Discovery of biomolecules from South African marine sponge symbionts through Metagenomic studies

Matlakala Reginah Mapatha

A thesis submitted in partial fulfilment of the requirements for the degree of
MAGISTER SCIENTIAE (M.Sc.)
Department of Biotechnology,
University of the Western Cape
Bellville

Supervisor: Prof. Marla Trindade

(Date)

Declaration

I, Matlakala Reginah Mapatha, hereby declare that "**Discovery of biomolecules from South African marine sponge symbionts through Metagenomics studies**" is my own work, that it has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete reference.

Date: …………………………….

Signed: ………………………….

Abstract

Marine sponges are known to harbour diverse microorganisms which have been proposed to be involved in the synthesis of most of the natural compounds derived from sponges. However, only about 1% of these microorganisms can currently be identified through culture-based approaches, limiting the discovery of novel natural compounds. To overcome this limitation, a metagenomics approach has been introduced for the discovery of natural compounds from symbiotic bacterial metagenomic DNA.

Natural products are low molecular weight compounds produced by a range of living organisms that may have pharmaceutical applications such as antibiotics, anticancer, anti-parasitic and anti-fungal agents. In this study the focus was on natural products that are commonly synthesized by nonribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs). NRPS and PKS encode two families of secondary metabolite enzymes from microorganisms that are of great interest to the pharmaceutical industry. Bacterial genes encoding biosynthetic enzymes of these systems are usually arranged in clusters, which simplifies their cloning and transfer into a heterologous host.

The aim of this study was to discover and identify biosynthetic pathways for novel natural products from symbionts from two South African marine sponges through metagenomic studies. A metagenomic fosmid library that was created in *Escherichia coli* EPI300 from marine sponge DNA was screened for the detection of the antibacterial activity against various pathogenic bacterial strains using agar overlay assays. The library was also screened using the PCR-based approach with degenerate primers for the isolation of the genes or gene clusters that encode for the production of the bioactive secondary metabolites such as nonribosomal peptides and polyketides.

Abstract

The PCR-based screening of the fosmid library resulted in the detection and identification of a gene cluster encoding for the Type I polyketide synthase. A biosynthetic pathway was proposed based on the annotation and sequence analysis of the gene cluster using antiSMASH 3.0. A 41.6kb gene cluster fragment containing a short PKS module was isolated from the marine sponge *Higginsia bidentifera* metagenomic library. The gene cluster consisted of 33 open reading frames (ORFs), with ORF19 encoding a PKS gene that consisted of KS-AT-DH domains, which showed 97% sequence similarity to that of a marine bacterium *Cobetia marina*. The detection of a PKS-like pathway from the sponge metagenomic library indicates the importance of using sequence-based screening for the discovery natural product biosynthetic pathways from uncultivated bacteria.

Key words: *metagenomics, functional screening, sequence-based screening, marine sponges, marine natural products, nonribosomal peptide synthases, polyketide synthases*.

UNIVERSITY *of the*
WESTERN CAPE
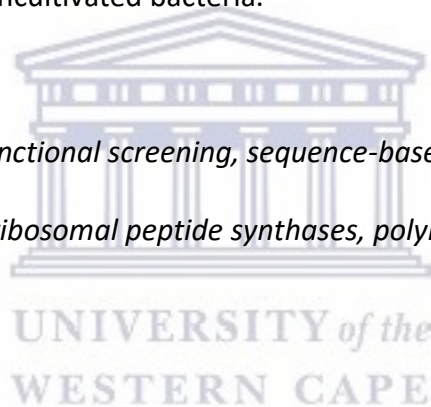
Acknowledgments

First, I would like to thank my supervisor Prof. Marla Trindade (Tuffin) for the opportunity she has given me to perform this research project within her laboratory and for allowing me to be part of the PharmaSea Research group at IMBM. I also thank her for reading and editing my thesis.

I would also like to thank Mr Lonnie van Zyl for sharing his scientific knowledge and the practical experience. And for providing the metagenomic library without which the research would not have been possible.

I thank Pepe and Marco Bonanno for the suggestions, sharing their practical experience, their support and for the encouragement.

I thank all the IMBM members for the support, the laughter, the entertainment and for keeping me company during those long hours.

I am especially thankful to my parents, Mr Lazarus and Mrs Grace Mapatha, and all my siblings for their prayers, their unconditional love and unending support and encouragement. This one is for you guys.

Dedications

This work is dedicated to my late cousin, Lukhetho Simon Ramaru. May his beautiful soul rest in eternal peace.

Table of Contents

Table of Contents

Table of Contents

Table of Contents

UNIVERSITY of the
WESTERN CAPE

http://etd.uwc.ac.za

Table of Figures

x

Table of Figures

List of Tables

Abbreviations

°C	Degree Celsius
µF	Microfarad
A domain	Adenylation domain
ACP	Acyl carrier protein
Amp	Ampicillin
antiSMASH	Antibiotics and secondary metabolites analysis shell
AT	Acyltransferase
ATCC	American type culture collection
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosome
BLAST	Basic Local Alignment Search Tool
bp	base pair
C domain	Condensation domain
CAM	Chloramphenicol
cm	Centimetre
CoA	Co-Enzyme A
contig	Contiguous
DEBS	6-deoxyerythronolide B synthase
DH domain	Dehydratase domain
DMSO	Dimethylsulfoxide
DNA	Deoxyribose nucleic acid
dNTPs	Deoxynucleotide triphosphates
eDNA	Environmental DNA
EDTA	Ethylenediaminetetracitic acid
ER	Enoylreductase
FAS	Fatty acid synthase
g	Gravitational force
HMW	High molecular weight
IMBM	Institute for Microbial Biotechnology and Metagenomics

Abbreviations

IPTG	Isopropyl- β-D-thiogalactopyranoside
kb	Kilobase pairs
KCl	Potassium chloride
KR	Keto Reductase
KS	Ketosynthase
KS_α	Ketosynthase alpha subunit
KS_β	Ketosynthase beta subunit
LB	Lysogeny broth
$MgCl_2$	Magnesium chloride
min	Minutes
ml	Millilitre
mlibrary	Metagenomic library
mM	Millimolar
MRSA	Methicillin resistant *Staphylococcus aureus*
NaCl	Sodium Chloride
NADH	Nicotinamide Adenine Dinucleotide, reduced
NCBI	National Centre for Biotechnology Information
ng/µl	Nanogram per microliter
NP	Natural product
NRP	Nonribosomal peptide
NRPS	Nonribosomal peptide synthase
OD	Optical density
ORF	Open Reading Frame
PCP	peptidyl carrier protein
PCR	Polymerase Chain Reaction
PK	Polyketide
PKS	Polyketide synthase
PRISM	PRediction Informatics for Secondary Metabolomes
RNA	Ribo nucleic acid
RNAse	Ribonuclease
rpm	Revolutions per minute

Abbreviations

rRNA	Ribosomal ribonucleic acid
SCUBA	Self-contained underwater breathing apparatus
SDS	Sodium Dodecyl sulphate
sec	Seconds
T domain	Thiolation domain
TAE	Tris-Acetate-EDTA buffer
TAR	Transformation associated recombination
TE	Tris-EDTA buffer
TE domain	Thioesterase domain
THN	Tetrahydronaphthalene
U	Unit
UV	Ultraviolet light
v/v	Volume per volume
w/v	Weight per volume
WHO	World Health Organisation
x-gal	5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside
μg	Microgram
μg/ml	Microgram per millilitre
μl	Microliter
μm	Micrometre
μM	Micromolar

Chapter 1

Literature Review

1.1 Introduction

Infectious disease continues to be one of the main causes of death worldwide accounting for more than half of all the fatalities, and the mortality rate seems to be increasing in the developed countries as well (Abdelmohsen et al., 2010). This calls for the development of novel and improved strategies for the treatment of these infectious diseases with minimised or no side effects to humans. There is a serious need for the discovery of novel and improved classes of antibacterial compounds with new modes of action, particularly for the treatment of infectious diseases caused by multiple resistant pathogenic microorganisms against the antibiotic agents that are already in clinical use, including the new generation of antibiotics (Xiong et al., 2013; Zhang et al., 2005). The emergence of multi-drug resistant pathogens creates a major challenge in the treatment of infectious diseases worldwide. The resistance can be due to the survival mechanisms developed by these pathogenic bacteria, which decreases the effectiveness of most antibiotics. Examples of multi-drug resistant pathogens include *Enterococcus faecium*, methicillin-resistant *Staphylococcus aureus* (MRSA), *Klebsiella pneumoniae*, *Bacillus anthracis*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa,* *Enterobacter* spp. and *Mycobacterium tuberculosis* (Genilloud, 2014; Gomes et al., 2013).

1

These pathogenic microorganisms continue to be a threat to human health since there has been a decline in antibiotic drug discovery, both from synthetic and natural sources.

1.1.1 Background on natural products

Natural products (NPs) are small low molecular weight chemical compounds that are produced by living environmental organisms in nature, such as animals, plants, fungi and bacteria (Bérdy, 2005). Microorganisms, which are ubiquitous in every habitat, are known to be an ideal source of natural products, as they can grow rapidly and can produce NPs in large scale. Microbial natural products have been an important source for the discovery of novel antibiotics for the past few decades and their analogs are introduced in the pharmaceutical industry and continue to be used for clinical purposes (Genilloud, 2014). Their biological activity, due to their enormous chemical diversity, has provided important leads for drug discovery and development, accounting for over 60% of the total market (Singh & Macdonald, 2010). These compounds normally do not directly play an essential role in the primary metabolism of the microorganism, such as normal growth and development of the cells, as they are mainly produced as secondary metabolites during the late stages of growth when there are low nutrient concentrations (Ridley & Khosla, 2009). Secondary metabolites offer a complementary function as a chemical defence mechanism against competitors for scarce nutrients and pathogens in the environment, and for the reproductive fitness of the organisms that produce them (Baker et al., 2007; Lorente et al., 2014). They possess a wide variety of structural diversity that varies among different species that produce them.

2

Microbial NPs used as pharmaceutical drugs include a wide range of bioactive compounds such as non-ribosomal peptides, polyketides, lantibiotics, siderophores and bacteriocins, amongst others. The golden era of natural compounds-driven medicine began when Alexander Fleming discovered the first antibiotic, penicillin, from the fungus *Penicillium rubens* in 1929 (Fleming, 1929). Thereafter, more antibiotics such as streptomycin, tetracycline and others were discovered for the treatment of bacterial and fungal infections. Most of these antibiotic classes were produced before 1970 and thereafter the new classes that are in clinical use were based on the chemical modifications of the existing ones (Genilloud, 2014).

NPs from easily cultured soil-inhabiting microorganisms such as *Actinomycetes* and *Bacilli* have been used for many decades as a source of many pharmaceutical agents for the treatment of human diseases, such as streptomycin, amphotericin B and erythromycin that can be used for the treatment of human diseases, such as bacterial infections, various cancers and fungal infections among others (Zhang et al., 2011; Handelsman et al., 1998; Ridley & Khosla, 2009). However, they have fallen out of favour since the traditional screening of soil microbes often results in rediscovering antibiotic compounds that are already known (Thaker et al., 2013). The search for novel metabolites has now shifted towards the oceans, which covers more than 70% of the earth's surface area (Rocha-Martin et al., 2014). The marine environment remains an untapped source of novel bacteria as well as novel bioactive compounds (Dharmaraj, 2010). The natural products from marine organisms are different from those of terrestrial organisms and are structurally diverse. The microorganisms from the marine environment are likely to possess diverse and unique biochemical and physiological characteristics as they live in a biologically competitive environment with extreme conditions such as alkaline pH, high temperature, high

3

pressure, low oxygen, toxic metals, no sunlight and low nutrient availability (Zhang et al., 2005). Most of the marine natural compounds have been isolated from microorganisms associated with marine invertebrates such as sponges, tunicates, bryozoans and molluscs.

Natural product discovery involves screening, extraction and identification of the compounds, and more recently is paralleled with genomic techniques to identify the biosynthetic pathways responsible for their synthesis (Monaco & Quinlan, 2014).

1.1.2 Marine sponges as a source of natural products

Marine sponges, of the phylum *Porifera* (meaning pore-bearing), are multicellular invertebrates with a simple body structure that lack a nervous system and have no muscle cells, but consist of different amoeboid cells that move freely in the matrix (Hochmuth & Piel, 2009a). These organisms are considered to be the most primitive of the metazoans of about 600 million years and can occupy up to 80% of the available surfaces in the polar regions (Hentschel et al., 2012). Marine sponges are also globally distributed in marine environments including tropical, polar and temperate oceans (Hentschel et al., 2012). Marine sponges are divided into four classes according to the nature of their skeletal spicules; *Hexactinellida* or glass sponges (Figure 1A) that are found in deep marine habitats, the *Demospongia* (Figure 1B) or demosponges which are the largest group found in marine environments and freshwater, the *Calcarea* (Figure 1C) or calcareous sponges that possesses calcium carbonate spicules and the recently recognised *Homoscleromorpha* which are encrusting sponges (Figure 1D) (Gazave et al., 2010).

4
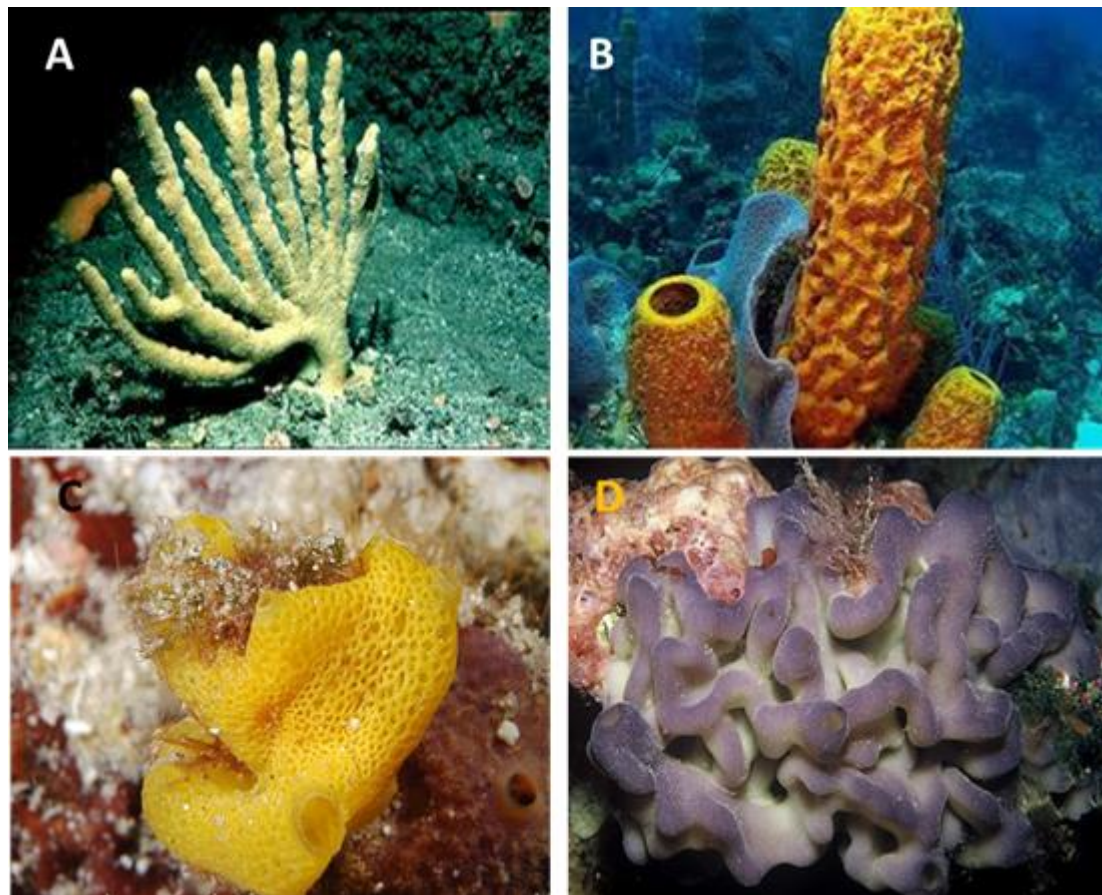
Figure 1.1: Examples of the four different classes of marine sponges. *Hexacteinella* or glass sponges (A), the *Demospongia* or demosponges (B), the *Calcarea* (C) and the *Homoscleromorpha* (D) (Taken from www.reproductivelearning.weebly.com/porifera.html and www.deepseaimages.com/dsilibrary. Viewed 29 August 2016 and http://www.place-hecataei-milesii.net/classification2/homoscleromorphes.html. Viewed 16 March 2018).

Marine sponges are sessile and filter-feeding animals that filter large amounts of water through their numerous tiny pores. Sea water containing bacteria, microalgae and other organic particles are drawn through the pores into the mesohyl layer of the sponge. These sponges absorb nutrients and remove a large percentage of microorganisms and other

5

particulate matter from the aqueous environment by pumping large volumes of seawater through their aquiferous system situated in the mesohyl during feeding (Kennedy et al., 2007). This makes marine sponges a host to a large number of diverse microorganisms in their tissues that can be as high as 40-60 % of the total sponge biomass and the bacterial population may reach up to 10^8-10^{10} cells per gram of the sponge wet weight (Anand et al., 2006; Ouyang et al., 2010). These can include bacteria, cyanobacteria, fungi and eukaryotic algae that can be associated extra- or intracellularly (Long et al., 2005). The mesohyl is an extracellular matrix populated by sponge cells and constitutes most of the sponge body (Uria & Piel, 2009). Some of the bacterial cells are ingested by the sponges' archaeocytes where they are digested through phagocytosis as food particles for the metabolic requirements of sponges, while some bacterial cells are able to resist the digestion in the mesohyl layer (Figure 1.2). Some of these bacterial cells which are housed in the mesohyl can be established as part of the sponge-specific symbionts, where they provide organic nutrients and chemical defences against eukaryotic predators, play a role in waste removal and stabilization of the sponge skeleton as well as chemically defending the sponges against microbial infection (Skariyachan et al., 2013). These symbiotic microorganisms benefit from marine sponges as they get a stable supply of nutrients, a habitat to thrive in and a supply of the scarce element nitrogen that is excreted as an end product in the form of ammonia by the marine sponges (Hentschel et al., 2012). Some of the bacterial cells can resist digestion in the mesohyl and become pathogenic to the marine sponges. This can be due to their ability to shield themselves from being recognised by the host sponge phagocytic cells by the production of slime capsules (Wilkinson et al., 1984).
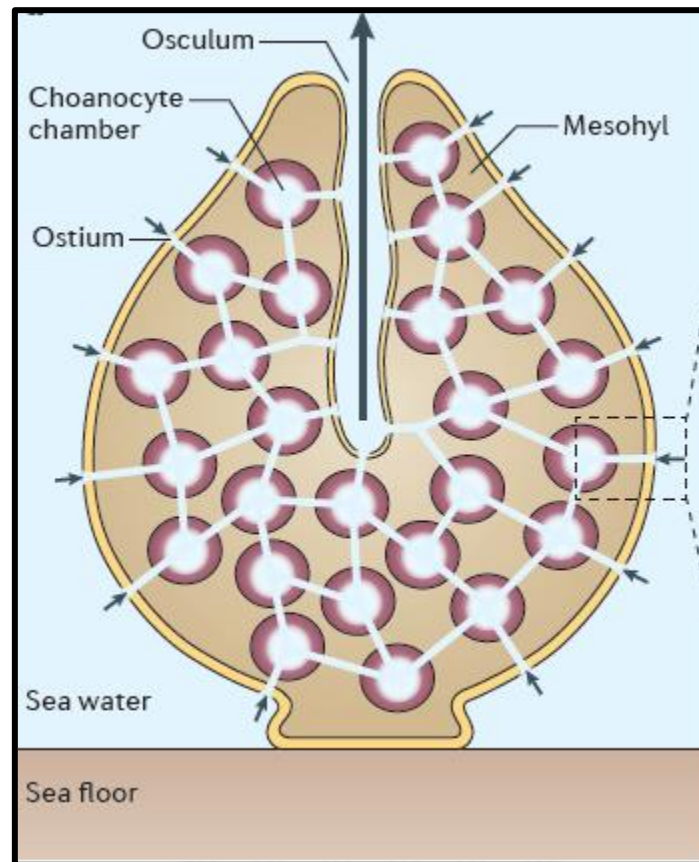
Figure 1.2: The schematic structure of a simple body plan of a marine sponge; where seawater enters the sponge cells through the ostia. The water is filtered to remove the microorganisms in the mesohyl where phagocytosis also takes place, and discharged from the mesohyl through the osculum (Adapted from Hentschel et al., 2012).

Because of their lack of physical defence mechanisms, marine sponges are exposed to an enormous pressure which has caused the evolution of structurally diverse chemical defences (Pimentel-Elardo et al., 2012). These biologically active compounds can play an important role in protecting the organisms from competitors, predators and pathogens in the environment. This makes marine sponges a very rich source of potentially valuable complex biologically active natural products with biotechnological interest as well as for drug discovery and

7

development (Hentschel et al., 2006). But most of these natural compounds cannot be further developed as marine sponges are difficult to collect and their collection cannot be maintained sustainably as it may lead to extinction.

There is evidence that the microbial symbionts associated with marine invertebrates such as sponges, tunicates and bryozoans are likely to be the true producers of many of the metabolites already identified as they are structurally related to secondary metabolites of microbial origin, which are very rare in animals. For example, Davidson et al., 2001 did a study to show that the bacterial symbiont *Endobugula sertula* of the bryozoan *Bugula neritina* was the true producer of the bryostatin isolated from this marine invertebrate (Davidson et al., 2001). Consequently, bryostatin is used in combination with other drugs, for the treatment of various cancers such as pancreatic and renal cancer, leukemia and melanoma (Sudek et al., 2007).
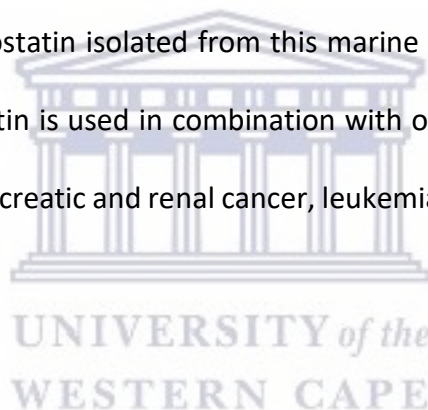
The majority of the secondary metabolites which are produced by marine invertebrate symbiotic bacteria are complex polyketides, nonribosomal peptides or polyketide-nonribosomal peptide hybrids (Piel 2004; Lane & Moore 2012). These natural metabolites may have antimicrobial activities against bacteria and fungi; and they can also be utilised as antitumor drugs, immunosuppressants, enzyme inhibitors, antiviral, antimalarial, anti-inflammatorials, antifouling, toxins and siderophores (Pimentel-Elardo et al., 2012; Schirmer et al., 2005). However, many marine sponge symbionts have not yet been studied due to the difficulties in cultivating them under normal laboratory conditions. This could be due to the complex nutritional requirements of these microorganisms, including the necessity to be co-

cultured. Due to these limitations, the development of metagenomics was introduced as an alternative strategy in order to isolate biosynthetic genes from these complex microorganisms without having to cultivate them (discussed in more details in Section 1.3).

1.1.3 Marine microbes associated with marine sponges

Studies using phylogenetic analysis of 16S rRNA genes have revealed that Alphaproteobacteria, Gammaproteobacteria, Actinobacteria, Cyanobacteria, Betaproteobacteria and Firmicutes are the most dominant bacterial phyla associated with marine sponges (Kim & Fuerst, 2006; Abe et al., 2012). Culture-independent molecular approaches revealed that actinobacteria are the most abundant microorganisms that are associated with marine invertebrates, of which the majority are known to produce bioactive NPs (Hentschel et al., 2002). For example, phylogenetic analysis of a 16S rRNA gene library showed that *Streptomyces* was the most abundant genus associated with marine sponge *Haliclona* spp. and also produce the largest number of natural products (Figure 1.3) (Khan et al., 2011; Valliappan et al., 2014). Another example is an analysis of 16S rRNA sequences that showed the presence of a novel *Actinomycete* genus *Salinispora* that was found to be associated with a marine sponge *Pseudoceratina clavata* (Kim et al., 2005). They produce a wide variety of secondary metabolites to help them survive the competition against other organisms for resources and environmental pressures (Singh & Macdonald, 2010). Microorganisms are an ideal source of natural products as they have a simple genome compared to plants and animals, making it easy to manipulate genes responsible for the production of secondary metabolites.

9
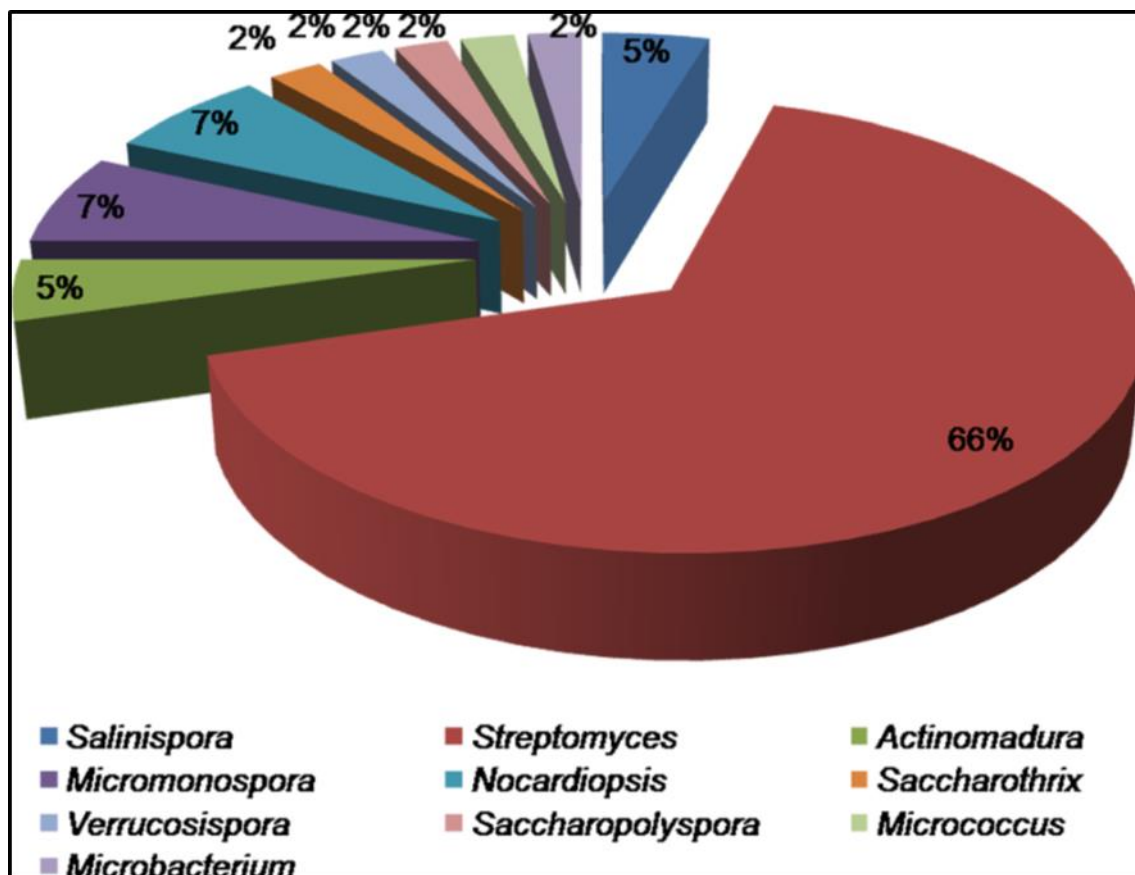
Figure 1.3: The distribution of marine invertebrate-associated Actinobacteria which are known to produce natural products. *Streptomyces* produces the largest number of natural products (66%), followed by *Micromonospora* and *Nocardiopsis* both at 7%, and *Salinispora* and *Actinomadura* both at 5%. The rest form a small part (2%) of the natural product producers (Taken from Valliappan et al., 2014).

Other marine bacterial genera such as *Pseudomonas* and *Bacillus* are also known to produce natural products (Singh & Macdonald, 2010). Marine *Bacillus* are known to produce a wide range of secondary metabolites which include polypeptides, macrolactones, fatty acids polyketides and isocoumarins with biological activities such as antimicrobial, anticancer and

10

antialgal (Mondol et al., 2013). *Pseudomonas* spp. have also been shown to produce valuable biologically active compounds. One example is a *Pseudomonas fluorescens* which was isolated from a marine sponge *Baikalospongia bacillifera* strain 28Bb-08*,* which was found to contain PKS gene clusters for the production of some novel bioactive compounds (Lipko et al., 2012).

