The isolation and characterization of novel natural products from marine bacterial symbionts

UNIVERSITY *of the*
WESTERN CAPE

Timothy Matsiko Ninsiima Klein

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 (Tuffin)

Date: 14th May 2015

Declaration

I, Timothy Matsiko Ninsiima Klein, hereby declare that "**The isolation and characterisation of novel natural products from marine bacterial symbionts**" 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.

Date: ……

Signed: ……

Abstract

Drug-resistant infections are a global health crisis and drastically hinder the treatment options to effectively combat disease. Today, natural products remain an important source of novel drug candidates. Micro-organisms, in addition to being a source of bioactive natural products, represent a sustainable source of these compounds. As the marine environment is largely underexplored, the oceans represent a potential source of novel NPs.

This study aimed at the discovery of novel NPs from bacteria associated with novel marine invertebrate species endemic to the South African coast, including a sponge *Spongia* (*Spongia*) sp. 001RSASPN and a tunicate, *Pseudodistoma africanum* Millar, 1954. The methodology comprised of culture-dependent and culture-independent strategies. The former involved the isolation of bacteria associated with the invertebrate species and subsequent screening for anti-microbial activity against a panel of indicator strains including a multi-drug resistant *E. coli* strain. Anti-bacterial activity was detected in 6.1% and 4% of bacterial isolates from the sponge and tunicate isolates respectively.

The culture-independent strategy involved the use of PCR to select bioactive strains likely to contain novel NRPS or PKS secondary metabolite pathways. An NRPS A-domain exhibiting low sequence identity (65%) to reference sequences in the NCBI database was amplified from isolate PE8-15, a strain belonging to the genus *Bacillus*. This predicted a novel NRPS pathway within this strain. In addition, this isolate exhibited the most diverse anti-microbial profile including anti-bacterial and anti-fungal activity (*A. fumigatus* ATCC 46645). Therefore, as the most promising candidate, the genome of

Abstract

PE8-15 was sequenced following which 10 secondary metabolite pathways including bacteriocins (5), NRPS (3), siderophore (1) and a terpene pathway were identified. The A-domain amplified from PE8-15 originated from Cluster 4, and NRPS pathway predicted to encode a lipopeptide. Lipopeptides are an important class of compounds with a range of industrial applications in the pharmaceutical, cosmetic as well as food industry.

The identification of potentially novel secondary metabolite pathways from even well-studied groups of organisms demonstrates the importance of sequence-based methods in natural product discovery. Furthermore, this study highlights the South African coast as a rich source of microbial natural products and should be exploited further for drug discovery.

Key words: *drug-resistance, natural products, marine invertebrates, micro-organisms, whole-genome sequencing, genome mining*, *NRPS, lipopepetide*.

Acknowledgements

Firstly, I would like to thank my supervisor Prof. Tuffin for the exciting opportunity to take part in this project and for all the excellent training I have received.

Secondly, I would like to thank Lonnie van Zyl for the laboratory training and the constructive discussions during the course of this project. Dr. Goodman and the rest of the IMBM staff (Carmen, Dr. Kirby and Manillize) for all their support.

I would also like to thank Dr. William Mavengere for his patience and laboratory training, it was a great pleasure.

Thanks to all the IMBM members that made this experience both entertaining and bearable.

Special thanks to Mapula, Wesley and Gerda for their help and input as well as Rele and Whitney for the company during the long hours.

Lastly, I would like to thank my family for all the support and encouragement especially my mom without whom this opportunity would not have been possible. Thank you.

Abbreviations

°C	Celsius
µF	Microfarad
µg	Microgram
µg/ml	Microgram per millilitre
µl	Microlitre
µm	Micrometre
µM	Micromolar
aa-tRNA	Amino acid transfer RNA
ACM	Activated charcoal medium
ACP-domain	Acyl carrier protein domain
A-domain	Adenylation domain
AIDS	Acquired immune deficiency syndrome
AMT-domain	Aminotransferase domain
antiSMASH	Antibiotics and secondary metabolite analysis shell
AT-domain	Acyl-transferase domain
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosome
BLAST	Basic local alignment tool
bp	Base pair
cDNA	Complementary DNA
C-domain	Condensation domain

Abbreviations

CI	Chloroform isoamyl alcohol
Contig	Contiguous
CTAB	Cetyltrimethylammonium bromide
Cy-domain	Cyclization domain
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytidine 5'-triphosphate
DEBS	6-deoxyerythronolide B synthase
dGTP	2'-deoxyguanosine 5'-triphosphate
dH_2O	Deionized water
DH-domain	Dehydratase domain
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphates
dTTP	2'-deoxythymidine 5'-triphosphate
E-domain	Epimerization domain
EDTA	Ethylenediaminetetraacetic acid
ER-domain	Enoylreductase domain
FAS	Fatty acid synthase
g	Gravitational force
g/L	Grams per litre
GYM	Glucose yeast malt medium
HGT	Horizontal gene transfer
HIV	Human immunodeficiency virus
HMA	High microbial abundance

Abbreviations

HPLC	High-performance liquid chromatography
IMBM	Institute for Microbial Biotechnology and Metagenomics
Kb	Kilobase pairs
KCl	Potassium chloride
Km	Kilometres
KR-domain	Ketoreductase domain
KS-domain	Ketosynthase domain
KS_{α}	Ketosynthase alpha subunit
KS_{β}	Ketosynthase beta subunit
Kv	Kilovolts
L	Litre
LB	Lysogeny broth
LMA	Low microbial abundance
M	Molar
Mbp	Megabase pair
M-domain	Methylation domain
Mg/ml	Milligrams per millilitre
Min	Minute(s)
ml	Millilitre
mM	Millimolar
modAA	Modified amino acid
MRSA	Methicillin resistant *Staphylococcus aureus*
MS	Mass spectroscopy

Abbreviations

NaPDos	Natural product domain seeker
NCBI	National Centre for Biotechnology Information
ng	Nanogram
nm	Nanometres
NMR	Nuclear magnetic resonance
NP	Natural product
NRP	Non-ribosomal peptide
NRPS	Non-ribosomal peptide synthase
OD	Optical density
ORF	Open reading frame
OSMAC	One strain many compounds
OTU	Operational taxonomic unit
PCI	Phenol/Chloroform/Isoamyl alcohol
PCI	Phenol chloroform isoamyl alcohol
PCP-domain	Peptidyl carrier protein domain
PCR	Polymerase chain reaction
PKS	Polyketide synthase
PPTase	Phosphopantetheine transferase
RNA	Ribonucleic acid
rpm	Revolutions per minute
rRNA	Ribosomal ribonucleic acid
SCUBA	Self-contained underwater breathing apparatus
Sec	Seconds

Abbreviations

SM	Secondary metabolite
TAE	Tris-Acetate-EDTA buffer
TAR	Transformation associated recombination
T-domain	Thiolation domain
TE	Tris-EDTA buffer
TE-domain	Thioesterase domain
U	Unit
UV	Ultraviolet
v/v	Volume per volume
Volts/cm	Volts per centimetre
VRE	Vancomycin-resistant enterococci
w/v	Weight per volume
WGS	Whole genome sequencing
x-gal	5-bromo-4-chloro-3-indolyl-beta-D-galacto- pyranoside

Table of contents

Table of contents

Table of contents

Table of contents

List of figures

List of tables

List of tables

Chapter 1 Literature review

1.1 Introduction

The rise in drug-resistant pathogens is currently an alarming global health concern (Spellberg *et al*., 2008). In the United States alone, approximately 100,000 individuals die annually from infections caused by antibiotic resistant bacteria (Wong *et al*., 2012). For example, methicillin-resistant *S. aureus* (MRSA) infections are responsible for more deaths than those from both AIDS and tuberculosis (Wong *et al*., 2012). In developing communities where sanitation is frequently inadequate, outbreaks of cholera (caused by *V. cholera* O1 and O139 strains) are a major public health concern particularly due to the occurrence of drug resistant strains during epidemics (Akoachere *et al*., 2013). Furthermore, in South Africa there is currently an increase in tuberculosis resistance as a result of the high prevalence of HIV infected individuals (O'Donnell *et al*., 2010).

Table 1.1: Frequently encountered drug-resistant bacteria (Levy, 2002)

Hospital	Community
Gram-negative	
Acinetobacter sp.	*E. coli*
Citrobacter sp.	*Neisseria gonorrhoeae*
Enterobacter sp.	*S. typhi*
Klebsiella sp.	*Salmonella typhimurium*
P. aeruginosa	
Serratia marcescens	
Gram-positive	
Enterococcus sp.: vancomycin-resistant enterococci (VRE)	*Enterococcus* sp.: VRE
coagulase-negative *Staphylococcus*	*Mycobacterium tuberculosis*
MRSA	MRSA
MRSA heterogeneously resistant to vancomycin	*Streptococcus pneumoniae*
	Streptococcus pyogenes

A major contributing factor to the spread of drug-resistant pathogenic strains is antibiotic misuse. A classic example on the misuse of antibiotics was demonstrated by Amidi and colleagues (1975) investigating the prescription of antibiotics among 40 doctors in Shiraz (Iran). The study involved a healthy individual claiming to have classical symptoms of viral urinary tract infection. Out of 40 doctors visited, 37 (92.5%) still prescribed antibiotics following the diagnosis of a viral infection and the majority of the drugs were prescribed at inadequate doses. More recent studies have shown antibiotic misuse to still be a major problem in health care facilities (Thriemer *et al*., 2013; Means *et al*., 2014). In addition to misuse, an increase in immuno-compromised patients and antibiotic treated food are major contributing factors in the spread of drug resistance

(Levy, 2002). Therefore, without strict measures in place to control the use of antibiotics, the spread of resistant pathogens will persist and compromise public health. In light of these, there is an urgent, but also continuous need for the discovery and development of new drug candidates to combat drug resistance.

1.2 Natural products and drug discovery

Natural products (NPs) are compounds produced by living organisms such as fungi, bacteria and plants (Ji *et al*., 2009). NPs are usually produced as a result of secondary metabolism and although not essential for growth, improve the survival of the producing organism (Williams *et al*., 1989). NPs have been exploited for many years in the treatment of a myriad of ailments and diseases due to the broad range of biological activities they possess. Clay tablets from Mesopotamia 2600 B.C (modern day Iraq) are amongst the oldest records depicting the use of NPs (oils from Cypress and Myrrh) as medicine which are still being used today as remedies against coughs and inflammation (Dias *et al*., 2012). In fact, the isolation of chemicals with useful properties from preparations of 'ethno-medicinal' sources was the foundation of the modern pharmaceutical industry (Pomponi, 2001). Today, NPs remain a major resource for drug discovery and during the past 30 years, 70% of anti-microbials and 60% of chemotherapeutics have been developed or analogously synthesized from NPs (Pomponi, 2001; Grüschow *et al*., 2011). In comparison to synthetic compounds, NPs offer various advantages over synthetic compounds. For example, NPs exhibit greater structural diversity than synthetic compounds and are naturally designed by nature to interact with biological targets such as proteins (Zhang and Demain, 2005).

Initially, plants represented a major source of medicinal compounds exemplified by taxol an anti-cancer drug from *Taxus brevifolia* and etoposide from *Podophyllum peltatum* (Proksch *et al*., 2003). However, microbes were recognised as a useful source of NPs following the discovery of penicillin in 1928 and its later use as an antibiotic (Davies and Davies, 2010; Berdy, 2005). The field of microbial NP discovery was later expanded by Selman Waksman whose work involved screening of soil microbes for antagonistic compounds (Woodruff, 2014; Reilly *et al*., 1945; Waksman and Woodruff, 1941). His work led to the discovery of numerous antibiotics commonly used today, a period referred to as the 'Golden era of antibiotics' and is depicted in figure 1.1 (Stallforth and Clardy, 2014).

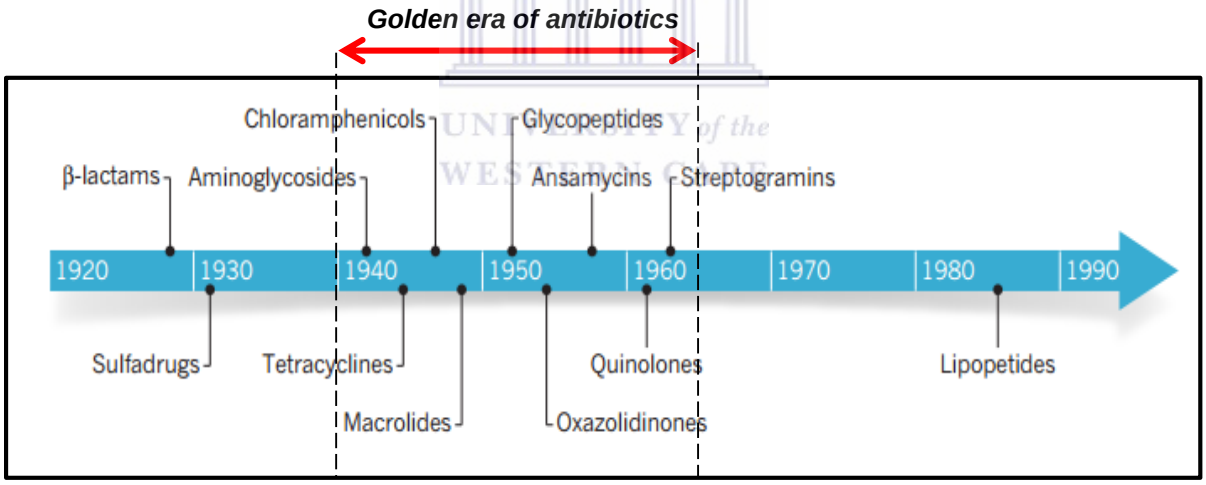

Figure 1.1: Timeline depicting the discovery of new antibiotic classes. The discovery of penicillin (β-lactams) pioneered the discovery of pharmaceutically relevant microbial antibiotics that peaked in 1950's. The period between the 1940's and 1960's saw the discovery of a large number of antibiotics and is a period referred to as the 'Golden era' of antibiotics. However, the period after the 1960's severely lacked any notable drug discoveries with the lipopeptides emerging as the only new class 20 years later (*Adapted from* Lewis, 2012).

1.3 Natural product classes of pharmaceutical relevance

Pharmaceutically relevant NPs belong to different chemical classes that differ not only in structure, but also in the mechanisms by which they are synthesized. Important classes of NPs include; terpenoids, alkaloids and various ribosomal peptides with a wide range of bioactive properties (Graça *et al.,* 2013). However, for the purpose of this study, NPs belonging to the polyketide and non-ribosomal peptide classes are of particular importance.

1.3.1 Polyketides

Polyketides are a large class of NPs with important biological activities that include; immuno-suppression, anti-cancer and anti-microbial (Chen *et al.,* 2006). As a result of polyketide structural and functional diversity (figure 1.2), compounds in this family are an attractive source of novel drug candidates.

Assembly of the polyketide backbone is orchestrated by the polyketide synthase (PKS) enzyme family (Kealey *et al*., 1998). These enzymes are categorized into three different types (type I, type II and type III) based on their products as well as biochemical features (Gao *et al*., 2010).

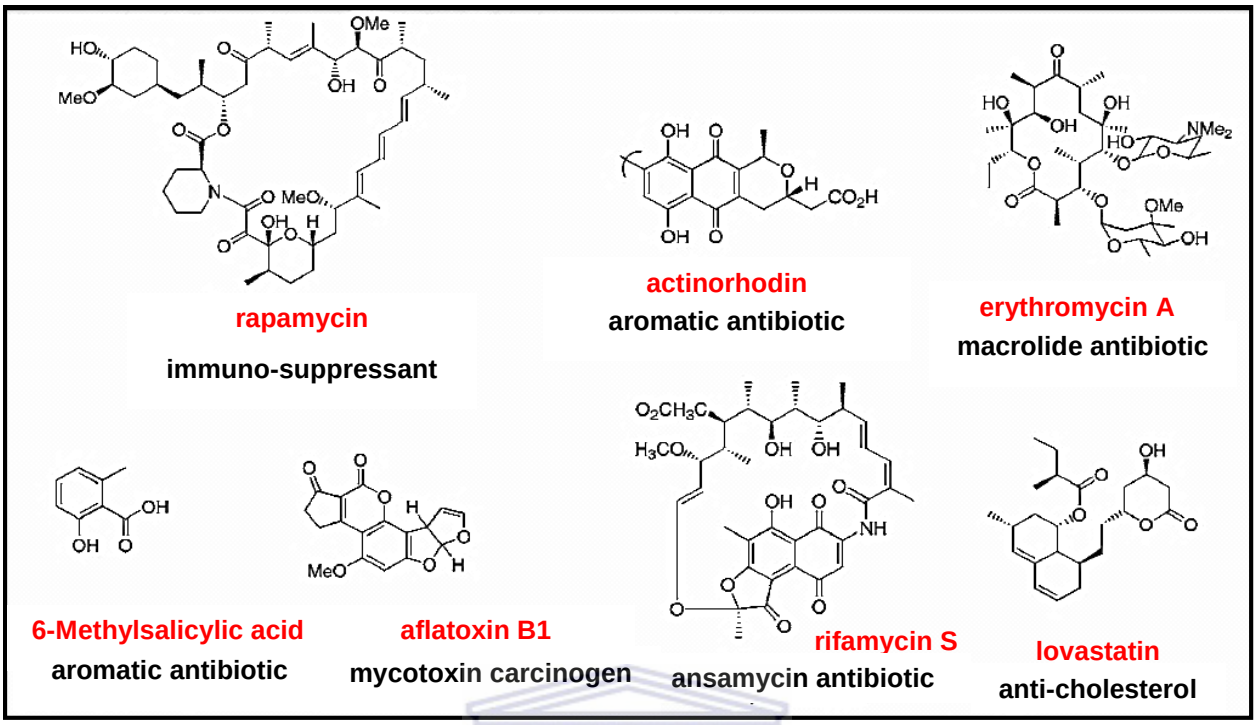

Figure 1.2: Structural and functional diversity of the polyketide compounds. Examples shown include: rapamycin (immuno-suppressant), actinorhodin (aromatic antibiotic), erythromycin A (macrolide antibiotic), 6-Methylsalicylic acid (aromatic antibiotic), aflatoxin (mycotoxin carcinogen), rifamycin S (ansamycin antibiotic) and lovastatin (anti-cholesterol) (*Taken from* Staunton and Weissman, 2001).

1.3.1.1 Type I polyketide synthases

Type I PKSs are a group of large multi-functional enzymes exhibiting modular architecture (Cheng *et al.,* 2003). A module is a catalytic unit (figure 1.3) consisting of domains (in varying combinations) needed for the successful incorporation of monomers and modification of the growing chain (Ruan *et al*., 1997; Bergendahl *et al*., 2002). Three compulsory domains are required for successful biosynthesis of type I polyketides and include; acyl-transferase (AT), acyl-carrier protein (ACP) and a ketosynthase (KS) (Schirmer *et al.,* 2005). Additional domains may be present that play

a role in modification of the polyketide chain. Domains that are involved in chain modification include; ketoreductases (KR), enoylreductases (ER) and dehydratases (DH) (Gao *et al*., 2010). The KR-domain plays a role in reducing the β-keto group (formed after each condensation reaction) to a hydroxyl group that is subsequently reduced by the DH-domain to form a double bond that can be further reduced by the ER-domain (Keatinge-Clay and Stroud, 2006).

Type I polyketide biosynthesis is initiated following the activation of a starter unit (usually a molecule with an acyl group, such as malonate or acetate) by the AT-domain and its subsequent transfer to the ACP-domain (Lim *et al*., 2012). The ACP-domain in turn is responsible for shuttling the acyl units onto the KS-domain which carries out a decarboxylative claisen condensation reaction between the starter and extender units (Crawford and Townsend, 2010). Activation of the ACP-domain requires the addition of a 4'-phosphopantetheine group as a post-translational modification by the phosphopantetheine transferase enzyme (PPTase) (Walsh *et al*., 1997). Once chain elongation is complete, the final product is released by the catalytic action of a thioesterase domain (TE) usually located at the C-terminal of the last module (Cane and Walsh, 1999).

A characteristic feature of type I PKSs is the single use of each domain involved and their role in synthesizing macrolide compounds (Novakova *et al.,* 2002). However, in contrast to bacterial type I PKS, fungal type I PKSs make use of each active site iteratively and are responsible for the production of aromatic compounds (Crawford and Townsend, 2010). Examples of bacterial type I PKS compounds include; rifamycin, rapamycin, epothilone and erythromycin (Khosla *et al*., 1999; Ginolhac *et al*., 2004).

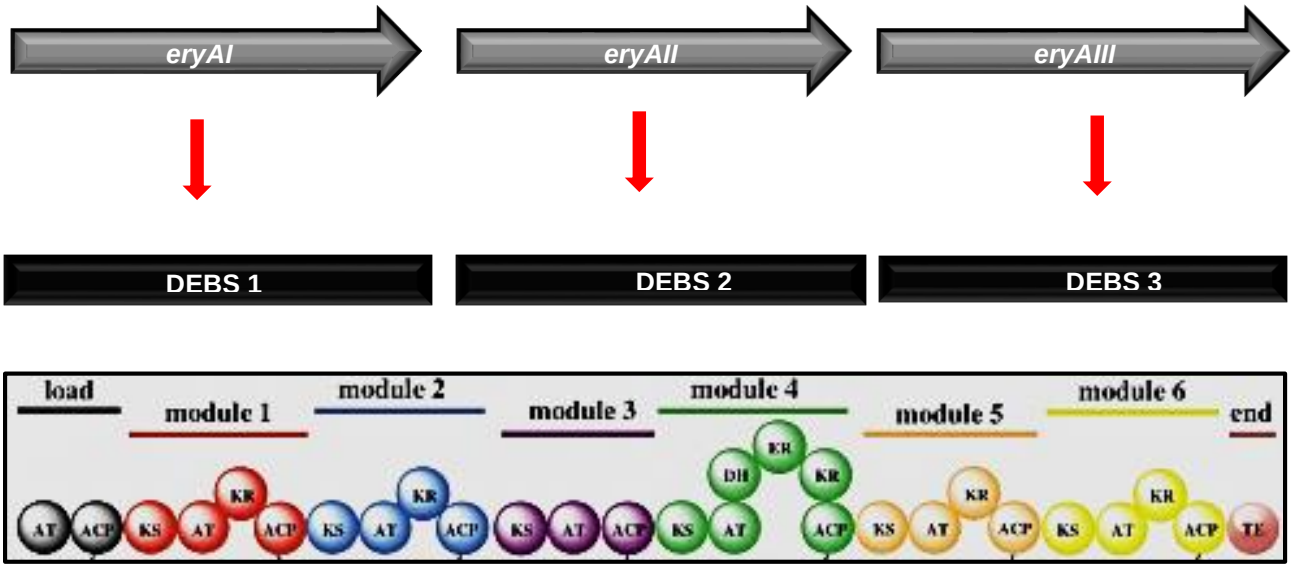

Figure 1.3: Modular organisation of the erythromycin gene cluster, a type I PKS. The PKS involved in the production erythromycin, is encoded by three genes *eryAI*, *eryAII* and *eryAIII*. The erythromycin genes code for the proteins DEBS 1, DEBS 2 and DEBS 3 respectively. Each module is uniquely colour coded and contains the compulsory domains KS, AT and ACP. Optional domains (KR and DH) involved in backbone modification, are present in all modules except module 3. A thioesterase (TE) domain is located downstream (C-terminal) of the final module (module 6) and functions as a chain terminator releasing the final product (*Adapted from* Staunton and Weissman, 2001).

1.3.1.2 Type II polyketide synthases

In contrast to the type I PKS family of enzymes, bacterial type II PKSs are mono-functional enzymes whose active sites are distributed among individual protein subunits (figure 1.4) that congregate to form complexes (Khosla *et al.,* 1999). Each active site in type II PKSs is used iteratively producing compounds with aromatic rings, the aromatic polyketides (Staunton and Weissman, 2001; Seow *et al*., 1997). The core set of domains required for the successful synthesis of aromatic polyketides include an acyl carrier protein (ACP) and two ketosynthase (KS) genes KS_{α} and KS_{β} (King *et al*., 2009).

The KS_{β} domain, also referred to as the chain length factor, plays a role in determining the length of the polyketide chain (Tang *et al*., 2003). Conversely, the KS_{α} domain condenses the acyl substrates whereas the ACP-domain serves as an anchor for the growing polyketide chain and is responsible for receiving the malonyl substrates (Revill *et al*., 1995). Although AT-domains are not present within type II PKS gene clusters, ACP-domains have been shown to exhibit self malonyl-transfer activity (Arthur *et al*., 2006). Important aromatic polyketides include doxorubicin, tetracenomycin and actinorhodin (Khosla *et al*., 1999; Summers *et al*., 1993).

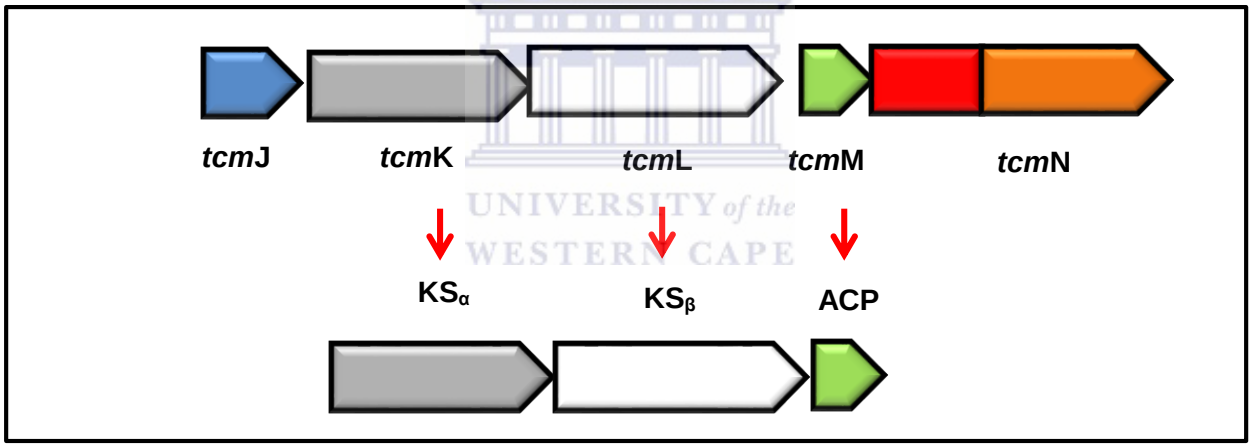

Figure 1.4: Illustration of type II PKS genetic organisation exemplified by the tetracenomycin gene cluster from *Streptomyces glaucescens*. Each domain, KS_{α}, KS_{β} and ACP (encoded by *tcm*K, *tcm*L and *tcm*M respectively), is located on an individual gene. This is in sharp contrast to the modular type I PKSs where domains are located on large proteins encoded by a single gene. The individual domains of type II PKS subsequently form dissociable protein complexes that are iteratively (repetitively) used during biosynthesis (*Adapted from* Seow *et al*., 1997).

1.3.1.3 Type III polyketide synthases

Type III PKSs make up a family of enzymes predominantly found in higher plants, however, bacterial type III PKSs were first discovered in *Streptomyces griseus* (Funa *et al*., 1999). In comparison to type I and II PKSs, this class of PKSs is relatively simple with a homodimeric structure where each monomer is responsible for the iterative loading of substrates, extension and carrying out cyclization reactions (Yu *et al*., 2012). In addition, type III PKSs, unlike the type I and type II PKS enzymes use free coA substrates as opposed to 4'-phosphopantetheine bound substrates (Saxena *et al*., 2003). Although relatively simple in comparison to other PKS classes, type III PKS produce diverse compounds such as chalcones, pyrones and stilbenes (Yu *et al*., 2012).

1.3.2 Non-ribosomal peptides

Non-ribosomal peptides (NRPs) constitute a large group of microbial NPs which exhibit useful biological activities including anti-inflammatory, anti-viral, biosurfactants, anti-microbial and anti-tumour (Stachelhaus *et al.*, 1998; Caboche *et al*., 2010). They are small polypeptides usually consisting of fewer than 50 amino acids, whose synthesis is engineered by a family of enzymes known as the non-ribosomal peptide synthases (NRPSs) (Neilan *et al.*, 1999; Walsh, 2008).

NRPs are built from proteinogenic amino acids in the 'L' form, although unusual components including hydroxyl acids, non-proteinogenic amino acids and amino acids in the 'D' form are often observed (Lee *et al.*, 2005; Marahiel *et al*., 1997). NRPs

currently in clinical use include; penicillin, cephalosporins and vancomycin that is an antibiotic used as a last line of defence against MRSA infections (Micek, 2007). In addition, polymyxin E is extensively used to combat lung infections caused by *P. aeruginosa* (Hancock and Chapple, 1999). Members of the gram-positive *Actinomycetes* and the genus *Bacilli* are major producers of NRPs (Finking and Marahiel, 2004).

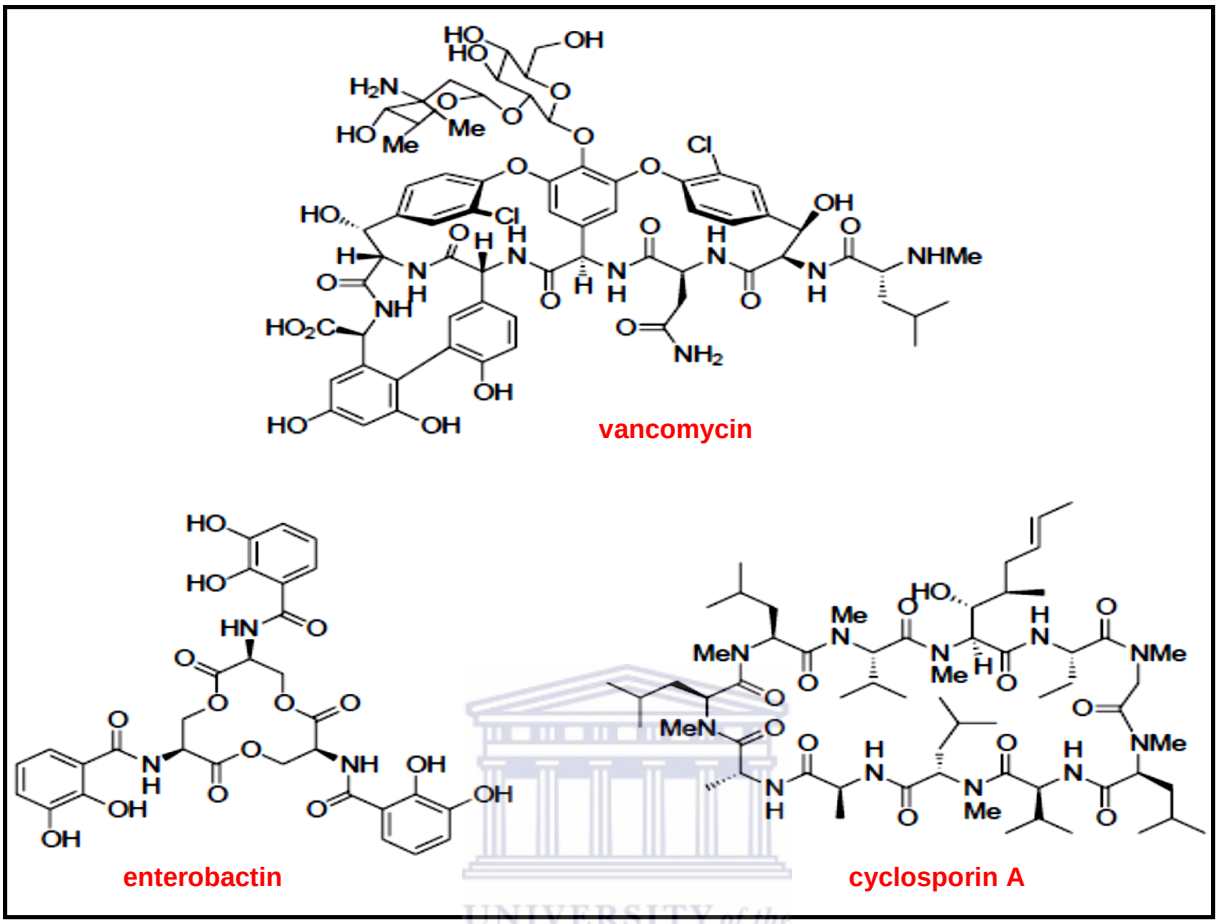

Figure 1.5: Examples of non-ribosomal peptides and their structural diversity. Vancomycin, a last line of defence antibiotic against MRSA, cyclosporin is a potent immune-suppressant (Borel *et al*., 1976) and enterobactin is an iron-chelating agent (siderophore) facilitating iron assimilation from the environment (Gehring *et al*., 1998). (*Taken from* Challis and Naismith, 2004).

1.3.2.1 Biosynthesis of non-ribosomal peptides

The ribosome-independent synthesis of NRPs (Neilan *et al.,* 1999) was first hypothesized in the 1960's when cell free extracts of peptide antibiotic producing organisms were studied. These studies demonstrated the restoration of antibiotic production even in the presence of RNAses and ribosome inhibitors (Gevers *et al.*, 1968; Mach *et al*., 1963).

NRPSs exhibit a modular organization whereby each module (figure 1.6) is a catalytic unit responsible for the incorporation of one amino acid into the backbone of the resulting product (Challis and Naismith, 2004). For successful biosynthesis of the peptide product, each NRPS module requires a set of core domains that include; a condensation (C), adenylation (A) and thiolation (T) or peptidyl carrier protein (PCP) (Felnagle *et al*., 2008). In addition, NRPSs may contain secondary domains responsible for the modification of the peptide chain and include; an epimerization (E) domain involved in altering the chirality of the amino acid, a methylation (M) domain involved in addition of methylation groups and a heterocyclization (Cy) domain (Lautru and Challis, 2004). Once chain elongation is complete, the final product is released by the catalytic action of the thioesterase domain (TE) (Ehrenreich *et al*., 2005).

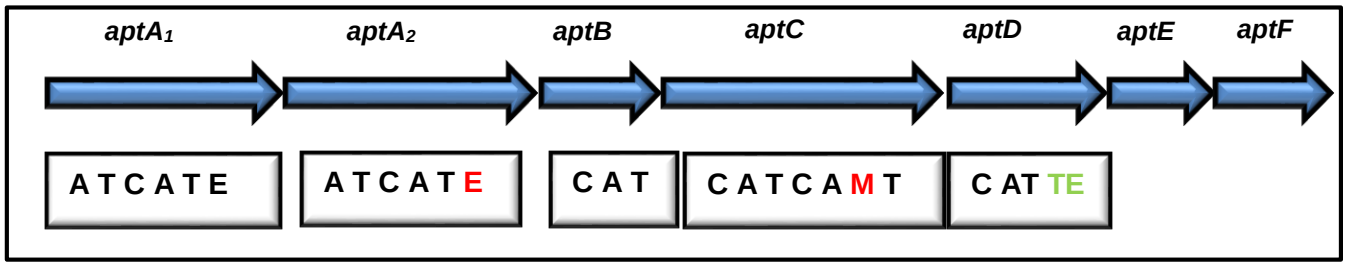

Figure 1.6: Modular organization of NRPSs exemplified by the anabaenopeptin gene cluster. Genes involved in NRP biosynthesis frequently form clusters and are further organized into modules containing the condensation, adenylation and thiolation domains respectively (C-A-T). Domains variably present within each module can include a methylation (M) and epimerization (E). The thioesterase domain (TE) plays a role in cyclisation and release of the final peptide chain (*Adapted from* Rouhiainen *et al*., 2010).

During NRP biosynthesis, the A-domain is responsible for substrate specificity (Weber *et al.*, 2000) and activating the cognate amino acid to an aminoacyl adenylate using ATP (Stachelhaus *et al.*, 1999). This mechanism is analogous to the aa-tRNA synthetase in ribosome-dependent protein synthesis (May *et al.*, 2002). The activated amino acid is subsequently transferred and covalently bound to the PCP-domain where it can undergo modifications by additional domains (Weber *et al.*, 2000). The final step during NRP biosynthesis is catalysed by the C-domain responsible for peptide-bond formation between activated amino acids bound to the PCP-domains of adjacent modules (Weber *et al.*, 2000).

1.3.3 Polyketide and non-ribosomal peptide hybrid compounds

This is an interesting class of NPs that consists of compounds assembled from both NRP and polyketide elements (Weinig *et al*., 2003). Although NRPs and polyketides differ structurally and in their building blocks, the respective biosynthetic enzymes exhibit structural and catalytic similarities that enable the synthesis of these hybrid products (Du *et al.,* 2001). Important examples of hybrid compounds include epothilone an anti-tumour agent (Chen *et al*., 2001), yersiniabactin a siderophore (Haag *et al*., 1993), barbamide an anti-mollusc compound (Chang *et al*., 2001) as well as rapamycin an immuno-suppressant (Law, 2005).

1.4 Natural products from the sea

Spanning approximately 70% of the earth's surface, oceans hold the largest share of the ecosystem (Glöckner and Joint, 2010). The rich biodiversity contained within the oceans (15 animal phyla exclusive to the oceans) makes them a unique and rich drug discovery reservoir (Leal *et al*., 2012).

Marine NP discovery initially was focused on the easily accessible macro-organisms (such as algae, soft corals and sponges) from which a range of bioactive NPs were isolated. For example, the discovery of unusual nucleosides from the sponge *Cryptotethia crypta* in the early 1950s led to the development of the potent anti-cancer and anti-viral analogues, Ara-C and Ara-A respectively (Bergmann and Feeney, 1951; Hu *et al*., 2011). Ziconotide, also of marine origin, is a peptide isolated from a marine mollusc (*Conus magus)* and is used as a non-addictive drug for the treatment of chronic

pain (McGivern, 2007). With marine invertebrates making up 60% of the marine animal diversity (Leal *et al*., 2012), there is numerous potential for the discovery of novel drug leads in the oceans.

However, marine NP discovery efforts gradually turned to the smaller life forms such as bacteria and fungi (Gerwick and Moore, 2012) that constitute a large portion of the marine biomass (Sogin *et al*., 2006). Of particular interest are the micro-organisms associated with eukaryotic hosts. These invertebrate-associated microbes are postulated to play a number of roles for their hosts such as chemical defence against predators and competitors (Haygood *et al*., 1999).

