| methyl laurate | 0.253 | | |

### 4.4.5 Biochemical characterization of the recombinant esterase 7N9

#### 4.4.5.1 Substrate specificity

The 7N9 esterase was found to have a higher specificity towards short chain fatty acids (Table 3). Fatty acid substrates ranging in carbon chain length from 16 (pNPP) to two (pNPA) carbon atoms were assessed; with a $V_{max}$ for pNPP being 1.507 mM/ml/min and $K_m$ 0.6275 mM, while pNPA had a $V_{max}$ of 2.731 mM/ml/min and a $K_m$ 0.1674 mM.

**Table 3:** Substrate specificity of the esterase

| Substrate | $V_{max}$ [mM/ml/min] | $K_m$ [mM] |
|---|---|---|
| pNPP | 1.507 | 0.6275 |
| pNPM | 1.515 | 0.4768 |
| pNPL | 1.596 | 0.4229 |
| pNPC | 2.653 | 0.2506 |
| pNPB | 2.722 | 0.1992 |
| pNPA | 2.731 | 0.1674 |

#### 4.4.5.2 Temperature dependency

The activity of the enzyme was assessed at different temperatures ranging from 4°C to 60°C (Figure 7). The enzyme displayed the highest activity at 4°C and 20°C with activity declining thereafter and no activity at 60°C, identifying it as a cold-adapted type of hormone sensitive esterase.

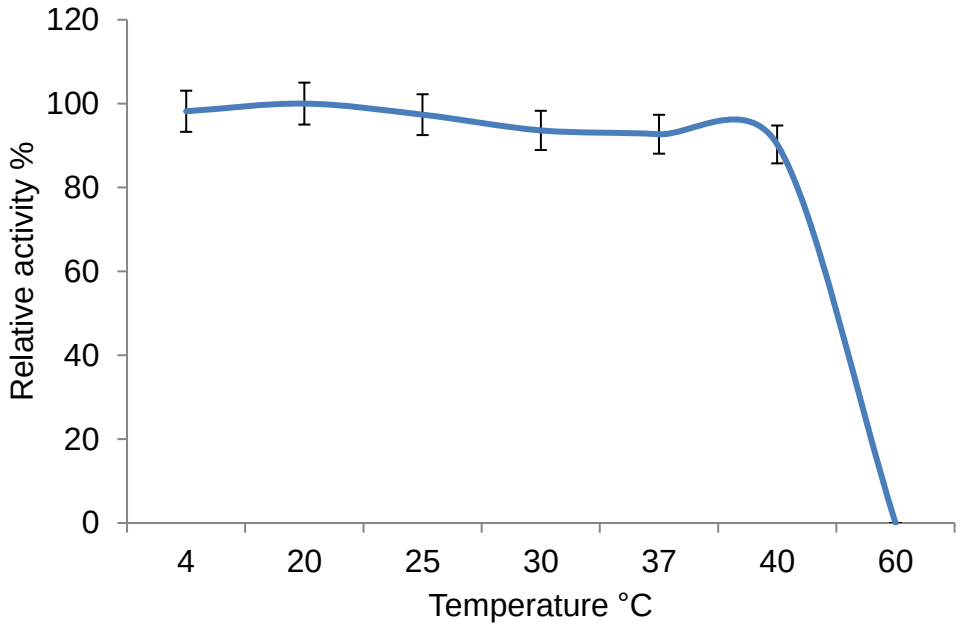



**Figure 7:** Temperature dependency of the esterase 7N9. The temperature dependency of the esterase 7N9 activity was tested at 4, 20, 25, 30, 37, 40 and 60°C. The values displayed are the means and standard deviations of triplicate measurements.

#### 4.4.5.3 pH dependency

The pH dependency of the esterase was tested at different pHs ranging from 5 to 10. Optimal activity was achieved at pH 8.0, higher and lower pHs lead to a decline in activity, nonetheless activity is seen for all pH values investigated (Figure 8). Interestingly the optimal observed pH is in line with normal pH conditions encountered in seawater, where the pH ranges from 7.5 to 8.4 (Chester R and Jickells TD, 2012).

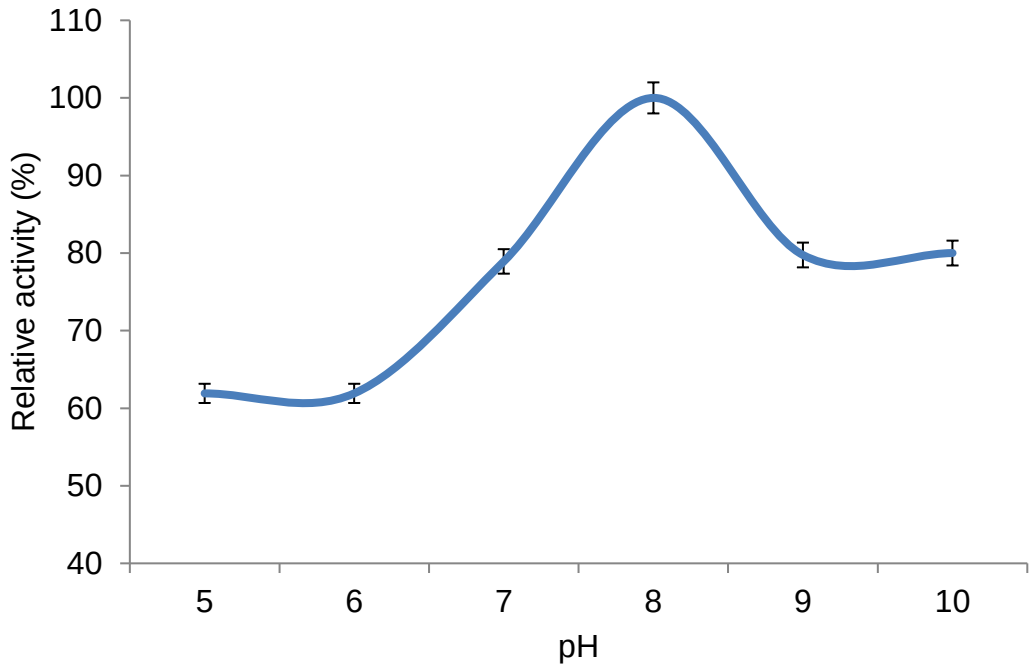



**Figure 8:** pH dependency of esterase activity. The effect of different pH levels on the activity of the esterase 7N9 was tested, investigated were the pH values 5, 6, 7, 8, 9 and 10. The values displayed are the means and standard deviations of triplicate measurements.

#### 4.4.5.4 Effect of metal ions on enzyme activity

The effect of different concentrations (1-5 mM) of various metal ions ($Ag^{+}$, $Cu^{2+}$, $K^{+}$, $Co^{2+}$, $Mg^{2+}$ and $Ba^{2+}$, as well as the heavy metal ions $Hg^{+}$ and $Pb^{2+}$) on enzyme activity was tested. Increasing activity was observed with increasing concentrations of $Cu^{2+}$, $Ag^{+}$ and $Ba^{2+}$; while a decrease in activity was observed for $K^{+}$, $Mg^{2+}$, $Co^{2+}$ and the heavy metals $Hg^{2+}$ and $Pb^{2+}$ (Figure 9). The increase in $Pb^{2+}$ concentration having the most detrimental effect on esterase activity, with only residual activity remaining at elevated levels of this heavy metal ion.

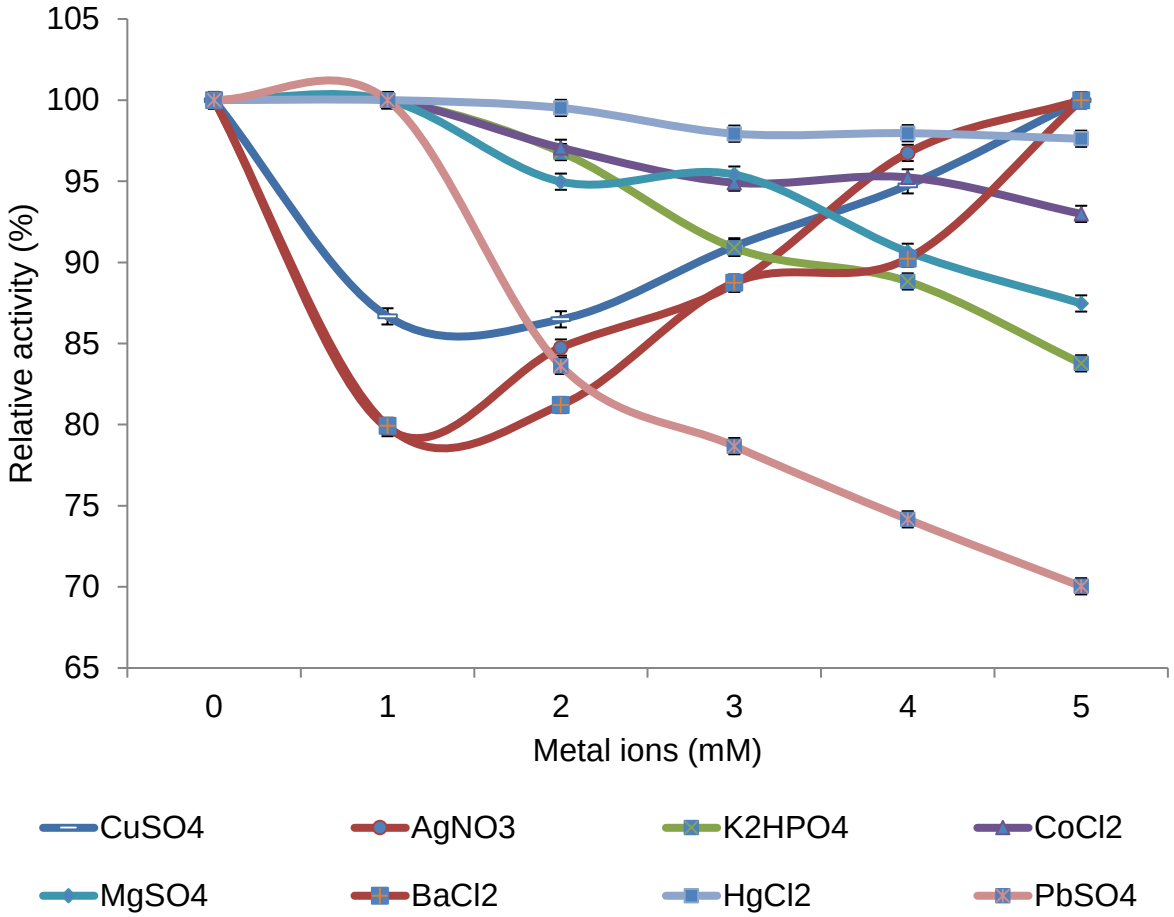



**Figure 9:** Effect of different concentrations of different metal ions onto the activity of the esterase. The inhibitory and beneficial effects of various metal ($Ag^{+}$, $Cu^{2+}$, $K^{+}$, $Co^{2+}$, $Mg^{2+}$ and $Ba^{2+}$) and heavy metal ions ($Hg^{+}$ and $Pb^{2+}$) at concentrations ranging from 1 to 5 mM on the esterase activity were tested. The values displayed are the means and standard deviations of triplicate measurements.

#### 4.4.5.5 Halotolerance

The halotolerance of the esterase activity in 7N9 was then investigated, by measuring activity at different percentages of sodium chloride, ranging from 1% to 24% (Figure 10). Good levels of activity were observed over the range of sodium chloride concentrations up to 16%, with still 87% of relative activity at that concentration and a more rapid decline thereafter. The overall salt concentration of sea water is typically around 3.5% (Chester R and Jickells TD, 2012) and therefore falls within the range of optimal activity of the enzyme.

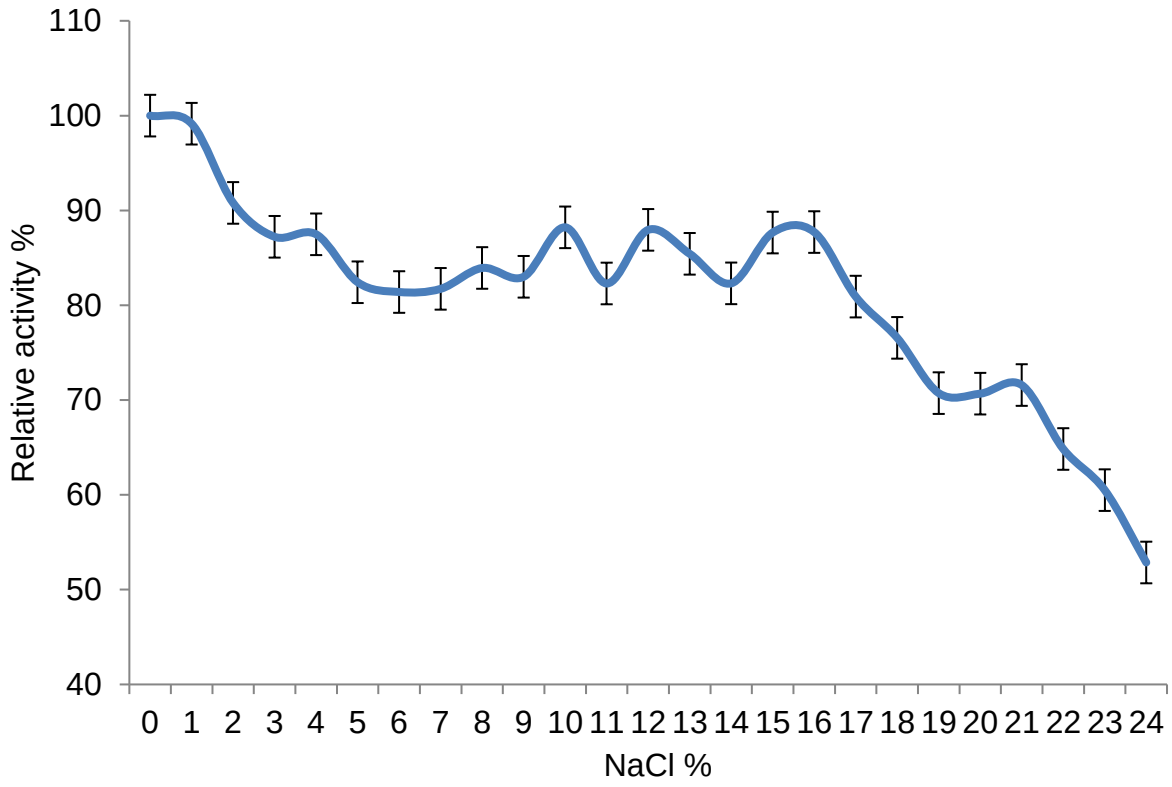



**Figure 10:** Halotolerance of the esterase activity**.** Sodium chloride concentrations from 1% to 24% were tested for their effect on the activity of the esterase 7N9. The values displayed are the means and standard deviations of triplicate measurements.

### 4.5 Discussion

The ever increasing demand for novel biocatalysts has resulted in the development of a range of different approaches to explore and exploit the genetic resources in various environmental ecosystems. One approach which has been successfully employed to this end is metagenomics which helps facilitate access to genetic resources from uncultured microorganism (Kennedy *et al.*, 2011; Baweja *et al.*, 2016; Parages *et al.*, 2016). Marine environments in particular are proving particularly interesting as a source for novel microbial biodiversity, with numerous examples of metagenomics based approaches being employed to identify novel biocatalysts with potential biotechnological applications (Kodzius and Gojobori, 2015; Popovic *et al.*, 2015). In this study a gene encoding a novel psychrophilic esterase (7N9) from the hormone sensitive lipase (Hsl) family IV was identified following the functional screening of a deep sea sponge *Stelletta normani*

metagenomic library and the recombinant enzyme has been biochemically characterized. Functional screening of marine sponge and sediment metagenomics libraries have resulted in the discovery of a variety of novel lipases including the recent reports of a cold-active salt tolerant esterase from artic sediment (De Santi *et al.*, 2016); and a high organic solvent tolerant and thermostable esterase from marine mud (Gao *et al.*, 2016). The 7N9 esterase was identified as the most active fosmid clone of 20 lipase active clones following the initial screening of approximately 14,000 clones, from the metagenomics library.

Following sequencing of the 7N9 harbouring fosmid (Figure 2) the esterase was heterologously expressed in *Escherichia coli* and the recombinant 7N9 protein was subsequently biochemically characterized. The esterase was found to be closely related to the E40 esterase (66% amino acid homology), which was itself isolated via a functional metagenomic approach from marine sediment retrieved from a depth of 154 m in the South China Sea (Li *et al.*, 2012; Li *et al.*, 2015a). 7N9 and E40 both possess the two highly conserved GDSAG and HGG motifs which group them into the correspondent subfamily of lipase family IV (Mohamed *et al.*, 2013; Ramnath *et al.*, 2017) (Figure 4). In contrast however the 7N9 esterase has a much lower optimal temperature (20°C) than the E40 esterase (45°C) and is therefore the first truly cold-adapted esterase in this lipase subfamily. As both enzymes were retrieved from metagenomic libraries one cannot say with certainty from what type of microorganism these esterases may have been isolated, but phylogenetic comparison and protein homology suggest a close relatedness to hypothetical proteins from the marine symbiont Candidatus genus *Entotheonella* (Figure 3 and 4). Interestingly a novel carboxylesterase Est06, isolated from a forest soil metagenomics library has also recently been reported to share 61% similarity with a hypothetical protein from Candidatus *Entotheonella* sp. TSY1 (Dukunde *et al.*, 2017). This talented bacterium is known to produce the majority of all known secondary metabolites found in the sponge *Theonella swinhoei* (Wilson *et al.*, 2014).

The 3D model of the 7N9 esterase was calculated using the 3D crystal structure of the closely related E40 esterase (Li *et al.*, 2015a) as template and subsequently *in silico* docking studies with different substrates and inhibitors were performed. Esterase 7N9 was found to have subtle differences in its CAP and catalytic domain when compared to E40 esterase.

This structural differences may contribute towards the different substrate specificities and the different temperature activity profiles which we have observed. Our 7N9 esterase was most catalytically active with pNPA (p-nitrophenylacetate) as a substrate, whereas E40 was found to be more active on pNPB (p-nitrophenylbutyrate). The *in silico* docking studies confirmed the high specificity towards short chain fatty acids (Figure 6), as well as towards the inhibitor Phenylmethansulfonic acid, which is most likely able to bind covalently to a serine residue (Table 2).

The recombinant 7N9 esterase was biochemically characterized with respect to its temperature and pH activity profiles, together with its halotolerance and the effect of metal ions on activity was also assessed. The enzyme can be classified as cold-active and slightly alkaliphilic, as highest activity was observed in the range of 4°C to 20°C and at pH 8.0 (Figure 7, 8). Metal ions were found to have a marked effect on the activity of the enzyme (Figure 9), with for example increases in the heavy metal ion $Pb^{2+}$ concentration from 1 to 5 mM resulted in a decrease in enzyme activity of almost 30%. In contrast increasing the $Ba^{2+}$ concentration from 1 to 5 mM resulted in a 25% increase in enzyme activity. In total the metal ions $Cu^{2+}$, $Ag^{+}$ and $Ba^{2+}$ were found to have a positive effect on enzyme activity at concentrations ranging from 1 to 5 mM while the addition of $Hg^{2+}$, $Pb^{2+}$, $Mg^{2+}$, $K^{+}$ and $Co^{2+}$ had detrimental effects (Figure 9).

Metal ions are known to have an effect on enzyme activity (Colak *et al.*, 2005) and therefore must be taken in account when enzymes for certain tasks are needed. Environmental increases in metal ions like $Cu^{2+}$ and $Pb^{2+}$ are known to be associated with oil spills (Moreno *et al.*, 2011) and are therefore of interest to the here investigated esterase to evaluate its use in potential oil removal applications. On the one hand $Cu^{2+}$ ions increase the enzyme activity, but $Pb^{2+}$ is detrimental on the other hand. Furthermore, oil spills in cold environments are becoming more abundant due to the increased industrial exploitation of these environment; thus specialized bioremediation strategies will be required to treat these spills in the future (Yang *et al.*, 2009). In addition in respect to oil spills and oil, saline industry wastewaters a certain halotolerance is also beneficial, as those wastewaters can contain up to 14% (*w*/*v*) sodium chloride (Margesin and Schinner, 2001). The halotolerance of the enzyme was assessed in the range of 0% to 24%, the enzyme losses in the range from

1% to 16% sodium chloride concentration only 13% of its activity towards higher salt concentrations (Figure 10) and therefore coupled with its cold activity and metal ion presence responsiveness it is potentially well suited for bioremediation processes in cold environments.