The uncultured sponge bacterial symbionts can be investigated by the metagenomic methods for their potential for producing novel pharmaceutical natural products and the related biosynthetic pathways (Li, 2009). This information can consequently provide insights about the location of unculturable symbiont producers in the host by linking the functional genes to the phylogenetic markers on a single insert (Uria et al., 2005). For example, phylogenetic analysis of sponge DNA resulted in the discovery of a new bacterial phylum *Poribacteria* which was found to contain a sponge specific PKS gene (Fieseler et al., 2004; Siegl & Hentschel, 2010). In their study using whole genome amplification, Siegl and Hentschel, 2010, identified a clone that was 98% identical to *Chloroflexi* associated with a marine sponge *Aplysina aerophoba.* PCR screening of this clone resulted in the isolation of a novel NRPS gene cluster (Siegl & Hentschel, 2010). The screening of the metagenomic library of the marine sponge *Plakortis simplex* has led to the discovery of a novel modular Type I PKS/FAS that is known to be specifically associated with marine sponge symbionts (Della Sala et al., 2013). Another classic example is the metagenomic analysis of the sponge metagenome which led to the isolation of the onnamide/theopederin antitumor PKS gene clusters from uncultured sponge symbionts (Piel et al., 2004).

1.2 Classes of natural products for drug discovery

Natural products include a wide range of biologically active secondary metabolites with pharmaceutical applications. These include, among others, bacteriocins, terpenes, siderophores, lantipeptides, polypetides and nonribosomal peptides. Bacteriocins are a class of ribosomal peptides produced by bacteria to inhibit the growth of other related bacterial strains (Papagianni, 2003). However, the main focus of this study will be on polyketides and nonribosomal peptides as they represent the most understudied of the microbial natural products from marine microorganisms with enormous chemical and biological diversity.

1.2.1 Polyketides

Polyketides are a large structurally diverse group of bioactive secondary metabolites which are mainly produced by microorganisms, and are considered the most important class of marine natural products (Shen, 2003). Many pharmaceutical natural products that are produced by marine sponge symbionts include among others macrolides, polyphenols, polyenes, enedynes, polyethers, erythromycin, tetracyclines, siderophores, rapamycin and mithramycin (Parsley et al., 2011). They are an important source of novel pharmaceutical therapeutics with many clinical applications that include antibiotics, antitumor agents, immunosuppressants, antiparasitics and cholesterol-lowering agents (Kim & Fuerst, 2006). Polyketides are enzymatically synthesized from carboxylic acid by polyketide synthase (PKS) enzymes (Ridley & Khosla, 2009). The polyketide synthetic pathways are classified into three major groups based on the mode of action of the enzymes and the organisation of their gene clusters; namely Type I, Type II and Type III PKSs. Most macrocyclic polyketides are mainly

12

produced by *Streptomyces* through the mechanism of Type I modular polyketide synthases (Schwecke et al., 1995).

1.2.1.1 Mechanism of Type I Polyketide Synthases

Type I PKSs are large, modular multifunctional enzymes that consist of repeated sets of modules which add a single building block to the growing peptide chain (Fieseler et al., 2007). Each module consists of at least three covalently linked catalytic domains, which are ketosynthase (KS), acyltransferase (AT) and acyl carrier protein (ACP) (Figure 1.4). The KS domain catalyses Claisen-type condensation for the chain-elongation reaction where the ketide chain is transferred from the ACP of the upstream module to the cysteine residue in the active site of the KS domain, while the AT domain incorporates the malonyl or methyl-malonyl-CoA to the corresponding ACP domain, and the ACP domain which activates the acyl-CoA and channels the growing polypeptide intermediates from one module to the next (Figure 1.5) (Nikolouli & Mossialos 2012; Shen 2003). There are other additional domains that are involved in the polypeptide chain modification, such as ketoreductases (KR), enoylreductases (ER) and dehydratases (DH). The KR domain plays a role in the reduction of the keto ester, which is then dehydrated by the DH domain and reduced again by the ER domain (Staunton & Weissman, 2001). The TE domain terminates and releases the fully-formed heptaketide intermediate (Staunton & Weissman, 2001).
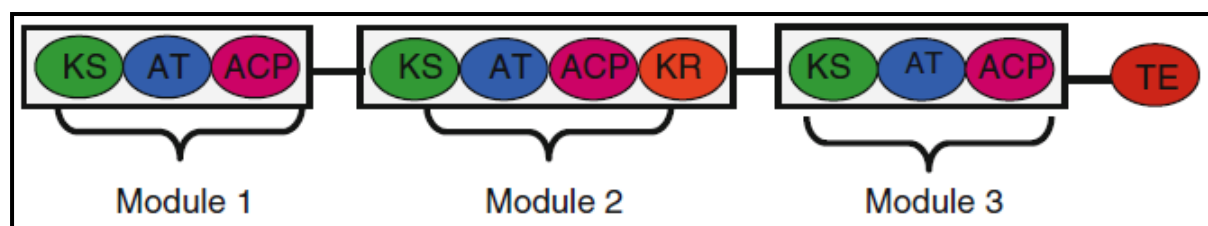
Figure 1.4: The simple modular structure of the polyketide synthase Type I enzyme consisting of a set of enzymatic domains that are responsible for the elongation of a polyketide chain. The additional domains for modification of the individual units can also be found within the modules, such as keto-reductase (KR) domain that contributes to the diversity of the polyketide natural products (Taken from Nikolouli & Mossialos, 2012).

The biosynthesis of Type I polyketides resembles that of fatty acids by the fatty acid synthase (FAS) enzymes in primary metabolism, but unlike FAS that uses malonyl-CoA as a substrate for the extension of fatty acids, Type I PKSs use propionyl-CoA, methylmalonyl-CoA and ethylmalonyl CoA as substrates for the polyketide assembly (Ridley & Khosla, 2009). Type I PKSs are responsible for the synthesis of polyketides with structural diversity, such as erythromycin, rapamycin or epothilone which can be used for the treatment of Gram positive bacterial infections, the treatment of tuberculosis or *Neisseria meningitidis* and exhibit antitumor activity, respectively (Ginolhac et al., 2004; Ridley & Khosla, 2009). The biosynthesis of erythromycin (shown on Figure 1.5) by 6-deoxyerythronolide synthase (DEBS) is one of the most studied Type I PKS pathways. Erythromycin biosynthesis starts with the loading module (AT and ACP) of DEBS 1 which accepts the starter unit propionate from the propionyl-CoA, and ends with the DEBS 3 that terminates the full-length polyketide chain from the ACP domain with the TE domain activity which catalyses the cyclisation of the fully-formed heptaketide intermediate resulting in 6-dEB (Staunton & Weissman, 2001).

14
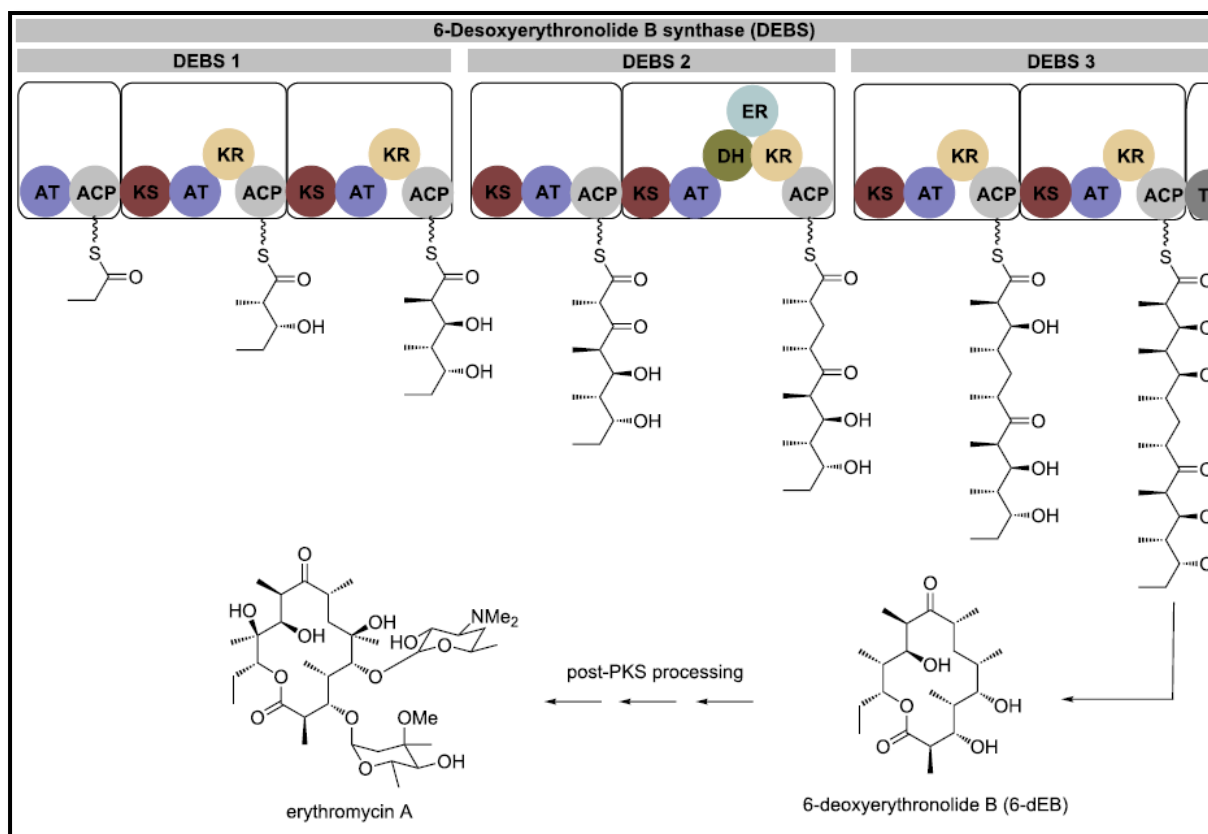
Figure 1.5: An example of a well-studied mechanism of a polyketide antibiotic erythromycin A biosynthesis in the actinomycete bacterium *Saccharopolyspora erythyraea* by Type I Polyketide synthase module. The loading module loads the propionyl CoA and transfers the activated propionyl to the ACP domain. The KS domain receives the growing polyketide chain from the ACP domain and catalyses the decarboxylation between the growing polyketide chain and the extender unit to generated β-ketoacyl thioester-ACP intermediate (Taken from Klopries et al., 2013).

1.2.1.2 Mechanism of Type II Polyketide synthases

Type II polyketide synthases are a dissociated mono-or bi-functional group of enzymes with one or two active sites which act in an iterative manner during synthesis (Liou, 2003; Ridley & Khosla, 2009). These enzymes are commonly found in bacteria, mainly actinomycetes but can also be found in plants and fungi, and usually code for aromatic polyketides (Snyder et

15

al., 2003). Type II PKS enzymes contain a single ketosynthase domain (KS_α), a chain length factor (KS_β) and an acyl carrier protein (ACP) domain that are mainly responsible for the iterative assembly of small phenolic aromatic polyketides such as tetracyclines, daunorubicin and tetracenomycin C. The KS_α subunit catalyses the formation of C-C bond from the activated acyl and malonyl building blocks, while the KS_β determines the chain length (Hertweck et al., 2007). Like Type I PKSs, Type II PKSs also use an ACP domain to activate the acetyl CoA substrate (Shen, 2003). As opposed to Type I PKSs, the genes that encode these protein domains are grouped together to show a KS/KS/ACP architecture (Shen, 2003).

Type II PKS biosynthesis is initiated by the decarboxylation of malonyl-ACP to acetyl-ACP by the KS_α, after which the acetyl group is transferred to the active site of the KS_α to provide the starter unit for the condensation (Figure 1.6). The ACP is loaded with another unit of malonate to give an acetoacetyl-ACP which is transferred from the ACP back to the KS_α through the condensation reaction of KS_α, and then the second cycle of chain extension can begin (Staunton & Weissman, 2001).
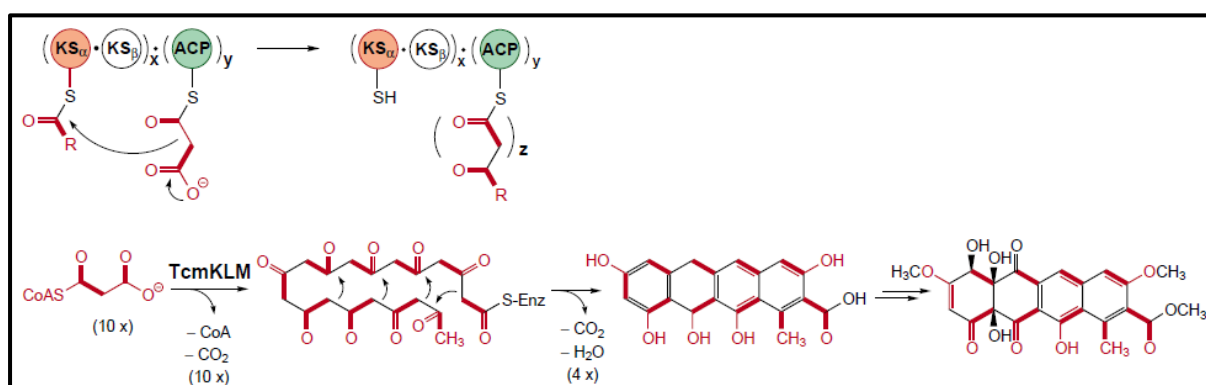
Figure 1.6: The schematic structure of an example of Type II PKS mechanism for the iterative biosynthesis of an aromatic polyketide tetracenomycin (Adapted from Shen, 2003).

1.2.1.3 Mechanism of Type III Polyketide Synthases

The Type III PKSs are the simplest and smallest group of homodimeric enzymes which are commonly found in plants, some bacteria and fungi (Gomes et al., 2013). Type III PKSs are composed of single ketosynthase domains that are responsible for the condensations between acyl CoA units. Because Type III PKSs lack an ACP domain, they directly utilise acyl CoA as a substrate to synthesise polyketide intermediates (Shen, 2003). The acetyl-CoA directly binds to the KS catalytic site which performs the Claisen condensation. They participate iteratively in the biosynthesis of small aromatic compounds such as naringerin, chalcone and tetrahydronaphthalene (THN), which can be a building block for a more complex structure of flaviolin (Figure 1.7) (Hertweck, 2009; Ridley & Khosla, 2009).
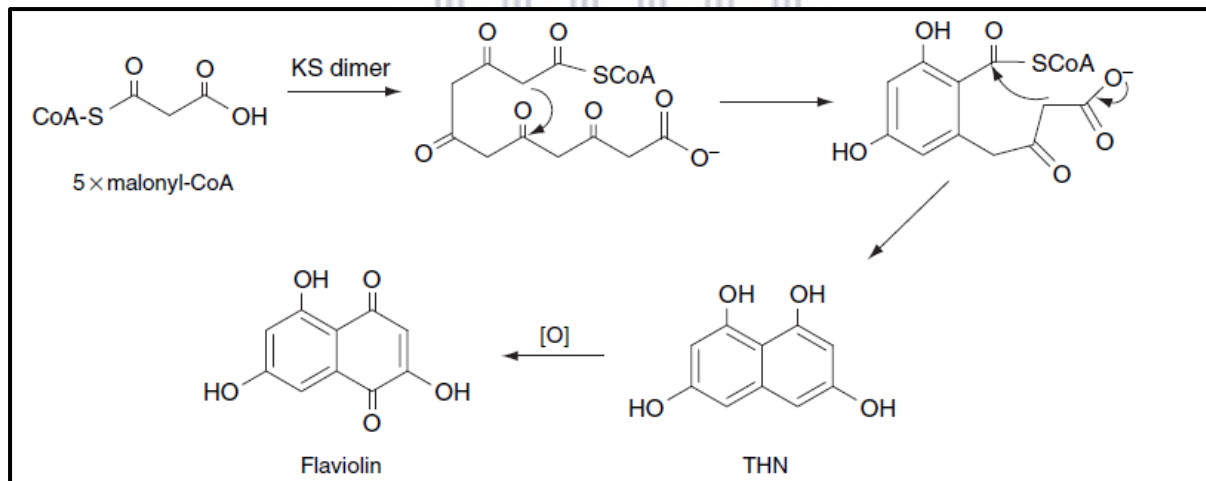
Figure 1.7: An example of the biosynthesis of tetrahydronaphthalene (THN) by Type III PKS. THN can be oxidised to flaviolin, the first Type III polyketide from a bacterium *Streptomyces coelicolor* (Taken from Ridley & Khosla, 2009).

17

1.2.2 Nonribosomal peptides

Nonribosomal peptides are an important and a large group of secondary metabolites which are biosynthesized by the nonribosomal peptide synthases (NRPS) by a variety of microorganisms (Neilan et al., 1999). They include well-known antibiotics such as penicillin, gramicidin, vancomycin, bleomycin, enterobactin, vibriobactin, cyanotoxins and cyclosporine A, with a wide variety of biological activity that includes anti-inflammatorials, anticancer agents, surfactants, immunosuppressants, siderophores and toxins (Tambadou et al., 2014).

The NRPS enzymes catalyse the synthesis of small peptides with a variety of structures such as linear or cyclic-branched, independent of the nucleic acid molecules. These peptides are synthesised by a sequential condensation of amino acids and mostly consist of less than twenty amino acid residues (Ansari et al., 2004).
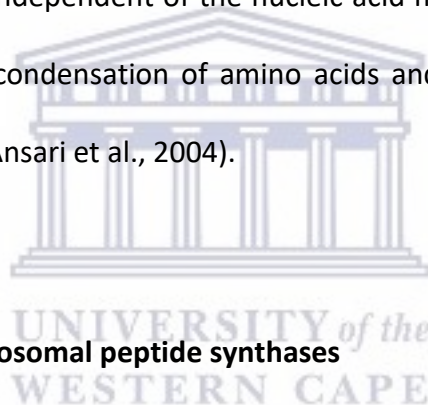
1.2.2.1 Mechanism of Nonribosomal peptide synthases

The nonribosomal peptide synthases (NRPS) are large multidomain modular enzymes that are involved in the biosynthesis of oligopeptide secondary metabolites independently from the ribosomal machinery produced by microorganisms (Nikolouli & Mossialos 2012). NRPS enzymes are organised into a series of catalytic units known as modules which contain three specific main domains, namely condensation (C), adenylation (A) and the peptidyl carrier protein (PCP) domains (Figure 1.8), that sequentially incorporate amino acid building blocks into a growing peptide chain. Each domain is responsible for the catalysis of one round of the addition of an amino acid building block (Pimentel-Elardo et al., 2012). Each module consists of about 1000 to 1500 amino acids, and is a functional building block (Du et al., 2001).
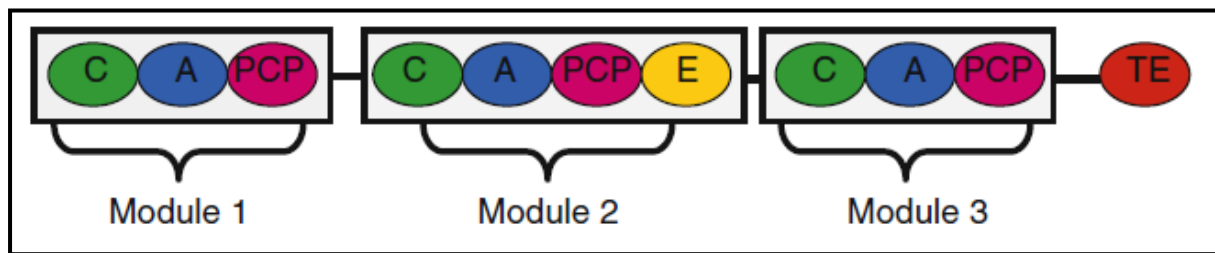
Figure 1.8: The structure of the NRPS enzyme organisation showing three modules consisting of three main domains (C, A and PCP). The condensation (C) domain is responsible for the formation of a peptide (C-N) bond between the elongated chain, the adenylation (A) domain activates the amino acids which are transferred to the PCP domain and the thioesterase (TE) domain terminates the chain elongation to release the NRPS product (Taken from Nikolouli & Mossialos 2012).

The adenylation (A) domain, which has a specific binding unit, is responsible for the recognition and activation of the amino acid monomers by using ATP and the transfer of the activated substrate to the peptidyl carrier protein (PCP) domain of the same module (Figure 1.9). The PCP domain is responsible for the propagation of the growing peptide chain and the condensation (C) domain at the N-terminus, which catalyses the formation of the peptidyl (C-N) bonds between the elongated chain and the activated amino acid from the upstream module (Nikolouli & Mossialos, 2012). The A domain is the most conserved region and it consists of approximately 550 amino acids (Schwarzer & Marahiel, 2001). The PCP domain consists of about 800-1000 amino acid residues, while the C domain is approximately 450 amino acid residues in length (Schwarzer & Marahiel, 2001). The growing peptide remains covalently attached to the enzyme during biosynthesis until it reaches its full length where the thioesterase (TE) domain catalyses the hydrolysis of the peptide or the intracellular

19

cyclisation and releases the final peptide product from the PCP domain of the last module (Ansari et al., 2004; Felnagle et al., 2008).

Figure 1.9: The schematic structure of an example of an antibiotic biosynthesised by the nonribosomal peptide synthase enzyme. 1) The amino acid is activated by the A domain. 2) The activated amino acid is transferred to the PCP domain. 3) The condensation of the amino acid that is bound to PCP domain. 4) The modification of the amino acid by the E domain. 5) Trans-esterification of the peptide chain from the PCP to the TE domain. 6). The release of the final peptide product from the enzyme through either hydrolysis or macrocyclization that is catalysed by the TE domain (Taken from Strieker et al., 2010).

1.2.3 Hybrid NRPS-PKS

The structural similarities between PKS and NRPS has allowed for the formation of hybrid clusters with an element from each class, that produce NRPS-PKS hybrid bioactive products (Du et al., 2001). The mixed NRPS/PKS gene clusters combine both the NRPS and PKS strategies to form structurally diverse hybrid peptide/polyketide natural products from short chain carboxylic acids and amino acids with useful therapeutic activities ranging from immunosuppression, cancer and infectious diseases (Schwarzer & Marahiel, 2001; Walsh, 2004; Walsh, 2008). Hybrid NPs can be divided into two classes based on the biosynthetic mechanisms by which the amino acid and the carboxylic acid monomers are incorporated into them. The first class represents compounds where amino acids and polyketide moieties are synthesised individually by NRPS and PKS enzymes and then combined to form a hybrid final product, while the second class of compounds are where the NRPS/PKS hybrid system mediates the direct transfer of the NRPS-bound peptidyl intermediate by a PKS module to form a hybrid peptide-polyketide product (Du et al., 2001; Nikolouli & Mossialos, 2012).

Examples of these include bleomycin, epothilones, cyclosporine and rapamycin (Du et al., 2001). Hybrid NRPS-PKS compounds have been mainly isolated from the marine actinomycetes (Mizuno et al., 2013). For example, rapamycin which is an immunosuppressant, was isolated from the bacterium *Streptomyces hygroscopicus* (Schwecke et al., 1995). Other examples include barbamide which was isolated from marine cyanobacteria. Other compounds such as onnamide A and theopederin (Figure 1.10) were identified by first isolating the entire gene clusters encoding for their biosynthesis through the metagenomic approach from the symbiotic bacteria of marine sponges, and subsequently

21

expressing in the cultivable bacteria for the production and isolation of these compounds (Lane & Moore, 2012; Piel, 2002; Uria & Piel, 2009).
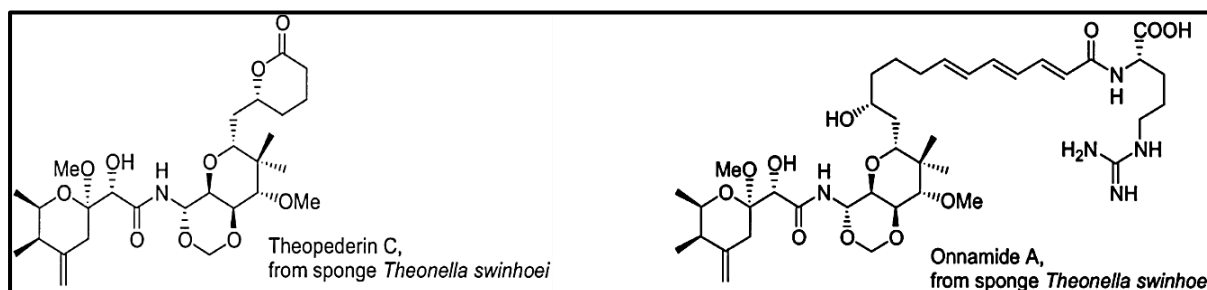
Figure 1.10: Examples of hybrid natural products that were isolated from the symbionts associated with the marine sponge *Theonella swinhoei.* (Adapted from Piel et al., 2004).

1.3   Culture-dependent studies

Previously, the discovery of new pharmaceutical drugs involved the cultivation of microorganisms from the environment, however it was not always successful due to the high rediscovery rate (Gomes et al., 2013). A culture-dependent approach involves directly culturing microorganisms from the environment for the isolation of secondary metabolites. For example, 24 different cultivable actinobacterial species were isolated from a marine sponge *Haliclona* sp. using different selective culture media and PCR screened for secondary metabolite genes encoding PKSs and NRPS using degenerate primers, which resulted in only two isolates that did not show the presence of either of the pathways of interest (Jiang et al., 2007). In another study, 238 bacterial isolates from the sponge *Aplysina aerophoba* were functionally screened, resulting in the inhibition of growth of the bacterial indicator strains by 27 isolates (Hentschel et al., 2001).

The growth conditions such as temperature, aeration, the pH of the media, the incubation time and media composition need to be taken into account when growing cultures as they can affect the production of the metabolite of interest. This is due to the fact that the majority of microorganisms may have needs that are too complex to be reproduced in the laboratory (Gomes et al., 2013). Sometimes even the slight change in the culture medium conditions can have an impact on the quality of a certain compound as well as the metabolic profile of that organism (Scherlach & Hertweck, 2009). Some microorganisms may require growth conditions that resemble that of their native environment in order to produce a maximum amount of active metabolites (Penesyan et al., 2010). An example is the production of a lasso peptide capistruin by *Burkholderia thailandensis* (strain E264). When the strain was initially cultured in M9 minimal media at 37°C, only trace amounts of capistruin was observed but when the strain was cultured in M20 medium at 42°C, the production was increased by approximately 300-fold (Knappe et al., 2008).

The limited ability to culture the majority of environmental microbes represents a major bottleneck in the classical culture-based screening approaches for microbial derived biologically active natural compounds, including those associated with marine invertebrates (Penesyan et al., 2010). Co-culturing can be an effective approach for exploring the microbial secondary metabolites as this can activate silent genes (Montasser & Luesch, 2012). This involves the mixed fermentation between two or more organisms of the same or different species. This approach can result in a large chemical diversity of natural products (Scherlach & Hertweck, 2009). For example, the co-culturing of a marine fungus (strain CNL-365) and a marine bacterium (strain CNJ-328) resulted in the production of a new chlorinated

benzophenone antibiotic, pestalone, which could not be produced when the fungus was cultured on its own (Cueto et al., 2001).

1.4   Metagenomic Studies

Studies have shown that the majority of microorganisms in the environment, whether they are associated with other organisms or free-living have not been cultivated in the laboratory using standard microbiological methods, and there are no general strategies developed for culturing them. Thus, there is an abundance of biosynthetic pathways from microbes that are unable to be accessed using culture-dependent approaches. This has led to the development of new approaches to study these microbes and to identify the genes that are required for the production of natural products, such as metagenomics.

Metagenomics is defined as the culture-independent analysis of microbial genomes contained in a particular environmental sample (Handelsman, 2004). The total environmental DNA can be directly extracted, fragmented and cloned into an easily cultured bacterial host to provide tools for studying the taxonomic diversity as well as the biosynthetic potential of the uncultivated species (Kennedy et al., 2010). These libraries or the uncloned metagenomic DNA can be directly sequenced to analyse the functional capabilities of the whole community and to identify genes that encode natural product biosynthetic pathways (Katz et al., 2015; Walker et al., 2014).

Research has shown that some PKS genes obtained from cultivable bacteria associated with marine sponges are different from the PKS genes derived from the metagenomic library of the same sponge (Kim & Fuerst, 2006). In their research, Kim and Fuerst have shown that the

PKS genes that were observed using metagenomic screening of a sponge metagenome, were not observed in cultivable bacteria, and that the PKSs from sponge metagenomes formed a marine sponge-specific PKS cluster. This suggests that metagenomic screening enables the isolation of genes from uncultivable microbes, which could not be accessed by using only the culture-based screening.

