Screening Bacterial Symbionts of Marine Invertebrates for Ribosomally Synthesized Natural Products

Thea Hanekom

A thesis submitted in partial fulfilment of the requirements for the degree of Magister Scientiae in the Department of Biotechnology, University of the Western Cape.

July 2016

Supervisor: Professor Marla Trindade

Abstract

Pharmaceutical research and development strategies rely on the constant discovery of novel natural products as potential drugs. Recent studies have shown that the microorganisms associated with sponges are the true producers of some previously isolated compounds. This study created a large collection of bacterial symbionts associated with the South African marine sponge, *Hamacantha esperioides*. The bioactivity assays performed, showed that 44 isolates produced compounds with antimicrobial or anti-inflammatory activity. The successful identification of novel species that produce potential natural products highlights the importance of cultivation-dependent methods. To further screen for natural products, a cultivation-independent approach was used. A sequenced-based method, based on the biosynthetic genes of polytheonamide, was developed to screen for proteusins in sponge metagenomic DNA and the genomes of bacterial symbionts. The degenerate primers could amplify the targeted genes from DNA known to contain homologues. Evaluation of the primers' specificity showed non-specific amplification of genes, some containing similar conserved domains as the target genes. This study demonstrated that the use and development of cultivation-dependent and -independent screens are important for the discovery of novel natural products from the symbiotic bacteria of South African sponges.

Key words: metagenomics, screen, symbiotic bacteria, sponges, ribosomal peptides

Declaration

I declare that '*Screening Bacterial Symbionts of Marine Invertebrates for Ribosomally Synthesized Natural Products'* is my own work, that it has not been submitted before for any degree or assessment in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by means of complete references.

Thea Hanekom

_____ July 2016

Acknowledgements

I would like to express my appreciation to the following people:

Prof. Marla Trindade, I am sincerely grateful for your professional guidance and support towards the successful completion of this study;

The management team of IMBM: Dr. Heide Goodman, Dr. Bronwyn Kirby-McCullough and Lonnie van Zyl, as well as previous and current members of IMBM, for your support, advice and camaraderie at the institute;

My family and friends, for all your love and support;

Gerdus Brand, thank you for your loving patience, motivation and support throughout the writing of this thesis;

My mentors at CSIR, Dr. Petrus van Zyl and Dr. Michael Crampton, for encouraging me to pursue this degree.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

Psalm 136: 1 "Give thanks to the Lord, for he is good! His faithful love endures forever."

Abbreviations

Media

Modified 172F	172
Actinomyces isolation media	AIM
Activated charcoal media	ACM
Blue-green 11 medium	BG11
Complex carbon agar	CCA
Streptomyces medium (modified DSMZ Media 65)	GYM
Lysogeny broth	LB
Medium A agar	MMA
Oatmeal agar (ISP Med 3) - modified	OMA
Planctomycetes medium	PMM
Pringsheim's cyanobacteria specific media	PCS
Reasoner´s 2 agar	R2A
Seawater Agar	SWA
Shivji NA ¼ strength	SNA
Sponge extract medium	SEM
Tryptocase soya agar pH 4.5	TS4
Tryptocase soya agar	TSA
Tryptone yeast extract medium	TY
ZoBell agar ¼ strength	ZBA
ZoBell agar ¼ strength with crystal violate solution (0.00001%)	ZBV

Other

ATP-binding cassette	ABC
Ampicillin	Amp
Base pairs	bp
Basic Local Alignment Search Tool	BLAST
Pre-computed BLAST of proteins	BLink
Conserved domain database	CDD
Colony forming units	CFU

Clusters of orthologous groups	COG
Hexadecyltrimethylammonium bromide	CTAB
Dimethyl sulfoxide	DMSO
Deoxyribonucleic acid	DNA
Double stranded DNA	dsDNA
Genomic DNA	gDNA
Metagenomic DNA	mDNA
Ethylenediaminetetraacetic acid	EDTA
Formate dehydrogenase	FDH
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	HEPES
High molecular weight	HMW
Identity	ID
Institute for Microbial Biotechnology and Metagenomics	IMBM
The International Union of Pure and Applied Chemistry	IUPAC
Kanamycin	Kan
Kilo base pairs	kb
Lanthionine	Lan
Methyllanthionine	MeLan
Linear azole-containing peptide(s)	LAP(s)
Low melting temperature	LMP
Maximum parsimonious	MP
Nif11 nitrogen fixing protein(s)	N11P(s)
Not available (or adjusted)	NA
No bands	NB
National Center for Biotechnology Information	NCBI
Next-generation sequencing	NGS
Nitrile hydratase(s)	NHase(s)
Nitrile hydratase-related leader peptide(s)	NHLP(s)
Non-redundant	nr
Non-redundant nucleotide	nr/nt
Non-ribosomal peptide synthetase(s)	NRPS(s)
Optical density (at 600 nm)	OD_{600}

One-strain-many-compounds	OSMAC
Phenol:chloroform:isoamyl alcohol solution	PCI
Protein cluster accession number	PCLA
Polymerase chain reaction(s)	PCR(s)
Protein database	PDB
Pulse field gel electrophoresis	PFGE
Polyketyde synthase(s)	PKS(s)
Pyrroloquinoline	PQQ
Polyvinylpyrrolidone	PVP
Polyvinylpolypyrrolidone	PVPP
Query coverage	QC
Reverse complement	RC
Ribonucleotide-diphosphate reductase	RDR
Reference sequences (NCBI)	RefSeq
Ribosomally synthesized and post-translationally modified peptide(s)	RiPP(s)
Ribonucleic acid	RNA
Radical S-adenosylmethionine	rSAM
Room temperature	RT
S-adenosylmethionine	SAM
Self-contained underwater breathing apparatus	SCUBA
Sodium dodecyl sulphate	SDS
Standard melting temperature	SMP
Spectinomycin	Spc
Subtree-pruning-regrafting	SPR
Tris-acetate-EDTA	TAE
Tris-borate-EDTA	TBE
Tris-EDTA	TE
Thiazole/ oxazole-modified microcins	TOMM(s)
Theonella swinhoei chemotype Y	TSY
World Wide Fund for Nature	WWF
Centrifugal force	xg

List of Tables and Figures

Tables

protein cluster, the taxonomical groups in which the protein cluster is conserved (NCBI databases) and the function of the protein according to the clusters of orthologous groups (COGs) were recorded. Proteins of the same colour are homologous to each other.

Table 25 Comparison of the binding efficiency of the primers to the Ocin1 amplicons.

Figures

Figure 1 The core genes in a polyketide or non-ribosomal peptide biosynthetic pathway encode modules, called non-ribosomal peptide synthetases or polyketide synthases. These modules consist of different combinations of domains with specific enzymatic functions for the biosynthesis of a non-ribosomal peptide or polyketide. This figure was adapted from Finking and Marahiel (2004**).**

Figure 2 The gene cluster (top) of melithiazol (bottom) includes PKSs, NRPSs and hybrid modules (Silakowski et *al.*, 1999).

Figure 3 Activation of the post-translational modification enzymes can be catalysed when the leader peptide (red coil) is recognized by the enzymes. The leader peptide also helps to orientate the core peptide (blue) to ensure that post-translational modifications occur at the correct positions (adapted from Yang & van der Donk, 2013).

Figure 4 Examples of lanthipeptides with a conserved methyllanthionine ring (red; Patton et *al.*, 2008).

Figure 5 Cypemycin and grisemysen are the only members, so far, of linaridins and have a similar structure to lanthipeptides as they contain a thioether cross link moiety (red; Claesen & Bibb, 2011).

Figure 6 Microviridin A (1), B (2) and C (3) are tricyclic depsipeptides with ω-ester (red) and ω-amide bonds (blue). This figure was adapted from (Murakami et *al.*, 1997).

Figure 7 Microcin J25 is also classified as a lasso peptide (Section 2.2.6.; Oman & van der Donk, 2009).

Figure 8 The general topology of lasso peptides consists of a ring that is formed with an isopeptide bond between the eighth or ninth residue (arrows) and the N-terminal residue of the peptide. The C-terminal of the peptide is threaded through the ring. The structure of lasso peptides is stable as the tail is trapped within the ring. Figures acquired from Knappe *et al.* (2008) and Maksimov *et al.* (2012).

Figure 9 Molecular structure of the LAP, microcin B17 (Oman & van der Donk, 2010). These peptides (and other TOMMs) are distinguished by their thiazole (blue) and oxazole (red) heterocycles.

Figure 10 The thiopeptides thiostrepton A (A), nosihepeptide (B) and GE2270A (C), with pyridine, hydroperidine or dehydropiperidine at the cores (red) of the peptides (Kelly et *al.*, 2009).

Figure 11 The cyanobactins, patellamide A and B, was the first members of their class to have their biosynthetic pathway characterised (Schmidt *et al.*, 2005). These peptides, like other TOMMs, are distinguished by their thiazole (blue) and oxazole (red) heterocycles.

Figure 12 The primary (A) and three dimensional structure of subtilosin A (B). The peptides in this class are distinguished by their head-to-tail cyclisation (N-C linkage, absent in B) and sulphur-alpha-carbon linkages (red lines) between Cys and other residues (Kawulka *et al.*, 2003).

Figure 13 The glycosin, sublancin, is glycosylated with a S-link to the free Cys residue of the peptide (red; Wang & van der Donk, 2011).

Figure 14 The structure of bottromycin A2 (A) and bottromycin D (B) with a difference in residue 13 (red). This figure was adapted from Shimamura *et al.* (2009) and Hou *et al.* (2012). These peptides are also modified to *contain* non-proteinogenic residues, such as *tert-Leu* (*tLeu*; blue).

Figure 15 Primary structure of polytheonamide A and B. There is a stereochemical difference between the two peptides at the 44th residue, a β,β dimethylmethionine sulfoxide (red). Adapted from Hamada *et al.* (2005).

agarose gel was prepared and a part of it was replaced with low melting point agarose gel (LMP).

Figure 25 Polytheonamides biosynthetic cluster and putative proteusin pathways of '*Entotheonella factor* TSY1 and TSY2'. This figure is adapted from the Extended Data Figure 4 of Wilson *et al.* (2014).

Figure 26 A map of the polytheonamide biosynthetic gene cluster (A) along with the hypothetical proteusin pathways on the genomes of other organisms (B-I). A) '*E. factor* TSY1' (KI932682.1), B) *Bradyrhizobium* japonicum USDA 6 (NC_017249.1), C) *Anabaena* variabilis ATCC 29413 (CP000117.1), D) *Desulfarculus* baarsii DSM 2075 (NC_014365.1), E) *Microcoleus* sp. PCC 7113 (NC_019738.1), F) Nostoc sp. PCC 7120 (NC_003272.1), G) *Rhizobium* leguminosarum bv. trifolii WSM2304 (NC_011368.1), H) *S.* meliloti SM11 (CP001830.1), I*) Stigmatella* aurantiaca DW43-1 (NC_014623.1). The open reading frames (arrows) are coded according to the predicted function of the product.

Figure 27 Class B radical SAM methyltransferases catalyse the C-β-methylation (highlighted in red) at un-activated carbons of RiPPs (Dunbar & Mitchell, 2013). These rSAM proteins include PoyB and PoyC with a cobalamin-binding domain (CBD).

Figure 28 An alignment of PoyA with the representative proteins of the protein family, TOMM_pelo (accession: TIGR03793), that include *proteins* from *A.* variabilis, *Nostoc spp.,* Pelotomaculum thermopropionicum and other species. The alignment was performed with the BioEdit Sequence Alignment Editor (Hall, 1999) and the Clustal W algorithm (Thompson *et al.,* 1994) and the similarity threshold was set to 80 % (highlighted residues).

Figure 29 Conserved domain hits (from NCBI) of PoyB (top) and PoyC (bottom). The arrows in red indicate the regions to which the Ocin1 primers were designed. The jagged edges for some of the conserved domains indicate only a partial hit to the query.

Figure 30 Clade A of the maximum parsimonious (MP) tree of the representative proteins of the rSAM_ocin_1 family (full tree is available in Appendix C; Figure 1). The presented clade is that of PoyB and PoyC (Clade C) and the seven closest related

proteins (Clade B). Analyses were conducted in MEGA5 (Tamura *et al.*, 2011). The reference sequences (RefSeq; NCBI), denoted with the prefix 'WP_' in their accession numbers, were used as the representatives of identical sequences.

Figure 31 An alignment of the rSAM_ocin_1 proteins (accession: TIGR03975) that are closely related to PoyB and PoyC, together with the bottromycins' rSAM_ocin_1 proteins. The alignment was performed with the BioEdit Sequence Alignment Editor (Hall, 1999) and the Clustal W algorithm (Thompson *et al.*, 1994) and the similarity threshold was set to 80 % (highlighted residues). The conserved region in the black blocks corresponds to the region selected for the design of the primers (Ocin1F and R). The conserved domains for PoyB (grey bars) together with the B12-binding sites or FeS/SAM binding sites (black arrows) are indicated.

Figure 32 An alignment of the rSAM_ocin_1 proteins (accession: TIGR03975) that are closely related to PoyB and PoyC, bottromycins' rSAM_ocin_1 proteins and representative proteins of the B12_rSAM_oligo family. Alignment was performed with the BioEdit Sequence Alignment Editor (Hall, 1999) and the Clustal W algorithm (Thompson *et al.*, 1994) and the similarity threshold was set to 80 % (highlighted residues). The conserved region in the black blocks corresponds to the region selected for the design of the primers (Ocin1F and R), as in Figure 8.

Figure 33 Gradient PCR of Ocin1 primers with *poyB-plasmid* DNA as the template. The annealing temperature for each lane was as follows: 1) 55.0 °C, 2) 55.6 °C, 3) 56.1 °C, 4) 57.4 °C, 5) 58.9 °C, 6) 60.3 °C, 7) 61.0 °C, 8) 62.3 °C, 9) 63.4 °C, 10) 64.4 °C, 11) 64.7 °C, 12) 65.0 °C. The negative control (lane 13), where no template DNA was added, had an annealing temperature of 60.3 °C. M1) λ phage DNA digested with the endonuclease PstI was used as the DNA ladder.