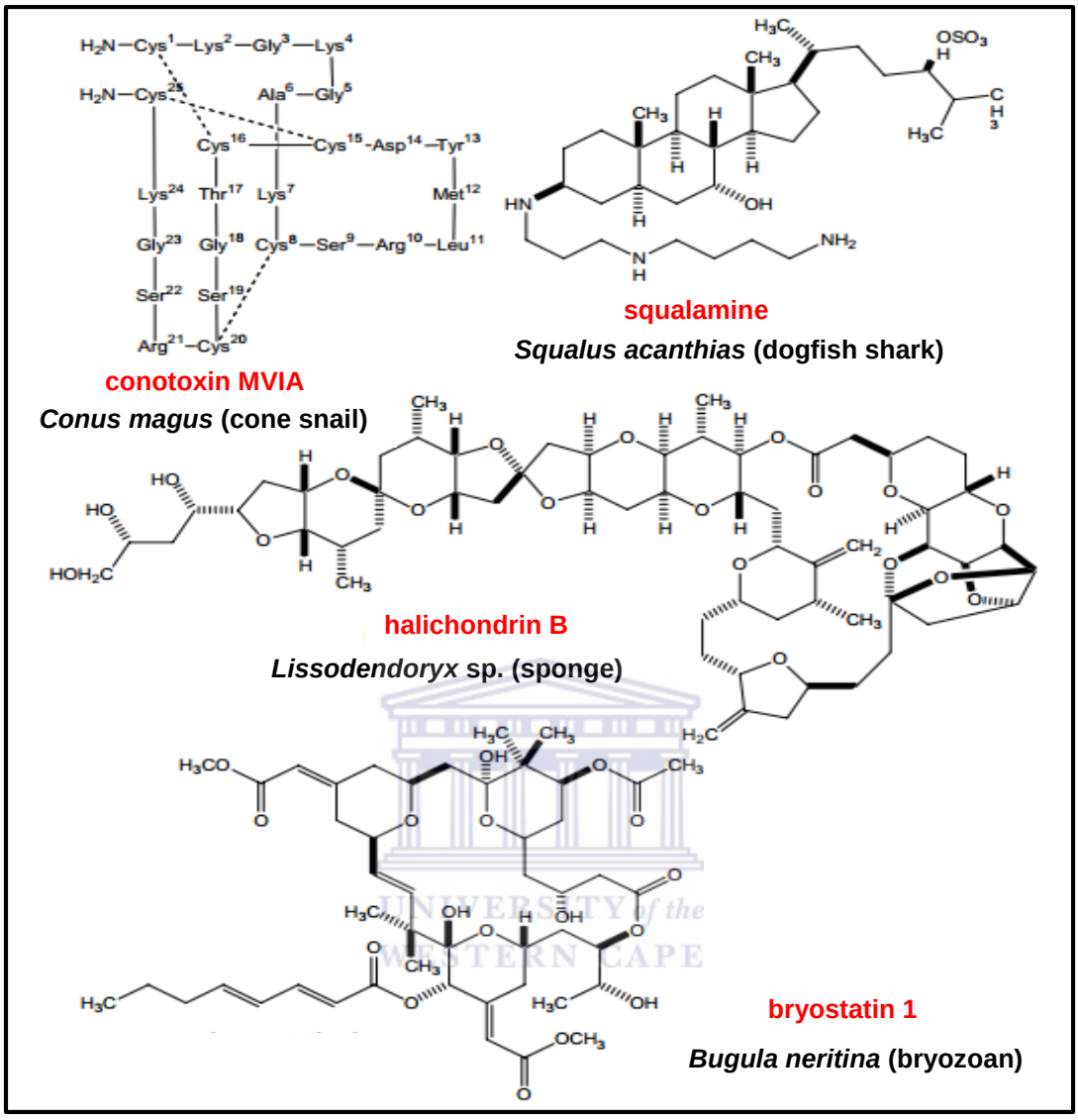

Figure 1.7: Natural products isolated from marine animals. Examples shown include conotoxin MVIIA (Ziconotide) from a cone snail (*Conus magus*), squalamine isolated from a dogfish shark (*Squalus acanthias*), halichondrin B of marine sponge origin (*Lissodendoryx* sp.) and bryostatin isolated from a bryozoan (*Bugula neritina*) (*Taken from* Proksch *et al*., 2003).

1.5 Marine invertebrates

Although marine NPs have been isolated from a range of marine organisms, for the purpose of this study, the phylum Porifera and subphylum Tunicata are of particular interest.

1.5.1 Sponges

Sponges are the oldest animals in the phylum Porifera dating back nearly 580 million years (Fieseler *et al.,* 2004). They are sessile filter-feeders making up a dominant benthic community in the ocean (Leys and Eerkes-Medrano, 2006). Although the majority of sponges inhabit the marine environment, approximately 1% of these invertebrates are found in fresh water bodies (Belarbi *et al.,* 2003).

Sponges have an aquiferous body through which they continuously filter large volumes of sea-water for nutrient acquisition (Kennedy *et al*., 2007). As water flows through specialised canals within the sponge, bacteria (among other micro-organisms) are transferred to the mesohyl tissue (figure 1.8) where they are able to establish a symbiotic relationship with the host (Kennedy *et al*., 2007). In high microbial abundance sponges (HMA), micro-organisms may constitute as much as 40% of the sponge biomass estimated to be 10^8 to 10^{10} bacteria per gram of tissue (Selvin *et al*., 2009). These high bacterial densities in HMAs are in contrast to low microbial abundance sponges (LMA) that exhibit microbial densities ranging between 10^5 to 10^6 bacteria per gram of tissue (Hentschel *et al.,* 2006).

Sponges were established as a source of potent biologically active compounds (NPs) following the discovery of the uncommon nucleosides spongouridin and spongothymidin from the sponge *Cryptotethia crypta* (Bergmann and Feeney, 1951; Hu *et al*., 2011; Kijjoa and Sawangwong, 2004). Following their discovery, numerous bioactive compounds have been isolated from marine sponges, and these animals contribute the largest number of novel marine invertebrate NPs (Mehbub *et al*., 2014).

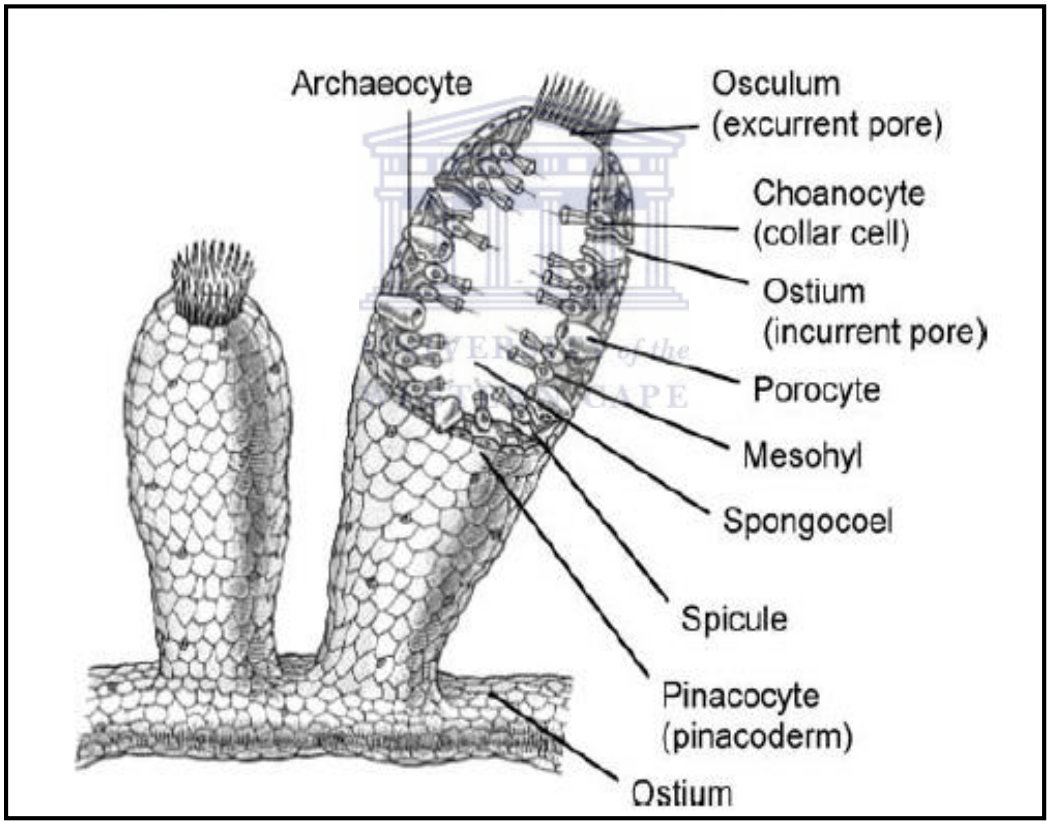

Figure 1.8: Simplified illustration of the sponge anatomy. The pinacoderm functions as a barrier between the outer and inner environment. The region between the inner and outer layer is the mesohyl layer. As the sea-water flows in through the incurrent pore, bacteria are transferred to the mesohyl tissue and subsequently ingested or form symbiotic associations with the sponge. Waste is released through the excurrent pore (*Taken from* Kennedy *et al*., 2007).

The phylum Porifera is comprised of three major taxonomic classes that include; Hexactinellida, Calcarea and the Demospongiae (Abdelmohsen *et al.,* 2014). The class Demospongiae is the largest class, consisting of over 80% of all known sponge species from which a myriad of novel compounds have been isolated (Thompson *et al*., 1985; Wang and Lavrov, 2008). Looking at the total number of NPs isolated from each order in the class Demospongiae (figure 1.9), it is clear that perhaps not all taxonomic groups of sponges are equally important sources of NPs. For example, in comparison to the orders Halichondria and Dictyoceratida, fewer NPs have been isolated from the orders Astrophorida and Hadromerida. Taking these observations into account, designing a more targeted sampling strategy might be more appropriate to improve NP discovery efforts. However, sampling bias could contribute to the differences in the number of NP discovered from different orders within the class Demospongiae.

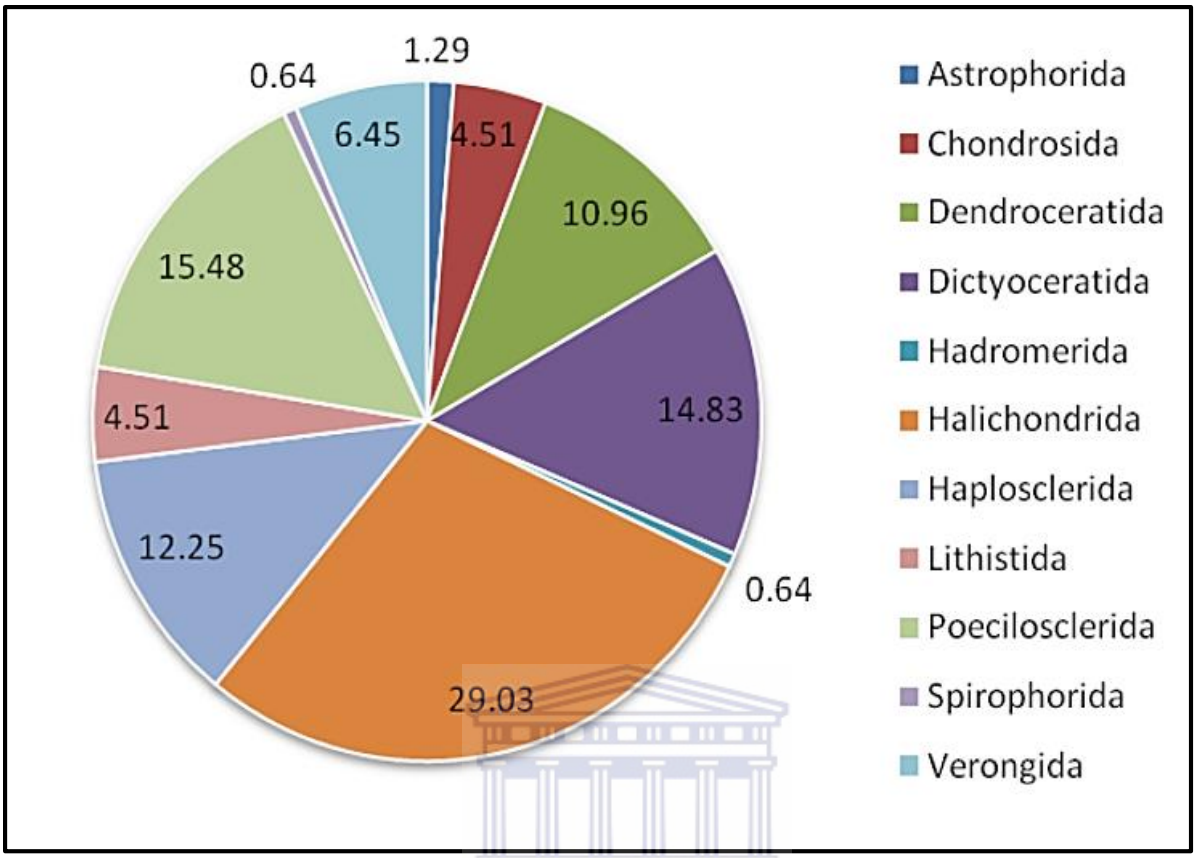

Figure 1.9: Percentage of natural products isolated from sponge-associated microbes represented according to the order in the class Demospongiae. The sponge orders Halichondrida, Poecilosclerida and Dictyoceratida contribute a total of 29.03%, 15.48% and 14.83% respectively of NPs from microbes associated with sponges. In certain sponge orders, fewer microbial NPs have been isolated and include the orders Chondrosida (4.51%), Astrophorida (1.29%), Hadromerida (0.64%), Lithistida (4.51%), and Spirophorida (0.64%) (*Taken from* Thomas *et al*., 2010).

1.5.2 Tunicates

Classified under the phylum Chordata, tunicates make up a diverse group of animals whose body is covered in a cellulose-containing tunic (Lemaire, 2011). Tunicates are the only animals with the ability to biosynthesise cellulose, a trait hypothesised to have been acquired from bacteria through horizontal gene transfer (HGT) (Sasakura *et al*., 2005). In a manner similar to sponges, tunicates obtain their food by a filter-feeding mechanism (Radford *et al*., 2000). They are usually sessile when adults and make use of an array of toxins for protection (Cooper and Yao, 2012).

In the context of drug discovery, tunicates are an important group of organisms because they are a source of numerous NPs with bioactive properties. For example, Didemnin B a strong anti-cancer compound that also possesses anti-viral activity (RNA and DNA viruses) was the first marine NP that progressed onto phase I clinical trials (Ankisetty *et al*., 2013).

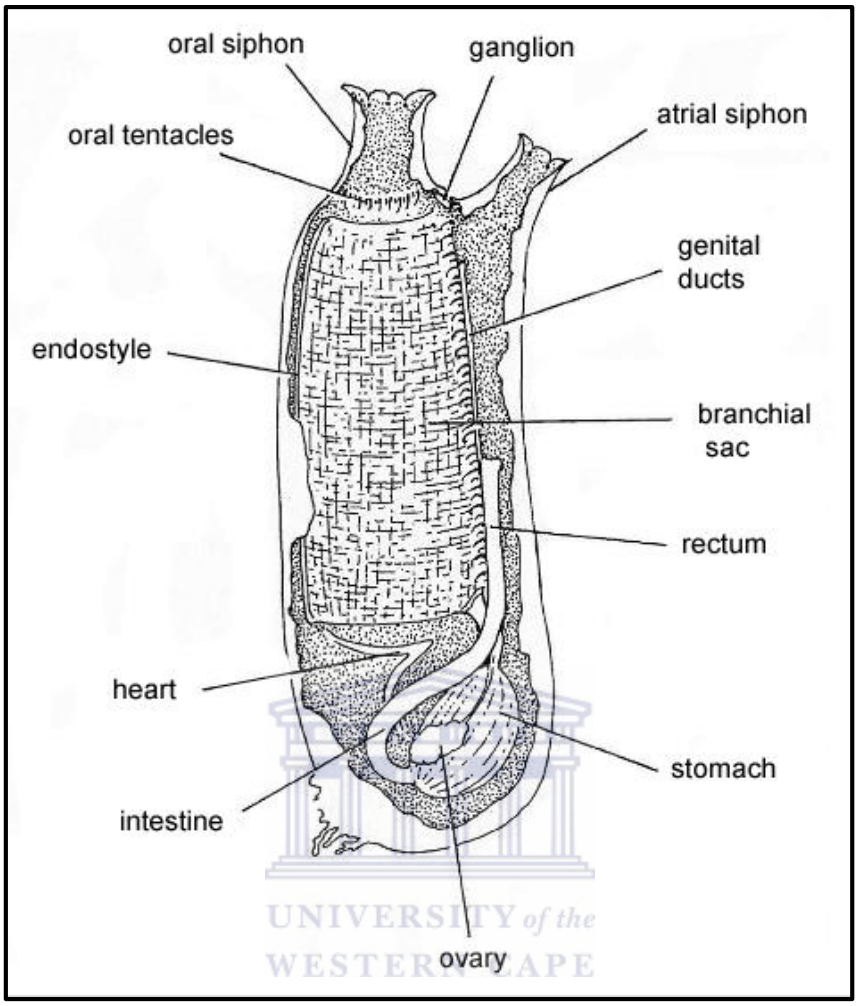

Figure 1.10: Simplified illustration of tunicate anatomy (*Ciona intestinalis* form Millar 1970). The oral siphon serves as an entry point for sea-water from which particulate matter (such as bacteria) is retained within the branchial sac and the waste excreted through the atrial siphon (Shenkar *et al*., 2015). (*Taken from* NOBANIS, 2015).

1.5.3 Marine invertebrate-associated bacteria

The filter-feeding nature of sessile marine invertebrates facilitates close contact between the hosts and micro-organisms. In addition to being a source of food, providing stability to the sponge skeleton and waste management, micro-organisms play a role in chemical defence to deter competitors and predators (Hentschel *et al*., 2002; Haygood *et al.,* 1999).

Due to the striking resemblance of compounds isolated from invertebrates to those considered exclusively of microbial origin, micro-organisms are thought to be the true producers of marine invertebrate NPs (Piel *et al*., 2004; Muscholl-Silberhorn *et al.,* 2008). For example, ET-743 (also known as Trabectedin/Yondelis), is an anti-cancer agent isolated from the tunicate *Ecteinascidia turbinata*. This compound shows structural similarity to the prokaryotic derived compounds, saframycin A, saframycin Mx1 and safracin B and strongly suggests bacterial origin as opposed to eukaryotic (tunicate) production as was initially suggested (Rath *et al*., 2011). A more recent study however, proved the bacterial origin of compounds in the marine sponge *Theonella swinhoei*. The marine sponge *T. swinhoei* produces an array of compounds of ribosomal origin (polytheonamides) as well as polyketides such as onnamide/theopederins (Piel *et al*., 2004; Hamada *et al*., 2005). Although originally thought to be produced by the sponge itself, Wilson and colleagues (2014) proved, using single-cell genomics and metagenomics that the polytheonamides and onnamides/theopederins are in fact produced by symbiotic bacteria. This symbiotic bacterium belongs to a novel and as yet uncultured bacterium of the candidate genus *Entotheonella*.

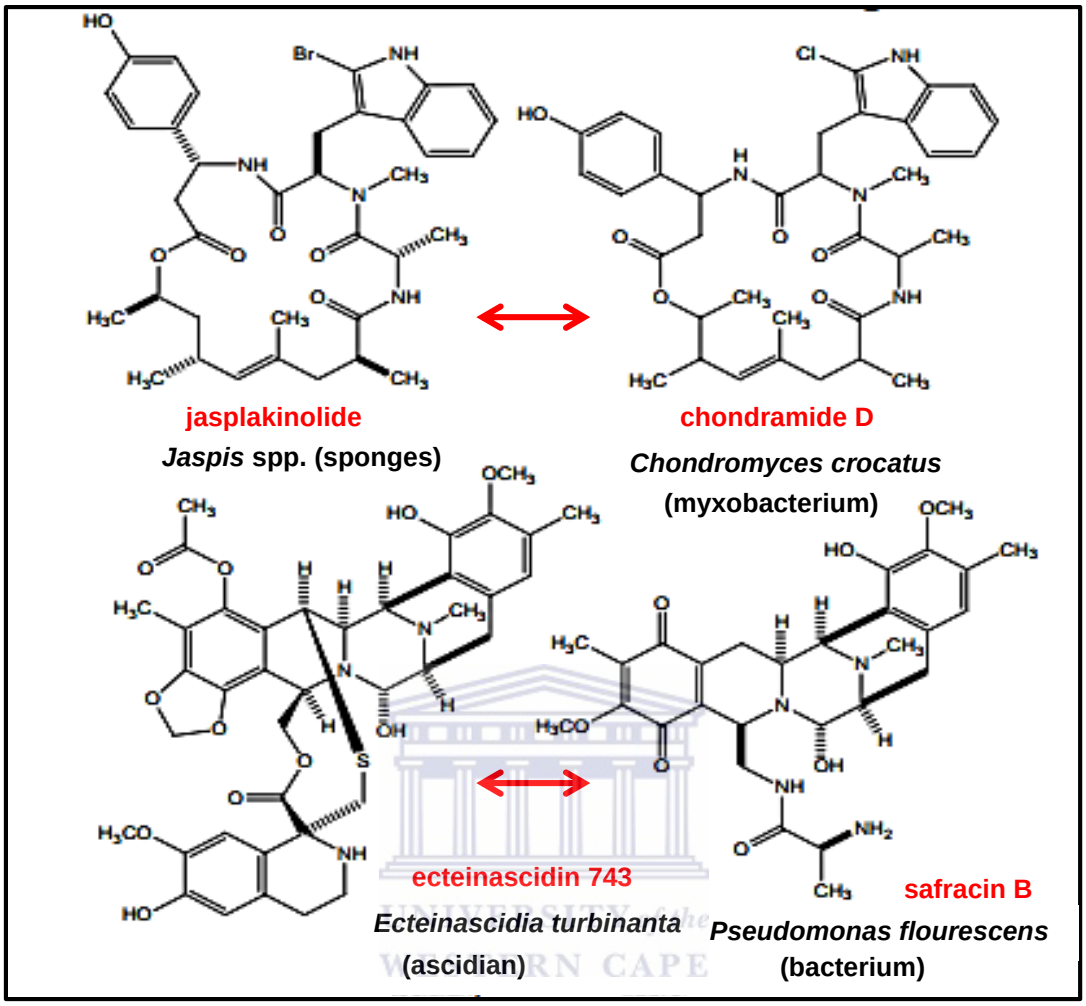

Figure 1.11: Structural similarity between marine compounds of microbial and invertebrate origin. Non-ribosomal peptides and polyketides were known exclusively to be of microbial origin. Therefore, the structural similarity of marine invertebrate compounds to those isolated from their symbiotic bacterial counterparts suggests microbes to be the true producers (*Taken from* Proksch *et al*., 2003).

1.5.3.1 Distribution of marine invertebrate-associated bacteria

It has been established that filter-feeders associate with micro-organisms within their tissues whose cell density frequently outnumbers that of the surrounding sea-water (Taylor *et al*., 2007). Marine invertebrates, in addition to selectively filtering out bacterial symbionts from the sea-water, are able to transmit specific microbial populations by vertical transmission to their progeny (Sharp *et al*., 2007). It is therefore expected that bacterial communities differ between invertebrates and do not necessarily reflect the communities in the surrounding sea-water.

Several studies have investigated the bacterial community structures and host-specificity of bacterial populations within marine invertebrate hosts. For example, Erwin and colleagues (2011) compared the bacterial communities in sea-water, two marine sponges (*Hymeniacidon heliophila* and *Haliclona tubifera*) and the tunicate *Didemnum* sp. Evidence of species-specific association between invertebrate hosts and microbes was evident in the two sympatric sponges that shared 4 operational taxonomic units (OTUs) of which 2 OTUs were shared with the ambient bacterioplankton communities. However, a similar study observed differences in the bacterial communities within the sponge *Cymbastela concentrica* harvested from tropical and temperate regions in Australia (Taylor *et al* 2005). These studies implicate other factors such as temperature as playing a role in shaping the bacterial community structures. Similar bacterial community studies were done on the tunicate *E. turbinata* by Peréz-Matos and colleagues (2007) that identified a consistent bacterial population between tunicates of the same species harvested in the Caribbean and Mediterranean Sea.

These studies, in addition to providing evidence of species-specific bacterial communities, implicate environmental factors (such as temperature) in shaping the bacterial communities.

In the context of NP discovery, results from bacterial community studies are important in formulating an appropriate sampling strategy. For example, novel invertebrate species would be considered a preferred source of novel NPs as they likely harbour unique microbial communities and hence unique functional diversity. Alternatively, because identical invertebrate species harvested from different geographical locations (for example temperate vs. tropical) can differ in their microbial communities, then previously studied organisms still hold potential to yield unique chemistry on condition that they are obtained from different geographical locations.

1.6 Bottlenecks in the discovery and development of natural products

Although various NPs with promising pharmaceutical relevance are being discovered, further study and their progression into clinical trials face various limiting factors. This section aims to highlight the major limitations encountered during NP discovery and the subsequent development into drugs.

The progression of NPs into the clinical trial stage requires a considerable amount of drug mass (usually kilogram amounts) and the lack of adequate drug mass is a major limiting factor (Tsukimoto *et al*., 2011). In some instances, obtaining sufficient quantities of the desired compound can only be achieved by harvesting large amounts of invertebrate tissue that is usually impractical and unrealistic. For example, to obtain

approximately 1 gram of ET-743 (Yondelis), 1 metric tonne of the tunicate *E. turbinate* is required (Proksch *et al*., 2003). This complicates the large-scale production of lead compounds for clinical development and is obviously environmentally detrimental and not a sustainable source of lead compounds.

It has been established that invertebrate-associated bacteria are possibly the true producers of NPs isolated from marine invertebrates. This is of significant importance because attempting to improve lead compound titres using large-scale fermentations of the producer micro-organism is a viable alternative to harvesting tonnes of invertebrate tissue. Unfortunately, only about 1% of bacteria are cultured *in vitro* and of the approximately 61 bacterial phyla, 31 lack cultivable representatives (Vartoukian *et al*., 2010). For example, a study investigating the variation of the microbial communities within the sponge *Aplysina aerophoba*, showed that only 0.15% of the bacteria were cultivated *in vitro* when compared to data obtained from microscopic analysis (Friedrich *et al*., 2001). This has led to the use of metagenomics to access the otherwise inaccessible genetic material of the uncultivable bacteria (Handelsman, 2004). However, although metagenomics is a powerful tool, the poor efficiency in obtaining intact fragments and inactivation of promoters in the heterologous hosts limit its effectiveness (Xiong *et al*., 2013). Therefore, isolation and characterisation of whole organisms is still an important tool for NP discovery especially with the advancement of sequencing tools as will be discussed.

Traditionally, bacterial isolations were based on physical characteristics such as colony morphology and colour. However, the isolation of bacteria based on physical characteristics easily contributes to redundancy (in the culture collection and NP library)

as organisms from the same genus might share morphological similarities but differ in their chemical profile (Ritacco *et al*., 2003). Alternatively, morphologically different strains might be responsible for the production of the same secondary metabolites (Macintyre *et al*., 2014). A high degree of redundancy in the culture collection and NP extract library is undesirable because it increases the probability of isolating known or similar compounds (Tulp and Bohlin, 2005).

In attempt to de-replicate compounds at an early stage of the discovery process, analytical tools such as nuclear magnetic resonance spectra (NMR), mass spectroscopy (MS), high performance liquid chromatography (HPLC) and ultra-violet spectra (UV) provide useful structural and chemical information. These tools are however not applicable to high-throughput processes and furthermore discriminate compounds based on physicochemical properties and not biological properties (Wong *et al*., 2012). This implies that compounds with a similar structure to known compounds but which exhibit novel biological activity might be disregarded.

1.7 Developments in the natural product discovery pipeline

Acknowledging the bottlenecks in NP discovery and development (section 1.6) is important in streamlining and maximising the output of the NP drug discovery process. This section discusses the strategies currently employed to overcome the various limitations associated with NP discovery. These have been divided into culture-independent and culture-dependent strategies.

1.7.1 Culture-independent methods

1.7.1.1 Whole-genome sequencing (WGS)

Past efforts in NP discovery from bacteria (particularly the genus *Streptomyces*) largely depended on chance, whereby isolates were grown under a range of growth conditions with an aim to stimulate secondary metabolite (SM) production (Ziemart *et al*., 2014; Jensen *et al*., 2014). In recent years, the advancement of high-throughput sequencing (HTS) technology (figure 1.12) has drastically increased the number of bacterial genomes available to about 3625 genomes by 2013 (Bertelli and Greub, 2013). Analysis of this data suggests that culture-based methods are not optimal for the activation of SM pathways and therefore NP discovery. For example, analysis of the complete *S. coelicolor* (A3) genome revealed the ability of this organism to produce an additional 20 compounds including coelichelin and coelimycin P1 (Aigle *et al*., 2014). Therefore, unlike bioactivity-based screening, WGS is superior as a NP discovery tool as it does not depend on the expression of SMs. Secondly, WGS analysis suggests that even previously well-studied organisms may still be a rich source of novel NPs (SMs). In addition, through WGS a complete snapshot of the SM pathways is available and allows for the selection of a particular class of SMs of interest to investigate. This is in contrast to chemical analysis where the discovery SMs is limited to those that are actively expressed.

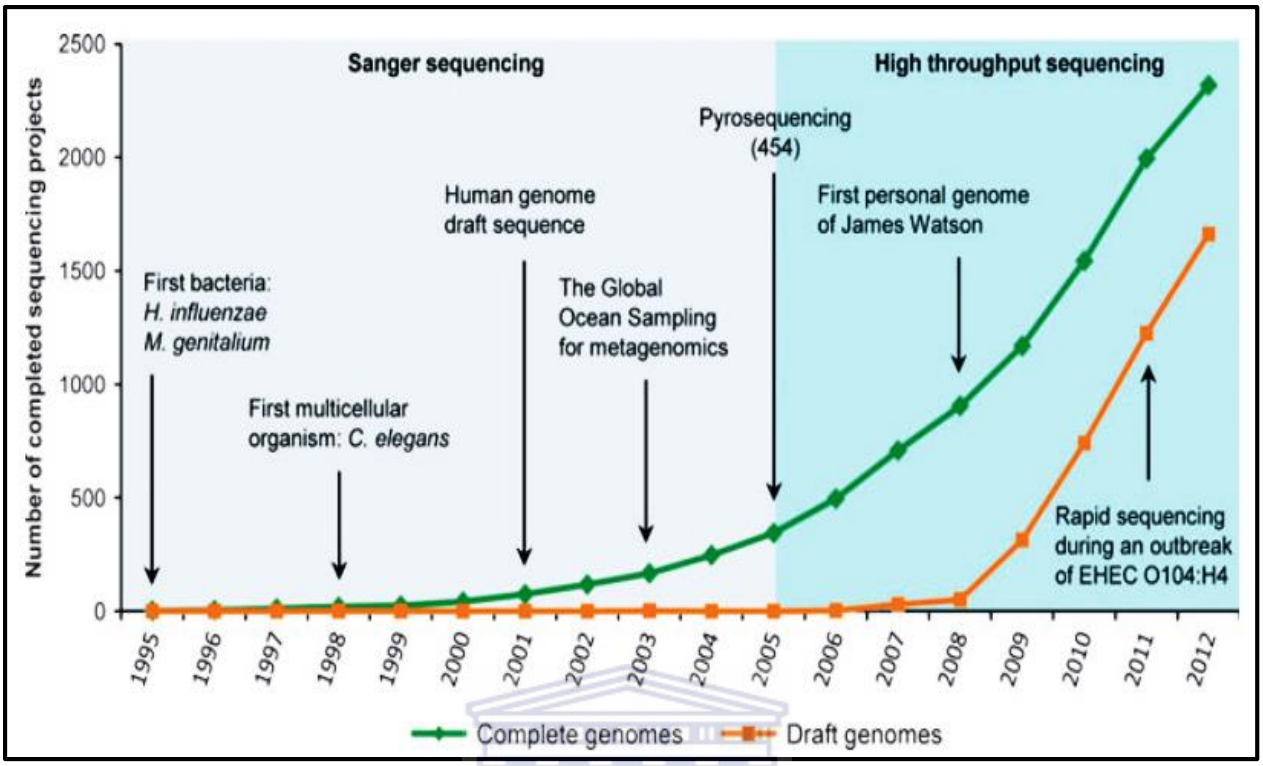

Figure 1.12: The advancement of DNA sequencing technologies between 1995 and 2012. Sanger sequencing techniques were prevalent prior to 2005 and were used to sequence the first bacterial genome (*H. influenza*) and the draft human genome. The development of high-throughput sequencing technologies such as pyrosequencing, however, resulted in the significant increase in genome sequence data from both draft and completed genomes (*Taken from* Bertelli and Greub, 2013).

The development of informatics-based tools, specifically designed for the detection and analysis of SM pathways from the vast sequence data generated by HTS greatly accelerates the NP discovery (table 1.2). For example, the 'Antibiotics and Secondary Metabolite Analysis Shell' (antiSMASH) software tool can rapidly search genome sequences for pathways belonging to a range of different SM classes including PKSs, NRPSs, bacteriocins, siderophores and terpenes (Blin *et al*., 2013). Specific for NRPS pathways, the NRPSpredictor2 tool is able to predict the amino acid specificity of NRPS

A-domains (Röttig *et al*., 2011). This is important when predicting possible monomer composition of a compound and comparison between a potentially novel pathway and known pathways. The NORINE database, dedicated exclusively to peptides of NRPS origin provides various computational analysis tools during NRPS characterisation (Caboche *et al*., 2008). For example, predicted amino acid monomers in an NRPS pathway of interest can be used to search for compounds of similar monomer composition. Additional bioinformatics tools for the computational analysis of SM pathways are summarised in table 1.2.

Table 1.2: Examples of computational tools for the analysis of secondary metabolite pathways at the genetic level

Tool	Function/Description	URL	Reference
antiSMASH2.0	Genome mining tool to detect secondary metabolite pathways in genome sequence data. Provides predicted structures for pathways detected.	http://antismash.secondarymetabolites.org/	(Blin *et al*., 2013)
NRPSpredictor2	Predict the substrate specificity of NRPS adenylation domains.	http://nrps.informatik.uni-tuebingen.de/Controller?cmd=SubmitJob	(Röttig *et al*., 2011)
NORINE	A database exclusive to natural products of non-ribosomal peptide origin. Facilitates structure based searches using predicted cognate amino acids recruited by the adenylation domain.	http://bioinfo.lifl.fr/norine/	(Caboche *et al*., 2008)
DoBiscuit	A manually curated database of secondary metabolites providing integrated information on biosynthetic pathways particularly polyketides.	http://www.bio.nite.go.jp/pks/top	(Ichikawa *et al*., 2013)
pep2path	A mass-spectroscopy based approach to match tandem MS data to the corresponding biosynthetic gene clusters	http://pep2path.sourceforge.net/	(Medema *et al*., 2014)

(exclusive to peptide natural products).

NP.searcher	A genome mining tool for the detection of NRPS, PKS, hybrid pathways as well as terpenoid genes.	http://dna.sherman.lsi.umich.edu/	(Li *et al*., 2009)
BAGEL	A web-based tool for the mining of genomes to detect ORFs associated with bacteriocin biosynthesis in microbial DNA sequences.	http://bagel2.molgenrug.nl/	(de Jong *et al*., 2006)

1.8 Culture-based approach

WGS data suggests that NPs frequently go undetected during conventional bioactivity-guided screening. Inadequate growth conditions that are unfavourable for the expression of SM pathways are considered a major contributing factor. For this reason, two major developments to traditional culture-based techniques are discussed here that aim at activating silent SM pathways and as a result may diversify NP libraries.

1.8.1 Co-cultivation

Co-cultivation involves the fermentation of two or more micro-organisms in the same vessel (mixed culture) mimicking the natural competition for resources (Marmann *et al*., 2014). The aim of co-cultivation is to trigger the expression of silent SM genes by creating a competitive environment similar to that of the natural habitat (Ola *et al*., 2013). Co-cultivation has successfully been used to activate a cryptic PKS pathway in *Aspergillus nidulans* following the mixed fermentation of this fungus with *Streptomyces rapamycinicus* (Nützmann *et al*., 2011). In addition, co-cultivation of the fungus *Emericella sp.* (strain CNL-878) and the marine isolate *Salinispora arenicola* (strain CNH-665), resulted in a drastic increase in production of a novel cyclic lipopeptide emericellamide A approximately 100-fold (Oh *et al*., 2007). Co-cultivation therefore presents as an ideal method to diversify NP libraries particularly due the low costs involved and its application at a large-scale.

1.8.2 One strain many compounds (OSMAC)

One Strain Many Compounds (OSMAC) is a culture-based approach that involves the manipulation of easily accessible culturing conditions such as temperature, type of culture vessel and media components (Bode *et al*., 2002). The aim of this approach is to induce the expression of silent SM pathways (Bills *et al*., 2008). A major advantage provided by OSMAC is that it eliminates bias during bioactivity-screening because the stimuli responsible for the activation of SM pathways differ between organisms. This approach therefore has the potential to further diversify NP libraries and improve the hit-rate.

1.9 An overview of marine biotechnology in South Africa

South Africa is bordered by a 3650 km linear coastline that extends from the south of Namibia in the west to the south of Mozambique in the east (Davies-Coleman and Beukes, 2004; Griffiths *et al*., 2010). Three main bio-geographical zones make up this coastline and include a warm temperate southeast coast, a cool temperate west coast and the subtropical east coast each harbouring unique biodiversity of endemic marine organisms (Davies-Coleman and Beukes, 2004). This provides the South African marine research community with rich and abundant sampling resources for NP discovery.