1.4.1   Metagenomic Techniques

The screening of metagenomic libraries is a process of identifying and isolating a particular metagenomic clone that contains the gene or function of interest from a large number of other clones. The metagenomics approach for the natural product discovery can be done using two screening techniques; namely function based and sequence based screening. Function based screening can only be done through metagenomic libraries, whereas sequenced based screening can include PCR based or hybridization probe based screening of libraries. Sequence based screening can also include the sequencing of metagenomic libraries and shotgun sequencing of uncloned metagenomic DNA.

1.4.1.1 The construction of metagenomic libraries

The metagenomics technique begins with the collection of an environmental sample (which can be soil, water or marine invertebrates), followed by directly extracting the total community DNA so that all the genomes in that community can be pooled. DNA is partially digested into large fragments using restriction enzymes or by pipetting after purification and

cloned into an artificial vector system, such as a fosmid, cosmid or bacterial artificial chromosomes (BACs), to generate metagenomic libraries whereby clones can be functionally screened for the unidentified natural products (Figure 1.11). The choice of cloning vectors can determine the success or the failure of the screening protocols. The advantage of using these vectors is that they allow cloning of large DNA insert sizes. High molecular weight DNA is needed for library construction as biosynthetic gene clusters encoding antimicrobial compounds can range from 10- 100 kb (Singh & Macdonald, 2010). The fosmid (the F1 origin-based cosmid vector) or cosmid libraries can only allow for the screening of up to 40 kb DNA inserts, while BACs can allow screening of large inserts of up to 200 kb. Although BAC vectors can allow heterologous expression of larger inserts than fosmids, sufficient extraction of HMW DNA can be a major drawback in BAC metagenomic application (Ouyang et al., 2010). The fosmid vectors replicate using a single-copy F-factor replicon and show a high stability when harbouring large inserts (Weiland et al., 2010). However, functional screening using fosmid vectors can only be done using *E. coli* as a host as a result of the F-factor. There are other shuttle vectors that have extended host range such as RK2-based vectors that can be used to increase the chances of heterologous gene expression. These vectors can allow for the conjugation of libraries from *E. coli* to a number of other hosts (Aakvik et al., 2009). Their ability to replicate in multiple hosts enables the screening to be conducted in the background of different regulatory and metabolic networks (Trindade et al., 2015). Libraries constructed in shuttle vectors need to be transformed into the multiple hosts; however, the levels of efficiency required are difficult to generate in non-*E. coli* hosts (Trindade et al., 2015).

The library is transfected into a suitable microbial host, usually *Escherichia coli*. The advantage of using *E. coli* as a host for expression is that it is commonly used in industrial fermentations

26

as it is easily cultivable under standard laboratory conditions and because of its high transfection efficiency (Daniel, 2004). Another advantage is that *E. coli* is able to grow rapidly on inexpensive substrates and its genetic material has been well characterised (Uria et al., 2005). The disadvantage of using *E. coli* as a host strain is that it may lack the necessary machinery to express the genes that are responsible for the formation of the biocatalysts of interest.

The vector, with the DNA insert that can contain functional genes or phylogenetic genes (Uria et al., 2005), then replicates in the host to produce many clones of the DNA fragments that can be screened for the presence of enzymatic and other bioactive compounds encoded by the metagenomic DNA fragments. The metagenomic libraries, that represent genome fragments from a heterogeneous mixture of species, strains and populations, need to be as large as possible and also represent all microbial cells present in the environmental sample to make sure that the majority of the biosynthetic pathways present in the symbionts are highly covered (van Elsas et al., 2008). The metagenomic library can be as big as two million clones in order to cover all the genomes in that particular sample (Singh & Macdonald, 2010). The constructed metagenomic libraries are then used for the detection of novel natural products based on the functional expression or sequence based mining.

1.4.1.2 Function- based screening

The function-based or activity-based screening approach involves the screening of metagenomic library clones directly for the genes encoding biological or functional activities without having to sequence them first, and so it allows for the identification of complete

genes or gene clusters and new classes of enzymes that have useful functions in the pharmaceutical industry (Kennedy et al., 2010; Angelov et al., 2009). Since the function-based approach does not rely on the previous knowledge of the gene sequence of the biosynthetic pathway of interest, it can be used to detect DNA sequences with completely new functions that are different from the currently known biocatalysts, and that would have been difficult to identify using sequence information alone (Yun & Ryu, 2005; McMahon et al., 2012). This approach has the potential for the direct identification of clones that are capable of inhibiting the growth of target pathogens and it can also be used in industries for the discovery of novel enzymes. As soon as the clone that contains the function of interest is identified, the fosmid DNA is extracted and sequenced in order to compare the insert sequence with other DNA sequences from other organisms and to identify genes that are responsible for the activity production (Kennedy et al., 2010).

However, the limitation of this approach is that it is dependent on the success of expression of the genes of interest in the heterologous screening host, their ability to be phenotypically detected and the clustering of all genes that are required for the functions in order to work efficiently (Schloss & Handelsman, 2003; Angelov et al., 2009). Most function-based approaches rely on *E. coli* as an expression host, which might not have the necessary elements to decode DNA from distantly related microorganisms and only a small number of positive clones may be obtained during a round of screening (Singh et al., 2009). This can be due to the lack of the regulatory elements that turn on the expression of gene clusters, or due to the promoter regions not being able to be recognized by the *E. coli* transcriptional machinery or expressed functional proteins might not be exported for activity (Kennedy et al., 2010). When screening for the genes that are responsible for the production of natural products, *E. coli*

might not be able to express these genes because of the lack of the necessary precursors that are responsible for the production of that particular metabolite. Alternative bacterial hosts from different taxonomic groups such as *Streptomyces lividans*, *Pseudomonas putida* and *Rhizobium leguminosarum* can be used in addition to *E. coli* by providing the machinery needed for the expression of genes from diverse microorganisms and increase the success rate of function-based screening (Kennedy et al., 2008). *Streptomyces* spp. are better hosts for the detection of secondary metabolites as their GC-rich genomes supports the large proportion of biosynthetic genes that synthesize numerous secondary metabolites (Singh & Macdonald, 2010).

The function-based approach has been used to detect antibiotic activities in metagenomic clone libraries of soil and marine invertebrates. This approach mainly relies on the top agar overlay assays for the functional screening of clones that produce antibacterial small molecules (Brady, 2007; Iqbal et al., 2014). During the functional screening process, the individual metagenomic clones are normally placed on an agar plate which contains the appropriate media for the growth of the host strain in a 96-well format, and the indicator strain in sloppy agar poured over the entire plate. The presence of the low molecular weight antibiotic substances is observed by inhibition zones of clearance around the clone colonies on the lawn of the test strains on the plates or colour production by the clones. For example, a clone producing a long-chain *N*-acyltyrosines antibiotic was identified from the eDNA cosmid libraries that were screened in *E. coli* using an agar overlay assay (Brady et al., 2004). Another example is the antibacterial activity screening of 5000 eDNA cosmid library clones in *E. coli*, which resulted in the observation of a clone which inhibited growth of *B. subtilis* after using an agar overlay assay (Courtois et al., 2003). Functional screening can also detect the

29

production of pigmentation by the clones in the media. This has led to the isolation of antibacterial active pigments such as violacein, indigo and turbomycin A and B from the soil metagenomic library (Rocha-Martin et al., 2014; Gillespie et al., 2002).
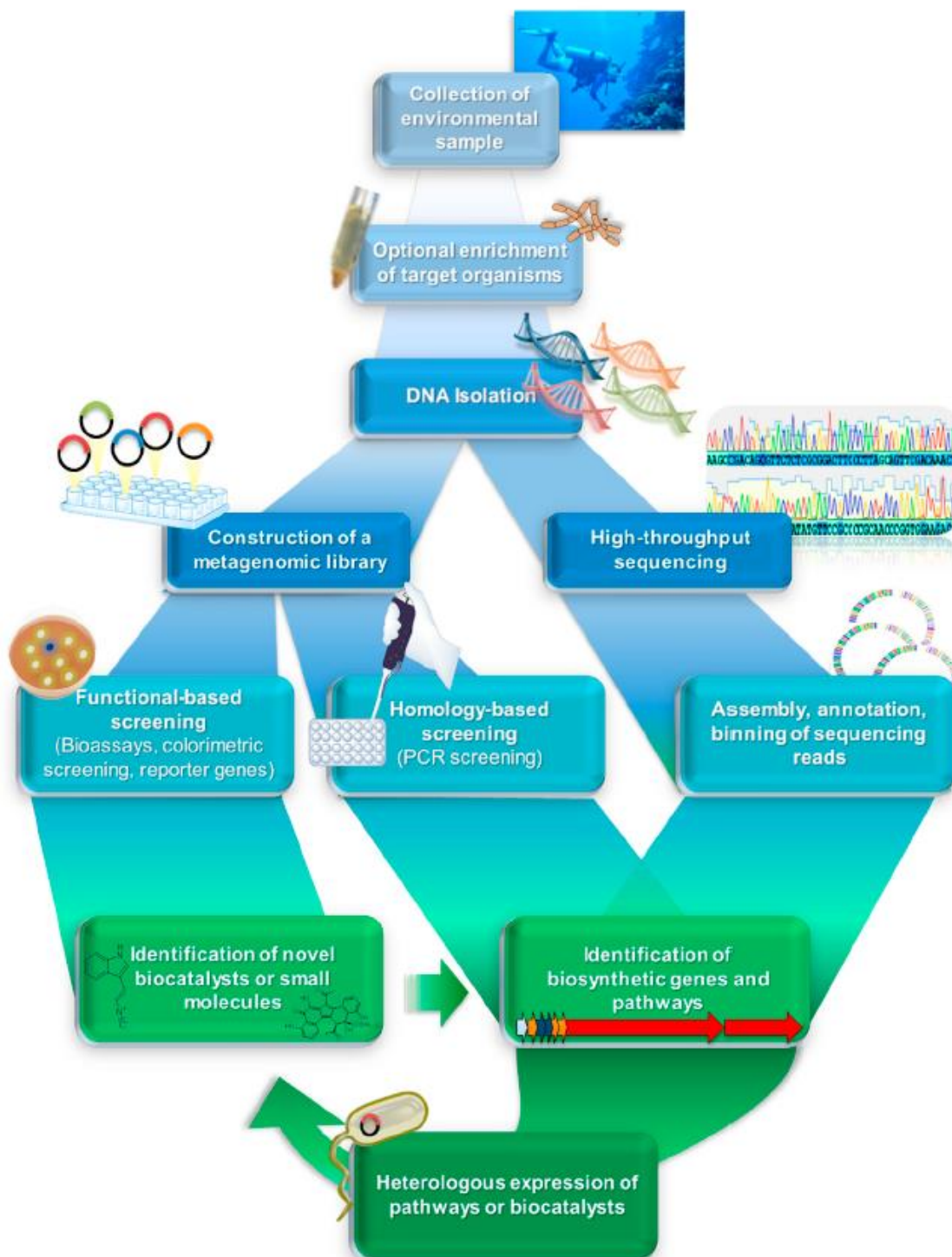
Figure 1.11: The schematic representation of the construction of metagenomic libraries from uncultured microorganisms from environmental samples, and the two strategies (functional and sequenced-based) that are used for screening them for novel natural product discovery (Taken from Wilson & Piel 2013).

31

1.4.1.3 Sequence-based screening

1.4.1.3.1 PCR or probe based screening

The sequence-based screening approach relies on the conserved sequences of the target genes to design degenerate PCR primers or hybridization probes to search for and identify the similar sequences present in the metagenomic library clones (Schloss & Handelsman, 2003). The PCR based approach can be used to study the microbial community in an environmental sample as well as the genes that encode for the novel bioactive compounds on the basis of the sequence homology. The main advantage of the sequence-based approach is that it does not depend on the expression of the biosynthetic genes of interest in a heterologous host for their detection. However, the main limitation of this screening strategy is that it can only be used to target sequences of the genes which are already known and cannot be used to identify novel genes or products that have different sequences (Iqbal et al., 2012; Daniel, 2004; Singh & Macdonald, 2010). This approach cannot guarantee the acquirement of full biosynthetic clusters that are responsible for the biosynthesis of the product of interest (Yun & Ryu, 2005). The fragments amplified during PCR screening can be used as hybridization probes for further screening of the metagenomic clones, where the clones are transferred from the agar plates onto the nylon membranes and hybridised with the labelled PCR products (Metsä-Ketelä et al., 1999; Ginolhac et al., 2004).

NP biosynthetic genes are targeted by PCR-based screening with degenerate primers that were designed from highly conserved motifs of the known genes to amplify gene fragments. Therefore only genes with regions that are similar to those of the primers can be recovered

from metagenomic library clones (Simon & Daniel, 2009). The PCR-based approach is normally used to identify modular PKS and NRPS biosynthetic genes from metagenomic library as these pathways are difficult to identify using functional-based screening (Wilson & Piel, 2013). This is because most of the natural products cannot be synthesised under normal laboratory conditions. If they are produced, it is usually in small amounts that cannot be detected through functional screening and it requires all the biosynthetic genes to be present and expressed in the host organism (Sun et al., 2012; Owen et al., 2012).

Despite the difficulties discussed, this approach has been applied successfully for the discovery and isolation of novel natural products, which include polyketides and nonribosomal peptides, with diverse therapeutic potential from marine invertebrates (particularly sponges, bryozoans and tunicates) metagenomic DNA by PCR-based screening. For example, the onnamide and theopederin biosynthetic gene clusters were isolated from the complex metagenomic DNA of the marine sponge *Theonella swinhoei* by PCR amplification of the KS domains of the Type I PKS (Piel et al., 2004). The analysis of these gene sequences provided evidence that the genes are from an uncultured microbial symbiont. Other examples include the isolation of the two polyketides psymberin and pederin gene clusters from the metagenomes of the marine sponge *Psammocinia* aff. *bulbosa* and the *Paederus* beetle's endosymbionts, respectively by using degenerate PCR amplification of the KS domains (Piel, 2002; Fisch et al., 2009).

1.4.1.3.2 Library sequencing

Library sequencing is another strategy for identifying novel biosynthetic genes from the DNA of the previously uncultivated bacteria. This will be further discussed in detail in Section 1.4.1.3.3.

1.4.1.3.3 Metagenomic DNA sequencing/shotgun sequencing

Metagenomic DNA sequencing involves sequencing of total DNA fragments isolated from bacterial populations of an environmental sample and is able to capture complete genomes of all organisms in that population. The main purpose of DNA sequencing is to determine the nucleotide sequence of genetic material of the entity of an organism which may include both coding and non-coding sequences of DNA, and to identify the location of those genes (Ridley, 2000). The advantage of metagenomic DNA sequencing is that it provides information about the gene sequences of the bacterial population without the need to assemble individual bacterial genomes. Sequencing of the metagenomic DNA and the analysis of the genome sequences can give information about the genes encoding target proteins of a particular environmental sample by comparing them with the known ones based on the similarity of the gene sequences (Nikolouli & Mossialos 2012). Metagenomic sequencing enables the sequencing of the whole genome of a particular bacterial species without culturing. This can enable the recovery of functional genes from the environmental sample that can be identified using PCR and then expressed in a heterologous host (Singh & Macdonald, 2010). The decrease in the cost of genome sequencing allows for the discovery of thousands of gene clusters encoding the biosynthetic enzymes and new metabolic pathways for biosynthesis of natural compounds of uncultured microorganisms (Ng et al., 2015; Medema et al., 2011). The

34

development of bioinformatics tools during the last decade have led to the identification and understanding of new NRPS and PKS gene clusters through genome mining (Jenke-Kodama & Dittmann, 2009; Blin et al., 2013). Genome annotation is an important process of analyzing the biological information of the assembled amino acid sequences (Stein 2001), in order to determine the properties of the predicted natural compounds. There are two main tasks in genome annotation: gene finding and protein function assignment.

Many algorithms have been developed for the bioinformatics analysis of secondary metabolite biosynthesis based on microbial genome sequences (Blin et al., 2013). These algorithms have greatly increased the rate of natural product discovery from marine microbes. An example is the Basic Local Alignment Search Tool (BLAST) algorithm on NCBI which determines the proteins which have a short region in common with known NRPS and PKS enzymes and can be used for the discovery of NRPS and PKS products (Nikolouli & Mossialos, 2012).

The antibiotics and secondary metabolite analysis shell (antiSMASH) (http://antismash.secondarymetabolites.org/) web server and stand-alone software was developed for the analysis of secondary metabolite gene clusters from nucleotide sequences (Blin et al., 2013; Medema et al., 2011; Weber et al., 2015). This automated genome mining tool is able to identify gene clusters encoding secondary metabolites from assembled contiguous sequences of bacterial and fungal origin generated from the metagenomic DNA, as well as providing information on genome function and chemical structures (Medema et al., 2011; Charlop-Powers et al., 2014). This software allows for the upload of the DNA sequence

files (either in FASTA, EMBL or GenBank format) to detect and identify PKS and NRPS biosynthetic gene clusters from uncultivated microbes, their expected natural products encoded and their functions.

PRediction Informatics for Secondary Metabolomes (PRISM) (http://magarveylab.ca/prism/) is another bioinformatics tool that can be used for genome mining. This computational web server can be used for the identification of gene clusters involved in the biosynthesis of secondary metabolites as well as the prediction of the chemical structures of NRPs and Type I and Type II polyketides encoded by those gene clusters (Skinnider et al., 2015). Like the antiSMASH web server, PRISM also allows for the input of DNA sequences either in FASTA or GenBank format and also give functional information about the biosynthetic pathway of interest.
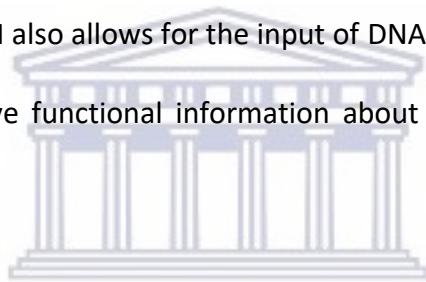
This information will then enable the researchers to exploit the potential of an organism to produce novel compounds (Lane & Moore, 2012). Once the specific biosynthetic genes or pathways have been identified from the metagenomic sequence, they can be selected, cloned and expressed in an appropriate cultivable host for functional screening as discussed in section 1.4.1.2.

1.5   The Aims and Objectives

The aim of this research is to explore the metagenomic library of the symbiotic community associated with South African marine sponges in order to discover the new biosynthetic enzymes, polyketide synthases and nonribosomal peptide synthases, which could lead to the identification of novel natural products for drug discovery and development.

The specific objectives of this study are:

1. To functionally screen for novel antimicrobial activities of the metagenomic clones against four different indicator strains to determine the production of antibacterial activity by observing the colour change and the presence of the zone of clearance around the metagenomic clones.

2. To use PCR-based screening with different degenerate primers specific for polyketide synthase and nonribosomal peptide synthase domains to amplify them and then isolate, sequence and characterize the amplicons in order to determine their biosynthetic pathways.

Chapter 2

Materials and Methods

2.1   Sample information

The sponge samples were collected from Algoa Bay, in Port Elisabeth, Eastern Cape, South Africa by SCUBA diving at depths of up to 23 m (longitude 34.006 S and latitude 25,7201 E). The sponge samples (PE8 and PE7) were identified as *Spongia* (Spongia) sp. 001RSASPN and *Higginsia bidentifera*, respectively (Figure 2.1). The sponge species was selected due to their endemism in South Africa and their association with microbial species that are known to produce biocatalysts and natural products (unpublished results from other IMBM projects).

Phylum:	Porifera	Phylum:	Porifera
Class:	Demospongiae	Class:	Demospongiae
Order:	Dictyoceratida	Order:	Axinellida
Family:	Spongidae	Family:	Stelligeridae
Genus:	Spongia	Genus:	Higginsia
Species:	Spongia (Spongia) sp. 001RSASPN	Species:	Higginsia bidentifera
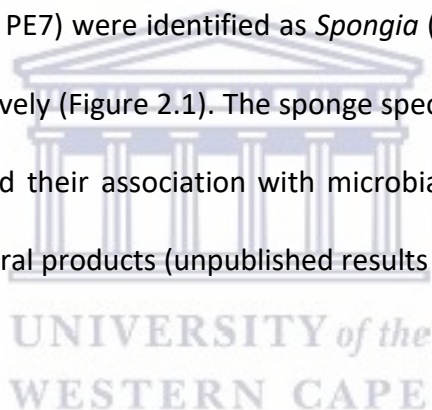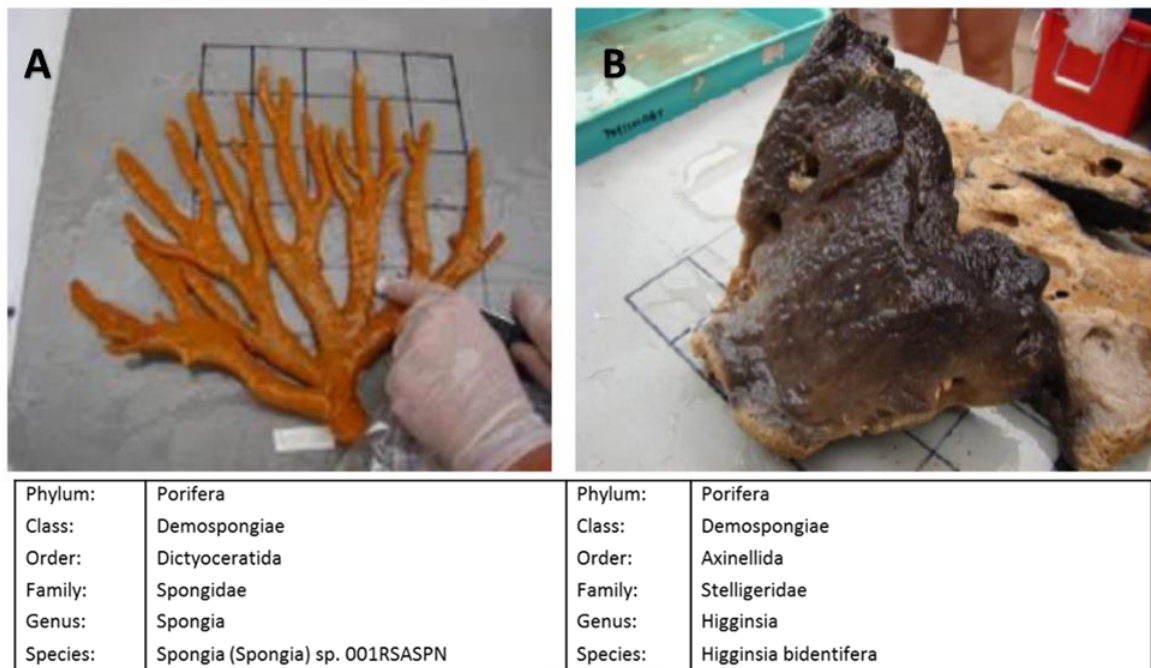
Figure 2.1: The classification of the marine sponge samples that were used in this study. *Spongia (Spongia)* sp. 001RSASPN (A) and *Higginsia bidentifera* (B).

2.2   General laboratory chemicals and reagents

The chemicals used in this study were supplied by Sigma Aldrich Chemical Company (Deissenhofen, Germany), Flukka, Kimix Chemical and Laboratory supplies (South Africa) and Merck Chemical and Laboratory supplies (Darmstadt, Germany). The DNA size markers and DNA modifying enzymes (polymerases and restriction endonucleases) were purchased from Life Technologies (California, USA) and New England Biolabs Inc. (Ipswich, MA, USA). Oligonucleotides for Polymerase Chain reaction (PCR) were purchased from Integrated DNA Technologies (IDT) (Iowa, USA).

The buffers and the stock solutions used in this study are listed in Appendix A.

2.3   Media

The media used are listed in Table 2.1. The recipes are from Sambrook and Russel (2001) unless otherwise stated. All media were autoclaved at 121°C for 20 minutes unless otherwise specified.

Table 2.1: The growth media used in the study

Constituent	1 Litre final Volume
LB Medium (Luria-Bertani Medium)	
Tryptone	10 g
Yeast extract	5 g
NaCl	10 g
LB agar	
Tryptone	10 g
Yeast extract	5 g
NaCl	10 g
Agar	15 g
SOC Medium	
Tryptone	20 g
Yeast Extract	5 g
NaCl	0.5 g
KCl	10 ml
$MgCl_2$	5 ml (filter-sterilized and added before use)
Glucose	20 ml (filter-sterilized and added before use)

After autoclaving, the media were cooled down to 55°C before adding antibiotics (Table 2.2) where necessary.

40

Table 2.2: Stocks and final concentrations of antibiotics used in the study

Antibiotics	Preparation
Ampicillin	100 mg/ml in distilled water stock 100 µg/ml final concentrations
Chloramphenicol	34 mg/ml in 100% ethanol stock 12.5 µg/ml final concentrations

2.4   Bacterial strains and plasmids

Table 2.3: Bacterial strains used in the study

Bacterial strain	Standard features	Supplier
E. coli EPI300™	*F⁻ mcrA Δ(mrr-hsdRMS-mcrBC) f80lacZ M15 ΔlacX74 recA1 endA1 araD139Δ(ara, leu)7697 galU galK 1-rpsl (StrR) nupG trfA*	Epicentre Biotechnology, an Illumina company (USA)
E. coli JM109	*endA1, recA1, gyrA96, thi, hsdR17 (rk-, mk+), relA1, supE44, Δ (lac-proAB),[F'traD36, proAB, laqIqZΔM15]*	Promega Corp. Madison, (USA)

Table 2.4: Plasmids used in the study

Plasmid	Description	Source
pCC2FOS	Fosmid vector containing a chloramphenicol resistance gene, linearized at the Eco72 I (blunt) site, dephosphorylated	Epicentre Biotechnology, an Illumina company (USA)
pGEM T- Easy	Cloning vector containing an ampicillin resistance gene, with T overhangs at the insertion site	Promega Corp. Madison, (USA)
pJET1.2/blunt	Suicide vector (Eco47IR), blunt DNA ends for ligation with insert, T4 promoter, Amp^R	Thermo Scientific (USA)
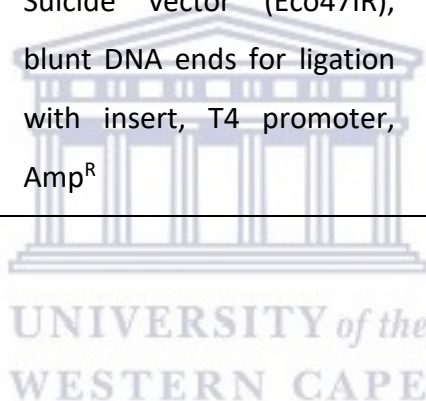
2.5 Extraction of DNA

2.5.1 Plasmid extraction for sequencing

Single *E. coli* colonies were picked and incubated overnight at 37°C with shaking at 200 rpm in LB medium (10 ml) in the presence of an appropriate antibiotic (normally 100 µg/ml ampicillin). Cells were harvested by centrifugation at 6000 x g for 10 min. Plasmid DNA was extracted using the Plasmid Mini Prep kit (Qiagen GmbH, Hilden, Germany), following the manufactures' instructions.

2.5.2   Fosmid DNA extraction

Single *E. coli* clones containing fosmid were picked from the agar plates and inoculated into 10 ml LB with 12.5 µg/ml CAM and 0,02% L-arabinose (w/v) (induction solution) before incubating at 37°C overnight with shaking. The cells were harvested by centrifugation at 6000 x g for 10 mins at 4°C. The supernatant was discarded. Cells were resuspended in 1 ml of ice cold GET buffer (50 mM glucose, 10 mM EDTA, 25 mM Tris-HCl, pH 8.0) and 2 µl of 10mg/ml RNase A (Thermo Scientific) was added. A volume of 1 ml of lysis solution (0.2 M NaOH, 1% SDS) was added and the tubes were incubated at room temperature for 5 mins. A volume of 1 ml of 3 M potassium acetate (pH5.5) was added and the cells were incubated on ice for 15 mins. The solution was centrifuged at 15,000 rpm at for 30 mins at 4 °C. The supernatant was transferred into fresh tubes and one volume of isopropanol was added before incubating at room temperature for 15 mins to precipitate DNA. The tubes were centrifuged at 15,000 rpm for 30 mins, the supernatant was discarded and the DNA pellets were washed by adding 1 ml 70% ice-cold ethanol and centrifuging at 15,000 rpm for 10 mins. Ethanol was removed and the DNA pellets were air dried. The DNA was resuspended in 50 µl of 1xTE buffer (pH8).