Thus in conclusion a metagenomic fosmid library from the deep sea sponge *Stelletta normani* was successfully functionally screened for novel lipolytic enzymes. We describe here a novel truly cold active esterase of the GDSAG subfamily of the hormone sensitive lipase family IV. The gene encoding the lipolytic function was identified by sequencing the harboring fosmid and successfully cloned into an overexpression vector and is heterologously expressed in *Escherichia coli*. The recombinant esterase is most active against short chain fatty acid like p-nitrophenylacetate. It displays close structural relatedness to a previously described esterase (E40) isolated from a marine sediment sample, despite its different physicochemical properties. Optimal enzyme activity is achieved at low temperatures (4°C to 20°C), at an alkaline pH (pH 8.0) and salt concentrations only have a minor influence on activity levels, resembling native physiological conditions of the environment from which the initial deep sea metagenomic sample was retrieved.

* Stephen A. Jackson and Alan D.W. Dobson conceived and designed the study; Erik Borchert performed the metagenomic experiments, the fosmid sequencing, subcloned the esterase and analyzed the data. Joseph Selvin and Seghal G. Kiran performed the biochemical characterization and the docking studies.

## 4.6. Bibliography


Alcaide, M., P. J. Stogios, Á. Lafraya, A. Tchigvintsev, R. Flick, R. Bargiela, T. N. Chernikova, O. N. Reva, T. Hai, C. C. Leggewie, N. Katzke, V. La Cono, R. Matesanz, M. Jebbar, K. E. Jaeger, M. M. Yakimov, A. F. Yakunin, P. N. Golyshin, O. V. Golyshina, A. Savchenko, M. Ferrer, and M. Consortium, 2015a, Pressure adaptation is linked to thermal adaptation in salt-saturated marine habitats: Environ Microbiol, v. 17, p. 332-45.

Alcaide, M., A. Tchigvintsev, M. Martínez-Martínez, A. Popovic, O. N. Reva, Á. Lafraya, R. Bargiela, T. Y. Nechitaylo, R. Matesanz, M. A. Cambon-Bonavita, M. Jebbar, M. M. Yakimov, A. Savchenko, O. V. Golyshina, A. F. Yakunin, P. N. Golyshin, M. Ferrer, and M. Consortium, 2015b, Identification and characterization of carboxyl esterases of gill chamber-associated microbiota in the deep-sea shrimp *Rimicaris exoculata* by using functional metagenomics: Appl Environ Microbiol, v. 81, p. 2125-36.

Arpigny, J. L., and K. E. Jaeger, 1999, Bacterial lipolytic enzymes: classification and properties: Biochem J, v. 343 Pt 1, p. 177-83.

Azam, F., 1998, Microbial control of oceanic carbon flux: the plot thickens: Science, v. 280, p. 694-696.

Baweja, M., L. Nain, Y. Kawarabayasi, and P. Shukla, 2016, Current technological improvements in enzymes toward their biotechnological applications: Front Microbiol, v. 7, p. 965.

Bernard, L., H. Schaefer, F. Joux, C. Courties, G. Muyzer, and P. Lebaron, 2000, Genetic diversity of total, active and culturable marine bacteria in coastal seawater.: Aquatic Microbial Ecology, v. 23, p. 1-11.

Borchert, E., S. A. Jackson, F. O'Gara, and A. D. Dobson, 2016, Diversity of natural product biosynthetic genes in the microbiome of the deep sea sponges *Inflatella pellicula*, *Poecillastra compressa*, and *Stelletta normani*: Front Microbiol, v. 7, p. 1027.

Bowie, J. U., R. Lüthy, and D. Eisenberg, 1991, A method to identify protein sequences that fold into a known three-dimensional structure: Science, v. 253, p. 164-70.

Cavicchioli, R., K. S. Siddiqui, D. Andrews, and K. R. Sowers, 2002, Low-temperature extremophiles and their applications: Curr Opin Biotechnol, v. 13, p. 253-61.

Chester R, and Jickells TD, 2012, M*arine Geochemistry* Marine Geochemistry: Blackwell Publishing, John Wiley & Sons, 411 p.

Colak, A., D. Sişik, N. Saglam, S. Güner, S. Canakçi, and A. O. Beldüz, 2005, Characterization of a thermoalkalophilic esterase from a novel thermophilic bacterium, *Anoxybacillus gonensis* G2: Bioresour Technol, v. 96, p. 625-31.

Colovos, C., and T. O. Yeates, 1993, Verification of protein structures: patterns of nonbonded atomic interactions: Protein Sci, v. 2, p. 1511-9.

De Santi, C., B. Altermark, M. M. Pierechod, L. Ambrosino, D. de Pascale, and N. P. Willassen, 2016, Characterization of a cold-active and salt tolerant esterase identified by functional screening of Arctic metagenomic libraries: BMC Biochem, v. 17, p. 1.

Dinsdale, E. A., R. A. Edwards, D. Hall, F. Angly, M. Breitbart, J. M. Brulc, M. Furlan, C. Desnues, M. Haynes, L. Li, L. McDaniel, M. A. Moran, K. E. Nelson, C. Nilsson, R. Olson, J. Paul, B. R. Brito, Y. Ruan, B. K. Swan, R. Stevens, D. L. Valentine, R. V. Thurber, L. Wegley, B. A. White, and F. Rohwer, 2008, Functional metagenomic profiling of nine biomes: Nature, v. 452, p. 629-32.

Dukunde, A., D. Schneider, M. Lu, S. Brady, and R. Daniel, 2017, A novel, versatile family IV carboxylesterase exhibits high stability and activity in a broad pH spectrum: Biotechnol Lett, v. 39, p. 577-587.

Ferrer, M., M. Martínez-Martínez, R. Bargiela, W. R. Streit, O. V. Golyshina, and P. N. Golyshin, 2016, Estimating the success of enzyme bioprospecting through metagenomics: current status and future trends: Microb Biotechnol, v. 9, p. 22-34.

Friesner, R. A., J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, and P. S. Shenkin, 2004, Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy: J Med Chem, v. 47, p. 1739-49.

Fu, C., Y. Hu, F. Xie, H. Guo, E. J. Ashforth, S. W. Polyak, B. Zhu, and L. Zhang, 2011, Molecular cloning and characterization of a new cold-active esterase from a deep-sea metagenomic library: Appl Microbiol Biotechnol, v. 90, p. 961-70.

Gao, W., K. Wu, L. Chen, H. Fan, Z. Zhao, B. Gao, H. Wang, and D. Wei, 2016, A novel esterase from a marine mud metagenomic library for biocatalytic synthesis of short-chain flavor esters: Microb Cell Fact, v. 15, p. 41.

Hårdeman, F., and S. Sjöling, 2007, Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment: FEMS Microbiol Ecol, v. 59, p. 524-34.

Jackson, S. A., B. Flemer, A. McCann, J. Kennedy, J. P. Morrissey, F. O'Gara, and A. D. Dobson, 2013, Archaea appear to dominate the microbiome of *Inflatella pellicula* deep sea sponges: PLoS One, v. 8, p. e84438.

Jegannathan, K. R., and P. H. Nielsen, 2013, Environmental assessment of enzyme use in industrial production - a literature review: Journal of Cleaner Production, v. 42, p. 228-240.

Kennedy, J., C. E. Codling, B. V. Jones, A. D. Dobson, and J. R. Marchesi, 2008, Diversity of microbes associated with the marine sponge, *Haliclona simulans*, isolated from Irish waters and identification of polyketide synthase genes from the sponge metagenome: Environ Microbiol, v. 10, p. 1888-902.

Kennedy, J., B. Flemer, S. A. Jackson, J. P. Morrissey, F. O'Gara, and A. D. Dobson, 2014, Evidence of a putative deep sea specific microbiome in marine sponges: PLoS One, v. 9, p. e91092.

Kennedy, J., N. D. O'Leary, G. S. Kiran, J. P. Morrissey, F. O'Gara, J. Selvin, and A. D. Dobson, 2011, Functional metagenomic strategies for the discovery of novel enzymes and biosurfactants with biotechnological applications from marine ecosystems: J Appl Microbiol, v. 111, p. 787-99.

Kodzius, R., and T. Gojobori, 2015, Marine metagenomics as a source for bioprospecting: Mar Genomics, v. 24 Pt 1, p. 21-30.

Li, P. Y., X. L. Chen, P. Ji, C. Y. Li, P. Wang, Y. Zhang, B. B. Xie, Q. L. Qin, H. N. Su, B. C. Zhou, Y. Z. Zhang, and X. Y. Zhang, 2015a, Interdomain hydrophobic interactions modulate the thermostability of microbial esterases from the hormone-sensitive lipase family: J Biol Chem, v. 290, p. 11188-98.

Li, P. Y., B. B. Xie, X. Y. Zhang, Q. L. Qin, H. Y. Dang, X. M. Wang, X. L. Chen, J. Yu, and Y. Z. Zhang, 2012, Genetic structure of three fosmid-fragments encoding 16S rRNA genes of the Miscellaneous Crenarchaeotic Group (MCG): implications for physiology and evolution of marine sedimentary archaea: Environ Microbiol, v. 14, p. 467-79.

Li, W., A. Cowley, M. Uludag, T. Gur, H. McWilliam, S. Squizzato, Y. M. Park, N. Buso, and R. Lopez, 2015b, The EMBL-EBI bioinformatics web and programmatic tools framework: Nucleic Acids Res, v. 43, p. W580-4.

Lüthy, R., J. U. Bowie, and D. Eisenberg, 1992, Assessment of protein models with three-dimensional profiles: Nature, v. 356, p. 83-5.

Maestro, S. ö., LLC, New York, NY, 2016, Schrödinger Release 2016-3.

Margesin, R., and F. Schinner, 2001, Potential of halotolerant and halophilic microorganisms for biotechnology: Extremophiles, v. 5, p. 73-83.

McWilliam, H., W. Li, M. Uludag, S. Squizzato, Y. M. Park, N. Buso, A. P. Cowley, and R. Lopez, 2013, Analysis Tool Web Services from the EMBL-EBI: Nucleic Acids Res, v. 41, p. W597-600.

Mirete, S., V. Morgante, and J. E. González-Pastor, 2016, Functional metagenomics of extreme environments: Curr Opin Biotechnol, v. 38, p. 143-9.

Mobarak-Qamsari, E., R. Kasra-Kermanshahi, and Z. Moosavi-Nejad, 2011, Isolation and identification of a novel, lipase-producing bacterium, *Pseudomonas aeruginosa* KM110: Iran J Microbiol, v. 3, p. 92-8.

Mohamed, Y. M., M. A. Ghazy, A. Sayed, A. Ouf, H. El-Dorry, and R. Siam, 2013, Isolation and characterization of a heavy metal-resistant, thermophilic esterase from a Red Sea brine pool: Sci Rep, v. 3, p. 3358.

Moreno, R., L. Jover, C. Diez, and C. Sanpera, 2011, Seabird feathers as monitors of the levels and persistence of heavy metal pollution after the Prestige oil spill: Environ Pollut, v. 159, p. 2454-60.

Ollis, D. L., E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I. Silman, and J. Schrag, 1992, The alpha/beta hydrolase fold: Protein Eng, v. 5, p. 197-211.

Panda, T., and B. S. Gowrishankar, 2005, Production and applications of esterases: Appl Microbiol Biotechnol, v. 67, p. 160-9.

Parages, M. L., J. A. Gutiérrez-Barranquero, F. J. Reen, A. D. Dobson, and F. O'Gara, 2016, Integrated (meta) genomic and synthetic biology approaches to develop new biocatalysts: Mar Drugs, v. 14.

Popovic, A., A. Tchigvintsev, H. Tran, T. N. Chernikova, O. V. Golyshina, M. M. Yakimov, P. N. Golyshin, and A. F. Yakunin, 2015, Metagenomics as a tool for enzyme discovery: Hydrolytic enzymes from marine-related metagenomes: Adv Exp Med Biol, v. 883, p. 1-20.

Ramnath, L., B. Sithole, and R. Govinden, 2017, Classification of lipolytic enzymes and their biotechnological applications in the pulping industry: Can J Microbiol, v. 63, p. 179-192.

Rao, T. E., M. Imchen, and R. Kumavath, 2017, Marine Enzymes: Production and Applications for Human Health: Adv Food Nutr Res, v. 80, p. 149-163.

Rhee, J. K., D. G. Ahn, Y. G. Kim, and J. W. Oh, 2005, New thermophilic and thermostable esterase with sequence similarity to the hormone-sensitive lipase family, cloned from a metagenomic library: Appl Environ Microbiol, v. 71, p. 817-25.

Santiago, M., C. A. Ramírez-Sarmiento, R. A. Zamora, and L. P. Parra, 2016, Discovery, molecular mechanisms, and industrial applications of cold-active enzymes: Front Microbiol, v. 7, p. 1408.

Sasso, F., A. Natalello, S. Castoldi, M. Lotti, C. Santambrogio, and R. Grandori, 2016, *Burkholderia cepacia* lipase is a promising biocatalyst for biofuel production: Biotechnol J, v. 11, p. 954-60.

Selvin, J., J. Kennedy, D. P. Lejon, G. S. Kiran, and A. D. Dobson, 2012, Isolation identification and biochemical characterization of a novel halo-tolerant lipase from the metagenome of the marine sponge *Haliclona simulans*: Microb Cell Fact, v. 11, p. 72.

Sharma, S., and S. S. Kanwar, 2014, Organic solvent tolerant lipases and applications: ScientificWorldJournal, v. 2014, p. 625258.

Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, and D. G. Higgins, 2011, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega: Mol Syst Biol, v. 7, p. 539.

Singh, R., M. Kumar, A. Mittal, and P. Mehta, 2016, Microbial enzymes: industrial process in 21s$^{t}$ Century: 3 Biotech, v. 6.

Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar, 2013, MEGA6: Molecular Evolutionary Genetics Analysis version 6.0: Mol Biol Evol, v. 30, p. 2725-9.

Van Domselaar, G. H., P. Stothard, S. Shrivastava, J. A. Cruz, A. Guo, X. Dong, P. Lu, D. Szafron, R. Greiner, and D. S. Wishart, 2005, BASys: a web server for automated bacterial genome annotation: Nucleic Acids Res, v. 33, p. W455-9.

Verger, R., 1997, 'Interfacial activation' of lipases: facts and artifacts: Trends in Biotechnology, v. 15, p. 32-38.

Vieites, J. M., M. E. Guazzaroni, A. Beloqui, P. N. Golyshin, and M. Ferrer, 2009, Metagenomics approaches in systems microbiology: FEMS Microbiol Rev, v. 33, p. 236-55.

Webb, B., and A. Sali, 2016, Comparative Protein Structure Modeling Using MODELLER: Curr Protoc Protein Sci, v. 86, p. 2.9.1-2.9.37.

Wilson, M. C., T. Mori, C. Rückert, A. R. Uria, M. J. Helf, K. Takada, C. Gernert, U. A. Steffens, N. Heycke, S. Schmitt, C. Rinke, E. J. Helfrich, A. O. Brachmann, C. Gurgui, T. Wakimoto, M. Kracht, M. Crüsemann, U. Hentschel, I. Abe, S. Matsunaga, J. Kalinowski, H. Takeyama, and J. Piel, 2014, An environmental bacterial taxon with a large and distinct metabolic repertoire: Nature, v. 506, p. 58-62.

Wirsen, C. O., and S. J. Molyneaux, 1999, A study of deep-sea natural microbial populations and barophilic pure cultures using a high-pressure chemostat: Appl Environ Microbiol, v. 65, p. 5314-21.

Yang, S., H. Jin, Z. Wei, R. He, Y. Ji, X. Li, and S. Yu, 2009, Bioremediation of oil spills in cold environments: A Review: Pedosphere, v. 19, p. 371-381.

# Chapter 5

# 5. General discussion

## 5.1 Secondary metabolites from deep sea sponges

The results of the study conducted and presented in chapter 2 indicate a comparable secondary metabolomic potential in deep sea sponges to that of shallow water sponges, highlighting their future potential in biotechnology and the wider health care sector. The applied 454 pyrosequencing approach employed here, using degenerate primer pairs targeting adenylation (AD) and ketosynthase (KS) domains of nonribosomal peptide synthetases and polyketide synthase gene clusters yielded a large number of potentially novel domains of these types and therefore indicates the presence of a potential "treasure trove" of novel secondary metabolites in the microbiome of deep sea sponges. Sequence similarities to gene clusters known to be involved in the synthesis of many different classes of antibiotics and toxins production genes were observed, for example these included lipopetides, glycopeptides, macrolides, streptogramins, depsipeptides, cyanoginosines, bacteriocins and hepatotoxins. An attempt was also made to affiliate the retrieved sequences to potential microbial producers. To achieve this the sequences were uploaded to MG-RAST (Meyer *et al.*, 2008) and subsequently compared to a previous 16S rRNA microbiome sequencing study of these deep sea sponge species (Kennedy et al., 2014). The affiliations of the AD and KS domain sequences appeared to be reasonable for large majorities of the generated data sets, as highly abundant phyla like *Proteobacteria* and *Actinobacteria* were identified; which is well in line with the huge secondary metabolite potential connected to these phyla (Gerth *et al.*, 1996; Wenzel and Müller, 2009; Chater *et al.*, 2010). Nonetheless and unexpectedly another bacterial phylum, *Cyanobacteria*, was identified as a prominent contributor of both AD and KS domains. *Cyanobacteria* are known producers of secondary metabolites, such as for example jamaicamides (Edwards *et al.*, 2004) and hectochlorin (Ramaswamy *et al.*, 2007), but all members of this phylum rely on photosynthesis for energy production, which is most unlikely to take place in the deep sea. A high rate of horizontal gene transfer and the common localization of PKS and NRPS gene clusters on 'genomic/pathogenicity islands' which are rich in mobile genetic elements may account for

the observed phenomena (Ridley *et al.*, 2008; Ziemert *et al.*, 2014). Due to the high horizontal transfer rate of secondary metabolite gene clusters a taxonomic identification can be more informative if accompanied by other approaches such as 16S rRNA sequencing, nonetheless the true origin of a specific cluster most likely remains hidden. The high rate of horizontal gene transfer of secondary metabolite gene clusters is well illustrated by the example of the pederin type of biosynthetic gene cluster which is known to be distributed widely in nature from beetles to sponges (Piel *et al.*, 2005). Pederin activity was first reported in 1919 from the beetle *Paedarus fuscipes* (Netolitzky, 1919; Frank and Kanamitsu, 1987) and 33 years later it was isolated by collecting 25 million specimens of the beetle and subsequently got its name Pederin (Narquizian and Kocienski, 2000). Derivatives of Pederin were later described from different marine sponges, like Mycalamide A&B from a marine sponge in New Zealand (Perry *et al.*, 1990), Onnamides from a Japanese sponge *Theonella* sp. (Matsunaga *et al.*, 1992; Kobayashi *et al.*, 1993) and Theopederin A-E from the same *Theonella* sponge species (Fusetani *et al.*, 1992) (Figure 1).

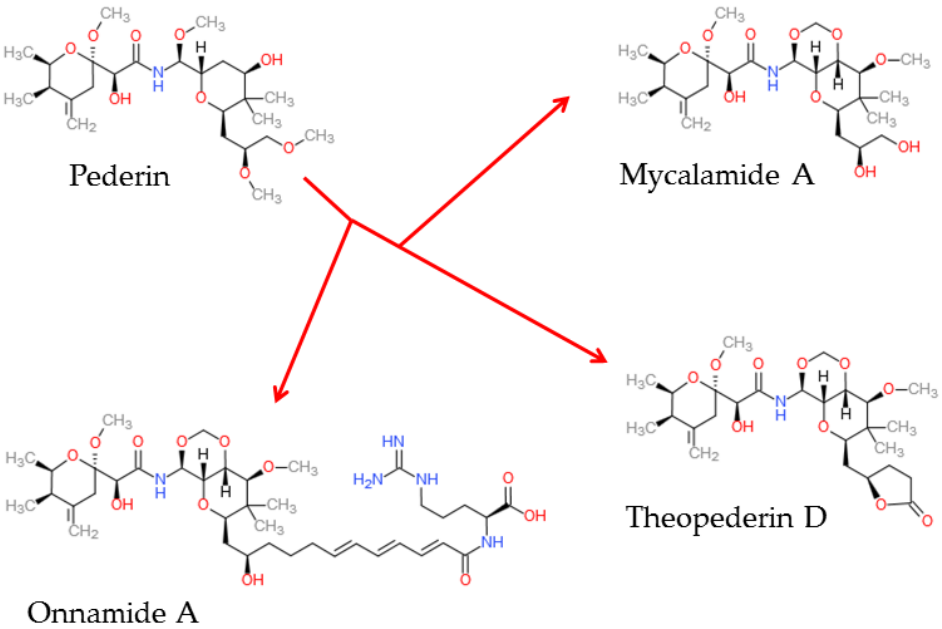



**Figure 1:** Pederin and its derivatives. Pederin is produced in the beetle *Paedarus fuscipes*, while Mycalamide D, Theopederin D and Onnamide A are representatives of secondary metabolites isolated from marine sponges (*Theonella*, *Mycale*).