Figure 34 Touchdown PCRs with DreamTaq (Lanes 1-9) and Phusion (Lanes 10-12) polymerase and 2 mM MgCl2 at annealing temperatures (Ta): 58 to 49 °C (Lanes 1-3), 60 to 51 °C (Lanes 4-6 and 10-12) and 62 to 53 °C (Lanes 7-9). The template DNA used for Lanes 1, 4, 7 and 11) *S. meliloti;* Lanes 2, 5, 8 and 12) *poyB-plasmid* and Lanes 3, 6, 9 and 10) no DNA. M1) 1 kb; M2) λ PstI and M3) 100 bp DNA markers.

Figure 35 Touchdown PCRs with DreamTaq polymerase and 3 mM MgCL2 at annealing temperature (Ta): 58 to 49 °C (Lanes 1-3), 60 to 51 °C (Lanes 4-8) and 62 to 53 °C (Lanes 9-13). Template DNA added: *S. meliloti* (Lanes 1, 4, 6, 10 and 12) and *poyB-plasmid* (Lanes 2, 5, 7, 9 and 11). The DNA ladders are as follows: M1) 1 kb; M2) λ PstI and 100 bp.

Figure 36 Different attempts of extracting mDNA with the modified urea lysis method. Lanes 1-4) the extracted mDNA. Lanes 5-6) the mDNA, from lanes 3 and 4, after purification with PVPP. The DNA markers are M1) λ DNA, about 48.5 kb; M2) λ Hind III and M3) λ PstI.

Figure 37 The validation PCR of the extracted mDNA with the universal 16S rRNA primers. Lanes 1, 3, 5 and 7) mDNA at concentrations: undiluted, 10 times diluted, 50 times diluted and 100 times diluted. Lanes 2, 4, 6 and 8) spiked-mDNA (mDNA with gDNA of *D. maris*). The spiked-mDNA was diluted the same as the mDNA. Lane 9) control reaction with gDNA of *D. maris* as template and lane 10) no template DNA. DNA marker λ PstI (M1).

Figure 38 Size selection of the DNA purified with formamaide. The fragments greater than 23 kb in size was cleaned (A) as well as fragments between approximately 9 to 23 kb (B).

Figure 39 Phusion PCR with 16S rRNA primers: E9F with U1510R (lanes 1 and 4), Ento271F with Ento1290R (lanes 3 and 6) and Ento735F with Ento1290R (lanes 2, 5 and 7). Sample PE05-A was added as template to the reactions in lanes 1, 3 and 7. Lane 2) A nested PCR was performed with the products of a PCR reaction mixture that was the same as lane 3. Lanes 4-6) Control reactions with *D. maris* gDNA as template. DNA marker λ PstI (M1). The control PCR without template DNA was electrophoresed on another gel, and no amplicons were produced.

Figure 40 Touchdown PCRs with DreamTaq and 3 mM MgCL2 and Phusion polymerase. Annealing temperature range used 58 to 49 °C. Template DNA added: Group 1 to 4 (lanes 1-4), *S. meliloti* (lane 5) and *poyB-plasmid* (Lane 6) and no DNA (lane 7). The 1 kb DNA ladder (M1) was used. Refer to Table 12 for the individual isolates included in the different groups.

Figure 41 DreamTaq touchdown PCR with annealing temperature ranges 58 to 49 °C (A-D) and 62 to 53°C (E and F). The number above each lane is the PE14-isolate number. The control reactions (green) with *poyB-plasmid* (PoyB) and gDNA from *S. meliloti* (Sm) where the 400 bp fragment represents the partially amplified rSAM_ocin_1 homologous gene. No template DNA added (Neg). DNA markers included M1) λ PstI; M2) 100 bp and M3) 1 kb. The negative control in B is the control for experiment A and B. Similarly, the negative control in E was for experiment E and F. The region between 250 to 600 bp (C and D) was cleaned and used in a second PCR with Ocin1 primers and selected fragments (white blocks) were sequenced and analysed. Unfortunately the loading dye used for some of the samples in A and B masked some of the DNA bands (bright lines at 1 500 and 250 bp region).

Figure 42 The iron-sulphur binding regions of poyB, poyC, (A and B) formate dehydrogenase subunit alpha (FDH; WP_017111042.1) and (C and D) ribonucleotide-diphosphate reductase (RDR; WP_011261821.1). The proteins and gene was aligned with the reverse-compliment sequence of primer Ocin1R. Alignment was performed with the BioEdit Sequence Alignment Editor (Hall, 1999) and the Clustal W algorithm (Thompson *et al.*, 1994) and the similarity threshold was set to 80 % (highlighted residues). Below the alignment of the nucleotides, the binding consensus is illustrated with a colon for an exact match and a period for a match with an ambiguous sequence.

Figure 43 Alignment of the sequenced amplicons with the DNA sequence of the top hit (accession numbers are as in Table 23). The binding sites of the primers are also indicated. Alignment was performed with the BioEdit Sequence Alignment Editor (Hall, 1999) and the Clustal W algorithm (Thompson *et al*., 1994). Matches in the gene, amplicon and primers are highlighted in yellow. RC indicates that the reverse compliment of a sequence was used.

Table of Contents

Chapter 1 – Literature Review

1. The need for natural product discovery and drug development

Humankind has for centuries exploited the medicinal properties of natural products by using them as traditional medicines. Plant extracts were used as the basis for many of these remedies. Classical natural product discovery relied on the analysis of these traditional medicines to identify the active components (Harvey, 2000). In addition to the analysis of traditional medicines, natural product discovery also depended on serendipitous findings.

Biotechnology is essentially the exploitation of organisms, or the biochemical compounds or proteins that they produce, to the benefit of various industries. The biochemical compounds are referred to as natural products. This review will focus on the discovery and development of pharmaceutically relevant natural products of microbial origin.

Natural products are produced by organisms across all three domains of life, namely eukaryotes, bacteria and archaea. So far the main source of natural products is terrestrial organisms, especially plants, fungi and soil bacteria. Fungi are well known producers of natural products such as the non-ribosomal peptides, penicillin and cephalosporin (Walsh & Fischbach, 2010). The golden era for drug discovery from microorganisms, especially for antibiotics, began after the discovery of penicillin (Li & Vederas, 2009; Silver, 2011). This era reached its peak in the 1960s and then followed a rapid decline. By the 1980s novel compounds were rarely discovered (Light, 2009; Silver, 2011).

The majority of the drugs available today are synthetic analogues or derivatives of natural products (Li & Vederas, 2009). This can be attributed to the limited supply of some natural products for research and clinical trials, as well as to the difficulty in obtaining natural products with acceptable pharmacological properties (Schmidt *et al.*, 2005). When the structures of natural products are known, the knowledge can be used as the foundation for combinatorial chemistry (Newman & Cragg, 2007) or combinatorial biosynthesis (Floss, 2006) to produce alternative products.

The number of effective antimicrobials is decreasing as more strains of common diseases exhibit resistance to even the antimicrobials that are considered to be a last resort. This necessitates the need to discover and develop new antimicrobial

compounds. Natural products are not only important for the development of antimicrobials, but they are also important for the development of other pharmaceuticals such as anti-inflammatory, neuro-active (Kokel *et al.*, 2010), antiviral (Férir *et al.*, 2013) and antimalarial drugs (Ang *et al.*, 2000). The discovery of novel natural products is also vital since no new chemical classes of drugs have been approved as pharmaceuticals since the 1980s (Silver, 2011). As a consequence, researchers have started to focus on the marine environment, which is an enormous source of biodiversity, but has not yet been exhaustively exploited as a source of natural products for the development of pharmaceuticals.

2. Classification of natural products

Prominent classes of natural products include alkaloids, terpenoids, non-ribosomal peptides, polyketides, as well as ribosomally synthesized and post-translationally modified peptides (RiPPs). Research efforts in the field of natural products has focussed primarily on polyketides and non-ribosomal peptides since they represent a substantial number of important pharmaceuticals (Cane, 1997).

2.1. Polyketides and non-ribosomal peptides

Polyketides and non-ribosomal peptides have a similar system of biosynthesis that is based on chain elongation. The core synthesis of the molecule is catalysed by transferring the molecule along consecutive multi-modular enzymes, referred to as polyketide synthases (PKSs) or non-ribosomal peptide synthetases (NRPSs). A PKS or NRPS module is a single polypeptide that consists of at least two individually folded protein domains, where each domain functions as an independent enzyme (Figure 1). Various auxiliary tailoring enzymes can modify the functional group of the new chain before it is passed onto the next module, or they modify the assembled chain into the mature compound. The auxiliary enzymes are encoded together with the multi-modular enzymes within the biosynthesis cluster of the natural product (Walsh, 2004; Finking & Marahiel, 2004).

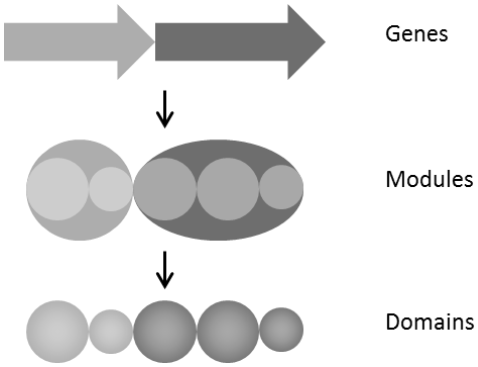


Figure 1 The core genes in a polyketide or non-ribosomal peptide biosynthetic pathway encode modules, called non-ribosomal peptide synthetases or polyketide synthases. These modules consist of different combinations of domains with specific enzymatic functions for the biosynthesis of a non-ribosomal peptide or polyketide. This figure was adapted from Finking and Marahiel (2004).

Due to the similarity in the structures of PKSs and NRPSs, there are molecules produced by hybrids of these enzymes, for example, melithiazol (Figure 2), which is passed between PKS and NRPS domains and hybrids thereof, during biosynthesis (Silakowski *et al.*, 1999).

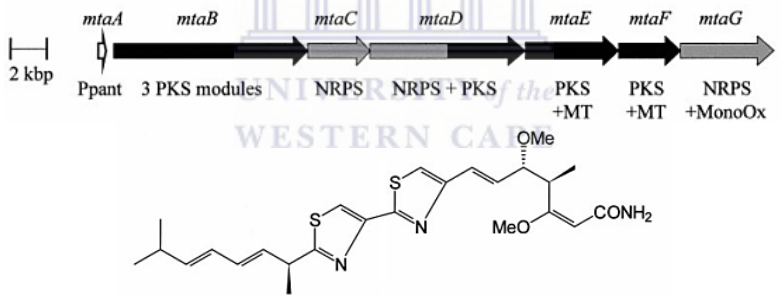


Figure 2 The gene cluster (top) of melithiazol (bottom) includes PKSs, NRPSs and hybrid modules (Silakowski *et al.*, 1999).

Polyketides are isoprenoids with biosynthetic pathways similar to fatty acids. Fatty acids and polyketides are synthesized from an acetyl coenzyme A molecule, but unlike fatty acids, polyketides can be synthesized from derivatives thereof (Ridley & Khosla, 2009). The PKSs catalyse the synthesis of the polyketides through a condensation reaction of the acyl thioester precursors that result in the formation of ketides. Polyketide synthesis can be classified into three groups, PKS I to III, based on the organisation of the gene cluster, as well as the structure and the function of the synthesising enzymes (Ridley & Khosla, 2009). It is important to note that type II PKSs, which synthesise aromatic polyketides, differs from the other two types. Type II PKSs

use the enzymatic modules repeatedly (Cane, 1997; Cane & Walsh, 1999), while a molecule synthesized by type I and III PKSs is passed in one direction along the modules. PKSs commonly function in conjunction with other enzymes such as fatty acid synthases, terpene synthases or, as mentioned before, NRPSs.

Non-ribosomal peptides are synthesized through the condensation of activated amino acids (Finking & Marahiel, 2004; Li *et al.*, 2009). There are five important domains for the elongation of the peptide. The condensation domain, the adenylation-domain, peptidyl carrier protein, thioesterase-domain and the Cy-domain (Finking & Marahiel, 2004). As in polyketide synthesis, non-ribosomal peptide synthesis relies on tailoring enzymes that modify the peptide into its mature structure.

2.2. Ribosomally synthesized and post-translationally modified peptides (RiPPs)

The scope of this review includes only classes of RiPPs produced by bacteria, as the research focus of this thesis will be to screen for RiPPs produced by sponge-associated bacteria. Information for the non-bacterial classes (toxoglossan venom peptides, amotoxins, phallotoxins, cyclotides, orbitides, pyrroloquinoline (PQQ), pantocin and the thyroid hormones T3 and T4) can be found in the review by Arnison *et al.* (2013).

Natural products that are synthesized through ribosomes and undergo post-translational modifications are collectively called RiPPs. There is also a restriction of only including peptides that are not larger than 10 kDa (Arnison *et al.*, 2013). It is important to note that due to a lack of consensus in the description and classification of RiPPs (McIntosh *et al.*, 2009; Arnison *et al.*, 2013), there are many terms within this class of natural products that describe the same concept. An example thereof is the core peptide of RiPPs that eventually becomes the mature and active peptide. It is sometimes referred to as the structural peptide, propeptide or prepeptide (Oman & van der Donk, 2010). Another example is the use of bacteriocins. Defining bacteriocins is not a new notion as Jack *et al.* (Jack *et al.*, 1995) already in 1995 explained: *"bacteriocins comprise a rather ill-defined potpourri of proteinaceous molecules"*. Bacteriocins are a collective name for antimicrobial peptides originally thought to be only produced by Gram negative bacteria, such as the Enterobacteriaceae family, but they are found to be also

produced by Gram positive lactic acid bacteria. Bacteriocins include, for example, members of the RiPP classes such as lanthipeptides (Zhang *et al.*, 2012), glycosins (Wang & van der Donk, 2011), sactipeptides and thiopeptides (Kelly *et al.*, 2009). Along with the confusion caused by terminology, there is also the matter of peptides being classified into more than one of the RiPP classes. Examples are microcins B17 and J25. These compounds are classified as microcins (Section 2.2.5.) as well as linear azole-containing peptides (LAPs, in the case of micorcin B17; Section 2.2.7.1.) or as lasso peptides (in the case of micorcin J25; Section 2.2.6.).