2.5.3   Gel extraction and PCR purification

PCR products were separated by gel electrophoresis and the DNA fragments excised from the gels with a sterile scalpel under a long wavelength transilluminator, according to their fragments size. The excised fragments were placed into sterile Eppendorf tubes. Purification of the DNA fragments or the PCR reactions were carried out using the NucleoSpin® Gel Clean-

43

up kit (Macherey-Nagel, Germany) or Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare Life Sciences, USA) following the manufacturers' instructions.

2.5.4   DNA Quantification by Spectrophotometry

The nucleic acid purity, based on a 260/280 ratio and the concentrations were determined by using the NanoDrop® ND-1000 (NanoDrop Technologies, Inc., USA), following the manufacturer's instructions.

2.5.5   DNA agarose gel electrophoresis

Agarose gel electrophoresis was used to separate the DNA fragments. The agarose gels were prepared in 0.8-2% [w/v] agarose in 1X TAE buffer (40 mM Tris, 20 mM acetic acid, 1mM EDTA), containing 0.5 µg/ml ethidium bromide to ensure DNA visualisation. DNA samples were mixed with 6X loading buffer (20% glycerol, 0.1 M EDTA, 1% SDS, 0.25% Bromophenol blue) before loading onto an agarose gel. The molecular markers of the appropriate size were used to estimate the size of the DNA bands. The agarose gel electrophoresis was performed in 1X TAE running buffer at 90-120 Volts for 1-1.5 hours. The DNA fragments were visualised under UV light and the gels were imaged using the digital imaging system Alpha Imager®HP 2000 (Alpha Innotech, San Leadro, USA).

2.6   Polymerase Chain Reaction

Polymerase chain reactions (PCR) were carried out in 0.2 ml thin-walled tubes in volumes of

25 to 50 µl, using the thermocycler that is equipped with a heated lid. A standard reaction

consisted of 1X PCR buffer, 0.2 mM of dNTP mix, 1 µM of each of the forward and reverse

primers, 1.25U thermostable DNA polymerase and the nuclease-free water to adjust the

volume. The reactions were carried out according to the manufacturers' instructions using a

Labnet MultiGene™ Gradient PCR Thermal Cycler (Labnet International, Inc.) or an automated

thermal cycler (T100 Thermal Cycler, BioRad). The primers used and the cycling conditions are

listed in Table 2.5.

Materials and Methods

Table 2.5: Primers used in the study and the PCR cycling conditions

Primer	Target region	Primer Sequence (5'-3')	Product size (bp)	Cycling conditions	Reference
MTF MTR	A-domain (NRPS)	CCNCGDATYTTNACYTG GCNGGYGGYGCNTAYGTNCC	900	Initial denaturation at 94°C for 5 min, 30 cycles 94°C for 1 min, 53°C for 30s, 72°C for 1 min and final extension at 72°C for 10 min.	(Vizcaino et al., 2005)
NRPS-A3 NRPS-A7R	A-domain (NRPS)	GCSTACSYSATSTACACSTCSGG SASGTCVCCSGTSCGGTAS	700	Touchdown: 95°C for 3 min, 14 cycles 95°C for 30s, 63°C (-1°C/cycle) for 30s, 72°C for 60s. 30 cycles of 95°C for 30s, 50°C for 30s, 72°C for 60s and final extension at 72°C for 5 min.	Gontang et al., 2010
KS1F1 KS1R1	Type I PKS (KS domain)	ATGGAYCCSCARCARCGBCT GCYTCGATSGGRTCNCCSA	700	Touchdown: Initial denaturation at 95°C for 3 min, 10 cycles 95°C for 30s, 77.3°C(-1°C/cycle) for 30s, 72°C for 60s. 20 cycles of 95°Cfor 30s, 62.3°C for 30s, 72°C for 50s and final extension at 72°C for 5 min.	(Russel Hill, unpublished)
KSDPQQF KSFGTGR	Type I PKS (KS domain)	MGNGARGCNNWNSMNATGGAYCCNCARCANMG GGRTCNCCNARNSWNGTNCCNGTNCCRTG	750	Touchdown: 95°C for 3 min, 14 cycles 95°C for 30s, 63°C (-1°C/cycle) for 30s, 72°C for 60s. 30 cycles of 95°C for 30s, 50°C for 30s, 72°C for 60s and final extension at 72°C for 5 min.	(Fieseler et al., 2007)
KS2aF1 KS2bR1	Type II PKS (KS domain)	TSGCSTGYTTCGAYGCSAT GCRTAGAACCASGCGAWSGA	1070	Touchdown: Initial denaturation at 95°C for 3 min, 10 cycles 95°C for 30s, 78°C(-1°C/cycle) for 30s, 72°C for 80s. 20 cycles of 95°C for 30s, 63°C for 30s, 72°C for 80s and final extension at 72°C for 5 min.	(Russel Hill, unpublished)
M13F M13R	Region flanking the MCS of pGem®-T Easy vector	CCCACTCACGACGTTGTAAAACG AGCGGATAACAATTTCACACAGG		Touchdown: Initial denaturation at 94°C for 2 min, 10 cycles 94°C for 30s, 65°C (-1°C/cycle) for 30s, 72°C for 50s. 25 cycles of 94°C for 30s, 55°C for 30s, 72°C for 50s and final extension at 72°C for 5 min.	(Norrander et al., 1983)

2.7   Restriction enzyme digestion

Restriction enzyme digestions were prepared in sterile 1.5 ml centrifge tubes in 10 – 30 μl reaction volumes and were incubated at 37°C overnight. Approximately 1U of the enzyme was used per microgram of plasmid or genomic DNA in the presence of an appropriate buffer as supplied by the manufacturer. Restriction enzymes were inactivated by incubating at 80°C for 20 mins.

2.8   Ligation of PCR products

Ligation reactions were carried out in 10-20 μl total volume. To each microcentrifuge tube, an appropriate cloning vector (50 ng/μl pGEM®-T Easy Vector System I or pJET1.2/blunt cloning vector) and an appropriate amount of purified PCR products, in a 2:1 or 3:1 ratio were combined with 1U of T4 DNA ligase, 1x ligation buffer and the UV-treated water to make up the required final volume. Reactions of 10-20 μl total volume were incubated at 4°C or 22°C for 1-16 hours. Ligation reactions were transformed directly into host *E. coli* cells by electroporation (2.9.2).

2.9   Transformation of the ligation products

2.9.1   Preparation of electrocompetent cells

The glycerol stocks of EPI300 or JM109 *E. coli* were streaked onto a LB agar plate and incubated at 37°C overnight. A single colony was transferred into 10 ml LB and incubated at 37°C overnight with shaking. All the following steps were carried out in the cold room (4°C)

47

or on ice. The competent cells were prepared by inoculating the 10 ml culture into 1 L of LB and grown at 37°C with shaking until an OD_{600} of 0.35-0.4 was reached. The cells were immediately chilled on ice for 30 minutes, with gentle swirling occasionally to ensure even cooling. The cells were harvested at 1000 x g for 20 minutes at 4°C. The supernatant was discarded and the cell pellet resuspended in 200 ml ice cold sterile distilled water before harvesting the cells at 1000 x g for 20 minutes at 4°C. The supernatant was discarded and the pellet resuspended in 100 ml of ice cold sterile distilled water. The cells were harvested at 1000 x g for 20 minutes at 4°C. The supernatant was discarded and the pellet resuspended in 40 ml of ice cold 10% [v/v] glycerol which was transferred into two 50 ml conical tubes. The cells were harvested at 1000 x g for 20 minutes at 4°C. The supernatant was carefully aspirated using a sterile Pasteur pipette, the tubes transferred onto ice and the cell pellets resuspended in 1 ml ice cold 10% [v/v] glycerol by swirling gently. The resuspended cells were aliquoted into sterile ice cold 1.5 ml microcentrifuge tubes on ice and snap frozen in liquid nitrogen before storing in the -80°C freezer until further use.

2.9.2   Transformation of electrocompetent cells

Electrocompetent EPI300 or JM109 *E. coli* cells were transformed directly with the ligation reaction. The microcentrifuge tube containing 50 µl electrocompetent cells was thawed on ice for about one minute. The ligation mixture (2 µl) was added to the thawed cells and mixed gently. The mixture was transferred into a chilled sterile 0.1 cm electroporation cuvette (BioRad). The electroporation was done using the BioRad MicroPulser™ under the following conditions: 1.8 KV, 25 µF and 200 Ω. One ml of LB was immediately added to the pulsed cells

and the mixture was transferred to a 1.5 ml microcentrifuge tube and incubated at 37°C for 1 hour while shaking. A 100 µl of the mixture was plated out onto an LB agar plate complemented with 100 µg/ml ampicillin (80 µg/ml X-Gal and 0.5 mM IPTG were added where a pGEM®-T Easy vector was used for ligations), and incubated at 37°C overnight.

2.9.3   Screening of transformants

The white colonies were picked, inoculated into LB containing 100 µg/ml ampicillin and incubated at 37°C overnight while shaking. The plasmid DNA was extracted as described in Section 2.5.1. The PCR with M13 vector primers (Table 2.5.) was used to screen for the clones containing the expected insert size. The samples that contained the expected product were sequenced and analysed.
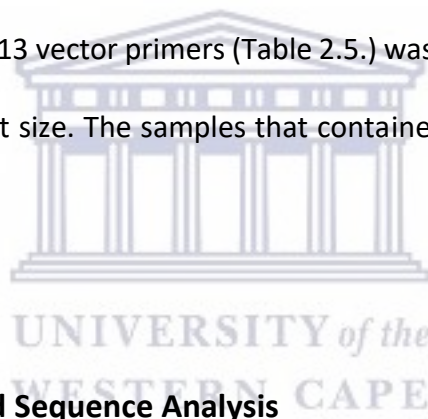
2.10   DNA Sequencing and Sequence Analysis

Sanger DNA sequencing was performed at the University of Stellenbosch sequencing facility (ABI PRISM 377 automated DNA sequencer). Purified PCR amplicons and cloned DNA inserts were sequenced in both forward and reverse directions using the appropriate primers. The nucleotide sequences and chromatograms were manually edited using the software package Chromas Lite 2.1 version (Technnelysium Pty Ltd, 2012). The DNA sequences were analysed using BLASTx and BLASTn (nr/nt) tools of the NCBI Genbank (National Centre for Biotechnology Information, Bethesda, MD) nucleotide collection database, to identify the closest matching sequences (http://www. ncbi.nlm.nih.gov/BLAST).

49

2.11 Fosmid DNA sequencing

The high quality fosmid DNA was extracted (Section 2.5.2) from a positive metagenomic clone and submitted to the University of the Western Cape next generation (MiSeq) sequencing facility. The sample was prepared using the Nextera XT kit v2 and sequenced using the Illumina MiSeq platform.

2.11.1 Fosmid sequence assembly

The fosmid sequence data was analysed using the CLC genomics work bench version 6.5 software package. Contiguous sequences (contigs) from the assembly were analysed for putative open reading frames (ORFs), and BLASTn and BLASTx were used to search the nucleotide and protein database for homologous nucleotide and protein sequences, respectively (Altschul et al., 1997). The reads were subjected to quality control and adapter trimming using the following parameters; quality score _ 0.5, maximum score _ 3 ambiguous bases, minimum sequence length of 50 nucleotides and removal of adapter sequences. The contigs were constructed *de novo* assembly using the following parameters; mismatch cost _ 2, insertion cost _ 3, length fraction _ 0.95 and a maximum contig length of 1000bp.

2.11.2 Annotation of secondary metabolites pathways

The annotation and detection of secondary metabolite pathways of the fosmid sequence data contigs were performed using antiSMASH version 3.0 software

50

(http://antismash.secondarymetabolites.org/) (Blin et al., 2013). The antiSMASH analysis

parameters used were as shown in Figure 2.2.

Figure 2.2: the antiSMASH genome annotation parameters used in this study.

2.12   Functional based screening

2.12.1 Preparation of glycerol stocks

The fosmid clone library aliquots that were stored at -80°Cwere thawed and diluted 1:1 with LB broth. Each 2 ml diluted culture was plated onto a Q-Tray (Genetix) containing LB agar supplemented with 12.5 μg/ml of chloramphenicol at a density of 1000 clones per plate. The Q-trays were incubated at 37°C overnight. The fosmid clone library (approximately 10,000) was transferred into 96 well microtiter plates containing LB medium supplemented with 12.5 μg/ml of chloramphenicol using a colony picking/gridding robotics system (Genetix QPix 2 XT, Molecular Devices, Hampshire, UK). The plates were sealed with breathable strips and incubated at 37°C overnight with gentle shaking. Replicates were made with a sterile 96 pin replicator into 96 deep well plates (10 clones per well) containing 2 ml LB supplemented with 12.5 μg/ml of chloramphenicol and 0.02% L-Arabinose for the induction of fosmid copy number. Fifty percent glycerol was added to a final concentration of 20% in each plate before they were sealed with an aluminium foil and stored at -80°C. The deep well plates were incubated at 37°C overnight with gentle shaking. The replica plates were used for the antibacterial activity screening.

2.12.2 Library screening for antibacterial activity

The agar-overlay assay was used to test for the antimicrobial activity of the fosmid library clones against four indicator strains, namely; *Bacillus cereus* ATCC 10702, *Pseudomonas putida* ATCC 12633, *Mycobacterium aurum* A+ (Medical Microbiology, University of Cape Town) and

Staphylococcus epidermidis ATCC 14990. The fosmid clones from the 96 deep-well plates were stab-inoculated onto the LB agar plates supplemented with 12.5 µg/ml chloramphenicol and 0.02% arabinose, using a 96 pin replicator. The plates were incubated at 30°C overnight and thereafter at room temperature (~25°C) for two to four days to allow the clones to mature. A single colony of each indicator strain was inoculated into 10 ml LB and grown overnight at 37°C overnight with shaking. The optical density of each of the four strains was measured at 600nm and the formula (OD_{600} x µl = 160) was used to calculate the amount of each strain needed for the assay. The appropriate amount of the culture was pipetted into 24 ml soft agar (0.75% agar), mixed well and carefully poured over the stab-inoculated colonies. The library plates were incubated at 30°C overnight before examining the colonies displaying zones of growth inhibition. The plates were then incubated at room temperature for up to seven days while inspecting for the zones of inhibition daily.
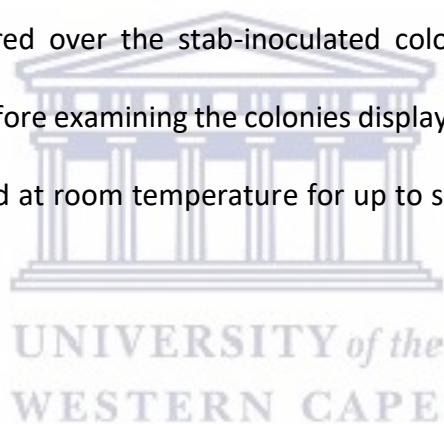
2.13   Sequence-based screening for PKS and NRPS genes

The metagenomic clones in 96 deep-well plates (Section 2.12.1) were screened for the genes encoding for the enzymes responsible for the production of antimicrobial compounds using touch-down PCR (Table 2.5) with degenerate primers designed for the amplification of the NRPS, Type I and Type II PKS genes. The fosmid DNA was extracted from the pooled metagenomic clones (Section 2.5.2) and used as a template for the PCR screening to allow for the screening of thousands of pooled clones in one reaction. Each well consisted of 10 clones and each pool consisted of 960 clones. After running the amplicons on an agarose gel, the bands with the

53

expected sizes were excised, purified (Section 2.5.3) and subcloned in a plasmid (pGEM®-T Easy) vector. The white colonies were picked and colony PCR was performed with M13 forward and reverse sequencing primers to confirm the presence of the inserts. PCR products were purified and sequenced using M13F/M13R primers (Section 2.10) to identify the insert sequences.

For the plates that showed a positive band on the gel after the first screening, two rows (representing 240 clones each) were pooled and rescreened further until the positive clones were detected (Figure 2.3). Clones identified to harbour the gene of interest were selected and inoculated (1:1000) from the glycerol microtiter plates, in LB broth and plated onto LB agar plates supplemented with chloramphenicol (12.5µg/ml) and an induction solution (0.02% L-arabinose). The plates were incubated at 37°C overnight and the colony PCR was carried out with degenerate primers to identify the particular positive library clone. The fosmid DNA was extracted and purified from an identified positive library clone and sequenced. The sequences were analysed using CLC (Section 2.11.1) and antiSMASH (Section 2.11.2).

54
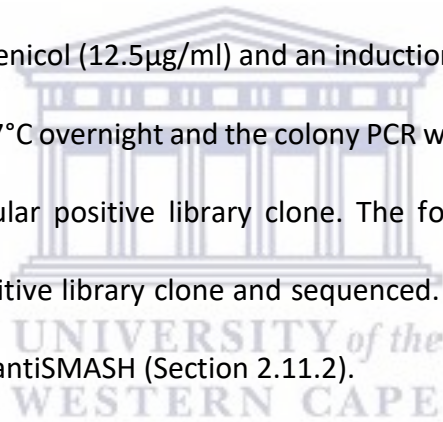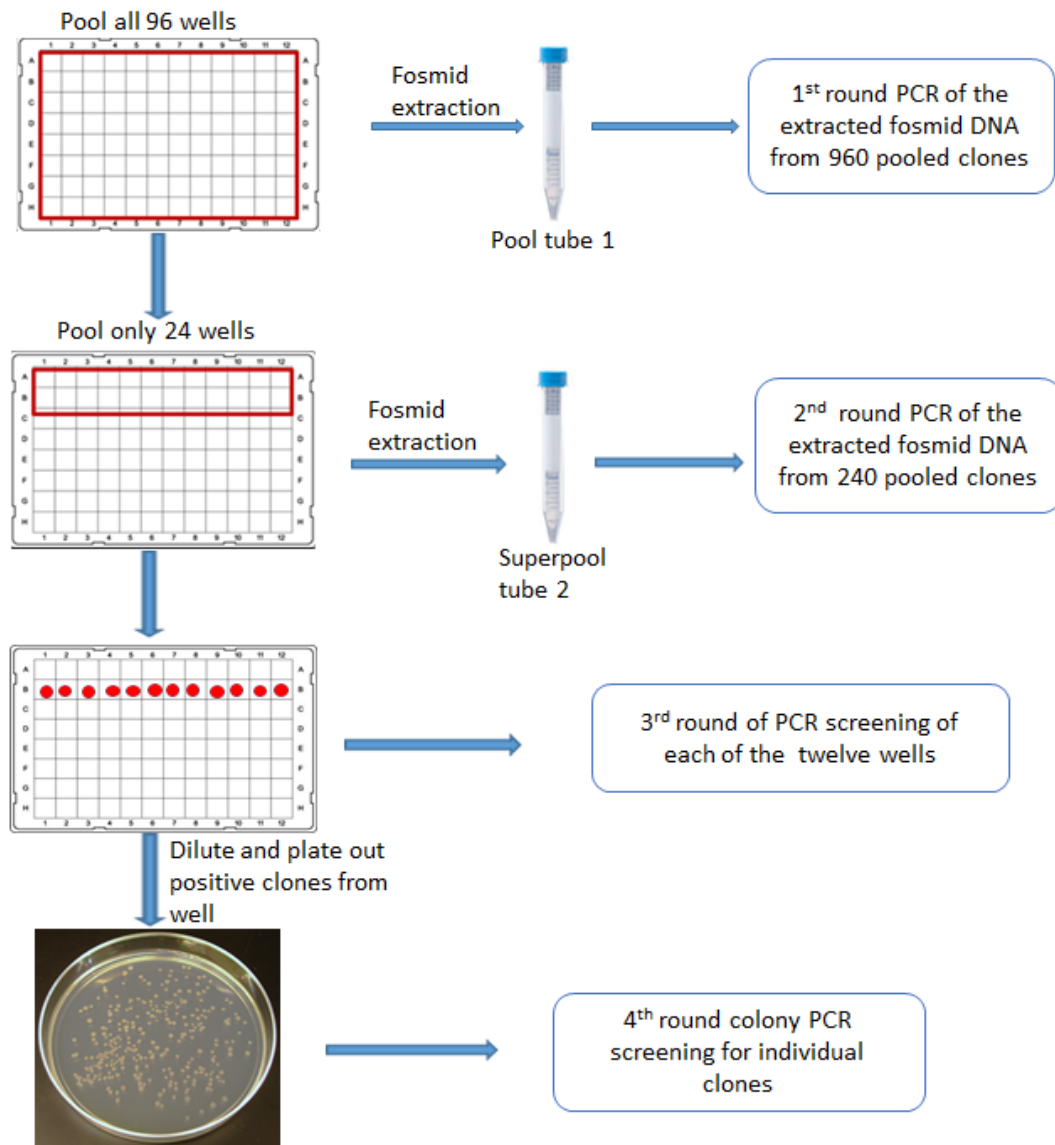
Figure 2.3: Screening of the sponge metagenomic library using pooling strategy for degenerate PCR. Fosmid clones in 96 wells (10 clones per well; 1st round) were pooled to extract DNA that was used as a template for PCR. After observing the positive band on the gel, the plates were pooled in rows of 24 wells (2nd round) and screened further until a positive band was observed. The wells with positive clones were diluted (1:1000) and plated out for further colony PCR screening (4th round) to detect individual positive clones.

Chapter 3

Results and Discussion

3.1   Introduction

As discussed in Chapter 1, metagenomics theoretically provides the means to access the full potential of the functional gene composition and novel biosynthetic pathways within uncultured microbes. Due to the limitations associated with both function- and sequenced-based screening, particularly for secondary metabolites (refer to chapter 1), this study employed both approaches to isolate and identify novel biosynthetic pathways from a marine sponge metagenome.

As part of a previous study in IMBM, metagenomic libraries of the South African endemic sponges, *Spongia* sp. 001RSASPN (PE8) and *Higginsia bidentifera* (PE7) (Figure 2.1), harvested from Algoa Bay were prepared, and formed the material used for screening here. The studies of many other marine sponges have shown complex bacterial communities that produce many pharmaceutically important natural compounds (Li, 2009; Zhang et al., 2009; Hentschel et al., 2012). For example, research has shown that the microbes associated with marine sponges produce medically relevant compounds which include ecteinascidin-743 (Yondelis®), dolastatin-10 (Adcetris®) and halichondrin B (Halaven®) (Webster & Taylor, 2012). For this reason, it was decided to investigate whether the microorganisms associated with these two marine sponges are capable of producing bioactive compounds. According to the culture-dependent work that

56

was conducted at IMBM, out of 184 isolates from the marine sponge PE7, nine showed antibacterial activity against the multi-drug resistant *E. coli* 1699 engineered strain (Cubist, Lexington, MA, USA), while only four had activity against *B. cereus*. Eleven of 210 isolates from the marine sponge PE8 showed activity against *E. coli* 1699, 30 against *B. cereus,* and a small number of isolates showed some activity against *M. smegmatis*, *S. epidermidis* and *P. putida* (Table 3.1). Some isolates from PE8 were further analysed for the presence of NRPS and PKS gene clusters using PCR screening, while PE7 isolates have not yet been analysed (unpublished IMBM data). These results gave some confidence that the bacteria associated with these sponges have the potential to produce bioactive compounds and could thus serve as appropriate material for a metagenomics analysis.
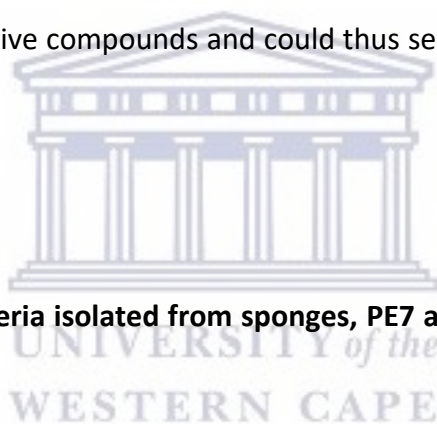
Table 3.1: The number of bacteria isolated from sponges, PE7 and PE8 and their antibacterial activity

Sponge ID	Isolate #	*E. coli* 1699	*M. smegmatis*	*S. epidermidis*	*P. putida*	*B. cereus*
PE7	184	9	2	1	3	4
PE8	210	11	5	8	9	30

The metagenomic libraries were constructed using the CopyControl™ fosmid library production kit following manufacturer's instructions (www.epicentre.com, Madison, WI) by colleagues in the IMBM. The fosmid is normally present in single copy inside the cell, however, can be induced to

57

high copy number by adding arabinose to the growth media, for gene expression and functional screening purposes.

Here, screening focused on isolating and identifying the biosynthetic gene clusters of two classes of compounds of interest, namely NRPs and PKSs. Isolating biosynthetic genes of sponge-derived natural products can be extremely challenging due to the high complexity and density of the sponge symbiotic microorganisms which contain various homologous genes from diverse pathways (Piel, 2010). However, these biosynthetic genes are often clustered on the chromosome, which facilitates their isolation from complex sponge-associated communities using metagenomic approaches (Uria, 2015). The libraries can be analysed to identify the biosynthetic genes of interest using one of two approaches: function-based screening and/or sequenced-based screening.

In function-based screening, the library clones are assessed for their ability to confer antibacterial activity. Such clones could then be interrogated to identify the biosynthetic pathways responsible for the activity produced (Kennedy et al., 2010). Upon identifying a positive clone with the ability to inhibit the growth of an indicator strain, the DNA insert within the specific clone is sequenced to allow for the identification of the genes responsible, using bioinformatics, for the observed phenotype.

The sequence-based approach employed in this project used PCR and degenerate primers designed to conserved regions on the domain of the respective synthases to screen the libraries (Schloss & Handelsman, 2003). This approach has been successfully used to access the complex microbial community and for the discovery and isolation of novel bioactive compounds such as

polyketides and nonribosomal peptides from marine symbiotic bacteria (Nikolouli & Mossialos, 2012). Freeman et al, (2012) isolated a biosynthetic pathway for polytheonamides A and B from novel uncultured bacterial symbionts of the marine sponge *Theonella swinhoei*. These large natural product toxins were shown to be of ribosomal origin and are comprised of nonproteinogenic amino acids (Freeman et al., 2012). Using sequence-based screening with degenerate primers, Piel and his colleagues, isolated and identified a biosynthetic gene cluster for an antitumor polyketide-nonribosomal hybrid onnamide from an unidentified symbiotic bacterium associated with the same marine sponge (Piel et al., 2004). Interestingly, it has more recently been shown that both the polytheonamide and onnamide pathways are produced by the same uncultured bacteria, *Candidatus* "Entotheonella" (Wilson et al 2014).

In these studies, for the amplification of NRPS and PKS genes from the metagenomic libraries, several primer sets were used. The MTF/MTR primers amplify the conserved A domain of NRPS genes. Their design was based on conserved peptide motifs A2 (KAGGAY(LV)) and A8 (QVKIRG) from peptide synthases which are involved in peptide synthesis and other adenylate-forming enzymes from various bacteria and fungi (Vizcaíno et al., 2005). These primers amplify fragments between 750 and 1000bp and were previously used for the detection of NRPS biosynthetic genes from both cultured and uncultured cyanobacterial samples (Neilan et al., 1999); the identification of NRPS from different isolates from the South China Sea sponges (Zhang et al., 2009); and intertidal mudflats (Tambadou et al., 2014).

The A3/A7R degenerate primers amplify a 700-750bp fragment. These primers recognize the conserved regions A3 and A7 in the A domains of actinomycetes NRPSs (Ayuso-Sacido &

Genilloud, 2005). Their specificity to detect the known NRPS gene sequences was tested on the gene cluster from *Amycolatopsis lactamdurans*; and was successfully used to detect a NRPS gene cluster with phylogenetic relation to an actinomycete, when screening the metagenomic library of the *Aplysina aerophoba* sponge (Pimentel-Elardo et al., 2012).

The PKSs contain a KS domain in each module which shows the highest degree of conservation among all domains and is adequate for the investigation of the PKS genes in sponge metagenomes (Atikana et al., 2013). Therefore, the KS domain is well suited for the phylogenetic analyses of PKS gene diversity (Zhang et al., 2009). The KSDPQQF/KSHGTGR primers target the two motifs DPQQ and HGTGT, which are highly conserved among the KS domain of Type I PKSs and have been widely used to study the diversity of PKSs from marine sponge metagenomes (Piel, 2002). Another set of primers used for the detection of Type I PKS genes in this study was KS1F1/KS1R1 (Blasiak and Hill, personal communication). These primers were designed to conserved regions of a protein sequence alignment of >300 KS domains to amplify Type I KS domains from *Streptomyces* sp. These primers were successfully utilised to amplify PKS genes from *Streptomyces* isolates from the tunicate *Pseudodistoma africanum* Millar, 1954 (IMBM, unpublished data).