To verify the secondary metabolite potential of deep sea sponges and to get more accurate insights into the actual microbes harbouring these clusters would require the use of

metagenomic fosmid or bacterial artificial chromosome clone (bac) libraries. These libraries prepared from deep sea sponge metagenomic DNA could be screened with PCR probes for specific AD and KS domains identified in this study and fosmids/bacs carrying these domains could be subsequently sequenced in their entirety to gain more information. Furthermore standard microbial cultivation approaches and screening for antimicrobial activities could also be applied, but bearing in mind the overall low numbers of microbes that can currently be cultured from environmental samples and the likelihood of even lower numbers of being able to be cultured from deep sea source; may render this approach undesirable.

In total the AD domains appeared not to be as abundant as the KS domains (1621 versus 14244), however one must also bear in mind that KS sequences are closely related to fatty acid synthases and sponges are known to be rich in these type of genes. Therefore one must be careful not to make definitive assumptions based on the numbers obtained. To circumvent this problem the sequences were further compared to reference database of true secondary metabolite affiliated AD and KS sequences, leading to the identification of 48 unique AD and 175 unique KS domains. Furthermore it should be noted that a rarefaction curve plateau was reached for only one of three of the sponge species analysed, indicating that deeper sequencing would be required to obtain a more complete overview of the true secondary metabolite potential of these sponges.

Following further analysis of an OTU network generated from the KS sequences of some individual sponge samples (Figure 2), remarkably only a small number of sequences were shared among the sponges and even among the same sponge species. This indicates the uniqueness of each sponge sample taken as it appears that each sponge has a large unique set of secondary metabolite genes, while only having a small set of shared genes. If this is a general trend in sponge microbiome secondary metabolite production then this needs to be further evaluated. It would be reasonable to assume that the differences observed may be as a result of depth or location dependent effects; as these sponge samples have been collected from different locations and depths. In any case these trends further highlight the potential of these deep sea sponge metagenomes as a good source of novel bioactive compounds, nonetheless it should also be remembered that it is extremely difficult and indeed expensive to retrieve samples from deep sea environments.

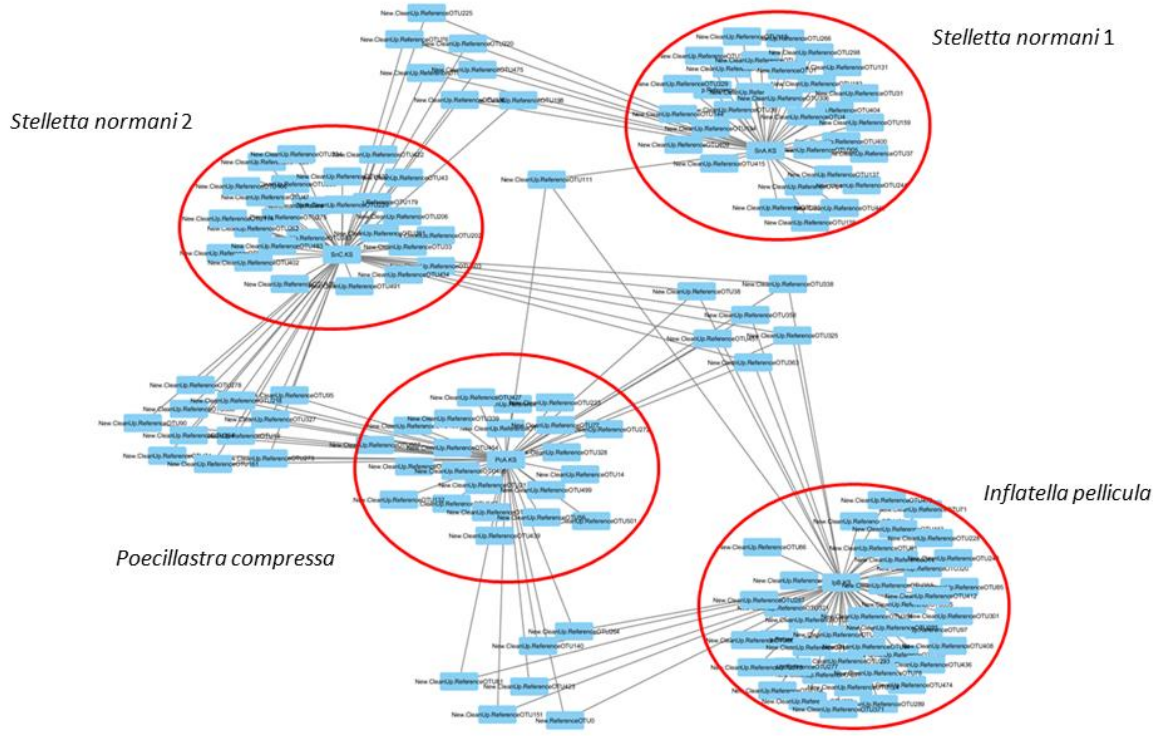



**Figure 2**: OTU network of representative ketosynthase sequences from individual sponge samples. The red circle indicates the 'core/unique' ketosynthase sequences of a given sponge sample. Sequences represented outside the circles are shared among the different sponge samples.

## 5.2 The significance of the genus *Pseudoalteromonas* spp.

The genus *Pseudoalteromonas* spp. can generally be divided into pigmented and non-pigmented strains. In the study presented in chapter three, non-pigmented *Pseudoalteromonas* spp. isolated from deep sea sponges were investigated with respect to their potential biotechnological applications. In addition a comparative genomic approach was applied to identify similarities and potential differences between free-living and host-associated *Pseudoalteromonas* strains. The isolates displayed various enzymatic activities and some of these (β-glucosidase, β-galactosidase and protease) were more closely investigated for their optimal temperature ranges. Somewhat surprisingly the targeted enzyme activities yielded different temperature optima, with only one being cold-adapted (β-glucosidase, 23°C optimal). A possible explanation for this observation could be that there is only a selective pressure for the β-glucosidase activity to be cold-adapted and the two other activities are not important for the strain in the conditions they are likely to encounter in the deep sea or perhaps there may be no selective pressure for protease and β-galactosidase activity. β-glucosidases are typically involved in the degradation of cellulose the most abundant polysaccharide on the planet (Klemm *et al.*, 2005) and act in conjunction with cellobiohydrolases and endoglucanases to achieve this. This type of enzyme is therefore likely to be important for marine *Pseudoalteromonas* species to increase the range of nutrient sources available to them and aid in their survival in the harsh deep sea environment. In an effort to gain further insights into the biotechnological potential of the isolates their genomes were sequenced and annotated. Subsequently four different putative β-glucosidase genes were identified in the most active isolate (EB27) and one of these genes was successfully subcloned and heterologously expressed in *Escherichia coli*. The recombinant β-glucosidase enzyme displayed the same optimal temperature pattern as the cultured isolate (data not shown in chapter 3). The β-galactosidase genes from SK20 and EB27 were further investigated and were found to be closely-related (99% and 92% protein identity) to a cold active β-galactosidase from *Pseudoalteromonas haloplanktis* (Hoyoux *et al.*, 2001). Attempts to subclone these β-galactosidase genes however proved unsuccessful. The number of protease genes in the genomes of the investigated genomes was quite high (39 to 48) and therefore it was difficult to pinpoint one specific gene as being responsible for the observed phenotype.

β-glucosidases, β-galactosidases and proteases were targeted initially, because all of these enzymes have certain applications in different industries (bioremediation, paper, pulp industry, winemaking, etc.) and there is an ongoing need for cold active enzymes from each of these enzyme classes. The most obvious reason to use cold adapted enzymes is to reduce costs via obtaining good reaction speeds at lower temperatures. Secondly enzymes can be used as biocatalysts in various chemical synthetic reactions, therefore reducing the amount of chemicals used and therefore making industrial processes more environmentally friendly. Furthermore more specific advances of cold adapted enzymes are their structural flexibility, promoting low substrate affinity and high specific activity (at low temperatures), reducing undesirable chemical side reactions that usually occur at higher temperatures and also facilitating an easy thermal inactivation of these enzymes, as they are normally thermolabile (Cavicchioli *et al.*, 2002; Santiago *et al.*, 2016).

The genomes of the three isolates and two reference strains were also compared at a whole genome level, because non-pigmented *Pseudoalteromonas* spp. have been described as being phylogenetically shallow, but we see a huge variability in plate based screenings for enzymatic activities. The genome comparison revealed that besides their phylogenetic shallowness the individual isolates each possess a huge number of unique genes, defining their overall pan-genome as open, consequently have a huge intraspecies genetic variability (Bosi *et al.*, 2017). The genomes shared 63% to 73% of all genes present and the number of unique genes per genome ranged from 8.5% to 20%, while approximately 20% to 25% of the genes were not annotated by the annotation pipeline (RAST) employed (Aziz *et al.*, 2008; Overbeek *et al.*, 2014; Brettin *et al.*, 2015) and were marked as 'hypothetical' or 'unknown'. Remarkably only ten genes are shared solely between the sponge isolates, including multidrug resistance genes, integrases, recombinases and cation efflux systems. This rather small number of in this case 'sponge isolate specific' genes renders a specific host adaptation or host-associated lifestyle of these isolates unlikely. In addition to this a conserved bacteriocin gene cluster was found in all isolates and the free living reference strains used, which contains a tetratricopeptide repeat (TPR) domain. Proteins containing TPR motifs can be found in virtually any type of organisms from bacteria to fungi to insects and plants and even in animals (Blatch and Lässle, 1999; Jernigan and Bordenstein, 2015). They mediate protein-protein interactions while being involved in many different cellular functions.

Besides cellular functions they are also believed to be involved in symbiosis (Siegl *et al.*, 2011; Reynolds and Thomas, 2016) and in bacterial mediated mammalian infection (Groshong *et al.*, 2014). Bacteriocins on the other hand are antimicrobial compounds which are produced as a means of defence or for communication purposes and differ from traditional antimicrobial compounds in their relatively narrow killing spectrum and are normally only active against bacteria that are closely related to the producing strain (Riley and Wertz, 2002). Furthermore bacteriocins encoding phage tail-like structures, especially from *Pseudoalteromonas* spp., have been associated with microbe-host interaction. These bacteriocins enable larvae of for example the tubeworm *Hydroides elegans* to settle onto a biofilm of *Pseudoalteromonas luteoviolacea* and trigger metamorphosis of the larvae (Shikuma *et al.*, 2014). This connection between the TPR domain and larvae settlement inducing capabilities of bacteriocins may indicate some kind of host-association characteristics for *Pseudoalteromonas* spp., but in general this cannot really be regarded as 'true symbiosis', and further work would need to be conducted to link this particular gene cluster found in the studied isolates with a role in host symbiosis. Nonetheless also *Pseudoalteromonas spongiae*, a sponge isolated *Pseudoalteromonas* spp. is able to induce larvae settlement in the marine tubeworm *Hydroides elegans*, its larvae settlement capabilities have not been tested with sponge larvae to date (Huang *et al.*, 2007).

In conclusion *Pseudoalteromonas* spp. are very versatile organisms to study being quite different from each other and are particularly interesting for its bioactivities (especially pigmented strains) and for enzymatic activities (especially non-pigmented strains) for industrial applications even considering that they are commonly isolated from cold-environments. In this respect a recent study Bosi *et al.*, 2017 where they compared on a large scale *Pseudoalteromonas* genomes from pigmented and non-pigmented strains highlighted the fact that HGT events are very common across *Pseudoalteromonas* spp. isolates, explaining to some extent the huge genomic variability found in the genus. They also reported the presence of at least one bacteriocin gene cluster in all investigated genomes and it might be interesting to evaluate the presence of a TPR domain in these clusters. Future research into the genus of *Pseudoalteromonas* spp. as model organisms for the expression of cold active enzymes (Papa *et al.*, 2007), for bioactive compounds (Bowman, 2007; Fehér *et al.*, 2010; Offret *et al.*, 2016) and enzymes with new biochemical traits (Yan *et al.*, 2009; Al Khudary *et*

*al.*, 2010; Yang *et al.*, 2016) is highly promising and many more discoveries from this rather young genus (Gauthier *et al.*, 1995) can be expected in the near future.

### 5.3 Metagenomic approaches to identify novel lipolytic enzymes

The true biotechnological potential of a given environmental sample is often quite difficult to assess and while a number of different approaches are available, each and every one of these has its own unique limitations. Cultivation of microbes from environmental samples is often used to isolate novel microorganisms, allowing them to be subsequently studied for various traits, such as for example secondary metabolite production, expression of enzymatic activities and adaptation to environmental stresses. Unfortunately it is now well established that only a very small proportion (0.1%-1%) of all microbes in any given environmental sample can be cultivated (Bernard *et al.*, 2000). With the rise of next generation sequencing technologies the sequencing of whole metagenomes, which allows the analysis of metagenomic DNA from all organisms in an environmental sample, has become increasingly affordable (Escobar-Zepeda *et al.*, 2015). The bottleneck to this approach is the wealth of data generated and the required time-consuming bioinformatics involved to decipher the information gained and to translate it into applicable knowledge. The obvious advantage is that problems associated with cultivation based approaches can to some extent become circumvented, as genomes from uncultivable microorganisms can be identified and investigated, nonetheless this is a pure *in silico* lead approach and preferably needs *in vitro* validation. Functional metagenomics based approaches provide a mechanism whereby genes identified *in silico* can be functionally expressed in a heterologous system, allowing for the heterologous protein to be biochemically characterised. For example *Escherichia coli* is routinely used as a heterologous host enabling the expression of a wide variety of metagenomic DNA from various environmental samples and their subsequent screening for the desired phenotypic trait (Schloss and Handelsman, 2003; Handelsman, 2004; Uchiyama and Miyazaki, 2009; Simon and Daniel, 2011). When coupled with a robust and high-throughput screening regime, this approach is an extremely effective way of isolating novel enzymes from otherwise inaccessible microbes. In addition, one of the major advantages of functional based metagenomics based approaches is that they have the potential to identify entirely new classes of genes encoding either known or indeed novel functions (Kennedy *et*

*al.*, 2011). However despite the success of functional metagenomics for the discovery of new enzymes, the approach can be limited to some extent by the ability of metagenomic clones to produce active enzymes. As previously mentioned many functional metagenomic approaches rely on the use of *E. coli* as a host to express metagenome encoded proteins. While a large number of genes derived from *Enterobacteriaceae* can readily be expressed in common *E. coli* host systems, many genes from more distantly related organisms may not be expressed. This can occur for example due to the promoter regions of these genes not being recognized by the *E. coli* transcriptional machinery or due to differences in codon usage; being expressed at low levels. Even where transcription and translation of foreign genes results in efficient protein expression, additional problems can occur when proteins need to be post-translationally modified or exported for activity. For these reasons, the availability of suitable heterologous expression hosts remains one of the main barriers to functional metagenomic based screening (Coughlan *et al.*, 2015; Mirete *et al.*, 2016). This is highlighted by our attempt to subclone and heterologously express five different protease genes (data not shown) from the afore mentioned sequenced *Pseudoalteromonas* sp. isolates (Chapter 3). While successful cloning of the protease genes was confirmed via gel electrophoresis for at least three of the proteases no phenotypic activity was observed, which is therefore most likely due to inefficient protein expression or missing post-translational modifications.

In chapter 4 a functional metagenomic approach was employed to investigate the biotechnological potential of the microbiome of the deep sea sponge *Stelletta normani*. A metagenomic fosmid library comprising of 14,000 clones with an average insert size of approximately ~35-40kb was constructed and screened for various enzymatic (protease, cellulase, lipase and amylase activity) and antimicrobial (antibacterial and antifungal) activities. Remarkably a high prevalence for lipolytic clones was observed, with more than 30 positive clones being found. Four of the most active lipase fosmid clones were chosen for sequencing in an attempt to identify the gene responsible for the phenotype. The gene from the most active fosmid clone was subsequently chosen for heterologous expression in the pBAD overexpression system in *Escherichia coli*. The gene showed close relatedness (66% protein identity) to a formerly identified esterase (E40) of the hormone-sensitive lipase family (Hsl) IV from another marine metagenomic library (Li *et al.*, 2015); with both esterases sharing the HGG and GDSAG motif of the GDSAG subfamily of the type IV Hsl family.

Furthermore the gene was closely related (61% protein identity) to sequences related to the metabolically versatile bacterium *Entotheonella* TSY1 and TSY2. This to date uncultured, symbiotic Gram-negative δ-proteobacterium is known for its secondary metabolite potential and detoxifying capabilities (Wilson *et al.*, 2014; Liu *et al.*, 2016; Keren *et al.*, 2017). Biochemical investigation revealed that our esterase is distinctly different from the aforementioned esterase E40, particularly with respect to its substrate specificity, temperature activity profile and halotolerance profile. The herein newly described esterase 7N9 is most active towards short-chain fatty acids (C2), is truly cold-adapted, with peak activity from 4°C to 20°C. It also displayed a wide halotolerance, losing only 12% of its activity when the salt concentration is increased from 0% to 10%. The esterase E40 was reported to be most active towards fatty acids with a side chain length of four carbon atoms, its optimal temperature is approximately 45°C and it is most active at a salt concentration of 3%, with rapid decline thereafter (Li *et al.*, 2015). Moreover the crystal structure of E40 (Li *et al.*, 2015) and Ramachandran plotting (Figure 3) were used to predict a 3D structure for 7N9 and subsequently perform *in silico* docking studies with several substrates and inhibitors.

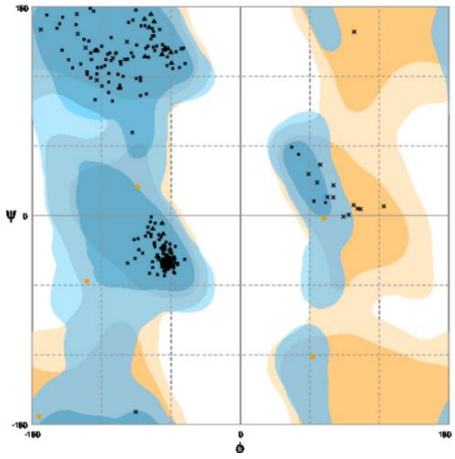


**Figure 3:** Ramachandran plot of 7N9 esterase with E40 esterase as template. Black dots within blue areas indicate residues in favoured regions (98% of all dots) and black dots in beige regions indicate allowed regions (2%).

The *in silico* docking studies confirmed the preference of 7N9 for short-chained fatty acids and showed high docking scores for the inhibitor phenylmethanesulfonic acid. This inhibitor is known to be able to bind covalently to the active site serine residue, confirming the likely involvement of this residue in the reaction mechanism of the esterase (Selvin *et al.*, 2012). The esterase 7N9 described in chapter 4 is a novel truly cold active and halotolerant

enzyme identified via a metagenomic approach from a deep sea sponge sample. The properties of this esterase make it suitable for industrial processes and environmental recovery projects. Possible applications of this type of enzyme may be in the bioremediation of oil spills and especially its cold adaptation is here of major importance, as the prevalence of oil spills in cold environments are becoming more abundant due to the increased industrial exploitation of these environments and search for new oil reservoirs in remote locations (Yang *et al.*, 2009).

## 5.4 Conclusions and future prospects

This thesis aimed to broaden our understanding of deep sea microorganisms and their relationship to sponges, while investigating the secondary metabolite potential for novel drug leads and the biochemical characteristics of deep sea enzymes and consequential providing input for future exploitation efforts of the deep sea environment. The secondary metabolite potential in the microbiome of deep sea sponges revealed here is notable and justifies further exploration, nonetheless it can be only regarded as a glimpse or "snap shot" of the true potential of deep sea sponges as more sponge samples would need to be investigated together with more deep sequencing, using other sequencing platforms and various other degenerate primer pairs, to allow an appreciation of the total extent and complexity of all the secondary metabolite gene clusters present in this ecosystem. Furthermore besides next generation sequencing approaches to target certain domains of secondary metabolite biosynthetic gene clusters other approaches would also need to be employed to get a better understanding of the entire gene clusters surrounding these domains to assess both their potential novelty and products produced from the clusters. This could be achieved through the screening of metagenomic libraries with probes generated for interesting domains obtained by NGS approaches. If an interesting secondary metabolite gene cluster is identified in a metagenomic library or a silence cryptic gene cluster in a bacterial genome the entire gene cluster could be subcloned via transformation-associated recombination (TAR) cloning into a suitable bacterial or yeast expression system to facilitate overexpression and subsequent biochemical characterization (Ongley *et al.*, 2013). Such TAR cloning based approaches have already successfully been used to subclone biosynthetic gene clusters from the obligate marine actinomycete *Salinispora* (Bonet *et al.*, 2015; Tang *et al.*, 2015).