The Tsitsikamma reserve and Algoa bay (South Africa) are habitats to numerous marine ascidians. Compounds belonging to acyclic amino alcohols have been isolated from these invertebrates exhibiting anti-microbial properties (Davies-Coleman and Beukes,

2004). In addition, soft coral distribution along the South African coast is relatively abundant with approximately 200 species (60-70% endemic) known and are a major component of the benthic community (Hooper and Davies-Coleman, 1995). Soft corals such as *Capnella thyrsoidea* are an important source of potent anti-inflammatory xenicane diterpenes such as tsitsixenicin A (Hooper and Davies-Coleman, 1995).

In addition to the abundant fauna of tunicates and soft corals, molluscs and sponges make up a diverse group of marine fauna along the South African coast (Griffiths *et al*., 2008). Previously uncharacterised alkaloids, tsitsikammamme A and tsitsikammamine B (bispyrroloiminoquinone alkaloids) were isolated from a latrunculid sponge (Hooper *et al*., 1996). Two additional pyrroloiminoquinone alkaloids, 14-bromodiscorhabdin C and 14-bromodihydrodiscorhabdin C were identified in the same study. Various novel NPs have been isolated from marine molluscs (gastropods) and algae from the South African coast exhibiting anti-oesophageal cancer activity (van Wyk *et al*., 2008; Mann *et al*., 2007). South African marine sponges have also proved a rich source of bioactive compounds exhibiting activity against oesophageal cancer cell lines (Whibley *et al*., 2005). Collectively, these studies highlight the potential for the discovery of yet unknown NPs from the South African coast.

It is obvious that the South African NP research has paid minimal attention to micro-organisms as sources of NPs. This study therefore provides an opportunity to address this research gap between micro-organisms and NP discovery in South Africa.

1.10 Research objectives

This study forms part of a broader collaborative project (PharmaSea) with the principal aim of isolating and characterising a novel NRPS or PKS pathway from bacteria associated with novel marine invertebrate species endemic to the South African coast.

1.10.1 Culture-dependent approach

Specific objectives include;

(i) Isolating bacteria associated with novel marine sponge and tunicate species endemic to the South African coast.

(ii) Functional screening of the marine invertebrate-associated bacteria for anti-microbial activity against a panel of indicator organisms.

(iii) To select preliminary candidate strains for whole genome sequencing based on the anti-microbial profile.

1.10.2 Culture-independent approach

Specific objectives include;

(i) Homology screening (specifically for NRPS and PKS genes) of bioactive isolates using PCR.

(ii) Selection of a suitable candidate strain based on anti-microbial profiles and PCR screening results.

(iii) Genome mining of genome sequence data for potentially novel NRPS or PKS pathways.

(iv) Use of bioinformatics tools to characterise a potentially novel pathway.

Chapter 2 Materials and methods

2.1 General chemicals and enzymes

All chemicals used in this study were supplied by Merck Chemicals and laboratory supplies (Darmstadt, Germany), Sigma Aldrich Chemical Company (Deissenhofen, Germany) and Kimix Chemical and Laboratory Supplies (South Africa). Culture media were supplied by Oxoid Ltd and Biolabs.

Oligonucleotides for use in the polymerase chain reaction (PCR) were synthesized by Inqaba Biotech (Johannesburg, South Africa) and Integrated DNA Technologies (Coralville, Iowa, USA).

DNA modifying enzymes (polymerases & restriction endonucleases) were purchased from Fermentas Life Sciences Ltd (Vilnius, Lithuania).

2.2 Bacterial strains used in this study

Table 2.1: Description of bacterial strains used in this study

Bacterial strains	Genotype	Supplier
E. coli EPI300	F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ- rpsL (StrR) nupG trfA tonA	Epicentre® an Illumina company (USA)

2.3 Plasmids used in this study

Table 2.2: Description of plasmid used in this study

Plasmid	Description/Relevant genotype	Supplier
pUC18	A high copy number cloning vector Contains the *bla* gene coding for a β-lactamase that confers ampicillin resistance. *LacZ* gene present coding for the N-terminal of β-galactosidase enabling (α)-complementation with the defective β-galactosidase Δ (*LacZ*) within the host facilitating blue/white clone selection	Thermo scientific

2.4 Specimen collection and taxonomic identification

2.4.1 Collection

Sponge and tunicate specimens were collected using SCUBA by Dr Shirley Parker-Nance (Just Blue, South Africa) from low profile rocky reefs in Algoa bay White sands reef (South Africa) at depths of 23-25 metres (figure 2.1). Sampling was carried out on the 23rd January and the 4th of April 2013 respectively. The sponge specimen was collected at the geographical coordinates longitude 34˚00.366S and latitude 25˚43.209E whereas the tunicate specimen was collected at longitutde 34000.406S and latitude 25˚43.117E.

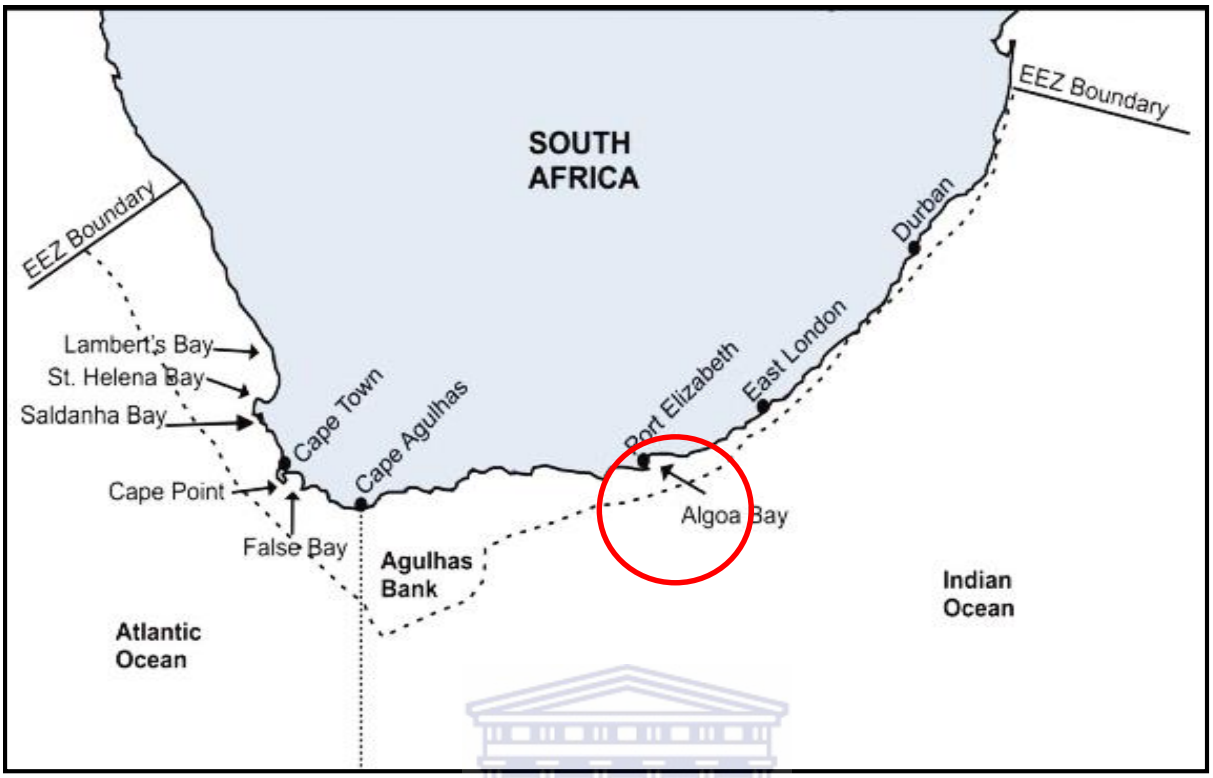

Figure 2.1: Geographical location of sampling site. Both specimens were collected from Algoa bay (demarcated by a red circle), White sands reef at depths ranging between 23-25 metres. The habitats consisted of low profile rocky reefs (*Adapted from* Griffiths *et al*., 2010).

2.4.2 Taxonomy

Taxonomic identification of the invertebrate specimens was undertaken by Dr. Shirley Parker-Nance (Just Blue, South Africa) and the results are summarised in figure 2.2. The specimens are novel species in their respective genera and are classified as *Spongia (Spongia)* sp. 001RSASPN and *Pseudodistoma africanum* Millar, 1954 (figure 2.2).

Sample ID:	PE8
Phylum:	Porifera
Class:	Demospongiae
Order:	Dictyoceratida
Family:	Spongidae
Genus:	Spongia
Subgenus:	Spongia (Spongia) Linnaeus, 1759
Species:	*Spongia (Spongia)* sp. 001RSASPN

Sample ID:	PE13
Phylum:	Chordata
Subphylum:	Urochordata-Tunicata
Class:	Ascidiacea
Order:	Aplousobranchia
Family:	Pseudodistomidae
Genus:	Pseudodistoma
Species:	*Pseudodistoma africanum* Millar, 1954

Figure 2.2: Taxonomic classification of marine invertebrate specimens used in this study.

2.5 Isolation of marine invertebrate-associated bacteria

Sponge and tunicate tissue was processed within 24 hours of collection by mechanically homogenising 1 gram of tissue (including both internal and external tissue) using a sterile mortar and pestle. From the resulting mixture, 1 ml was transferred to 9 ml of sterile sea-water and a dilution series (10^{-1}-10^{-6}) performed. From each dilution, 100 µl was plated onto different media representing different nutrient sources (Appendix B) and incubated at 15˚C for a period of 6-8 weeks. Growth was monitored regularly and colonies picked based on physical characteristics including colour, colony size and morphology. Strains were subsequently re-streaked onto the medium from which they were isolated.

2.6 Functional screening for anti-bacterial activity

The agar-overlay assay was used to screen for anti-bacterial activity of invertebrate-associated bacteria against a panel of indicator organisms consisting of; *Escherichia coli* 1699 (a multi-drug resistant strain, Cubist), *Bacillus cereus* ATCC 10702, *Staphylococcus epidermidis* ATCC 14990 and *Mycobacterium smegmatis* LR222. Briefly, isolates to be screened were grown on their respective isolation medium for 14 days at 15˚C. Prior to anti-bacterial screening, the indicator organisms were grown up in 10 ml of LB broth overnight at 37°C shaking at 250 rpm. To ensure an equal amount of cells was used for each experiment, the following formulae were used; OD_{600} x X µl = 4 (*E. coli 1699)* and OD_{600} x X µl = 160 (*B. cereus*, *S. epidermis* and *M. smegmatis)* where OD_{600} refers to optical density measured at 600 nm. The appropriate volume of cell culture was added to 6 ml of sloppy agar and gently distributed over the colonies to

be screened and incubated at 37˚C overnight. Anti-bacterial activity was confirmed by a zone of clearance surrounding the colony.

2.6.1 Effect of a change in culture medium on anti-microbial activity

Selected bioactive isolates were further investigated for anti-bacterial activity on glucose yeast malt medium (GYM). Briefly, isolates were grown for 14 days at 15°C on GYM agar and the agar-overlay assay performed as described in section 2.6.

In a collaborative effort with the Medina Foundation (Granada, Spain), the screening platform was extended to include clinical strains including *Escherichia coli* MB2884, *Staphylococcus aureus* MRSA MB5393, *Pseudomonas aeruginosa* PA01, and *Pseudomonas aeruginosa* PA01 synergy with IMP. In addition, *Aspergillus fumigatus* ATCC 46645 and *Candida albicans* MY1055 were included as eukaryotic indicator strains. All strains isolated from the specimens described in section 2.4 were grown in four different liquid media including GYM, Zobell, Tryptic soy and activated charcoal medium (ACM). Extracts (as described in section 2.6.2.5) were subsequently screened against the panel of indicator strains listed above.

2.6.2 Extraction of anti-microbial compounds

2.6.2.1 Well-diffusion assay

Prior to the preparation of crude extracts, the well-diffusion assay was used to detect anti-microbial compounds in the cell-free culture broth. Briefly, isolates were grown in 25 ml of medium for 14 days at 15°C while shaking at 250 rpm in a 100 ml conical flask. Cells were separated from the broth by centrifugation at 7000 g for 5 minutes at 15°C. The broth (20 ml) was concentrated at 46°C to approximately 10 ml. Test strains were grown overnight at 37°C in LB broth and adjusted to an OD_{600} between 0.3-0.4 and 100 µl of culture spread onto solid LB agar containing wells created using the wide end of a yellow pipette tip. The concentrated broth (50 µl) was subsequently loaded into the wells and allowed to stand at room temperature for 30-60 minutes and then incubated overnight at 37°C. MilliQ water was used as a negative control for anti-microbial activity.

Once anti-microbial activity was detected in the broth, an equal volume of acetone was added to the culture broth and allowed shake for 1 hour at room temperature. The mixture was subsequently concentrated at 46°C to half its original volume.

2.6.2.2 Column preparation

Columns were prepared using 15 ml falcon tubes. Briefly, 3-4 holes were created at the bottom of the 15 ml tube and plugged with glass-wool (approximately 1 cm) as illustrated in figure 2.3. The resin (2 ml) was allowed to pack on top of the glass-wool.

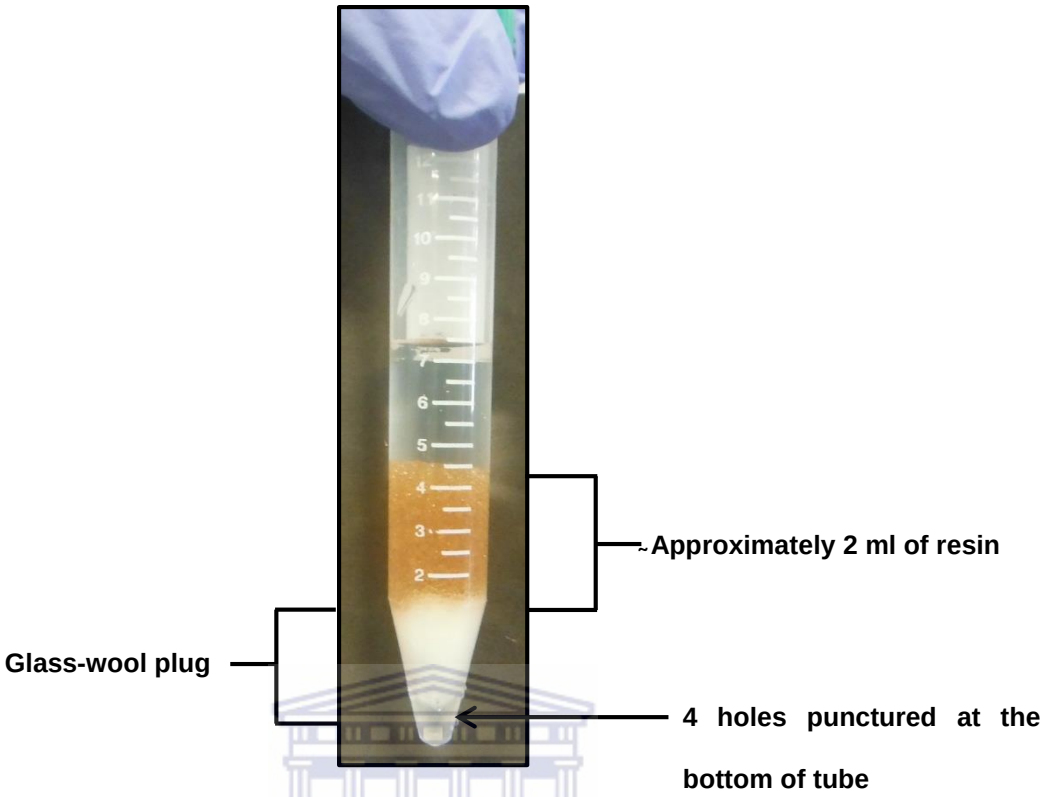

Figure 2.3: Assembly of the ion-exchange columns used in this study.

2.6.2.3 Anionic-exchange resin preparation

Amberlite IR120 (Na^+) resin columns were prepared as described above (figure 2.3). The resin was washed once with 10 ml of MilliQ water and allowed to flow through completely then with 15% HCl followed by an additional two washes with MilliQ water each at a volume of 10 ml. The concentrated acetone extract (section 2.6.2.1) was applied to the resin and allowed to drip through the column by gravity flow, followed by a single wash with 5 ml of MilliQ water. Bound material was eluted and collected with 5 ml of acetone and subsequently dried at 46°C (completely) in the absence or presence of vacuum. The dried extract was re-suspended in 500 µl of MilliQ water. Extracts were

tested for anti-microbial activity using the well-diffusion assay as described in section 2.6.2.1.

2.6.2.4 Cationic-exchange resin preparation

Amberlite IR 910 (Cl^-) resin was packed into the column and allowed to settle above the glass-wool plug at a volume of 2 ml (figure 2.3). Beads were washed with 10 ml of MilliQ water and allowed to flow through completely. The beads were washed with 0.1 M NaOH then two additional washes (10 ml) with MilliQ water. The concentrated acetone extract (section 2.6.2.1) was then allowed to flow through the resin followed by a wash with 5 ml of MilliQ water. Bound material was eluted and collected with 5 ml of acetone and subsequently dried at 46°C (completely) in the absence or presence of vacuum. The dried extract was re-suspended in 500 µl of MilliQ water. Extracts were tested for anti-microbial activity using the well-diffusion assay as described in section 2.6.2.1.

2.6.2.5 Hydrophobic-resin column preparation

The column was prepared as described above (section 2.6.2.2) and packed with 2 ml of Sepabeads® (SP-207). The resin was washed once with 5 ml of acetone followed by a single wash with 5 ml of methanol. Once the methanol had run through the resin, the column was washed once with 10 ml of MilliQ water. To the cell culture (10 ml), an equal volume of acetone was added and shaken for 1 hour at room temperature. This mixture was concentrated to half the volume at 46°C and allowed to run through the

resin by gravity flow. Once the culture had completely run through the column, the resin was washed once with 5 ml MilliQ water. Material bound to the column was eluted using 5 ml of acetone and collected. The eluent was dried to completion at 46°C and re-suspended in 200 µl of MilliQ water.

2.7 Construction of growth curves

A single colony was inoculated into 25 ml of the desired liquid medium in a conical flask (150 ml volume) and grown overnight at 15°C whilst shaking at 250 rpm. The overnight culture was adjusted to an OD_{600} of 0.2 and 1 ml transferred to 100 ml of sterile medium in a 250 ml conical and grown at room temperature while shaking at 150 rpm. Readings (OD_{600}) were taken at appropriate intervals until no further change in the OD_{600} was observed. The OD_{600} readings were subsequently plotted against time to obtain a growth curve.

2.8 DNA extraction

2.8.1 Small-scale genomic DNA extraction

A single colony was inoculated into sterile medium (10-20 ml) and allowed to grow at room temperature until turbid. Cell mass was harvested by centrifugation at 7,000 g for 5 minutes and the supernatant discarded. The resulting pellet was re-suspended in 500 µl of lysis buffer and incubated at 37˚C overnight. SDS at a final concentration of 1% [v/v] and 1 µl of RNAse A (20 mg/ml) were added to the mixture, gently inverted and

incubated at 65˚C for 30 minutes. An equal volume of phenol/chloroform/isoamyl alcohol (PCI) in the ratio of 25:24:1 was added, thoroughly mixed and centrifuged at 10,000 g for 10 minutes. The aqueous phase was transferred to a new tube and an equal volume of chloroform/isoamyl alcohol (CI) in a ratio of 24:1 added. The tube was gently inverted (3-6 times) and centrifuged at 10,000 g for 10 minutes and the aqueous phase (top layer) was transferred to a new micro-centrifuge tube. DNA was precipitated using one volume of isopropanol and gently inverted to mix thoroughly and centrifuged at 10,000 g for 10 minutes. Isopropanol was discarded and the pellet washed with 1 ml of 70% ethanol and subsequently air-dried. DNA was re-suspended in 50-100 µl of DNAse-free sterile water.

2.8.2 Large-scale genomic DNA extraction

A single colony was inoculated into 100-400 ml of culture medium at room temperature until turbid following which cells were harvested at 10,000 g for 5 minutes. The resulting pellet was re-suspended in 9.5 ml of TE buffer (pH 8) to which 0.5 ml of 10% [v/v] SDS, 100 µl of 20 mg/ml proteinase K and 1U of 10 mg/ml RNAse A was added, thoroughly mixed and incubated at 37˚C for 1 hour. To this mixture, 1.8 ml of 5 M NaCl and 1.5 ml of a 10% [w/v] CTAB/0.7 M NaCl solution was added, thoroughly mixed by inversion and incubated at 65˚C for 20 minutes. An equal volume of CI (24:1) was added and gently but thoroughly mixed followed by centrifugation at 10,000 g for 10 minutes at 4˚C. The aqueous layer was transferred to a new micro-centrifuge tube to which an equal volume of PCI (25:24:1) was added and gently mixed. Following phase separation at 10,000 g for 10 minutes, DNA was precipitated with a 0.6 volume of isopropanol. The

DNA was collected either by centrifugation or spooling and subsequently washed once with 1 ml of 70% ethanol for 5 minutes at 10,000 g. The ethanol was discarded following which the air-dried DNA pellet was re-suspended in 200-500 µl of TE buffer.

2.8.3 Hard lysis DNA extraction

DNA from *Streptomyces* isolates was extracted using a hard lysis method. Briefly, the cell mass was harvested by centrifugation at 10,000 g for 5 minutes. The supernatant was discarded and cells washed twice with TE buffer (pH 8) by centrifugation at 10,000 g for 5 minutes at 4°C. The supernatant was discarded and the cell pellet re-suspended in 100 µl of breaking buffer following which 0.3-0.4 g of acid-washed glass beads (212-300 µm, 50-70 U.S sieve) were added and the mixture briefly vortexed at maximum speed. To this mixture 100 µl of PCI in the ratio 25:24:1 was added and bead-beating performed for 3 minutes. TE buffer (pH 8) was added at a volume of 200 µl gently mixed and centrifuged at 10,000 g for 5 minutes at 4°C. The resulting supernatant was transferred to a new micro-centrifuge tube to which 1 ml of ice-cold 100% ethanol was added and kept on ice for 10 minutes. Centrifugation was performed as above following which the DNA pellet was re-suspended in 1 ml of ice-cold 70% ethanol and 10 µl of 4 M ammonium acetate. This was kept on ice for 10 minutes and centrifuged as above. The resulting DNA pellet was air-dried and re-suspended in 50 µl of TE buffer (pH 8).

2.8.4 Plasmid extraction using the alkaline lysis method

A single colony was inoculated into 10 ml of LB supplemented with ampicillin at a final concentration of 100 µg/ml and grown overnight at 37˚C shaking at 250 rpm. From the overnight culture, 2 ml was transferred to a sterile micro-centrifuge tube and cells harvested at 16,000 g for 30 seconds. The resulting supernatant was discarded and the cell pellet re-suspended in 100 µl of solution 1 (Appendix A) and placed on ice for 5 minutes. Solution 2 (Appendix A) at a volume of 200 µl, was added and mixed thoroughly by inversion and chilled on ice for 5 minutes. This was followed by the addition of 150 µl of solution 3 (Appendix A) and the mixture briefly vortexed and kept on ice for 5 minutes. Cell debris and chromosomal DNA was subsequently pelleted for 3 minutes at 16,000 g at room temperature. The supernatant was transferred to a sterile micro-centrifuge tube and the plasmid DNA precipitated using 800 µl of 95% ethanol for 2 minutes at room temperature. Plasmid DNA was collected by centrifugation at 16,000 g for 1 minute at room temperature and washed with 1ml of 70% ethanol. The plasmid DNA was air-dried and re-suspended in 50 µl of TE buffer (pH 8).

2.8.5 Extraction of sequencing grade plasmid DNA

To ensure good quality plasmid DNA suitable for sequencing, plasmid extractions were performed using the Qiaprep® miniprep kit (Qiagen). Briefly, a single colony was inoculated into 10 ml of LB supplemented with ampicillin at a final concentration of 100 µg/ml. Cultures were grown overnight at 37˚C while shaking at 250 rpm. Cells were harvested by centrifugation at 16,000 g for 30 seconds. Plasmid DNA was extracted according to the manufacturers' instructions.

2.8.6 Gel extraction and PCR cleanup

DNA and PCR amplicons resolved in agarose gel were purified using the NucleoSpin® gel and PCR cleanup kit (Macherey-Nagel) according to the manufacturers' instructions.

2.9 Total RNA extraction

A single colony was inoculated into 10 ml of medium and grown overnight at room temperature while shaking at 150 rpm. The OD_{600} of the overnight culture was adjusted between 0.1-0.3 and 1 ml transferred to a conical flask (250 ml volume) containing 50 ml of medium. This culture was grown at room temperature (shaking at 150 rpm) to the appropriate growth stage (mid-logarithmic or stationary) and cells immediately harvested by centrifugation at 7000 g at room temperature for 5 minutes. The supernatant was discarded and cells immediately re-suspended in 1 ml TRIzol reagent. Total RNA was subsequently extracted as per the manufacturers' instructions.

2.10 Nucleic acid analysis

Nucleic acid purity (based on 260/280 and 260/230 ratios) and the concentration of nucleic acid solutions were determined by spectrophotometric analysis using the NanoDrop® ND-1000 (NanoDrop technologies, Inc., USA).

2.10.1 DNA agarose gel electrophoresis

DNA was resolved using agarose gel electrophoresis. Agarose gels were prepared in 1X TAE buffer at a concentration of 0.8-1% [w/v] of agarose. To ensure DNA visualisation, the agarose gel was stained with ethidium bromide staining solution at a final concentration of 0.5 µg/ml. Prior to loading onto the agarose gel, DNA samples were prepared by the addition of 6X DNA loading buffer. Electrophoresis was performed in 1X TAE running buffer at 10-12 Volts/cm for 1 hour. DNA size was estimated using lambda DNA completely digested with *Pst* endonuclease as molecular weight marker. DNA was visualised under ultra violet (UV) light (302 nm) and images obtained using the digital imaging system Alpha Imager® HP 2000 (Alpha Innotech, USA). DNA to be excised and subsequently purified for further analysis, was visualised using long wavelength UV (365 nm) for no more than 10 seconds.

2.10.2 RNA agarose gel electrophoresis

RNA was resolved in an agarose gel prepared as described by Aranda and colleagues (2012). Briefly, to the 1% [w/v] agarose and 1X TAE buffer mixture, bleach at a final concentration of 0.5% [v/v] was added and incubated at room temperature for 5 minutes. The agarose mixture was subsequently heated and once cooled; ethidium bromide staining solution (final concentration 0.5 µg/ml) was added to the gel and allowed to set. Electrophoresis was performed in 1X TAE running buffer at 10-12 Volts/cm for 1 hour. Size estimation was performed using lambda DNA completely digested with *Pst* endonuclease as molecular weight marker. RNA was visualised under

UV light (302 nm) and images obtained using the digital imaging system Alpha Imager® HP 2000 (Alpha Innotech, USA).

2.11 Cloning of PCR amplicons

2.11.1 Preparation of electrocompetent cells

A single colony of *E. coli* EPI300 was inoculated into 10 ml LB and grown overnight at 37˚C while shaking at 250 rpm. Electrocompetent *E. coli* EPI300 cells were prepared by inoculating 400 ml of LB with 4 ml of the overnight culture and grown at 37˚C while shaking at 250 rpm. Cells were grown to an OD_{600} between 0.35-0.4 and immediately kept on ice for 40 minutes while occasionally swirling. All subsequent steps were performed in a cold room (4˚C) or on ice. Once cooled, cells were harvested at 7,000 g for 20 minutes at 4˚C and the resulting supernatant discarded. The cell pellet was re-suspended in 200 ml of ice-cold sterile demineralised water and harvested at 4˚C for 10 minutes at 7,000 g. The supernatant was discarded and cell pellet re-suspended in 40 ml ice-cold sterile distilled water and harvested at 7000 g for 5 minutes at 4°C (repeated 5-6 times). Once thoroughly washed, the cell pellet was re-suspended in 40 ml of ice-cold sterile 10% glycerol and harvested at 10,000 g for 20 minutes at 4˚C. The supernatant was discarded and cell pellet re-suspended in 1 ml of ice-cold 10% glycerol. Aliquots (70 µl) were transferred to chilled sterile 1.5 ml micro-centrifuge tubes on ice and electrocompetent cells snap frozen in liquid nitrogen and stored at - 80˚ C until further use.

2.11.2 Preparation of pUC18 cloning vector

The cloning vector (pUC18) was prepared for subsequent ligation reactions by digestion with 1U of fast-digest *Sma*I *e*ndonuclease per µg of vector DNA in a 20 µl reaction as per manufacturers' recommendations. The digested vector DNA was resolved in agarose gel (0.8% [w/v]) and the linearized vector subsequently purified as described in section 2.8.6.

2.11.3 Blunt-end DNA ligation

Linearized pUC18 vector was ligated with insert DNA (blunt-end ligation) in the ratios 1:5 or 1:3 (vector to insert). The reaction mixture consisted of 1U of T4 DNA ligase, 1X T4 DNA ligase buffer and nuclease free water to a final volume of 20 µl. The ligation reaction was performed overnight at 16°C. Blunt-ended inserts were created where necessary using T4 DNA polymerase as per manufacturers' protocol.

2.11.4 Transformation by electroporation

Electrocompetent *E. coli* EPI300 cells were transformed directly with the ligation mixture. Briefly, an aliquot of electrocompetent cells (70 µl) was thawed for approximately 1 minute on ice and 5 µl of the ligation mix added to the thawed cells. This mixture was kept on ice for 30-60 seconds and subsequently transferred to a chilled 0.1 cm sterile electroporation cuvette (Bio-Rad). Cells were electroporated using the Bio-Rad MicroPulser™ (USA) with the following parameters; 1 pulse of 1.8 Kv, 25 µF and 200 Ohms (Ω). Immediately after electroporation, 1 ml of LB was added to the

cells and the mixture transferred to a sterile 2 ml micro-centrifuge tube and incubated at 37°C for 1 hour. Electroporated cells (100 µl) were spread onto LB plates containing ampicillin at a final concentration of 100 µg/ml of and 80 µg/ml X-gal and incubated overnight at 37°C.

2.11.4.1 Screening of clones

Based on the principle of blue/white selection, white colonies were picked and plasmid DNA extracted as described in section 2.8.4. Using the M13 vector primers, PCR was used to screen for clones containing the expected insert size (table 2.3).

2.12 Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) was carried out in 25-50 µl reactions. Each reaction consisted of 0.25U of a thermo-stable DNA polymerase (Dream*Taq* DNA polymerase), 1X reaction buffer, 0.2 mM dNTP mix (dTTP, dGTP, dCTP and dATP) and forward/reverse primers at a final concentration of 1 µM. Nuclease free water was used to adjust the reaction volume. All reactions were carried out in an automated thermal cycler (T100 thermal cycler, Bio-Rad). Cycling parameters and primer sequences are listed in table 2.3.

2.12.1 Reverse transcription PCR

Following total RNA extraction, complementary DNA (cDNA) was synthesised using the iScript cDNA synthesis kit (Bio-Rad) according to the manufacturers' specifications from 500 ng of RNA (quantified by spectrophotometric analysis). Prior to cDNA synthesis, RNA samples were DNAse treated using RNAse-free DNAse I according to the manufactures' instructions. For downstream PCR analysis, the volume of the cDNA did not exceed 1/10th of the reaction volume as recommended by the supplier. PCR reactions were set-up as described in section 2.12. PCR cycling parameters are listed in table 2.3.

Table 2.3: Primers and PCR cycling parameters

Primer	Target sequence	Sequence (5' - 3')	Product size (b	Thermal cycling parametres	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 30 sec, 72°C for 1 min and final extension step at 72°C for 10 min.	(Vizcaino *et al* ., 2005)
E9F U1510R	16S rRNA gene	GAGTTTGATCCTGGCTCAG GGTTACCTTGTTGTTACACTT	1500	Initial denaturation at 94°C for 5 mins, 30 cycles of 94°C for 30 sec, 55°C for 30 sec, 72°C for 1.30 min and final extension at 72°C for 10 min.	(Farelly *et al* ., 1995) (Reysenbach *et al* ., 1995).
FAS-F FAS-R	ORF 48 (PE8-15-cluster 4)	TCGGTTCCACGAAGGATGAG GCTTTCGTTCCAGTACCGTG	856	Initial denaturation at 95° C for 3 min, 30 cycles of 95°C for 30 sec, 60°C for 30 sec, 72°C for 1 min and final extension at 72°C for 5 min	This study
NRPS1-F NRPS1-R	ORF 39 (PE8-15-cluster 4)	GGAGGAGTTTACGTTCCGCT CCGTCAGGGGAATAACGGAC	950	Initital denaturation at 95°C for 5 min, 30 cycles of 95°C for 30 sec, 61°C for 30 sec, 72°C for 1 min and final extension at 72°C for 5 min.	This study
KS1F1 KS1R1	Type I PKS (KS-domain) (*Streptomyces*)	ATGGAYCCSCARCARCGBCT GCYTCGATSGGRTCNCCSA	700	**Touchdown PCR**: Initial denaturation at 95°C for 3 min, 10 cycles of 95°C for 30 sec, 77.3°C (-1°C/cycle) for 30 sec, 72°C for 60 sec. 20 cycles of 95°C for 30 sec 62.3°C for 30 sec, 72°C for 50 sec and 72°C for 5 min.	(Hill, unpublished)
KS2aF1 KS2bR1	Type II PKS (KS-domain) *(Streptomyces)*	TSGCSTGYTTCGAYGCSAT GCRTAGAACCASGCGAWSGA	1070	**Touchdown PCR**: Initial denaturation at 95°C for 3 min, 10 cycles of 95°C for 30 sec, 78°C for 30 sec, 72°C for 80 sec. 20 cycles of 95°C for 30 sec, 63°C for 30 sec, 72°C for 80 sec and a final extension at 72°C for 5 min.	(Hill, unpublished)
M13 F M13 R	Region flanking the pUC18 MCS	CCCACTCACGACGTTGTAAAACG AGCGGATAACAATTTCACACAGG		**Touchdown PCR**: Initial denaturation at 94°C for 2 mins, 10 cycles of 94°C for 30 sec, 65°C (-1°C/cycle) for 30 sec and 72°C for 50 sec. 25 cycles of 94°C for 30 sec, 55°C for 30 sec and 72°C for 50 sec.	Norrander *et al* ., 1983

2.13 Sequence and phylogenetic analysis

Cloned inserts and PCR amplicons were sequenced at the University of Stellenbosch sequencing facility (ABI PRISM 377 automated DNA sequencer) in both the forward and reverse direction. Chromatograms were manually edited using the Chromas Lite version 2.1 software package (Technelysium Pty Ltd, 2012). Curated nucleotide sequences (forward and reverse) were assembled using the DNAMAN version 4.13 software package (Lynnon Biosoft) to construct a consensus sequence.

Phylogenetic analysis of assembled 16S rRNA sequences was performed against the 16S ribosomal RNA sequences (Bacteria and Archaea) database (NCBI) using the BLAST tool (Altschul *et al*., 1990). All other sequence identity analysis was performed using the BLASTx tool against the non-redundant protein database (NCBI). Sequence alignments were done using MEGA version 6 (Tamura *et al*., 2013).

2.14 Whole genome sequencing

Total genomic DNA was extracted (section 2.8.2) and DNA submitted to the University of the Western Cape sequencing facility. The sample (1 ng) was prepared using the nextera XT kit v2 (500 cycles) and sequencing reactions performed on the Illumina MiSeq platform.

2.14.1 Genome sequence assembly

Genome sequence data was curated and analysed using the CLC genomics work bench version 6.5 software package. Sequence reads were merged and matched to the PHIX174 genome used as positive control for the sequencing run. Prior to the *de novo* assembly, reads were subjected to quality control and adapter trimming using the following parameters; quality score_ 0.05, maximum _ 3 ambiguous bases, minimum sequence length of 50 nucleotides and removal of adapter sequences (Appendix D) Contiguous sequences (contigs) were constructed by *de novo* assembly using the following parameters; mismatch cost _ 2, insertion cost _ 3, deletion cost _ 3, length fraction _ 0.95, similarity fraction _ 0.95 and a minimum contig length of 1000bp.

2.14.2 Genome mining for secondary metabolite pathways

Detection and annotation of SM biosynthetic pathways in genome sequence data (contigs) was performed with antiSMASH2.0 (Blin *et al*., 2013). The antiSMASH analysis parameters are shown in figure 2.4.

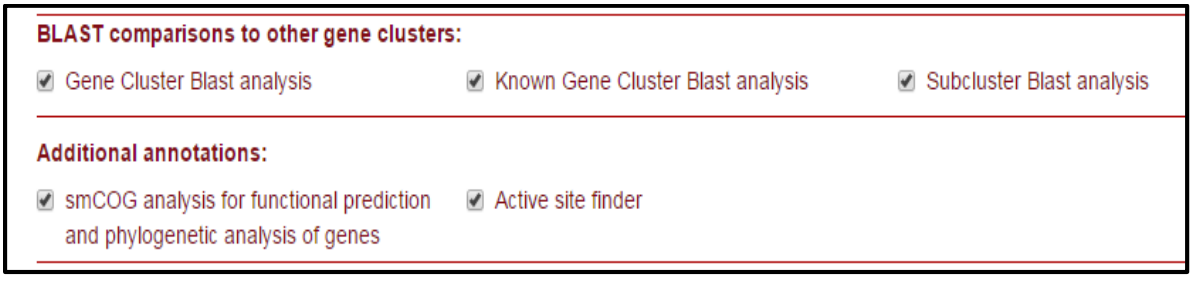

Figure 2.4: antiSMASH2.0 genome analysis parameters used in this study.

2.14.3 Computational characterisation of secondary metabolite pathways

Further bioinformatics analysis of SM pathways was performed using; NRPSpredictor2 (Röttig *et al*., 2011), Natural product domain seeker (NaPDos) *(*Ziemert *et al*., 2012), and MEGA 6 version 6.0 (Tamura *et al*., 2013).

2.14.4 Primer Design

Specific primers were designed using the NCBI primer-BLAST tool (Ye *et al*., 2012). Default settings were used for primer design with the exception of the product length (adjusted to minimum 800bp and maximum 1000bp).