The interest in researching RiPPs as pharmaceutical drugs has increased over the last 15 years, mainly due to the availability of sequencing techniques (Arnison *et al.*, 2013). Another driving factor for the increase in research is the structural diversity that RiPPs display, with more than 17 classes identified thus far (Arnison *et al.*, 2013). Some of the diversity is brought about by post-translational modifications that are so different to traditional ribosomal peptide synthesis that the active product is sometimes not recognised as being a product of ribosomal synthesis (Kersten *et al.*, 2011).

2.2.1. General structure and synthesis of RiPPs

The biosynthetic genes of RiPPs, arranged within an operon, are involved in the synthesis, modification, export and regulation of RiPPs. They include precursor gene(s) of the peptide, maturation enzymes and auxiliary genes. Examples of maturation enzymes are cyclohydratases, dehydrogenases, radical S-adenosylmethionine dependent methyltransferase (rSAMs), thiazole-oxazole modification enzymes and lanthionine synthetases. Auxiliary genes include those that encode a docking scaffold, ATP-binding cassette (ABC) transporters and anti-toxin genes with an immunity function for the producing organism. With the sequencing of RiPP biosynthetic pathways it was noticed that the auxiliary enzymes are shared between different classes (Wieland Brown *et al.*, 2009; Freeman *et al.*, 2012). It should be noted that there are instances where not all of the biosynthetic genes are encoded within the operon (Haft *et al.*, 2010; Kelly *et al.*, 2009).

Precursor peptides are about 20 to 60 amino acid residues in length (Kawulka *et al.*, 2003). For most of the RiPP classes the precursor can be divided into the leader and

core peptide. The core peptide will be modified by the post-translational enzymes into the active natural product. This is flanked by a leader peptide at the N-terminus. However, there are exceptions to this rule, for example, the leader peptide of bottromycins (Arnison *et al.*, 2013) seems to have been shifted to the C-terminus of the core peptide. The sequence of leader peptides is usually highly conserved for all members of the same class of RiPP. Although the functions of many leader peptides are still elusive, they are known to assist the post-translational process (Figure 3) by activating the maturation enzyme and orientating the core peptide (Patton *et al.*, 2008; Yang & van der Donk, 2013; Dunbar & Mitchell, 2013; Knerr & van der Donk, 2012).

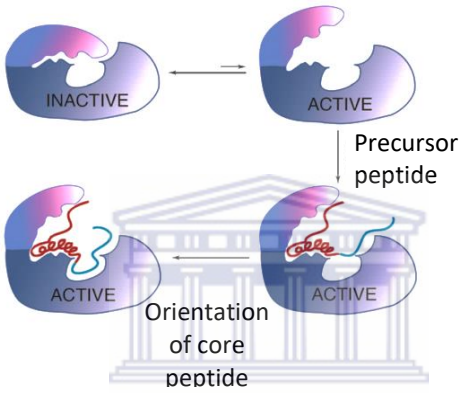


Figure 3 Activation of the post-translational modification enzymes can be catalysed when the leader peptide (red coil) is recognized by the enzymes. The leader peptide also helps to orientate the core peptide (blue) to ensure that post-translational modifications occur at the correct positions (adapted from Yang & van der Donk, 2013).

Various maturation enzymes will modify the structure of the peptide either at the molecular level, by modification of a functional group, or at the macromolecular level, by forming for example rings (cyaobactins, thiopeptides, etc.) or even knots (lassopeptides). The peptide is activated after post-translational modifications, for example, when homologues of ABC transport proteins activate the core peptide by cleaving the leader peptide from the core peptide as it is exported from the cell (McIntosh *et al.*, 2009).

One of the main differences between peptides synthesized by the ribosome and NRPSs lies with the monomers that are utilised. For ribosomal peptides the organisation of the monomers, the proteinogenic amino acids, are genetically encoded and transcribed. RiPPs could contain non-proteinogenic amino acids, but they are not incorporated

directly during transcription. The presence of non-proteinogenic amino acids are rather due to post-translational modifications of proteinogenic amino acids. For non-ribosomal peptides it is only the modules (NRPSs) that are genetically encoded. The organisation of the monomers is then inferred from the modules and the monomers incorporated are not restricted to only the proteinogenic amino acids (Caboche *et al.*, 2010).

A short introduction on the main characteristics of the bacterial RiPP classes will now be presented.

2.2.2. Lanthipeptides

Lanthipeptides have been researched for more than 50 years and are the archetypes of RiPPs. They were formerly known as lanthionine-containing antibiotics or lantibiotics, but are now referred to as lanthipeptides because they have shown bioactivities other than antimicrobial activities. These peptides are recognized by their lanthionine or methyllanthionine thioether (Lan or MeLan) bonds (Figure 4) between Cys and dehydrated Thr or Ser residues (Flühe & Marahiel, 2013). There are four classes within this group and peptides are classified according to the enzymes involved in the catalysis of the thioether crosslink formation.

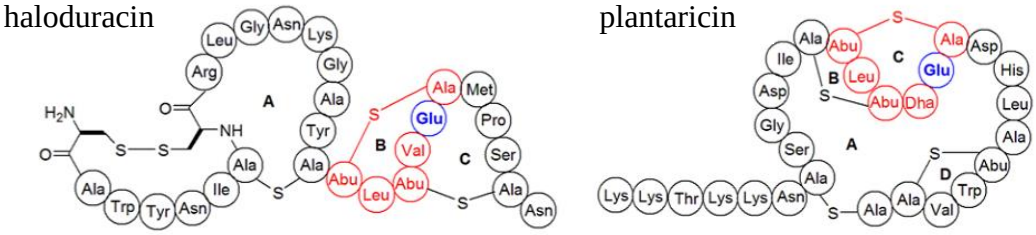


Figure 4 Examples of lanthipeptides with a conserved methyllanthionine ring (red; Patton *et al.*, 2008).

2.2.3. Linaridins

The rare peptides of this class was previously considered to be lanthipeptides since they have the same thioether crosslink moiety, but with the recent discovery of the biosynthetic gene cluster of cypemycin (Claesen & Bibb, 2010), it was shown that the moiety is not formed by the same enzymes as those of the lanthipeptides. Only two members are known in this class (Figure 5), cypemycin and grisemycin and their

biosynthesis has been elucidated by Claesen and Bibb (Claesen & Bibb, 2010; Claesen & Bibb, 2011).

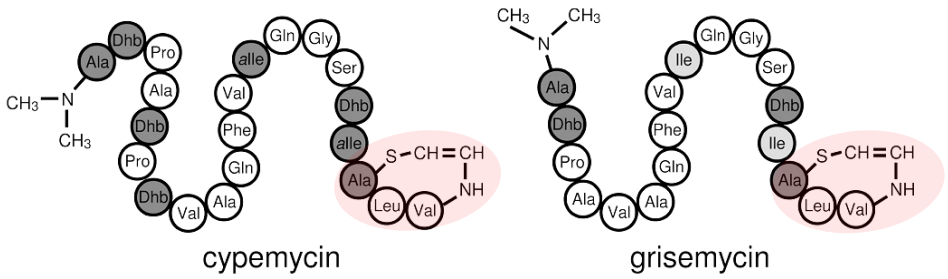


Figure 5 Cypemycin and grisemysen are the only members, so far, of linaridins and have a similar structure to lanthipeptides as they contain a thioether cross link moiety (red; Claesen & Bibb, 2011).

2.2.4. Microviridins

Microviridins are a family of tricyclic depsipeptides. They are distinguished by having two conserved ω-ester bonds between the residues Thr-Asp and Ser-Glu (Figure 6), as well as an ω-amide bond between the residues Lys-Glu (McIntosh *et al.*, 2009). They also have an unique cage-like structure which lends it high stability against proteases (Ziemert *et al.*, 2008).

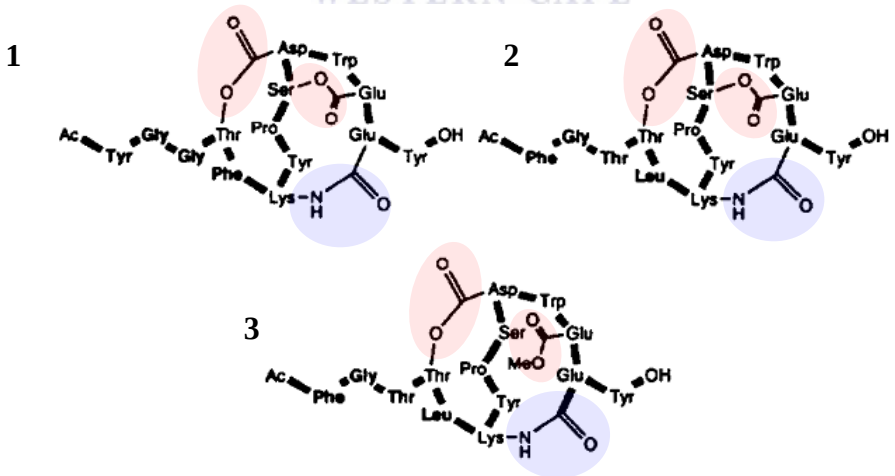


Figure 6 Microviridin A (1), B (2) and C (3) are tricyclic depsipeptides with ω-ester (red) and ω-amide bonds (blue). This figure was adapted from (Murakami *et al.*, 1997).

2.2.5. Microcins

These peptides are mainly produced by *Escherichia coli*. Peptides are classified within this class when they have a molecular mass of less than 10 kDa, which is small in

contrast to the size of other antimicrobial peptides produced by Gram-negative bacteria. There are only fourteen members in this class (Arnison *et al.*, 2013), of which the two peptides, J25 (Figure 7) and microcin B17 (Figure 9), are prominent as they are structurally related to other larger RiPP families, namely, lasso peptides (Section 2.2.6.) and LAPs (Section 2.2.7.1.).

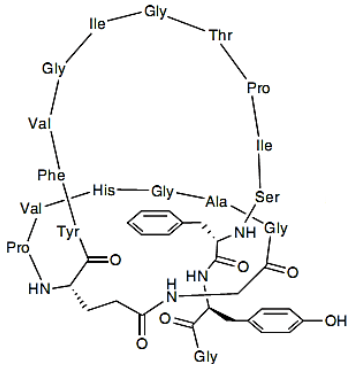


Figure 7 Microcin J25 is also classified as a lasso peptide (Section 2.2.6.; Oman & van der Donk, 2009).

2.2.6. Lasso Peptides

Lasso peptides are classified according to their exceptional knot-like structure. They consist of 16 to 21 residues. A ring-structure is created through an isopeptide bond between the N-terminal Gly and residue 8 or 9, which is either an Asp or Glu (Figure 8). The C-terminal residues are threaded through and entrapped within the ring which contributes to the stable structure of these peptides (Knappe *et al.*, 2008), known as the lasso fold.

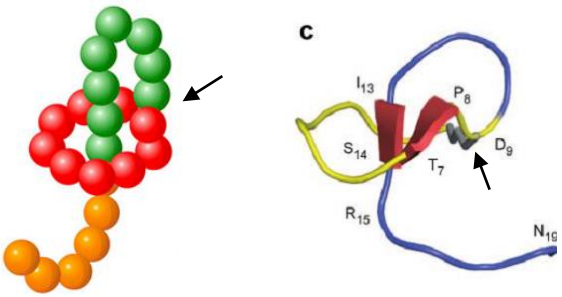


Figure 8 The general topology of lasso peptides consists of a ring that is formed with an isopeptide bond between the eighth or ninth residue (arrows) and the N-terminal residue of the peptide. The C-terminal of the peptide is threaded through the ring. The structure of lasso peptides is stable as the tail is trapped within the ring. Figures acquired from Knappe *et al.* (2008) and Maksimov *et al.* (2012).

2.2.7. Thiazole/oxazole-modified microcins (TOMMs)

These RiPPs are divided into three classes: linear azole-containing peptides (LAPs), thiopeptides and cyanobactins (Melby *et al.*, 2011). They are related by having thiazole and (methyl)oxazole heterocycles which are formed when cyclodehydratase and/or dehydrogenase catalyse the cyclisation of Cys, Ser and Thr residues within the core peptide.

2.2.7.1. Linear Azole-containing Peptides (LAPs)

The post-translational modifications by cyclodehydratase and/or dehydrogenase of the Cys, Ser and Thr residues within the core peptide of LAPs results in the characteristic thiazole and (methyl)oxazole heterocycles (Figure 9; Oman & van der Donk, 2009). Peptides within this class are also classified within other classes, as mentioned before. They are distinguished from other azol(in)e RiPPs as they are linear and do not form macrocyclised structures.

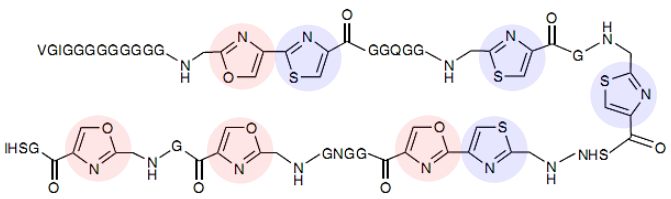


Figure 9 Molecular structure of the LAP, microcin B17 (Oman & van der Donk, 2010). These peptides (and other TOMMs) are distinguished by their thiazole (blue) and oxazole (red) heterocycles.

2.2.7.2. Thiopeptides

Thiopeptides have only recently been classified as RiPPs when four independent research groups published the biosynthetic gene clusters of a few thiopeptides (Wieland Brown *et al.*, 2009; Morris *et al.*, 2009; Liao *et al.*, 2009; Kelly *et al.*, 2009). These peptides are distinguished from other thio-containing RiPPs as they are highly modified with many non-proteinogenic residues that are the result of post-translational modification of proteinogenic residues. They have one or more macro-cycles with a nitrogenous heterocyclic core, namely pyridine, hydroperidine or dehydropiperidine (Figure 10).