The KS2aF1/KS2bR1 primer set was designed based on conserved regions of an amino acid alignment of 20 characterized Type II KS gene clusters from actinomycetes (Blasiak and Hill, personal communication). These primers amplify a region across consecutive KS_α and KS_β genes. The size range for this primer set is 1200-1400bp depending on the variable interdomain region extension. These primers were selected because they were successfully used for the

amplification of a Type II PKS gene sequence from a *Streptomyces* isolate cultured from the tunicate *P. africanum* Millar, 1954 (IMBM, unpublished data).

This chapter describes metagenomic screening for biosynthetic (PKS and NRPS) gene clusters from marine sponge metagenomic libraries. The main aim was to isolate large, contiguous regions of DNA from uncultured microorganisms associated with a marine sponge to access intact genetic pathways that are involved in natural product biosynthesis. Two approaches were employed:

1) Functional screening of a marine sponge fosmid library in *E. coli* EPI300 for antibacterial activity. The agar overlay assay was employed, using four different indicator strains: Gram positive; *Bacillus cereus, Staphylococcus epidermidis* and *Mycobacterium aurum,* and a Gram negative; *Pseudomonas putida*.

2) Sequenced-based screening of the metagenomic library clones for novel biosynthetic genes and pathways of the NRPS, Type I and Type II PKS genes by using PCR with degenerate primers designed to amplify a conserved region for the domains of each class as described above.

3.2 Functional screening of the fosmid libraries for antibacterial activity

The metagenomic fosmid library clones of two sponges, PE7 and PE8, were screened for their ability to express putative antimicrobial pathways by detecting their ability to inhibit the growth of various indicator strains using soft agar overlay assays, against *B. cereus*, *P. putida*, *M. aurum* and *S. epidermidis,* and their ability to produce pigmented compounds. The induction solution (0.2% L-arabinose) was added to the LB growth medium to increase the fosmid copy number for the host and to enhance the chances of the expression of the secondary metabolites from the cloned genes in the host and the likelihood of detecting the production of antibiotic activities. However, the disadvantage of this is that if a clone produces a compound toxic to *E. coli* this may result in the loss of clones harbouring the antibacterial activity due to increased compound concentrations. In other words, although maintaining low copy numbers may enable the host to survive toxicity, it might decrease the likelihood of detecting the production of antimicrobial activity (Trindade et al., 2015).
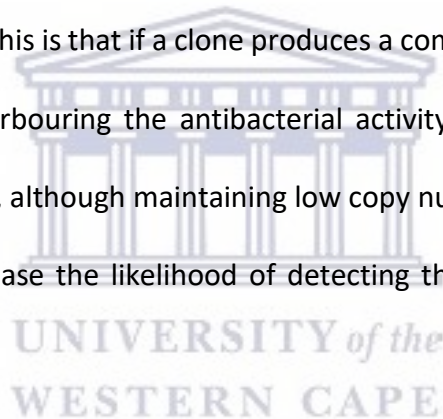
After functionally screening approximately 10,000 clones of each metagenomic library in *E. coli*, no clone was observed that had the ability to produce antibacterial activity against any of the above-mentioned indicator strains or that produced any pigmentation under the experimental conditions employed (Figure 3.1). The plates were incubated at room temperature for up to one week to allow the production of the secondary metabolites, and were examined daily for the formation of the inhibition zones around the library clone colonies. *E. coli* EPI300, containing just the fosmid backbone, was used as a negative control in all the activity assay plates.

The metagenomic library in principle consists of DNA fragments from a large number of marine sponge bacteria. However, there are several reasons why the functional screening was unsuccessful:

i) Given that marine sponges harbour very dense and diverse microbial communities, the 10 000 clones (per sponge) screened does not represent an exhaustive mining of the functional diversity and it is highly unlikely to cover the entire population of bacteria and biosynthetic pathways (Katz et al., 2015).

If we assume that the average bacterial genome is 5 Mb in size, and that there are 1,000 species present in the sponge, this would require a library of 1,000 x 5,000,000bp = 5×10^9bp to cover each genome once. This means that 10,000 clones at an average insert size of 30kb equals 3×10^8bp or enough to cover 60 bacterial genomes once (Gabor et al., 2004). To improve the chances of cloning an intact NRP or PKS pathway, which can range in size from 25kb to 150kb, would require exceptionally high coverage of any one genome. Thus, not only was the number of clones screened inadequate to cover the number of species present, it is also inadequate in terms of the chance of cloning an intact pathway. If the sponge community is not particularly diverse as has been found in the case of so-called low microbial abundance (LMA) sponges (Croué et al., 2013), then the number of clones screened could be considered sufficient (60 fold coverage of 1 bacterial species) to identify such pathways. To answer these questions, a study of the microbial diversity for each sponge would have to be conducted to address whether these are LMA or high microbial abundance sponges (more bacterial diversity). A bioinformatic analysis of the average

63

size of secondary metabolite pathways found in the organisms identified would also have to be undertaken to estimate the chance of cloning an intact pathway using the pCCFOS1 vector system.

ii) Since *E. coli* (EPI300) was used as a host, it is most likely that the strain may not have allowed for the expression of the secondary metabolites from the DNA insert, or that the laboratory conditions were not favourable for the host to express the biosynthetic genes. In this study, a fosmid vector was used (pCC2FOS) which is limited to capturing approximately 30 kb inserts, thus has less chance of containing the full biosynthetic gene cluster necessary for the production of antimicrobial compounds. Many biosynthetic gene clusters for structurally complex metabolites such as polyketides and nonribosomal peptides are larger than what a fosmid vector can accommodate and therefore cannot be expressed functionally in their entirety (Kim et al., 2010). The use of BAC libraries for the functional screening may increase the expression of cloned genes as well as the antibiotic detection in a heterologous host. This can offer a high chance of isolating intact biosynthetic pathways because BAC vectors can allow for screening of large DNA inserts of up to 100 kb and can also be maintained at low copy number (Kakirde et al., 2010).

iii) In some cases the biosynthetic gene clusters remain silent under standard laboratory conditions. The *E. coli* EPI 300 may not contain the necessary machinery for gene expression such as promoter and ribosome binding site recognition, and post-translational processing (Walker et al., 2014). The precursor molecules might be absent in *E. coli* because some enzymes use molecules produced by other metabolic pathways in the cell as building blocks for producing

peptides or polyketides (Milshteyn et al., 2014; Yuzawa et al., 2012). In a similar study conducted by Iqbal et al. (2014), a *Ralstonia metallidurans*-based library was screened using the top agar overlay assay against *B. subtilis*, and resulted in six positive hits. When they transformed the fosmid DNA extracted from the six positive hits into *E. coli* EPI300 and performed the same overlay assays, they could not observe any antibacterial activity. To overcome these limitations, other bacterial host strains such as *Streptomyces, Pseudomonas* or *Bacillus* could be used as expression hosts together with shuttle vectors since they are known to naturally produce many antibacterial compounds (Brady et al., 2010; McMahon et al., 2012; Martinez et al., 2004). This can be achieved by constructing the metagenomic library in *E. coli* as it is the most efficient host because of its high cloning efficiency, and transforming into an alternative host for antibacterial activity or pigmentation screening (Kakirde et al., 2010). For example, a broad host range shuttle fosmid vector such as pCCERI (developed at BIOMERIT, UCC), modified from pCC1FOS (Epicentre Biotechnologies) which can facilitate conjugation into strains such as *P. putida* and *Streptomyces sp.*, can be used for the construction of the metagenomic library. The use of multiple expression hosts increases the chances of foreign gene expression and the likelihood of discovering novel secondary metabolites.

iv) The metabolites are produced in small amounts that may not be able to be detected during functional screening (Chiang et al., 2010).
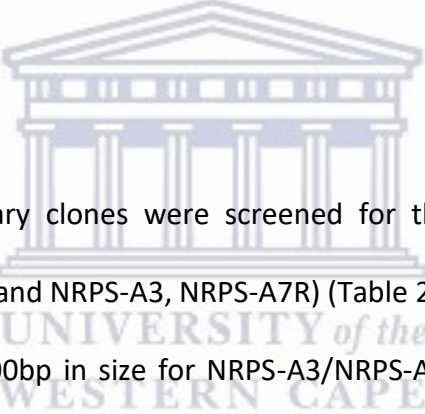
Figure 3.1: Antibacterial activity detection of the sponge metagenomic library clones against indicator strains using the top agar overlay method. The library clones were stab-inoculated onto LB-plates and overlaid with top agar containing indicator strains. No zones of inhibition were observed around the individual colonies.

Due to the expectation that the number of function-based hits would be low, the libraries were also screened in parallel using the PCR-based screening with degenerate primers described above (Table 2.2).

3.3 Sequence-based screening for the discovery of secondary metabolite genes

3.3.1 Identification of NRPS biosynthetic genes

The fosmid DNA was extracted from the pooled fosmid clones representing 960 clones per 96 well microtiter plate. The PCR screening with pools allows for the screening of thousands of clones in one reaction which is highly economic in terms of laboratory consumables, storage space as well as labour (Hrvatin & Piel, 2007), and can also give results with much clearer bands despite low fosmid copy numbers, compared to screening single clones individually (Hrvatin & Piel, 2007).

The sponge metagenomic library clones were screened for the NRPS genes by PCR using degenerate primers (MTF, MTR and NRPS-A3, NRPS-A7R) (Table 2.5). The amplification products of approximately 750bp and 800bp in size for NRPS-A3/NRPS-A7R and MTF/MTR primer sets respectively, were expected after separating the products on the 1% agarose gel. The NRPS positive clone of a *Bacillus* isolate from the marine sponge PE8 was used as a positive control during PCR screening. Of about 20,000 metagenomic fosmid clones that were screened using the above-mentioned primer sets, no target PCR product was observed on the agarose gels (Figure 3.2) even after optimising the PCR conditions by using different annealing temperatures. In another parallel study in the IMBM laboratory, the culture-dependent analysis of symbionts isolated from the PE8 sponge resulted in the detection of NRPS gene clusters in five of the *Bacillus* sp. using the MTF/MTR primers (IMBM, unpublished data). Theoretically the genomes of these

isolates should be represented in the metagenomic library, therefore it was unexpected that no amplicons were obtained. There are several reasons for the absence of amplification products. Firstly, it could be that the NRPS gene clusters were not present or were not well represented within the metagenomic library clones screened. In other words, the NRPS gene might have been contained in a very rare member of the microbe species whose genome was not well represented in the fosmid library.

Another reason for the lack of target gene product detection is that the extraction method might not have been sufficient to lyse the cell membranes of the microbes that contain the target NRPS gene clusters, as the single extraction method can underestimate the total number of bacterial cells present (Singh et al., 2009; Ekkers et al., 2012). The metagenomic library may only represent a small fraction of the microbial genomes present in the marine sponges and thus not reaching all the low abundance microorganisms. This could explain the detection of NRPS genes by culture-based screening and the lack of NRPS gene detection using the metagenomic screening of the marine sponge. The culture-independent approach can miss out on genes that can be identified using the culture-dependent approach. For example, Kim and Fuerst were able to detect the PKS gene responsible for the production of rifamycin using culture-dependent approach but could not detect the same gene when using the metagenomic approach (Kim & Fuerst, 2006).
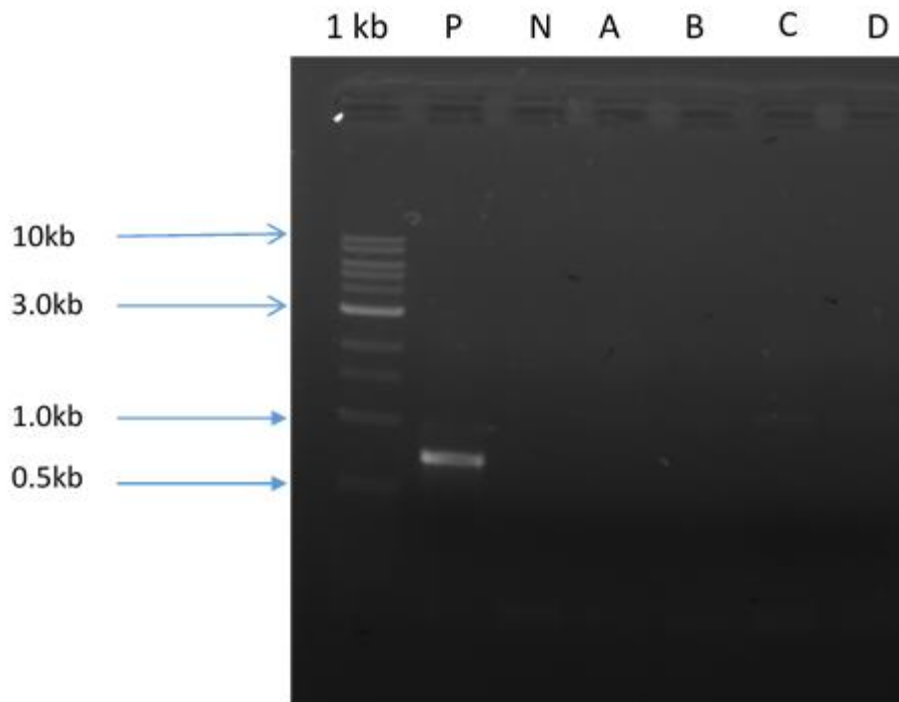
Figure 3.2: The amplification of NRPS genes from the sponge m-library fosmid DNA. The expected product size of approximately 750bp was observed for the positive control on 1% TAE agarose gel. Lane 1: 1000bp DNA molecular marker. Lane 2: Positive control (a NRPS fragment from a bioactive isolate associated with *Spongia* sp. 001RSASPN). Lane 3: Negative control. Letters A to D represent the pooled fosmid clone DNA extracted from four parts of a 96 well plate, each number representing 240 clones).

3.3.2 Identification of PKSII biosynthetic genes

Bacterial Type II PKSs contain a set of three genes; the alpha ketoacyl-synthase (KS_α and KS_β) subunits which catalyse the condensation between the acyl thioester species and the growing carbon skeleton; and a chain length factor (Seow et al., 1997; Metsä-Ketelä et al., 1999). The KS_α can be used as a marker for the screening of Type II polyketide compounds (Sun et al., 2012).

To isolate a fosmid clone with a Type II PKS domain, PCR screening with degenerate primers (KS2aF1/KS2bR1) was carried out with the fosmid DNA from 96-well pools as a template. The anticipated PCR product size of approximately 1200bp was observed and excised from the agarose gel (Figure 3.3). The cleaned PCR products were cloned into the pGEM-T Easy vector and transformed into electrocompetent *E. coli* cells. Two transformants were selected and the plasmids were sequenced using M13 sequencing primers to identify the PCR product cloned. The nucleotide sequences of the selected clones were matched with proteins (BLASTx) in the NCBI nr database. The deduced amino acid sequences were both shown to be closely related to those of the NADH-quinone oxidoreductase subunit M of *Pseudomonas stutzeri* with 94% identity. No presence of the KS genes was detected from the library clones as there were no amino acid sequences identified that resembled those of Type II KS domains. For this reason, no further dereplication of the fosmid clones was necessary and the fosmid clones were not analysed further.
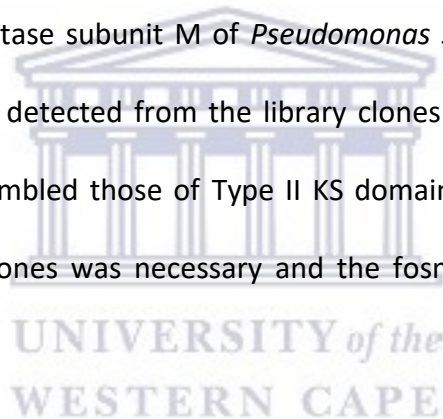
These are false positive fragments as the amplification did not result in KS-specific domain products. According to homology analysis of the fragments, there could have been some nonspecific annealing of the degenerate primers to a non-target locus in the metagenomic DNA at low stringency conditions (Deng et al., 2010). The nonspecific annealing becomes even higher when amplifying fragments from GC-rich genes due to their complex secondary structures which results in low percentage of amplification success (Deng et al., 2010). The addition of DMSO or glycerol in the reaction can also help to facilitate specific amplification in the template, especially for the GC-rich genes (Deng et al., 2010).

Increasing the annealing temperature can enhance specific amplification and prevent non-specific annealing of the primers (Strien et al., 2013). However, using mixed primers can require a decrease in annealing temperature, which leads to nonspecific amplification. A semi-nested touchdown PCR, using two sets of primers, can be useful in enhancing the specificity and product formation by the amplification of the extended nucleotide sequence followed by the amplification of the second region located within the first amplification (Chen et al., 2014), which can also help eliminate the false positive amplification.
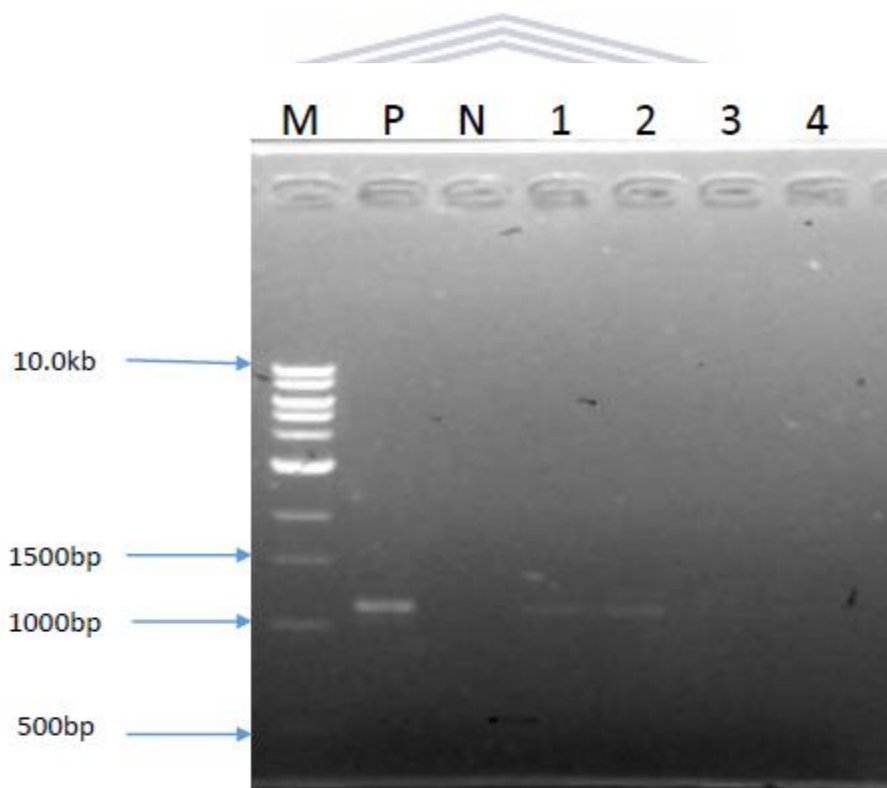
Figure 3.3: The amplification of Type II PKS genes from the sponge m-library fosmid DNA. The expected product size of approximately 1200bp was observed on 1% TAE agarose gel. Lane M: 1000bp DNA

molecular marker. Lane P: Positive control, PE13-181 (a Type II KS positive *Streptomyces* strain isolated from a tunicate, *Pseudodistoma africanum*, PE13). Lane N: Negative control (nuclease-free water as a template). Lanes 1-4: Pooled fosmid clone DNA extracted from four parts of a 96 well plate, each number representing 240 clones).

3.3.3 Identification of Type I PKS biosynthetic genes

Type I PKS genes were screened by using the two sets of degenerate primers (KSDPQQF/KSHGTGTR and KS1F1/KS1R1) for the amplification of the highly conserved regions of ketosynthase (KS) domains (Table 2.5). No amplification was obtained after screening the PE7 and PE8 metagenomic library clones with the KS1F1/KS1R1 primers. However, amplification was observed after PCR-screening about 10,000 clones of each sponge metagenomic library with the KSDPQQF/KSHGTGTR primers. The amplicons appeared as a single band of the correct size after gel electrophoresis (Figure 3.4) from the first pool PCR.

After obtaining the amplification products from the first pool (960 clones per 96 well microtiter plate) using the KSDPQQF/KSHGTGTR primers, the amplicons were subcloned and sequenced to confirm the presence of the KS sequence. This led to the rescreening of the super-pools which were prepared from two rows each of the same positive plate, representing about 240 clones per reaction. The fosmid DNA of each of the wells the two rows that showed the presence of a band on the gel was later extracted and used as a template for rescreening. The well with a positive band was then diluted to 1:1000 and screened further with PCR to isolate a single positive fosmid clone.

72
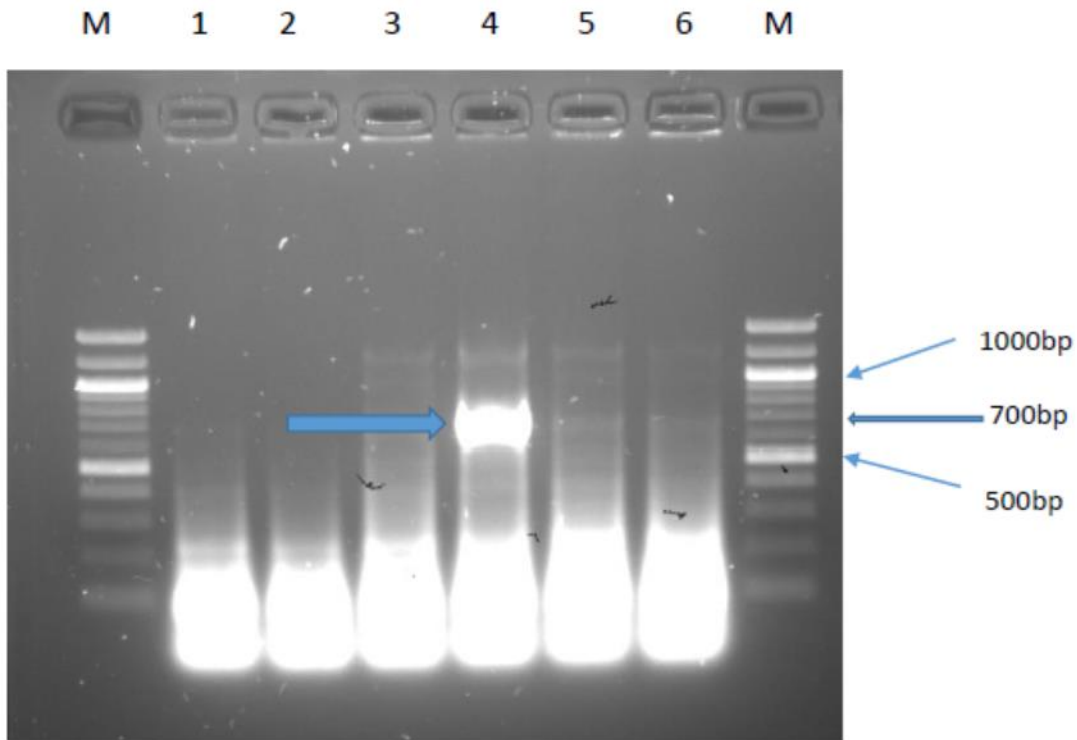
Figure 3.4: The PCR amplification of the KS genes of de-replicated fosmid pools. Lane M: 100pb DNA molecular marker (ranges from 100-1517bp). Lane 1-6: Fosmid library clones of two clone pools from wells E7-E12 of microtiter plate number 74 and 75. E10 (indicated by an arrow on lane 4) is a positive fosmid clone with the expected product size of approximately 700bp.

The Type I PKS PCR amplification products of approximately 700bp in size which were consistent with the expected size, were sub-cloned into a pGEM-T Easy vector and transformed into electrocompetent *E. coli* EPI300 cells. Plasmid inserts from two randomly selected transformants were sequenced using the M13 forward and reverse sequencing primers to ensure that the

amplicons were of the intended gene targets. The sequence analysis is presented in Table 3.2 and Figure 3.5. The different percentage identity of the KS genes to the NCBI hits suggests the presence of different PKS genes produced by a sponge symbiont. The sequence alignment showed differences in length and homology.

Table 3.2: Sequence analysis of the two pGEM-T Easy clones containing the putative Type 1 PKS amplicons resulting from screening the PE7 m-library with the KSDPQQF/KSHGTGTR primers.

Clone	Top BLAST hit	Query Coverage	E- Value	Identity	Query length (bp)	Hit Length (aa)	Accession number
1	Polyketide synthase, partial [*Cobetia marina*]	95%	8e-139	92%	667	227	AFY62963.1
2	Putative PKS [*Cobetia marina*]	98%	7e-72	86%	382	227	AFY62963.1
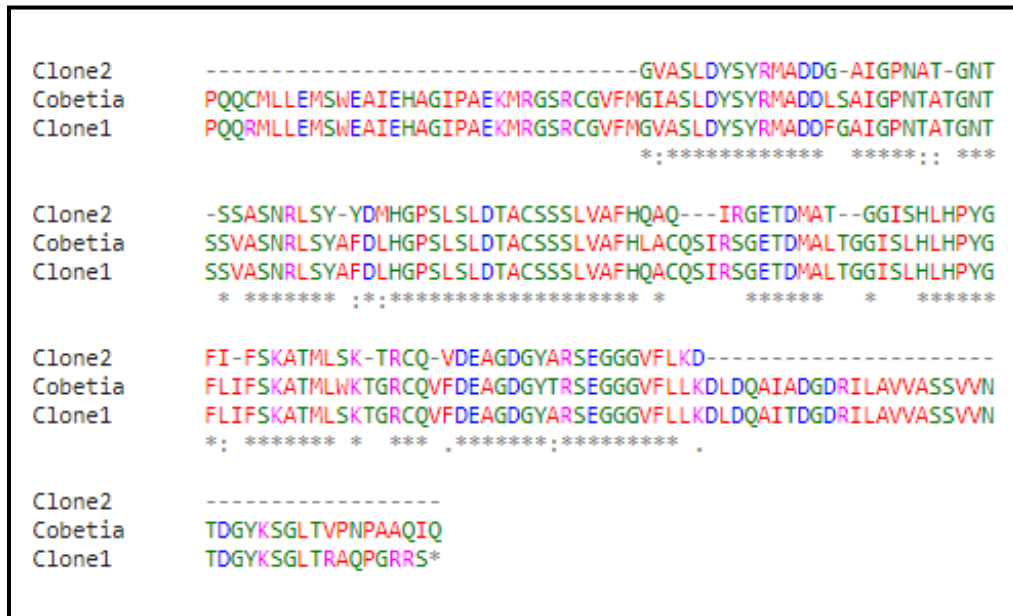
Figure 3.5: The alignment of the putative amino acid sequence of the KS amplicon detected in the sponge m-library against the closest result from the BLASTx analysis, a putative PKS from *C. marina* (Accession number AFY62963.1). The KS gene sequences were translated using sequence manipulation suite (SnapGene3) and aligned using Clustal Omega (www.ebi.ac.uk/Tools/msa/clustalO). KS clone1 and clone2 refer to the two KS positive clones with fragments from the sponge m-library.

BLASTx analysis of the two putative KS amplicons revealed that they were amplified from the PKS modules, possibly from the same pathway or alternatively different pathways. The DNA sequences of the PCR products were found to have high amino acid sequence identity (95% and 98%, respectively) to the PKS domain sequence from a marine Gammaproteobacterium, *Cobetia marina,* previously isolated from a *Mycale* sp. marine sponge from the coast of Fujian, China (Table 3.2). This suggests that a highly similar PKS pathway might also present in a marine sponge-specific symbiont present in the *Higginsia bidentifera* sponge in Algoa Bay, South Africa.

75

Having confirmed the presence of Type I PKS genes in the metagenomic library, the recovery of the fosmid clones which carry the biosynthetic gene clusters was pursued. Clone 1 above represents part of the fragment from the fosmid clone PE7-E10.

Based on the results of the PCR product sequence analysis, the positive pool (representing 1000 clones) was screened further (by dereplication of wells E7-E12 of plates 74 and 75 using the KSDPQQF/KSHGTGTR primers) to identify the fosmid/s which harbour the KS domains. A positive hit was observed from a pool that was representative of two fosmid clones from plates 74 and 75 (Figure 3.4). The two clones (plate 74 well E10 and plate 75 well E10) were screened separately to identify which one the amplicon originated from. The fosmid in the E10 well in plate 75 (Figure 3.6) was found to be positive and was selected for further analysis.