Chapter 3 Results and discussion

Whole genome sequencing as a tool for the discovery of novel secondary metabolite pathways from marine invertebrate-associated bacteria

3.1 Introduction

Marine invertebrates are an important source of novel NPs from which various pharmaceutically relevant compounds have been isolated (Gunasekera *et al*., 1990; Bergmann and Feeney, 1951). NPs isolated from marine invertebrates display high similarity those of microbial origin (Velho-Pereira and Furtado, 2012). As a result, it was hypothesised that micro-organisms are the true producers of NPs otherwise attributed to the invertebrates (Haygood *et al*., 1999). Evidence for this is presented by Wilson and colleagues (2014) discussed in section 1.5.3. The significance of these findings is evident from an environmental perspective as bioactive compounds frequently occur at such low concentrations within the invertebrate tissue that an unrealistic amount of tissue is required (Mendola, 2003). All the economic feasibility of aqua-culture in mass production of marine NPs has previously been investigated (Sipkema *et al*., 2005), this approach is hindered by an extensive range factors (Sipkema *et al*., 2005; Mendola, 2003). In contrast, micro-organisms are a sustainable source of compounds because large-scale microbial fermentation is relatively more practical in comparison to aqua-farming (Sabdono, 2008). Furthermore, improved compound titres can be achieved through heterologous expression of SM pathways in well characterised microbial hosts (Ongley *et al*., 2013).

The re-discovery of known microbial NPs is a major bottleneck in NP discovery and has sparked interest in novel or previously uncultivated micro-organisms as a source of novel compounds (Williams, 2008). With the advancement of HTS, microbial genome data currently available has increased substantially (Bertelli and Greub, 2013). Analysis of these data shows that micro-organisms with experimentally characterised SM profiles contain several pathways with yet unknown products. This implies that the 1% of cultivated microbial diversity, thought to contain limited NP novelty still holds the potential to yield unknown compounds.

In this chapter, the isolation of bacteria associated with two novel marine invertebrate species harvested from Algoa Bay (South Africa) is described. Invertebrate-associated bacteria are subsequently screened for anti-microbial activity with the incorporation of the OSMAC strategy described previously (section 1.8.1.2). Bioactive isolates are further screened by PCR for NRPS and PKS genes. In addition, the informatics based characterisation of a potentially novel NRPS pathway (identified using WGS) is described.

3.2 Anti-bacterial activity of marine invertebrate-associated bacteria

The sponge and tunicate specimens, *Spongia* sp. 001RSASPN and *P. africanum* Millar, 1954 respectively, are new species in their respective genera. To our knowledge, these species have previously not been studied in the context of NP discovery. In an attempt to diversify the bacteria isolated, several growth media representing a range of nutritional components were used (Appendix B).

Using the agar-overlay assay (section 2.6), anti-bacterial activity was detected by a zone of clearance around a colony (figure 3.1). The inclusion of a multi-drug resistant indicator strain (*E. coli* 1699) is important for the detection of bioactive compounds with mechanisms of action different to those of known compounds (Hentschel *et al*., 2001). Therefore, strains exhibiting activity against *E. coli* 1699 were of particular interest in this study. A complete list of antibiotics to which *E. coli* 1699 exhibits resistance are listed in table F.1 (Appendix F).

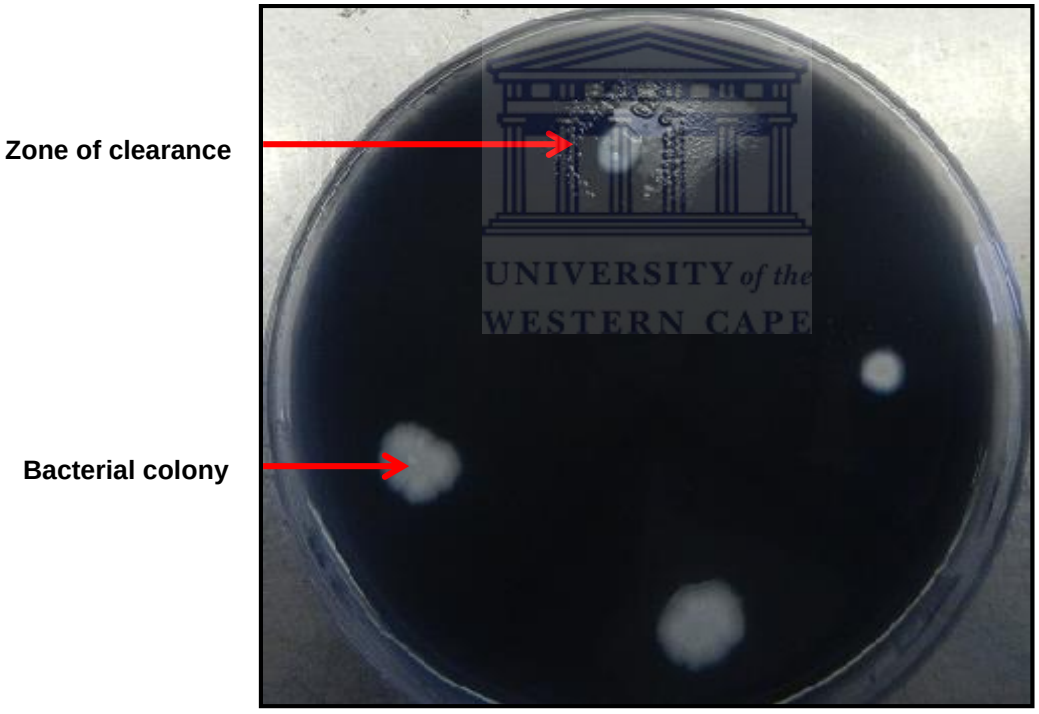

Figure 3.1: Detection of anti-bacterial activity in marine invertebrate-associated bacteria using the agar-overlay assay. Semi-solid agar (sloppy agar) is inoculated with the indicator strains and subsequently spread in a thin layer over the isolates. A zone of clearance surrounding the bacterial colony is indicative of the production of an anti-bacterial agent.

3.2.1 Isolation and anti-bacterial activity of sponge isolates

Based on colony morphology, a total of 212 bacterial strains associated with *Spongia* sp. 001RSASPN were isolated. Of these isolates, 131 were screened for anti-microbial activity and 6.1% (8 isolates) exhibited inhibitory activity against at least one indicator strain (table 3.1). Inhibitory activity was frequently observed against *B. cereus* whereas *M. smegmatis* was less frequently inhibited. A total of 3 isolates were active against the multi-drug resistant *E. coli* 1699 strain.

3.2.2 Isolation and anti-bacterial activity of tunicate isolates

From the tunicate *P. africanum* Millar 1954, a total of 184 bacterial strains were isolated of which 148 were screened for anti-bacterial activity. Anti-bacterial activity was exhibited by 6 isolates (4%) against at least one indicator strain (table 3.1). Inhibitory activity was observed against *B. cereus*, *S. epidermidis* and *E. coli* 1699 whereas *M. smegmatis* was not inhibited by any of the isolates.

Table 3.1: Anti-bacterial activity exhibited by marine invertebrate-associated bacteria

Strain ID	Source	*M. s*	*S. e*	*E. c*	*B. c*
PE8-15	Sponge	-	+	-	+
PE8-16	Sponge	-	+	-	+
PE8-17	Sponge	-	-	-	+
PE8-18	Sponge	-	-	-	+
PE8-20	Sponge	-	-	+	-
PE8-149(B)	Sponge	+	-	-	-
PE8-152	Sponge	-	-	+	-
PE8-153	Sponge	-	-	+	-
PE13-172	Tunicate	-	+	-	+
PE13-173	Tunicate	-	-	+	+
PE13-88	Tunicate	-	-	-	+
PE13-169	Tunicate	-	+	+	-
PE13-46	Tunicate	-	+	-	-
PE13-181	Tunicate	-	-	+	-
Total		**1**	**5**	**6**	**7**

(*E.c*) = *E. coli* 1699 **(*B.c*)** = *B. cereus* ATCC 10702 **(*S.e*)** = *S. epidermidis* ATCC 14990 **(*M.s*)** = *M. smegmatis* LR222.

3.2.3 Effect of growth medium on anti-bacterial activity (secondary screening)

OSMAC involves the manipulation of growth conditions with the aim of activating silent SM pathways (section 1.8.2). This approach was adopted in this study to investigate the anti-bacterial activity of the bioactive sponge isolates on solid GYM medium (**Note**: anti-bacterial activity of the sponge isolates was previously investigated on solid only (ACM)).

When grown on solid GYM medium, isolates PE8-15, PE-16, PE817, PE18 and PE8-149(B) all inhibited the multi-drug resistant *E .coli* 1699 strain. The change in anti-bacterial profile implicates expression of a different bioactive compound. In contrast, a loss of activity against *S. epidermidis* was observed in isolates PE8-15 and PE8-16 when grown on solid GYM medium. In addition, a change in medium did not affect the anti-bacterial profiles of isolates PE8-20 and PE8-152 (against *E. coli* 1699) and PE8-149(B) (against *M. smegmatis*).

Table 3.2: Effect of growth medium on the anti-bacterial activity

Strain	*E. c*		*B. c*		*S. e*		*M.s*		*A.f*	
	GYM	ACM	GYM	ACM	GYM	ACM	GYM	ACM	Zobell	ACM
PE8-15	+	-	+	+	-	+	ND	ND	+	-
PE8-16	+	-	+	+	-	+	ND	ND	-	-
PE8-17	+	-	+	+	ND	ND	ND	ND	-	-
PE8-18	+	-	+	+	ND	ND	ND	ND	-	-
PE8-20	+	+	ND	ND	ND	ND	ND	ND	-	-
PE8-152	+	+	ND	ND	ND	ND	ND	ND	-	-
PE8-149 (B)	+	-	ND	ND	ND	ND	+	+	-	-

(*E.c*) = *E. coli* 1699 **(*B.c*) =** *B. cereus* ATCC 10702 **(*S.e*) =** *S. epidermidis* ATCC 14990 **(*M.s*) =** *M. smegmatis* LR222 **(*A.f*) =** *A. fumigatus* ATCC 46645. **(+) -** Denotes a change in the anti-microbial activity as a result of a change in growth medium and **(ND) -** Indicates that anti-microbial activity was 'Not determined' experimentally

3.2.3.1 The effect of liquid medium on anti-bacterial activity

In an attempt to further diversify the bioactivity profile using OSMAC, selected isolates were fermented in liquid ACM and GYM media. The concentrated fermentation broth was tested for anti-bacterial activity as described in section 2.6.2. Anti-bacterial activity was detected against all indicator strains (figure 3.2) including *M. smegmatis* LR222. Inhibition of *M. smegmatis* was interesting as none of selected strains that were

screened previously exhibited this activity with the exception of PE8-149(B). The new anti-bacterial profiles observed in the liquid medium further demonstrate the production of a different bioactive compound(s).

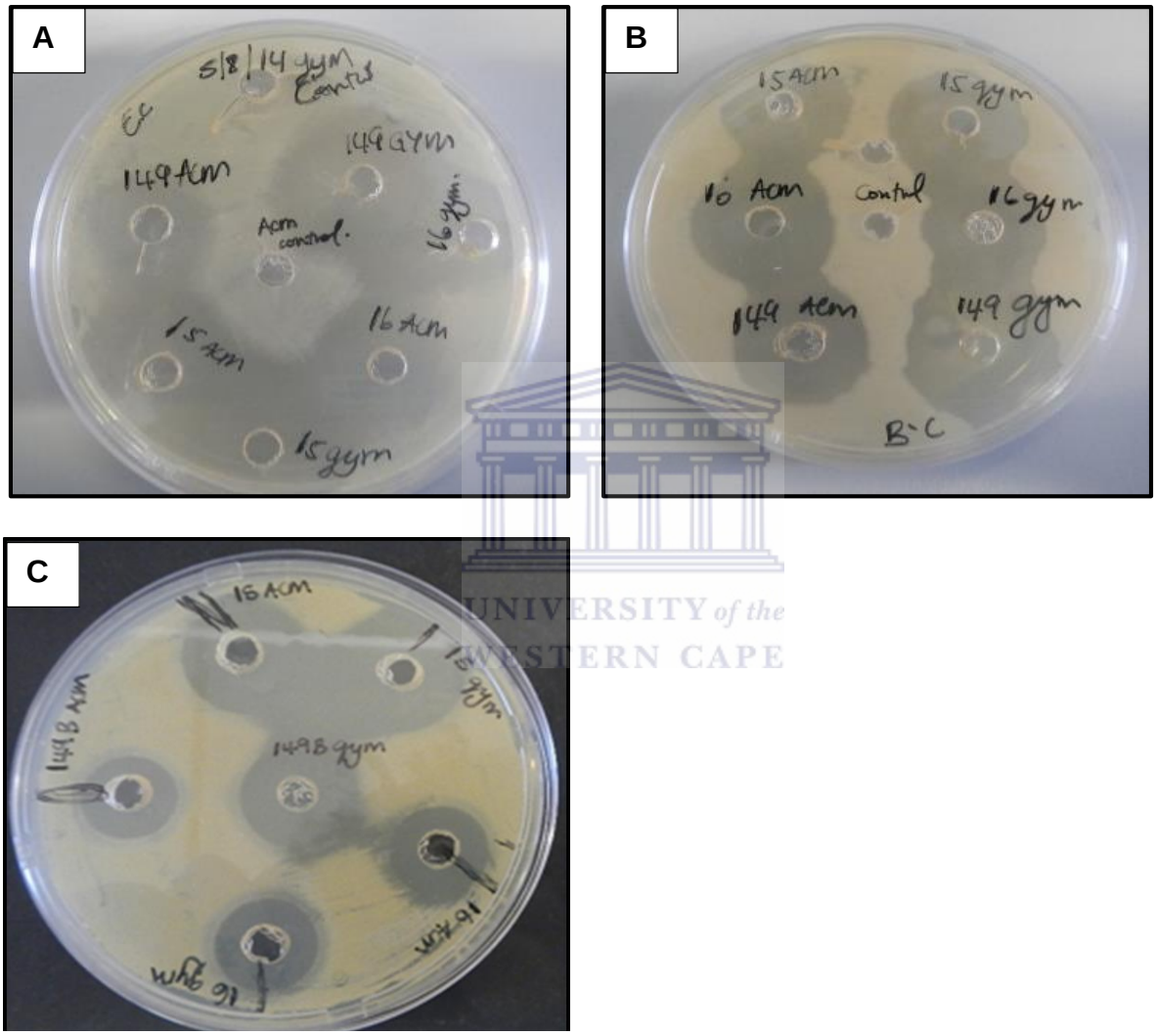

Figure 3.2: Detection of anti-bacterial activity in the fermentation broth of isolates PE8-15, 16 and 149(B) using the well-diffusion assay. Zones of inhibition indicate bacterial growth inhibition against the indicator strain. Photographs shown indicate anti-bacterial activity against *E. coli 1699* **(A)**, *B. cereus* ATCC 10702 **(B)** and *S. epidermidis* ATCC 14990 **(C)**. Activity against *M. smegmatis* LR222 is not shown in above. Sterile ACM and GYM media were used a negative control.

3.2.3.2 Isolation of bioactive compounds

In a collaborative effort with the Medina Foundation (Granada, Spain), extracts were prepared from liquid cultures of PE8-15 (using four different media including Zobell, TSA, GYM and ACM) and screened against a panel of clinical indicator strains including fungi and yeast. These screening efforts identified anti-fungal activity (against *A. fumigatus* ATCC 46645) in extracts prepared from PE8-15 grown in Zobell medium. No bioactivity was detected in extracts prepared from TSA, GYM and ACM.

The above extracts were prepared using a hydrophobic resin (Sepabeads®) and therefore the compound isolation procedure was selective for compounds with hydrophobic properties. To reduce the bias towards hydrophobic compounds, two additional resins (cationic/anionic-exchange resins) were incorporated into the extract preparation procedure. ACM fermentation broth extracts prepared using ionic-exchange resins did not exhibit any anti-bacterial activity (figure 3.3). In contrast, anti-bacterial activity was detected in both ionic-exchange extracts from the GYM fermentation broth (figure 3.3). Anti-bacterial activity was however more pronounced in the cationic-exchange resin derived extracts. These results suggest that the bioactive within the ACM broth differed to that of the GYM broth as activity would have been retained in the ACM ionic-exchange extracts if similar compound(s) were present. As part of an on-going collaboration between IMBM and the Medina Foundation (Granada, Spain), bioactive extracts are being investigated further to identify the active compounds.

NPs are known for their diverse biological activities. It is therefore a logical approach to diversify the screening platform to incorporate a range of targets as done in this study. In addition, a diverse screening platform may facilitate the identification of novel

mechanisms of action by even previously characterised compounds (Tulp and Bohlin, 2005).

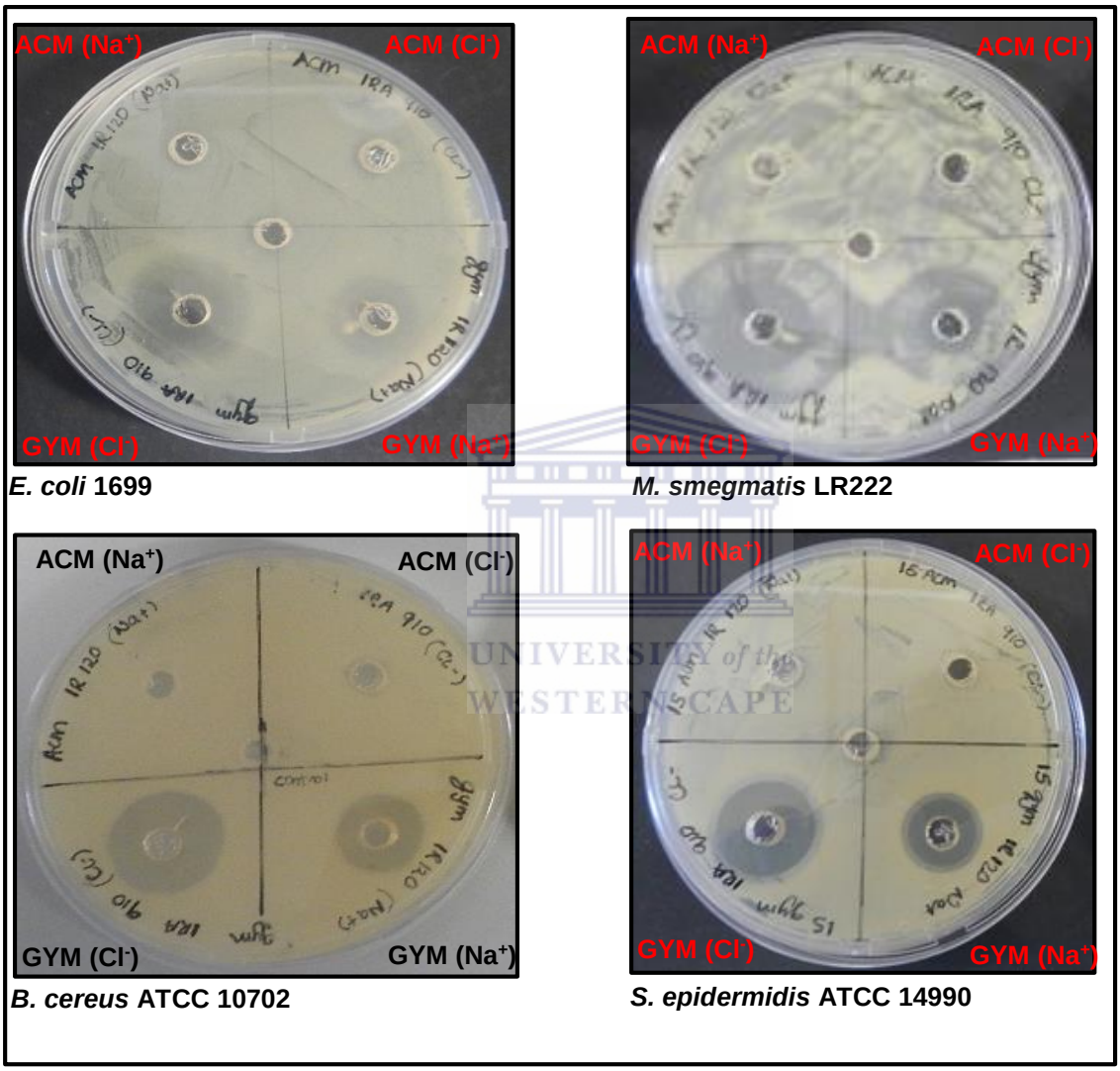

Figure 3.3: Anti-bacterial activity exhibited by ionic-exchange extracts prepared from the PE8-15 fermentation broths. Only extracts obtained from GYM media retained activity against the indicator strains. Extracts obtained from the cationic-exchange fraction (lower left quadrant) showed more potent activity against test strains in comparison to extracts obtained from the anionic-exchange column (lower right quadrant. The well in the centre of the plate represents the negative control (MilliQ water).

Antimicrobial activities shown include; (A) – *E. coli* 1699, (B) – *M. smegmatis* LR 222, (C) – *B. cereus*, (D) – *S. epidermidis*.

3.2.4 Phylogenetic Identification of bioactive isolates

Phylogenetic characterisation of bioactive isolates from *Spongia* sp. 001RSASPN and *P. africanum* Millar, 1954 was determined by 16S rRNA gene sequencing. Closest relatives were determined by a BLASTn analysis against the NCBI 16S rRNA sequence database (bacteria/archaea) presented in table 3.3.

3.2.4.1 Spongia sp. 001RSASPN associated bioactive isolates

Bioactive strains isolated from *Spongia* sp. 001RSASPN all belonged to the genus *Bacillus* (Phylum-Firmicutes). This genus consists of organisms found ubiquitously in a diverse range of environments perhaps due to their spore-forming ability (Nicholson *et al*., 2000). Furthermore, organisms in the genus *Bacillus* are known for their production of various bioactive compounds and possibly play a role in host-defence within *Spongia (Spongia)* sp. 001RSASPN. Bioactive *Bacillus* isolates have previously been isolated from various marine sponges (Zhang *et al*., 2009; Pabel *et al*., 2003: Hentschel *et al*., 2001). Although the sponge bioactive strains showed a very high sequence identity (>99%) to their top BLAST hits, the reference strains were not studied in the context of SM production. On this basis, these strains were considered as preliminary candidates for further analysis.

3.2.4.2 P. africanum Millar, 1954 associated bioactive isolates

In contrast to the sponge isolates, bioactive bacteria associated with *P. africanum* Millar 1954 belonged to three different genera including *Pseudovibrio* (Phylum-Proteobacteria), *Bacillus* (Phylum-Firmicutes) and *Streptomyces* (Phylum-Actinobacteria). The dominant bioactive isolates (under the conditions tested) were the four *Pseudovibrio* strains that show 99% sequence identity to *Pseudovibrio ascidiaceicola.*

An important characteristic of organisms in the genus *Pseudovibrio* is their production of bioactive SMs exemplified by the heptylprodigiosins from *P. denitrificans* strain Z143-1 (Sertan-de Guzman *et al*., 2007) and tropodithietic acid from *Pseudovibrio* sp. D323 (Penesyan *et al*., 2011). A subsequent BLAST analysis against the NCBI nucleotide collection (nr/nt) database was performed. This analysis indicated that the *Pseudovibrio* strains in this study shared 99% identity to *Pseudovibrio* strains isolated from various marine invertebrate organisms of which some exhibited anti-microbial activity (O'Halloran *et al*., 2011; Heindl *et al*., 2010).

The *Streptomyces* isolate (PE13-181) was highly similar to *Streptomyces flavogriseus* ATCC 33331 whose genome has previously been sequenced in which a potentially novel carbapenem antibiotic was identified (Blanco, 2012). As a result of the high similarity to a sequenced organism, PE13-181 was not considered for further analysis.

Table 3.3: Identification of bioactive isolates associated with marine invertebrates in this study

Strain ID	Closest relative (BLAST)	16S rRNA fragment length (bp)	Identity (%)	Source	Accession no.
PE8-15	*Bacillus mycoides* strain NBRC 101228	1416	99	*Spongia sp*. 001RSASPN	NR_113990.1
PE8-16	*Bacillus mycoides* strain NBRC 101228	1421	99	*Spongia sp*. 001RSASPN	NR_113990.1
PE8-17	*Bacillus cereus* strain ATCC 14579	1428	99	*Spongia sp*. 001RSASPN	NR_074540.1
PE8-18	*Bacillus anthracis* str. *Ames* strain *Ames*	669	99	*Spongia sp*. 001RSASPN	NR_074453.1
PE8-20	*Bacillus mycoides* strain NBRC 101228	1420	99	*Spongia sp.* 001RSASPN	NR_113990.1
PE8-149(B)	*Bacillus mycoides* strain NBRC 101228	1427	99	*Spongia sp*. 001RSASPN	NR_113990.1
PE8-152	*Bacillus weihenstephanensis* KBAB4 strain KBAB4	699	99	*Spongia sp*. 001RSASPN	NR_074926.1
PE8-153	*Bacillus cereus* ATCC 14579	1376	99	*Spongia sp*. 001RSASPN	NR_074540.1
PE13-172	*Pseudovibrio ascidiaceicola* strain NBRC 100514	1351	99	*P. africanum* Millar, 1954	NR_113916.1
PE13-173	*Pseudovibrio ascidiaceicola* strain NBRC 100514	766	99	*P. africanum* Millar, 1954	NR_113916.1

PE13-88	*Bacillus aerius* strain 24K	803	99	*P. africanum* Millar, 1954	NR_118439.1
PE13-169	*Pseudovibrio ascidiaceicola* strain NBRC 100514	1351	99	*P. africanum* Millar, 1954	NR_113916.1
PE13-46	*Pseudovibrio ascidiaceicola* strain NBRC 100514	1352	99	*P. africanum* Millar, 1954	NR_113916.1
PE13-181	*Streptomyces flavogriseus* strain ATCC 33331	1385	99	*P. africanum* Millar, 1954	NR_074559.1

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

3.3.1 Abundance and distribution of secondary metabolite pathways within the genus Bacillus

Bioactive bacteria isolated from *Spongia* sp. 001RSASPN all belonged to the genus *Bacillus* (section 3.2.4.1). Species within the *Bacillus* genus are generally known for their versatility in SM production. For example, genome analysis of *B. amyloliquefaciens* FZB42 showed that 8.5% of the genome was involved in secondary metabolism (Chen *et al*., 2007). Genomes available on Genbank for species closely related to the bioactive strains isolated from the *Spongia* sp. 001RSASPN were analysed to determine the distribution and abundance of SM pathways (figure 3.4 and 3.5 respectively). A major benefit of performing such an analysis prior to PCR screening is that it facilitates prioritisation for SM genes likely to be present within an organism(s).

NRPS and bacteriocin pathways were the most abundant class of SMs and were present in all genomes analysed. In comparison, terpene and siderophore biosynthetic genes were moderately distributed in all species with the exception of *B. mycoides* which lacked a siderophore biosynthetic pathway. Hybrid pathways (NRPS/transAT-PKS and NRPS/bacteriocin) were only detected in *B. subtilis* strains whereas the lantipeptide class of SMs was only present in *B. cereus*, *B. thuringiensis* and *B. subtilis* strains. The least abundant class, the phosphonates, were only present in *B. subtilis* strains. Although it would be ideal to represent each genome equally for such an

analysis, *B. mycoides* and *B. weihenstephanensis* lacked multiple strains with sequenced genomes and were therefore under-represented.

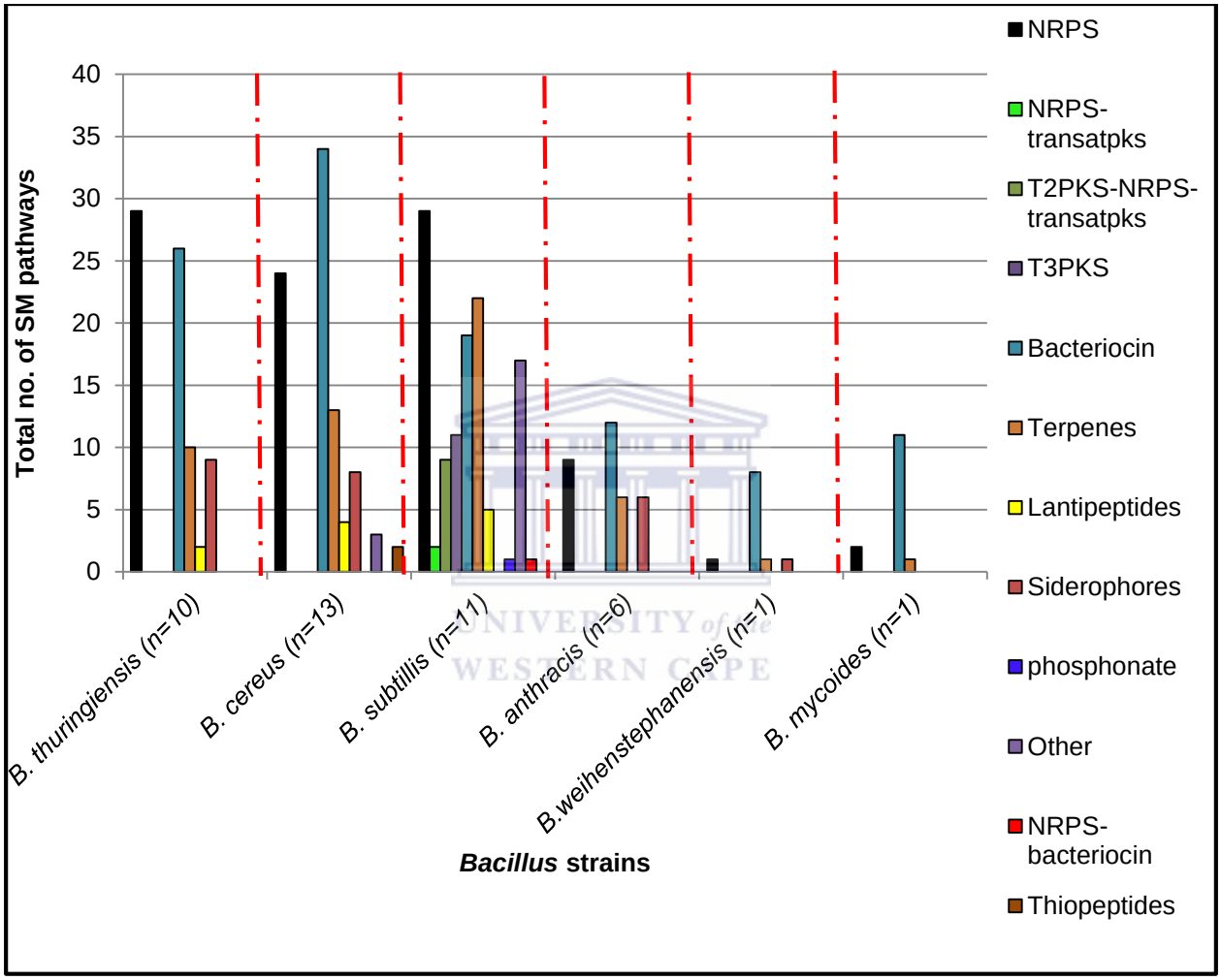

Figure 3.4: Distribution of secondary metabolite classes in selected *Bacillus* strains. *Bacillus* strains for this analysis were selected based on their relatedness to the bioactive *Bacillus* isolates in this study determined by 16S rRNA sequencing and analysis (section 3.2.4.1). Genome mining was done using the antiSMASH2.0 software. The number of strains analysed for each *Bacillus* species were as follows; *B. thuringiensis*; n = 10, *B. cereus*; n = 13, *B. subtilis*; n = 11, *B. anthracis*; n = 6, *B. weihenstephanensis*; n = 1, *B. mycoides*; n = 1. Accession numbers for each strain analysed are listed in the Appendix.

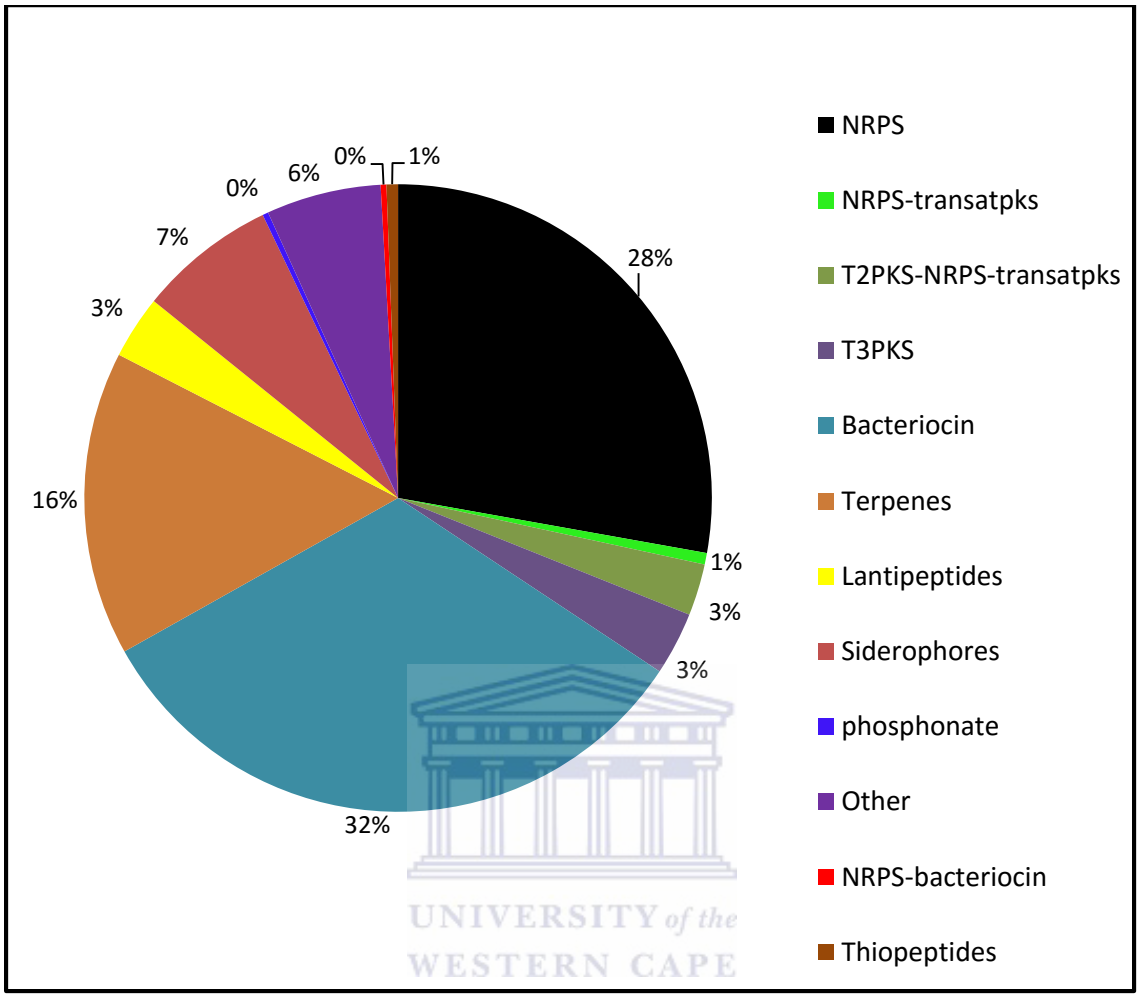

Figure 3.5: The average abundance of SM pathways in selected *Bacillus* strains. *Bacillus* strains for this analysis were selected based on their relatedness to the bioactive *Bacillus* isolates in this study determined by 16S rRNA sequencing and analysis (section 3.2.4.1). *In silico* analysis of these genomes was done using antiSMASH2.0. The number of strains analysed for each *Bacillus* species were as follows; *B. thuringiensis*; n = 10, *B. cereus*; n = 13, *B. subtilis*; n = 11, *B. anthracis*; n = 6, *B. weihenstephanensis*; n = 1, *B. mycoides*; n = 1. NRPS (shaded black) and bacteriocins (shaded blue) genes were present in all genomes analysed and represented the majority of SM classes (28% and 32% respectively).

3.3.2 PCR screening for NRPS genes

Due to a relatively high abundance of NRPS genes in *Bacillus* species (figure 3.5), the bioactive *Bacillus* strains were screened by PCR for NRPS genes using degenerate primers targeting the A-domain (Vizcaino *et al*., 2005). The use of degenerate primers is significant as the amplification is not limited to known gene sequences. NRPS genes were detected in all isolates screened (figure 3.6). A BLAST analysis (Altchul *et al.,* 1990) confirmed the correct amplification of the NRPS A-domain in each isolate. These data are summarised in table 3.4.

All NRPS amplicons exhibited high sequence identity (> 90%) to their top two BLAST hits with the exception of clone 1 from PE8-15. Clone 1 (PE8-15) shared 65% and 50% sequence identity respectively to the top two BLAST hits. The high sequence identity (>90%) of A-domain sequences suggests that they possibly originate from previously characterised pathways or from organisms whose genomes are sequenced (Gontang *et al*., 2010). Furthermore, all BLAST hits originated from sequences in *B. cereus* and *B. weihenstephanensis* NBRC 101238 = DSM 11821 with the exception of clone 1 of PE8-15, whose sequences were more identical to those originating from *B. oceanisediminis* 2691 and *Cyanothece sp.* PCC 7424.

A variation in the nucleotide identity between clones from the same organism was observed in isolates PE8-149(B), PE8-20 and PE8-15 (table 3.4). Although this could be due to PCR errors introduced by the non-proof reading DNA polymerase (Dream*Taq*), it could also suggest amplification of multiple A-domains within the same organism.