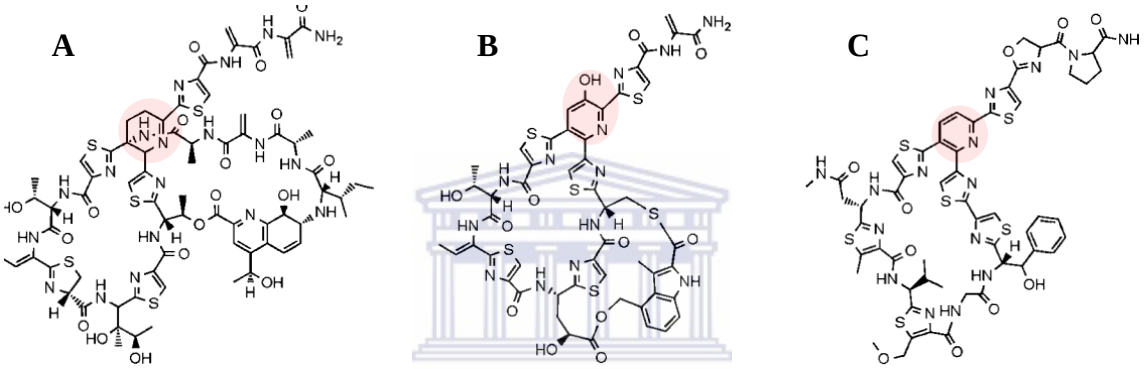
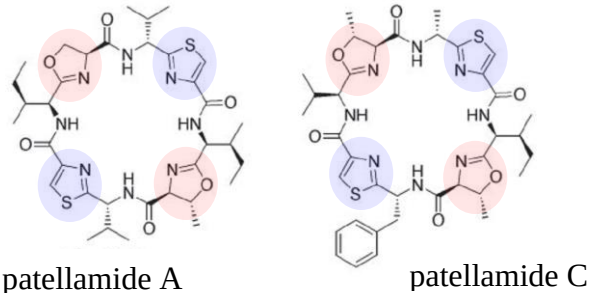


Figure 10 The thiopeptides thiostrepton A (A), nosihepeptide (B) and GE2270A (C), with pyridine, hydroperidine or dehydropiperidine at the cores (red) of the peptides (Kelly *et al.*, 2009).

2.2.7.3. Cyanobactins

The diverse group of cyanobactins are produced by members of the photosynthetic bacterial phylum, Cyanobacteria. They are distinguished as small N- to C-terminal (head-to-tail) cyclised peptides. These peptides are usually of a low molecular weight, with six to eight residues, although some have been reported to consist of up to 20 residues (Donia & Schmidt, 2011). Some of the post-translational modifications of this class include the formation of thiazole and oxazole heterocyclic moieties (Leikoski *et al.*, 2010), as mentioned before. Although hundreds of cyanobactins have been discovered through chemical extraction and cultivation-dependent methods, genome mining is becoming an effective tool for the discovery of novel cyanobactins (Donia & Schmidt, 2011; Donia *et al.*, 2008). The first biosynthetic pathway for this group was

elucidated in 2005 by Schmidt *et al.* (Schmidt *et al.*, 2005) for the production of patellamide A and C (Figure 11).



patellamide A patellamide C

Figure 11 The cyanobactins, patellamide A and B, was the first members of their class to have their biosynthetic pathway characterised (Schmidt *et al.*, 2005). These peptides, like other TOMMs, are distinguished by their thiazole (blue) and oxazole (red) heterocycles.

2.2.8. Sactipeptides

Sactipeptides are macrocyclic peptides with unique bonds between the sulphur of Cys residues and the α-carbon of another (non-conserved) residue. Although this type of bond is not found in any other RiPP classes, it is a familiar chemical feature of non-ribosomal peptides.

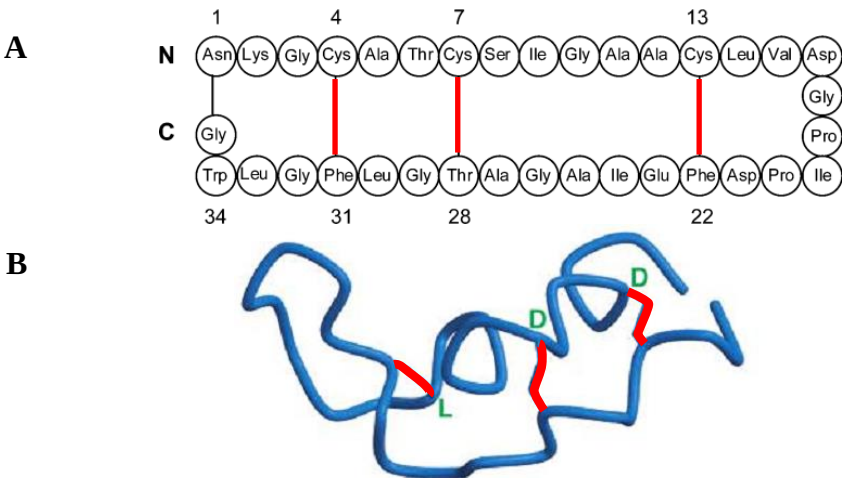


Figure 12 The primary (A) and three dimensional structure of subtilosin A (B). The peptides in this class are distinguished by their head-to-tail cyclisation (N-C linkage, absent in B) and sulphur-alpha-carbon linkages (red lines) between Cys and other residues (Kawulka *et al.*, 2003).

2.2.9. Bacterial head-to-tail cyclized peptides

Bacterial head-to-tail cyclised peptides form rigid structures that are able to withstand degradation. They are separated from other N- to C-terminal cyclised peptides as they have a large core peptide, with 35-70 amino acid residues, and different enzymes for macrocyclisation that catalyse the formation of a peptide bond (Figure 12). Their biosynthetic enzymes are divided into three groups according to their chemical properties. Another contrast to other RiPP classes are that the leader peptides are not as conserved and vary in length (van Belkum *et al.*, 2011).

2.2.10. Glycocins

The two members of the glycocins class, sublancin and glycocin F, are glycosylated at Cys residues. The glycosylation of these peptides are characterised by the bond between the sulphide of Cys and the sugar. These peptides show antimicrobial activity, however, their mechanism of action still needs to be elucidated (Arnison *et al.*, 2013). Even though these peptides are glycosylated, this chemical feature is not important for the antimicrobial activity of sublancin (Wang & van der Donk, 2011). It was found that the glycosylation of the peptide rather plays a role in the immunity of the producing organism.

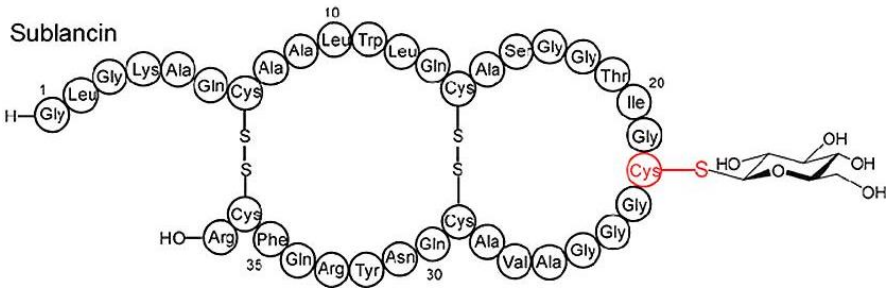


Figure 13 The glycosin, sublancin, is glycosylated with a S-link to the free Cys residue of the peptide (red; Wang & van der Donk, 2011).

2.2.11. Bottromycins

Bottromycins have unique modifications and structural characteristics. They have rare non-proteinogenic residues, such as *tert*-Leu and D-amino acids, which are also found within the proteusin class. The full structure of the first member of this family, bottromycin A_2 (Figure 14), has been elucidated by Shimamura *et al.* (2009). This is

more than 50 years after it was originally isolated in 1957 (Waisvisz *et al.*). Another distinctive characteristic of bottromycins are that they lack a leader peptide at the N-terminus of the core peptide, but rather have a follower peptide, of which the full function is unclear.

A

B

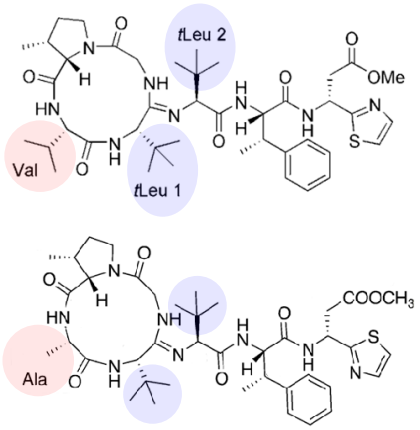


Figure 14 The structure of bottromycin A_2 (A) and bottromycin D (B) with a difference in residue 13 (red). This figure was adapted from Shimamura *et al.* (2009) and Hou *et al.* (2012). These peptides are also modified to contain non-proteinogenic residues, such as *tert*-Leu (*t*Leu; blue).

2.2.12. Proteusins

The proteusins play a central role in the research presented in this thesis and will be discussed in more detail. The only proteusins that have been isolated thus far are the polytheonamides (A, B and C; Hamada *et al.*, 2005). Polytheonamide A and B can change spontaneously into one another due to oxidation (Hamada *et al.*, 2005; Freeman *et al.*, 2012), since the only difference between them is the stereochemistry of the 44th residue (Figure 15). They were extracted from the lithistid sponge, *Theonella swinhoei*, and showed strong anti-tumour activity (Hamada *et al.*, 1994). It was the bioactive properties and the novel structure of polytheonamides (Hamada *et al.*, 2005) that captured the interest of researchers. Initially it was believed that polytheonamides were produced by a NRPS system. It was also thought that a symbiotic bacterium of the host sponge might be the true producing organisms, since other non-ribosomal peptides were produced by the symbionts of lithistid sponges (Hamada *et al.*, 2005).

Polytheonamide B consists of 48 residues. It is this great size and the high amount of post-translational modifications (Inoue *et al.*, 2010) observed, that led researchers to

think that it is the largest non-ribosomal peptide isolated. However, the ribosomal origin of the proteins were discovered recently (Freeman *et al.*, 2012). The extensive post-translational modifications (Figure 15) include the hydroxylations (4 residues) and C-methylation (21 residues). The peptide contains 18 D-amino acid residues and the non-proteinogenic residues include *β*-methylisoleucine, *β*-methylglutamine, *β*-hydroxyvaline, *γ*-*N*-methylasparagine, *tert*-leucine, *γ*-*N*-methyl-*β*-hydroxyasparagine and *β*,*β*-dimethylmethionine sulfoxide (residue 44; Hamada *et al.*, 2005; Mori *et al.*, 2007; Freeman *et al.*, 2012). The N-terminal residue is di-methylated, which is an uncommon moiety for ribosomal peptides, but it has been observed in the RiPP cypemycin (Freeman *et al.*, 2012; McIntosh *et al.*, 2009; Martin-Visscher *et al.*, 2008; Claesen & Bibb, 2010). A unique feature of polytheonamides is the numerous D-amino acids that alternate with the L-amino acids throughout the peptide (Figure 15). Polytheonamides have a β-sheet as a secondary structure (Figure 16) that forms a barrel. This barrel ultimately acts as an ion channel across membranes which weakens the resting membrane potential that contributes to the cytotoxicity of polytheonamides (Iwamoto *et al.*, 2010). This bioactivity allows them to be used as very potent anti-tumour therapeutics.

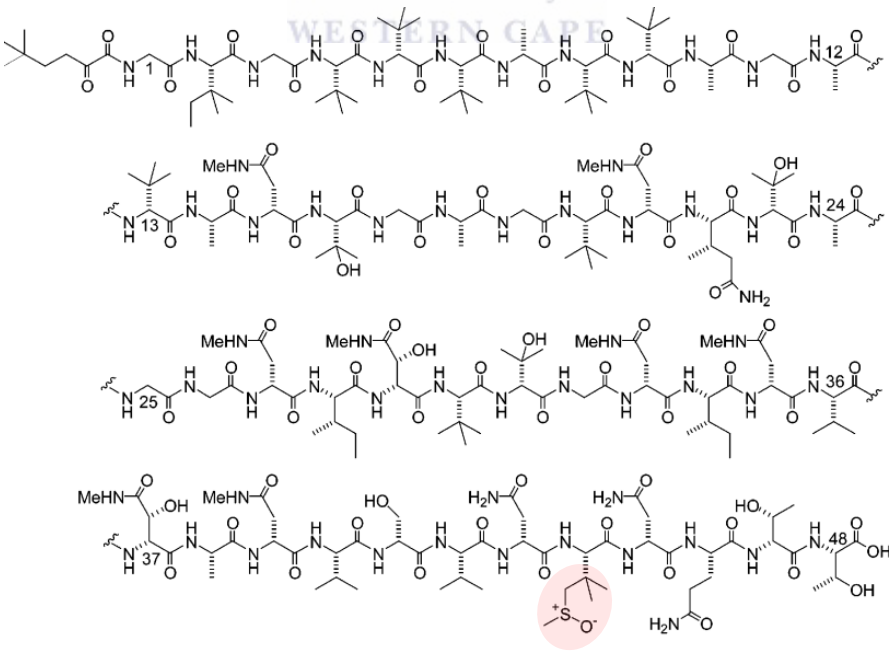


Figure 15 Primary structure of polytheonamide A and B. There is a stereochemical difference between the two peptides at the 44th residue, a β,β dimethylmethionine sulfoxide (red). Adapted from Hamada *et al.* (2005).

Polytheonamides are often compared to the non-ribosomal peptide gramicidin A as they have similar structural and functional characteristics. Both peptides are linear with alternating D- and L- amino acids. They form right-handed β-helixes (Figure 16) within environments that are similar to membranes (Mori *et al.*, 2007) and consequently functions as ion channels with selectivity for monovalent cations (Iwamoto *et al.*, 2010).