Additionally the aforementioned high likelihood of horizontal gene transfer of natural product gene clusters should be monitored if further samples and results become available, to be able to estimate more precisely the true potential of the microbiome of deep sea sponges. The different enzymatic activities described from *Pseudoalteromonas* spp. isolates and from the metagenomic screening of environmental DNA from the deep sea sponge *Stelletta normani* has shed light on the possible applications of enzymes from the deep sea for industrial purposes, as some of them possess useful traits such as cold

adaptation and halotolerance. Desirable enzymatic activities from the isolated *Pseudoalteromonas* spp. would be worth further investigation, particularly the cold active β-glucosidase already subcloned into *Escherichia coli*. β-glucosidases are key enzymes in the breakdown of cellulose, one of the most abundant polysaccharides on the planet and can be used for various applications. β-glucosidases are normally the rate limiting enzyme in cellulose breakdown catalysing the final step where cellobiose and other oligosaccharides are converted to glucose, because unfortunately it is itself inhibited by glucose (Xiao et al., 2004), therefore screening for novel β-glucosidases and tailoring these enzymes for increased performance is important (Sørensen *et al.*, 2013). Due to its oligosaccharide reducing capabilities β-glucosidases are used for various different industrial applications such as hydrolysis of bitter compounds in juice and wine, release of aromatic compounds from fruits and fermenting products and other food related applications (Singh *et al.*, 2016). Besides food flavour applications this type of enzyme is also important in the production of biofuels, as it can be used in the conversion of lignocellulosic substrates to fermentable sugars (Li *et al.*, 2013).

Thus the results presented here form a good basis for future studies on not only the secondary metabolite potential of the microbiota of deep sea sponge, but also of their potential role in host-microbe association/relations particularly in relation to marine invertebrates in the deep sea environment as well as the potential of deep sea bacteria to produce enzymes of industrial interest with sought after biochemical traits.

## 5.5 Bibliography


Al Khudary, R., R. Venkatachalam, M. Katzer, S. Elleuche, and G. Antranikian, 2010, A cold-adapted esterase of a novel marine isolate, *Pseudoalteromonas arctica*: gene cloning, enzyme purification and characterization: Extremophiles, v. 14, p. 273-85.

Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, and O. Zagnitko, 2008, The RAST Server: rapid annotations using subsystems technology: BMC Genomics, v. 9, p. 75.

Bernard, L., H. Schaefer, F. Joux, C. Courties, G. Muyzer, and P. Lebaron, 2000, Genetic diversity of total, active and culturable marine bacteria in coastal seawater.: Aquatic Microbial Ecology, v. 23, p. 1-11.

Blatch, G. L., and M. Lässle, 1999, The tetratricopeptide repeat: a structural motif mediating protein-protein interactions: Bioessays, v. 21, p. 932-9.

Bonet, B., R. Teufel, M. Crüsemann, N. Ziemert, and B. S. Moore, 2015, Direct capture and heterologous expression of *Salinispora* natural product genes for the biosynthesis of enterocin: J Nat Prod, v. 78, p. 539-42.

Bosi, E., M. Fondi, V. Orlandini, E. Perrin, I. Maida, D. de Pascale, M. L. Tutino, E. Parrilli, A. Lo Giudice, A. Filloux, and R. Fani, 2017, The pangenome of (Antarctic) *Pseudoalteromonas* bacteria: evolutionary and functional insights: BMC Genomics, v. 18, p. 93.

Bowman, J. P., 2007, Bioactive compound synthetic capacity and ecological significance of marine bacterial genus *Pseudoalteromonas*: Mar Drugs, v. 5, p. 220-41.

Brettin, T., J. J. Davis, T. Disz, R. A. Edwards, S. Gerdes, G. J. Olsen, R. Olson, R. Overbeek, B. Parrello, G. D. Pusch, M. Shukla, J. A. Thomason, R. Stevens, V. Vonstein, A. R. Wattam, and F. Xia, 2015, RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes: Sci Rep, v. 5, p. 8365.

Cavicchioli, R., K. S. Siddiqui, D. Andrews, and K. R. Sowers, 2002, Low-temperature extremophiles and their applications: Curr Opin Biotechnol, v. 13, p. 253-61.

Chater, K. F., S. Biró, K. J. Lee, T. Palmer, and H. Schrempf, 2010, The complex extracellular biology of *Streptomyces*: FEMS Microbiol Rev, v. 34, p. 171-98.

Coughlan, L. M., P. D. Cotter, C. Hill, and A. Alvarez-Ordóñez, 2015, Biotechnological applications of functional metagenomics in the food and pharmaceutical industries: Front Microbiol, v. 6, p. 672.

Edwards, D. J., B. L. Marquez, L. M. Nogle, K. McPhail, D. E. Goeger, M. A. Roberts, and W. H. Gerwick, 2004, Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*: Chem Biol, v. 11, p. 817-33.

Escobar-Zepeda, A., A. Vera-Ponce de León, and A. Sanchez-Flores, 2015, The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics: Front Genet, v. 6, p. 348.

Fehér, D., R. Barlow, J. McAtee, and T. K. Hemscheidt, 2010, Highly brominated antimicrobial metabolites from a marine *Pseudoalteromonas* sp: J Nat Prod, v. 73, p. 1963-6.

Frank, J. H., and K. Kanamitsu, 1987, *Paederus*, sensu lato (Coleoptera: *Staphylinidae*): natural history and medical importance: J Med Entomol, v. 24, p. 155-91.

Fusetani, N., T. Sugawara, and S. Matsunaga, 1992, Bioactive marine metabolites. 41. Theopederins A-E, potent antitumor metabolites from a marine sponge, *Theonella* sp: Journal of Organic Chemistry, v. 57, p. 3828-3832.

Gauthier, G., M. Gauthier, and R. Christen, 1995, Phylogenetic analysis of the genera *Alteromonas*, *Shewanella*, and *Moritella* using genes coding for small-subunit rRNA sequences and division of the genus *Alteromonas* into two genera, *Alteromonas* (emended) and *Pseudoalteromonas* gen. nov., and proposal of twelve new species combinations: Int J Syst Bacteriol, v. 45, p. 755-61.

Gerth, K., N. Bedorf, G. Höfle, H. Irschik, and H. Reichenbach, 1996, Epothilons A and B: antifungal and cytotoxic compounds from *Sorangium cellulosum* (Myxobacteria). Production, physico-chemical and biological properties: J Antibiot (Tokyo), v. 49, p. 560-3.

Groshong, A. M., D. E. Fortune, B. P. Moore, H. J. Spencer, R. A. Skinner, W. T. Bellamy, and J. S. Blevins, 2014, BB0238, a presumed tetratricopeptide repeat-containing protein, is

required during *Borrelia burgdorferi* mammalian infection: Infect Immun, v. 82, p. 4292-306.

Handelsman, J., 2004, Metagenomics: application of genomics to uncultured microorganisms: Microbiol Mol Biol Rev, v. 68, p. 669-85.

Hoyoux, A., I. Jennes, P. Dubois, S. Genicot, F. Dubail, J. M. François, E. Baise, G. Feller, and C. Gerday, 2001, Cold-adapted beta-galactosidase from the Antarctic psychrophile *Pseudoalteromonas haloplanktis*: Appl Environ Microbiol, v. 67, p. 1529-35.

Huang, Y. L., S. Dobretsov, H. Xiong, and P. Y. Qian, 2007, Effect of biofilm formation by *Pseudoalteromonas spongiae* on induction of larval settlement of the polychaete *Hydroides elegans*: Appl Environ Microbiol, v. 73, p. 6284-8.

Jernigan, K. K., and S. R. Bordenstein, 2015, Tandem-repeat protein domains across the tree of life: PeerJ, v. 3, p. e732.

Kennedy, J., B. Flemer, S. A. Jackson, J. P. Morrissey, F. O'Gara, and A. D. Dobson, 2014, Evidence of a putative deep sea specific microbiome in marine sponges: PLoS One, v. 9, p. e91092.

Kennedy, J., N. D. O'Leary, G. S. Kiran, J. P. Morrissey, F. O'Gara, J. Selvin, and A. D. Dobson, 2011, Functional metagenomic strategies for the discovery of novel enzymes and biosurfactants with biotechnological applications from marine ecosystems: J Appl Microbiol, v. 111, p. 787-99.

Keren, R., B. Mayzel, A. Lavy, I. Polishchuk, D. Levy, S. C. Fakra, B. Pokroy, and M. Ilan, 2017, Sponge-associated bacteria mineralize arsenic and barium on intracellular vesicles: Nat Commun, v. 8, p. 14393.

Klemm, D., B. Heublein, H. P. Fink, and A. Bohn, 2005, Cellulose: fascinating biopolymer and sustainable raw material: Angew Chem Int Ed Engl, v. 44, p. 3358-93.

Kobayashi, J., F. Itagaki, H. Shigemori, and T. Sasaki, 1993, Three new onnamide congeners from the Okinawan marine sponge *Theonella* sp.: Journal of Natural Products, v. 56, p. 976-981.

Li, D., X. Li, W. Dang, P. L. Tran, S. H. Park, B. C. Oh, W. S. Hong, J. S. Lee, and K. H. Park, 2013, Characterization and application of an acidophilic and thermostable β-glucosidase from *Thermofilum pendens*: J Biosci Bioeng, v. 115, p. 490-6.

Li, P. Y., X. L. Chen, P. Ji, C. Y. Li, P. Wang, Y. Zhang, B. B. Xie, Q. L. Qin, H. N. Su, B. C. Zhou, Y. Z. Zhang, and X. Y. Zhang, 2015, Interdomain hydrophobic interactions modulate the thermostability of microbial esterases from the hormone-sensitive lipase family: J Biol Chem, v. 290, p. 11188-98.

Liu, F., J. Li, G. Feng, and Z. Li, 2016, New Genomic Insights into "*Entotheonella*" Symbionts in *Theonella swinhoei*: Mixotrophy, Anaerobic Adaptation, Resilience, and Interaction: Front Microbiol, v. 7, p. 1333.

Matsunaga, S., N. Fusetani, and Y. Nakao, 1992, Eight new cytotoxic metabolites closely related to onnamide A from two marine sponges of the genus *Theonella*: Tetrahedron, v. 48, p. 8369-8376.

Meyer, F., D. Paarmann, M. D'Souza, R. Olson, E. M. Glass, M. Kubal, T. Paczian, A. Rodriguez, R. Stevens, A. Wilke, J. Wilkening, and R. A. Edwards, 2008, The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes: BMC Bioinformatics, v. 9, p. 386.

Mirete, S., V. Morgante, and J. E. González-Pastor, 2016, Functional metagenomics of extreme environments: Curr Opin Biotechnol, v. 38, p. 143-9.

Narquizian, R., and P. J. Kocienski, 2000, The pederin family of antitumor agents: structures, synthesis and biological activity: Ernst Schering Res Found Workshop, p. 25-56.

Netolitzky, F., 1919, Kaefer als Nahrungs- und Heilmittel: Koleopterologische Rundschau, v. 8.

Offret, C., F. Desriac, P. Le Chevalier, J. Mounier, C. Jégou, and Y. Fleury, 2016, Spotlight on antimicrobial metabolites from the marine bacteria *Pseudoalteromonas*: Chemodiversity and Ecological Significance: Mar Drugs, v. 14.

Ongley, S. E., X. Bian, B. A. Neilan, and R. Müller, 2013, Recent advances in the heterologous expression of microbial natural product biosynthetic pathways: Nat Prod Rep, v. 30, p. 1121-38.

Overbeek, R., R. Olson, G. D. Pusch, G. J. Olsen, J. J. Davis, T. Disz, R. A. Edwards, S. Gerdes, B. Parrello, M. Shukla, V. Vonstein, A. R. Wattam, F. Xia, and R. Stevens, 2014, The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST): Nucleic Acids Res, v. 42, p. D206-14.

Papa, R., V. Rippa, G. Sannia, G. Marino, and A. Duilio, 2007, An effective cold inducible expression system developed in *Pseudoalteromonas haloplanktis* TAC125: J Biotechnol, v. 127, p. 199-210.

Perry, N., J. Blunt, M. Munro, and A. Thompson, 1990, Antiviral and antitumor agents from a New Zealand sponge, *Mycale* sp. 2. Structures and solution conformations of mycalamides A and B: Journal of Organic Chemistry, v. 55, p. 223-227.

Piel, J., D. Butzke, N. Fusetani, D. Hui, M. Platzer, G. Wen, and S. Matsunaga, 2005, Exploring the chemistry of uncultivated bacterial symbionts: antitumor polyketides of the pederin family: J Nat Prod, v. 68, p. 472-9.

Ramaswamy, A. V., C. M. Sorrels, and W. H. Gerwick, 2007, Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*: J Nat Prod, v. 70, p. 1977-86.

Reynolds, D., and T. Thomas, 2016, Evolution and function of eukaryotic-like proteins from sponge symbionts: Mol Ecol, v. 25, p. 5242-5253.

Ridley, C. P., H. Y. Lee, and C. Khosla, 2008, Evolution of polyketide synthases in bacteria: Proc Natl Acad Sci U S A, v. 105, p. 4595-600.

Riley, M. A., and J. E. Wertz, 2002, Bacteriocins: evolution, ecology, and application: Annu Rev Microbiol, v. 56, p. 117-37.

Santiago, M., C. A. Ramírez-Sarmiento, R. A. Zamora, and L. P. Parra, 2016, Discovery, molecular mechanisms, and industrial applications of cold-active enzymes: Front Microbiol, v. 7, p. 1408.

Schloss, P. D., and J. Handelsman, 2003, Biotechnological prospects from metagenomics: Curr Opin Biotechnol, v. 14, p. 303-10.

Selvin, J., J. Kennedy, D. P. Lejon, G. S. Kiran, and A. D. Dobson, 2012, Isolation identification and biochemical characterization of a novel halo-tolerant lipase from the metagenome of the marine sponge *Haliclona simulans*: Microb Cell Fact, v. 11, p. 72.

Shikuma, N. J., M. Pilhofer, G. L. Weiss, M. G. Hadfield, G. J. Jensen, and D. K. Newman, 2014, Marine tubeworm metamorphosis induced by arrays of bacterial phage tail-like structures: Science, v. 343, p. 529-33.

Siegl, A., J. Kamke, T. Hochmuth, J. Piel, M. Richter, C. Liang, T. Dandekar, and U. Hentschel, 2011, Single-cell genomics reveals the lifestyle of *Poribacteria*, a candidate phylum symbiotically associated with marine sponges: ISME J, v. 5, p. 61-70.

Simon, C., and R. Daniel, 2011, Metagenomic analyses: past and future trends: Appl Environ Microbiol, v. 77, p. 1153-61.

Singh, G., A. K. Verma, and V. Kumar, 2016, Catalytic properties, functional attributes and industrial applications of β-glucosidases: 3 Biotech, v. 6.

Sørensen, A., M. Lübeck, P. S. Lübeck, and B. K. Ahring, 2013, Fungal Beta-glucosidases: a bottleneck in industrial use of lignocellulosic materials: Biomolecules, v. 3, p. 612-31.

Tang, X., J. Li, N. Millán-Aguiñaga, J. J. Zhang, E. C. O'Neill, J. A. Ugalde, P. R. Jensen, S. M. Mantovani, and B. S. Moore, 2015, Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining: ACS Chem Biol, v. 10, p. 2841-9.

Uchiyama, T., and K. Miyazaki, 2009, Functional metagenomics for enzyme discovery: challenges to efficient screening: Curr Opin Biotechnol, v. 20, p. 616-22.

Wenzel, S. C., and R. Müller, 2009, *Myxobacteria*--'microbial factories' for the production of bioactive secondary metabolites: Mol Biosyst, v. 5, p. 567-74.

Wilson, M. C., T. Mori, C. Rückert, A. R. Uria, M. J. Helf, K. Takada, C. Gernert, U. A. Steffens, N. Heycke, S. Schmitt, C. Rinke, E. J. Helfrich, A. O. Brachmann, C. Gurgui, T. Wakimoto, M. Kracht, M. Crüsemann, U. Hentschel, I. Abe, S. Matsunaga, J. Kalinowski, H. Takeyama, and J. Piel, 2014, An environmental bacterial taxon with a large and distinct metabolic repertoire: Nature, v. 506, p. 58-62.

Xiao, Z., X. Zhang, D. J. Gregg, and J. N. Saddler, 2004, Effects of sugar inhibition on cellulases and beta-glucosidase during enzymatic hydrolysis of softwood substrates: Appl Biochem Biotechnol, v. 113-116, p. 1115-26.

Yan, B. Q., X. L. Chen, X. Y. Hou, H. He, B. C. Zhou, and Y. Z. Zhang, 2009, Molecular analysis of the gene encoding a cold-adapted halophilic subtilase from deep-sea psychrotolerant bacterium *Pseudoalteromonas* sp. SM9913: cloning, expression, characterization and function analysis of the C-terminal PPC domains: Extremophiles, v. 13, p. 725-33.

Yang, J., Y. Yu, B. L. Tang, S. Zhong, M. Shi, B. B. Xie, X. Y. Zhang, B. C. Zhou, Y. Z. Zhang, and X. L. Chen, 2016, Pilot-scale production and thermostability improvement of the M23 protease pseudoalterin from the deep sea bacterium *Pseudoalteromonas* sp. CF6-2: Molecules, v. 21.

Yang, S., H. Jin, Z. Wei, R. He, Y. Ji, X. Li, and S. Yu, 2009, Bioremediation of oil spills in cold environments: A Review: Pedosphere, v. 19, p. 371-381.

Ziemert, N., A. Lechner, M. Wietz, N. Millán-Aguiñaga, K. L. Chavarria, and P. R. Jensen, 2014, Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*: Proc Natl Acad Sci U S A, v. 111, p. E1130-9.