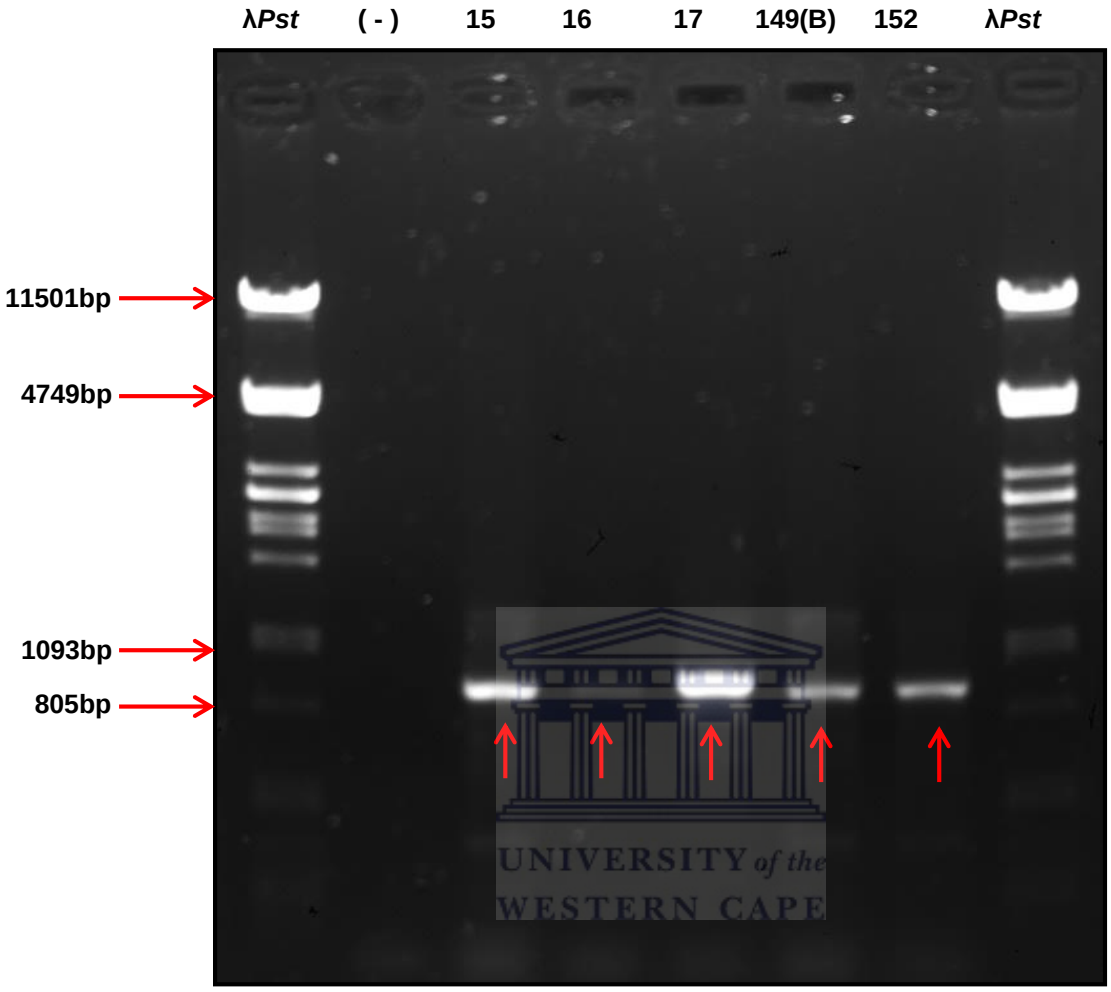

Figure 3.6: PCR amplification of NRPS genes (specifically the A-domain) in bioactive isolates associated with *Spongia* sp. 001RSASPN. The expected amplicon size was approximately 900bp. Lanes 1 and 8: DNA molecular marker (lambda DNA digested with *Pst*I). Lane 2: Negative control (Milli-Q water as a template). Lane 3-7: Isolates PE8-15, 16, 17, 149(B), 152 respectively. PCR amplicons were resolved in a 1% [w/v] TAE agarose gel.

Table 3.4: Sequence identity of NRPS amplicons

Isolate	Clone	Top BLAST hits (top 2 hits)	Identity (%)	Query cover	Accession no.
PE8_152	Clone 3	AMP binding protein *B. cereus*	99	95%	WP_002032315.1
		Putative non-ribosomal peptide synthetase [*Bacillus weihenstephanensis* NBRC 101238 = DSM 11821].	99	95%	GAE38254.1
PE8_152	Clone 2	AMP binding protein *B. cereus*	99	93%	WP_002032315.1
		Amino acid adenylation domain-containing protein *B. cereus*	99	93%	WP_016120260.1
PE8_152	Clone 1	AMP binding protein *B. cereus*	99	93%	WP_002032315.1
		Putative non-ribosomal peptide synthetase [*Bacillus Weihenstephane9nsis* NBRC 101238 = DSM 11821].	99	93%	GAE38254.1
PE8_149 (B)	Clone 3	AMP binding protein *B. cereus*	99	96%	WP_002032315.1
		Putative non-ribosomal peptide synthetase [*Bacillus weihenstephanensis* NBRC 101238 = DSM 11821].	99	96%	GAE38254.1
PE8_149 (B)	Clone 4	Amino acid adenylation domain-containing protein *B. cereus*	91	95%	WP_016127041.1
		Amino acid adenylation domain-containing protein *B. cereus*	91	95%	WP_016101927.1
PE8_149 (B)	Clone 5	AMP binding protein *B. cereus*	93	96%	WP_002032315.1
		Putative non-ribosomal peptide synthetase [*Bacillus*	93	96%	GAE38254.1

		weihenstephanensis NBRC 101238 = DSM 11821].			
PE8_20	Clone 3	AMP binding protein *B. cereus*	98	99%	WP_002032315.1
		Putative non-ribosomal peptide synthetase [*Bacillus weihenstephanensis* NBRC 101238 = DSM 11821].	98	99%	GAE38254.1
PE8_20	Clone 4	AMP binding protein *B. cereus*	98	99%	WP_002032315.1
		Putative non-ribosomal peptide synthetase [*Bacillus weihenstephanensis* NBRC 101238 = DSM 11821].	98	99%	GAE38254.1
PE8_20	Clone 5	MULTISPECIES: hypothetical protein (*Bacillus cereus* group)	99	51%	WP_002086912.1
		Hypothetical protein C174 15052 *Bacillus weihenstephanensis* FSL H7-687	99	51%	ETT76802.1
PE8_15	Clone 1	Hypothetical protein *Bacillus oceanisediminis*	65	90%	WP_019379468.1
		Amino acid adenylation protein (*Cyanothece sp.* PCC 7424	50	90%	WP_015956918.1
PE8_15	Clone 6	AMP binding protein *B. cereus*	98	98 %	WP_002032315.1
		Putative non-ribosomal peptide synthetase [*Bacillus weihenstephanensis* NBRC 101238 = DSM 11821].	98	98 %	GAE38254.1

3.3.3 PCR screening for PKS genes

Organisms in the genus *Streptomyces* are a well-known source of bioactive compounds particularly from the polyketide class (Berdy, 2005). Three isolates associated with *P. africanum* Millar, 1954 were identified as *Streptomyces* strains. Of these three isolates, only PE13-181 exhibited anti-bacterial activity against at least one indicator organism (*E. coli* 1699) whereas PE13-182 and PE13-175 did not exhibit any activity under the conditions tested.

Nevertheless, because *Streptomyces* species are an important source of polyketides and due to the well-known “cryptic pathways” phenomena in this genus, all three strains were screened with degenerate primers for type I and type II PKS genes. As depicted in figure 3.7, type I PKS genes were detected in isolates PE13-181 and PE13-182. Type II PKS genes were only detected in isolate PE13-181 (figure 3.8). Although PKS genes were detected in isolate PE13-182, this strain previously did not exhibit any anti-bacterial activity under the conditions tested. These results therefore highlight the importance of integrating bioactivity-guided screening with sequence-based screening. The absence of amplification in isolate PE13-175 can be explained by PCR bias as well as undetectable levels of product amplification. However, complete absence of PKS pathways in this organism cannot be ruled out.

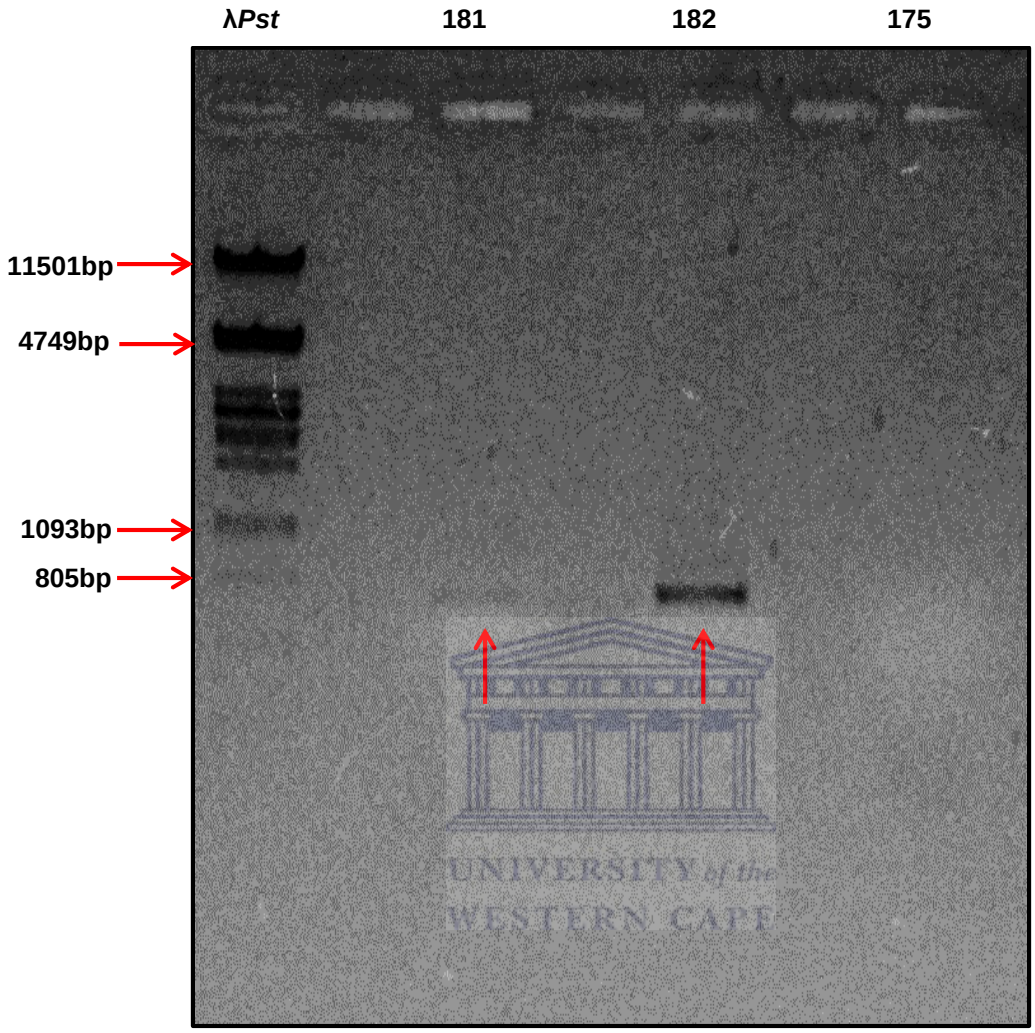

Figure 3.7: Amplification of type I PKS genes in *Streptomyces* isolates associated with *P. africanum* Millar, 1954. The expected amplicon size was approximately 700bp. Lane 1: DNA molecular marker (lambda DNA digested with *Pst*I). Lane 3**:** isolate PE13-181. Lane 5: isolate PE13-182. Lane 7: isolate PE13-175. Amplicons were resolved in a 1% [w/v] TAE agarose gel.

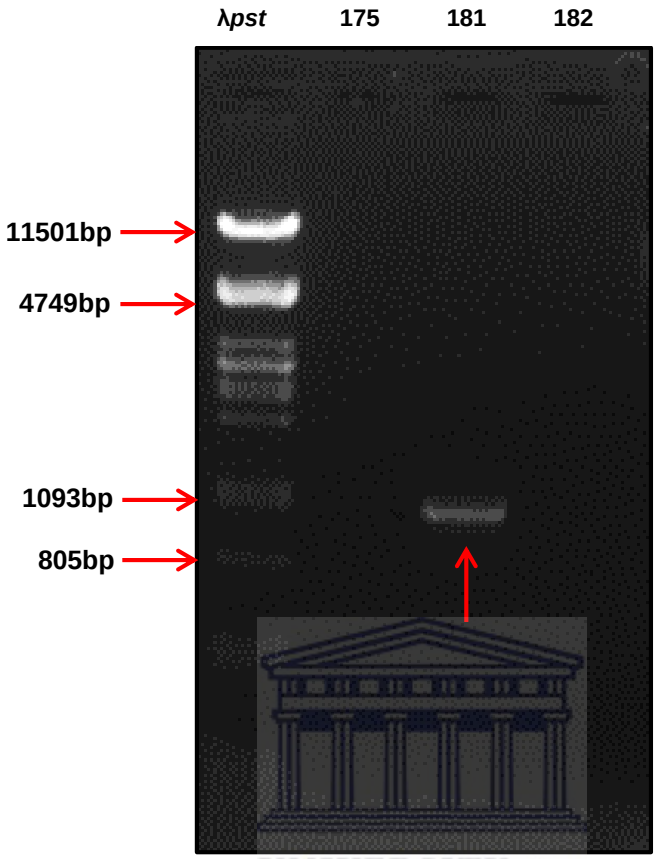

Figure 3.8: Amplification of type II PKS genes in the *Streptomyces* isolates associated with *P. africanum* Millar, 1954. The expected product size was approximately 1070bp. Lane 1: DNA molecular marker (lambda DNA digested with *Pst*I). Lane 2-4: Isolates PE13-175, 181 and 182 respectively. PCR products were resolved in a 1% [w/v] TAE agarose gel.

Table 3.5: Sequence analysis of PKS (KS) sequences

Isolate	PKS type	Clone	Top BLAST hits	Identity (%)	Query cover	Accession no.
PE13-181	Type I	Clone 2	Polyketide synthase, partial [*Streptomyces* sp. ID05-A0431]	94	98%	BAH68103.1
			Polyketide synthase, partial [*Streptomyces* sp. ID05-A0096]	94	98%	BAH67688.1
PE13-181	Type I	Clone 5	Polyketide synthase, partial [*Streptomyces* sp. ID05-A0431]	94	96%	BAH68103.1
			Polyketide synthase, partial [*Streptomyces* sp. ID05-A0096]	94	96%	BAH67688.1
PE13-182	Type I	Clone 1	Polyketide synthase, partial [*Streptomyces* sp. ID05-A0096]	99	99%	BAH67686.1
			Type I polyketide synthase, partial [*Streptomyces* sp. CNR-885 PL04]	98	95%	ACZ54341.1
PE13-182	Type I	Clone 3	Polyketide synthase, partial [*Streptomyces* sp. ID05-A0096]	98	99%	BAH67686.1
			Type I polyketide synthase, partial [*Streptomyces* sp. CNR-885 PL04]	97	95%	ACZ54341.1
PE13-182	Type I	Clone 4	Polyketide synthase, partial [*Streptomyces* sp. ID05-A0431]	99	100%	BAH68102.1
			Type I polyketide synthase, partial [*Streptomyces* sp. CNR-885	98	95%	ACZ54340.1

PE13-181	Type II	Polyketide beta-ketoacyl synthase WhiE-KS [*Streptomyces* sp. PAMC26508]	100	78%	WP_015575864.1
		MULTISPECIES: beta-ACP synthase [*Streptomyces*]	100	78%	WP_014158007.1

In summary, isolates were considered unsuitable for further analysis (WGS) based on the following criteria:

(I) *Isolates that were closely related to an organism(s) whose genome has been sequenced and studied for the purpose of SM discovery (for example PE13-181).*

(II) *Isolates exhibiting a limited range of anti-microbial activity*

(III) *Isolates with NRPS and PKS amplicons exhibiting sequence identity greater than 75%.*

(IV) *Isolates closely related to strains previously isolated from marine invertebrates* ***(****PE13-172, PE13-173, PE13-169 and PE13-46).*

Table 3.6: Summary of PCR screening results

Isolate	Genus	Type I PKS	Type II PKS	NRPS
PE8-15	*Bacillus*	ND	ND	+
PE8-16	*Bacillus*	ND	ND	+
PE8-17	*Bacillus*	ND	ND	+
PE8-149(B)	*Bacillus*	ND	ND	+
PE8-152	*Bacillus*	ND	ND	+
PE13-181	*Streptomyces*	+	+	ND
PE13-182	*Streptomyces*	+	-	ND
PE13-175	*Streptomyces*	-	-	ND

(ND) - Indicates genes whose presence was not determined experimentally.

3.4 Whole genome sequencing (WGS)

Among all the bioactive isolates, PE8-15 was selected based on the selection criteria outlined above. This isolate exhibited the most diverse bioactivity profile with inhibition against gram-positive, gram-negative as well as eukaryotic organisms (fungi). In addition, the low sequence identity of the NRPS amplicon from clone 1 (PE8-15) predicts the presence of a potentially novel NRPS pathway.

Following quality trimming, a total of 4,625,328 reads averaging 193.7bp in length were retrieved. *De novo* assembly resulted in 101 contigs and a total genome size of 5,911,306bp with an average GC-content of 36%. This genome size was slightly larger

in comparison to the genome from the *B. mycoides* strain 219298 (the closest relative to PE8-15 based on 16S rRNA sequence analysis) with a genome size of 5.6753Mbp.

3.4.1 Genome mining of isolate PE8-15

Using antiSMASH2.0, the genome of PE8-15 was mined for SM specific pathways. A total of 10 pathways were identified (figure 3.9) belonging to four different classes. The genome predominantly contained bacteriocin (5) and NRPS (3) pathways. In contrast, siderophore and terpene pathways were the least abundant with one pathway of each detected. These results are consistent with the *in silico* analysis of SM abundance in selected *Bacillus* strains (section 3.3.1).

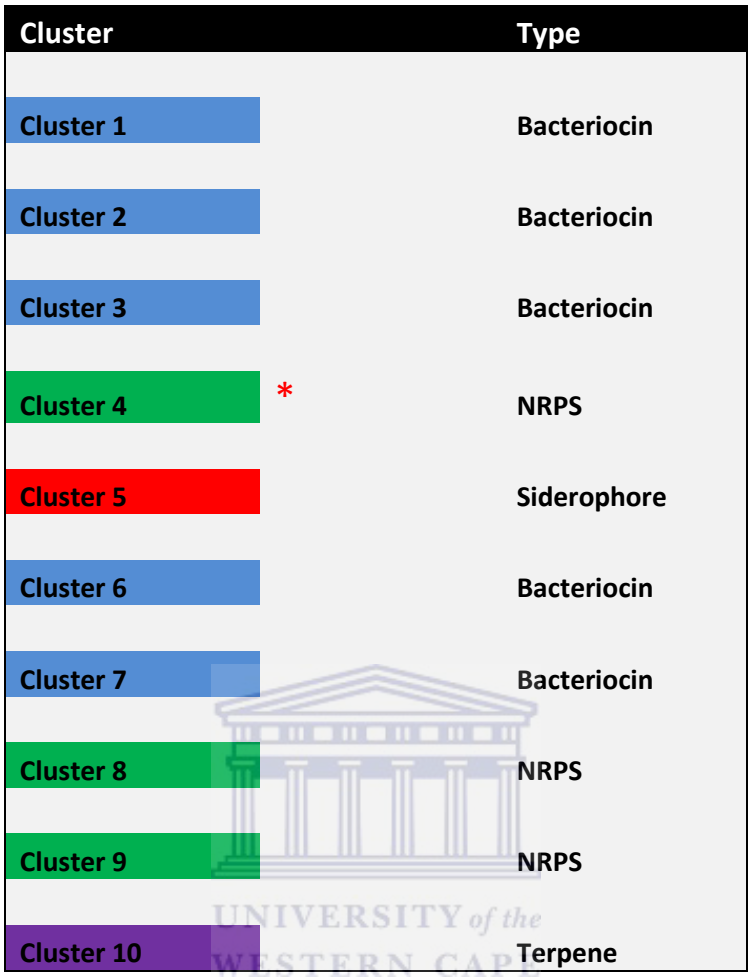

Figure 3.9: Secondary metabolite pathways identified in the genome of isolate PE8-15 using the antiSMASH2.0 software. A total of 10 SM pathways were detected of which the bacteriocin and NRPS classes were the most abundant with a total of 5 and 3 pathways respectively. Siderophore and terpene pathways were the least abundant with only one of each detected. The NRPS amplicon from clone 1 (PE8-15) originated from Cluster 4 (highlighted by a red asterisk).

3.4.2 Identification of an NRPS pathway in isolate PE8-15

Using clone 1 from PE8-15 as a query sequence, contig_18 was identified to contain Cluster 4 highlighted in figure 3.9. Cluster 4 consists of 4 NRPS genes (ORF36, 39, 41 and 43) all oriented in the same direction (figure 3.10). Located downstream of the

NRPS gene cluster are 4 adjacent PKS or fatty acid synthase (FAS) related genes (ORF46, 47, 48 and 49) separated from the NRPS genes by two ORFs (44 and 45). The PKS/FAS genes exhibited type II architecture in which each domain is located on discrete ORFs. The presence of both NRPS and PKS/FAS components suggest a hybrid NRPS-PKS/FAS pathway.

A PPTase and TE are located upstream of the NRPS genes in ORF30 and 29 respectively. SM pathways frequently contain two TE-domains of which one is integrated into the final module also known as TEI, whereas the additional TE (TE II) is discretely situated within the gene cluster (Claxton *et al*., 2009). TEI is involved in the catalytic release of the final peptide product (Hutchinson, 2003). In contrast, TEII is known to play various roles that among others, includes the removal of atypical acyl units from the carrier proteins (Claxton *et al*., 2009).

Analysis of Cluster 4 identified the presence of a discrete TE (TEII) in ORF29 although a typical C-terminal (final module) TEI is absent. Despite product release often carried out by a TEI, the discrete TE (TEII) is capable of this activity. For example, the TEII LovG situated in the lovastatin biosynthetic gene cluster, was shown to release dihydromonacolin L acid in addition to removal of anomalous intermediate products (Xu *et al*., 2013). Therefore, due to the lack of an integrated TE (TEI) within Cluster 4, it is possible that the discrete TE (ORF29) is involved in product release.

An MbtH-like protein (ORF45) is situated in the region between the NRPS and PKS/FAS gene clusters. These proteins are commonly associated with NRPS and siderophore biosynthetic gene clusters (Baltz, 2011). Although the specific role played

by MbtH-like proteins is still unclear, they have been implicated in forming complexes with A-domains and playing a role in the adenylation reactions (Herbst *et al*., 2013).

The presence of two transposase genes flanking Cluster 4, suggest that this cluster is a mobile genetic element. Mizuno and colleagues (2013) previously identified a hybrid NRPS/PKS related to the bleomycin family of antibiotics flanked by transposase genes. This was subsequently hypothesized to have been acquired by HGT. Comparison of the GC-content of the hybrid gene cluster and the plasmid in which the hybrid cluster was identified, indicated a higher GC-content in the hybrid gene cluster. It was therefore suggested that the bleomycin-like gene cluster was acquired through HGT.

To determine the possibility of HGT in the acquisition of Cluster 4, a GC-content analysis between Cluster 4 and the PE8-15 genome was conducted. This analysis showed a relatively small difference in GC-content between Cluster 4 (35%) and the PE8-15 genome (36%) therefore suggesting that HGT did not play a role in acquisition of Cluster 4. It is however possible that Cluster 4 was acquired from a closely related organism (as opposed to a distantly related organism) and hence the minimal difference in GC-content. Alternatively, if Cluster 4 was acquired a long time ago from a distantly related organism, the GC-content could ameliorate to that of its receiving host (Lawrence and Ochman, 1997). The latter possibly explains why the A-domain amplified from PE8-15 (clone 1) showed significant sequence similarity (50%) to a sequence originating from *Cyanothece* (a genus in the phylum Cyanobacteria). It is also possible that Cluster 4 is mobile specifically between specific *Bacillus* species as opposed to distantly related organisms. Additional genes associated with Cluster 4 are summarised in table 3.7

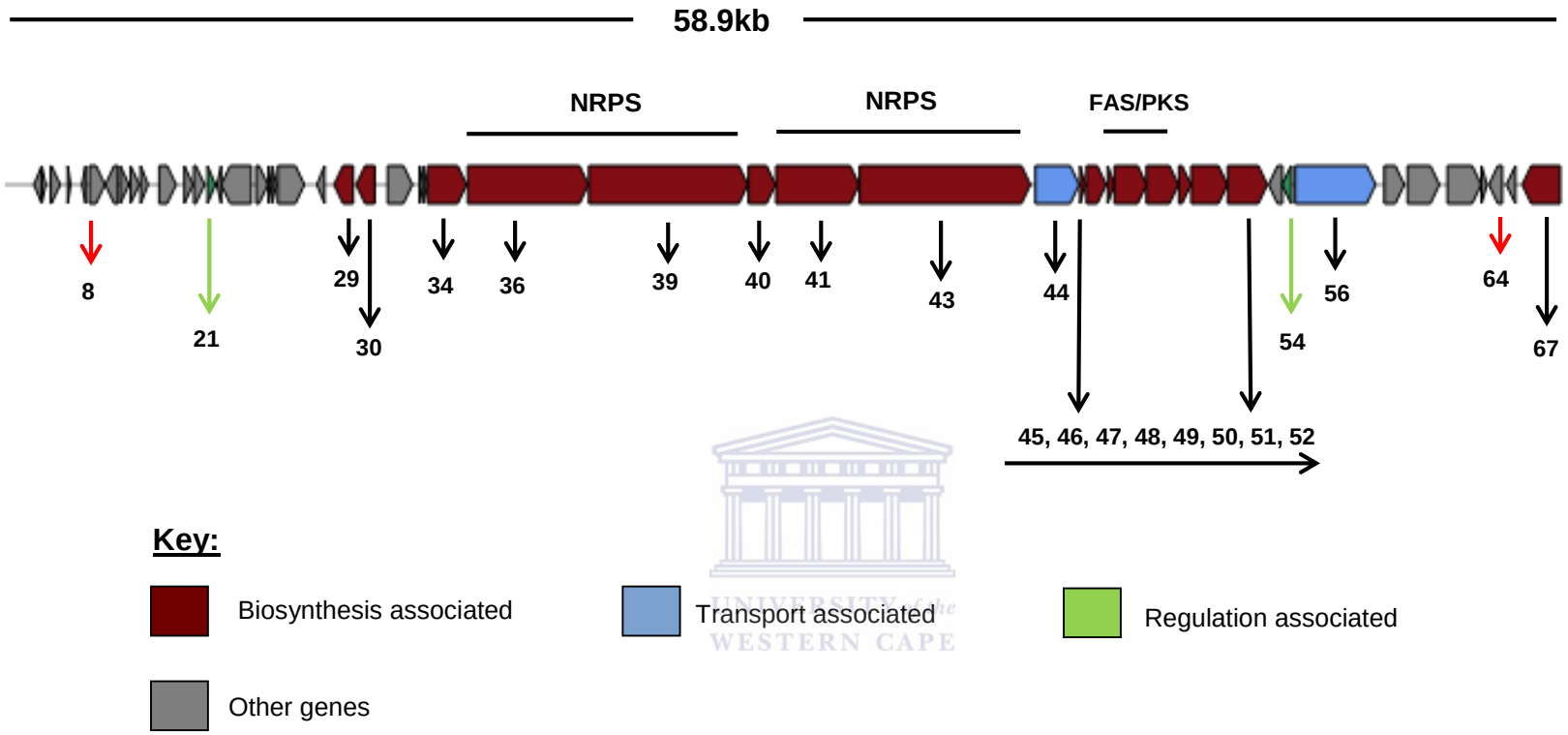

Figure 3.10: Prediction of ORFs on contig_18 using antiSMASH2.0. Each ORF is assigned a number and colour coded according to their role as described in the key. ORFs coding for putative transposase genes are identified with a red arrow and are seen to flank either side of Cluster 4. Transcriptional regulatory genes are identified with a green arrow.

Table 3.7: Predicted ORFs within Cluster 4 and their putative functions

ORF	Gene	Top BLAST hit	Proposed function	I / S (%)	Accession no.
8	Transposase	Transposase [*Bacillus cereus*]	DNA repair, replication and recombination	82/83	WP_002191895.1
21	ArsR family transcriptional regulator	Transcriptional repressor [*Bacillus cereus*]	Arsenical resistance operon repressor	99/100	WP_016095069.1
29	Thioesterase	Hypothetical protein [*Bacillus oceanisediminis*]	Predicted thioesterase involved in NRPS biosynthesis	68/81	WP_019379475.1
30	Phosphopantetheinyl transferase	Hypothetical protein [*Bacillus oceanisediminis*]	Transfers the 4'-phosphopantetheine group from coenzyme A (CoA) to the invariant serine resulting in a holo-ACP	46/69	WP_026041532.1
34	Dehydrogenase	Hypothetical protein [*Laceyella sacchari*]	-	43/64	WP_022737585.1
36	NRPS	Hypothetical protein [*Bacillus oceanisediminis*]	NRP biosynthesis	56/72	WP_019379472.1
39	NRPS	Hypothetical protein [*Bacillus oceanisediminis*]	NRP biosynthesis	54/71	WP_019379471.1
40	Dioxygenase TauD/TfdA	Hypothetical protein [*Bacillus*	Taurine catabolism	70/85	WP_019379470.1

		oceanisediminis]			
41	NRPS	Hypothetical protein [*Bacillus oceanisediminis*]	NRP biosynthesis	53/74	WP_019379469.1
43	NRPS	Hypothetical protein [*Bacillus oceanisediminis*]	NRP biosynthesis	62/77	WP_019379468.1
44	ABC transporter ATP-binding protein	Hypothetical protein [*Bacillus oceanisediminis*]	Multi-drug resistance	75/86	WP_019379467.1
45	MbtH-like protein	Hypothetical protein [*Bacillus oceanisediminis*]	Synthesis of antibiotics, siderophores and glycopeptidolipids	75/82	WP_019379466.1
46	Short-chain-dehydrogenase (reductase/SDR) (KR)	3-oxoacyl-ACP synthase [*Paenibacillus mucilaginosus*]	FabG, 3-ketoacyl-acyl carrier protein reductase. NADH dependent reduction of ACP in *de novo* fatty acid synthesis	54/74	WP_014651400.1
47	Acyl carrier protein	Hypothetical protein ER50_05790 [*Bacillus safensis*]	Phosphopantetheine attachment site	43/74	KEP30945.1
48	Beta-ketoacyl synthase	Beta-ketoacyl-ACP synthase [*Bacillus gaemokensis*]	FabF-elongation during fatty acid synthesis	56/72	KEK21730.1
49	Beta-ketoacyl synthase	Hypothetical protein BAGA_26265 [*Bacillus gaemokensis*]	FabB-catalyse a claissen-like condensation in addition to fatty	37/58	KEK21729.1

			acid synthesis and degradation, polyketide synthesis		
50	[3-hydroxymyristoyl] acetylglucosamine	Beta-hydroxyacyl-(acyl-carrier-protein) Dehydratase [*Planctomycete* KSU-1]	FabZ- responsible for the dehydration of beta-hydroxyacyl-ACP to trans-2-acyl-ACP. This is the third step of bacterial or plastid type II fatty acid biosynthesis	49/70	WP_007222374.1
51	Amino transferase class-III	4-aminobutyrate aminotransferase, [*Bacillus oceanisediminis*].	Members of this family act on basic amino acids and the subsequent intermediates are involved in transamination and decarboxylation	86/92	WP_033196446.1
52	Aldehyde dehydrogenase	Hypothetical protein [*Bacillus oceanisediminis*]	Involved in the oxidation of a wide range of endogenous/exogenous aliphatic and aromatic aldehydes to their corresponding carboxylic acids and are also involved in detoxification	76/73	WP_019379464.1

54	padR family transcriptional regulator	Hypothetical protein [*Bacillus oceanisediminis*]	Transcription regulator	81/77	WP_019379462.1
56	AcrB/AcrD/AcrF family protein	Hypothetical protein [*Bacillus oceanisediminis*]	Cation/multidrug efflux pump	74/89	WP_019379461.1
64	Transposase	Transposase [*Bacillus cereus*]	DNA repair, replication and recombination	95/99	WP_002090249.1
67	Beta-lactamase	Penicillin-binding protein [*Bacillus* sp. FJAT-13831]		74/86	WP_017151866.1

3.4.3 Modular organisation of Cluster 4: a putative hybrid NRPS-PKS/FAS pathway

NRPSs are organised into modules where each module is responsible for the incorporation of one amino acid into the growing chain (section 1.3.2.1). Cluster 4 consists of 5 NRPS modules (figure 3.11) each containing the classic set of compulsory domains *i.e.* C, A and T-domain. Additional modification domains present in modules 1 and 2 include an epimerization domain. An odd observation was the presence of two C-domains in module 2. The presence of a carboxy-terminal located C-domain has previously been identified in the microcystin synthase gene cluster where it is involved in peptide-bond formation (Tillett *et al*., 2000). However, as there 5 A-domains and therefore amino acid substrates, the function of the additional C-domain in Cluster 4 requires further investigation. The presence of a C-domain in the first module, is reminiscent of lipopeptide pathways such as iturin A (Tsuge *et al*., 2001), surfactin (Bruner *et al*., 2002) and lichenysin A (Konz *et al*., 1999). The C-domain in the first module is suggested to link the first amino acid to a fatty acid moiety (Miao *et al*., 2006).

As previously discussed, PKS/FAS genes in close proximity to the NRPS cluster are suggestive of a hybrid pathway between the NRPS and PKS/FAS. The domains associated with PKS/FAS biosynthesis (2 KS, ACP and KR) are situated on individual ORFs consistent with the type II class of PKS and FAS enzymes. A major difference between polyketide and fatty acid biosynthesis is the complete and sequential reduction of fatty acids by the KR, DH and ER-domains (Crawford and Townsend, 2010). The presence of both a KR and DH-domain (ORF34) and the absence of an ER-domain suggest that a polyketide is incorporated instead. The role of the ER-domain can

however be provided by a *trans*-acting ER as seen in the type I PKS (PpsD) involved in the production of surface-exposed lipids in *Mycobacterium tuberculosis* (Siméone *et al*., 2007).

Immediately downstream of the PKS/FAS gene cluster, an ORF predicted to encode an unusual aminotransferase class-III gene (AMT) was observed. The presence of an AMT-domain is rather interesting as it has previously only been identified in the iturins, hepatotoxins (microcystin and nodularin) and prodigiosins (Aron *et al*., 2005; Moffitt and Neilan, 2004). The proposed role of this domain is the conversion of an activated fatty acid moiety into a β-amino fatty acid as observed in the biosynthesis of mycosubtilin. Mycosubtilin is a lipopeptide in the iturin family whose peptide moiety is modified by the addition of a β-amino fatty acid (Duitman *et al*., 1999).

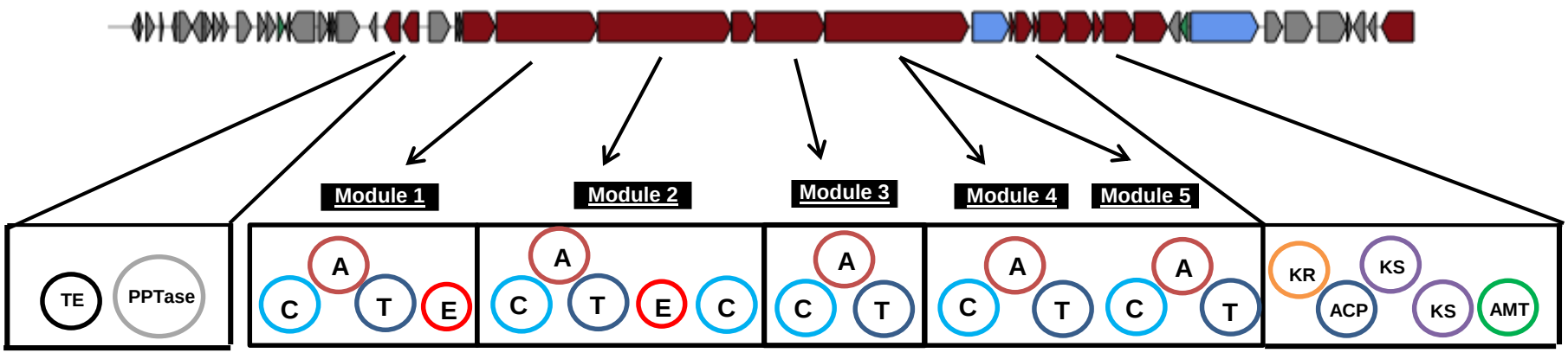

Figure 3.11: Modular organisation of Cluster 4, a putative hybrid NRPS-PKS/FAS pathway. The NRPS genes contained a total of 5 modules with each of the three core domains C-A-T present. Modules 1 and 2 each have an epimerization domain and module 2 contains an additional C-domain. The PKS/FAS like domains exhibited typical type II architecture with each core domain (KSα-KSβ-ACP) located on a discrete ORF. In addition to the core domains, the PKS/FAS cluster has a ketoreductase (KR) and an aminotransferase (AMT) domain present. Thioesterase (TE) and phosphopantetheine transferase (PPTase) genes were situated upstream of module 1.

3.4.4 Prediction of A-domain specificity in Cluster 4

The A-domain constitutes part of the compulsory domains (C-A-T) required for the successful incorporation of the cognate amino acid into the growing chain. Following the co-crystallisation of the gramicidin S A-domain with its phenylalanine substrate (Conti *et al*., 1997), amino acids involved in substrate interaction were identified. Amino acids at specific positions lining the active site were used as a predictor of the A-domain substrate specificity (Stachelhaus *et al*., 1999; Challis *et al*., 2000). This forms the basis of the NRPSpredictor2 algorithm used to predict substrate specificity of the A-domain using the specific signature of amino acids lining the active site (Röttig *et al*., 2011).

The predicted cognate amino acids for the 5 different modules in Cluster 4 are listed in table 3.8. These included; non-proteinogenic amino acids (ornithine) in modules 2 and 3 and proteinogenic amino acids (asparagine, threonine and leucine) modules 1, 4 and 5 respectively. The incorporation of non-proteinogenic amino acids (for example ornithine) is characteristic of NRPS enzymes (Hubbard *et al*., 2000).