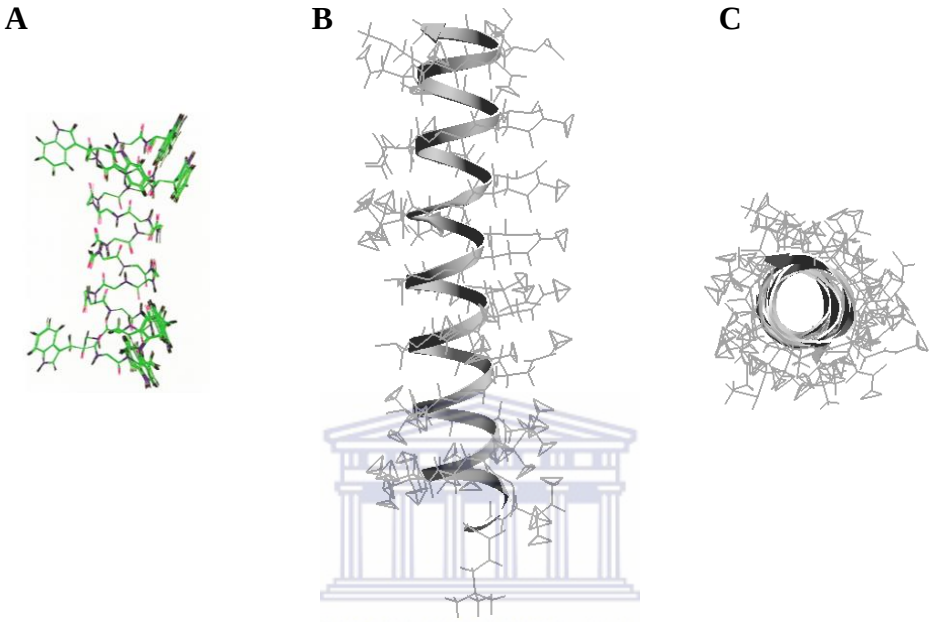


Figure 16 (A) The structure of the homo-dimer, gramicidin A (Ketchem *et al.*, 1993). (B) A three dimensional representation of the right-handed helix of polytheonamide B. (C) The top (C-terminus) view of the channel that is formed by the helical structure. (The graphics of B and C was created with the protein database (PDB), accession number: 2RQO, and the Swiss PDB-Viewer tool (Guex & Peitsch, 1997)).

3. The marine environment as a source for natural products

3.1. Marine natural product research

Marine natural products were first discovered in the 1950s (Matsunaga *et al.*, 1985; Leal *et al.*, 2012), but only started to gain popularity in the 1980s, which led to an increase in the number of discoveries (Davies-Coleman, 2005). The discoveries were primarily made by marine natural product chemists who examined the extracts of macroorganisms. Marine natural products are screened with a wide variety of bioassays to determine what biological activities they have.

The progress made with the discovery of marine natural products, and the subsequent development of drugs, was primarily dependent on time and technology. Even though

many marine natural products were discovered between the 1960s and the 1980s, there was a period from 1980 to 2000 when most pharmaceutical companies abandoned their bioprospecting projects. The discoveries made during this period were mainly due to academic research. Over the last 10 years there was a renewed interest in bioprospecting the marine environment. The new interest in bioprospecting for marine natural products is partly due to improved analytical technologies and to the development of high-throughput methods (Leal *et al.*, 2012). The improvements included advances in techniques used for the isolation, purification, structure determination and high-throughput bioactivity analysis of natural products. The refinements enabled the use of a smaller sample size for analysis, as the new equipment and techniques are more sensitive. Some of the improved techniques include chromatography, mass spectrometry and nuclear magnetic resonance (Harvey, 2008; Walsh & Fischbach, 2010). The progress made with the discovery of marine natural products also becomes evident in the type of compounds isolated. Early discoveries were of simpler non-polar compounds and this progressed to smaller and more complex compounds (Ianora *et al.*, 2012). Along with the improvements in laboratory techniques, the methods of sampling also improved from relying on hand-picking and trawling to SCUBA diving and advanced depth collection with submersibles. One of the major influences that fuelled marine natural products research is advances in sequencing techniques that made the discovery of the biosynthetic pathway of the natural products more readily available.

3.1.1. Examples of marine natural products

In 2013 eight marine natural products were available on the market (Martins *et al.*, 2014). These include trabectedin (Yondelis®) and cytarabine (Cytosar-U®) that is used for the treatment of cancer as well as the anit-viral drug iota-carrageenan (Carragelose®). The latter is a sulphated polysaccharide that was extracted from red algae, specifically *Chondrus crispus* and *Gigartina stellata* (Pietra, 1990). The nucleoside cytarabine, or Ara-C, was isolated from the marine sponge *Cryptotethia crypta* (Sipkema *et al.*, 2005), while trabectedin, an alkaloid, was isolated from the tunicate *Ecteinascidia turbinata* (Wright *et al.*, 1990).

Three other marine natural products will be used as examples to illustrate some of the achievements and limitations encountered with marine drug discovery. The examples

are bryostatins (a polyketide), the conopeptide MVIIA (a RiPP) and discodermin A (suggested to be a non-ribosomal peptide). The conopeptide is already available on the market while the others are still being evaluated in clinical trials.

The bryostatins, a family of polyketides, was studied as early as 1968 when bioassays were performed with the chemical extracts of *Bugula neritina* (phylum Bryozoa). Bryostatin 1 was only isolated in 1981 by fractionation of the extracts that showed bioactivity against cancerous cell lines (Pettit *et al.*, 2002). One of the major limitations of marine drug discovery, a shortage in the supply of the compound, hindered preclinical and clinical trials. *B. neritina* have been studied since 1758 (Mendola, 2003). The knowledge of this organism acquired over the centuries, was used to develop aquaculture strategies and protocols in the 1990s to provide large-scale supplies of bryostatin 1. Since the cost of aquacultures is significant, alternative methods of production were also investigated such as total- and semi-chemical synthesis (Evans *et al.*, 1999; Wender *et al.*, 1998). The latter process produced bryostatins which retained the activity of the natural product (Faulkner, 2000). Bryostatin 1 was the second marine natural compound to enter clinical trials and it is still the subject of numerous phase II clinical trials (Hussain *et al.*, 2012; Ianora *et al.*, 2012). A series of analyses by Haygood and Davidson on the microbial symbionts of *B. neritina* led to their proposal that these symbionts are the producers of bryostatins (Haygood *et al.*, 1999; Faulkner, 2000; Davidson *et al.*, 2001) and in 2007 a putative biosynthesis pathway was identified from an uncultivated microbial symbiont of *B. neritina* (Sudek *et al.*, 2007).

The first marine natural product to pass clinical trials that was licenced for marketing in 2004 (Ianora *et al.*, 2012; Kornprobst, 2014), is a toxin peptide produced by the cone snail, *Conus magnus*. This peptide is known as conotoxin MVIIA, ziconotide or Prailt® and is classified as a RiPP. Conopeptides have been discovered since the early 1980s within the venom of these snails, however, only after a long period of clinical trials and after approval of the licence (1999-2004; Svenson, 2013) was this natural product available for the treatment of chronic pain. The specificity of the product reduces the probability of adverse effects and it is this property that made it so attractive for development as a therapeutic (Shen *et al.*, 2000).

During the early years of bioprospecting the marine environment, invertebrates were favourable sources of different classes of natural products. Various peptides have been extracted, but their bioactive properties were unknown. It was only in 1980 that the importance of peptides as natural products was revealed when peptides from the tunicate *Lissoclinum patell*a showed some bioactivity (Matsunaga *et al.,* 1985). In 1985 the research group of Matsunaga *et al.* (1985) isolated discodermin A. It was the first bioactive peptide isolated from a sponge. They proposed that this antimicrobial non-ribosomal peptide is a product of a symbiotic microorganism of the sponge, *Discodermia kiiensis*. This was due to it having a unique structure with multiple D-amino acids and a *tert*-Leu residue. Until then, the latter type of residue has only been found in bottromycin. Bottromycin was extracted from a terrestrial *Streptomyces bottropensis* (Waisvisz *et al.*, 1957). D-amino acids are more prevalent within bacterial proteins, which is another indication of the symbiotic origin of these peptides. Since then, other peptides containing these unusual residues have been discovered, namely polytheonamides and yaku'amides. It is the ribosomal origin of bottromycins and polytheonamides that suggest that discodermins (and yaku'amides) are produced through ribosomal synthesis (i.e. RiPPs; Freeman *et al.*, 2012).

3.2. The marine environment is suitable for the evolution/synthesis of novel compounds

As the field of marine natural products expanded, scientists were intrigued by the biochemical diversity of marine organisms' metabolites that were unparalleled in terrestrial organisms. The factors that influence the uniqueness and abundance of marine natural products can be ascribed to be either the taxonomical or ecological influence (Jensen & Fenical, 1996).

The taxonomical influence on the novelty and diversity of marine natural products lies in the divergence of terrestrial and marine organisms long ago. This divergence is generally at species level (Jensen & Fenical, 1996). Genera with species of terrestrial origin which have been encountered during bioprospecting for marine natural products include *Pseudomonas*, *Alteromonas*, *Vibrio*, *Bacillus*, and *Streptomyces*. Even though there are biochemical differences at species level, species from the same genera share similarities in their metabolism and thus they have the ability to produce similar

compounds (Jensen & Fenical, 1996). This is why there are marine compounds with structural similarities to those of terrestrial origin, such as the afore-mentioned discodermins and bottromycins.

The ecological difference between the marine and terrestrial environment drives the adaptation of micro- and macroorganisms in order to ensure their survival (Li & Qin, 2005). These biochemical and physiological adaptions are seen through the structural diversity and novelty of their secondary metabolites (Jensen & Fenical, 1996; Faulkner, 2000; Simmons *et al.*, 2005). For example, the high salinity of the ocean creates an environment where halogens, especially chlorine and bromine, are more readily available. Various halogenated natural products have been discovered (Simmons *et al.*, 2005). Also, many marine macroorganisms are physically soft and their chemical defence mechanisms compensates for their lack of structural defence. This can be through the production of natural products or hosting microorganisms with the ability to produce advantageous natural products.

3.3. Sponges and their symbiotic relationship with microorganisms

The symbiosis of microorganisms with macroorganisms is complex (Taylor *et al.*, 2007; Piel *et al.*, 2004; Mohamed *et al.*, 2008). The host macroorganism creates a unique microenvironment with various physical, chemical, and biological factors. These factors determine the selection of organisms who can survive within the specific environment and maintain the mutualistic relationship (Jensen & Fenical, 1996). This selection may be the driving force for the production of novel natural products.

An example of this type of relationship is found within sponges. Sponges (phylum Porifera) are sessile invertebrates and as their name suggests, they draw in the surrounding water through pores and filter it (Taylor *et al.*, 2007; Kennedy *et al.*, 2007). When filtering the water, they ingest the nutrients and the microorganisms therein. Some microorganisms are retained and live in symbiosis with the sponge and others pass through the feeding chambers without being ingested. The symbiotic microorganisms are found to be associated with the external and internal tissue of the macroorganism host (Lee *et al.*, 2001; Kennedy *et al.*, 2008). Up to about 40 % of the

sponge biomass could consist of the symbiotic microorganisms (Fusetani & Matsunaga, 1993).

The symbiotic relationship of sponges and microorganisms is not fully comprehended, but it comes across as a mutualistic relationship. In this relationship the sponge provides a safe, nutrient rich environment for the microorganisms compared to the open water. The role of the microorganisms includes the supply of nutrients (Bondarev *et al.*, 2013), the removal of waste products from the sponge and the strengthening of the sponge structure as they constitute a high percentage of the biomass of the host. Another function of sponge symbionts is their production of secondary metabolites, many of which have widely diverse bioactive properties (Kennedy *et al.*, 2007; Lee *et al.*, 2001). For instance, antimicrobial compounds can be produced to protect the host macroorganism (Bondarev *et al.*, 2013; Faulkner, 2000) as was suggested for the *Pseudovibrio* species (Bondarev *et al.*, 2013).

This is why, at the end of the 20th century, the research focus shifted from sponges to their symbiotic microorganisms. Marine sponges have long been recognised as rich sources of natural products for drug development (Pettit *et al.*, 2002; Mayer *et al.*, 2010; Ouyang *et al.*, 2010; Laport *et al.*, 2009), however it has been proven that the symbiotic microorganisms are the true producers of many known sponge natural products (Muscholl-Silberhorn *et al.*, 2008; Taylor *et al.*, 2007).

4. Methods for the biodiscovery of natural products

In the previous sections a few examples of the marine natural products were discussed. In this section we will look at some of the methods and techniques that have been used to discover these compounds. The limitations and advantages of the methods are also discussed.

4.1. Bioassay screens

Bioassay screens are used with cultivation-dependent and -independent methods for the discovery of natural products. When a biological activity is observed, the compound (or crude extract) is referred to as being bioactive. Bioactive extracts are purified to determine the biochemical properties of the responsible compound(s).

Before the turn of the century, natural product research was mostly focussed on antimicrobial and anti-tumour bioactive compounds (Ianora *et al.*, 2012). Since then a wider range of bioassay screens have been included as the applications of the compounds have extended to more than just the pharmaceutical industry. However, the focus of this study is to discover pharmaceutical-relevant natural products. Other pharmaceutical-relevant assays that have been developed include anti-inflammatory (Svenson, 2013), neuro-active (Kokel *et al.*, 2010), antiviral (Férir *et al.*, 2013) and antimalarial screens (Ang *et al.*, 2000). With these, natural products (crude extract or refined compounds) are tested against different microbial strains or specific cell lines (Imhoff *et al.*, 2011). Examples of screens used are those employed at *The Centre for Marine Bioactives and Drug Discovery* (MabCent-SFI), at the University of Tromsø, Norway (Table 1).

Table 1 Bioassays employed by MabCent-SFI platforms.[1]

Activity	**Target cell line or strain**
Toxicity	MRC-5 (normal lung fibroblast)
Anticancer	HT-29 (colon adenocarcinoma) MCF-7 (breast cancer cells) A2058 (melanoma) IPC-81 (leukaemia) NF-κB (A549) Kinase inhibition
Antibacterial	*E. coli* ATCC 25922 *Staphylococcus aureus* ATCC 9144 Methicillin-resistant *S. aureus* (MRSA) *Psuedomons aeruginosa* ATCC 27853 *Enterococcus faecalis*
Inhibition of biofilm formation	*Staphylococcus epidermidis* ATCC 35984
Immunomodulation (anti-inflammatory)	TNF-α IL-1β (THP-1) NF-κB (U937)
Diabetes	PPARc PTB-1B

[1] Adapted from (Svenson, 2013)

4.2. Traditional and cultivation-dependent screening methods

The initial method to discover marine natural products was to screen the extracts of macroorganisms, such as sponges, for bioactivity. This is sometimes referred to as the

"grind and find" approach. This type of screening method led to the discovery of natural products such as bryostatins, discodermins and polytheonamides as discussed in Sections 3.1. and 2.2.12. Cultivation-based screens have been attempted for many of the macroorganisms from which bioactive compounds were discovered through the "grind and find" approach, but difficulties were usually encountered with the cultivation of the symbiotic microorganisms.