# 6. Bibliography

## 6.1 Publication list

### 6.1.1 Original Research Article

Borchert E, Jackson SA, O'Gara F, Dobson AD: **Diversity of Natural Product Biosynthetic Genes in the Microbiome of the Deep Sea Sponges *Inflatella pellicula*, *Poecillastra compressa*, and *Stelletta normani***. Front Microbiol 2016, **7**:1027. (published)

Borchert E, Knobloch S, Dwyer E, O'Flynn S, Jackson SA, Jóhannsson R, Marteinsson VT, O'Gara F, Dobson ADW: **Biotechnological potential of cold adapted *Pseudoalteromonas* spp. isolated from 'deep sea' sponges.** Mar Drugs 2017 (published)

Borchert E, Selvin J, Kiran GS, Jackson SA, O'Gara F, Dobson ADW: **Characterization of a novel cold active deep sea esterase from a metagenomic library from the sponge *Stelletta normani*.** Front Microbiol *2017* (submitted)

### 6.1.2 Review

Jackson SA, Borchert E, O'Gara F, Dobson ADW: **Metagenomics for the discovery of novel biosurfactants of environmental interest from marine ecosystems**. Curr Opin Biotechnol 2015, **33**:176-182. (published)

### 6.1.3 Book Chapter

Borchert E, Jackson SA, O'Gara F, Dobson ADW: Psychrophiles: From Biodiversity to Biotechnology 2nd Edition 20017, **Chapter 23: Psychrophiles as a source of novel antimicrobials.** Springer Verlag, Berlin Heidelberg (accepted)

Steinert G, Huete-Stauffer C, Aas-Valleriani N, Borchert E, Bhushan A, Campbell A, Mares MCD, Costa AM, Gutleben J, Knobloch S, Lee RG, Munroe S, Naik D, Peters EE, Stokes E, Wang W, Einarsdóttír E, Sipkema D: Grand Challenges in Marine Biotechnology 2017, Section 4 'Grand projects': **BluePharmTrain – A European Sponge Biotechnology Project.** Springer Verlag, Berlin Heidelberg (submitted)

# 7. Appendix

## 7.1 Supplementary material Chapter 2

### S1 Table: Primers used in this study

| AD | Sequence | Sample |
|---|---|---|
| A16A3F | CGTATCGCCTCCCTCGCGCCATCAG TCACGTACTA GCSTACSYSATSTACACSTCSGG | BD243 |
| B16A7R | CTATGCGCCTTGCCAGCCCGCTCAG TCACGTACTA SASGTCVCCSGTSCGGTAS | |
| A19A3F | CGTATCGCCTCCCTCGCGCCATCAG TGTACTACTC GCSTACSYSATSTACACSTCSGG | BD130 |
| B19A7R | CTATGCGCCTTGCCAGCCCGCTCAG TGTACTACTC SASGTCVCCSGTSCGGTAS | |
| A50A3F | CGTATCGCCTCCCTCGCGCCATCAG ACTAGCAGTA GCSTACSYSATSTACACSTCSGG | BD226 |
| B50A7R | CTATGCGCCTTGCCAGCCCGCTCAG ACTAGCAGTA SASGTCVCCSGTSCGGTAS | |
| A62A3F | CGTATCGCCTCCCTCGCGCCATCAG TACGTCATCA GCSTACSYSATSTACACSTCSGG | BD92 |
| B62A7R | CTATGCGCCTTGCCAGCCCGCTCAG TACGTCATCA SASGTCVCCSGTSCGGTAS | |
| A47A3R | CGTATCGCCTCCCTCGCGCCATCAG TGTGAGTAGT GCSTACSYSATSTACACSTCSGG | BDV1267 |
| B47A7R | CTATGCGCCTTGCCAGCCCGCTCAG TGTGAGTAGT SASGTCVCCSGTSCGGTAS | |
| A15A3F | CGTATCGCCTCCCTCGCGCCATCAG ATACGACGTA GCSTACSYSATSTACACSTCSGG | BDV1379 |
| B15A7R | CTATGCGCCTTGCCAGCCCGCTCAG ATACGACGTA SASGTCVCCSGTSCGGTAS | |
| A52A3F | CGTATCGCCTCCCTCGCGCCATCAG AGTATACATA GCSTACSYSATSTACACSTCSGG | BDV1346 |
| B52A7RKS | CTATGCGCCTTGCCAGCCCGCTCAG AGTATACATA SASGTCVCCSGTSCGGTAS | |
| A38KSiF | CGTATCGCCTCCCTCGCGCCATCAG TACACGTGAT GCIATGGAYCCICARCARMGIVT | BD243 |
| B38KSiR | CTATGCGCCTTGCCAGCCCGCTCAG TACACGTGAT GTICCIGTICCRTGISCYTCIAC | |
| A41KSiF | CGTATCGCCTCCCTCGCGCCATCAG TAGTGTAGAT GCIATGGAYCCICARCARMGIVT | BD130 |
| B41KSiR | CTATGCGCCTTGCCAGCCCGCTCAG TAGTGTAGAT GTICCIGTICCRTGISCYTCIAC | |
| A70KSiF | CGTATCGCCTCCCTCGCGCCATCAG TGAGTCAGTA GCIATGGAYCCICARCARMGIVT | BD92 |
| B70KSiR | CTATGCGCCTTGCCAGCCCGCTCAG TGAGTCAGTA GTICCIGTICCRTGISCYTCIAC | |
| A60KSiF | CGTATCGCCTCCCTCGCGCCATCAG CTACGCTCTA GCIATGGAYCCICARCARMGIVT | BDV1267 |
| B60KSiR | CTATGCGCCTTGCCAGCCCGCTCAG CTACGCTCTA GTICCIGTICCRTGISCYTCIAC | |

| | | |
|---|---|---|
| A37KSiF | CGTATCGCCTCCCTCGCGCCATCAG TACACACACT GCIATGGAYCCICARCARMGIVT | BDV1379 |
| B37KSiR | CTATGCGCCTTGCCAGCCCGCTCAG TACACACACT GTICCIGTICCRTGISCYTCIAC | |

**S2 Table: Sequence alignment output of KS sequences with NaPDos database**

| Query id | Database match id | % id. | align length | e-value | pathway product | class |
|---|---|---|---|---|---|---|
| New.CleanUp.ReferenceOTU0_IpB.KS_8412 | StiE_Q8RJY2_1KSB | 47 | 208 | 2.00E-44 | stigmatellin | modular |
| New.ReferenceOTU0_PcA.KS_7856 | CurA_AAT70096_mod | 54 | 90 | 1.00E-32 | curacin | KS |
| New.ReferenceOTU0_PcA.KS_7856 | CurA_AAT70096_mod | 37 | 83 | 1.00E-32 | curacin | KS |
| New.CleanUp.ReferenceOTU1_PcA.KS_6826 | StiG_Q8RJY0_1KSB | 63 | 105 | 6.00E-55 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU1_PcA.KS_6826 | StiG_Q8RJY0_1KSB | 68 | 68 | 6.00E-55 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU103_PcA.KS_6079 | StiC_Q8RJY4_1KSB | 68 | 139 | 2.00E-49 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU103_PcA.KS_6079 | StiC_Q8RJY4_1KSB | 66 | 29 | 2.00E-49 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU103_PcA.KS_6079 | StiC_Q8RJY4_1KSB | 53 | 17 | 2.00E-49 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU104_PcA.KS_4856 | CurL_AAT70107_mod | 75 | 135 | 8.00E-41 | curacin | modular |
| New.CleanUp.ReferenceOTU106_PcA.KS_3967 | EpoE_Q9L8C6_1mod | 58 | 203 | 4.00E-57 | epothilone | modular |
| New.CleanUp.ReferenceOTU106_PcA.KS_3967 | EpoE_Q9L8C6_1mod | 81 | 16 | 4.00E-57 | epothilone | modular |
| New.CleanUp.ReferenceOTU11_PcA.KS_8003 | CurA_AAT70096_mod | 51 | 128 | 4.00E-44 | curacin | KS |
| New.CleanUp.ReferenceOTU11_PcA.KS_8003 | CurA_AAT70096_mod | 85 | 26 | 4.00E-44 | curacin | KS |
| New.CleanUp.ReferenceOTU11_PcA.KS_8003 | CurA_AAT70096_mod | 45 | 40 | 4.00E-44 | curacin | KS |
| New.CleanUp.ReferenceOTU113_PcA.KS_4694 | CALO5_12183629_i | 57 | 197 | 1.00E-56 | calicheamicin | iterative |
| New.CleanUp.ReferenceOTU113_PcA.KS_4694 | CALO5_12183629_i | 69 | 29 | 1.00E-56 | calicheamicin | iterative |
| New.CleanUp.ReferenceOTU116_PcA.KS_4922 | MtaB_Q9RFL0_1KSB | 76 | 29 | 2.00E-13 | myxothiazol | KS1 |
| New.CleanUp.ReferenceOTU116_PcA. | MtaB_Q9RFL0_1KS | 54 | 28 | 2.00E- | myxothiazol | KS1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| KS_4922 | B | | | 13 | | |
| New.CleanUp.ReferenceOTU122_PcA.KS_6145 | MtaB_Q9RFL0_2KSB | 65 | 104 | 1.00E-55 | myxothiazol | modular |
| New.CleanUp.ReferenceOTU122_PcA.KS_6145 | MtaB_Q9RFL0_2KSB | 82 | 39 | 1.00E-55 | myxothiazol | modular |
| New.CleanUp.ReferenceOTU122_PcA.KS_6145 | MtaB_Q9RFL0_2KSB | 62 | 40 | 1.00E-55 | myxothiazol | modular |
| New.CleanUp.ReferenceOTU127_PcA.KS_6364 | StiG_Q8RJY0_1KSB | 61 | 157 | 1.00E-64 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU127_PcA.KS_6364 | StiG_Q8RJY0_1KSB | 63 | 30 | 1.00E-64 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU127_PcA.KS_6364 | StiG_Q8RJY0_1KSB | 65 | 31 | 1.00E-64 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU131_PcA.KS_7377 | EpoE_Q9L8C6_1mod | 47 | 159 | 4.00E-30 | epothilone | modular |
| New.CleanUp.ReferenceOTU134_PcA.KS_6272 | EpoD_Q9L8C7_2mod | 47 | 232 | 5.00E-52 | epothilone | modular |
| New.CleanUp.ReferenceOTU136_PcA.KS_8120 | MxaB_Q93TX0_1KSB | 61 | 163 | 2.00E-49 | myxalamid | modular |
| New.CleanUp.ReferenceOTU136_PcA.KS_8120 | MxaB_Q93TX0_1KSB | 63 | 19 | 2.00E-49 | myxalamid | modular |
| New.CleanUp.ReferenceOTU138_PcA.KS_5049 | StiG_Q8RJY0_1KSB | 50 | 165 | 6.00E-42 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU138_PcA.KS_5049 | StiG_Q8RJY0_1KSB | 59 | 39 | 6.00E-42 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU139_PcA.KS_6442 | MtaB_Q9RFL0_2KSB | 54 | 89 | 4.00E-42 | myxothiazol | modular |
| New.CleanUp.ReferenceOTU139_PcA.KS_6442 | MtaB_Q9RFL0_2KSB | 59 | 56 | 4.00E-42 | myxothiazol | modular |
| New.CleanUp.ReferenceOTU139_PcA.KS_6442 | MtaB_Q9RFL0_2KSB | 46 | 35 | 4.00E-42 | myxothiazol | modular |
| New.CleanUp.ReferenceOTU153_PcA.KS_6288 | TylGIII_O33956_1mod | 57 | 75 | 1.00E-29 | tylosin | modular |
| New.CleanUp.ReferenceOTU153_PcA.KS_6288 | TylGIII_O33956_1mod | 68 | 34 | 1.00E-29 | tylosin | modular |
| New.CleanUp.ReferenceOTU19_PcA.KS_4305 | CurJ_AAT70105_mod | 63 | 181 | 3.00E-64 | curacin | modular |
| New.CleanUp.ReferenceOTU19_PcA.KS_4305 | CurJ_AAT70105_mod | 52 | 40 | 3.00E-64 | curacin | modular |
| New.CleanUp.ReferenceOTU2_PcA.KS_6820 | EpoD_Q9L8C7_4mod | 78 | 80 | 2.00E-55 | epothilone | modular |
| New.CleanUp.ReferenceOTU2_PcA.KS_6820 | EpoD_Q9L8C7_4mod | 56 | 71 | 2.00E-55 | epothilone | modular |
| New.CleanUp.ReferenceOTU2_PcA.KS_6820 | EpoD_Q9L8C7_4mod | 72 | 25 | 2.00E-55 | epothilone | modular |
| New.CleanUp.ReferenceOTU21_PcA. | EpoD_Q9L8C7_4m | 61 | 177 | 7.00E- | epothilone | modul |

| | | | | | | |
|---|---|---|---|---|---|---|
| KS_2730 | od | | | 51 | | ar |
| New.CleanUp.ReferenceOTU25_PcA.KS_8098 | MycAIII_Q83WE8_1KSB | 53 | 45 | 9.00E-09 | mycinamicin | modular |
| New.CleanUp.ReferenceOTU29_PcA.KS_5493 | EpoD_Q9L8C7_4mod | 67 | 199 | 2.00E-72 | epothilone | modular |
| New.CleanUp.ReferenceOTU34_PcA.KS_7090 | StiH_Q8RJX9_1KSB | 44 | 212 | 8.00E-42 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU37_PcA.KS_7809 | EpoD_Q9L8C7_3mod | 53 | 90 | 3.00E-46 | epothilone | modular |
| New.CleanUp.ReferenceOTU37_PcA.KS_7809 | EpoD_Q9L8C7_3mod | 64 | 74 | 3.00E-46 | epothilone | modular |
| New.CleanUp.ReferenceOTU40_PcA.KS_6246 | KirAII_CAN89632_5T | 46 | 168 | 1.00E-51 | kirromycin | trans |
| New.CleanUp.ReferenceOTU40_PcA.KS_6246 | KirAII_CAN89632_5T | 53 | 55 | 1.00E-51 | kirromycin | trans |
| New.CleanUp.ReferenceOTU41_PcA.KS_7883 | Stro2780_2 | 45 | 147 | 4.00E-26 | salinilactam | modular |
| New.CleanUp.ReferenceOTU41_PcA.KS_7883 | Stro2780_2 | 71 | 17 | 4.00E-26 | salinilactam | modular |
| New.CleanUp.ReferenceOTU42_PcA.KS_5408 | StiA_Q8RJY6_1KSB | 70 | 117 | 8.00E-43 | stigmatellin | KS1 |
| New.CleanUp.ReferenceOTU43_PcA.KS_5085 | SpnA_Q9ALM6_1KSB | 66 | 134 | 1.00E-46 | spinosad | modular |
| New.CleanUp.ReferenceOTU48_PcA.KS_6738 | TylGI_O33954_2mod | 71 | 134 | 6.00E-78 | tylosin | modular |
| New.CleanUp.ReferenceOTU48_PcA.KS_6738 | TylGI_O33954_2mod | 77 | 75 | 6.00E-78 | tylosin | modular |
| New.CleanUp.ReferenceOTU49_PcA.KS_7041 | EpoF_Q9L8C5_1mod | 78 | 59 | 9.00E-24 | epothilone | modular |
| New.CleanUp.ReferenceOTU50_PcA.KS_8029 | EpoE_Q9L8C6_1mod | 60 | 213 | 3.00E-66 | epothilone | modular |
| New.CleanUp.ReferenceOTU51_PcA.KS_6732 | JamE_AAS98777_KS1 | 59 | 126 | 4.00E-63 | jamaicamide | KS |
| New.CleanUp.ReferenceOTU51_PcA.KS_6732 | JamE_AAS98777_KS1 | 75 | 60 | 4.00E-63 | jamaicamide | KS |
| New.CleanUp.ReferenceOTU58_PcA.KS_7548 | StiE_Q8RJY2_1KSB | 58 | 142 | 1.00E-51 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU58_PcA.KS_7548 | StiE_Q8RJY2_1KSB | 62 | 32 | 1.00E-51 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU58_PcA.KS_7548 | StiE_Q8RJY2_1KSB | 67 | 21 | 1.00E-51 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU6_PcA.KS_5303 | LipC_ABB05104_1KSB | 70 | 150 | 8.00E-60 | lipomycin | modular |
| New.CleanUp.ReferenceOTU61_PcA.KS_6415 | ChlB1_AAZ77673_i | 54 | 120 | 5.00E-46 | chlorothricin | iterative |
| New.CleanUp.ReferenceOTU61_PcA. | ChlB1_AAZ77673_i | 74 | 43 | 5.00E- | chlorothricin | iterativ |

| | | | | | | |
|---|---|---|---|---|---|---|
| KS_6415 | | | | 46 | | e |
| New.CleanUp.ReferenceOTU61_PcA.KS_6415 | ChlB1_AAZ77673_i | 67 | 27 | 5.00E-46 | chlorothricin | iterative |
| New.CleanUp.ReferenceOTU62_PcA.KS_4285 | AveA3_Q9S0R4_3mod | 54 | 135 | 6.00E-49 | avermectin | modular |
| New.CleanUp.ReferenceOTU62_PcA.KS_4285 | AveA3_Q9S0R4_3mod | 65 | 54 | 6.00E-49 | avermectin | modular |
| New.CleanUp.ReferenceOTU62_PcA.KS_4285 | AveA3_Q9S0R4_3mod | 81 | 21 | 6.00E-49 | avermectin | modular |
| New.CleanUp.ReferenceOTU64_PcA.KS_6418 | NosB_Q9RAH3_H | 63 | 178 | 1.00E-61 | nostopeptolide | hybrid KS |
| New.CleanUp.ReferenceOTU64_PcA.KS_6418 | NosB_Q9RAH3_H | 67 | 30 | 1.00E-61 | nostopeptolide | hybrid KS |
| New.CleanUp.ReferenceOTU66_PcA.KS_6004 | JamK_AAS98782_mod | 67 | 220 | 2.00E-69 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU67_PcA.KS_4944 | JamE_AAS98777_KS1 | 62 | 163 | 1.00E-57 | jamaicamide | KS |
| New.CleanUp.ReferenceOTU68_PcA.KS_6006 | CurI_AAT70104_mod | 59 | 192 | 8.00E-61 | curacin | modular |
| New.CleanUp.ReferenceOTU69_PcA.KS_8131 | JamL_AAS98783_mod | 58 | 158 | 2.00E-44 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU69_PcA.KS_8131 | JamL_AAS98783_mod | 68 | 22 | 2.00E-44 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU73_PcA.KS_7226 | ChlB1_AAZ77673_i | 53 | 152 | 3.00E-49 | chlorothricin | iterative |
| New.CleanUp.ReferenceOTU73_PcA.KS_7226 | ChlB1_AAZ77673_i | 57 | 35 | 3.00E-49 | chlorothricin | iterative |
| New.CleanUp.ReferenceOTU78_PcA.KS_4842 | VicC_BAD08359_1KSB | 72 | 65 | 2.00E-30 | vicenistatin | modular |
| New.CleanUp.ReferenceOTU78_PcA.KS_4842 | VicC_BAD08359_1KSB | 79 | 28 | 2.00E-30 | vicenistatin | modular |
| New.CleanUp.ReferenceOTU79_PcA.KS_4095 | ChlA2_AAZ77694_2KSB | 64 | 211 | 3.00E-72 | chlorothricin | modular |
| New.CleanUp.ReferenceOTU81_PcA.KS_3207 | JamK_AAS98782_mod | 64 | 123 | 6.00E-46 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU81_PcA.KS_3207 | JamK_AAS98782_mod | 84 | 19 | 6.00E-46 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU86_PcA.KS_5188 | PimS1_Q9X993_3KSB | 50 | 208 | 2.00E-46 | pimaricin | modular |
| New.CleanUp.ReferenceOTU90_PcA.KS_7319 | SpnD_Q9ALM3_3KSB | 67 | 98 | 5.00E-61 | spinosad | modular |
| New.CleanUp.ReferenceOTU90_PcA.KS_7319 | SpnD_Q9ALM3_3KSB | 61 | 87 | 5.00E-61 | spinosad | modular |
| New.CleanUp.ReferenceOTU92_PcA.KS_6863 | SpnD_Q9ALM3_1KSB | 69 | 213 | 3.00E-70 | spinosad | modular |
| New.CleanUp.ReferenceOTU95_PcA. | EpoC_Q9L8C8_H | 62 | 159 | 3.00E- | epothilone | hybrid |

| | | | | | | |
|---|---|---|---|---|---|---|
| KS_4447 | | | | 61 | | KS |
| New.CleanUp.ReferenceOTU95_PcA.KS_4447 | EpoC_Q9L8C8_H | 58 | 24 | 3.00E-61 | epothilone | hybrid KS |
| New.CleanUp.ReferenceOTU95_PcA.KS_4447 | EpoC_Q9L8C8_H | 52 | 27 | 3.00E-61 | epothilone | hybrid KS |
| New.CleanUp.ReferenceOTU96_PcA.KS_3644 | MxaB_Q93TX0_1KSB | 50 | 132 | 1.00E-51 | myxalamid | modular |
| New.CleanUp.ReferenceOTU96_PcA.KS_3644 | MxaB_Q93TX0_1KSB | 68 | 53 | 1.00E-51 | myxalamid | modular |
| New.CleanUp.ReferenceOTU96_PcA.KS_3644 | MxaB_Q93TX0_1KSB | 44 | 27 | 1.00E-51 | myxalamid | modular |
| New.CleanUp.ReferenceOTU100_SnC.KS_3847 | PikAII_Q9ZGI4_1KSB | 57 | 179 | 1.00E-49 | pikromycin | modular |
| New.CleanUp.ReferenceOTU115_SnC.KS_27131 | ChlA5_AAZ77698_1KSB | 64 | 203 | 7.00E-74 | chlorothricin | modular |
| New.CleanUp.ReferenceOTU118_SnA.KS_20033 | TetE_BAE93730_3mod | 74 | 76 | 3.00E-45 | tetronomycin | modular |
| New.CleanUp.ReferenceOTU118_SnA.KS_20033 | TetE_BAE93730_3mod | 61 | 77 | 3.00E-45 | tetronomycin | modular |
| New.CleanUp.ReferenceOTU118_SnA.KS_20033 | TetE_BAE93730_3mod | 73 | 15 | 3.00E-45 | tetronomycin | modular |
| New.CleanUp.ReferenceOTU118_SnA.KS_20033 | TetE_BAE93730_3mod | 75 | 12 | 3.00E-45 | tetronomycin | modular |
| New.CleanUp.ReferenceOTU123_SnC.KS_9934 | AveA2_Q9S0R7_3mod | 57 | 164 | 6.00E-50 | avermectin | modular |
| New.CleanUp.ReferenceOTU133_SnA.KS_26669 | CurI_AAT70104_mod | 59 | 179 | 5.00E-50 | curacin | modular |
| New.CleanUp.ReferenceOTU135_SnC.KS_6797 | SpnD_Q9ALM3_3KSB | 68 | 144 | 2.00E-55 | spinosad | modular |
| New.CleanUp.ReferenceOTU135_SnC.KS_6797 | SpnD_Q9ALM3_3KSB | 72 | 18 | 2.00E-55 | spinosad | modular |
| New.CleanUp.ReferenceOTU138_SnC.KS_32467 | JamM_AAS98784_H | 60 | 100 | 2.00E-56 | jamaicamide | hybrid KS |
| New.CleanUp.ReferenceOTU138_SnC.KS_32467 | JamM_AAS98784_H | 67 | 64 | 2.00E-56 | jamaicamide | hybrid KS |
| New.CleanUp.ReferenceOTU138_SnC.KS_32467 | JamM_AAS98784_H | 50 | 24 | 2.00E-56 | jamaicamide | hybrid KS |
| New.CleanUp.ReferenceOTU144_SnA.KS_31766 | CurI_AAT70104_mod | 65 | 143 | 2.00E-52 | curacin | modular |
| New.CleanUp.ReferenceOTU149_SnA.KS_20234 | CALO5_12183629_i | 54 | 218 | 3.00E-60 | calicheamicin | iterative |
| New.CleanUp.ReferenceOTU161_SnC.KS_33548 | AveA4_Q9S0R3_2mod | 64 | 120 | 9.00E-42 | avermectin | modular |
| New.CleanUp.ReferenceOTU161_SnC.KS_33548 | AveA4_Q9S0R3_2mod | 85 | 13 | 9.00E-42 | avermectin | modular |
| New.CleanUp.ReferenceOTU165_SnC. | ChlB1_AAZ77673_i | 60 | 94 | 9.00E- | chlorothricin | iterativ |