76
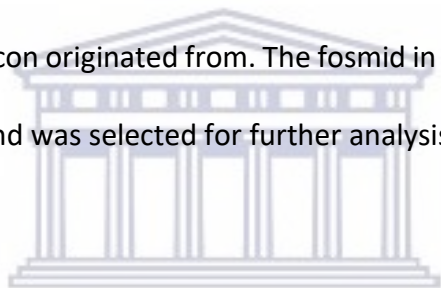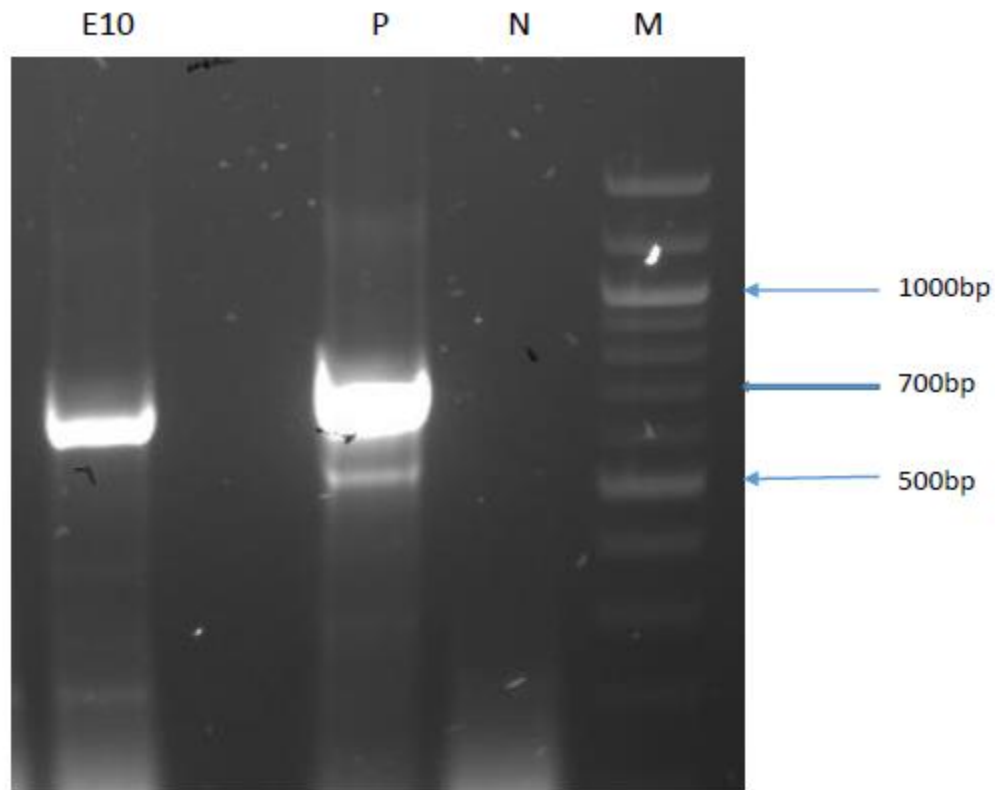
Figure 3.6: An agarose gel electrophoresis (1%) of the amplification (touchdown PCR) of a positive clone with KSDPQQF/ KSHGTGR primers and a product size of approximately 700bp. Lane E10: Positive clone from well E10 of the microtiter plate number 75. Lane P: Positive control (genomic DNA of *Thalassomonas actinarium*, a marine bacterium known to encode Type I PKS pathway). Lane N: Negative control (nuclease-free water as a template). Lane M: 100bp DNA molecular marker.

3.3.3.1 Fosmid Sequencing and bioinformatics analysis

The KS-positive fosmid clone (PE7-E10) was sequenced using next generation (MiSeq) technology as described in Section 2.6.5. The fosmid sequence generated a total of 42390bp. The sequence

reads were trimmed and CLC Genomics Workbench 4.9 was used for *de novo* assembly. Following the vector trimming, the length of the insert was approximately 34209bp.

A discontiguous megaBLAST (BLASTn) analysis of the fosmid PE7-E10 sequence revealed that a large part of the fosmid sequence (with 53% query cover) showed 84% sequence identity to the marine bacterium *C. marina* strain JCM 21022, a Gammaproteobacterium of family *Halomonadaceae* (Table 3.3) (Tang et al., 2017). This suggests that the insert could possibly be from a bacterium that is closely related to *C. marina* species, which was associated with the South African marine sponge *Higginsia bidentifera*.

Table 3.3: Top hits from BLASTn analysis of the fosmid PE7-E10 clone sequence.

Species hit	Query cover	Identities	% Identity	E-value	Accession Number
C. marina strain JCM 21022, complete genome	53%	4689/5571	84	0.0	CP017114.1
H. chromatiredurens strain AGD 8-3	5%	1180/1420	83	0.0	CP014226.1

The fosmid PE7-E10 insert sequence was annotated and shown to contain thirty-three putative open reading frames (ORFs) (Figure 3.7). Twelve of these ORFs show high similarity to biosynthetic genes and one to a regulatory gene (Figure 3.8, Table 3.4). When mined for secondary metabolite biosynthetic genes using antiSMASH version 3.0.5, the presence of a Type I PKS gene cluster, containing a single KS module, was revealed.
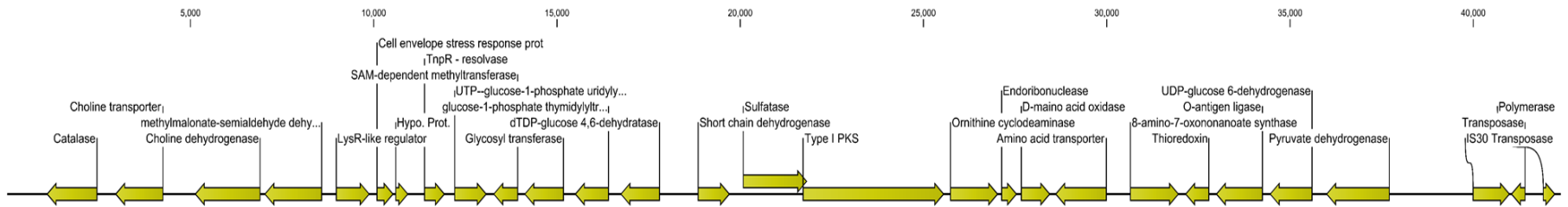
Figure 3.7: The gene cluster organisation of the fosmid clone PE8-E10.
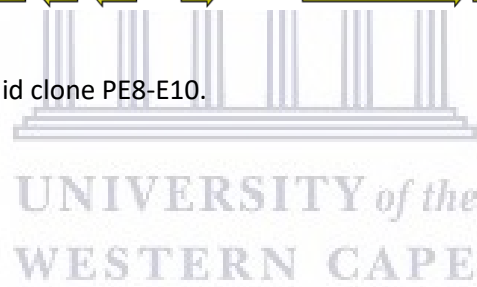
By structural and functional analogy, Type I PKS modules consist of three main catalytic domains (AT, KS and an ACP). The AT domain incorporates the malonyl or methylmalonyl-CoA and transfers the activated substrate to the ACP domain, while the KS domain forms the C-C bond of the elongated chain (Amoutzias et al., 2016). One of the most immediate observations was that the polyketide synthase was approx. 1500 amino acids shorter than the synthases it showed highest sequence identity to. On further inspection, it was observed that the acyl carrier protein (ACP) domain was not present in the synthase. This has not yet been seen within other Type I PKS gene clusters, although it was observed for Type II PKS gene clusters. Shen 2003 established that there are novel Type II PKSs that can be ACP-independent, such as the macrotetrolide gene cluster. These PKSs that lack an ACP domain can directly use acyl CoAs as substrates to catalyse both C-C and C-O bonds for the production of macrotetrolide (Shen, 2003). Before exploring whether this PKS represents a novel Type 1 gene, as had been demonstrated for the Type II PKSs, further sequence analysis was performed.

In any cloning process, there is always the possibility of a chimera. A chimera is represented by the fusion of two pieces of DNA that are not naturally contiguous. In the case of a metagenomic library, this could, in theory, occur with fragments from two genetically distinct species, or even from the same species, especially if it is a dominant phylotype (Liu et al., 2014). To investigate whether a chimera had been generated and thereby resulting in the shorter PKS gene, a BLASTp analysis of the ORF sequences that resulted from antiSMASH analysis (Figure 3.8) was performed. The analysis indicated that the genes on this cluster have high sequence identities with the protein sequences of PKS gene clusters from marine *Cobetia* species, sharing up to 100% (Table

81

3.4). The fosmid PKS gene which encodes for KS-AT-DH is smaller in size than that of *Cobetia* species which encodes for KS-AT-DH-ER-KR-ACP, according to antiSMASH (Figure 3.9). A comparison between the PKS gene cluster of fosmid PE7-E10 with the PKS gene clusters of other *Cobetia* species (Figure 3.10) indicated a high degree of conservation in the synteny and sequence of some of the genes up-and downstream of the PKS. This immediately ruled out the possibility that the fosmid insert is a chimera. Interestingly, ORFs19 (truncated PKS) and 24 (8-amino-7-oxononanoate) are present in all the *Cobetia* species; however, unlike for all the *Cobetia* genomes, these two genes on the fosmid are separated by 4 additional genes (ORFs20-23) and show high sequence identity to a number of other Gammaproteobacteria, and not *Cobetia* (Table 3.4). It was also interesting to note the presence of two transposase genes (*tnpA* and *tnpR*) on the fosmid cluster that are not on the other *Cobetia* species. Collectively these findings suggest that there has been some form of gene rearrangement in the genome of the symbiont that the fosmid gene cluster originates from, possibly through a mobile genetic element that was acquired from another microbe through horizontal gene transfer (HGT) (Mizuno et al., 2013). HGT allows bacteria to evolve and be able to adapt to the symbiotic lifestyle in the environment (Giordano et al., 2015). Through this, it is possible that a part of the polyketide synthase gene which contains ER, KR and ACP domains in the genome of *Cobetia* sp. UCD-24C, *C. amphilecti* and *C. marina* JCM21022, was possibly deleted in the genome of the symbiont that the fosmid gene cluster originates from. Furthermore, additional genes (ORF24 and ORFs10-15) were observed within the fosmid gene cluster on either side of the PKS gene which are part of the O-antigen biosynthetic gene cluster (Iguchi et al., 2015). The O-antigen pathway is responsible for the production of the polysaccharides which form part of the lipopolysaccharides that are present in

82

the outer membrane of the Gram negative bacteria and are essential for their survival in the environment or their host organism (Iguchi et al., 2015). It therefore appears that a rearrangement and/or HGT event between an O-antigen-like biosynthetic pathway, a PKS-like gene cluster and other Gammaproteobacteria genes has occurred in the genome from which the fosmid originates.

Since the PKS operon on the fosmid is not complete, but is virtually identical, at the nucleotide level, to the operon present on the *Cobetia* sp. UCD-24C genome, all further analyses were conducted for this operon.

Prediction Informatics for Secondary Metabolomes (PRISM) web application that was designed to efficiently identify biosynthetic gene clusters and predict the structures of the encoded NRPs and Type I and II PKs, as well as the substrates for the enzyme domains (Skinnider et al., 2015), was used to analyse the gene cluster from *C. marina* JCM 21022 (49bp). PRISM predicted propionic acid as the substrate used by this PKS gene cluster, with the highest confidence (Figure 3.11). A methylmalonate semialdehyde dehydrogenase was observed located upstream of the fosmid PKS I gene (Figure 3.7). This gene encodes the enzyme that catalyses the NAD-dependent oxidation of 2-methyl-3-oxopropanoate and CoA to produce propionyl-CoA (Talfournier et al., 2011). Propionic acid is produced biologically as its coenzyme A ester, propionyl-CoA, from the metabolic breakdown of fatty acids. However, this gene was not observed in the *Cobetia* sp. UCD-24C genome, but it was seen on the *C. marina* JCM 21022 (results not shown). This suggests that the PKS pathway from some *Cobetia* species could possibly be responsible for the production of polyketide compounds using propionyl-CoA as a precursor.
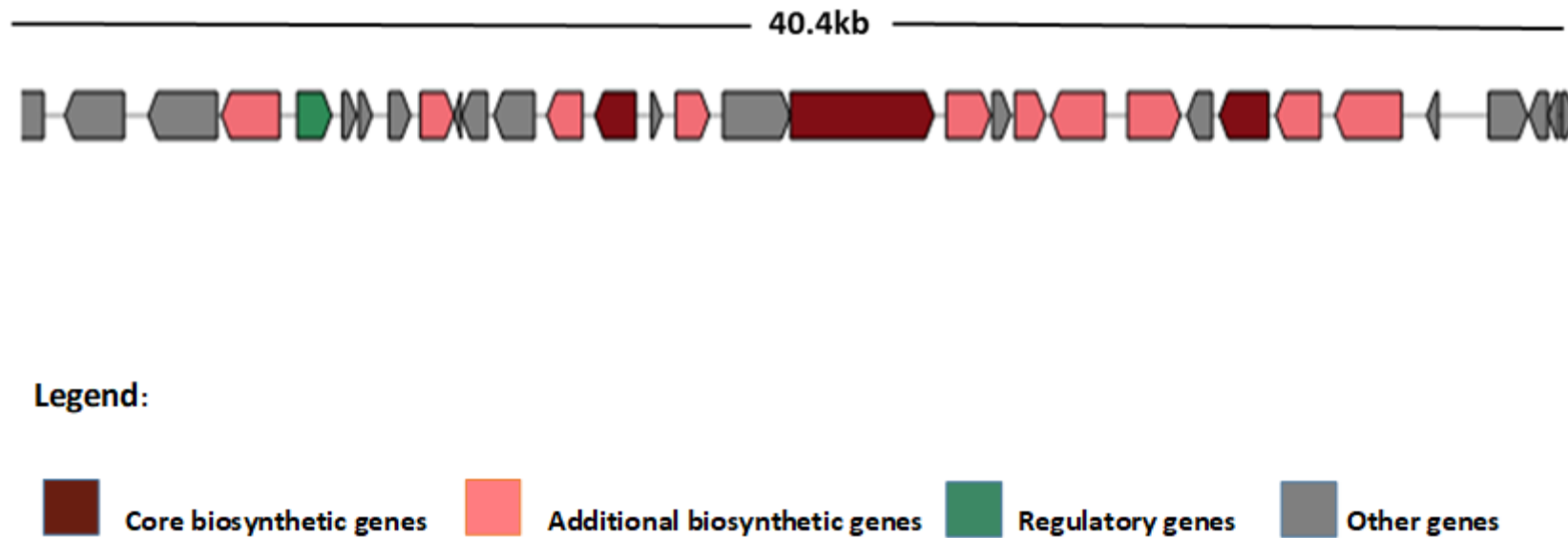
Figure 3.8: antiSMASH analysis of the sequenced insert of fosmid clone PE7-E10. The insert shows the molecular organisation of a Type 1 PKS biosynthetic pathway, when compared with related clusters found in other bacterial genomes. The PKS biosynthetic gene cluster consisted of 33 ORFs, indicated above with different colours according to their functions as described in the legend. The putative ORFs were further analysed in Table 3.4.

Table 3.4: The top BLASTx hits of predicted ORFs of the fosmid clone and their proposed functions.

ORFs	Amino acids	Strand	Proposed Function	Closest Relative	Amino Acids (Hit)	Query cover	Identity (%)	Accession number
2	439	-	Catalase	*Cobetia* sp. UCD-24C	505	98%	98	WP_054557206.1
3	512	-	Choline transporter	*Cobetia* sp. UCD-24C	535	100%	99	WP_054557121.1
4	596	-	Choline dehydrogenase	*Cobetia* sp. UCD-24C	553	93%	97	WP_05455203.1
5	505	-	Methylmalonate-semialdehyde dehydrogenase (acylating)	*Cobetia amphilecti*	505	100%	95	WP_043336175.1
6	298	+	LysR family transcriptional regulator	*Cobetia crustatorum*	298	100%	98	WP_024952721.1

7	122	+	Hypothetical protein	*Cobetia crustatorum*	122	100%	100	WP_024952722.1
8	114	+	Hypothetical protein	*Cobetia sp.* UCD-24C	114	100%	95	WP_054557115.1
9	187	+	Hypothetical protein	*Cobetia crustatorum*	187	100%	87	WP_024952724.1
10	294	+	UTP-glucose-1-phosphate uridylyl transferase	*Cobetia sp.* UCD-24C	294	100%	95	WP_054556681.1
12	224	-	SAM-dependent methyltransferase	*Cobetia sp.* UCD-24C	213	95%	94	KMP75567.1
13	350	-	hypothetical protein	Cobetia crustatorum	350	100%	96	WP_024950971.1
14	308	-	Glucose-1-phosphate thymidyl transferase	*Cobetia sp.* UCD-24C	308	100%	99	WP_054557113.1
15	357	-	dUTP-glucose 4,6-dehydratase	*Cobetia sp.* UCD-24C	357	100%	99	WP_054557112.1

16	92	+	HTH-like domain-containing protein	*Bradyrhizobium sp.* err11	279	57%	43	SCB55847.1
17	291	+	Short-chain dehydrogenase/reductase	*Cobetia sp.* UCD-24C	251	86%	98	WP_054557327.1
18	583	+	Hypothetical protein	*Cobetia crustatorum*	583	100%	97	WP_051510682.1
19	1246	+	Type I polyketide synthase	*Cobetia sp.* UCD-24C	2758	100%	97	WP_054557323.1
20	385	+	Ornithine cyclodeaminase	*Rouxiella chamberiensis*	379	97%	81	WP_045046927.1
21	156	+	Hypothetical protein	*Bacillus cereus*	141	90%	99	WP_060757431.1
22	265	+	D-amino-acid oxidase	Gamma proteobacterium (18)	431	95%	74	WP_027977722.1
23	467	-	Amino acid permease-associated protein	*Halomonas sp.* KM-1	467	100%	73	WP_010627551.1

87

24	457	+	8-amino-7-oxononanoate synthase	*Cobetia sp*. UCD-24C	444	96%	96	WP_054557326.1
25	216	-	thioredoxin	*Cobetia sp.* UCD-24C	216	100%	95	WP_054557317.1
26	425	-	Hypothetical protein BFX80-06595	*Cobetia marina*	414	95%	82	AOM02895.1
27	388	-	UDP-glucose 6-dehydrogenase	*Cobetia sp*. UCD-24C	388	100%	97	WP_054557315.1
28	576	-	Pyruvate dehydrogenase	*Cobetia* marina	576	100%	94	WP_054556154.1
29	98	-	transposase	*Cobetia crustatorum*	336	71%	83	WP_024951836.1
30	338	+	IS30 family transposase	*Cobetia amphilecti*	338	100%	96	WP_043337400.1
31	163	-	transposase	*Cobetia crustatorum*	275	80%	87	WP_024953295.1

32	95	-	transposase	*Cobetia crustatorum*	95	100%	92	WP_024953294.1
33	112	+	polymerase	*Cobetia marina*	416	100%	56	AOM02885.1
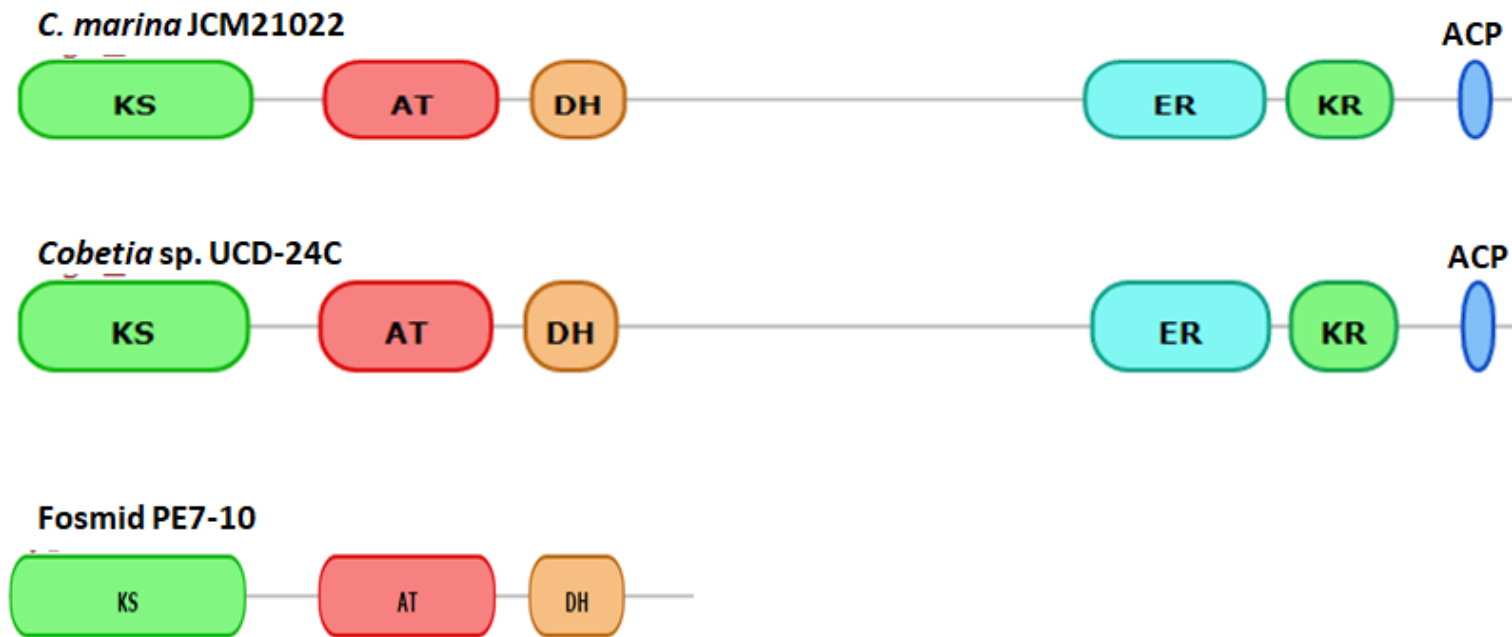
Figure 3.9: Comparison of the PKS domain from the fosmid PE7-E10 with those from *C. marina* JCM21022 and *Cobetia* sp. UCD-24C from the GenBank, which

were shown to be the closest hits after BLASTn and BLASTx analysis, respectively. The fosmid PE7-E10 PKS domain is smaller than that of two *Cobetia* species

which have three other domains (ER-KR-ACP), which were not observed on fosmid PE7-E10.
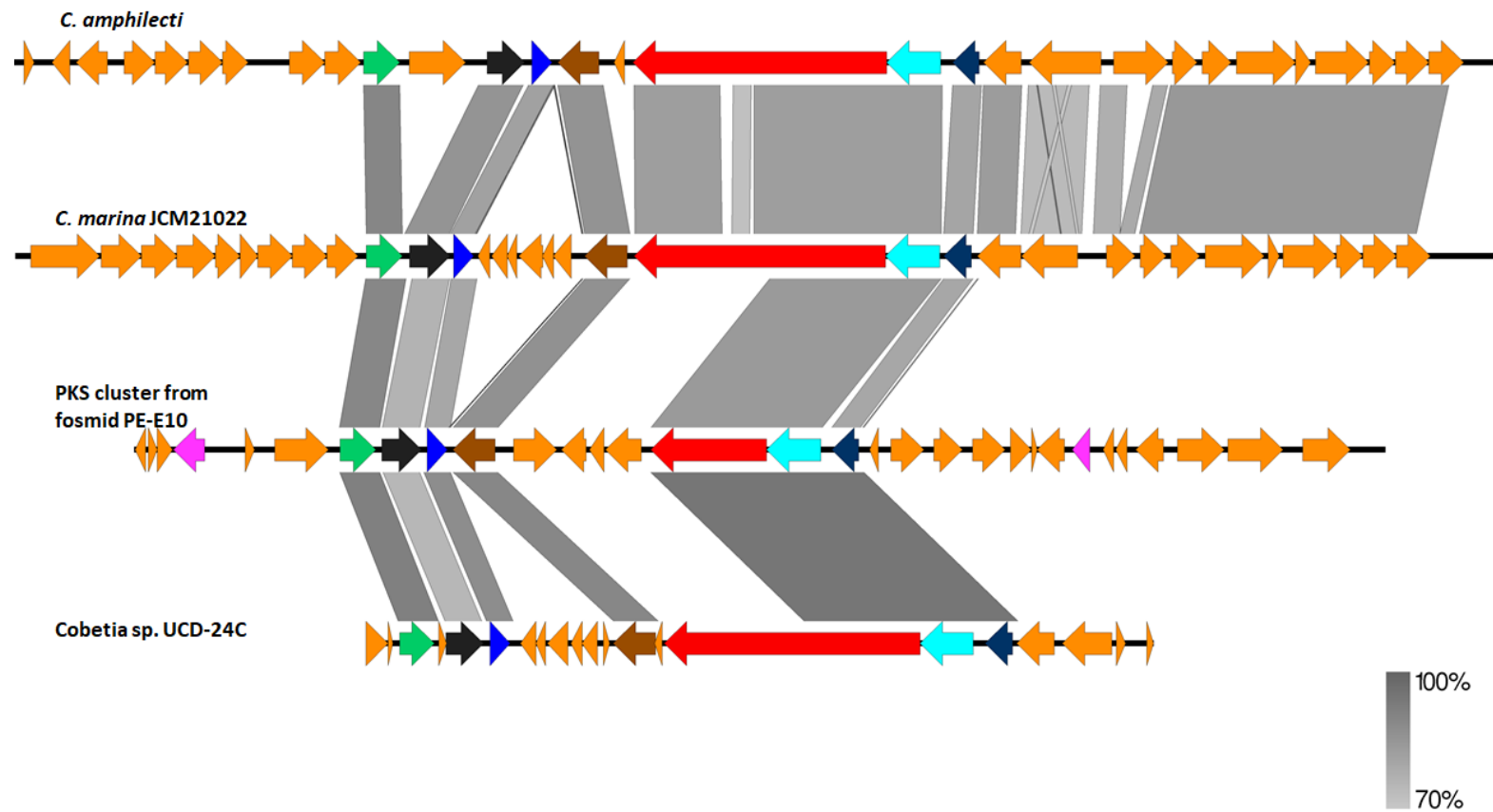
Figure 3.10: Comparison of the PKS gene cluster of fosmid PE7-E10 from the sponge m-library with a section of PKS gene cluster of *Cobetia* sp. UCD-24C, *C. amphilecti* and *C. marina* JCM21022 generated with Easyfig. Annotation: Green – nucleotide sugar dehydrogenase; Black – O-antigen ligase domain-

containing protein; Blue – DsbA family oxidoreductase; Brown – 8-amino-7-oxononanoate synthase; Red – Type I PKS; Light blue – hypothetical protein; Dark

blue - short-chain dehydrogenase; light purple - transposase genes; Orange – other ORFs identified by antiSMASH; are shown in (*Tnp*R and *Tnp*A).

Table 3.5: The possible products for the biosynthetic pathway from the *C. marina* JCM 21022, according to PRISM. Propionic acid is proposed to be the precursor used for the product biosynthesis by this pathway with the highest confidence.

Name(s)	Scaffold	ECFP6	ECFP6	SMILES
Propionic acid	Scaffold_1	1.00	1.00	O=C(CC)O
Butan-2-one	Scaffold_1	0.54	0.21	O=C(C)CC
2-Oxobutananoic acid, 3-methylpyruvic acid, 2-Ketobutyric acid, FEMA 3723	Scaffold_1	0.53	0.50	CCC(=O)C(O)=O
n-Butyric acid	Scaffold_1	0.53	0.64	O=C(CCC)O
Malonic acid	Scaffold_1	0.46	0.54	O=C(O)CC(O)=O

3.4   Conclusion

The results show the presence of a potential incomplete PKS gene cluster with one module that consisted of three domains, namely, KS, AT and DH. It was shorter than that of other *Cobetia* species that were compared with it, which had all six domains (KS, AT, DH, ER, KR and T). This suggests that the fosmid pathway originally looked like the PKS biosynthetic pathway of *C. marina* species before it had some domains (ER, KR and ACP) deleted possibly during a transposition event. Therefore, this observed pathway likely does not produce a polyketide in the originating host.