Cultivation-dependent screening methods, typically employed for the discovery of marine natural products, generally consist of four stages. Initially organisms are isolated from an environment and fermented. The fermentation products are screened for a desired bioactivity (examples given in Section 4.1) and the bioactive compound is then isolated. Finally the biochemical characteristics of the compound are determined.

Cultivation-dependent screening methods have been successfully applied to discover bioactive compounds from microbial symbionts of sponges. For example, *Pseudovibrio* species have been isolated from Irish sponges by O'Halloran *et al*. (2011) and 85 % of the isolates displayed antimicrobial activity against at least one of the indicator strain that were tested. Similarly Santos *et al.* (2010) isolated diverse bacterial species (Firmicutes, Alpha- and Gamma-Proteobacteria) with antimicrobial activities from Brazilian sponges.

4.2.1. Limitations and significance of cultivation-dependent methods

The limitations of cultivation-dependent methods are mainly linked to the type of organisms isolated, that is in turn limited by the media and conditions used for isolation. Generally, the organisms isolated from any environment do not represent the true microbial community within that environment (Ward *et al.*, 1990). Many attempts have been made (for example: Schmidt *et al.*, 2000; Bode *et al.*, 2002; Muscholl-Silberhorn *et al.*, 2008; Connon & Giovannoni, 2002) to improve the cultivation media and techniques employed, but it is difficult to mimic the natural environment of the organisms entirely. Initially, routine media (usually being rich in nutrients) was used for the isolation of marine microorganisms. It was soon realised that many of the species are adapted to have limited access to nutrients in their habitat and cannot metabolise the complex carbons (Connon & Giovannoni, 2002). This is why a wide array of media

(frequently the oligotrophic type) is continuously being developed (Connon & Giovannoni, 2002) to select for microorganisms based on their known physiological requirements. Oligotrophic media have successfully promoted the isolation of a more diverse collection of organisms. However, the use of nutrient rich and complex media is still common practice as the cultivation conditions of most microorganisms are still elusive (Bernan *et al.*, 1997). It has been suggested that the cultivation of some organisms might not be possible until the symbiotic relationships of organisms (macro and micro) are understood (Bondarev *et al.*, 2013; Querellou *et al.*, 2010).

Another limitation is when cultivation conditions used for screening do not activate the biosynthetic pathways of natural products (Jensen & Fenical, 1996; Newman & Cragg, 2012). Consequently, many of the products remain undiscovered during the screening process. With the genomic age the full genome sequence of organisms became available. Analysis of the genomes showed that many of the natural products have not been discovered with the standard laboratory cultivation approaches. This concept was exemplified when the full genome of the model organism *Streptomyces coelicolor* A3(2) was published in 2002 (Bentley *et al.*, 2002). Examining the genome, 18 gene clusters for unprecedented secondary metabolites of various classes were discovered. These types of gene clusters are known as "cryptic", "silent" (Imhoff *et al.*, 2011) or "orphan" pathways. Thus by changing the cultivation conditions the activation of cryptic pathways has been observed with the production of novel natural products (Qu *et al.*, 2011). An approach to discover the products of orphan pathways is the one-strain-many-compounds (OSMAC) approach (Bode *et al.*, 2002) where matrices of screens are performed under multiple conditions.

One of the major impediments in the development of drugs from natural products is the production of the compounds in large quantities for evaluation in trials and marketing (Proksch *et al.*, 2002). The complexity of the compounds make the chemical synthesis, and thus large-scale production thereof, complicated and generally not feasible (Li & Qin, 2005; Kennedy *et al.*, 2007). This creates the need for an effective, conservational and economically feasible supply of natural products. One such approach is the semi-synthesis of analogues of the natural product. An example is where cyanosafracin B, produced by *Pseudomonas fluorescence*, is used for the semi-synthesis of trabectedin

(Yondelis®; Cuevas & Francesch, 2009). Another example is the production of Ara-C, the analogue of the nucleoside cytarabine (Taylor *et al.*, 2007; Sipkema *et al.*, 2005; Dunlap *et al.*, 2007; Laport *et al.*, 2009).

There are currently numerous organisms that can still be exploited for the production of novel natural products if they could be cultivated. The development and improvement of cultivation-dependent methods are thus still imperative and have the distinct advantage that the producing organisms are cultured under laboratory conditions and are available for further investigations and scaled up production. Ultimately the bioactive compound can be obtained in large quantities, for development and marketing purposes, once the large-scale fermentation of the organisms is optimised. It is important to note that some cultivation-independent methods are based on cultivation-dependent methods. An example is the functional screening of heterologously expressed libraries (refer to the next section). Also, when a biosynthetic pathway of a natural product has been identified using sequence-based cultivation-independent methods, the pathway can be expressed in a heterologous host to characterise the biosynthetic genes and the product. This underlines the importance of cultivation-dependent methods.

4.3. Cultivation-independent screening methods

Cultivation-independent screening methods use molecular and *in silico* techniques that do not require the isolation and cultivation of organisms for the discovery of natural products. Traditionally cultivation-independent methods were used to access the genomes of uncultured organisms through metagenomics, but the methods expanded to include the exploration of their transcriptomes and proteomes (Figure 17). The information generated from these methods is stored on open access database servers for comparative analysis with bioinformatics tools. It is also important to note that, although the focus is on uncultured organisms, cultured organisms are also investigated with cultivation-independent methods, since sequencing and analysis of their genomes could result in the identification of biosynthetic pathways for natural products that have not been discovered with the cultivation-dependent methods (Bentley *et al.*, 2002; Udwary *et al.*, 2007).

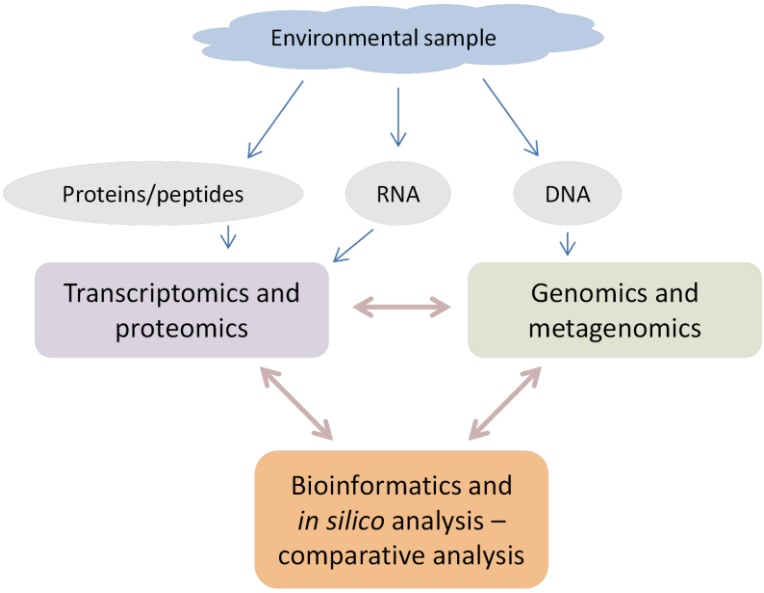


Figure 17 An overview of the cultivation-independent methods used for the screening of natural products.

4.3.1. Metagenomics

In 1998 Handelsman *et al.* introduced the term metagenomics (Steele & Streit, 2005), which is essentially the study of the genomes of a mixed community of organisms within an environment. Some of the first metagenomic DNA (mDNA) libraries were constructed from seawater (Béjà *et al.*, 2000) and marine sponges (Schleper *et al.*, 1998). Since many marine natural products are biosynthesized by the (as of yet) uncultivable symbiotic microorganisms of invertebrates (Haygood *et al.*, 1999; Piel *et al.*, 2004; Proksch *et al.*, 2002), researches use metagenomics (Figure 18) to circumvent the limitations of cultivation-dependent methods and gain access to the biosynthetic pathways of (known and unknown) natural products.

Metagenomic screening methods require the construction of mDNA libraries that are screened either with function-based assays or with sequence-based analysis. Functional screening is based on the heterologous expression of the genes by the surrogate organism. These organisms include the workhorse of the laboratory, *E. coli*, but other hosts have also been included such as *Streptomyces* (Courtois *et al.*, 2003; Li & Qin, 2005; Ikeda *et al.*, 2014; Wang *et al.*, 2010); *Bacillus* (Eppelmann *et al.*, 2001; Biver, 2013) and various Proteobacteria (Charlop-Powers *et al.*, 2013; Li & Qin, 2005). Functional screening can be done by observing phenotypical changes of a sought after

activity (Section 4.1). This has been successfully employed for the discovery of natural products with antimicrobial screening (Rondon *et al.*, 2000; Courtois *et al.*, 2003). Another method is where the surrogate host (or a mutant surrogate host) is dependent on the targeted gene for survival within specific conditions (Wang *et al.*, 2010; Simon & Daniel, 2011). Many natural products have been identified with functional screens of metagenomes derived from soil and marine environments (Courtois *et al.*, 2003; Brady *et al.*, 2001). This method is not readily employed for the discovery of RiPPs, as researchers rather take advantage of the available genomic methods.

Sequenced-based screening of a metagenome can be performed with homology-based techniques and sequencing. Advances in sequencing methods (next generation sequencing) enable the screening of mDNA without the construction of libraries (Suenaga, 2012; Heidelberg *et al.*, 2010). Probes are used for homology-based screening methods, such as oligonucleotide primers for PCR (Więckowicz *et al.*, 2011; Donia & Schmidt, 2011). This requires prior knowledge of the biosynthetic gene(s) and thus novel genes and natural products could go unnoticed. However, another homology-based method, retrobiosynthetic analysis (Piel, 2011), is not dependent on prior knowledge of the biosynthetic gene(s), but it is based rather on the structure and biochemical characteristics of the compound. This knowledge can be used to deduce the sequence of the structural or biosynthetic genes and then degenerate primers are designed to target those sequences.

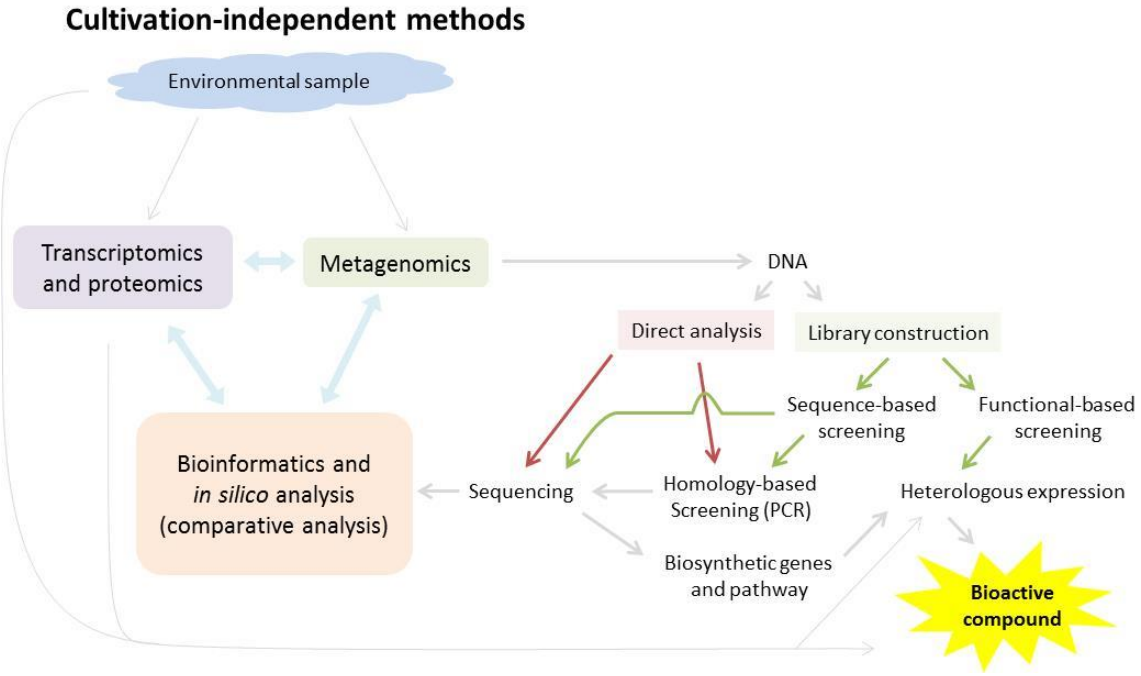


Figure 18 Metagenomics (a cultivation-independent method) is used to screen the genomes of uncultured organisms for natural products.

4.3.2. Bioinformatic and *in silico* analyses - comparative analyses

The interdisciplinary field of bioinformatics relies on computational techniques to combine the data generated by chemical analysis and molecular "omics" methods in databases. The omics methods include genomics, metagenomics, transcriptomics and proteomics. The data is analysed with specialised algorithms to correlate the genotype, chemotype and phenotype of a natural product.

A major limitation of bioinformatics is the need for adequate technologies to process the data. For instance, with the mass inflow of sequencing data, genomes are not always accurately annotated (Weber, 2014) and classic gene-identification algorithms do not always recognise the biosynthetic genes of natural products (Haft, 2009; de Jong *et al.*, 2010). Various natural product databases and automated bioprospecting tools have been established (Fedorova *et al.*, 2012) to circumvent this limitation. The biosynthetic genes of microbial natural products are usually clustered together (Robinson, 1991; Jack *et al.*, 1995; Walsh, 2000). This enables the use of *in silico* tools to perform comparative analysis to assist and automate the screening of natural product pathways (Brady *et al.*, 2001). Many of the tools focus on the discovery of non-ribosomal peptides and

polyketides (Weber, 2014; Caboche *et al.*, 2007; Ichikawa *et al.*, 2013; Weber *et al.*, 2009), but the tools available for the exploration of RiPPs are also gaining interest, as it can be seen that the RiPP-related tools were developed predominantly in the last few years (Table 2).

There is also a great need for inter-disciplinary research programs. Pep2Path is such a software package and it can be used for the automation of peptidogenomics (Medema *et al.*, 2014). Peptidogenomics is the correlation of mass spectrometry data to the biosynthetic gene clusters of known and putative peptides. Pep2Path is comprised of the two algorithms, NRP2Path and RiPP2Path, respectively for non-ribosomal peptides and RiPPs. The Pep2Path algorithms also make use of databases and other tools, such as NORINE (Caboche *et al.*, 2007), NRPSpredictor2 (Rottig *et al.*, 2011), Genbank, and antiSMASH (Medema *et al.*, 2011; Blin *et al.*, 2014) to achieve the goal of peptidogenomics. This software is a valuable tool that makes natural product discovery and the characterisation thereof a high-throughput process.