| | | | | | | |
|---|---|---|---|---|---|---|
| KS_34237 | | | | 47 | | e |
| New.CleanUp.ReferenceOTU165_SnC.KS_34237 | ChlB1_AAZ77673_i | 66 | 47 | 9.00E-47 | chlorothricin | iterative |
| New.CleanUp.ReferenceOTU165_SnC.KS_34237 | ChlB1_AAZ77673_i | 43 | 54 | 9.00E-47 | chlorothricin | iterative |
| New.CleanUp.ReferenceOTU166_SnC.KS_32005 | EpoD_Q9L8C7_4mod | 62 | 152 | 5.00E-63 | epothilone | modular |
| New.CleanUp.ReferenceOTU166_SnC.KS_32005 | EpoD_Q9L8C7_4mod | 64 | 47 | 5.00E-63 | epothilone | modular |
| New.CleanUp.ReferenceOTU178_SnC.KS_27407 | StiE_Q8RJY2_1KSB | 49 | 188 | 3.00E-40 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU179_SnA.KS_20095 | JamK_AAS98782_mod | 52 | 183 | 8.00E-47 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU179_SnA.KS_20095 | JamK_AAS98782_mod | 89 | 9 | 8.00E-47 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU180_SnC.KS_23940 | FurC1_ABB88521_KSB | 67 | 90 | 6.00E-37 | 5-alkenyl-3,3(2h)-furanone | modular |
| New.CleanUp.ReferenceOTU180_SnC.KS_23940 | FurC1_ABB88521_KSB | 83 | 23 | 6.00E-37 | 5-alkenyl-3,3(2h)-furanone | modular |
| New.CleanUp.ReferenceOTU203_SnC.KS_18676 | StiE_Q8RJY2_1KSB | 50 | 210 | 3.00E-52 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU204_SnC.KS_22809 | EcoE_AAX98188_3KSB | 65 | 195 | 4.00E-73 | eco-02301 | modular |
| New.CleanUp.ReferenceOTU204_SnC.KS_22809 | EcoE_AAX98188_3KSB | 57 | 23 | 4.00E-73 | eco-02301 | modular |
| New.CleanUp.ReferenceOTU217_SnC.KS_13158 | JamE_AAS98777_KS1 | 58 | 95 | 8.00E-28 | jamaicamide | KS |
| New.CleanUp.ReferenceOTU219_SnC.KS_31427 | AveA4_Q9S0R3_2mod | 51 | 156 | 1.00E-33 | avermectin | modular |
| New.CleanUp.ReferenceOTU221_SnA.KS_26896 | JamE_AAS98777_KS1 | 57 | 191 | 7.00E-58 | jamaicamide | KS |
| New.CleanUp.ReferenceOTU221_SnA.KS_26896 | JamE_AAS98777_KS1 | 72 | 18 | 7.00E-58 | jamaicamide | KS |
| New.CleanUp.ReferenceOTU223_SnC.KS_28590 | LipC_ABB05104_1KSB | 68 | 115 | 5.00E-58 | lipomycin | modular |
| New.CleanUp.ReferenceOTU223_SnC.KS_28590 | LipC_ABB05104_1KSB | 63 | 60 | 5.00E-58 | lipomycin | modular |
| New.CleanUp.ReferenceOTU225_SnC.KS_31358 | CurJ_AAT70105_mod | 60 | 155 | 2.00E-60 | curacin | modular |
| New.CleanUp.ReferenceOTU225_SnC.KS_31358 | CurJ_AAT70105_mod | 59 | 61 | 2.00E-60 | curacin | modular |
| New.CleanUp.ReferenceOTU230_SnC.KS_17755 | JamE_AAS98777_KS1 | 52 | 203 | 3.00E-59 | jamaicamide | KS |
| New.CleanUp.ReferenceOTU230_SnC.KS_17755 | JamE_AAS98777_KS1 | 55 | 20 | 3.00E-59 | jamaicamide | KS |
| New.CleanUp.ReferenceOTU232_SnA. | TylGIII_O33956_1 | 53 | 188 | 2.00E- | tylosin | modul |

| | | | | | | |
|---|---|---|---|---|---|---|
| KS_34475 | mod | | | 46 | | ar |
| New.CleanUp.ReferenceOTU232_SnA.KS_34475 | TylGIII_O33956_1mod | 82 | 33 | 2.00E-46 | tylosin | modular |
| New.CleanUp.ReferenceOTU235_SnC.KS_14184 | StiG_Q8RJY0_1KSB | 72 | 87 | 7.00E-38 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU235_SnC.KS_14184 | StiG_Q8RJY0_1KSB | 71 | 21 | 7.00E-38 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU236_SnC.KS_8939 | EpoD_Q9L8C7_4mod | 55 | 230 | 9.00E-65 | epothilone | modular |
| New.CleanUp.ReferenceOTU239_SnC.KS_25668 | CurA_AAT70096_mod | 69 | 68 | 3.00E-42 | curacin | KS |
| New.CleanUp.ReferenceOTU239_SnC.KS_25668 | CurA_AAT70096_mod | 52 | 86 | 3.00E-42 | curacin | KS |
| New.CleanUp.ReferenceOTU239_SnC.KS_25668 | CurA_AAT70096_mod | 69 | 16 | 3.00E-42 | curacin | KS |
| New.CleanUp.ReferenceOTU240_SnC.KS_23492 | AveA4_Q9S0R3_2mod | 76 | 137 | 2.00E-62 | avermectin | modular |
| New.CleanUp.ReferenceOTU240_SnC.KS_23492 | AveA4_Q9S0R3_2mod | 68 | 28 | 2.00E-62 | avermectin | modular |
| New.CleanUp.ReferenceOTU246_SnC.KS_28319 | SpnA_Q9ALM6_1KSB | 65 | 134 | 3.00E-45 | spinosad | modular |
| New.CleanUp.ReferenceOTU254_SnA.KS_34590 | ChlB1_AAZ77673_i | 51 | 173 | 3.00E-45 | chlorothricin | iterative |
| New.CleanUp.ReferenceOTU254_SnA.KS_34590 | ChlB1_AAZ77673_i | 56 | 36 | 3.00E-45 | chlorothricin | iterative |
| New.CleanUp.ReferenceOTU254_SnA.KS_34590 | ChlB1_AAZ77673_i | 73 | 11 | 3.00E-45 | chlorothricin | iterative |
| New.CleanUp.ReferenceOTU257_SnA.KS_35444 | LnmJ_AF484556_4T | 56 | 63 | 2.00E-25 | leinamycin | trans |
| New.CleanUp.ReferenceOTU257_SnA.KS_35444 | LnmJ_AF484556_4T | 51 | 37 | 2.00E-25 | leinamycin | trans |
| New.CleanUp.ReferenceOTU257_SnA.KS_35444 | LnmJ_AF484556_4T | 71 | 17 | 2.00E-25 | leinamycin | trans |
| New.CleanUp.ReferenceOTU257_SnA.KS_35444 | LnmJ_AF484556_4T | 38 | 32 | 2.00E-25 | leinamycin | trans |
| New.CleanUp.ReferenceOTU258_SnC.KS_36616 | StiF_Q8RJY1_1KSB | 53 | 92 | 1.00E-24 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU262_SnA.KS_21832 | CALO5_12183629_i | 64 | 81 | 6.00E-27 | calicheamicin | iterative |
| New.CleanUp.ReferenceOTU262_SnA.KS_21832 | CALO5_12183629_i | 69 | 13 | 6.00E-27 | calicheamicin | iterative |
| New.CleanUp.ReferenceOTU264_SnA.KS_20245 | JamE_AAS98777_KS1 | 63 | 84 | 6.00E-21 | jamaicamide | KS |
| New.CleanUp.ReferenceOTU266_SnC.KS_29268 | ChlB1_AAZ77673_i | 61 | 180 | 2.00E-59 | chlorothricin | iterative |
| New.CleanUp.ReferenceOTU271_SnC. | AveA4_Q9S0R3_2 | 60 | 108 | 3.00E- | avermectin | modul |

| | | | | | | |
|---|---|---|---|---|---|---|
| KS_7918 | mod | | | 39 | | ar |
| New.CleanUp.ReferenceOTU271_SnC.KS_7918 | AveA4_Q9S0R3_2mod | 86 | 21 | 3.00E-39 | avermectin | modular |
| New.CleanUp.ReferenceOTU272_SnC.KS_11826 | EpoD_Q9L8C7_4mod | 62 | 133 | 2.00E-48 | epothilone | modular |
| New.CleanUp.ReferenceOTU275_SnA.KS_12135 | MtaB_Q9RFL0_1KSB | 56 | 109 | 3.00E-49 | myxothiazol | KS1 |
| New.CleanUp.ReferenceOTU275_SnA.KS_12135 | MtaB_Q9RFL0_1KSB | 56 | 48 | 3.00E-49 | myxothiazol | KS1 |
| New.CleanUp.ReferenceOTU275_SnA.KS_12135 | MtaB_Q9RFL0_1KSB | 38 | 65 | 3.00E-49 | myxothiazol | KS1 |
| New.CleanUp.ReferenceOTU276_SnC.KS_25657 | JamE_AAS98777_KS1 | 42 | 161 | 4.00E-35 | jamaicamide | KS |
| New.CleanUp.ReferenceOTU276_SnC.KS_25657 | JamE_AAS98777_KS1 | 58 | 33 | 4.00E-35 | jamaicamide | KS |
| New.CleanUp.ReferenceOTU277_SnC.KS_16211 | AveA4_Q9S0R3_3mod | 65 | 213 | 2.00E-73 | avermectin | modular |
| New.CleanUp.ReferenceOTU278_SnC.KS_36608 | AveA4_Q9S0R3_2mod | 46 | 194 | 3.00E-33 | avermectin | modular |
| New.CleanUp.ReferenceOTU281_SnC.KS_19700 | AveA2_Q9S0R7_4mod | 79 | 150 | 2.00E-58 | avermectin | modular |
| New.CleanUp.ReferenceOTU30_SnA.KS_23211 | StiG_Q8RJY0_1KSB | 66 | 106 | 2.00E-59 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU30_SnA.KS_23211 | StiG_Q8RJY0_1KSB | 57 | 84 | 2.00E-59 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU300_SnC.KS_12556 | AveA4_Q9S0R3_2mod | 82 | 57 | 8.00E-34 | avermectin | modular |
| New.CleanUp.ReferenceOTU300_SnC.KS_12556 | AveA4_Q9S0R3_2mod | 71 | 35 | 8.00E-34 | avermectin | modular |
| New.CleanUp.ReferenceOTU38_SnC.KS_26004 | EpoD_Q9L8C7_4mod | 66 | 168 | 3.00E-63 | epothilone | modular |
| New.CleanUp.ReferenceOTU47_SnC.KS_9398 | JamE_AAS98777_KS1 | 51 | 165 | 2.00E-47 | jamaicamide | KS |
| New.CleanUp.ReferenceOTU48_SnC.KS_36821 | CurL_AAT70107_mod | 56 | 148 | 8.00E-53 | curacin | modular |
| New.CleanUp.ReferenceOTU48_SnC.KS_36821 | CurL_AAT70107_mod | 57 | 49 | 8.00E-53 | curacin | modular |
| New.CleanUp.ReferenceOTU51_SnC.KS_16295 | JamE_AAS98777_KS1 | 57 | 149 | 5.00E-61 | jamaicamide | KS |
| New.CleanUp.ReferenceOTU51_SnC.KS_16295 | JamE_AAS98777_KS1 | 49 | 63 | 5.00E-61 | jamaicamide | KS |
| New.CleanUp.ReferenceOTU55_SnC.KS_22107 | StiG_Q8RJY0_1KSB | 57 | 157 | 4.00E-75 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU55_SnC.KS_22107 | StiG_Q8RJY0_1KSB | 75 | 67 | 4.00E-75 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU57_SnC. | Stro2778_1 | 67 | 89 | 1.00E- | salinilactam | modul |

| | | | | | | |
|---|---|---|---|---|---|---|
| KS_26481 | | | | 58 | | ar |
| New.CleanUp.ReferenceOTU57_SnC.KS_26481 | Stro2778_1 | 52 | 106 | 1.00E-58 | salinilactam | modular |
| New.CleanUp.ReferenceOTU57_SnC.KS_26481 | Stro2778_1 | 71 | 17 | 1.00E-58 | salinilactam | modular |
| New.CleanUp.ReferenceOTU65_SnC.KS_28498 | ChlB1_AAZ77673_i | 60 | 181 | 4.00E-69 | chlorothricin | iterative |
| New.CleanUp.ReferenceOTU65_SnC.KS_28498 | ChlB1_AAZ77673_i | 62 | 45 | 4.00E-69 | chlorothricin | iterative |
| New.CleanUp.ReferenceOTU88_SnC.KS_31623 | EpoD_Q9L8C7_4mod | 60 | 143 | 2.00E-59 | epothilone | modular |
| New.CleanUp.ReferenceOTU88_SnC.KS_31623 | EpoD_Q9L8C7_4mod | 58 | 57 | 2.00E-59 | epothilone | modular |
| New.CleanUp.ReferenceOTU90_SnC.KS_12662 | MtaD_Q9RFK8_1KSB | 51 | 175 | 1.00E-52 | myxothiazol | hybrid KS |
| New.CleanUp.ReferenceOTU90_SnC.KS_12662 | MtaD_Q9RFK8_1KSB | 88 | 16 | 1.00E-52 | myxothiazol | hybrid KS |
| New.CleanUp.ReferenceOTU92_SnC.KS_17813 | AveA4_Q9S0R3_2mod | 56 | 171 | 5.00E-60 | avermectin | modular |
| New.CleanUp.ReferenceOTU92_SnC.KS_17813 | AveA4_Q9S0R3_2mod | 71 | 45 | 5.00E-60 | avermectin | modular |
| New.CleanUp.ReferenceOTU99_SnC.KS_31079 | SpnC_Q9ALM4_2KSB | 60 | 142 | 4.00E-55 | spinosad | modular |
| New.CleanUp.ReferenceOTU99_SnC.KS_31079 | SpnC_Q9ALM4_2KSB | 68 | 34 | 4.00E-55 | spinosad | modular |
| New.CleanUp.ReferenceOTU101_IpB.KS_8118 | StiE_Q8RJY2_1KSB | 59 | 129 | 1.00E-58 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU101_IpB.KS_8118 | StiE_Q8RJY2_1KSB | 68 | 60 | 1.00E-58 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU108_IpB.KS_6810 | EcoA_AAX98184_2KSB | 82 | 28 | 2.00E-17 | eco-02301 | modular |
| New.CleanUp.ReferenceOTU108_IpB.KS_6810 | EcoA_AAX98184_2KSB | 53 | 32 | 2.00E-17 | eco-02301 | modular |
| New.CleanUp.ReferenceOTU108_IpB.KS_6810 | EcoA_AAX98184_2KSB | 50 | 20 | 2.00E-17 | eco-02301 | modular |
| New.CleanUp.ReferenceOTU108_IpB.KS_6810 | EcoA_AAX98184_2KSB | 59 | 17 | 2.00E-17 | eco-02301 | modular |
| New.CleanUp.ReferenceOTU111_IpB.KS_7399 | LnmJ_AF484556_2T | 54 | 79 | 1.00E-36 | leinamycin | trans |
| New.CleanUp.ReferenceOTU111_IpB.KS_7399 | LnmJ_AF484556_2T | 44 | 77 | 1.00E-36 | leinamycin | trans |
| New.CleanUp.ReferenceOTU111_IpB.KS_7399 | LnmJ_AF484556_2T | 71 | 17 | 1.00E-36 | leinamycin | trans |
| New.CleanUp.ReferenceOTU112_IpB.KS_4072 | SpnD_Q9ALM3_3KSB | 70 | 93 | 1.00E-61 | spinosad | modular |
| New.CleanUp.ReferenceOTU112_IpB. | SpnD_Q9ALM3_3K | 51 | 117 | 1.00E- | spinosad | modul |

| | | | | | | |
|---|---|---|---|---|---|---|
| KS_4072 | SB | | | 61 | | ar |
| New.CleanUp.ReferenceOTU113_IpB.KS_5760 | CurJ_AAT70105_mod | 54 | 167 | 2.00E-35 | curacin | modular |
| New.CleanUp.ReferenceOTU115_IpB.KS_6919 | EpoE_Q9L8C6_1mod | 58 | 160 | 4.00E-42 | epothilone | modular |
| New.CleanUp.ReferenceOTU115_IpB.KS_6919 | EpoE_Q9L8C6_1mod | 48 | 23 | 4.00E-42 | epothilone | modular |
| New.CleanUp.ReferenceOTU120_IpB.KS_7317 | MxaB_Q93TX0_1KSB | 58 | 162 | 6.00E-61 | myxalamid | modular |
| New.CleanUp.ReferenceOTU120_IpB.KS_7317 | MxaB_Q93TX0_1KSB | 65 | 43 | 6.00E-61 | myxalamid | modular |
| New.CleanUp.ReferenceOTU120_IpB.KS_7317 | MxaB_Q93TX0_1KSB | 73 | 11 | 6.00E-61 | myxalamid | modular |
| New.CleanUp.ReferenceOTU122_IpB.KS_3399 | Sare1246_1 | 63 | 107 | 1.00E-45 | rifamycin | modular |
| New.CleanUp.ReferenceOTU122_IpB.KS_3399 | Sare1246_1 | 78 | 36 | 1.00E-45 | rifamycin | modular |
| New.CleanUp.ReferenceOTU125_IpB.KS_5475 | MtaE_Q9RFK7_1KSB | 59 | 209 | 3.00E-68 | myxothiazol | modular |
| New.CleanUp.ReferenceOTU133_IpB.KS_10209 | StiG_Q8RJY0_1KSB | 60 | 195 | 6.00E-61 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU136_IpB.KS_8219 | StiG_Q8RJY0_1KSB | 51 | 196 | 1.00E-42 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU14_IpB.KS_9834 | StiE_Q8RJY2_1KSB | 50 | 110 | 3.00E-37 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU14_IpB.KS_9834 | StiE_Q8RJY2_1KSB | 58 | 52 | 3.00E-37 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU14_IpB.KS_9834 | StiE_Q8RJY2_1KSB | 42 | 24 | 3.00E-37 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU141_IpB.KS_4989 | CurM_AAT70108_mod | 75 | 122 | 4.00E-63 | curacin | modular |
| New.CleanUp.ReferenceOTU141_IpB.KS_4989 | CurM_AAT70108_mod | 47 | 74 | 4.00E-63 | curacin | modular |
| New.CleanUp.ReferenceOTU15_IpB.KS_7033 | CurI_AAT70104_mod | 63 | 146 | 5.00E-53 | curacin | modular |
| New.CleanUp.ReferenceOTU15_IpB.KS_7033 | CurI_AAT70104_mod | 44 | 41 | 5.00E-53 | curacin | modular |
| New.CleanUp.ReferenceOTU151_IpB.KS_3662 | CALO5_12183629_i | 50 | 217 | 1.00E-53 | calicheamicin | iterative |
| New.CleanUp.ReferenceOTU154_IpB.KS_8928 | NosB_Q9RAH3_H | 73 | 123 | 4.00E-75 | nostopeptolide | hybrid KS |
| New.CleanUp.ReferenceOTU154_IpB.KS_8928 | NosB_Q9RAH3_H | 58 | 96 | 4.00E-75 | nostopeptolide | hybrid KS |
| New.CleanUp.ReferenceOTU155_IpB.KS_5402 | EpoD_Q9L8C7_3mod | 56 | 115 | 7.00E-55 | epothilone | modular |
| New.CleanUp.ReferenceOTU155_IpB. | EpoD_Q9L8C7_3m | 69 | 77 | 7.00E- | epothilone | modul |