A-domains within module 3 (ornithine) and module 4 (threonine) exhibited the highest sequence identity (70% and 100% respectively) to their closest active site signature sequence. Lowest sequence identity was observed in A-domains from module 2 (ornithine) and 5 (leucine) both exhibiting 50% identity to the closest active site signature sequence. The low identity to known active site sequence signatures could be indicative of A-domains with specificity to yet unknown substrates not represented in NRPSpredictor2 reference database. However, the degeneracy of the amino acid specificity signature could account for the low identity. For example, 4 different amino acid selectivity signatures are known for A-domains specific for leucine whereas

tyrosine and cysteine specific A-domains have 3 and 2 selectivity signatures respectively (Stachelhaus *et al*., 1999). Therefore, amino acid specificity signatures are not expected to be identical to the signatures of previously characterised A-domains, and could explain the low sequence identity observed in the ornithine and leucine selective A-domains.

Amino acid positions in the A-domain specificity signature sequence exhibit variable degrees of conservation. For example, amino acids at positions 235 and 517 in the active site are strictly conserved as observed in table 3.8. Asparagine (235) and Lysine (517) are invariantly conserved and play a role in stabilising the α-amino and α-carboxylate groups of the cognate amino acid respectively (Stachelhaus *et al*., 1999). Consistent with the observation that the residues Asp235 and Lys517 are invariantly conserved, all the A-domains in Cluster 4 had conserved residues at these positions as shown in table 3.8. All other residue positions in the active site (236, 239, 278, 299, 301, 322, 330 and 331) are either moderately or highly variant (Stachelhaus *et al*., 1999) and could further account for the low sequence identities observed by the A-domains particularly in modules 2 and 5 both at 50%.

Table 3.8: Prediction of A-domain amino acid specificity using NRPSpredictor2

Module	Amino acid specificity	Identity (%)	A-domain active site amino acids									
			235	236	239	278	299	301	322	330	331	517
Module 1	Asparagine	60	D	S	V	L	I	G	E	V	W	K
Module 2	Ornithine	50	D	P	E	N	A	G	F	T	T	K
Module 3	Ornithine	70	D	A	K	N	L	G	L	S	N	K
Module 4	Threonine	100	D	F	W	N	I	G	M	V	H	K
Module 5	Leucine	50	D	Y	F	Y	M	S	Q	V	E	K

3.4.5 C-domain classification

The C-domain plays a role in peptide-bond formation between amino acids during NRP biosynthesis (Samel *et al*., 2007). Classification of C-domains is based on the stereochemistry of the amino acids between which they catalyse peptide-bond formation. C-domain subtypes include the LcL and DcL C-domains where; LcL domains catalyse the peptide-bond formation between L-amino acids and the DcL domains catalyse peptide-bond formation between a D and L-amino acid (Rausch *et al*., 2007). In addition, classification of C-domains can be based on functional subtypes and include; the heterocyclization (Cyc) and dual E/C domains. Cyc-domains, in addition to catalysing peptide-bond formation, play a role in cyclization of the peptide product (Rausch *et al*., 2007). The dual E/C domains however, are involved in both epimerization and peptide-bond formation (Balibar *et al*., 2005). Lastly, the starter C-domains are involved in the acylation of the first amino acid with a fatty acid or polyketide moiety (Miao *et al*., 2006). The Natural Product Domain Seeker (NaPDos) server was used to classify the C-domains in Cluster 4 into their respective sub-types as summarised in table 3.9. The C-domains in module 1, 3 and 4 were classified as LcL whereas the C-domain in module 2 classified as DcL.

Although the first module of an NRPS usually lacks a C-domain, the first module of Cluster 4 contained a C-domain. It is therefore possible that the C-domain in the first module of Cluster 4 belongs to the starter C-domain class. The false clustering of the C-domain in module 1 as an LcL subtype is not unlikely as starter and LcL C-domains share a close evolutionary relationship (Rausch *et al*., 2007). Interestingly, the C-domain in module 5 clustered with a recently discovered C-domain subtype, the

modified amino acid (modAA) domain. The ‘modified AA’ C-domain subtype (previously identified in bleomycin and microcystin pathways), although not experimentally characterised, is hypothesised to play a role in amino acid modification (Ziemert *et al*., 2012). For example, the dehydration of serine to dehydroalanine is a proposed modification of the modAA C-domain subtype although further investigation is required (Tillett *et al*., 2000).

In addition to functional classification, the NaPDos server is able to identify potential pathways from which the C-domain originates. C-domains in Cluster 4 were predicted to originate from three different pathways namely; microcystin (module 1, 2 and 5), actinomycin (module 3) and syringomycin (module 4). KS or C-domains with amino acid identity greater than 85% are likely to belong to pathways that produce similar or identical compounds to those of the reference pathway (Gontang *et al*., 2010). All domains showed less than 40% identity at the amino acid level (table 3.9) to their respective reference pathways. Therefore, C-domains in Cluster 4 are likely to be involved in the biosynthesis of a novel compound. Alternatively, the low identity could be as a result of limited sequence data available within the NaPDos reference database.

Table 3.9: C-domain subtypes as determined by the NaPDos server

Module	C-domain	Class	Pathway	Identity (%)
Module 1	1	LcL	Microcystin	32
Module 2	1	DcL	Microcystin	33
Module 2	1	DcL	Microcystin	35
Module 3	1	LcL	Actinomycin	23
Module 4	1	LcL	Syringomycin	32
Module 5	1	ModAA	Microcystin	39

3.4.5.1 Conserved C-domain motifs

The conserved C-domain conserved sequence motif HHxxxDGxS/C (referred to as the His-motif) is implicated in the catalytic reaction of peptide-bond formation (Bergendahl *et al*., 2002). Alignment of the C-domain amino acid sequences from Cluster 4, resulted in the identification of an intact His-motif in all C-domains with the exception of module 3 and 5 (table 3.10). The C-domain in module 3 contained a 'Phe' residue in the place of the 'Gly' and the first and second 'His' residues were replaced by 'Leu' and 'Ser' respectively. In contrast, the C-domain His-motif in module 5 consisted of a 'Cys' residue in place of the first 'His' residue.

Mutational analysis by Stachelhaus and colleagues (1998), saw the abolition of the condensation reaction between the Pro-CAT (proline specific module) and Phe-ATE (phenylalanine specific domain) modules following the exchange of the second 'His'

(HHxxxDG) with a 'Val' residue. Deviation from the conserved His-motif has been observed in a number of NRPS pathways including pAMDE1 gene cluster (Mizuno *et al*., 2013) as well as the bleomycin, tallysomycin and the zorbamycin gene clusters (Galm *et al*., 2011). The C-domains that show deviation from this motif are thought to be inactive or play alternative roles during biosynthesis (Galm *et al*., 2011).

Table 3.10: Conserved amino acid motifs in each C-domain

Module	C-domain motif								
	H	**H**	**x**	**x**	**x**	**D**	**G**	**x**	**S/C**
Module 1	H	H	L	A	I	D	G	W	S
Module 2	H	H	L	L	M	D	G	W	S
Module 2	H	H	L	L	L	D	G	W	S
Module 3	L	S	L	N	T	D	F	I	S
Module 4	H	H	I	I	F	D	G	W	S
Module 5	C	H	L	I	F	D	G	W	S

3.4.6 Conserved PCP-domain motifs

The conserved sequence motif "LGG H/D S L/I" is required for the attachment of the 4'-phosphopanteitheine co-factor that facilitates activation of the PCP-domain (Vollenbroich *et al*., 1993) was present in all PCP-domains from Cluster 4. The serine residue in this conserved motif is necessary for the attachment of the co-factor and was

conserved in all PCP-domains and is indicative of functional PCP-domains. Slight deviation from the conserved PCP-domain motif was observed in module 1, where a 'Met' residue replaced the 'Leu' residue. The core of the conserved motif (H/D S L/I) is considered as the binding site of the 4'-phosphopanteitheine co-factor and the effect of the observed substitutions (Met in place of Leucine) that lie outside of the core motif are yet unknown.

Table 3.11: Conserved 4'-phosphopantetheine binding site in PCP-domains

Module	PCP motif					
	L	**G**	**G**	**H/D**	***S***	**L/I**
Module 1 (PCP1)	M	G	G	D	S	I
Module 2 (PCP1)	L	G	G	D	S	I
Module 3 (PCP1)	L	G	G	H	S	L
Module 4 (PCP1)	L	G	G	D	S	I
Module 5 (PCP2)	L	G	G	D	S	L

3.4.7 Cluster 4 summary: a putative lipopeptide pathway

Cluster 4 consists of 4 NRPS genes with a combined total of 5 modules. Located in close proximity, are PKS/FAS associated genes situated on an individual ORFs consistent with the type II class of these enzyme families. The presence of both NRPS and PKS/FAS genes suggests a possible hybrid pathway. In addition to consisting of

both NRPS and PKS/FAS, various characteristics of Cluster 4 are consistent with lipopeptide pathways described below.

Although usually absent, the first module in Cluster 4 contains a C-domain (Lautru and Challis, 2004). This unusual C-domain has previously been proposed to play a role in the coupling of a fatty acid group to an amino acid group (Miao *et al*., 2006). Furthermore, Cluster 4 contains an AMT-domain that is known to play a role in the conversion of fatty acids into β-amino fatty acid groups (Duitman *et al*., 1999). The AMT-domain has been found in only a few pathways as mentioned previously including cyanobacterial hepatotoxins, prodigiosins and lipopeptides (iturin family). Prodigiosin production is a characteristic feature of organisms in the genus *Serratia* and are also produced by a few other organisms for example those in the order *Actinomycetales* (Khanafari *et al*., 2006; Williams, 1973). Therefore, Cluster 4 is likely to encode a lipopeptide because these compounds are frequently produced by organisms in the genus *Bacillus* (Raaijmakers *et al*., 2010). Three lipopeptide families commonly found in the genus *Bacillus* include; iturins, surfactins and fengycins (Raaijmakers *et al*., 2010). A distinguishing characteristic of the different lipopeptide families is the modification of their fatty acid moiety, where surfactins and fengycins both incorporate a β-hydroxy fatty acid whereas the iturins incorporate a β-amino fatty acid (Raaijmakers *et al*., 2010). Therefore, the presence of an AMT-domain, associated with the conversion of fatty acids into β-amino fatty acid groups likely suggests the production of an iturin type lipopeptide.

ORF44 was annotated as a transporter protein by antiSMASH2.0. A BLAST analysis of this ORF identified a hypothetical protein sharing 75% identity as the closest related

sequence originating from (*B. oceanisediminis*). This hypothetical protein contains a PvdE region that functions as an ABC-type siderophore export system. A TransportDB analysis (Ren *et al*., 2004) yielded similar results although the closest related sequence was a pyoverdin export ABC transporter from *Thermosynechococcus elongatus* BP-1 at 42% identity. This discrepancy is likely due to insufficient reference sequence data within the database. These two analyses however both implicate a siderophore type export system related to pyoverdin. Pyoverdin is a siderophore produced by *Pseudomonas aeruginosa* (Cox and Adams, 1985).

Siderophores are low molecular weight chelating compounds specific for ferric iron secreted by micro-organisms under iron-limiting conditions (Neilands, 1995). Their biosynthesis follows either an NRPS-dependent or NRPS-independent mechanism (Miethke and Marahiel, 2007). AntiSMASH2.0 is able to discriminate between NRPS-dependent/independent siderophore pathways, although the software classifies NRPS-dependent siderophores as NRPS. The presence of a transporter associated with siderophore export could therefore be suggestive of Cluster 4 being involved in the production of an iron-chelating compound. Lipopeptides have previously been identified that exhibit siderophore activity such as the serobactins (Rosconi *et al*., 2013), marinobactins and aquachelins (Martinez and Butler, 2007). Further experimentation using a siderophore specific activity assay is however necessary to confirm this hypothesis.

The second transport associated ORF; ORF56 was classified into the AcrB/AcrD/AcrF family of proteins sharing 74% sequence identity (antiSMASH2.0). This family of proteins among other substrates is associated with acriflavine resistance (Visalli *et al*.,

2003). A further TransportDB analysis of this ORF shows a close relationship to the acriflavine resistance protein corresponding to the previous analysis. The AcrB and AcrD homologs are also involved in the efflux of hydrophobic inhibitors such as fatty acids and bile salts in *E. coli* (Ma *et al*., 1995). Therefore, because the ORF56 transporter in Cluster 4 shows homology towards AcrA/AcrB, it is likely that the compound encoded by Cluster 4 is hydrophobic in nature. This is consistent with the hypothesis that a lipopeptide type of compound is produced by Cluster 4.

3.4.8 Gene cluster comparison

BLASTn analysis consistently indicated that the Cluster 4 NRPS genes shared highest sequence identity to NRPS genes from *B. oceanisediminis* 2691 (table 3.7). The *B. oceanisediminis* 2691 genome, publicly available, was subsequently submitted to antiSMASH2.0 to determine whether a pathway similar to Cluster 4 was present. This analysis led to the identification of a NRPS pathway in *B. oceanisediminis* 2691 exhibiting a similar genetic organisation and composition to Cluster 4 although several differences were evident (figure 3.12).

The absence of the PKS/FAS genes in *B. oceanisediminis* 2691 was a notable difference because the resulting product possibly lacks the hybrid properties as seen in Cluster 4. Both gene clusters contain a *padR* and a*snC* transcriptional regulator genes, however, an additional *padR* transcriptional regulator is present in *B. oceanisediminis* 2691. This suggests that although these two clusters possibly utilise similar regulatory mechanisms, they could require different activation triggers. Additional genes

associated with biosynthesis present in *B. oceanisediminis* 2691 but absent from Cluster 4 include; methyltransferase, isochorismatase, oxidoreductase, 2 additional ABC transporters and an ABC-2 type transporter. The presence of tailoring enzymes such as methyltransferase, isochorismatase and oxidoreductase are indicative of different product modifications that provide further support to the notion that two different products are encoded by the NRPS pathway in *B. oceanisediminis* 2691.

3.4.8.1 Comparison of A-domain specificities

Comparison of the A-domain specificity between the two pathways is summarised in (table 3.12). Several differences in A-domain specificity were evident. For example, module 1 in Cluster 4 is specific for asparagine whereas the A-domain in ORF266 (corresponding to module 1 in the *B. oceanisediminis* 2691) was specific for histidine. The fifth module in Cluster 4 and *B. oceanisediminis* 2691 is specific for leucine and isoleucine respectively and all other modules share the same substrate specificities. This provides further evidence that these two gene clusters unlikely encode the same compound.

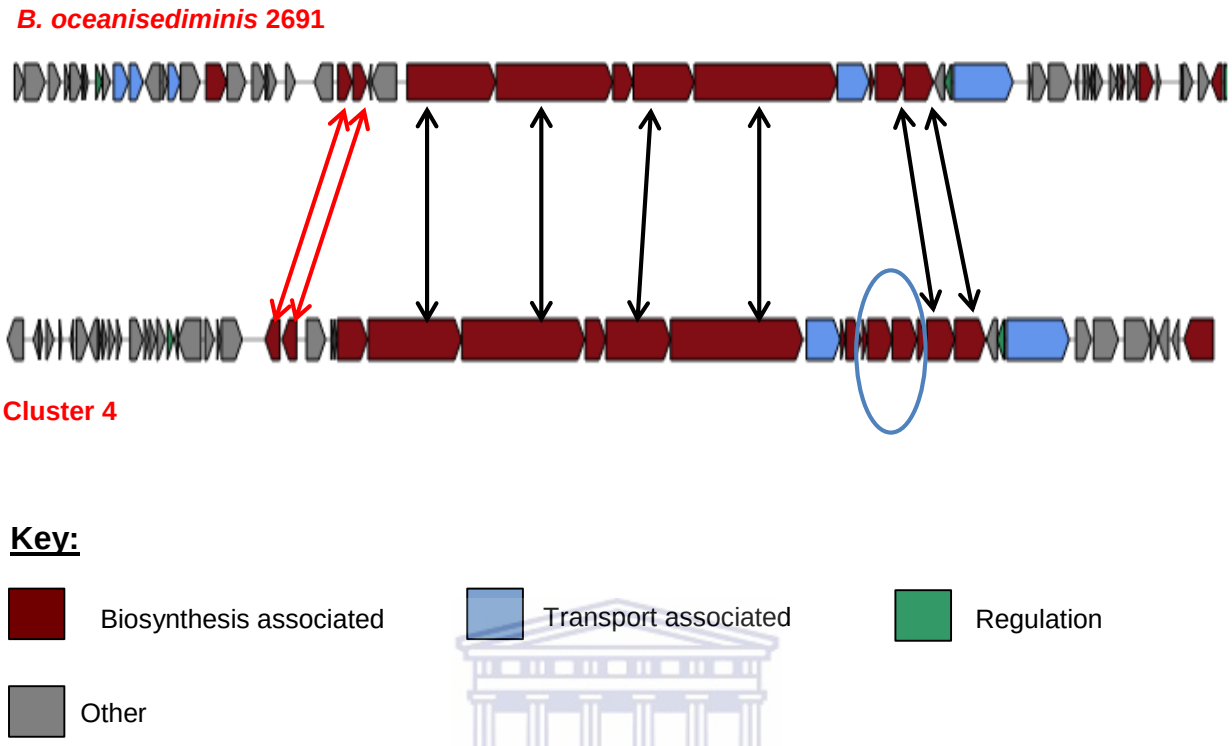

Figure 3.12: Comparison of Cluster 4 and an NRPS gene cluster identified in *B. oceanisediminis* 2691. Genes present in both clusters are highlighted with an arrow. One major difference is the absence of the PKS/FAS genes from *B. oceanisediminis* 2691. Secondly, the ORFs predicted to encode the thioesterase and phosphopantetheinyl transferase are oriented in opposite directions between the two clusters and are highlighted with red arrows.

Table 3.12: Comparison of A-domain substrate specificities

Cluster 4				*B. oceanisediminis* 2691			
Module	**Amino acid substrate**	**Identity (%)**	**Active site signature**	**ORF**	**Amino acid substrate**	**Identity (%)**	**Active site signature**
Module 1	Asparagine	60	DSVLIGEVWK	ORF266	Histidine	70	DSVLIAEVWK
Module 2	Ornithine	50	DPENAGFTTK	ORF263	Ornithine	50	DPENAGFTTK
Module 3	Ornithine	70	DAKNLGLSNK	ORF261	Ornithine	70	DAKNLGLSNK
Module 4	Threonine	100	DFWNIGMVHK	ORF259 (1)	Threonine	100	DFWNIGMVHK
Module 5	Leucine	50	DYFYMSQVEK	ORF259 (2)	Isoleucine	50	DYFFMSQVEK

3.4.9 Expression analysis of Cluster 4

As discussed above, Cluster 4 shows characteristics typical of a hybrid gene cluster due to the presence of both NRPS and PKS/FAS genes in close proximity. This section focuses on the experimental approach taken to determine whether Cluster 4 is expressed in any of the culture conditions under which biological activity was detected. Of particular interest was determining the expression of Cluster 4 in Zobell medium as extracts from this medium exhibited anti-fungal activity. Cluster 4 is hypothesised to encode a lipopeptide, a group compounds that typically exhibit anti-fungal activity (Pathak and Keharia, 2014). Therefore, if Cluster 4 is expressed in Zobell medium this could provide strong evidence for the production of a lipopeptide. Specific primers were subsequently designed for ORF39 (NRPS) and ORF48 (KS) as representative genes of Cluster 4 (figure 3.13).

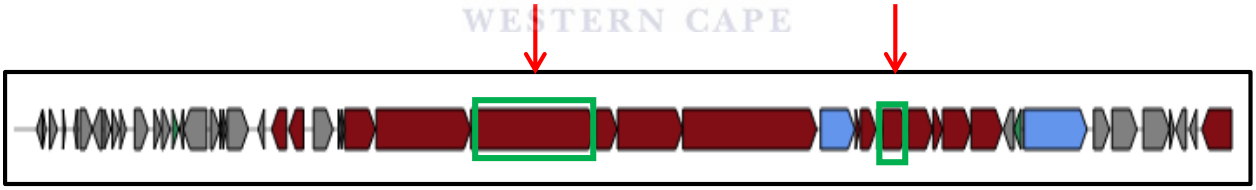

Figure 3.13: Specific primers were designed to ORF39 (NRPS) and ORF48 (FAS/PKS) that are demarcated in green above.

Total RNA was extracted from PE8-15 cultures grown to mid-logarithmic and stationary phase in ACM and Zobell media. Although SMs are frequently produced during the stationary phase of microbial growth (Sánchez *et al*., 2010), exceptions have been observed (Nieselt *et al*., 2010). Therefore, in order to avoid bias, expression of

representative genes from Cluster 4 was investigated in both logarithmic and stationary phase cultures of PE8-15. Following amplification of cDNA, ORF39 and ORF48 genes were detected in Zobell medium during mid-logarithmic growth stage (figure 3.14). This is contrast to ORF39 and ORF48 that were not detected in ACM medium (figure 3.15). The observation that Cluster 4, a putative iturin lipopeptide pathway, is expressed during the mid-logarithmic stage is surprising because the iturin lipopeptides are known to be expressed during late stationary phase (Raaijmakers *et al*., 2010). However, depending on the growth conditions, changes in expression patterns may be observed.

These results unfortunately are not sufficient to associate the anti-fungal activity with Cluster 4. This is because the anti-fungal activity of PE8-15 was detected in a 2 week old culture grown in McCartney vials whereas transcription of Cluster 4 genes was detected in mid-log cultures grown in a conical flask. The difference in culture vessels used could affect the growth kinetics and subsequently the expression of SM pathways. Therefore, subsequent experimentation is required to screen for anti-fungal activity in both mid-log and stationary-phase cultures and provide more concrete correlation between Cluster 4 and the observed biological activity.

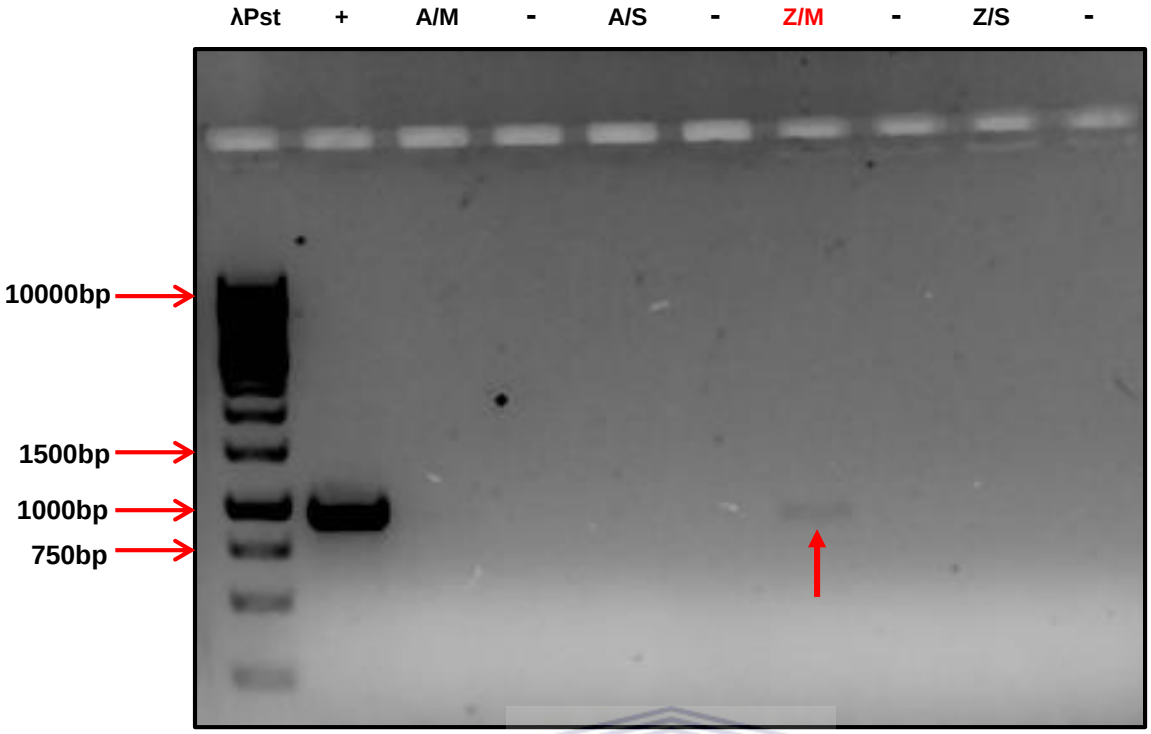

Figure 3.14: Reverse transcription PCR to detect the expression of ORF39 (NRPS) in isolate PE8-15. Lane 1: 1kb DNA ladder. Lane 2: Positive control (PE8-15 genomic DNA). Lane 3: Amplification of cDNA prepared from an ACM mid-log culture (A/M). Lane 5: Amplification of cDNA prepared from an ACM stationary-phase culture (A/S). Lane 7: Amplification of cDNA prepared from a Zobell mid-log culture (Z/M). Lane 9: Amplification of cDNA prepared from a Zobell stationary culture (Z/S). Lane 4, 6, 8, 10: Represent negative controls for each reaction. Expected PCR product size is 950b. Amplicons were resolved in a 1% [w/v] TAE agarose gel.

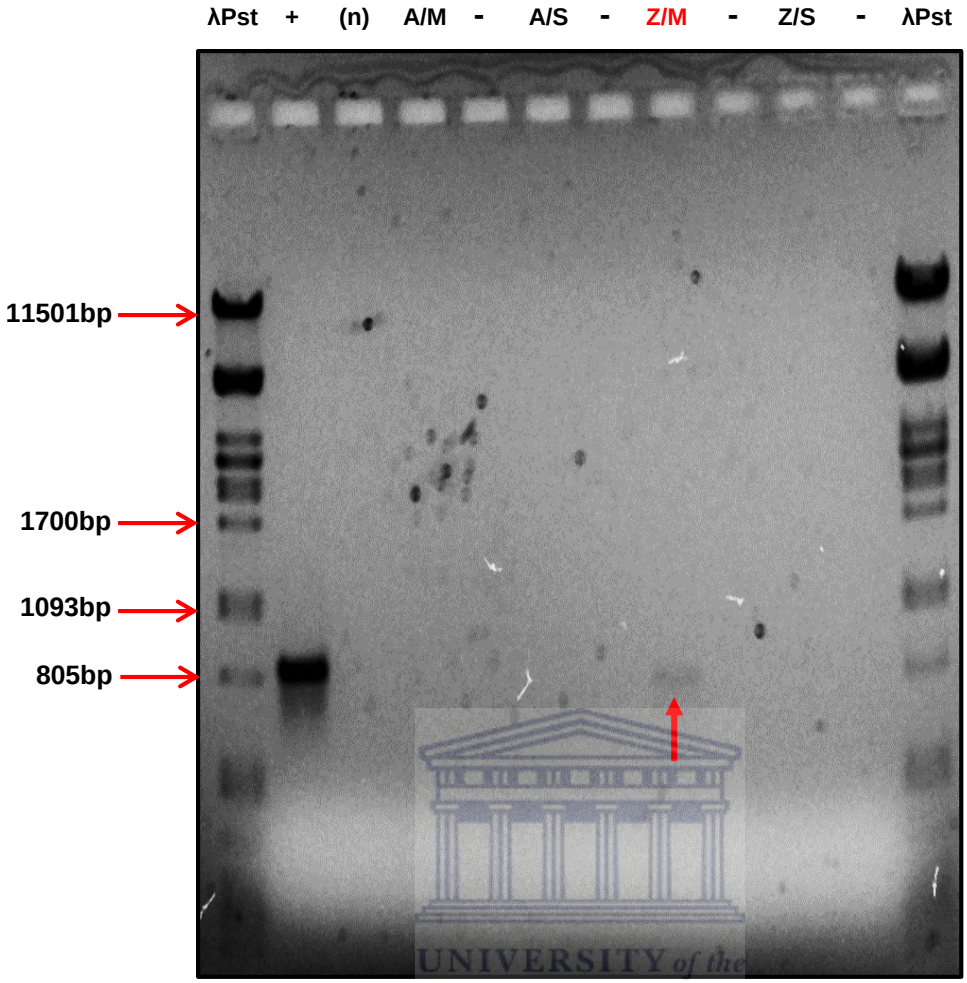

Figure 3.15: Reverse transcription PCR to detect the expression of ORF48 (PKS/FAS gene). Lane 1 and 12: DNA molecular marker (lambda DNA digested with *Pst*I). Lane 2: Positive control (PE8-15 genomic DNA). Lane 3: Negative control (water as a template). Lane 4: Amplification of cDNA prepared from an ACM mid-log culture. Lane 6: Amplification of cDNA prepared from an ACM stationary phase culture. Lane 8: Amplification of cDNA prepared from a Zobell mid-log culture. Lane 10: Amplification of cDNA prepared from a Zobell stationary culture. Lane 5, 7, 9, 11: Represent negative controls for each reaction. Expected product size is 856bp. Amplicons were resolved in a 1% [w/v] TAE agarose gel.

3.5 Conclusion

We report here a potentially novel hybrid NRPS-PKS/FAS gene cluster hypothesised to be involved in the biosynthesis of a lipopeptide. In addition, 10 additional SM pathways were identified by genome mining and represent an untapped reservoir of potentially novel bioactive compounds. We demonstrate the significance of sequence-based methods (PCR and WGS) in NP discovery as they are able to identify the full complement of an organisms SM repertoire in a more efficient and cost-effective manner circumventing costly and time-consuming analytical methods such as NMR and MS. Furthermore, although well-known for their ability to produce an array of SMs, these results show that the genus *Bacillus* is still a promising source of novel chemistry in NP discovery.

Chapter 4 General discussion and final conclusion

4.1 General discussion and final conclusion

Drug-resistance among clinically relevant micro-organisms is currently a major health crisis globally (Walsh and Amyes, 2004). Drug-resistant pathogens have serious negative implications on patient health care. For example, infections caused by drug-resistant pathogens result in extended hospital stays (increased medical costs) and are associated with higher mortality rates among patients in comparison to drug susceptible infections (Cosgrove, 2006). Furthermore, treatment options for drug-resistant infections are significantly limited as the available drugs are rendered ineffective (Falagas and Bliziotis, 2007). The lack of treatment options and a drastic decline in the discovery of novel drugs, underpins the urgent and continuous need for new drug candidates.

NPs are compounds produced by living organisms including bacteria, fungi and plants (Ji *et al*., 2009) usually as a result of secondary metabolism (Williams *et al*., 1989). They are important for drug discovery as they exhibit diverse biological activities that include anti-microbial, anti-cancer and anti-inflammatory properties. The “Golden era of antibiotics”, a period when the majority of antibiotics in use today were discovered, highlights the importance of micro-organisms as producers of bioactive NPs relevant to drug discovery (Lewis, 2012). The drugs discovered during this period were predominantly isolated from soil-dwelling bacteria and fungi. Unfortunately, discovery of novel compounds in these environments diminished over time (Daniel, 2004), and this has led to increased interest in alternative sources such as marine-dwelling organisms.

Invertebrates such as sponges, tunicates, soft-corals and bryozoans constitute the largest portion of the marine animal diversity (Leal *et al*., 2012). These animals are known for their production of bioactive NPs and hence important for drug discovery

(Perdicaris *et al*., 2013; Ankisetty *et al*., 2013) However, a major bottleneck in marine drug development is that these compounds often occur at very low concentrations within the invertebrate tissue, thus requiring large amounts (tonnes) of invertebrate biomass to obtain sufficient bioproduct (Proksch *et al*., 2002; Proksch *et al*., 2003). Therefore, marine invertebrates are not a sustainable and environmentally friendly source of NPs.

Marine invertebrates often form symbiotic relationships with a diverse consortium of micro-organisms. Interestingly, NPs originally isolated from marine invertebrates, show high similarity to those exclusively of microbial origin (Piel *et al*., 2004). This subsequently led to the hypothesis that marine invertebrate-associated bacteria are the true producers of these compounds. Experimental evidence for this hypothesis was presented by Wilson and colleagues (2014) described in section 1.5.3. In comparison to invertebrates, micro-organisms represent an alternative and most importantly a sustainable source of marine NPs.

The current study was aimed at identifying novel NPs from micro-organisms associated with novel marine invertebrate species endemic to the South African coast. The methodology consisted of two components; a culture-dependent and a culture-independent strategy.

The culture-based approach involved the isolation of bacteria associated with the invertebrate specimens (section 2.4.2) using conventional microbiology methods. A wide range of media were used to access as much of the microbial and chemical diversity as possible. Furthermore, the bacterial isolates were screened for anti-

microbial activity against a panel of indicator strains including the multi-drug resistant *E. coli* 1699 strain. The significance of inhibitory activity against this strain is that because it is resistant to the majority of standard antibiotics (Appendix F), the compounds produced by the bacteria likely have a novel mechanism of action (Hentschel *et al*., 2001). This study addresses two major bottlenecks in drug discovery (section 1.6); (i) NP discovery from a sustainable source (ii) NP discovery from underexplored environments less likely to result in the re-discovery of known compounds.

The primary screening process involved screening isolates on the medium on which they were isolated. This was a major limitation in this study as laboratory conditions are often not conducive for SM expression. This theory stems from the analysis of microbial genome data that has revealed the presence of numerous silent SM pathways in even well-studied micro-organisms (Aigle *et al*., 2014; Ikeda *et al*., 2003). Silent SM pathways encode potentially novel compounds and are therefore of great importance to NP discovery. OSMAC (section 1.8.2) is aimed at stimulating the expression of these silent SM pathways and formed an integral part of the secondary screening process. Among the *Bacillus* bioactive isolates, the number of strains active against the multi-drug resistant *E. coli* strain doubled from 3 to 6 strains following the use of OSMAC (section 3.2.3). However, this only represented a very small number of different combinations which could have been tested and there is much wider range of media and treatments that could be used to maximise SM production (Bertrand *et al*., 2013; Doull *et al*., 1994; Rateb *et al*., 2011). Therefore, these findings indicate that laboratory conditions may indeed be sub-optimal for microbial SM production. As shown in this study, OSMAC is necessary to improve discovery of bioactive compounds from micro-organisms.

The second strategy employed was a culture-independent approach that involved screening of bioactive bacteria for SM genes by PCR and WGS of a suitable strain. PCR played an important role in the de-replication and identification of bacterial isolates likely to produce novel SMs.

PCR as a de-replication tool works on the principle of the degree of sequence similarity that SM genes exhibit (usually conserved domains such as the A-domain or KS) towards characterised pathways (Gontang *et al*., 2010). A high sequence similarity to known pathways usually suggests that the pathway from which the query sequence originates is involved in the production of a similar compound. In this study, the need for de-replication arose from the identification of isolates that shared very high 16S rRNA sequence similarity (>98%) to reference sequences in the database. Furthermore, the majority of the bioactive isolates (*Bacillus*) shared the same closest relative according to the BLAST analysis yet displayed different anti-microbial profiles. An NRPS amplicon amplified from the *Bacillus* strain PE8-15 exhibited low sequence similarity (65% and 50% respectively) to its reference sequences in the database. Based on the principle of PCR de-replication, this revealed a potentially novel NRPS pathway within PE8-15. The rest of the NRPS amplicons derived from the other *Bacillus* isolates showed a high sequence similarity (>90%) to their reference sequences. It is not conclusive evidence that NRPS amplicons with a high sequence similarity to their reference sequences originate from previously characterised pathways. Nonetheless, the pathways of origin are publicly available and could be under investigation.

Traditionally, de-replication, involves the rapid identification of known compounds present in a mixture by chromatography (HPLC), MS and NMR analyses. These

techniques however, require significant amounts of the compound being analysed and are therefore severely limited by the levels of expression of a compound by the organism (Lin *et al*., 2008). Although recent advancements have improved the sensitivity of analytical equipment, for example the use of microcoil and microtube probes for NMR (Dalisay and Molinski, 2009) these tools may not be readily available to research groups. In addition, most analytical methods (NMR, LC-MS and UV spectra) are not high-throughput and thus make the initial de-replication process rather time and labour intensive (Wong *et al*., 2012). Lastly, these analytical methods de-replicate compounds based on patterns (physicochemical properties) and do not account for novel biological properties (Wong *et al*., 2012). This could result in compounds belonging to well characterised classes but still exhibiting novel mechanisms of action being undetected (Wong *et al*., 2012).

In comparison, PCR offers various advantages as a de-replication tool over the use of more conventional chemical analysis; (i) it facilitates selective screening for NPs genes (ii) detects SM associated genes independent of their expression and therefore not limited by repressed pathways (silent pathways) (iii) does not require the need for costly and time-consuming extract preparation and chemical analysis for the initial de-replication process (iv) it de-replicates isolates based on the SM genes. Therefore, bacterial strains are selected based on their potential to produce novel SMs as opposed to their phylogenetic affiliation. Furthermore, PCR facilitates the prioritisation of strains for further investigation thus likely to reduce the number of strains to be analysed (Liu *et al*., 2012).

Exhibiting the most diverse anti-microbial profile and containing a potentially novel NRPS pathway, the genome of isolate PE8-15 was sequenced (WGS). Genome mining identified a potentially novel NRPS pathway. Informatics-based characterisation of Cluster 4 predicted the possible production of a lipopeptide compound (chapter 3).

The discovery of a potentially novel lipopeptide is significant as these compounds have numerous industrial applications that extend beyond therapeutic use. For example, in the cosmetic industry, lipopeptides are used as dermatological products due to their low skin irritation properties (Mandal *et al*., 2013). In the food industry, lipopeptides are used as emulsifiers (Meena and Kanwa, 2015). Lipopeptides have also been shown to have important application in the agriculture sector due their antagonistic properties against phytopathogens (Cawoy *et al*., 2015). This group of compounds can also act as biosurfactants (microbial-derived surfactants) and are an environmentally friendly alternative for environmental clean-ups compared to their synthetic counterparts (Varadavenkatesan and Murty, 2013).