The use of bioinformatics tools is an integral part of the research and the discovery of novel marine natural products or biosynthetic pathways. It is also important to confirm the outcomes from *in silico* analyses with experimental procedures. This includes the functional characterisation of the identified putative genes as well as the production of the compounds encoded, determination of structural characterisation and analysis of bioactivities.

Table 2 Examples of the web-based databases and tools available for screening natural products. A short description, the class of natural products covered, and the year of release and updating is also indicated. Note that this is a non-exhaustive list.

Name	**Short description**	**Class covered**	**Year**	**Reference**
BAGEL **BA**cteriocin **GE**nome mining too**L**	Web-based bacteriocin genome mining tool.	RiPPs (bacteriocins)	2006, 2010, 2013	(de Jong *et al.*, 2006; de Jong *et al.*, 2010; van Heel *et al.*, 2013)
BACTIBASE	Web-based database and tool for bacteriocin characterization.	RiPPs (bacteriocins)	2007, 2010	(Hammami *et al.*, 2007; Hammami *et al.*, 2010)
antiSMASH **anti**biotics and **S**econdary **M**etabolite Analysis **SH**ell	The first comprehensive pipeline capable of identifying and annotating biosynthetic loci covering the whole range of known secondary metabolite compound classes in bacterial and fungal genome sequences.	polyketides, non-ribosomal peptides, terpenes, RiPPs (lantibiotics, bacteriocins), nucleosides, beta-lactams, siderophores, melanins, etc.	2011, 2014	(Medema *et al.*, 2011; Blin *et al.*, 2014)
Pep2Path	Software for automated mass spectrometry-guided genome mining of peptidic natural products.	non-ribosomal peptides, RiPPs	2014	(Medema *et al.*, 2014)
RiPPquest	A combinatorial tool that performs genome- and metabolome-mining of sequenced microorganisms.	RiPPs (specifically lanthipeptides)	2014	(Mohimani *et al.*, 2014)

4.3.3. Limitations and significance of cultivation-independent methods

4.3.3.1. Functional screening

Heterologous expression is one of the limitations of functional-based methods, although this is not limited to the field of natural product discovery. The level of heterologous expression is generally not sufficient for biodiscovery, for instance, quantitative analysis showed that only 40 % of genes are expressed in *E. coli* (Gabor *et al.*, 2004). However, as mentioned before, this is why different approaches, including a wider range of surrogate hosts, are continuously developed to circumvent and improve gene expression. The selection of bioassays should also be carefully considered as it could restrain the range of compounds that will be discovered (Suenaga, 2012). Even though there are limitations, functional screening of DNA libraries is still one of the principal methods used for the discovery of natural products, especially with high-throughput screens that are being developed such as those at MabCent-SFI (Table 1). One of the major advantages is that once a novel compound is detected, it is already in a known expression system and thus the afore-mentioned supply problem for down-stream development should not be a limiting factor.

4.3.3.2. Sequence-dependent screening

Sequence-based screening methods are limited by the lack of extensive sequencing and computational technologies to assemble and annotate the (meta-)genomes (Heidelberg *et al.*, 2010; Mavromatis *et al.*, 2012), but these technologies are consistently being improved. Some of the solutions for this include better coverage of genomes by deep (second generation) and ultra-deep (third generation) sequencing (Querellou *et al.*, 2010). The problems of genome annotation are discussed by Poptsova and Gogarten (2010). Overall, annotations are based on comparative analysis and due to the advancements of genomic and proteomic techniques, the accumulation of experimental information as well as development of advanced algorithms, the annotations of (meta-)genomes are refined continuously. The use of sequence-based methods, compared to functional-based screening, has the advantage of revealing the biosynthetic pathways of known natural products and predicting the hypothetical pathways of novel natural products. Knowledge of the biosynthetic pathway is important for development and production of the product.

4.3.3.3. Cultivation-independent methods

It should also be noted that many of the cultivation-independent methods have mainly been developed over the last 10 years and thus hold great potential for the discovery of natural products from the marine environment. These methods, especially those related to genomics and proteomics, are powerful tools for the discovery of novel natural products from uncultured organisms.

The biosynthetic genes of a natural product can be identified using culture-independent molecular techniques. The compound can be produced on a large-scale, through heterologous expression of the biosynthetic genes, to provide enough material for the determination of the biochemical properties and structure of the compound, as well as to provide material for clinical trials and marketing.

4.3.4. Discovery of RiPPs with combinatorial methods

There is no standard method for the discovery of novel RiPPs. However, there are numerous publications available on the methods used. Early discoveries were made with traditional and cultivation-dependent methods. Recent methods are broadly based on homology analysis and it is common practice to use a combination of the methods discussed earlier. The methods can be considered to belong to one of three main methodologies, namely genome mining, reverse genomics and peptidogenomics (Table 3). A few examples will be discussed to demonstrate these methods.

Whether it is screening a single organism's genome or a consortium thereof from a specific environment (constituting a metagenome), genome mining is probably the most widely used approach for the screening of RiPPs. An example of how this approach is implemented will be discussed with the pioneering work of Schmidt *et al*. (2005). With genome mining the biosynthesis genes of the cyanobactins, patellamide A and C, were determined. A library of the metagenome of the ascidian symbiont, *Prochloron didemni*, was created and initially screened with NRPS-specific PCR primers. This screen provided no insight into the biosynthesis of the peptides and thus the ribosomal origin was pursued. To design the primers the draft genome of *P. didemni* was manually examined (mined) by means of tBLASTn searches for all the possible sequences of the core peptide of patellamide A. Further genome mining of the flanking regions revealed

the whole biosynthetic cluster. The functions of the genes were confirmed by heterologous expression in *E. coli*. A similar approach was used by Kelly *et al*. (2009) to elucidate the biosynthesis of the thiopeptide, thiostrepton A, when they screened the genome of *Streptomyces laurentii*. Numerous RiPPs have been discovered with genome mining, for example lanthipeptides (Begley *et al.*, 2009), linaridins (Claesen & Bibb, 2010; Claesen & Bibb, 2011), microviridins (Weiz *et al.*, 2011; Ziemert *et al.*, 2010), microcins (Duquesne *et al.*, 2007; Claesen & Bibb, 2010; Haft *et al.*, 2010), lasso peptides (Knappe *et al.*, 2008; Maksimov *et al.*, 2012), cyanobactins (Schmidt & Donia, 2009; Wang *et al.*, 2011), sactipeptides (Murphy *et al.*, 2011), and other TOMMs (Wieland Brown *et al.*, 2009; Liao *et al.*, 2009; Morris *et al.*, 2009; Engelhardt *et al.*, 2010; Schmidt *et al.*, 2005), as well as bacteriocins (Więckowicz *et al.*, 2011; Haft, 2009).

Reverse genomics, or retrobiosynthetic analysis, combines the structural and biochemical characteristics of a peptide to infer the genes involved in the biosynthesis thereof. Degenerate primers are designed to these genes and used for the screening of (meta-)genomes by PCR. This approach has been successfully employed for the identification and characterisation of the biosynthetic pathways of RiPPs such as polytheonamides. As discussed before (Section 2.2.12), this peptide was thought to be produced by NRPSs due to the unusual incorporation of unique, non-proteinogenic residues. It is only 18 years after it was first isolated from the sponge *T. swinhoei*, that their ribosomal origin has been elucidated (Freeman *et al.*, 2012). As for patellamides, the metagenome of *T. swinhoei* was first screened with NRPS-specific PCR primers, but the biosynthetic genes were not detected. Degenerate primers were designed to the amino acid sequence of the core peptide and subsequently the precursor gene was discovered. Using other genomic techniques the whole biosynthetic cluster was determine (Freeman *et al.*, 2012).

In the subsequent study by this group (Wilson *et al.*, 2014; which occurred concurrent to the research in this thesis) the producing organism was isolated. Two new bacterial phylotypes of the candidate genus '*Entotheonella'* were isolated from the sponge. This genus was so unique that it was suggested that these species are part of a new phylum 'Tectomicrobia'. '*Entotheonella'* species are also not confined to *T. swinhoei*, as they

have been detected in a wide range of sponges of various locations (Wilson *et al.*, 2014; Schmidt *et al.*, 2000; Brück *et al.*, 2008). Sequencing of the two species from *T. swinhoei* revealed that '*Entotheonella'* species are rich producers of natural products from various classes of NRPs, PKs and RiPPs. The biosynthetic pathways of many of the known natural products from *T. swinhoei* were also identified in these genomes. The polytheonamide biosynthetic cluster is encoded within the genome of '*Entotheonella factor* TSY1' along with other putative proteusin pathways. This shows that research of proteusin peptides is at its infancy and that many studies will follow this pioneering work.

Another approach used for the discovery of RiPPs is a combination of genome mining together with data from proteomics. This is also referred to as peptidogenomics. This approach is not readily used for the discovery of RiPPs, but holds great potential, as demonstrated by the ioneering work of Kersten *et al.* (2011). With this approach the mass spectrometry data obtained with proteomics is compared to sequenced genomes of microorganisms (or even metagenomes) to determine the biosynthetic pathway. The recently developed bioinformatics tools, such as RiPPquest as well as above-mentioned Pep2Path and antiSMASH, are automatizing this approach (Table 2).

Table 3 Approaches used for the discovery of RiPPs. It is important to note that the list is not exhaustive. Genome mining can be centred on the precursors or the post translational modification (PTM) enzymes.

Method	Description/ target	RiPP	Class	Reference
Genome mining (genomics)	Precursors	Astexin 1 Cypemycin Bottromycins	Lasso peptide Linaridin Bottromycins	(Maksimov e*t al.*, 2012; Claesen & Bibb, 2010; Gomez-Escribano *et al.*, 2012)
	PTM enzymes	Lichenicidin, Putative sactipeptides	Lanthipeptide Sactipeptides	(Begley *et al.*, 2009; Murphy *et al.*, 2011)
Retrobiosynthetic analysis (reverse genomics)		Microviridin B and J Nocathiacin Polytheonamides	Microviridins Thiopeptides Proteusins	(Ziemert *et al.*, 2008; Wei *et al.*, 2011; Freeman *et al.*, 2012)
Peptidogenomics		AmfS SGR-1832 SSV-2083	Lanthipeptides Linaridins Lasso peptides	(Kersten *et al.*, 2011)

From the literature that was studied it can be seen that the methods used for the discovery of RiPPs have advanced from simple (nevertheless important) genome mining strategies to combinatorial methods that attempt to determine the genotype, chemotype and phenotype of known and unknown peptides.

5. Aims and objectives

Cultivation-independent methods have become paramount to the screening of marine natural products, especially RiPPs. However, more researchers are using a combination of cultivation-dependent and -independent screening methods. The combined techniques reduce the limitations associated with the methods when they are used independently. A few researchers have successfully used this approach for the screening of the symbiotic bacteria of invertebrates (Kennedy *et al.*, 2008; Engelhardt *et al.*, 2010; O'Halloran *et al.*, 2011; Öztürk *et al.*, 2013; Bondarev *et al.*, 2013).

It is clear that there is no standard method for the screening of marine natural products and RiPPs. Numerous approaches have been presented and the most promising approach appears to employ a combination of methods to exploit the biotechnological potential of the symbiotic microorganisms of marine sponges. This can be achieved by combining cultivation-dependent methods, to isolate and cultivate the symbionts, and cultivation-independent sequencing methods, to screen for the biosynthetic genes of RiPPs.

In the light of the literature study presented in the previous sections, the aims of this thesis will now be described.

The aim of the cultivation-dependent section of this study was to isolate symbiotic bacteria from a South African marine sponge. The isolates were then screened for antimicrobial activities.

The aim of the cultivation-independent section of this study was to extract metagenomic DNA from a South African marine sponge, create a library from the extracted DNA and develop a genomic screening method for RiPPs, specifically polytheonamides or proteusins.

Chapter 2 - Isolation of microbial symbionts and their bioactivity profiles

1. Introduction

Metagenomic diversity studies have indicated that only a small fraction of the true diversity of a community is represented by the cultivated organisms in any given environment (Ward *et al.*, 1990). Metagenomics has also showed that many of the uncultivated organisms are resources for novel natural products. Therefore, there is clearly a need for the development and improvement of cultivation techniques to enable researchers to gain access to the currently uncultivated organisms.

The rise in resistance to current antibiotics necessitates the need for the discovery of antimicrobial compounds. Terrestrial resources are becoming exhausted as many of the known antimicrobial compounds are rediscovered (Hou *et al.*, 2012; Fischbach & Walsh, 2009). This is why the marine environment attracts the attention of scientists. Marine organisms display unique adaptions to their environments, which is mirrored in the structural diversity and novelty of their secondary metabolites (Jensen & Fenical, 1996; Faulkner, 2000; Simmons *et al.*, 2005). Examples are the incorporation of halogens (chlorine and bromine; Simmons *et al.*, 2005) as well as unique peptide structures which are cyclic or linear with unusual or rare amino acids, such as D-amino acids and *tert*-Leu (Matsunaga *et al.*, 1985).

Marine sponges are recognised as rich sources of natural products (secondary metabolites) for drug development (Pettit *et al.*, 2002; Mayer *et al.*, 2010; Ouyang *et al.*, 2010; Laport *et al.*, 2009). The structural characteristics of the sponge natural products led to the hypothesis that the symbiotic microorganisms in the sponges are the true producers of the compounds. Some of the compounds consist of multiple D-amino acids that are generally associated with microbial peptide synthesis. There are also structural similarities between natural products extracted from sponges and those extracted from terrestrial microorganisms (Fusetani & Matsunaga, 1993). For many sponge natural products, for example discodermins (Matsunaga *et al.*, 1985) and polytheonamides (Freeman *et al.*, 2012), it has been proven that the symbiotic microorganisms are the true producers (Muscholl-Silberhorn *et al.*, 2008; Taylor *et al.*, 2007).