This study again demonstrated the significance of using the sequence-based methods, which included sequencing and bioinformatics analytical programs such as antiSMASH and PRISM, in the discovery of biosynthetic gene clusters from metagenomic libraries. These methods are

cost-effective and not time consuming as one is able to screen more than one clone at a time by using pooled PCR with primers which can specifically recognize the target biosynthetic pathway. The presence of a PKS positive fosmid clone indicates that the culture-independent techniques can be successfully used to mine for the biosynthetic genes from sponge-associated microbiota. However, it could not be established whether there is a full-length version of the pathway or if the microbes associated with the marine sponge *Higginsia bidentifera* have the capacity to produce secondary metabolites that can be used for drug discovery. More metagenomic studies need to be conducted in order to uncover symbiotic bacterial species as well as novel biologically active natural products. This is in contrast to the culture-based methods that tend to enrich for those microorganisms that produce compounds which are already known and are abundant in nature.
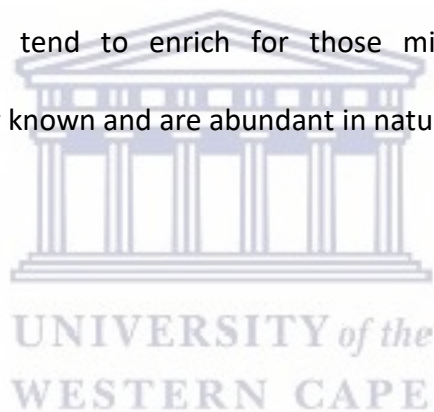
Chapter 4

General Discussion and Final Conclusion

Multi-drug resistance by most human pathogens has been identified by the WHO as one of the greatest threats to human health. Antimicrobial resistance reduces the possibility of effectively treating infectious diseases and leads to increased human mortality (Qin et al., 2011).

Marine invertebrates such as sponges, bryozoans and tunicates are a rich source of secondary metabolites with diverse structures, most of which are produced during secondary metabolism as a defence mechanism against predators and biofouling because they are sessile animals. These secondary metabolites are of importance in the pharmaceutical industry for their broad range of applications as antibacterial, antifungal, antimalarial, antitumor and antiviral agents (Barone et al., 2014). However, their development as pharmaceutical drugs has been limited by the difficulty in achieving a sustainable supply and the low amounts of these natural products in the invertebrate's tissue (Kim, 2015; Singh et al., 2009). Most of these marine NPs are nonribosomal peptides and polyketides that are synthesised by the large nonribosomal peptide synthases (NRPS) and polyketide synthases (PKS), respectively. Evidence shows that many of these NPs are produced by the associated symbiotic bacteria instead of the invertebrates themselves (Uria & Piel, 2009). Studying the

95

NRPS and PKS enzymes can lead to the evolutionary events that lead to their diverse structures and also help to discover new pharmaceutical natural products (Hochmuth & Piel, 2009b).

However, the complexity and the difficulty to cultivate a wide diversity of the microbial symbionts have hindered the attempt to access the biosynthetic pathways of these NPs and their biotechnological potential (Kurnia et al., 2017). This has led to the development of the culture-independent approaches in order to access complex microbial communities.

Studies have shown that some biosynthetic genes discovered using culture-based screening were different from those that were retrieved using culture-independent approaches of symbiotic bacteria (Kim & Fuerst, 2006). This shows that the culture-independent approach in combination with the culture based approach, holds a promise for better and rapid exploitation of the microbial chemical diversity associated with marine sponges. Metagenomic screening and DNA sequencing technologies have allowed researchers to gain access to these microbes and to understand the biosynthetic pathways of natural products or bioactive secondary metabolites they produce.

The main aim of this study was to detect and study the biosynthetic genes responsible for the secondary metabolism natural products from the uncultured symbiotic bacteria of two South African marine sponges using metagenomic approaches. Two screening approaches, namely function based and sequence-based, were used in this study for novel genes or gene product discovery.
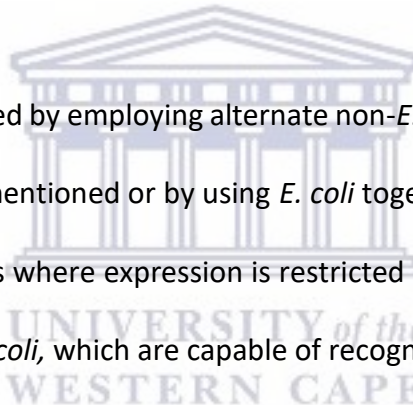
Functional metagenomics is a potentially powerful approach for screening libraries for the identification of bioactive natural products encoded by environmental DNA in a high-

throughput manner (Milshteyn et al., 2014). It allows for the recovery of complete gene clusters for NP biosynthesis. The advantage of the function based screening is that after the novel NP is recovered there is a certainty that it is accurately biosynthesized and that it can be produced by the host strain (Felczykowska et al., 2015). However, the limitations of this approach are that it requires the genes or biosynthetic pathways, which are often from uncultured microorganisms, to be cloned and transferred into a heterologous screening host for expression (Ginolhac et al., 2004). In order for the drug discovery approaches to be successful, the new hosts that can activate transcription and produce molecules encoded by a diverse set of clusters need to be used (Katz et al., 2015). *E. coli* has been used in most studies as a host strain of choice. However, *E. coli* only contains half of the RNA polymerase sigma factors that are found in *Pseudomonas putida* or *Streptomyces*, for example. The use of other plasmids such as the novel broad-host shuttle fosmid pCCERI, which can also facilitate conjugation into alternative hosts such as *P. putida* and *S. lividans*, could significantly increase the expression efficiency during functional screening. If successful, this approach can lead to the discovery of completely novel genes or gene clusters that may encode for unknown enzyme structures or functions.

In this study, fosmid clones in *E. coli* were functionally screened for antimicrobial activity against a number of indicator strains and for pigmentation (Section 2.8.2). The main aim of this approach was to identify fosmid clones which carried the phenotypes that can produce bioactive natural products. However, in this study no positive fosmid clones were observed during function based screening.

97

Possibly, genes from the same biosynthetic gene cluster were located in different library clones due to random fragmentation of mDNA during metagenomic library construction. The pharmaceutically applicable NPs such as polyketides and nonribosomal peptides are synthesized by large gene clusters that can reach up to 100kb in size (Singh & Macdonald, 2010), which exceeds the insert size limitations of fosmids. There is a possibility that clones were unable to produce activity because the fosmid vectors used can only carry about 40kb DNA inserts, which affects the success of the expression of a pathway. Consequently, the improvement of the library construction methods is needed to increase the functional metagenomic screening efficiency.

The expression can be improved by employing alternate non-*E. coli* host strains for functional screening such as the above mentioned or by using *E. coli* together with vectors that contain additional tRNA genes in cases where expression is restricted by codon usage. Heterologous sigma factors expression in *E. coli,* which are capable of recognising heterologous promoters, can help enhance the mDNA expression (Rashid & Stingl, 2015). Lastly, specific chaperone proteins which help with the correct folding of biosynthetic enzymes can also be expressed to enhance the heterologous expression and enable production in *E. coli* (Milshteyn, Jessica S Schneider, et al., 2014). There is also a possibility that there were no antibiotic and pigment expressing clones represented in the library, or at least in the number of clones that were screened.

Screening of the sponge library using primers specific for the A domain of the NRPS genes did not result in any positive fosmid clones. This might indicate that these genes are rare in the

symbiotic bacteria associated with these two sponge species or that the NRPS genes vary significantly from the primers that were used for their detection. The PCR-based screening of the marine sponge *H. bidentifera* metagenomic library using the KS domain specific primers proved to be successful, which resulted in the amplification of two 700bp KS products. BLASTx analysis of the two fragments showed sequence identity, with 92% and 86%, respectively, to a gene found in *C. marina* (AFY62963.1) from a marine sponge, which encodes a PKS-like biosynthetic pathway. The fact that these KS genes are different suggests the presence of two different genes that could have come from the same *Cobetia* strain or from two different strains that were associated with the marine sponge.

In this study, *Cobetia* species seems to be associated with *H. bidentifera*, but other studies have reported it to be associated with *Mycale* sp. marine sponge which is widely distributed along the coast of Fujian in China and some in seawater from the Pacific Ocean (Su et al., 2014; Ibacache-Quiroga et al., 2017). Some sponge microbial symbionts may be specific for that host and this is evidenced by the fact that distantly related sponges from different geographical locations house similar microbial species that have not been seen in other sponge hosts (Webster & Taylor, 2012). For example, the sponge *Halichondria panacea* species from different marine locations (Adriatic Sea, North Sea, and the Baltic Sea) were shown to harbour the same genera, dominated by *Rhodobacter*, within their mesohyl (Brinkmann et al., 2017). Some sponges may always harbour a particular bacterial species that are likely to consistently produce a specific group of compounds as they need to adapt to their unique habitats (Thomas, 2010). Some sponge species may select specific microbial symbionts, for example, Candidatus phylum 'Poribacteria' was reported to be specific to marine demosponges (Hentschel et al., 2006).

A fosmid clone was sequenced and analysed for the presence of the polyketide biosynthetic genes. A discontiguous megaBLAST (BLASTn) analysis of the fosmid insert, showed the nucleotide sequence to be most similar to fragments of a *C. marina* strain JCM 21022 (CP017114), however at amino acid level (BLASTx) the proteins were most similar encoded by the pathway from *Cobetia* sp. UCD-24C (Table 3.4). The PE7-E10 fosmid PKS gene cluster was smaller than those of the other *Cobetia* species that it was compared to. The gene cluster annotation (Figure 3.8) revealed the fosmid pathway to be an O-antigen that was possibly interrupted by a shortened PKS element during transposition. This is confirmed by the presence of the genes that code for the enzymes such as UTP-glucose-1-phosphate uridylyl transferase, glucose-1-phosphate thymidyl transferase and dUTP-glucose 4,6-dehydratase, which form part of an O-antigen pathway. The O-antigen is a component of lipopolysaccharides that are found in the outer membrane of Gram negative bacteria (Cao et al., 2015).

It has been shown that some bacterial symbionts lose genes because they fail to provide a meaningful function or some ancestral gene sequences that are not essential anymore (Giordano et al., 2015). These bacteria continuously live within their hosts and can obtain the required metabolic compounds from their host, so they can end up discarding the corresponding biosynthetic genes or pathways (Moran, 2002).

The presence of the transposable insertion elements on the fosmid insert indicates that this part of the PKS biosynthetic gene cluster is a mobile genetic element that was fragmented during the genome rearrangement, which resulted in the loss of the three PKS domains. These transposable insertion elements are proposed to play an important role in the evolution of bacterial genomes for the symbiotic relationships between the sponges host and the bacteria

(Thomas et al., 2010). For example, they can disrupt the non-essential genes or by rearranging other genes to generate new pathways that may confer additional advantages to the symbiont (Thomas et al., 2010).

Although *in silico* prediction tools enabled us to identify a clone with biosynthetic pathway, we could not establish the end product that the cluster was responsible for due to the rearranged pathway. Irrespective, the PKS gene cluster can range from 10 to 120kb in size, with an average of about 50kb (Giordano et al., 2015), which cannot be represented in just one fosmid clone. As with the functional screening, the insert size is also a limitation in sequence-based screening and it does not guarantee the attainment of complete forms of biosynthetic genes or gene clusters (Yun & Ryu, 2005).

Based on the proposition that due to the size of the pathway in the other *Cobetia* species genomes (48kbp), it would have spanned more than one fosmid and therefore would most likely not have resulted in the expression of the natural compound, genome walking during sequence-based screening would also be necessary before attempting functional screening in order to recover the complete gene cluster which can be able to produce antimicrobial activity (Wilson & Piel, 2013). This can be achieved by using the PCR-based screening with primers designed based on the end sequences of the positive fosmid, which recognise the target biosynthetic pathway (Uria, 2015). This screening process can be repeated to screen for the additional parts of the pathway until the entire biosynthetic gene cluster is retrieved (Uria, 2015). Once the entire biosynthetic gene cluster is retrieved, it can be expressed in a heterologous host and screened for the antimicrobial activity. Thus, functional expression of these biosynthetic pathways can only be done after first detecting and isolating them using PCR-based screening.

101

The other limitation is that the sequence-based approach only gives a sequence which still needs to be cloned and expressed in a heterologous host in order to be able to assess the activity of the product and its structure. Heterologous expression of the gene clusters from metagenomic clones can give access to the encoded NPs and their activity. However, large mDNA fragments which are large enough to capture the entire biosynthetic gene cluster, cannot be easily cloned for heterologous expression. Therefore, sequence-based screening together with transformation-assisted recombination (TAR) can be used to access the complete biosynthetic gene clusters after recovering using genome-walking (Feng et al., 2011). For those biosynthetic pathways that are too large to be contained in one fosmid clone, the isolation of more than one overlapping fosmid clone from the library is required. This can be achieved by using PCR-based screening with primers designed from the end sequences of the isolated fosmid, which recognise the target biosynthetic pathway (Uria, 2015). The gene sequences of interest derived from the overlapping fosmid clones are reassembled into a single fragment, transformed and expressed in *Saccharomyces cerevisiae*. The TAR cloning system has been successfully used to clone an 89kb NRPS gene cluster from three overlapping eDNA cosmid clones (Kim et al., 2010).

The development of high-throughput sequencing methods, efficient bioinformatic analysis and routine heterologous gene expression will in future enable us to exploit the hidden pharmacological potential of a bacterial community that is so far underexplored. Future studies should include further screening of the sponge libraries to recover more fragments of the PKS biosynthetic gene cluster, as Table 3.2 shows that there might be another fosmid clone containing a PKS pathway that is similar to the isolated one. The full sequence can be analysed using bioinformatics for the ORFs and for the domain annotation. Alternatively the

102

whole metagenome can be sequenced and a bioinformatic approach taken to reconstruct the pathways. To prove that the expressed biosynthetic pathway encodes for proteins that have functional activity, the activity assay can be done using the suitable indicator strains.

In this study, the importance of sequence-based metagenomics in biosynthetic cluster identification from marine sponge symbionts was highlighted. Metagenomics is an important strategy that allows for the discovery of biosynthetic gene clusters that could not be discovered using culture-based strategies, because many diverse symbionts are unable to be cultured under laboratory conditions. Mining for PKS systems from metagenomes can be challenging because sponge-associated microbes may contain numerous individual genomes with numerous genes from different pathways. The data represented suggests that one of the limitations to sequence-based screening is ending up with inactive biosynthetic pathways. In conclusion, we could not establish whether the South African marine sponge, *H. bidentifera* (PE7), has the potential to be a good source of novel secondary metabolites, but that it may be associated with uncultured microbial symbionts that contain inactive PKS biosynthetic genes. This urges the continued investigation of marine sponges from the South African coasts, as they are yet to be exploited as a rich source of clinically relevant natural compounds.

Appendix A

Buffers and Stock solutions

50 X TAE buffer (pH 8)

242 g Tris base

18.6 g EDTA

57.1 ml glacial acetic acid

Sterile dH_2O to a final volume of 1 L

Ampicillin stock

100 mg/ml ampicillin in sterile dH_2O

Stored at -20°C

Ammonium acetate (4M)

15.42 g ammonium acetate

50 ml of sterile dH_2O

Chloramphenicol stock

34 mg/ml chloramphenicol in 100% ethanol

Stored at -20°C

DNA loading buffer

20% [v/v] glycerol

0.1 M EDTA

1% [w/v] SDS

EDTA (0.5 M, pH 8)

EDTA salt 186.1 g

NaOH pellets 20 g in sterile dH_2O

Buffers and Stock solutions

<u>Ethidium bromide staining solution</u>

20 g ethidium bromide staining powder

20 ml of sterile dH_2O

<u>GTE (Solution 1)</u>

50 mM EDTA

25 mM Tris-HCl

10 mM EDTA

<u>NaOH/SDS (Solution 2)</u>

0.2 M NaOH

1% [w/v] SDS

<u>Potassium Acetate pH 4.8 (Solution 3)</u>

29.5 ml glacial acetic acid

KOH pellets to pH 4.8

100 ml of sterile dH_2O

<u>TE buffer (pH 8)</u>

1 M Tris-HCl 1 ml

0.5 M EDTA 200 µl

Sterile dH_2O to a final volume of 100 ml

<u>X-gal</u>

0.6 g X-gal

30 ml dimethylformamide

Stored at -20°C

References

Aakvik, T., Degnes, K.F., Dahlsrud, R., Schmidt, F., Dam, R., Yu, L., Volker, U., Ellingsen, T.E. & Valla, S. 2009. A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species. *FEMS Microbiology Letters*, 296(2): 149–158.

Abdelmohsen, U.R., Pimentel-Elardo, S.M., Hanora, A., Radwan, M., Abou-El-Ela, S.H., Ahmed, S. & Hentschel, U. 2010. Isolation, phylogenetic analysis and anti-infective activity screening of marine sponge-associated actinomycetes. *Marine Drugs*, 8(3): 399–412.

Abe, T., Sahin, F.P., Akiyama, K., Naito, T., Kishigami, M., Miyamoto, K., Sakakibara, Y. & Uemura, D. 2012. Construction of a Metagenomic Library for the Marine Sponge Halichondria okadai. *Bioscience, Biotechnology, and Biochemistry*, 76(4): 633–639.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, 25(17): 3389–3402.

Amoutzias, G., Chaliotis, A. & Mossialos, D. 2016. Discovery Strategies of Bioactive Compounds Synthesized by Nonribosomal Peptide Synthetases and Type-I Polyketide Synthases Derived from Marine Microbiomes. *Marine Drugs*, 14(80): 1–20. http://www.mdpi.com/1660-3397/14/4/80.

Anand, T.P., Bhat, A.W., Shouche, Y.S., Roy, U., Siddharth, J. & Sarma, S.P. 2006. Antimicrobial activity of marine bacteria associated with sponges from the waters off the coast of South East India. *Microbiological research*, 161(3): 252–62.

References

http://www.ncbi.nlm.nih.gov/pubmed/16765842.

Angelov, A., Mientus, M., Liebl, S. & Liebl, W. 2009. A two-host fosmid system for functional
screening of (meta)genomic libraries from extreme thermophiles. *Systematic and applied microbiology*, 32(3): 177–85. http://www.ncbi.nlm.nih.gov/pubmed/19285378.

Ansari, M.Z., Yadav, G., Gokhale, R.S. & Mohanty, D. 2004. NRPS-PKS: a knowledge-based
resource for analysis of NRPS/PKS megasynthases. *Nucleic acids research*, 32(Web Server issue): W405–W413.

Atikana, A., Naim, M.A. & Sipkema, D. 2013. Detection of Keto Synthase (KS) Gene Domain
in Sponges and Bacterial Sponges. *Annales Bogorienses*, 17(2): 27–33.

Ayuso-Sacido, A. & Genilloud, O. 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. *Microbial ecology*, 49(1): 10–24. http://www.ncbi.nlm.nih.gov/pubmed/15614464.

Baker, D.D., Chu, M., Oza, U. & Rajgarhia, V. 2007. The value of natural products to future
pharmaceutical discovery. *Natural product reports*, 24(6): 1225–44. http://www.ncbi.nlm.nih.gov/pubmed/18033577.

Barone, R., De Santi, C., Palma Esposito, F., Tedesco, P., Galati, F., Visone, M., Di Scala, A. &
De Pascale, D. 2014. Marine metagenomics, a valuable tool for enzymes and bioactive compounds discovery. *Frontiers in Marine Science*, 1(38): 1–6. http://journal.frontiersin.org/article/10.3389/fmars.2014.00038/abstract.

Bérdy, J. 2005. Bioactive microbial metabolites. *The Journal of antibiotics*, 58(1): 1–26.

Blin, K., Medema, M.H., Kazempour, D., Fischbach, M. a., Breitling, R., Takano, E. & Weber, T.

107

References

2013. antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Research*, 41(W1): W204–W212. http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkt449.

Brady, S.F. 2007. Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. *Nature protocols*, 2(5): 1297–1305. http://www.ncbi.nlm.nih.gov/pubmed/17546026.

Brady, S.F., Chao, C.J., Clardy, J. & Al, B.E.T. 2004. Long-Chain N -Acyltyrosine Synthases from Environmental DNA. *Applied and Environmental Microbiology*, 70(11): 6865–6870.

Brady, S.F., Simmons, L., Kim, J.H. & Schmidt, E.W. 2010. Metagenomic approaches to natural products from free living and symbiotic organisms. *Natural product reports*, 26(11): 1488–1503.

Brinkmann, C., Marker, A. & Kurtböke, D. 2017. An Overview on Marine Sponge-Symbiotic Bacteria as Unexhausted Sources for Natural Product Discovery. *Diversity*, 9(4): 40. http://www.mdpi.com/1424-2818/9/4/40.

Cao, B., Tian, Z., Wang, S., Zhu, Z., Sun, Y., Feng, L. & Wang, L. 2015. Structural comparison of O-antigen gene clusters of Legionella pneumophila and its application of a serogroup-specific multiplex PCR assay. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 108(6): 1405–1423.

Charlop-Powers, Z., Milshteyn, A. & Brady, S.F. 2014. Metagenomic small molecule discovery methods. *Current Opinion in Microbiology*, 19: 70–5. http://www.sciencedirect.com/science/article/pii/S1369527414000691.

Chen, R., Li, C., Pei, X., Wang, Q., Yin, X. & Xie, T. 2014. Isolation an Aldehyde Dehydrogenase

References

Gene from Metagenomics Based on Semi-nest Touch-Down PCR. *Indian Journal of Microbiology*, 54(1): 74–79.

Chiang, Y., Chang, S., Oakley, B.R. & Wang, C.C.C. 2010. Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms. *Current Opinion in Chemical Biology*, 15(1): 137–143. http://dx.doi.org/10.1016/j.cbpa.2010.10.011.

Courtois, S., Cappellano, C.M., Ball, M., Francou, F., Normand, P., Martinez, A., Kolvek, S.J., Hopke, J., Osburne, M.S., August, P.R., Nalin, R., Jeannin, P., Simonet, P., Pernodet, J., Lyon, C.B. & Cedex, V. 2003. Recombinant Environmental Libraries Provide Access to Microbial Diversity for Drug Discovery from Natural Products. *Applied and Environmental Microbiology*, 69(1): 49–55.

Croué, J., West, N.J., Escande, M.-L., Intertaglia, L., Lebaron, P. & Suzuki, M.T. 2013. A single betaproteobacterium dominates the microbial community of the crambescidine-containing sponge Crambe crambe. *Scientific Reports*, 3(1): 2583. http://www.nature.com/articles/srep02583.

Cueto, M., Jensen, P.R., Kauffman, C., Fenical, W., Lobkovsky, E. & Clardy, J. 2001. Pestalone, a new antibiotic produced by a marine fungus in response to bacterial challenge. *Journal of Natural Products*, 64(11): 1444–1446.

Daniel, R. 2004. The soil metagenome--a rich resource for the discovery of novel natural products. *Current opinion in biotechnology*, 15(3): 199–204. http://www.ncbi.nlm.nih.gov/pubmed/15193327.

Davidson, S.K., Allen, S.W., Lim, G.E., Anderson, C.M. & Haygood, M.G. 2001. Evidence for the

References

Biosynthesis of Bryostatins by the Bacterial Symbiont 'Candidatus Endobugula sertula' of the Bryozoan Bugula neritina. *Applied and Environmental Microbiology*, 67(10): 4531–4537.

Deng, J., Wei, M., Yu, B. & Chen, Y. 2010. Efficient amplification of genes involved in microbial secondary metabolism by an improved genome walking method. *Applied Microbiology and Biotechnology*, 87(2): 757–764.

Dharmaraj, S. 2010. Marine Streptomyces as a novel source of bioactive substances. *World Journal of Microbiology and Biotechnology*, 26(12): 2123–2139.

Du, L., Sánchez, C. & Shen, B. 2001. Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. *Metabolic engineering*, 3(1): 78–95.

Ekkers, D.M., Cretoiu, M.S., Kielak, A.M. & Van Elsas, J.D. 2012. The great screen anomaly-a new frontier in product discovery through functional metagenomics. *Applied Microbiology and Biotechnology*, 93(3): 1005–1020.

van Elsas, J.D., Speksnijder, a J. & van Overbeek, L.S. 2008. A procedure for the metagenomics exploration of disease-suppressive soils. *Journal of microbiological methods*, 75(3): 515–22. http://www.ncbi.nlm.nih.gov/pubmed/18778739.

Felczykowska, A., Krajewska, A., Zielińska, S., Łoś, J.M., Bloch, S.K. & Nejman-Faleńczyk, B. 2015. The most widespread problems in the function-based microbial metagenomics. *Acta Biochimica Polonica*, 62(1): 161–166.

Felnagle, E.A., Jackson, E.E., Chan, Y.A., Podevels, A.M., Berti, D., Mcmahon, M.D. & Thomas, M.G. 2008. Nonribosomal peptide synthatases involved in the production of medically relevant natural products. *Mol Pharm*, 5(2): 191–211.

References

Feng, Z., Kim, J.H. & Brady, S.F. 2011. Fluostatins produced by the heterologous expression of a tar reassembled environmental DNA derived Type II PKS gene cluster. *Journal of American Chem Society*, 132(34): 11902–11903.

Fieseler, L., Hentschel, U., Grozdanov, L., Schirmer, A., Wen, G., Platzer, M., Hrvatin, S., Butzke, D., Zimmermann, K. & Piel, J. 2007. Widespread occurrence and genomic context of unusually small polyketide synthase genes in microbial consortia associated with marine sponges. *Applied and environmental microbiology*, 73(7): 2144–2155. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1855645&tool=pmcentrez &rendertype=abstract.

Fieseler, L., Horn, M., Wagner, M. & Hentschel, U. 2004. Discovery of the Novel Candidate Phylum Poribacteria in Marine Sponges. *Applied and environmental microbiology*, 70(6): 3724–3732.

Fisch, K.M., Gurgui, C., Heycke, N., van der Sar, S. a, Anderson, S. a, Webb, V.L., Taudien, S., Platzer, M., Rubio, B.K., Robinson, S.J., Crews, P. & Piel, J. 2009. Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. *Nature chemical biology*, 5(7): 494–501. http://www.ncbi.nlm.nih.gov/pubmed/19448639.

Fleming, A. 1929. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B.influenzae. *British journal of experimental pathology*, 10(1923): 226–236.

Freeman, M.F., Gurgui, C., Helf, M.J., Morinaka, B.I., Uria, A.R., Oldham, N.J., Sahl, H.-G., Matsunaga, S. & Piel, J. 2012. Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. *Science*, 338: 387–390.

References

Gabor, E.M., Alkema, W.B.L. & Janssen, D.B. 2004. Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environmental Microbiology*, 6(9): 879–886.

Gazave, E., Lapébie, P., Renard, E., Vacelet, J., Rocher, C., Ereskovsky, A. V., Lavrov, D. V. & Borchiellini, C. 2010. Molecular phylogeny restores the supra-generic subdivision of homoscleromorph sponges (Porifera, Homoscleromorpha). *PLoS ONE*, 5(12).

Genilloud, O. 2014. The re-emerging role of microbial natural products in antibiotic discovery. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 106(1): 173–188.

Gillespie, D.E., Brady, S.F., Bettermann, A.D., Cianciotto, N.P., Liles, M.R., Rondon, M.R., Clardy, J., Goodman, R.M. & Handelsman, J. 2002. Isolation of Antibiotics Turbomycin A and B from a Metagenomic Library of Soil Microbial DNA. *Applied and Environmental Microbiology*, 68(9): 430.

Ginolhac, A., Jarrin, C., Gillet, B., Robe, P., Pujic, P., Tuphile, K., Bertrand, H., Vogel, T.M., Perrière, G., Simonet, P. & Nalin, R. 2004. Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. *Applied and Environmental Microbiology*, 70(9): 5522–5527.

Giordano, D., Coppola, D., Russo, R., Denaro, R., Giuliano, L., Lauro, F.M., di Prisco, G. & Verde, C. 2015. Marine Microbial Secondary Metabolites: Pathways, Evolution and Physiological Roles. *Advances in microbial physiology*, 66: 357–428. http://www.sciencedirect.com/science/article/pii/S0065291115000028.

Gomes, E.S., Schuch, V. & Lemos, E.G. de M. 2013. Biotechnology of polyketides: New breath

References

of life for the novel antibiotic genetic pathways discovery through metagenomics. *Brazilian Journal of Microbiology*, 44(4): 1007–1034.

Gurgui, C. & Piel, J. 2010. *Metagenomics*. W. R. Streit & R. Daniel, eds. Totowa, NJ: Humana Press. http://link.springer.com/10.1007/978-1-60761-823-2.

Handelsman, J. 2004. Metagenomics : Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews*, 68(4): 669–685.

Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J. & Goodman, R.M. 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & biology*, 5(10): R245–R249.