This study demonstrates that WGS is a superior tool in the discovery of SM pathways compared to alternative methods such as DNA library screening. DNA library preparation requires the use of vectors into which DNA fragments are cloned. Currently available vectors such as fosmids, although suitable for longer inserts, are limited to an insert size of less than 40kb (Williams *et al*., 2012). However, complete SM pathways can reach up to lengths upwards of 90kb in size (Ichikawa *et al*., 2013). In the context of the current study, fosmid vectors would not be appropriate as Cluster 4 spans a total length of approximately 59kb. Alternative vector systems such as the bacterial artificial chromosomes (BAC) are able to carry inserts up to 300kb in size (Shizuya *et al*., 1992).

However, BAC vectors require the need to trim unwanted sequences from the inserts and because the DNA is randomly fragmented, genes from the same cluster are frequently situated on different clones (Kouprina and Larionov, 2006). Furthermore, in terms of preparation time and labour intensity, WGS uses next generation sequencing platforms that are fast and negate the need for tedious and time consuming DNA library preparation.

As previously discussed, SMs (or NPs) are often produced at low concentrations by the host organism. Improving compound titres can be achieved by large-scale fermentation of the native producer (Ongley *et al*., 2013). However, the native host may exhibit poor growth properties inappropriate for large-scale fermentation (Lopanik, 2014). Alternatively, large-scale fermentation can be achieved by heterologous expression of the genes of interest in a non-native host (Chen *et al*., 2010). This requires the cloning of the complete gene cluster into a suitable vector and subsequent transfer to a non-native host. We propose transformation-associated recombination (TAR) as a suitable approach to completely clone Cluster 4. A major advantage of TAR cloning is that it facilitates selective isolation of the complete pathway of interest from the genome and is not limited by the size of the insert DNA (Noskov *et al*., 2003). This system has been successfully used to clone the complete taromycin A gene cluster (Yamanaka *et al*., 2014).

A major advantage of heterologous expression (as opposed to large-scale fermentation of the native host), is that well-studied expression hosts are readily available and may not require extensive optimisation for large-scale fermentation. Furthermore, environmental isolates of interest are usually not fully characterised and detailed

information for example on their potential pathogenicity is often lacking. Therefore, the use of a well-studied heterologous host is an advantage due to the strict safety regulations in place for the industrial production of pharmaceuticals (Zobel *et al*., 2015). For example, an appropriate host for the expression of Cluster 4 (NRPS pathway predicted to encode a lipopeptide) would be *B. subtilis* that has previously been used as a heterologous host for NRPS pathways (Eppelmann *et al*., 2001).

Future work would entail the expression of Cluster 4 by heterologous expression using the previously described TAR cloning system. Once successfully expressed, the purified product encoded by Cluster 4 will be screened for a range of biological activities with clinical relevance. Most importantly however, structural-elucidation of the compound biosynthesised by Cluster 4 will be performed and analysed for novelty. Furthermore, gene-knockout studies may be necessary to confirm which genes within Cluster 4 that are actively involved in biosynthesis. In addition to complete characterisation of Cluster 4, it would be important to investigate the other SM pathways in PE8-15 for novel compounds/bioactivity. Furthermore, it would be worthwhile to re-screen isolates that previously did not exhibit any bioactivity. This will however involve the use of OSMAC as this was not incorporated into the initial screening process. Lastly, the non-bioactive strains will be screened by PCR using a range of primers targeting different classes of SM pathways.

The current study highlights the importance of the South African coastline and endemic marine invertebrate species as a source of bioactive micro-organisms potentially producing clinically relevant biomolecules. This is in contrast to previous studies where the marine invertebrates were the sole focus for marine bio prospecting (section 1.9).

Currently, there is only one established academic drug discovery and development platform in Africa, the H3-D Drug Discovery & Discovery Centre at the University of Cape Town (South Africa) (H3-D, 2015). A major focus of this organisation is to link basic and clinical studies. These initiatives are particularly important for developing communities such as those in Sub-Saharan Africa where the increase of drug-resistant pathogens (e.g. the multi-drug resistant TB strain), high prevalence of HIV/AIDS, wide-spread poverty and high medical costs are major challenges in the healthcare sector. Therefore, as micro-organisms are considered a sustainable source of NPs, this readily accessible resource in South Africa should be further exploited for the expansion of the drug discovery platform in South Africa.

Appendices

Appendix A

Buffers and stock solutions

Buffers and stock solutions used in this study are listed below. Solutions marked with an asterisk (*****) were autoclaved at 121°C for 20 minutes at 15 psi. Alternatively, solutions/buffers not amenable to autoclaving were sterilized using 0.22 µm filters.

Table A.1: Buffers and stock solutions used in this study

Buffers and stock solutions	
50 X TAE buffer (*)	**NaCl solution (5 M) (*)**
242 g Tris base	146.1 g
18.6 g EDTA	500 ml of sterile dH_2O
57.1 ml glacial acetic acid	
sterile dH_2O to a final volume of 1 L	**NaOH/SDS (solution 2)***
	0.2 M NaOH
Ampicillin stock	1% [w/v] SDS
100 mg/ml ampicillin in sterile dH_2O	
Stored at -20° C	**Potassium acetate pH 4.8 (solution 3)**
	29.5 ml glacial acetic acid
Ammonium acetate (4 M)	KOH pellets to pH 4.8
15.42 g	100 ml of sterile dH_2O
50 ml of sterile dH_2O	
	TE buffer pH 8 (10 mM Tris-HCl; 1 mM) (*)
Breaking buffer	10 mM Tris-HCL
2% Triton X100	1 mM EDTA
1% [w/v] SDS	
100 mM NaCl	**Wolfe's mineral solution**
10 mM Tris-HCl pH 8	Nitrilotriacetic acid – 1.5 g
1 mM EDTA	$MgSO_4$-$7H_2O$ – 3 g

CTAB/NaCl solution (*)

10% [w/v] CTAB

0.7 M NaCl

DEPC treated water (*)

0.1% diethyl pyrocarbonate

Final volume of 1L dH2O

DNA loading buffer

20% [v/v] glycerol

0.1 M EDTA

1% [w/v] SDS

0.25% bromophenol blue

Ethidium bromide staining solution

200 mg ethidium bromide powder

20 ml of sterile dH_2O

GTE (solution 1) (*)

50 mM glucose

25 mM Tris-HCl

10 mM EDTA

Lysis buffer

25 mM Tris-HCl pH 8

50 mM glucose

10 mM EDTA

25 mg/ml lysozyme powder

$MnSO_4$-H_2O – 0.5 g

NaCl – 1 g

$FeSO_4$-$7H_2O$ – 0.1 g

$CaCl_2$ – 0.1 g

$CoCl_2$ – $6H_2O$ – 0.1 g

$ZnSO_4$-$7H_2O$ – 0.1 g

$CuSO_4$-$5H_2O$ – 0.01 g

$AIK(SO)_4$-$12H_2O$ – 0.01 g

Boric acid – 0.01 g

Wolfe's vitamin solution

Pyridoxine hydrochloride – 10 mg

Thiamine-HCl - 5 mg

Riboflavin – 5 mg

Nicotinic acid – 5 mg

Calcium D-(+)-pantothenate – 5 mg

p-Aminobenzoic acid – 5 mg

Thioctic acid – 5 mg

Biotin – 2 mg

Folic acid – 2 mg

Vitamin B12 – 0.1 mg

dH_2O – 1 L

X-gal

0.6 g x-gal

30 ml dimethylformamide

Stored at -20° C

Appendix B

Isolation and growth media used in this study

Culture media used in this study are listed in table B.1 and table C.1 and unless stated otherwise, were autoclaved at 121°C for 20 minutes (15 psi). Agar was added at a concentration of 15 g/L for the preparation of solid media and pH adjusted using 32% HCl and1 or 10 M NaOH.

Table B.1: List of culture media used for the of isolation marine invertebrate associated bacteria

Media	Constituents	Quantity/L of dH_20
	HEPES	2.38 g
	Sodium pyruvate	3 g
	Yeast extract	0.1 g
	Soybean peptone	3 g
	$NaNO_3$	0.34 g
	K_2HPO_4	0.1 g
	$MgSO_4$* 7 H_2O	0.15 g
Activated charcoal Media (pH 7.0)	Activated charcoal	3 g
	Mannitol	15 g
	K_2HPO4	0.3 g
	$MgSO_4$* 7 H_2O	0.3 g
	10% Sodium molybdate solution	0.1 g
	10% $FeCl_3$	0.05 g
	NaCl	18 g

	$MgCl_2$	2 g
	KCl	0.525 g
Ashby's nitrogen free media (pH 7.5)	$CaCl_2$	0.3 g

	Solution 1 - $NaNO_3$ (15 g/L)	100 ml
	Solution 2 - K_2HPO_4 (2 g/L)	10 ml
	Solution 3 - $MgSO_4$*7 H_2O (3.75 g/L)	10 ml
	Solution 4 - $CaCl_2$*2 H_2O (1.8 g/L)	10 ml
	Solution 5 - Citric acid (0.3 g/L)	10 ml
	Solution 6 - Ammonium ferric citrate (0.3 g/L)	10 ml
	Solution 7 - Na_2EDTA (0.05 g/L)	10 ml
	Solution 8 - Na_2CO_3 (1 g/L)	10 ml
	Wolfe's mineral solution	1 ml
	NaCl	18 g
	$MgCl_2$	2 g
BG 11 (pH 7.5)	KCl	0.525 g

	D-Glucose	4 g
	Yeast extract	4 g
	Malt extract	10 g
	$CaCO_3$	2 g
	NaCl	24 g
	$MgCl_2$	5.3 g
	KCl	0.7 g
Glucose-Yeast-Malt (GYM)	$CaCl_2$	0.1 g

	Glucose	10 g
	Yeast extract	5 g
	Starch	10 g
	Tryptone	5 g
	$MgSO_4$* 7 H_2O	2 g
	$CaSO_4$ * 2 H_2O	2 g
	NaCl	18 g
	$MgCl_2$	2 g
	KCl	0.525 g

Modified 172F (pH 7.5)	$CaCl_2$	0.075 g
Oatmeal (ISP Media 3)	Jungle oats easy (no sugar)	40 g
	Glucose	1 g
	$(NH_4)_2SO_4$	0.25 g
	Peptone	0.15 g
	Yeast extract	0.15 g
	NaCl	18 g
	$MgCl_2$	2 g
	KCl	0.525 g
Planctomycetes medium	$CaCl_2$	0.075 g
	KNO_3	0.2 g
	$(NH_4)_2 HPO_4$	0.02 g
	$MgSO_4$* 7 H_2O	0.01 g
	$CaCl_2$ * 2 H_2O	0.005 g
	$FeCl_2$	0.5 g
	NaCl	18 g
	$MgCl_2$	2 g
	KCl	0.525 g
Pringsheim's Cyanobacteria specific media	$CaCl_2$	0.075 g
	Casein acid hydrolysate	0.5 g
	Dextrose	0.5 g
	K_2HPO_4	0.3 g
	$MgSO_4$ anhydrous	0.024 g
	Proteose peptone	0.5 g
	Sodium pyruvate	0.3 g
	Starch soluble	0.5 g
Reasoners 2 agar (pH 7.2)	Yeast extract	0.5 g
	dH_2O	250 ml
	Sea water	750 ml
	Wolfe's mineral solution	5 ml
Seawater agar	Wolfe's vitamin solution	10 ml

	Peptone	5 g
	Beef extract	3 g
	NaCl	18 g
	$MgCl_2$	2 g
	KCl	0.525 g
Shivji's nutrient agar (pH 7.5)	$CaCl_2$	0.075 g
	Sponge extract	10 g
	NaCl	18 g
	$MgCl_2$	2 g
	KCl	0.525 g
Sponge extract media	$CaCl_2$	0.075 g
	TSB	3 g
	NaCl	18 g
	$MgCl_2$	2 g
	KCl	0.525 g
Tryptocase soya agar (pH7.5)	$CaCl_2$	0.075 g
	TSB	3 g
	NaCl	18 g
	$MgCl_2$	2 g
	KCl	0.525 g
Tryptocase soya agar (pH4.5)	$CaCl_2$	0.075 g
	Yeast extract	1.25 g
	Peptone	3.75 g
	NaCl	18 g
	$MgCl_2$	2 g
	KCl	0.525 g
Zobell 1/4 strength (ZBA)	$CaCl_2$	0.075 g
	Yeast extract	1.25 g
	Peptone	3.75 g
	NaCl	18 g
	$MgCl_2$	2 g

	KCl	0.525 g
Zobell + 0.0001% crystal violet	$CaCl_2$	0.075 g

Appendix C

Table C.1: General purpose culture media used in this study

Media	Constituents	Quantity/L of dH_20
Lysogeny broth (LB)		
	Tryptone	10 g
	Yeast	5 g
	NaCl	10 g
Sloppy agar (semi-solid LB)		
	Tryptone	10 g
	Yeast extract	5 g
	NaCl	10 g
	Agar	7.5 g

Appendix D

Table D.1: Nextera sequencing adapter sequences

Adapter	Sequence
N504 internal	AATGATACGGCGACCACCGAGATCTACACAGAGTAGATCGTCGGCAGCGTC
N504 end	AATGATACGGCGACCACCGAGATCTACACAGAGTAGATCGTCGGCAGCGTC
N706 internal	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGG
N706 end	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGG
Transposase	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
Transposase	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG

Appendix E

Table E.1: Accession numbers for *Bacillus* strains used for the *in silico* analysis of secondary metabolite distribution

B. anthracis	*B. thuringiensis*	*B. cereus*	*B. subtilis*	*B. weihenstephanensis*	*B. mycoides*
NC_003997.1	NC_005957.1	NC_004722.1	NC_000964.3	NC_010184.1	NZ_CP007626.1
NC_005945.1	NC_020376.1	NC_012472.1	NC_020507.1		
NC_012659.1	NC_014171.1	NC_011658.1	NC_014976.1		
NC_012581.1	NC_01877.1	NC_011773.1	NC_018520.1		
NC_007530.2	NC_018500.1	NC_003909.8	NC_020244.1		
NC_017729.1	NC_018508.1	NC_011725	NC_017196.1		
	NC_008600.1	NC_006274.1	NC_016047.1		
	NC_020238.1	NC_016779.1	NC_014479.1		
	NC_017200.1	NC_018491.1	NC_020832.1		
	NC_017208.1	NC_011772.1	NC_019896.1		
		NC_016771.1	NC_017195.1:		
		NC_011969.1			
		NC_014335.1			

Appendix F

Table F.1: List of antibiotics to which the multi-drug resistant *E. coli* 1699 exhibits resistance

Compound	MIC (µg/mL)	Target/MOA	Antibiotic classification
A54145CB-181234	> 512	membrane	
Calcimycin (A23187)	64	membrane	ionophore
Daptomycin	> 512	membrane	lipopeptide
Gramicidin	128	membrane	polypeptide
Polymyxin B	1	membrane	polypeptide (cationic)
Ampicillin	> 256	cell wall	aminopenicillin
Aztreonam	≤ 0.03	cell wall	monocyclic beta-lactam
Cephalosporin C	64	cell wall	cephalosporin
Penicillin G	> 256	cell wall	beta-lactam
Ristocetin	> 512	cell wall	aminoglycoside
Teicoplanin	> 512	cell wall	glycopeptide
Vancomycin	512	cell wall	glycopeptide
Aclacinomycin A	> 512	DNA interaction	anthracycline
Actinomycin A	> 256	DNA interaction	polypeptide (toxic)
Actinomycin D	256	DNA interaction	polypeptide (toxic)
Bleomycin A2	> 64	DNA interaction	glycopeptide
Coumermycin A1	64	DNA interaction	aminocoumarin
Daunorubicin	> 256	DNA interaction	anthracycline
Gliotoxin	32	DNA interaction	mycotoxin
Mitomycin C	1	DNA interaction	aziridine-containing
Streptonigrin	2	DNA interaction	quinone-containing
Streptozotocin	> 64	DNA interaction	glucosamine
Chromomycin A3	> 512	gyrase	glycoside
Nalidixic Acid	> 256	gyrase	naphthyridone
Novobicin	> 256	gyrase	aminocoumarin

Apramycin	64	protein synthesis	aminoglycoside
Dibekacin	> 256	protein synthesis	aminoglycoside
Gentamycin	128	protein synthesis	aminoglycoside
Kanamycin	> 256	protein synthesis	aminoglycoside
Kasugamycin	> 64	protein synthesis	aminoglycoside
Neomycin	16	protein synthesis	aminoglycoside
Netilmycin	128	protein synthesis	aminoglycoside
Streptomycin	> 256	protein synthesis	aminoglycoside
Streptothricin	256	protein synthesis	aminoglycoside
Tobramycin	> 256	protein synthesis	aminoglycoside
Puromycin	> 64	protein synthesis	aminonucleoside
Spectinomycin	> 256	protein synthesis	aminocyclitol
Chloramphenicol		protein synthesis	acetamide
Chlortetracycline		protein synthesis	tetracyclide
Erythromycin	64	protein synthesis	macrolide
Lincomycin	512	protein synthesis	lincosamide
Spiramycin	256	protein synthesis	macrolide
Tetracycline	> 256	protein synthesis	tetracycline
Thiostrepton	> 512	protein synthesis	oligopeptide
Tylosin	512	protein synthesis	macrolide
Virginamycin	64	protein synthesis	streptogramin
Rifampin	> 256	RNA polymerase	rifamycin
Rifamycin SV	> 64	RNA polymerase	rifamycin
Albomycin	high	iron metabolism	cyclic polypeptide
Trimethoprim	> 400	DHFR	diaminopyrimidine

References

Abdelmohsen, U.R., Yang, C., Horn, H., Hajjar, D., Ravasi, T. & Hentschel, U. (2014) 'Actinomycetes from Red Sea sponges: sources for chemical and phylogenetic diversity', *Marine Drugs,* vol.12, no.5, pp.2771-2789.

Aigle, B., Lautru, S., Spiteller, D., Dickschat, J.S., Challis, G.L., Leblond, P. & Pernodet, J. (2014) 'Genome mining of *Streptomyces ambofaciens*', *Journal of Industrial Microbiology & Biotechnology,* vol.41, no.2, pp.251-263.

Akoachere, J.T.K., Masalla, T.N. & Njom, H.A. (2013) 'Multi-drug resistant toxigenic *Vibrio cholerae* O1 is persistent in water sources in New Bell-Douala, Cameroon', *BMC Infectious Diseases,* vol.13, no.1, pp.1-12.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) 'Basic local alignment search tool', *Journal of Molecular Biology,* vol.215, no.3, pp.403-410.

Amidi, S., Solter, S., Rashidian, B., Zokaian, A. & Razmjoian, F. (1975) 'Antibiotic use and abuse among physicians in private practice in Shiraz, Iran', *Medical Care*, vol.13. no.4. pp.341-345.

Ankisetty, S., Khan, S.I., Avula, B., Gochfeld, D., Khan, I.A. & Slattery, M. (2013) 'Chlorinated didemnins from the Tunicate *Tridid*emnum *solidum'*, *Marine Drugs,* vol.11, no.11, pp.4478-4486.

Aranda, P.S., LaJoie, D.M. & Jorcyk, C.L. (2012) 'Bleach gel: a simple agarose gel for analysing RNA quality', *Electrophoresis,* vol.33, no.2, pp.366-369.

References

Aron, Z.D., Dorrestein, P.C., Blackhall, J.R., Kelleher, N.L. & Walsh, C.T. (2005) 'Characterization of a new tailoring domain in polyketide biogenesis: the amine transferase domain of MycA in the mycosubtilin gene cluster', *Journal of the American Chemical Society,* vol.127, no.43, pp.14986-14987.

Arthur, C.J., Szafranska, A.E., Long, J., Mills, J., Cox, R.J., Findlow, S.C., Simpson, T.J., Crump, M.P. & Crosby, J. (2006) 'The malonyl transferase activity of type II polyketide synthase acyl carrier proteins', *Chemistry & Biology,* vol.13, no.6, pp.587-596.

Balibar, C.J., Vaillancourt, F.H. & Walsh, C.T. (2005) 'Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains', *Chemistry & Biology*, vol.12, no.11, pp.1189-1200.

Baltz, R.H. (2011) 'Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolite discovery', *Journal of Industrial Microbiology & Biotechnology,* vol.38, no.11, pp.1747-1760.

Belarbi, E.H., Contreras Gómez, A., Chisti, Y., García Camacho, F. & Molina Grima, E. (2003) 'Producing drugs from marine sponges', *Biotechnology Advances,* vol.21, no.7, pp.585-598.

Berdy, J. (2005) 'Bioactive microbial metabolites', *The Journal of Antibiotics,* vol.58, no.1, pp.1-26.

References

Bergendahl, V., Linne, U. & Marahiel, M.A. (2002) 'Mutational analysis of the C-domain in nonribosomal peptide synthesis', *European Journal of Biochemistry,* vol.269, no.2, pp.620-629.

Bergmann, W. & Feeney, R.J. (1951) 'Contributions to the study of marine products. XXXII. The nucleosides of sponges. I. 1', *The Journal of Organic Chemistry,* vol.16, no.6, pp.981-987.

Bertelli, C. & Greub, G. (2013) 'Rapid bacterial genome sequencing: methods and applications in clinical microbiology', *Clinical Microbiology and Infection,* vol.19, no.9, pp.803-813.

Bertrand, S., Schumpp, O., Bohni, N., Monod, M., Gindro, K. & Wolfender, J. (2013) '*De novo* production of metabolites by fungal co-culture of *Trichophyton rubrum* and *Bionectria ochroleuca*', *Journal of Natural Products*, vol.76, no.6, pp.1157-1165.

Bills, G., Platas, G., Fillola, A., Jimenez, M., Collado, J., Vicente, F., Martin, J., Gonzalez, A., Bur-Zimmermann, J. & Tormo, J. (2008) 'Enhancement of antibiotic and secondary metabolite detection from filamentous fungi by growth on nutritional arrays', *Journal of Applied Microbiology,* vol.104, no.6, pp.1644-1658.

Blanco, G. (2012) 'Comparative analysis of a cryptic thienamycin-like gene cluster identified in *Streptomyces flavogriseus* by genome mining', *Archives of Microbiology,* vol.194, no.6, pp.549-555.

References

Blin, K., Medema, M.H., Kazempour, D., Fischbach, M.A., Breitling, R., Takano, E. & Weber, T. (2013) 'antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers', *Nucleic Acids Research,* vol.41, Web server issue, pp.W204-12.

Bode, H.B., Bethe, B., Höfs, R. & Zeeck, A. (2002) 'Big effects from small changes: possible ways to explore nature's chemical diversity', *ChemBioChem,* vol.3, no.7, pp.619-627.

Borel, J.F., Feurer, C., Gubler, H. & Stähelin, H. (1994) 'Biological effects of cyclosporin A: a new antilymphocytic agent', *Agents and Actions,* vol.43, no.3-4, pp.179-186.

Bruner, S.D., Weber, T., Kohli, R.M., Schwarzer, D., Marahiel, M.A., Walsh, C.T. & Stubbs, M.T. (2002) 'Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE', *Structure,* vol.10, no.3, pp.301-310.

Caboche, S., Leclere, V., Pupin, M., Kucherov, G. & Jacques, P. (2010) 'Diversity of monomers in nonribosomal peptides: towards the prediction of origin and biological activity', *Journal of Bacteriology,* vol.192, no.19, pp.5143-5150.

Caboche, S., Pupin, M., Leclere, V., Fontaine, A., Jacques, P. & Kucherov, G. (2008) 'NORINE: a database of nonribosomal peptides', *Nucleic Acids Research,* vol.36, Database issue, pp.D326-31.

Cane, D.E. & Walsh, C.T. (1999) 'The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases', *Chemistry & Biology,* vol.6, no.12, pp.R319-R325.

References

Cawoy, H., Debois, D., Franzil, L., De Pauw, E., Thonart, P. & Ongena, M. (2015) 'Lipopeptides as main ingredients for inhibition of fungal phytopathogens by *Bacillus subtilis/amyloliquefaciens*', *Microbial Biotechnology,* vol.8, no.2, pp.281-295.

Challis, G.L. & Naismith, J.H. (2004) 'Structural aspects of non-ribosomal peptide biosynthesis', *Current Opinion in Structural Biology,* vol.14, no.6, pp.748-756.

Challis, G.L., Ravel, J. & Townsend, C.A. (2000) 'Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains', *Chemistry & Biology,* vol.7, no.3, pp.211-224.

Chang, Z., Flatt, P., Gerwick, W.H., Nguyen, V., Willis, C.L. & Sherman, D.H. (2001) 'The barbamide biosynthetic gene cluster: a novel marine Cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit', *Gene*, vol.296, no.1, pp.235-247.

Chen, H., O'Connor, S., Cane, D.E. & Walsh, C.T. (2001) 'Epothilone biosynthesis: assembly of the methylthiazolylcarboxy starter unit on the EpoB subunit', *Chemistry & Biology,* vol.8, no.9, pp.899-912.

Chen, X.H., Koumoutsi, A., Scholz, R., Eisenreich, A., Schneider, K., Heinemeyer, I., Morgenstern, B., Voss, B., Hess, W.R. & Reva, O. (2007) 'Comparative analysis of the complete genome sequence of the plant growth–promoting bacterium *Bacillus amyloliquefaciens* FZB42', *Nature Biotechnology,* vol.25, no.9, pp.1007-1014.

Chen, X.H., Vater, J., Piel, J., Franke, P., Scholz, R., Schneider, K., Koumoutsi, A., Hitzeroth, G., Grammel, N., Strittmatter, A.W., Gottschalk, G., Sussmuth, R.D. &

References

Borriss, R. (2006) ‘Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42’, *Journal of Bacteriology,* vol.188, no.11, pp.4024-4036.

Chen, Y., Smanski, M.J. & Shen, B. (2010) ‘Improvement of secondary metabolite production in *Streptomyces* by manipulating pathway regulation’, *Applied Microbiology and Biotechnology*, vol.86, no.1, pp.19-25.

Cheng, Y.Q., Tang, G.L. & Shen, B. (2003) ‘Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis’, *Proceedings of the National Academy of Sciences of the United States of America,* vol.100, no.6, pp.3149-3154.

Claxton, H.B., Akey, D.L., Silver, M.K., Admiraal, S.J. & Smith, J.L. (2009) ‘Structure and functional analysis of RifR, the type II thioesterase from the rifamycin biosynthetic pathway’, *The Journal of Biological Chemistry,* vol.284, no.8, pp.5021-5029.

Conti, E., Stachelhaus, T., Marahiel, M.A. & Brick, P. (1997) ‘Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S’, *The EMBO Journal,* vol.16, no.14, pp.4174-4183.

Cooper, E.L. & Yao, D. (2012) ‘Diving for drugs: tunicate anticancer compounds’, *Drug Discovery Today,* vol.17, no.11, pp.636-648.

Cosgrove, S.E. (2006) ‘The relationship between antimicrobial resistance and patient outcomes: mortality, length of hospital stay, and health care costs’, *Clinical Infectious Diseases: An official publication of the Infectious Diseases Society of America,* vol.42 Suppl 2, pp.S82-9.

References

Cox, C.D. & Adams, P. (1985) ‘Siderophore activity of pyoverdin for *Pseudomonas aeruginosa*’, *Infection and Immunity,* vol.48, no.1, pp.130-138.

Crawford, J.M. & Townsend, C.A. (2010) ‘New insights into the formation of fungal aromatic polyketides’, *Nature Reviews Microbiology,* vol.8, no.12, pp.879-889.

Dalisay, D.S. & Molinski, T.F. (2009) ‘NMR quantitation of natural products at the nanomole scale’, *Journal of Natural Products,* vol.72, no.4, pp.739-744.

Daniel, R. (2004) ‘The soil metagenome–a rich resource for the discovery of novel natural products’, *Current Opinion in Biotechnology*, vol.15, no.3, pp.199-204.

Davies, J. & Davies, D. (2010) ‘Origins and evolution of antibiotic resistance’, *Microbiology and Molecular Biology Reviews,* vol.74, no.3, pp.417-433.

Davies-Coleman, M.T. & Beukes, D.R. (2004) ‘Ten years of marine natural products research at Rhodes University’, *South African Journal of Science,* vol.100, no.11 & 12, pp.p.539-544.

de Jong, A., van Hijum, S.A., Bijlsma, J.J., Kok, J. & Kuipers, O.P. (2006) ‘BAGEL: a web-based bacteriocin genome mining tool’, *Nucleic Acids Research,* vol.34, Web server issue, pp.W273-9.

Demain, A.L. & Zhang, L. (2005) ‘*Natural products*: *drug discovery and therapeutic medicine’*, New Delhi, Humana Press.

Dias, D.A., Urban, S. & Roessner, U. (2012) ‘A historical overview of natural products in drug discovery’, *Metabolites,* vol.2, no.2, pp.303-336.

References

Doull, J.L., Singh, A.K., Hoare, M. & Ayer, S.W. (1994) 'Conditions for the production of jadomycin B by *Streptomyces venezuelae* ISP5230: Effects of heat shock, ethanol treatment and phage infection', *Journal of Industrial Microbiology*, vol.13, no.2, pp.120-125.

Du, L., Sánchez, C. & Shen, B. (2001) 'Hybrid peptide–polyketide natural products: biosynthesis and prospects toward engineering novel molecules', *Metabolic Engineering,* vol.3, no.1, pp.78-95.

Duitman, E.H., Hamoen, L.W., Rembold, M., Venema, G., Seitz, H., Saenger, W., Bernhard, F., Reinhardt, R., Schmidt, M., Ullrich, C., Stein, T., Leenders, F. & Vater, J. (1999) 'The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase', *Proceedings of the National Academy of Sciences of the United States of America,* vol.96, no.23, pp.13294-13299.

Ehrenreich, I.M., Waterbury, J.B. & Webb, E.A. (2005) 'Distribution and diversity of natural product genes in marine and freshwater Cyanobacterial cultures and genomes', *Applied and Environmental Microbiology,* vol.71, no.11, pp.7401-7413.

Eppelmann, K., Doekel, S. & Marahiel, M.A. (2001) 'Engineered biosynthesis of the peptide antibiotic bacitracin in the surrogate host *Bacillus subtilis*', *The Journal of Biological Chemistry*, vol.276, no.37, pp.34824-34831.

Erwin, P.M., Olson, J.B. & Thacker, R.W. (2011) 'Phylogenetic diversity, host-specificity and community profiling of sponge-associated bacteria in the northern Gulf of Mexico', *PLoS One,* vol.6, no.11, pp.e26806.

References

Falagas, M.E. & Bliziotis, I.A. (2007) 'Pandrug-resistant gram-negative bacteria: the dawn of the post-antibiotic era?', *International Journal of Antimicrobial Agents,* vol.29, no.6, pp.630-636.

Farrelly, V., Rainey, F.A. & Stackebrandt, E. (1995) 'Effect of genome size and rrn gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species', *Applied and Environmental Microbiology,* vol.61, no.7, pp.2798-2801.

Felnagle, E.A., Jackson, E.E., Chan, Y.A., Podevels, A.M., Berti, A.D., McMahon, M.D. & Thomas, M.G. (2008) 'Nonribosomal peptide synthetases involved in the production of medically relevant natural products', *Molecular Pharmaceutics,* vol.5, no.2, pp.191-211.

Fieseler, L., Horn, M., Wagner, M. & Hentschel, U. (2004) 'Discovery of the novel candidate phylum 'Poribacteria in marine sponges', *Applied and Environmental Microbiology,* vol.70, no.6, pp.3724-3732.

Finking, R. & Marahiel, M.A. (2004) 'Biosynthesis of nonribosomal peptides 1', *Annual Review of Microbiology.,* vol.58, pp.453-488.

Friedrich, A.B., Fischer, I., Proksch, P., Hacker, J. & Hentschel, U. (2001) 'Temporal variation of the microbial community associated with the Mediterranean sponge *Aplysina aerophoba*', *FEMS Microbiology Ecology,* vol.38, no.2-3, pp.105-113.

Funa, N., Ohnishi, Y., Fujii, I., Shibuya, M., Ebizuka, Y. & Horinouchi, S. (1999) 'A new pathway for polyketide synthesis in microorganisms', *Nature,* vol.400, no.6747, pp.897-899.

References

Galm, U., Wendt-Pienkowski, E., Wang, L., Huang, S., Unsin, C., Tao, M., Coughlin, J.M. & Shen, B. (2011) 'Comparative analysis of the biosynthetic gene clusters and pathways for three structurally related antitumor antibiotics: bleomycin, tallysomycin, and zorbamycin', *Journal of Natural Products,* vol.74, no.3, pp.526-536.

Gao, X., Wang, P. & Tang, Y. (2010) 'Engineered polyketide biosynthesis and biocatalysis in *Escherichia coli*', *Applied Microbiology and Biotechnology,* vol.88, no.6, pp.1233-1242.

Gehring, A.M., Mori, I. & Walsh, C.T. (1998) 'Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF', *Biochemistry,* vol.37, no.8, pp.2648-2659.

Gerwick, W.H. & Moore, B.S. (2012) 'Lessons from the past and charting the future of marine natural products drug discovery and chemical biology', *Chemistry & Biology,* vol.19, no.1, pp.85-98.

Gevers, W., Kleinkauf, H. & Lipmann, F. (1968) 'The activation of amino acids for biosynthesis of gramicidin S', *Proceedings of the National Academy of Sciences of the United States of America,* vol.60, no.1, pp.269-276.

Ginolhac, A., Jarrin, C., Gillet, B., Robe, P., Pujic, P., Tuphile, K., Bertrand, H., Vogel, T.M., Perriere, 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,* vol.70, no.9, pp.5522-5527.

References

Glöckner, F.O. & Joint, I. (2010) 'Marine microbial genomics in Europe: current status and perspectives', *Microbial Biotechnology,* vol.3, no.5, pp.523-530.

Gontang, E.A., Gaudencio, S.P., Fenical, W. & Jensen, P.R. (2010) 'Sequence-based analysis of secondary-metabolite biosynthesis in marine Actinobacteria', *Applied and Environmental Microbiology,* vol.76, no.8, pp.2487-2499.

Graça, A.P., Bondoso, J., Gaspar, H., Xavier, J.R., Monteiro, M.C., de la Cruz, M., Oves-Costales, D., Vicente, F. & Lage, O.M. (2013) 'Antimicrobial activity of heterotrophic bacterial communities from the marine sponge *Erylus discophorus* (Astrophorida, Geodiidae)', *PLoS One,* vol.8, no.11, pp.e78992.

Griffiths, C.L., Robinson, T.B., Lange, L. & Mead, A. (2010) 'Marine biodiversity in South Africa: an evaluation of current states of knowledge', *PLoS One,* vol.5, no.8, pp.e12008.

Grüschow, S., Rackham, E.J. & Goss, R.J. (2011) 'Diversity in natural product families is governed by more than enzyme promiscuity alone: establishing control of the pacidamycin portfolio', *Chemical Science,* vol.2, no.11, pp.2182-2186.

Gunasekera, S.P., Gunasekera, M., Longley, R.E. & Schulte, G.K. (1990) 'Discodermolide: a new bioactive polyhydroxylated lactone from the marine sponge *Discodermia dissoluta*', *The Journal of Organic Chemistry,* vol.55, no.16, pp.4912-4915.

H3-D — Africa's first integrated drug discovery and development centre. (2015) [ONLINE] Available at: http://h3d.co.za/. [Accessed; 7th May 2015].

References

Haag, H., Hantke, K., Drechsel, H., Stojiljkovic, I., Jung, G. & Zahner, H. (1993) 'Purification of yersiniabactin: a siderophore and possible virulence factor of *Yersinia enterocolitica*', *Journal of General Microbiology,* vol.139, no.9, pp.2159-2165.

Hamada, T., Matsunaga, S., Yano, G. & Fusetani, N. (2005) 'Polytheonamides A and B, highly cytotoxic, linear polypeptides with unprecedented structural features, from the marine sponge, *Theonella swinhoei*', *Journal of the American Chemical Society,* vol.127, no.1, pp.110-118.

Handelsman, J. (2004) 'Metagenomics: application of genomics to uncultured microorganisms', *Microbiology and Molecular Biology Reviews,* vol.68, no.4, pp.669-685.

Hancock, R.E. & Chapple, D.S. (1999) 'Peptide antibiotics', *Antimicrobial Agents and Chemotherapy,* vol.43, no.6, pp.1317-1323.

Haygood, M.G., Schmidt, E.W., Davidson, S.K. & Faulkner, D.J. (1999) 'Microbial symbionts of marine invertebrates: opportunities for microbial biotechnology', *Journal of Molecular Microbiology and Biotechnology,* vol.1, no.1, pp.33-43.

Heindl, H., Wiese, J., Thiel, V. & Imhoff, J.F. (2010) 'Phylogenetic diversity and antimicrobial activities of bryozoan-associated bacteria isolated from Mediterranean and Baltic Sea habitats', *Systematic and Applied Microbiology,* vol.33, no.2, pp.94-104.

Hentschel, U., Hopke, J., Horn, M., Friedrich, A.B., Wagner, M., Hacker, J. & Moore, B.S. (2002) 'Molecular evidence for a uniform microbial community in sponges from different oceans', *Applied and Environmental Microbiology,* vol.68, no.9, pp.4431-4440.

References

Hentschel, U., Schmid, M., Wagner, M., Fieseler, L., Gernert, C. & Hacker, J. (2001) 'Isolation and phylogenetic analysis of bacteria with antimicrobial activities from the Mediterranean sponges *Aplysina aerophoba* and *Aplysina cavernicola*', *FEMS Microbiology Ecology,* vol.35, no.3, pp.305-312.

Hentschel, U., Usher, K.M. & Taylor, M.W. (2006) 'Marine sponges as microbial fermenters', *FEMS Microbiology Ecology,* vol.55, no.2, pp.167-177.

Herbst, D.A., Boll, B., Zocher, G., Stehle, T. & Heide, L. (2013) 'Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes', *The Journal of Biological Chemistry,* vol.288, no.3, pp.1991-2003.

Hooper, G.J. & Davies-Coleman, M.T. (1995) 'New Metabolites from the South African soft coral *Capnella thyrsoidea*', *Tetrahedron,* vol.51, no.36, pp.9973-9984.