| | | | | | | |
|---|---|---|---|---|---|---|
| KS_5402 | od | | | 55 | | ar |
| New.CleanUp.ReferenceOTU164_IpB.KS_8206 | JamL_AAS98783_mod | 59 | 180 | 4.00E-46 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU168_IpB.KS_4544 | MxaC_Q93TW9_3KSB | 55 | 110 | 2.00E-33 | myxalamid | modular |
| New.CleanUp.ReferenceOTU170_IpB.KS_6099 | StiB_Q8RJY5_1KSB | 50 | 105 | 2.00E-49 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU170_IpB.KS_6099 | StiB_Q8RJY5_1KSB | 57 | 81 | 2.00E-49 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU176_IpB.KS_8873 | JamK_AAS98782_mod | 50 | 216 | 2.00E-52 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU177_IpB.KS_5623 | StiG_Q8RJY0_1KSB | 59 | 195 | 1.00E-53 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU177_IpB.KS_5623 | StiG_Q8RJY0_1KSB | 73 | 15 | 1.00E-53 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU181_IpB.KS_2713 | SpnC_Q9ALM4_2KSB | 61 | 49 | 1.00E-32 | spinosad | modular |
| New.CleanUp.ReferenceOTU181_IpB.KS_2713 | SpnC_Q9ALM4_2KSB | 74 | 34 | 1.00E-32 | spinosad | modular |
| New.CleanUp.ReferenceOTU181_IpB.KS_2713 | SpnC_Q9ALM4_2KSB | 64 | 36 | 1.00E-32 | spinosad | modular |
| New.CleanUp.ReferenceOTU183_IpB.KS_8569 | CurJ_AAT70105_mod | 48 | 196 | 3.00E-33 | curacin | modular |
| New.CleanUp.ReferenceOTU184_IpB.KS_8894 | CurK_AAT70106_mod | 64 | 112 | 6.00E-44 | curacin | modular |
| New.CleanUp.ReferenceOTU184_IpB.KS_8894 | CurK_AAT70106_mod | 37 | 52 | 6.00E-44 | curacin | modular |
| New.CleanUp.ReferenceOTU186_IpB.KS_6682 | JamK_AAS98782_mod | 67 | 127 | 1.00E-60 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU186_IpB.KS_6682 | JamK_AAS98782_mod | 62 | 86 | 1.00E-60 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU187_IpB.KS_7576 | CurI_AAT70104_mod | 71 | 89 | 1.00E-53 | curacin | modular |
| New.CleanUp.ReferenceOTU187_IpB.KS_7576 | CurI_AAT70104_mod | 69 | 39 | 1.00E-53 | curacin | modular |
| New.CleanUp.ReferenceOTU187_IpB.KS_7576 | CurI_AAT70104_mod | 48 | 61 | 1.00E-53 | curacin | modular |
| New.CleanUp.ReferenceOTU187_IpB.KS_7576 | CurI_AAT70104_mod | 50 | 18 | 1.00E-53 | curacin | modular |
| New.CleanUp.ReferenceOTU188_IpB.KS_7041 | JamJ_AAS98781 | 51 | 108 | 2.00E-53 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU188_IpB.KS_7041 | JamJ_AAS98781 | 49 | 73 | 2.00E-53 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU188_IpB.KS_7041 | JamJ_AAS98781 | 62 | 40 | 2.00E-53 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU191_IpB. | LipA_ABB05102_1 | 54 | 71 | 1.00E- | lipomycin | modul |

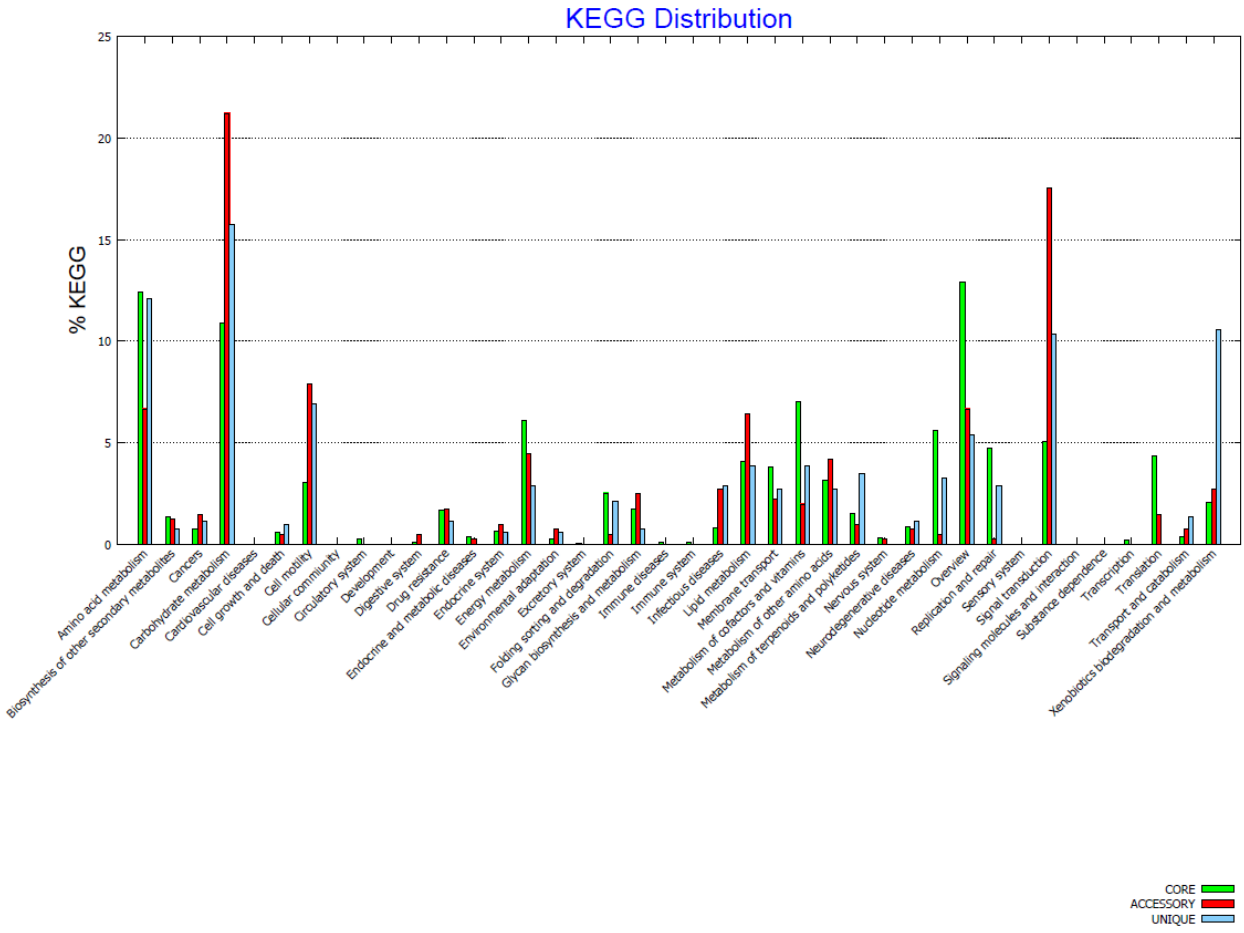
| | | | | | | |
|---|---|---|---|---|---|---|
| KS_7271 | KSB | | | 12 | | ar |
| New.CleanUp.ReferenceOTU20_IpB.KS_9626 | FurD2_ABB88522_KSB | 59 | 169 | 2.00E-55 | 5-alkenyl-3,3(2h)-furanone | modular |
| New.CleanUp.ReferenceOTU204_IpB.KS_7291 | CurI_AAT70104_mod | 71 | 85 | 4.00E-32 | curacin | modular |
| New.CleanUp.ReferenceOTU205_IpB.KS_7293 | EpoE_Q9L8C6_1mod | 49 | 94 | 3.00E-22 | epothilone | modular |
| New.CleanUp.ReferenceOTU205_IpB.KS_7293 | EpoE_Q9L8C6_1mod | 75 | 16 | 3.00E-22 | epothilone | modular |
| New.CleanUp.ReferenceOTU205_IpB.KS_7293 | EpoE_Q9L8C6_1mod | 36 | 191 | 2.00E-18 | epothilone | modular |
| New.CleanUp.ReferenceOTU206_IpB.KS_5133 | StiA_Q8RJY6_1KSB | 65 | 207 | 2.00E-71 | stigmatellin | KS1 |
| New.CleanUp.ReferenceOTU211_IpB.KS_6536 | CurA_AAT70096_mod | 68 | 209 | 2.00E-69 | curacin | KS |
| New.CleanUp.ReferenceOTU211_IpB.KS_6536 | CurA_AAT70096_mod | 73 | 11 | 2.00E-69 | curacin | KS |
| New.CleanUp.ReferenceOTU213_IpB.KS_2471 | PimS2_Q9EWA1_2KSB | 62 | 199 | 2.00E-65 | pimaricin | modular |
| New.CleanUp.ReferenceOTU215_IpB.KS_788 | JamM_AAS98784_H | 73 | 41 | 7.00E-14 | jamaicamide | hybrid KS |
| New.CleanUp.ReferenceOTU223_IpB.KS_9401 | StiG_Q8RJY0_1KSB | 66 | 121 | 1.00E-57 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU223_IpB.KS_9401 | StiG_Q8RJY0_1KSB | 59 | 54 | 1.00E-57 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU23_IpB.KS_8804 | StiI_Q8RJX8_1KSB | 44 | 224 | 4.00E-39 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU230_IpB.KS_8268 | MxaD_Q93TW8_1KSB | 60 | 127 | 9.00E-39 | myxalamid | modular |
| New.CleanUp.ReferenceOTU230_IpB.KS_8268 | MxaD_Q93TW8_1KSB | 91 | 11 | 9.00E-39 | myxalamid | modular |
| New.CleanUp.ReferenceOTU234_IpB.KS_4135 | JamL_AAS98783_mod | 66 | 80 | 7.00E-61 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU234_IpB.KS_4135 | JamL_AAS98783_mod | 64 | 67 | 7.00E-61 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU234_IpB.KS_4135 | JamL_AAS98783_mod | 40 | 63 | 7.00E-61 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU234_IpB.KS_4135 | JamL_AAS98783_mod | 82 | 22 | 7.00E-61 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU237_IpB.KS_5662 | CALO5_12183629_i | 55 | 196 | 4.00E-67 | calicheamicin | iterative |
| New.CleanUp.ReferenceOTU237_IpB.KS_5662 | CALO5_12183629_i | 71 | 31 | 4.00E-67 | calicheamicin | iterative |
| New.CleanUp.ReferenceOTU239_IpB.KS_7627 | MxaB_Q93TX0_1KSB | 67 | 144 | 1.00E-61 | myxalamid | modular |
| New.CleanUp.ReferenceOTU239_IpB. | MxaB_Q93TX0_1KS | 55 | 38 | 1.00E- | myxalamid | modul |

| | | | | | | |
|---|---|---|---|---|---|---|
| KS_7627 | B | | | 61 | | ar |
| New.CleanUp.ReferenceOTU239_IpB.KS_7627 | MxaB_Q93TX0_1KSB | 57 | 21 | 1.00E-61 | myxalamid | modular |
| New.CleanUp.ReferenceOTU24_IpB.KS_9898 | AveA2_Q9S0R7_4mod | 69 | 157 | 1.00E-57 | avermectin | modular |
| New.CleanUp.ReferenceOTU241_IpB.KS_8150 | EcoA_AAX98184_2KSB | 49 | 220 | 2.00E-48 | eco-02301 | modular |
| New.CleanUp.ReferenceOTU243_IpB.KS_7133 | JamK_AAS98782_mod | 49 | 72 | 1.00E-43 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU243_IpB.KS_7133 | JamK_AAS98782_mod | 65 | 46 | 1.00E-43 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU243_IpB.KS_7133 | JamK_AAS98782_mod | 58 | 57 | 1.00E-43 | jamaicamide | modular |
| New.CleanUp.ReferenceOTU248_IpB.KS_3755 | EpoE_Q9L8C6_1mod | 64 | 112 | 3.00E-58 | epothilone | modular |
| New.CleanUp.ReferenceOTU248_IpB.KS_3755 | EpoE_Q9L8C6_1mod | 57 | 49 | 3.00E-58 | epothilone | modular |
| New.CleanUp.ReferenceOTU248_IpB.KS_3755 | EpoE_Q9L8C6_1mod | 52 | 46 | 3.00E-58 | epothilone | modular |
| New.CleanUp.ReferenceOTU251_IpB.KS_9798 | CurI_AAT70104_mod | 59 | 161 | 4.00E-55 | curacin | modular |
| New.CleanUp.ReferenceOTU251_IpB.KS_9798 | CurI_AAT70104_mod | 73 | 30 | 4.00E-55 | curacin | modular |
| New.CleanUp.ReferenceOTU254_IpB.KS_5431 | StiC_Q8RJY4_1KSB | 64 | 173 | 2.00E-56 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU258_IpB.KS_8813 | MxaB_Q93TX0_1KSB | 69 | 80 | 3.00E-62 | myxalamid | modular |
| New.CleanUp.ReferenceOTU258_IpB.KS_8813 | MxaB_Q93TX0_1KSB | 61 | 90 | 3.00E-62 | myxalamid | modular |
| New.CleanUp.ReferenceOTU258_IpB.KS_8813 | MxaB_Q93TX0_1KSB | 43 | 46 | 3.00E-62 | myxalamid | modular |
| New.CleanUp.ReferenceOTU259_IpB.KS_9796 | ChlB1_AAZ77673_i | 65 | 139 | 6.00E-60 | chlorothricin | iterative |
| New.CleanUp.ReferenceOTU259_IpB.KS_9796 | ChlB1_AAZ77673_i | 66 | 41 | 6.00E-60 | chlorothricin | iterative |
| New.CleanUp.ReferenceOTU27_IpB.KS_5488 | LnmJ_AF484556_2T | 47 | 135 | 1.00E-31 | leinamycin | trans |
| New.CleanUp.ReferenceOTU28_IpB.KS_8803 | LnmI_AF484556_2T | 64 | 108 | 4.00E-29 | leinamycin | trans |
| New.CleanUp.ReferenceOTU28_IpB.KS_8803 | LnmI_AF484556_2T | 64 | 14 | 4.00E-29 | leinamycin | trans |
| New.CleanUp.ReferenceOTU38_IpB.KS_9035 | LipD_ABB05105_2KSB | 61 | 71 | 1.00E-35 | lipomycin | modular |
| New.CleanUp.ReferenceOTU38_IpB.KS_9035 | LipD_ABB05105_2KSB | 79 | 58 | 1.00E-35 | lipomycin | modular |
| New.CleanUp.ReferenceOTU4_IpB.KS | EpoC_Q9L8C8_H | 64 | 141 | 1.00E- | epothilone | hybrid |

| | | | | | | |
|---|---|---|---|---|---|---|
| _8418 | | | | 47 | | KS |
| New.CleanUp.ReferenceOTU45_IpB.KS_5599 | Sare1250_2 | 51 | 102 | 4.00E-23 | rifamycin | modular |
| New.CleanUp.ReferenceOTU46_IpB.KS_5452 | EpoD_Q9L8C7_4mod | 66 | 219 | 7.00E-81 | epothilone | modular |
| New.CleanUp.ReferenceOTU47_IpB.KS_9400 | EpoD_Q9L8C7_4mod | 69 | 48 | 3.00E-16 | epothilone | modular |
| New.CleanUp.ReferenceOTU48_IpB.KS_5455 | JamE_AAS98777_KS1 | 57 | 142 | 1.00E-61 | jamaicamide | KS |
| New.CleanUp.ReferenceOTU48_IpB.KS_5455 | JamE_AAS98777_KS1 | 61 | 62 | 1.00E-61 | jamaicamide | KS |
| New.CleanUp.ReferenceOTU5_IpB.KS_9568 | NosB_Q9RAH3_H | 51 | 193 | 1.00E-48 | nostopeptolide | hybrid KS |
| New.CleanUp.ReferenceOTU52_IpB.KS_6840 | StiE_Q8RJY2_1KSB | 50 | 181 | 1.00E-46 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU52_IpB.KS_6840 | StiE_Q8RJY2_1KSB | 55 | 29 | 1.00E-46 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU6_IpB.KS_4876 | EpoD_Q9L8C7_4mod | 66 | 192 | 2.00E-69 | epothilone | modular |
| New.CleanUp.ReferenceOTU62_IpB.KS_9351 | JamE_AAS98777_KS1 | 47 | 225 | 7.00E-46 | jamaicamide | KS |
| New.CleanUp.ReferenceOTU70_IpB.KS_7847 | ChlB1_AAZ77673_i | 56 | 153 | 2.00E-49 | chlorothricin | iterative |
| New.CleanUp.ReferenceOTU70_IpB.KS_7847 | ChlB1_AAZ77673_i | 77 | 35 | 2.00E-49 | chlorothricin | iterative |
| New.CleanUp.ReferenceOTU71_IpB.KS_8182 | StiC_Q8RJY4_1KSB | 69 | 144 | 4.00E-57 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU71_IpB.KS_8182 | StiC_Q8RJY4_1KSB | 77 | 13 | 4.00E-57 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU78_IpB.KS_8907 | StiG_Q8RJY0_1KSB | 46 | 151 | 4.00E-30 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU78_IpB.KS_8907 | StiG_Q8RJY0_1KSB | 39 | 150 | 1.00E-19 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU79_IpB.KS_5779 | StiD_Q8RJY3_1KSB | 51 | 126 | 7.00E-35 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU79_IpB.KS_5779 | StiD_Q8RJY3_1KSB | 60 | 42 | 7.00E-35 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU79_IpB.KS_5779 | StiD_Q8RJY3_1KSB | 38 | 130 | 5.00E-15 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU81_IpB.KS_5588 | MxaF_Q93TW6_1KSB | 66 | 82 | 3.00E-41 | myxalamid | KS1 |
| New.CleanUp.ReferenceOTU81_IpB.KS_5588 | MxaF_Q93TW6_1KSB | 71 | 62 | 3.00E-41 | myxalamid | KS1 |
| New.CleanUp.ReferenceOTU86_IpB.KS_6372 | CurI_AAT70104_mod | 69 | 189 | 7.00E-65 | curacin | modular |
| New.CleanUp.ReferenceOTU89_IpB.K | EpoD_Q9L8C7_4m | 73 | 164 | 1.00E- | epothilone | modul |

| | | | | | | |
|---|---|---|---|---|---|---|
| S_7279 | od | | | 70 | | ar |
| New.CleanUp.ReferenceOTU89_IpB.KS_7279 | EpoD_Q9L8C7_4mod | 70 | 20 | 1.00E-70 | epothilone | modular |
| New.CleanUp.ReferenceOTU92_IpB.KS_5346 | StiG_Q8RJY0_1KSB | 45 | 220 | 3.00E-43 | stigmatellin | modular |
| New.CleanUp.ReferenceOTU97_IpB.KS_3095 | CurA_AAT70096_mod | 60 | 93 | 6.00E-52 | curacin | KS |
| New.CleanUp.ReferenceOTU97_IpB.KS_3095 | CurA_AAT70096_mod | 64 | 45 | 6.00E-52 | curacin | KS |
| New.CleanUp.ReferenceOTU97_IpB.KS_3095 | CurA_AAT70096_mod | 63 | 41 | 6.00E-52 | curacin | KS |
| New.CleanUp.ReferenceOTU97_IpB.KS_3095 | CurA_AAT70096_mod | 67 | 27 | 6.00E-52 | curacin | KS |

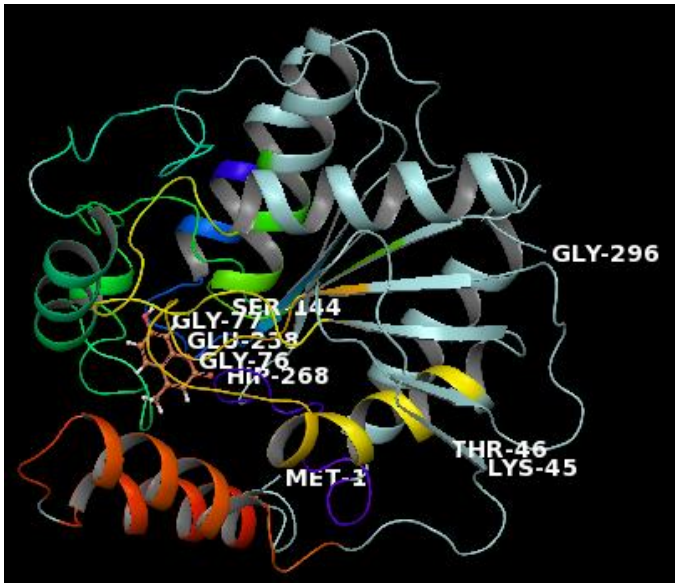
## 7.2 Supplementary material Chapter 3



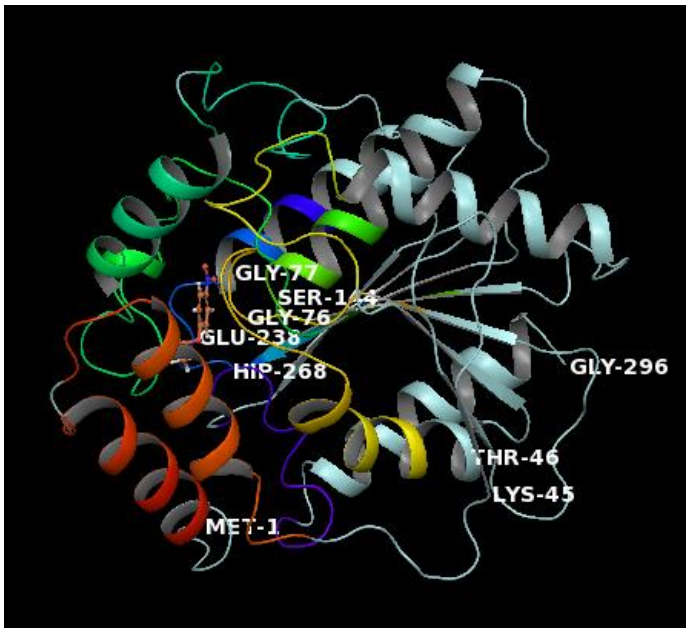

**Supplementary figure 1**: KEGG distribution of the proteins identified by BPGA from the five investigated *Pseudoalteromonas* sp. genomes.