Hentschel, U., Hopke, J., Horn, M., Anja, B., Wagner, M., Hacker, J., Bradley, S., Friedrich, A.B. & Moore, B.S. 2002. Molecular Evidence for a Uniform Microbial Community in Sponges from Different Oceans Molecular Evidence for a Uniform Microbial Community in Sponges from Different Oceans. *Applied and Environmental Microbiology*, 68(9): 4431–4440.

Hentschel, U., Piel, J., Degnan, S.M. & Taylor, M.W. 2012. Genomic insights into the marine sponge microbiome. *Nature reviews. Microbiology*, 10(9): 641–654. http://www.ncbi.nlm.nih.gov/pubmed/22842661 27 May 2014.

Hentschel, U., Schmid, M., Wagner, M., Fieseler, L. & Gernert, C. 2001. Isolation and phylogenetic analysis of bacteria with antimicrobial activities from the Mediterranean sponges Aplysina aerophoba and Aplysina cavernicola. *FEMS Microbiology Ecology*, 35: 305–312.

Hentschel, U., Usher, K.M. & Taylor, M.W. 2006. Marine sponges as microbial fermenters.

113

References

FEMS Microbiology Ecology, 55(2): 167–177.

Hertweck, C. 2009. The biosynthetic logic of polyketide diversity. *Angewandte Chemie (International ed. in English)*, 48(26): 4688–4716. http://www.ncbi.nlm.nih.gov/pubmed/19514004.

Hertweck, C., Luzhetskyy, A., Rebets, Y. & Bechthold, A. 2007. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Natural product reports*, 24(1): 162–190.

Hochmuth, T. & Piel, J. 2009a. Polyketide synthases of bacterial symbionts in sponges--evolution-based applications in natural products research. *Phytochemistry*, 70(15–16): 1841–9. http://www.ncbi.nlm.nih.gov/pubmed/19443000.

Hochmuth, T. & Piel, J. 2009b. Polyketide synthases of bacterial symbionts in sponges - Evolution-based applications in natural products research. *Phytochemistry*, 70(15–16): 1841–1849. http://dx.doi.org/10.1016/j.phytochem.2009.04.010.

Hrvatin, S. & Piel, J. 2007. Rapid isolation of rare clones from highly complex DNA libraries by PCR analysis of liquid gel pools. *Journal of Microbiological Methods*, 68(2): 434–436.

Iguchi, A., Iyoda, S., Kikuchi, T., Ogura, Y., Katsura, K., Ohnishi, M., Hayashi, T. & Thomson, N.R. 2015. A complete view of the genetic diversity of the Escherichia coli O-antigen biosynthesis gene cluster. *DNA Research*, 22(1): 101–107.

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

Iqbal, H. a, Feng, Z. & Brady, S.F. 2012. Biocatalysts and small molecule products from

References

metagenomic studies. *Current opinion in chemical biology*, 16(1–2): 109–16. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3335775&tool=pmcentrez &rendertype=abstract.

Jenke-Kodama, H. & Dittmann, E. 2009. Evolution of metabolic diversity: Insights from microbial polyketide synthases. *Phytochemistry*, 70(15–16): 1858–1866. http://dx.doi.org/10.1016/j.phytochem.2009.05.021.

Jiang, S., Sun, W., Chen, M., Dai, S., Zhang, L., Liu, Y., Lee, K.J. & Li, X. 2007. Diversity of culturable actinobacteria isolated from marine sponge Haliclona sp. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 92(4): 405–416.

Kakirde, K.S., Parsley, L.C. & Liles, M.R. 2010. Size does matter: Application-driven approaches for soil metagenomics. *Soil Biology and Biochemistry*, 42(11): 1911–1923. http://dx.doi.org/10.1016/j.soilbio.2010.07.021.

Katz, M., Hover, B.M. & Brady, S.F. 2015. Culture-independent discovery of natural products from soil metagenomes. *Journal of Industrial Microbiology & Biotechnology*. http://link.springer.com/10.1007/s10295-015-1706-6.

Kennedy, J., Flemer, B., Jackson, S. a, Lejon, D.P.H., Morrissey, J.P., O'Gara, F. & Dobson, A.D.W. 2010. Marine metagenomics: new tools for the study and exploitation of marine microbial metabolism. *Marine drugs*, 8(3): 608–28. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2857354&tool=pmcentrez &rendertype=abstract.

Kennedy, J., Marchesi, J.R. & Dobson, A.D.W. 2008. Marine metagenomics : strategies for the

115

discovery of novel enzymes with biotechnological applications from marine environments. *Microbila Cell Factories*, 7(27): 1–8.

Kennedy, J., Marchesi, J.R. & Dobson, A.D.W. 2007. Metagenomic approaches to exploit the biotechnological potential of the microbial consortia of marine sponges. *Applied microbiology and biotechnology*, 75(1): 11–20. http://www.ncbi.nlm.nih.gov/pubmed/17318533.

Khan, S.T., Komaki, H., Motohashi, K., Kozone, I., Mukai, A., Takagi, M. & Shin-Ya, K. 2011. Streptomyces associated with a marine sponge Haliclona sp.; biosynthetic genes for secondary metabolites and products. *Environmental Microbiology*, 13(2): 391–403.

Kim, J.H., Feng, Z., Bauer, J.D., Kallifidas, D., Calle, P.Y. & Brady, S.F. 2010. Cloning large natural product gene clusters from the environment: Piecing environmental DNA gene clusters back together with TAR. *Biopolymers*, 93(9): 833–844.

Kim, K.T., Garson, M.J. & Fuerst, J. a. 2005. Marine actinomycetes related to the 'Salinospora' group from the Great Barrier Reef sponge Pseudoceratina clavata. *Environmental Microbiology*, 7(4): 509–518.

Kim, S.-K. 2015. *Springer Handbook of Marine Biotechnology*. S.-K. Kim, ed. Berlin, Heidelberg: Springer Berlin Heidelberg. http://link.springer.com/10.1007/978-3-642-53971-8.

Kim, T.K. & Fuerst, J.A. 2006. Diversity of polyketide synthase genes from bacteria associated with the marine sponge Pseudoceratina clavata: culture-dependent and culture-independent approaches. *Environmental microbiology*, 8(8): 1460–1470. http://www.ncbi.nlm.nih.gov/pubmed/16872408.

Klopries, S., Sundermann, U. & Schulz, F. 2013. Quantification of N-acetylcysteamine activated

methylmalonate incorporation into polyketide biosynthesis. *Beilstein Journal of Organic Chemistry*, 9: 664–674.

Knappe, T.A., Linne, U., Zirah, S., Rebuffat, S., Xie, X. & Marahiel, M.A. 2008. Isolation and structural characterization of capistruin, a lasso peptide predicted from the genome sequence of Burkholderia thailandensis E264. *Journal of the American Chemical Society*, 130(17): 11446–11454. http://www.ncbi.nlm.nih.gov/pubmed/18671394.

Kurnia, N.M., Uria, A.R., Kusnadi, Y., Dinawati, L., Zilda, D.S., Hadi, T.A., Setyahadi, S. & Felix, F. 2017. Metagenomic Survey of Potential Symbiotic Bacteria and Polyketide Synthase Genes in an Indonesian Marine Sponge. *HAYATI Journal of Biosciences*, 24(1): 6–15. http://linkinghub.elsevier.com/retrieve/pii/S1978301916304429.

Lane, A.L. & Moore, B.S. 2012. A sea of bisynthesis: marine natural products meet the molecular age. *Natural product reports*, 28(2): 411–428.

Li, Z. 2009. Advances in marine microbial symbionts in the china sea and related pharmaceutical metabolites. *Marine drugs*, 7(2): 113–29. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2707038&tool=pmcentrez&rendertype=abstract.

Liou, G. 2003. Building-block selectivity of polyketide synthases. *Current Opinion in Chemical Biology*, 7(2): 279–284. http://linkinghub.elsevier.com/retrieve/pii/S1367593103000164.

Lipko, I.A., Kalyuzhnaya, O. V., Kravchenko, O.S. & Parfenova, V. V. 2012. Identification of polyketide synthase genes in genome of Pseudomonas fluorescens strain 28Bb-06 from freshwater sponge Baikalospongia bacillifera. *Molecular Biology*, 46(4): 609–611.

117

References

http://link.springer.com/10.1134/S0026893312040073.

Liu, J., Hua, Z.S., Chen, L.X., Kuang, J.L., Li, S.J., Shu, W.S. & Huang, L.N. 2014. Correlating microbial diversity patterns with geochemistry in an extreme and heterogeneous environment of mine tailings. *Applied and Environmental Microbiology*, 80(12): 3677–3686.

Long, P.F., Dunlap, W.C., Battershill, C.N. & Jaspars, M. 2005. Shotgun cloning and heterologous expression of the patellamide gene cluster as a strategy to achieving sustained metabolite production. *Chembiochem : a European journal of chemical biology*, 6(10): 1760–5. http://www.ncbi.nlm.nih.gov/pubmed/15988766 11 March 2016.

Lorente, A., Makowski, K. & Albericio, F. 2014. Bioactive Marine Polyketides as Potential and Promising Drugs. *Annals of Marine Biology and Research*, 1(1): 1–10.

Martinez, A., Kolvek, S.J., Yip, C.L.T., Hopke, J., Brown, K.A., MacNeil, I.A. & Osburne, M.S. 2004. Genetically Modified Bacterial Strains and Novel Bacterial Artificial Chromosome Shuttle Vectors for Constructing Environmental Libraries and Detecting Heterologous Natural Products in Multiple Expression Hosts. *Applied and Environmental Microbiology*, 70(4): 2452–2463.

McMahon, M.D., Guan, C., Handelsman, J. & Thomas, M.G. 2012. Metagenomic analysis of Streptomyces lividans reveals host-dependent functional expression. *Applied and Environmental Microbiology*, 78(10): 3622–3629.

Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. a, Weber, T., Takano, E. & Breitling, R. 2011. antiSMASH: rapid identification, annotation and

References

analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic acids research*, 39(Web Server issue): W339-46. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3125804&tool=pmcentrez &rendertype=abstract.

Metsä-Ketelä, M., Salo, V., Halo, L., Hautala, A., Hakala, J., Mäntsälä, P. & Ylihonko, K. 1999. An efficient approach for screening minimal PKS genes from Streptomyces. *FEMS microbiology letters*, 180(1): 1–6. http://www.ncbi.nlm.nih.gov/pubmed/10547437.

Milshteyn, A., Schneider, J.S. & Brady, S.F. 2014. Mining the Metabiome: Identifying Novel Natural Products from Microbial Communities. *Chemistry & Biology*, 21(9): 1211–1223. http://linkinghub.elsevier.com/retrieve/pii/S1074552114002804.

Milshteyn, A., Schneider, J.S. & Brady, S.F. 2014. Review Mining the Metabiome: Identifying Novel Natural Products from Microbial Communities. *Chemistry & Biology*, 21(9): 1211–1223. http://dx.doi.org/10.1016/j.chembiol.2014.08.006.

Mizuno, C.M., Kimes, N.E., López-Pérez, M., Ausó, E., Rodriguez-Valera, F. & Ghai, R. 2013. A Hybrid NRPS-PKS Gene Cluster Related to the Bleomycin Family of Antitumor Antibiotics in Alteromonas macleodii Strains. *PLoS ONE*, 8(9): 1–12.

Monaco, R.R. & Quinlan, R.F. 2014. Novel Natural Product Discovery from Marine Sponges and their Obligate Symbiotic Organisms. *Biorxiv*.

Mondol, M.A.M., Shin, H.J. & Islam, M.T. 2013. Diversity of secondary metabolites from marine Bacillus species: chemistry and biological activity. *Marine drugs*, 11(8): 2846–72. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3766869&tool=pmcentrez &rendertype=abstract.

119

References

Montasser, R. & Luesch, H. 2012. Marine natural products : a new wave of drugs ? *Future Med Chem*, 3(12): 1475–1489.

Moran, N.A. 2002. Microbial MinimalismGenome Reduction in Bacterial Pathogens. *Cell*, 108(5): 583–586.

Neilan, B.A., Dittmann, E., Rouhiainen, L., Amanda, R., Schaub, V., Sivonen, K., Börner, T., Rouhiainen, L.E.O., Bass, R.A. & Bo, T. 1999. Nonribosomal Peptide Synthesis and Toxigenicity of Cyanobacteria. *Journal of Bacteriology*, 181(July): 4089–4097.

Ng, T.B., Chueng, R.C.F., Wong, J.H., Bekhit, A.A. & Beckit, A.E.-D. 2015. Antibacterial products of marine organisms. *Applied Microbiology and Biotechnology*, 99: 4145–4173.

Nikolouli, K. & Mossialos, D. 2012. Bioactive compounds synthesized by non-ribosomal peptide synthetases and type-I polyketide synthases discovered through genome-mining and metagenomics. *Biotechnology Letters*, 34(8): 1393–1403.

Ouyang, Y., Dai, S., Xie, L., Kumar, M.S.R., Sun, W., Sun, H., Tang, D. & Li, X. 2010. Isolation of high molecular weight DNA from marine sponge bacteria for BAC library construction. *Marine Biotechnology*, 12(3): 318–325.

Owen, J.G., Robins, K.J., Parachin, N.S. & Ackerley, D.F. 2012. A functional screen for recovery of 4'-phosphopantetheinyl transferase and associated natural product biosynthesis genes from metagenome libraries. *Environmental Microbiology*, 14(5): 1198–1209.

Papagianni, M. 2003. Ribosomally synthesized peptides with antimicrobial properties: Biosynthesis, structure, function, and applications. *Biotechnology Advances*, 21(6): 465–499.

Parsley, L.C., Linneman, J., Goode, A.M., Becklund, K., George, I., Goodman, R.M., Lopanik,

References

N.B. & Liles, M.R. 2011. Polyketide synthase pathways identified from a metagenomic library are derived from soil Acidobacteria. *FEMS Microbiology Ecology*, 78(1): 176–187.

Penesyan, A., Kjelleberg, S. & Egan, S. 2010. Development of novel drugs from marine surface associated microorganisms. *Marine drugs*, 8(3): 438–59. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2857370&tool=pmcentrez &rendertype=abstract.

Piel, J. 2002. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles. *Proceedings of the National Academy of Sciences of the United States of America*, 99(22): 14002–14007. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=137826&tool=pmcentrez &rendertype=abstract.

Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N. & Matsunaga, S. 2004. Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge Theonella swinhoei. *Proceedings of the National Academy of Sciences of the United States of America*, 101(46): 16222–7. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=528957&tool=pmcentrez &rendertype=abstract.

Pimentel-Elardo, S.M., Grozdanov, L., Proksch, S. & Hentschel, U. 2012. Diversity of nonribosomal peptide synthetase genes in the microbial metagenomes of marine sponges. *Marine drugs*, 10(6): 1192–202. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3397433&tool=pmcentrez &rendertype=abstract.

References

Qin, S., Müller, W.E.G. & Cooper, E.L. 2011. Marine biotechnology: A new vision and strategy for Europe. *European Science Foundation*, 2011(September): 639140. http://www.ncbi.nlm.nih.gov/pubmed/23394961.

Rashid, M. & Stingl, U. 2015. Contemporary molecular tools in microbial ecology and their application to advancing biotechnology. *Biotechnology Advances*, 33(8): 1755–1773. http://linkinghub.elsevier.com/retrieve/pii/S0734975015300380.

Ridley, C.P. & Khosla, C. 2009. Polyketides. *Applied Microbiology: Industrial*: 472–481.

Ridley, M. 2000. *Genome : the autobiography of a species in 23 chapters*. http://books.google.com/books?id=7O775xBxfMoC&pg=PA5&dq=genome+the+autobiography+inauthor:ridley&hl=&cd=1&source=gbs_api%5Cnpapers3://publication/uuid/5953B26C-F19F-4A96-8C27-539CB29EE18D.

Rocha-Martin, J., Harrington, C., Dobson, A.D.W. & O'Gara, F. 2014. *Emerging strategies and integrated systems microbiology technologies for biodiscovery of marine bioactive compounds.* http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4071589&tool=pmcentrez&rendertype=abstract.

Della Sala, G., Hochmuth, T., Costantino, V., Teta, R., Gerwick, W., Gerwick, L., Piel, J. & Mangoni, A. 2013. Polyketide genes in the marine sponge Plakortis simplex: A new group of mono-modular type I polyketide synthases from sponge symbionts. *Environmental Microbiology Reports*, 5(6): 809–818.

Sang Chun, J., Kim, D., Yoon, J., Oh, T. & Lee, C. 2006. Sequence-Based Screening for Putative Polyketide Synthase Gene-Harboring Clones from a Soil Metagenome Library. *Journal of*

References

Microbiology and Biotechnology, 16(1): 153–157.

Scherlach, K. & Hertweck, C. 2009. Triggering cryptic natural product biosynthesis in microorganisms. *Organic & biomolecular chemistry*, 7(9): 1753–1760. http://dx.doi.org/10.1039/B821578B.

Schirmer, A., Gadkari, R., Reeves, C.D., Ibrahim, F., Delong, E.F., Richard, C. & Hutchinson, C.R. 2005. Metagenomic Analysis Reveals Diverse Polyketide Synthase Gene Clusters in Microorganisms Associated with the Marine Sponge Discodermia dissoluta. *Applied and environmental microbiology*, 71(8): 4840–4849.

Schloss, P.D. & Handelsman, J. 2003. Biotechnological prospects from metagenomics. *Current Opinion in Biotechnology*, 14(3): 303–310. http://linkinghub.elsevier.com/retrieve/pii/S0958166903000673.

Schwarzer, D. & Marahiel, M.A. 2001. Multimodular biocatalysts for natural product assembly. *Naturwissenschaften*, 88(3): 93–101.

Schwecke, T., Aparicio, J.F., Molnár, I., König, a, Khaw, L.E., Haydock, S.F., Oliynyk, M., Caffrey, P., Cortés, J. & Lester, J.B. 1995. The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proceedings of the National Academy of Sciences of the United States of America*, 92(17): 7839–43. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=41241&tool=pmcentrez&rendertype=abstract.

Seow, K.T., Meurer, G., Gerlitz, M., Wendt-Pienkowski, E., Hutchinson, C.R. & Davies, J. 1997. A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: A means to access and use genes from uncultured microorganisms. *Journal of*

References

Bacteriology, 179(23): 7360–7368.

Shen, B. 2003. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Current Opinion in Chemical Biology*, 7(2): 285–295.

Siegl, A. & Hentschel, U. 2010. PKS and NRPS gene clusters from microbial symbiont cells of marine sponges by whole genome amplification. *Environmental microbiology reports*, 2(4): 507–13. http://www.ncbi.nlm.nih.gov/pubmed/23766222.

Simon, C. & Daniel, R. 2009. Achievements and new knowledge unraveled by metagenomic approaches. *Applied Microbiology and Biotechnology*, 85(2): 265–276.

Singh, B.K. & Macdonald, C. a. 2010. Drug discovery from uncultivable microorganisms. *Drug discovery today*, 15(17–18): 792–9. http://www.ncbi.nlm.nih.gov/pubmed/20656054.

Singh, J., Behal, A., Singla, N., Joshi, A., Birbian, N., Singh, S., Bali, V. & Batra, N. 2009. Metagenomics: Concept, methodology, ecological inference and recent advances. *Biotechnology journal*, 4(4): 480–94. http://www.ncbi.nlm.nih.gov/pubmed/19288513.

Skariyachan, S., G Rao, A., Patil, M.R., Saikia, B., Bharadwaj Kn, V. & Rao Gs, J. 2013. Antimicrobial potential of metabolites extracted from bacterial symbionts associated with marine sponges in coastal area of Gulf of Mannar Biosphere, India. *Letters in applied microbiology*, 58(3): 231–241. http://www.ncbi.nlm.nih.gov/pubmed/24138171.

Skinnider, M.A., Dejong, C.A., Rees, P.N., Johnston, C.W., Li, H., Webster, A.L.H., Wyatt, M.A. & Magarvey, N.A. 2015. Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). *Nucleic Acids Research*, 43(20): 9645–9662.

Snyder, R. V., Gibbs, P.D.L., Palacios, A., Abiy, L., Dickey, R., Lopez, J. V. & Rein, K.S. 2003. Polyketide synthase genes from marine dinoflagellates. *Marine Biotechnology*, 5(1): 1–

References

12.

Staunton, J. & Weissman, K.J. 2001. Polyketide biosynthesis: a millennium review. *Natural Product Reports*, 18(4): 380–416. http://xlink.rsc.org/?DOI=a909079g.

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

Strien, J., Sanft, J. & Mall, G. 2013. Enhancement of PCR amplification of moderate gc-containing and highly GC-rich DNA sequences. *Molecular Biotechnology*, 54(3): 1048–1054.

Sudek, S., Lopanik, N.B., Waggoner, L.E., Hildebrand, M., Anderson, C., Liu, H., Patel, A., Sherman, D.H. & Haygood, M.G. 2007. Identification of the putative bryostatin polyketide synthase gene cluster from 'Candidatus *Endobugula sertula*', the uncultivated microbial symbiont of the marine bryozoan *Bugula neritina*. *Journal of natural products*, 70(1): 67–74. http://www.ncbi.nlm.nih.gov/pubmed/17253852.

Sun, W., Peng, C., Zhao, Y. & Li, Z. 2012. Functional gene-guided discovery of type II polyketides from culturable actinomycetes associated with soft coral scleronephthya sp. *PLoS ONE*, 7(8): 1–11.

Talfournier, F., Stines-Chaumeil, C. & Branlant, G. 2011. Methylmalonate-semialdehyde dehydrogenase from Bacillus subtilis: Substrate specificity and coenzyme A binding. *Journal of Biological Chemistry*, 286(25): 21971–21981.

Tambadou, F., Lanneluc, I., Sable, S., Klein, G.L., Doghori, I., Sopena, V., Sandrine, D., Barthelemy, C., Thiery, V. & Romain, C. 2014. Novel nonribosomal peptide synthetase (NRPS) genes sequenced from intertidal mudflat bacteria. *Federation of European*

References

Microbiological Societies, 357: 123–130. http://femsle.oxfordjournals.org/content/femsle/357/2/123.full.pdf.

Tang, X., Xu, K., Han, X., Mo, Z. & Mao, Y. 2017. Complete genome of Cobetia marina JCM 21022 T and phylogenomic analysis of the family Halomonadaceae *. *Chinese Journal of Oceanology and Limnology*.

Thaker, M.N., Wang, W., Spanogiannopoulos, P., Waglechner, N., King, A.M., Medina, R. & Wright, G.D. 2013. Identifying producers of antibacterial compounds by screening for antibiotic resistance. *Nature biotechnology*, 31(10): 922–7. http://www.ncbi.nlm.nih.gov/pubmed/24056948.

Thomas, T., Rusch, D., DeMaere, M.Z., Yung, P.Y., Lewis, M., Halpern, A., Heidelberg, K.B., Egan, S., Steinberg, P.D. & Kjelleberg, S. 2010. Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *The ISME Journal*, 4(12): 1557–1567. http://www.nature.com/doifinder/10.1038/ismej.2010.74.

Trindade, M., van Zyl, L.J., Navarro-Fernández, J. & Abd Elrazak, A. 2015. Targeted metagenomics as a tool to tap into marine natural product diversity for the discovery and production of drug candidates. *Frontiers in Microbiology*, 6(890): 1–14. http://journal.frontiersin.org/Article/10.3389/fmicb.2015.00890/abstract.

Uria, A. & Piel, J. 2009. Cultivation-independent approaches to investigate the chemistry of marine symbiotic bacteria. *Phytochemistry Reviews*, 8(2): 401–414. http://link.springer.com/10.1007/s11101-009-9127-7.

Uria, A.R. 2015. Capturing Natural Product Biosynthetic Pathways from Uncultivated Symbiotic Bacteria of Marine Sponges through Metagenome Mining: A Mini-Review.

References

Squalen Bulletin of Marine & Fisheries Postharvest &Biotechnology, 10(1): 35–49.

Uria, A.R., Fawzya, Y.N. & Chasanah, E. 2005. Novel molecular methods for discovery and engineering of biocatalysts from uncultured marine microorganisms. *Journal of Coastal Development*, 8(2): 53–71.

Valliappan, K., Sun, W. & Li, Z. 2014. Marine actinobacteria associated with marine organisms and their potentials in producing pharmaceutical natural products. *Applied Microbiology and Biotechnology*: 7365–7377.

Vizcaíno, J.A., Sanz, L., Cardoza, R.E., Monte, E. & Gutiérrez, S. 2005. Detection of putative peptide synthetase genes in Trichoderma species: Application of this method to the cloning of a gene from T. harzianum CECT 2413. *FEMS Microbiology Letters*, 244(1): 139–148.

Walker, A.W., Duncan, S.H., Louis, P. & Flint, H.J. 2014. Phylogeny, culturing, and metagenomics of the human gut microbiota. *Trends in Microbiology*, 22(5): 267–274. http://dx.doi.org/10.1016/j.tim.2014.03.001.

Walsh, C.T. 2004. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science (New York, N.Y.)*, 303(5665): 1805–10. http://www.ncbi.nlm.nih.gov/pubmed/15031493.

Walsh, C.T. 2008. The chemical versatility of natural-product assembly lines. *Accounts of chemical research*, 41(1): 4–10. http://www.ncbi.nlm.nih.gov/pubmed/17506516.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Bruccoleri, R., Lee, S.Y., Fischbach, M.A., Rolf, M., Wohlleben, W., Breitling, R., Takano, E. & Medema, M.H. 2015. antiSMASH 3 . 0 — a comprehensive resource for the genome mining of biosynthetic gene clusters.

References

Nucleic Acids Research, 1: 1–7.

Webster, N.S. & Taylor, M.W. 2012. Minireview Marine sponges and their microbial symbionts : love. *Environmental Microbiology*, 14(2): 335–346.

Weiland, N., Löscher, C., Metzger, R. & Schmitz, R. 2010. Consruction and screening of marine Metagenomic libraries. In *Metagenomics: Methods and Protocols*. 51–65. http://link.springer.com/10.1007/978-1-60761-823-2.

Wilkinson, C.R., Garrone, R. & Vacelet, J. 1984. Marine sponges discriminate between food bacteria and bacterial symboints: electron microscope radioautography and it in situ evidence. *The Royal Society of London*, 220: 519–528.

Wilson, M.C. & Piel, J. 2013. Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chemistry & biology*, 20(5): 636–47. http://www.ncbi.nlm.nih.gov/pubmed/23706630.

Xiong, Z.-Q., Wang, J.-F., Hao, Y.-Y. & Wang, Y. 2013. Recent advances in the discovery and development of marine microbial natural products. *Marine drugs*, 11(3): 700–17. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3705366&tool=pmcentrez &rendertype=abstract.

Yun, J. & Ryu, S. 2005. Screening for novel enzymes from metagenome and SIGEX, as a way to improve it. *Microbial cell factories*, 4(8). http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1079935&tool=pmcentrez &rendertype=abstract.

Yuzawa, S., Kim, W., Katz, L. & Keasling, J.D. 2012. Heterologous production of polyketides by modular type I polyketide synthases in Escherichia coli. *Current Opinion in Biotechnology*,
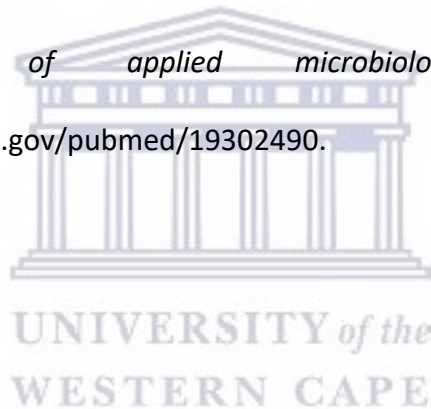
References

23(5): 727–735. http://dx.doi.org/10.1016/j.copbio.2011.12.029.

Zhang, H., Boghigian, B. a, Armando, J. & Pfeifer, B. a. 2011. Methods and options for the heterologous production of complex natural products. *Natural product reports*, 28: 125– 151.

Zhang, L., An, R., Wang, J., Sun, N., Zhang, S., Hu, J. & Kuai, J. 2005. Exploring novel bioactive compounds from marine microbes. *Current opinion in microbiology*, 8(3): 276–281. http://www.ncbi.nlm.nih.gov/pubmed/15939350.

Zhang, W., Zhang, F., Li, Z., Miao, X., Meng, Q. & Zhang, X. 2009. Investigation of bacteria with polyketide synthase genes and antimicrobial activity isolated from South China Sea sponges. *Journal of applied microbiology*, 107(2): 567–75. http://www.ncbi.nlm.nih.gov/pubmed/19302490.

References