Hooper, G.J., Davies-Coleman, M.T., Kelly-Borges, M. & Coetzee, P.S. (1996) 'New alkaloids from a South African latrunculid sponge', *Tetrahedron Letters,* vol.37, no.39, pp.7135-7138.

Hu, G., Yuan, J., Sun, L., She, Z., Wu, J., Lan, X., Zhu, X., Lin, Y. & Chen, S. (2011) 'Statistical research on marine natural products based on data obtained between 1985 and 2008', *Marine Drugs,* vol.9, no.4, pp.514-525.

Hubbard, B.K., Thomas, M.G. & Walsh, C.T. (2000) 'Biosynthesis of L-p-hydroxyphenylglycine, a non-proteinogenic amino acid constituent of peptide antibiotics', *Chemistry & Biology,* vol.7, no.12, pp.931-942.

References

Hutchinson, C.R. (2003) 'Polyketide and non-ribosomal peptide synthases: falling together by coming apart', *Proceedings of the National Academy of Sciences of the United States of America,* vol.100, no.6, pp.3010-3012.

Ichikawa, N., Sasagawa, M., Yamamoto, M., Komaki, H., Yoshida, Y., Yamazaki, S. & Fujita, N. (2013) 'DoBISCUIT: a database of secondary metabolite biosynthetic gene clusters', *Nucleic Acids Research,* vol.41, Database issue, pp.D408-14.

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. & Ōmura, S. (2003) 'Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*', *Nature Biotechnology,* vol.21, no.5, pp.526-531.

Jensen, P.R., Chavarria, K.L., Fenical, W., Moore, B.S. & Ziemert, N. (2014) 'Challenges and triumphs to genomics-based natural product discovery', *Journal of Industrial Microbiology & Biotechnology,* vol.41, no.2, pp.203-209.

Ji, H., Li, X. & Zhang, H. (2009) 'Natural products and drug discovery', *EMBO Reports,* vol.10, no.3, pp.194-200.

Kealey, J.T., Liu, L., Santi, D.V., Betlach, M.C. & Barr, P.J. (1998) 'Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts', *Proceedings of the National Academy of Sciences of the United States of America,* vol.95, no.2, pp.505-509.

Keatinge-Clay, A.T. & Stroud, R.M. (2006) 'The structure of a ketoreductase determines the organization of the β-carbon processing enzymes of modular polyketide synthases', *Structure,* vol.14, no.4, pp.737-748.

Kennedy, J., Marchesi, J.R. & Dobson, A.D. (2007) 'Metagenomic approaches to exploit the biotechnological potential of the microbial consortia of marine sponges', *Applied Microbiology and Biotechnology,* vol.75, no.1, pp.11-20.

Khanafari, A., Assadi, M.M. & Fakhr, F.A. (2006) 'Review of prodigiosin, pigmentation in *Serratia marcescens*', *Online Journal of Biological Sciences,* vol.6, no.1, pp.1-13.

Khosla, C., Gokhale, R.S., Jacobsen, J.R. & Cane, D.E. (1999) 'Tolerance and specificity of polyketide synthases', *Annual Review of Biochemistry,* vol.68, no.1, pp.219-253.

Kijjoa, A. & Sawangwong, P. (2004) 'Drugs and cosmetics from the sea', *Marine Drugs,* vol.2, no.2, pp.73-82.

King, R.W., Bauer, J.D. & Brady, S.F. (2009) 'An environmental DNA-derived type II polyketide biosynthetic pathway encodes the biosynthesis of the pentacyclic polyketide erdacin', *Angewandte Chemie International Edition,* vol.48, no.34, pp.6257-6261.

Konz, D., Doekel, S. & Marahiel, M.A. (1999) 'Molecular and biochemical characterization of the protein template controlling biosynthesis of the lipopeptide lichenysin', *Journal of Bacteriology,* vol.181, no.1, pp.133-140.

References

Kouprina, N. & Larionov, V. (2006) 'TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution', *Nature Reviews Genetics,* vol.7, no.10, pp.805-812.

Lautru, S. & Challis, G.L. (2004) 'Substrate recognition by nonribosomal peptide synthetase multi-enzymes', *Microbiology (Reading, England),* vol.150, no.6, pp.1629-1636.

Law, B.K. (2005) 'Rapamycin: an anti-cancer immunosuppressant?', *Critical Reviews in Oncology/Hematology,* vol. 56, no.1, pp.47-60.

Lawrence, J.G. & Ochman, H. (1997) 'Amelioration of bacterial genomes: rates of change and exchange', *Journal of Molecular Evolution,* vol.44, no.4, pp.383-397.

Leal, M.C., Puga, J., Serôdio, J., Gomes, N.C. & Calado, R. (2012) 'Trends in the discovery of new marine natural products from invertebrates over the last two decades–where and what are we bioprospecting?', *PLoS One,* vol.7, no.1, pp.e30580.

Lee, B.N., Kroken, S., Chou, D.Y., Robbertse, B., Yoder, O.C. & Turgeon, B.G. (2005) 'Functional analysis of all nonribosomal peptide synthetases in *Cochliobolus heterostrophus* reveals a factor, NPS6, involved in virulence and resistance to oxidative stress', *Eukaryotic Cell,* vol.4, no.3, pp.545-555.

Lemaire, P. (2011) 'Evolutionary crossroads in developmental biology: the tunicates', *Development (Cambridge, England),* vol.138, no.11, pp.2143-2152.

References

Levy, S.B. (2002) ‘The 2000 Garrod lecture. Factors impacting on the problem of antibiotic resistance’, *The Journal of Antimicrobial Chemotherapy,* vol.49, no.1, pp.25-30.

Lewis, K. (2012) ‘Antibiotics: recover the lost art of drug discovery’, *Nature,* vol.485, no.7399, pp.439-440.

Leys, S.P. & Eerkes-Medrano, D.I. (2006) ‘Feeding in a calcareous sponge: particle uptake by pseudopodia’, *The Biological Bulletin,* vol.211, no.2, pp.157-171.

Li, M.H., Ung, P.M., Zajkowski, J., Garneau-Tsodikova, S. & Sherman, D.H. (2009) ‘Automated genome mining for natural products’, *BMC Bioinformatics,* vol.10, no.1, pp.185.

Lim, J., Sun, H., Fan, J., Hameed, I.F., Lescar, J., Liang, Z. & Yang, D. (2012) ‘Rigidifying acyl carrier protein domain in Iterative Type I PKS CalE8 does not affect its function’, *Biophysical Journal,* vol.103, no.5, pp.1037-1044.

Lin, Y., Schiavo, S., Orjala, J., Vouros, P. & Kautz, R. (2008) ‘Microscale LC-MS-NMR platform applied to the identification of active Cyanobacterial metabolites’, *Analytical Chemistry,* vol.80, no.21, pp.8045-8054.

Liu, Q., Liu, C., Yu, J., Yan, J. & Qi, X. (2012) ‘Analysis of the ketosynthase genes in *Streptomyces* and its implications for preventing reinvestigation of polyketides with bioactivities’, *Journal of Agricultural Science*, vol.4, no.7, pp.p262.

Lopanik, N.B. (2014) ‘Chemical defensive symbioses in the marine environment’, *Functional Ecology*, vol.28, no.2, pp.328-340.

References

Ma, D., Cook, D.N., Alberti, M., Pon, N.G., Nikaido, H. & Hearst, J.E. (1995) 'Genes acrA and acrB encode a stress-induced efflux system of *Escherichia coli*', *Molecular Microbiology,* vol.16, no.1, pp.45-55.

Mach, B., Reich, E. & Tatum, E.L. (1963) 'Separation of the Biosynthesis of the Antibiotic Polypeptide Tyrocidine from Protein Biosynthesis', *Proceedings of the National Academy of Sciences of the United States of America,* vol.50, no.1, pp.175-181.

Macintyre, L., Zhang, T., Viegelmann, C., Martinez, I.J., Cheng, C., Dowdells, C., Abdelmohsen, U.R., Gernert, C., Hentschel, U. & Edrada-Ebel, R. (2014) 'Metabolomic Tools for Secondary Metabolite Discovery from Marine Microbial Symbionts', *Marine Drugs,* vol.12, no.6, pp.3416-3448.

Mandal, S.M., Barbosa, A.E. & Franco, O.L. (2013) 'Lipopeptides in microbial infection control: scope and reality for industry', *Biotechnology Advances,* vol.31, no.2, pp.338-345.

Mann, M.G., Mkwananzi, H.B., Antunes, E.M., Whibley, C.E., Hendricks, D.T., Bolton, J.J. & Beukes, D.R. (2007) 'Halogenated monoterpene aldehydes from the South African marine alga *Plocamium corallorhiza*', *Journal of Natural Products,* vol.70, no.4, pp.596-599.

Marahiel, M.A., Stachelhaus, T. & Mootz, H.D. (1997) 'Modular peptide synthetases involved in nonribosomal peptide synthesis', *Chemical Reviews,* vol.97, no.7, pp.2651-2674.

References

Marmann, A., Aly, A.H., Lin, W., Wang, B. & Proksch, P. (2014) 'Co-cultivation—A powerful emerging tool for enhancing the chemical diversity of microorganisms', *Marine Drugs*, vol. 12, no.2, pp.1043-1065.

Martinez, J.S. & Butler, A. (2007) 'Marine amphiphilic siderophores: marinobactin structure, uptake, and microbial partitioning', *Journal of Inorganic Biochemistry,* vol.101, no.11, pp.1692-1698.

May, J.J., Kessler, N., Marahiel, M.A. & Stubbs, M.T. (2002) 'Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases', *Proceedings of the National Academy of Sciences of the United States of America,* vol.99, no.19, pp.12120-12125.

McGivern, J.G. (2007) 'Ziconotide: a review of its pharmacology and use in the treatment of pain', *Neuropsychiatric disease and treatment,* vol.3, no.1, pp.69-85.

Means, A.R., Weaver, M.R., Burnett, S.M., Mbonye, M.K., Naikoba, S. & McClelland, R.S. (2014) 'Correlates of Inappropriate Prescribing of Antibiotics to Patients with Malaria in Uganda', *PLoS One,* vol.9, no.2, pp.e90179.

Medema, M.H., Paalvast, Y., Nguyen, D.D., Melnik, A., Dorrestein, P.C., Takano, E. & Breitling, R. (2014) 'Pep2Path: Automated mass spectrometry-guided genome mining of peptidic natural products', *PLoS Computational Biology,* vol.10, no.9, pp.e1003822.

Meena, K.R. & Kanwar, S.S. (2015) 'Lipopeptides as the antifungal and antibacterial agents: applications in food safety and therapeutics', *BioMed Research International,* vol.2015.

Mehbub, M.F., Lei, J., Franco, C. & Zhang, W. (2014) 'Marine sponge derived natural products between 2001 and 2010: Trends and opportunities for discovery of bioactives', *Marine Drugs,* vol.12, no.8, pp.4539-4577.

Mendola, D. (2003) 'Aquaculture of three phyla of marine invertebrates to yield bioactive metabolites: process developments and economics', *Biomolecular Engineering,* vol.20, no.4, pp.441-458.

Miao, V., Brost, R., Chapple, J., She, K., Coëffet-Le Gal, M. & Baltz, R.H. (2006) 'The lipopeptide antibiotic A54145 biosynthetic gene cluster from *Streptomyces fradiae*', *Journal of Industrial Microbiology and Biotechnology*, vol.33, no.2, pp.129-140.

Micek, S.T. (2007) 'Alternatives to vancomycin for the treatment of methicillin-resistant *Staphylococcus aureus* infections', *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America,* vol.45 Suppl 3, pp.S184-90.

Miethke, M. & Marahiel, M.A. (2007) 'Siderophore-based iron acquisition and pathogen control', *Microbiology and Molecular Biology Reviews,* vol.71, no.3, pp.413-451.

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,* vol.8, no.9, pp.e76021.

Moffitt, M.C. & Neilan, B.A. (2004) 'Characterization of the nodularin synthetase gene cluster and proposed theory of the evolution of cyanobacterial hepatotoxins', *Applied and Environmental Microbiology,* vol.70, no.11, pp.6353-6362.

References

Muscholl-Silberhorn, A., Thiel, V. & Imhoff, J.F. (2008) 'Abundance and bioactivity of cultured sponge-associated bacteria from the Mediterranean Sea', *Microbial Ecology,* vol.55, no.1, pp.94-106.

NaPDoS. (2012) 'Natural product domain seeker' Available from: http://napdos.ucsd.edu/, [Accessed on: 10th January 2015].

Neilan, B.A., Dittmann, E., Rouhiainen, L., Bass, R.A., Schaub, V., Sivonen, K. & Borner, T. (1999) 'Nonribosomal peptide synthesis and toxigenicity of *Cyanobacteria*', *Journal of Bacteriology,* vol.181, no.13, pp.4089-4097.

Neilands, J.B. (1995) 'Siderophores: structure and function of microbial iron transport compounds', *The Journal of Biological Chemistry,* vol.270, no.45, pp.26723-26726.

Nicholson, W.L., Munakata, N., Horneck, G., Melosh, H.J. & Setlow, P. (2000) 'Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments', *Microbiology and Molecular Biology Reviews,* vol.64, no.3, pp.548-572.

Nieselt, K., Battke, F., Herbig, A., Bruheim, P., Wentzel, A., Jakobsen, O.M., Sletta, H., Alam, M.T., Merlo, M.E., Moore, J., Omara, W.A., Morrissey, E.R., Juarez-Hermosillo, M.A., Rodriguez-Garcia, A., Nentwich, M., Thomas, L., Iqbal, M., Legaie, R., Gaze, W.H., Challis, G.L., Jansen, R.C., Dijkhuizen, L., Rand, D.A., Wild, D.L., Bonin, M., Reuther, J., Wohlleben, W., Smith, M.C., Burroughs, N.J., Martin, J.F., Hodgson, D.A., Takano, E., Breitling, R., Ellingsen, T.E. & Wellington, E.M. (2010) 'The dynamic architecture of the metabolic switch in *Streptomyces coelicolor*', *BMC Genomics,* vol.11, pp.10-2164-11-10.

References

Tunicates - Nobanis. [ONLINE] Available at: http://www.nobanis.org/marine-identification-key/tunicates. [Accessed 1st May 2015].

Norrander, J., Kempe, T. & Messing, J. (1983) 'Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis', *Gene*, vol.26, no.1, pp.101-106.

Noskov, V.N., Kouprina, N., Leem, S.H., Ouspznski, I., Barrett, J.C. & Larionov, V. (2003) 'A general cloning system to selectively isolate any eukaryotic or prokaryotic genomic region in yeast', *BMC Genomics,* vol.4, no.1, pp.16.

Novakova, R., Bistakova, J., Homerova, D., Rezuchova, B. & Kormanec, J. (2002) 'Cloning and characterization of a polyketide synthase gene cluster involved in biosynthesis of a proposed angucycline-like polyketide auricin in *Streptomyces aureofaciens* CCM 3239', *Gene,* vol.297, no.1, pp.197-208.

Nützmann, H.W., Reyes-Dominguez, Y., Scherlach, K., Schroeckh, V., Horn, F., Gacek, A., Schumann, J., Hertweck, C., Strauss, J. & Brakhage, A.A. (2011) 'Bacteria-induced natural product formation in the fungus *Aspergillus nidulans* requires Saga/Ada-mediated histone acetylation', *Proceedings of the National Academy of Sciences of the United States of America,* vol.108, no.34, pp.14282-14287.

O'Halloran, J., Barbosa, T., Morrissey, J., Kennedy, J., O'Gara, F. & Dobson, A. (2011) 'Diversity and antimicrobial activity of *Pseudovibrio* spp. from Irish marine sponges', *Journal of Applied Microbiology,* vol.110, no.6, pp.1495-1508.

O'Donnell, M.R., Jarand, J., Loveday, M., Padayatchi, N., Zelnick, J., Werner, L., Naidoo, K., Master, I., Osburn, G. & Kvasnovsky, C. (2010) 'High incidence of hospital

admissions with multidrug-resistant and extensively drug-resistant tuberculosis among South African health care workers', *Annals of Internal Medicine,* vol.153, no.8, pp.516-522.

Oh, D., Kauffman, C.A., Jensen, P.R. & Fenical, W. (2007) 'Induced production of emericellamides A and B from the marine-derived fungus *Emericella* sp. in competing co-culture', *Journal of Natural Products,* vol.70, no.4, pp.515-520.

Ola, A.R., Thomy, D., Lai, D., Brötz-Oesterhelt, H. & Proksch, P. (2013) 'Inducing secondary metabolite production by the endophytic fungus *Fusarium tricinctum* through coculture with *Bacillus subtilis*', *Journal of Natural Products,* vol.76, no.11, pp.2094-2099.

Ongley, S.E., Bian, X., Neilan, B.A. & Müller, R. (2013) 'Recent advances in the heterologous expression of microbial natural product biosynthetic pathways', *Natural Product Reports,* vol.30, no.8, pp.1121-1138.

Pabel, C.T., Vater, J., Wilde, C., Franke, P., Hofemeister, J., Adler, B., Bringmann, G., Hacker, J. & Hentschel, U. (2003) 'Antimicrobial activities and matrix-assisted laser desorption/ionization mass spectrometry of *Bacillus* isolates from the marine sponge *Aplysina aerophoba*', *Marine Biotechnology,* vol.5, no.5, pp.424-434.

Pathak, K.V. & Keharia, H. (2014) 'Identification of surfactins and iturins produced by potent fungal antagonist, *Bacillus subtilis* K1 isolated from aerial roots of banyan (*Ficus benghalensis*) tree using mass spectrometry', *3 Biotech,* vol.4, no.3, pp.283-295.

References

Penesyan, A., Tebben, J., Lee, M., Thomas, T., Kjelleberg, S., Harder, T. & Egan, S. (2011) 'Identification of the antibacterial compound produced by the marine epiphytic bacterium *Pseudovibrio* sp. D323 and related sponge-associated bacteria', *Marine Drugs,* vol.9, no.8, pp.1391-1402.

Perdicaris, S., Vlachogianni, T. & Valavanidis, A. (2013) 'Bioactive natural substances from marine sponges: new developments and prospects for future pharmaceuticals', *Natural Products Chemistry & Research*, vol.1, no.3, pp.1-8.

Pérez-Matos, A.E., Rosado, W. & Govind, N.S. (2007) 'Bacterial diversity associated with the Caribbean tunicate *Ecteinascidia turbinata*', *Antonie van Leeuwenhoek,* vol.92, no.2, pp.155-164.

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,* vol.101, no.46, pp.16222-16227.

Pomponi, S.A. (2001) 'The oceans and human health: the discovery and development of marine-derived drugs', *Oceanography,* vol.14, no.1, pp.78-87.

Proksch, P., Edrada, R. & Ebel, R. (2002) 'Drugs from the seas–current status and microbiological implications', *Applied Microbiology and Biotechnology*, vol.59, no.2-3, pp.125-134.

Proksch, P., Edrada-Ebel, R. & Ebel, R. (2003) 'Drugs from the sea-opportunities and obstacles', *Marine Drugs,* vol.1, no.1, pp.5-17.

References

Raaijmakers, J.M., De Bruijn, I., Nybroe, O. & Ongena, M. (2010) 'Natural functions of lipopeptides from *Bacillus* and *Pseudomonas*: more than surfactants and antibiotics', *FEMS Microbiology Reviews,* vol.34, no.6, pp.1037-1062.

Radford, J.L., Hutchinson, A.E., Burandt, M. & Raftos, D.A. (2000) 'Effects of metal-based environmental pollutants on tunicate hemocytes', *Journal of Invertebrate Pathology,* vol.76, no.4, pp.242-248.

Rateb, M.E., Houssen, W.E., Harrison, W.T., Deng, H., Okoro, C.K., Asenjo, J.A., Andrews, B.A., Bull, A.T., Goodfellow, M. & Ebel, R. (2011) 'Diverse metabolic profiles of a *Streptomyces* strain isolated from a hyper-arid environment', *Journal of Natural Products*, vol.74, no.9, pp.1965-1971.

Rath, C.M., Janto, B., Earl, J., Ahmed, A., Hu, F.Z., Hiller, L., Dahlgren, M., Kreft, R., Yu, F. & Wolff, J.J. (2011) 'Meta-omic characterization of the marine invertebrate microbial consortium that produces the chemotherapeutic natural product ET-743', *ACS Chemical Biology,* vol.6, no.11, pp.1244-1256.

Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D.H. (2007) 'Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution', *BMC Evolutionary Biology,* vol.7, pp.78.

Reilly, H.C., Schatz, A. & Waksman, S.A. (1945) 'Antifungal properties of antibiotic substances', *Journal of Bacteriology*, vol.49, no.6, pp.585-594.

References

Ren, Q., Kang, K.H. & Paulsen, I.T. (2004) 'TransportDB: a relational database of cellular membrane transport systems', *Nucleic Acids Research,* vol.32, Database issue, pp.D284-8.

Revill, W.P., Bibb, M.J. & Hopwood, D.A. (1995) 'Purification of a malonyltransferase from *Streptomyces coelicolor* A3 (2) and analysis of its genetic determinant', *Journal of Bacteriology,* vol.177, no.14, pp.3946-3952.

Reysenbach, A., Pace, N. & Robb, F. (1995) 'Archaea: a laboratory manual—thermophiles', *New York, Cold Spring Harbour Press.*Pp.101-107.

Ritacco, F., Haltli, B., Janso, J., Greenstein, M. & Bernan, V. (2003) 'Dereplication of *Streptomyces* soil isolates and detection of specific biosynthetic genes using an automated ribotyping instrument', *Journal of Industrial Microbiology and Biotechnology,* vol.30, no.8, pp.472-479.

Rosconi, F., Davyt, D., Martínez, V., Martínez, M., Abin-Carriquiry, J.A., Zane, H., Butler, A., Souza, E.M. & Fabiano, E. (2013) 'Identification and structural characterization of serobactins, a suite of lipopeptide siderophores produced by the grass endophyte *Herbaspirillum seropedicae*', *Environmental Microbiology,* vol.15, no.3, pp.916-927.

Röttig, M., Medema, M.H., Blin, K., Weber, T., Rausch, C. & Kohlbacher, O. (2011) 'NRPSpredictor2--a web server for predicting NRPS adenylation domain specificity', *Nucleic Acids Research,* vol.39, Web server issue, pp.W362-7.

References

Rouhiainen, L., Jokela, J., Fewer, D.P., Urmann, M. & Sivonen, K. (2010) 'Two alternative starter modules for the non-ribosomal biosynthesis of specific anabaenopeptin variants in *Anabaena* (Cyanobacteria)', *Chemistry & Biology,* vol.17, no.3, pp.265-273.

Ruan, X., Pereda, A., Stassi, D.L., Zeidner, D., Summers, R.G., Jackson, M., Shivakumar, A., Kakavas, S., Staver, M.J., Donadio, S. & Katz, L. (1997) 'Acyltransferase domain substitutions in erythromycin polyketide synthase yield novel erythromycin derivatives', *Journal of Bacteriology,* vol.179, no.20, pp.6416-6425.

Sabdono, A. (2008) 'Bacterial symbionts of reef's invertebrates: a marine natural drug's factory', *Journal of Coastal Development,* vol.12, no.1, pp.13-19.

Samel, S.A., Schoenafinger, G., Knappe, T.A., Marahiel, M.A. & Essen, L. (2007) 'Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase', *Structure,* vol.15, no.7, pp.781-792.

Sánchez, S., Chávez, A., Forero, A., García-Huante, Y., Romero, A., Sánchez, M., Rocha, D., Sánchez, B., Ávalos, M. & Guzmán-Trampe, S. (2010) 'Carbon source regulation of antibiotic production', *The Journal of Antibiotics,* vol.63, no.8, pp.442-459.

Sasakura, Y., Nakashima, K., Awazu, S., Matsuoka, T., Nakayama, A., Azuma, J. & Satoh, N. (2005) 'Transposon-mediated insertional mutagenesis revealed the functions of animal cellulose synthase in the ascidian *Ciona intestinalis*', *Proceedings of the National Academy of Sciences of the United States of America,* vol.102, no.42, pp.15134-15139.

References

Saxena, P., Yadav, G., Mohanty, D. & Gokhale, R.S. (2003) 'A new family of type III polyketide synthases in *Mycobacterium tuberculosis*', *The Journal of Biological Chemistry,* vol.278, no.45, pp.44780-44790.

Schirmer, A., Gadkari, R., Reeves, C.D., Ibrahim, F., DeLong, E.F. & 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*, vol.71, no.8, pp.4840-4849.

Selvin, J., Shanmugha Priya, S., Seghal Kiran, G., Thangavelu, T. & Sapna Bai, N. (2009) 'Sponge-associated marine bacteria as indicators of heavy metal pollution', *Microbiological Research,* vol.164, no.3, pp.352-363.

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 Bacteriology,* vol.179, no.23, pp.7360-7368.

Sertan-de Guzman, A.A., Predicala, R.Z., Bernardo, E.B., Neilan, B.A., Elardo, S.P., Mangalindan, G.C., Tasdemir, D., Ireland, C.M., Barraquio, W.L. & Concepcion, G.P. (2007) '*Pseudovibrio denitrificans* strain Z143-1, a heptylprodigiosin-producing bacterium isolated from a Philippine tunicate', *FEMS Microbiology Letters,* vol.277, no.2, pp.188-196.

Sharp, K.H., Eam, B., Faulkner, D.J. & Haygood, M.G. (2007) 'Vertical transmission of diverse microbes in the tropical sponge *Corticium* sp.', *Applied and Environmental Microbiology,* vol.73, no.2, pp.622-629.

References

Shenkar, N., Gittenberger, A., Lambert, G., Rius, M., Moreira Da Rocha, R., Swalla, B.J., Turon, X. (2015) 'Ascidiacea World Database', Accessed at http://www.marinespecies.org/ascidiacea on 1st May 2015

Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y. & Simon, M. (1992) 'Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector', *Proceedings of the National Academy of Sciences of the United States of America,* vol.89, no.18, pp.8794-8797.

Siméone, R., Constant, P., Guilhot, C., Daffe, M. & Chalut, C. (2007) 'Identification of the missing trans-acting enoyl reductase required for phthiocerol dimycocerosate and phenolglycolipid biosynthesis in *Mycobacterium tuberculosis*', *Journal of Bacteriology,* vol.189, no.13, pp.4597-4602.

Sipkema, D., Osinga, R., Schatton, W., Mendola, D., Tramper, J. & Wijffels, R.H. (2005) 'Large-scale production of pharmaceuticals by marine sponges: Sea, cell, or synthesis?', *Biotechnology and Bioengineering*, vol.90, no.2, pp.201-222.

Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., Arrieta, J.M. & Herndl, G.J. (2006) 'Microbial diversity in the deep sea and the underexplored 'rare biosphere'', *Proceedings of the National Academy of Sciences of the United States of America,* vol.103, no.32, pp.12115-12120.

Spellberg, B., Guidos, R., Gilbert, D., Bradley, J., Boucher, H.W., Scheld, W.M., Bartlett, J.G., Edwards, J.,Jr & Infectious Diseases Society of America (2008) 'The epidemic of antibiotic-resistant infections: a call to action for the medical community from the

Infectious Diseases Society of America', *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America,* vol.46, no.2, pp.155-164.

Stachelhaus, T., Mootz, H.D., Bergendahl, V. & Marahiel, M.A., (1998) 'Peptide bond formation in nonribosomal peptide biosynthesis catalytic role of the condensation domain', *Journal of Biological Chemistry,* vol.273, no.35, pp.22773-22781.

Stachelhaus, T., Mootz, H.D. & Marahiel, M.A. (1999) 'The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases', *Chemistry & Biology,* vol.6, no.8, pp.493-505.

Stallforth, P. & Clardy, J. (2014) 'Atlas for drug discovery', *Proceedings of the National Academy of Sciences of the United States of America,* vol.111, no.10, pp.3655-3656.

Staunton, J. & Weissman, K.J. (2001) 'Polyketide biosynthesis: a millennium review', *Natural Product Reports,* vol.18, no.4, pp.380-416.

Stewart, E.J. (2012) 'Growing unculturable bacteria', *Journal of Bacteriology*, vol.194, no.16, pp.4151-4160.

Summers, R.G., Wendt-Pienkowski, E., Motamedi, H. & Hutchinson, C.R. (1993) 'The tcmVI region of the tetracenomycin C biosynthetic gene cluster of *Streptomyces glaucescens* encodes the tetracenomycin F1 monooxygenase, tetracenomycin F2 cyclase, and, most likely, a second cyclase', *Journal of Bacteriology,* vol.175, no.23, pp.7571-7580.

References

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013) 'MEGA6: Molecular Evolutionary Genetics Analysis version 6.0', *Molecular Biology and Evolution,* vol.30, no.12, pp.2725-2729.

Tang, Y., Tsai, S. & Khosla, C. (2003) 'Polyketide chain length control by chain length factor', *Journal of the American Chemical Society,* vol.125, no.42, pp.12708-12709.

Taylor, M.W., Schupp, P.J., De Nys, R., Kjelleberg, S. & Steinberg, P.D. (2005) 'Biogeography of bacteria associated with the marine sponge *Cymbastela concentrica*', *Environmental Microbiology,* vol.7, no.3, pp.419-433.

Taylor, M.W., Radax, R., Steger, D. & Wagner, M. (2007) 'Sponge-associated microorganisms: evolution, ecology, and biotechnological potential', *Microbiology and Molecular Biology Reviews,* vol.71, no.2, pp.295-347.

Thomas, T.R.A., Kavlekar, D.P. & LokaBharathi, P.A. (2010) 'Marine drugs from sponge-microbe association—A review', *Marine Drugs,* vol.8, no.4, pp.1417-1468.

Thompson, J., Walker, R. & Faulkner, D. (1985) 'Screening and bioassays for biologically-active substances from forty marine sponge species from San Diego, California, USA', *Marine Biology,* vol.88, no.1, pp.11-21.

Thriemer, K., Katuala, Y., Batoko, B., Alworonga, J., Devlieger, H., Van Geet, C., Ngbonda, D. & Jacobs, J. (2013) 'Antibiotic prescribing in DR Congo: a knowledge, attitude and practice survey among medical doctors and students', *PLoS One,* vol.8, no.2, pp.e55495.

References

Tillett, D., Dittmann, E., Erhard, M., von Döhren, H., Börner, T. & Neilan, B.A. (2000) 'Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide–polyketide synthetase system', *Chemistry & Biology*, vol.7, no.10, pp.753-764.

Tsuge, K., Akiyama, T. & Shoda, M. (2001) 'Cloning, sequencing, and characterization of the iturin A operon', *Journal of Bacteriology,* vol. 183, no. 21, pp. 6265-6273.

Tsukimoto, M., Nagaoka, M., Shishido, Y., Fujimoto, J., Nishisaka, F., Matsumoto, S., Harunari, E., Imada, C. & Matsuzaki, T. (2011) 'Bacterial production of the tunicate-derived antitumor cyclic depsipeptide didemnin B', *Journal of Natural Products,* vol.74, no.11, pp.2329-2331.

Tulp, M. & Bohlin, L. (2005) 'Rediscovery of known natural compounds: nuisance or goldmine?', *Bioorganic & Medicinal Chemistry,* vol.13, no.17, pp.5274-5282.

van Wyk, A.W., Gray, C.A., Whibley, C.E., Osoniyi, O., Hendricks, D.T., Caira, M.R. & Davies-Coleman, M.T. (2008) 'Bioactive Metabolites from the South African marine mollusk *Trimusculus costatus*', *Journal of Natural Products,* vol.71, no.3, pp.420-425.

Varadavenkatesan, T. & Murty, V.R. (2013) 'Production of a lipopeptide biosurfactant by a novel *Bacillus* sp. and its applicability to enhanced oil recovery', *ISRN Microbiology,* vol.2013.

Vartoukian, S.R., Palmer, R.M. & Wade, W.G. (2010) 'Strategies for culture of 'unculturable'bacteria', *FEMS Microbiology Letters,* vol.309, no.1, pp.1-7.

References

Velho-Pereira, S. & Furtado, I. (2012) 'Antibacterial activity of halophilic bacterial bionts from marine invertebrates of Mandapam-India', *Indian Journal of Pharmaceutical Sciences,* vol.74, no.4, pp.331.

Visalli, M.A., Murphy, E., Projan, S.J. & Bradford, P.A. (2003) 'AcrAB multidrug efflux pump is associated with reduced levels of susceptibility to tigecycline (GAR-936) in *Proteus mirabilis*', *Antimicrobial Agents and Chemotherapy,* vol.47, no.2, pp.665-669.

Vizcaino, J., Sanz, L., Cardoza, R., Monte, E. & Gutierrez, 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,* vol.244, no.1, pp.139-148.

Vollenbroich, D., Kluge, B., D'Souza, C., Zuber, P. & Vater, J. (1993) 'Analysis of a mutant amino acid-activating domain of surfactin synthetase bearing a serine-to-alanine substitution at the site of carboxylthioester formation', *FEBS Letters,* vol.325, no.3, pp.220-224.

Waksman, S.A. & Woodruff, H.B. (1941) '*Actinomyces antibioticus*, a new soil organism antagonistic to pathogenic and non-pathogenic bacteria', *Journal of Bacteriology,* vol.42, no.2, pp.231-249.

Walsh, C.T., Gehring, A.M., Weinreb, P.H., Quadri, L.E. & Flugel, R.S. (1997) 'Post-translational modification of polyketide and nonribosomal peptide synthases', *Current Opinion in Chemical Biology,* vol.1, no.3, pp.309-315.

Walsh, F.M. & Amyes, S.G. (2004) 'Microbiology and drug resistance mechanisms of fully resistant pathogens', *Current Opinion in Microbiology,* vol. 7, no. 5, pp. 439-444.

Wang, X. & Lavrov, D.V. (2008) 'Seventeen new complete mtDNA sequences reveal extensive mitochondrial genome evolution within the Demospongiae', *PLoS One,* vol.3, no.7, pp.e2723.

Weber, T., Baumgartner, R., Renner, C., Marahiel, M.A. & Holak, T.A. (2000) 'Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases', *Structure,* vol.8, no.4, pp.407-418.

Weinig, S., Hecht, H., Mahmud, T. & Müller, R. (2003) 'Melithiazol biosynthesis: further insights into myxobacterial PKS/NRPS systems and evidence for a new subclass of methyl transferases', *Chemistry & Biology,* vol.10, no.10, pp.939-952.

Whibley, C.E., Keyzers, R.A., Soper, A.G., Davies-Coleman, M.T., Samaai, T. & Hendricks, D.T. (2005) 'Antiesophageal cancer activity from Southern African marine organisms', *Annals of the New York Academy of Sciences,* vol.1056, no.1, pp.405-412.

Williams, D.H., Stone, M.J., Hauck, P.R. & Rahman, S.K. (1989) 'Why are secondary metabolites (natural products) biosynthesized?', *Journal of Natural Products,* vol.52, no.6, pp.1189-1208.

Williams, P.G. (2008) 'Panning for chemical gold: marine bacteria as a source of new therapeutics', *Trends in Biotechnology,* vol.27, no.1, pp.45-52.

Williams, L.J., Tabbaa, D.G., Li, N., Berlin, A.M., Shea, T.P., Maccallum, I., Lawrence, M.S., Drier, Y., Getz, G., Young, S.K., Jaffe, D.B., Nusbaum, C. & Gnirke, A. (2012)

'Paired-end sequencing of fosmid libraries by Illumina', *Genome Research,* vol.22, no.11, pp.2241-2249.

Williams, R.P. (1973) 'Biosynthesis of prodigiosin, a secondary metabolite of *Serratia marcescens*', *Applied Microbiology,* vol.25, no.3, pp.396-402.

Wilson, M.C., Mori, T., Rückert, C., Uria, A.R., Helf, M.J., Takada, K., Gernert, C., Steffens, U.A., Heycke, N. & Schmitt, S. (2014) 'An environmental bacterial taxon with a large and distinct metabolic repertoire', *Nature,* vol.506, no.7486, pp.58-62.

Wong, W.R., Oliver, A.G. & Linington, R.G. (2012) 'Development of antibiotic activity profile screening for the classification and discovery of natural product antibiotics', *Chemistry & Biology,* vol.19, no.11, pp.1483-1495.

Woodruff, H.B. (2014) 'Selman A. Waksman, winner of the 1952 nobel prize for physiology or medicine', *Applied and Environmental Microbiology*, vol.80, no.1, pp.2-8.

Xiong, Z., Wang, J., Hao, Y. & Wang, Y. (2013) 'Recent advances in the discovery and development of marine microbial natural products', *Marine Drugs,* vol.11, no.3, pp.700-717.

Xu, W., Chooi, Y.H., Choi, J.W., Li, S., Vederas, J.C., Da Silva, N.A. & Tang, Y. (2013) 'LovG: the thioesterase required for dihydromonacolin L release and lovastatin nonaketide synthase turnover in lovastatin biosynthesis', *Angewandte Chemie (International ed.in English),* vol.52, no.25, pp.6472-6475.

Yamanaka, K., Reynolds, K.A., Kersten, R.D., Ryan, K.S., Gonzalez, D.J., Nizet, V., Dorrestein, P.C. & Moore, B.S. (2014) 'Direct cloning and refactoring of a silent

lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A', *Proceedings of the National Academy of Sciences of the United States of America,* vol.111, no.5, pp.1957-1962. .

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. & Madden, T.L. (2012) 'Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction', *BMC Bioinformatics,* vol.13, pp.134-2105-13-134.

Yu, D., Xu, F., Zeng, J. & Zhan, J. (2012) 'Type III polyketide synthases in natural product biosynthesis', *IUBMB Life,* vol.64, no.4, pp.285-295.

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,* vol.107, no.2, pp.567-575.

Ziemert, N., Podell, S., Penn, K., Badger, J.H., Allen, E. & Jensen, P.R. (2012) 'The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity', *PLoS One,* vol.7, no.3, pp.e34064.

Zobel, S., Kumpfmüller, J., Süssmuth, R.D. & Schweder, T. (2015) '*Bacillus subtilis* as heterologous host for the secretory production of the non-ribosomal cyclodepsipeptide enniatin', *Applied Microbiology and Biotechnology*, vol.99, pp.681-691.