## 7.3 Supplementary material Chapter 4

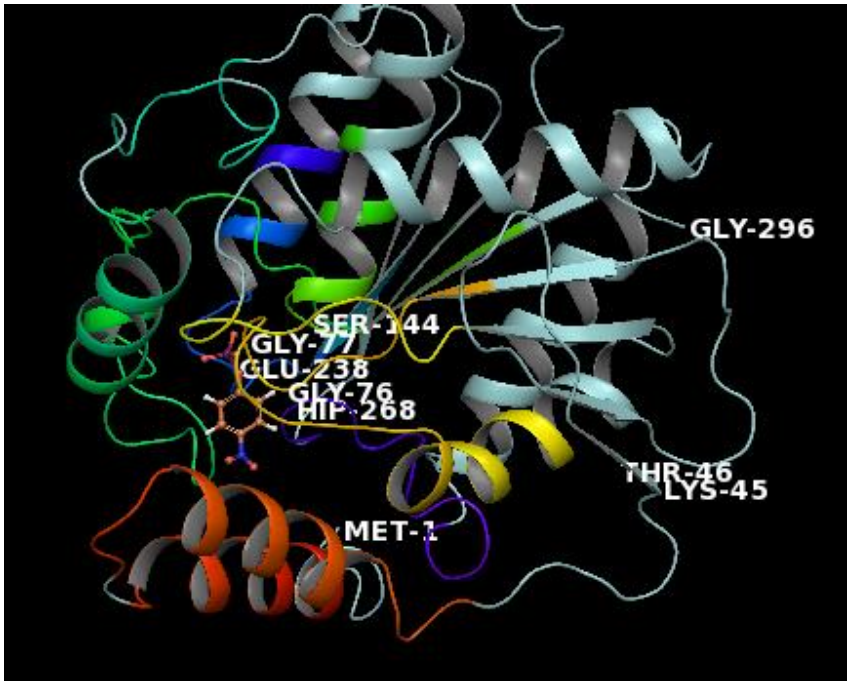
### 7.3.1 3D binding models: Substrates



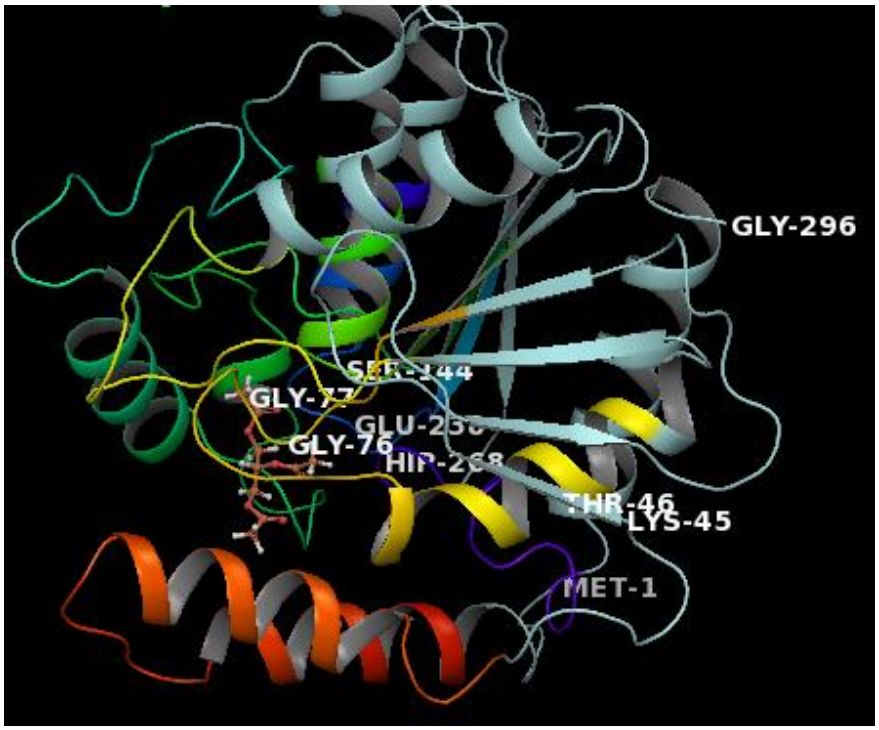

**4-Methylumbelliferone**



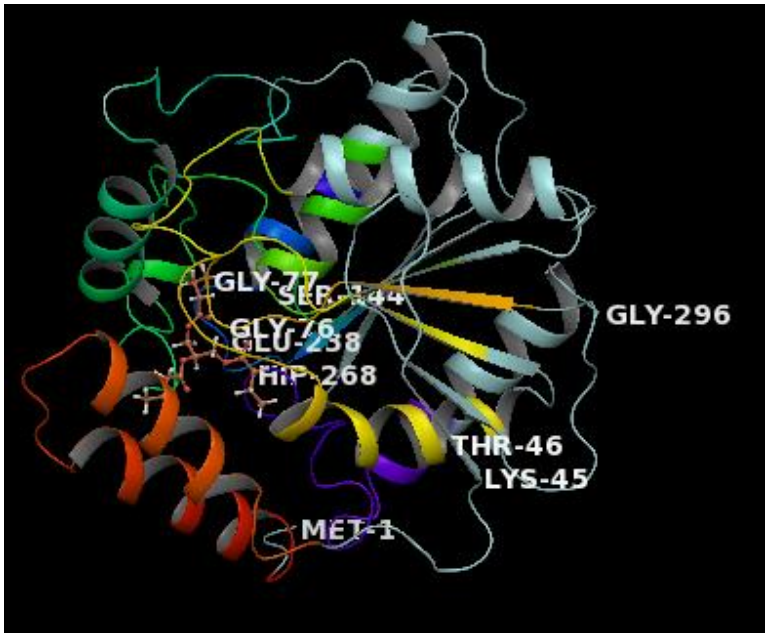

**4-Nitrophenyl acetate**



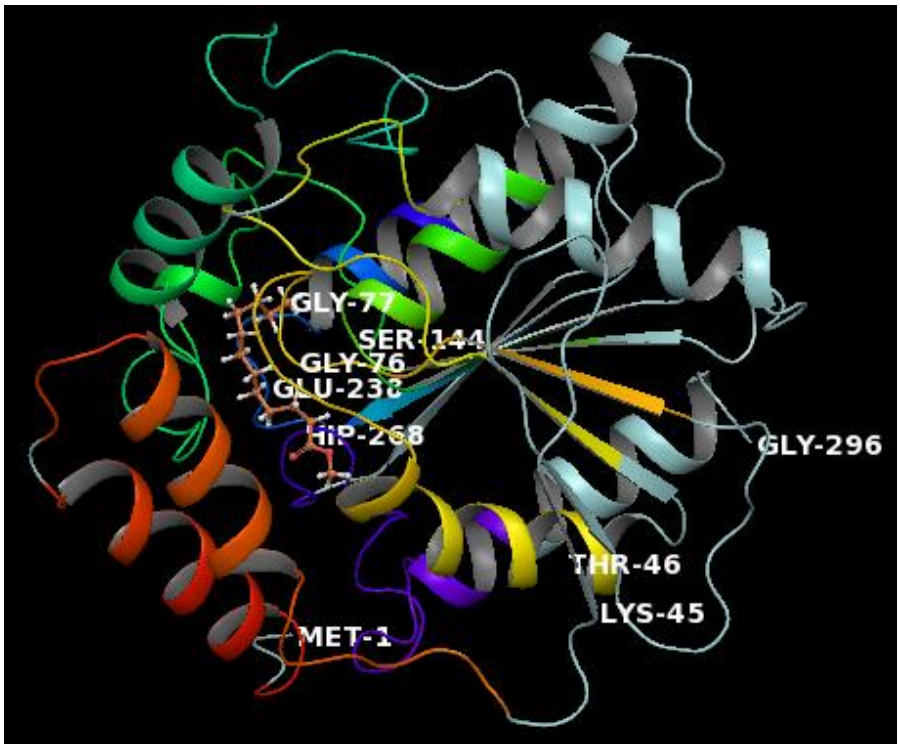

**4-Nitrophenyl phosphate**




**Triacetin**

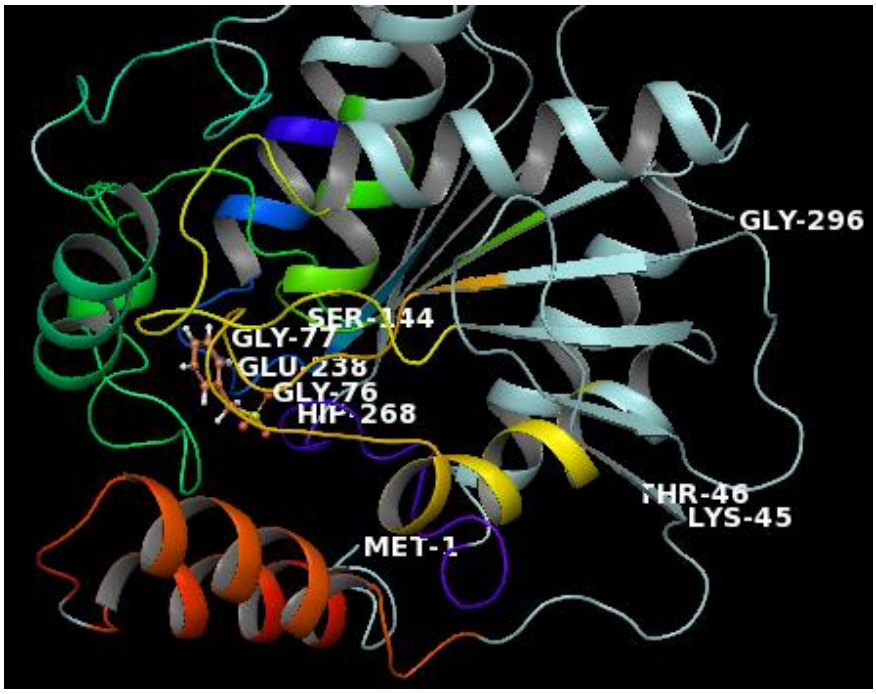



**Tributyrin**

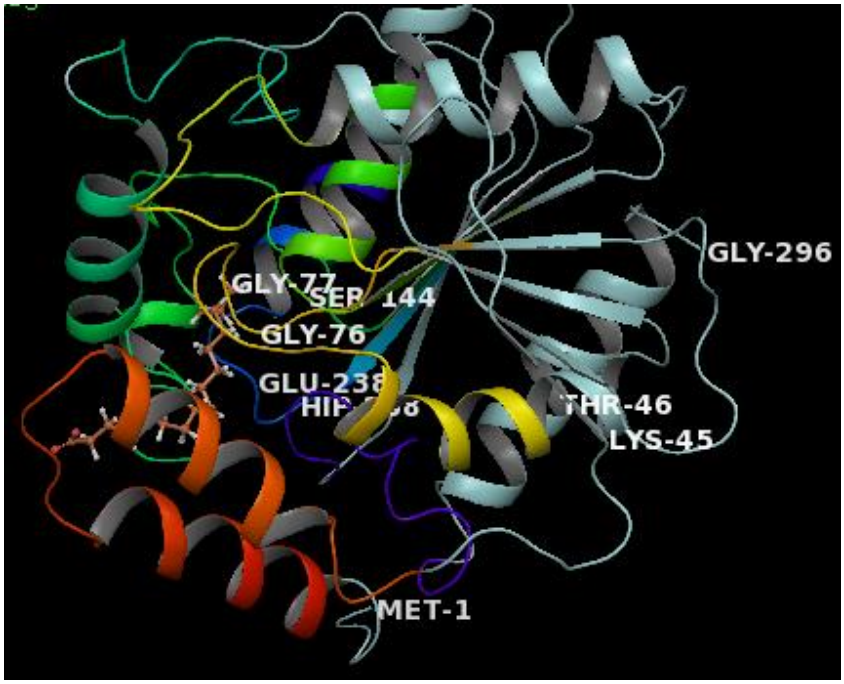



**Methyl laurate**

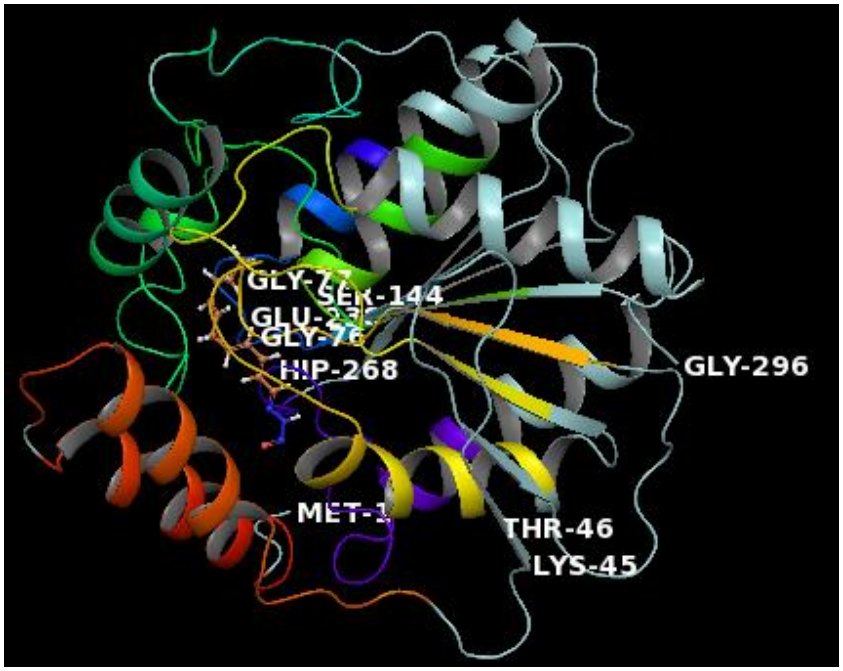
### 7.3.2 3D binding models: Inhibitors



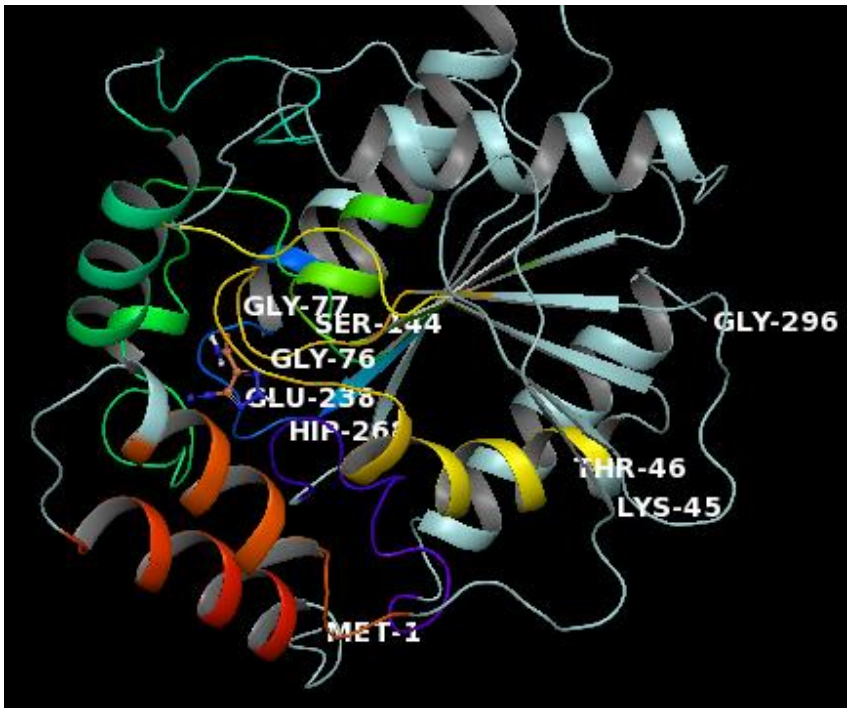

**Phenylmethansulfonic acid**



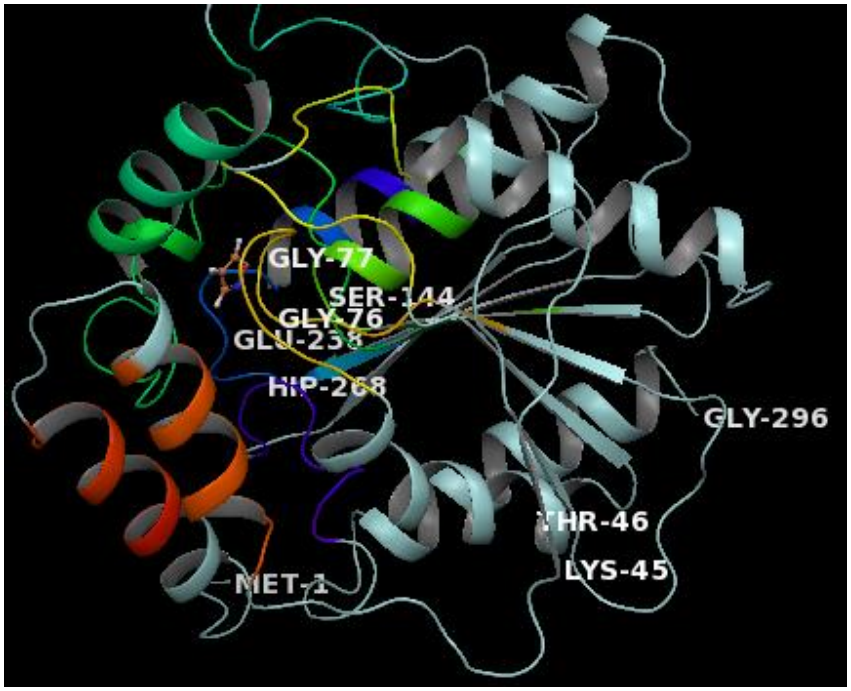

**Oleic acid**




**Triacsin C**




**5-Carbamoyl-2H-1,2,3-triazole-4-diazonium**




**Isoxazole**