Therefore, marine bacteria are a source of novel natural products which could be exploited for pharmaceutical applications. More importantly, with cultivation-dependent

methods the producing organisms are cultured under laboratory conditions and consequently the supply limitation could be circumvented. There is a great need for effective, conservational and economically feasible supply of natural products. This could be achieved, for example, by discovering a cultivable microorganism for fermentation or implementing semi-synthesis of analogues of a natural product. Semi-synthesis is employed to supply trabectedin (Yondelis®; Cuevas & Francesch, 2009). The antibiotic cyanosafracin B, produced by *P. fluorescens*, is processed to obtain analogues of the natural products, ecteinascidins. Ecteinascidins were originally extracted from the tunicate *E. turbinata* (Cuevas & Francesch, 2009).

This chapter describes the first study of the isolation, cultivation and antimicrobial screening of the microbial symbionts of the South African sponge, *Hamacantha (*Vomerula*) esperioides*. No other research has been done on this sponge species (to the knowledge of the author) other than the description of the type species by Ridley and Dendy (Ridley & Dendy, 1886).

2. Materials and methods

2.1. Sponge material (PE14)

The *H. esperioides* sponge (referred to as PE14 throughout) used in this study was collected by Dr. Shirley Parker-Nance by SCUBA diving on 4 April 2013. The site of collection is at the White Sands reef in Port Elizabeth, on the eastern coast of South Africa (S34 ° 00.406 ', E25 ° 43.117 '), at a depth of 23 to 25 m. The average water temperature was 15 °C. The material was shipped overnight at 4 °C and processed for cultivation upon delivery. The material for genomic studies was stored in 10 % (v/v) glycerol at -20 °C.

2.2. Media

A wide range of media (Table 4) was used for the isolation and cultivation of symbiotic microorganisms of the sponge PE14. The isolation media was selected to promote the growth of slow growing marine organisms and to prevent bias towards the fast growing organisms. Details of the composition of the media are given in Appendix A.

Table 4 Media used for the isolation of sponge symbionts.

Medium	**Abbr.**	**pH**	**Selectivity**
Modified 172F	172	7.5	Actinobacteria
Actinomycete isolation medium	AIM	8.1	Actinobacteria
Activated charcoal medium	ACM	7	Pyruvate as the primary carbon source
Blue-green 11	BG11	NA	Nitrogen Free Medium; cyanobacteria
Complex carbon agar	CCA	7.5	General media
Streptomyces medium (modified DSMZ Medium 65)	GYM	7.2	Actinobacteria; Streptomycetes
Medium A agar	MAA	10.0 - 10.5	Halo-tolerant and halophilic bacteria
Oatmeal agar (modified ISP Medium 3)	OMA	NA	Actinobacteria
Planctomycetes medium	PMM	NA	Planctomycetes bacteria
Pringsheim's biphasic soil-water medium	PCS	NA	Cyanobacteria
Reasoner's 2A agar	R2A	NA	Slow growing aquatic bacteria
Seawater Agar	SWA	6.5	Low nutrients content; slow growing bacteria
Shivji NA ¼ strength	SNA	7.5	Selective for firmicute bacteria
Sponge extract medium	SEM	NA	Nutrients required from sponges
Tryptic soya broth	TSA	7.5	Gram-positive bacteria
Tryptic soya broth pH 4.5	TS4	4.5	Gram-positive acidophiles
ZoBell agar ¼ strength	ZBA	7.5	Marine bacteria, primarily Gram-negative
ZoBell agar ¼ strength with 0.0001% (v/v) Crystal violet solution	ZBV	7.5	Gram-negative bacteria

NA: pH was not adjusted.

2.3. Microbiological techniques

2.3.1. Cultivation of sponge symbiotic microorganisms

*S*ponge material was crushed with a sterile pestle and mortar and a serial dilution of the liquid collected was made with filter-sterilised seawater. Dilutions (10^{-1} to 10^{-6}) were spread-plated on the different solid media (Table 4). Cycloheximide and nalidixic acid were added to isolation media GYM, 172, OMA, MAA at 100 µg.ml^{-1}and 50 µg.ml^{-1}, respectively, during the preparation of the plates. Plates were incubated at 15 °C and checked periodically for growth over a period of 4 months.

The liquid medium, blue-green 11 (BG11), was used for the isolation of cyanobacteria. After incubation of the liquid culture for 3 months, the culture was diluted and plated onto R2A medium for isolation of bacteria.

When growth of a colony was observed on the isolation medium, it was transferred and cultured until pure. The number of media used for continuous culturing was reduced after isolation (Table 5). Glycerol stocks of 50 % (v/v) were prepared and stored at -80 °C. The selection of colonies was based on the difference in morphology and growth rate.

Table 5 Colonies picked from the isolation media were further maintained on specific culturing medium.

Isolation medium	Continuous culturing medium
TSA	TSA
TS4	
SNA	
GYM	GYM
172	
OMA	
AIM	
R2A	R2A
BG11	
ZBA	ZBA
ZBV	
SWA	
CCA	
SEM	
ACM	ACM
MAA	MAA
PCS	PCS
PMM	PMM

2.3.2. Antimicrobial Assay

For the antimicrobial assay, the isolates were spot-inoculated on TSA, ZBA and GYM agar media and incubated for two weeks at 15 °C. The indicator strains, *E. coli* 1699 (Cubist Pharmaceuticals), *Bacillus cereus* ATCC10702, *S. epidermidis* ATCC14990, *Mycobacterium smegmatis* LR222 and *Pseudomonas putida* ATCC27853, were cultivated overnight in Lysogeny broth (LB: 10 $g.L^{-1}$ tryptone, 5 $g.L^{-1}$ yeast extract, 5 $g.L^{-1}$ NaCl) at 37 °C. The optical density at 600 nm (OD_{600}) was measured for each strain to normalise (Equation 1) the amount of cells added to 7 ml sloppy LB agar (0.7 % (w/v) agar). The plates that were spot-inoculated with the isolates were overlaid with the indicator strains and incubated overnight at 30 °C, after which zones of inhibition were noted. Some of the isolates showing antimicrobial activity were selected for further analysis.

$$\text{Volume A } (\mu l) = \frac{4}{OD_{600}}$$

$$\text{Volume B } (\mu l) = \frac{160}{OD_{600}}$$

Equation 1 Normalisation of the amount of cells added to the sloppy agar (7 ml) in the overlay assay for *E. coli* (Volume A) and *B. cereus*, *S. epidermidis*, *M. smegmatis* or *P. putida* (Volume B).

2.3.3. Preparation of extracts for further screening

Extracts from liquid cultures in TSA, ZBA, GYM and ACM media were prepared for bioassay screening by PharmaSea partners.

Fractionation columns (SPE columns) were prepared by first puncturing holes into the bottom of a 15 ml falcon tube and then covering it with glass wool. The tube was filled with 10 ml of SP207ss resin (Sigma-Aldrich). The SPE columns were then equilibrated by washing with 50 ml of acetone, followed by a wash of 50 ml methanol and a final wash with 100 ml MilliQ water.

Isolates were cultured at room temperature on an orbital shaker for 2 to 3 days. Extracts were prepared by acetone extraction where an equal volume of acetone was added to the culture and incubated on an orbital shaker for 1 hour. The acetone was evaporated in a fume hood before the crude extract was applied to the SPE columns. The columns were then washed with 50 ml MilliQ water. The organic extract was eluted with 50 ml of acetone. The eluate was collected and the acetone evaporated in a CentriVap centrifugal vacuum concentrator (Labconco). The dried extract pellet was then shipped to the PharmaSea partners for screening.

Screening was performed against pathogenic and drug-resistant bacteria, yeast, and fungi (MEDINA, Spain). An anti-inflammatory activity screen and a cell toxicity (MarBio, Norway) screen was performed as part of the evaluation for preclinical safety. Anti-fungal screens were against the type strains *Aspergillus fumigatus* ATCC 46645 (Monteiro *et al.*, 2012) and *Candida albicans* MY155 (Martín *et al.*, 2013). Anti-inflammatory screens were against TNF-α, IL-1β and NF-κ cell lines (Svenson, 2013). The cell toxicity screens were against MRC-5 cell lines.

2.4. Molecular techniques

2.4.1. Genomic DNA extraction

The small scale preparation method from Wilson (2001) was used to extract the genomic DNA of the PE14 isolates. Cultures used for extraction were grown in 5 ml of liquid ZBA broth at 15 °C until the stationary growth phase was reached.

2.4.2. Amplification of 16S rRNA genes

The 16S rRNA gene was partially amplified with the universal bacterial 16S rRNA primers E9F (GAGTTTGATCCTGGCTCAG; Ah Tow, 2003; Farrelly *et al.*, 1995) and U1510R (GGTTACCTTGTTACGACTT; Ah Tow, 2003; Reysenbach & Pace, 1995). Primers E9F (forward) and U1510R (reverse) were synthesized by Inqaba Biotech (Pretoria, South Africa) and purified on a reverse-phase cartridge.

Phusion High-Fidelity DNA polymerase (Fermentas, Thermo Scientific) together with the 5X Phusion HF Buffer was used according to the manufacturer's instructions. All the PCR reaction mixtures were made to a total volume of 20 µl. The primers were

added to a concentration of 1 mM. DNA was added to a concentration of 50 to 100 ng per 20 µl reaction. The control reaction of the PCR contained no template DNA.

The PCR reaction was performed by denaturing the DNA initially at 98 °C for 1 minute. This was followed by 35 cycles where the DNA was denatured at 98 °C for 10 seconds and the primers were annealed at 50 °C for 20 seconds after which extension was performed for 20 seconds at 72 °C. A final extension step of 5 minutes was performed at 72 °C.

2.4.3. Qualitative and quantitative analysis of nucleic acids

2.4.3.1. Spectrophotometric and fluorometric quantification

The NanoDrop 1000 Spectrophotometer (Thermo Scientific) was used to determine the quality and quantity of DNA by measuring the absorbance peaks at wavelengths of 280, 260 and 230 nm. The measurements were taken according to the manufacturer's instructions. If it was required to obtain the exact quantity of double stranded (dsDNA), the DNA was quantitated with Qubit dsDNA Broad Range Assay kit according to the manufacturer's instructions for fluorometric quantification in a Qubit 2.0 Fluorometer (Invitrogen; Life Technologies).

2.4.3.2. Agarose gel electrophoresis

To determine if the genes were successfully amplified by PCR the DNA was separated on agarose gels by electrophoresis. Agarose gels (1 % (w/v)) were prepared with TAE buffer (40 mM Tris-base, 20 mM acetic acid, 1 mM EDTA, pH 8.4) and ethidium bromide (10 $ng.ml^{-1}$) was added before pouring the gel into a mould. Electrophoresis was performed within TAE buffer at a constant voltage of 80 V. The AlphaImager HP System was used to visualise agarose gels by UV illumination at 304 nm. The wavelength was set to 365 nm if the DNA was cut and purified for further processing.

2.4.4. Cloning and sequencing

2.4.4.1. Preparation of PCR products

The NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) or the Illustra GFX PCR DNA and Gel Band Purification kit (GE Healthcare) was used to purify selected DNA fragments from agarose according to the specific manufacturer's instructions.

2.4.4.2. Preparation of electrocompetent cells

Electrocompetent cells of *E. coli* GeneHogs (Invitrogen) were prepared according to Dower *et al.* (1988). The following modifications were made: the culture was incubated until an optical density at 600 nm (OD_{600}) of 0.5 to 0.7 was reached and the washing step of the culture was performed with 10 % glycerol and not with HEPES.

2.4.4.3. Transformation by electroporation

The plasmids that were used included pJET1.2 (Thermo Scientific), pET21a and pET28a (Novagen). The pJET1.2 plasmid was the vector for the prepared PCR product. The pET21a and pET28a (Novagen) plasmids were used as controls for the transformation (without inserted PCR product).

The ligation mixture was treated with chloroform as prescribed (CloneJET PCR Cloning Kit; Thermo Scientific) and drop-dialysis was performed before transformation to ensure that the DNA is in a low ionic strength solution. Drop-dialysis was performed by floating a 0.025 µm DNA filter paper (Millipore, VSWP01300 MF-Millipore) on UV treated MiliQ water. An aliquot of 5 to 10 µl DNA was pipetted onto the filter and incubated for 20 minutes.

Cells, incubated on ice for 15 minutes with 1 µl of the DNA, were transferred to a cooled electroporation cuvette (BioRad) with a 1 mm gap. Electroporation was performed with a BioRad MicroPulser under the following conditions: 1.8 kV, 25 µF and 200 Ω. After electroporation, LB was added to the cuvette to a total volume of 1 ml. The transformed cells were transferred to a clean tube and incubated for 60 minutes in a water bath with a temperature of 37 °C. The culture (100 µl), diluted appropriately in LB, was spread onto LB plates. The appropriate antibiotics, depending on the plasmid that was used, were added during the preparation of the LB plates. For the culturing of transformants with pJET1.2 and pET21a, 50 $\mu g.ml^{-1}$ of ampicillin was added. For pET28a, 100 $\mu g.ml^{-1}$ of kanamycin was added. The plates were then incubated overnight at 37 °C. The transformation efficiency was calculated as CFU per ng of plasmid DNA added.

2.4.4.4. Plasmid extraction

For the extraction of the plasmids from the transformants, colonies were picked and inoculated into 5 ml of LB broth with the appropriate antibiotic. A modified alkaline extraction of the plasmids was performed as described by Birnboim and Doly (1979). Briefly; after pelleting the culture, cells were resuspended in 100 µl of the first solution (50 mM glucose, 25 mM Tris-HCl, 10 mM EDTA, pH 8.0) and 200 ng/µl RNase A. The second (lysis) solution (0.2 N NaOH, 1 % (w/v) SDS) was freshly prepared and 500 µl was added to the mixture, inverted and cooled before 150 µl of the third (neutralization) solution of 5 M potassium acetate (pH 4.8) was added. The mixture was centrifuged to remove the cell debris and chromosomal DNA. The supernatant was transferred to a clean tube and the DNA was precipitated with absolute ethanol. This was followed by washing the plasmid DNA with 70 % (v/v) ethanol and evaporating the excess ethanol before the DNA was resuspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0).

2.4.4.5. Confirmation of transformation by amplification

Amplification of inserts with the pJET1.2 primer pair (pJET1.2F: CGACTCACTATAGGGAGAGCGGC, pJET1.2R: AAGAACATCGATTT TCCATGGCAGT) was performed to check whether the extracted plasmids contained the inserts of the expected size. The PCR reaction mixtures were made to a total volume of 20 µl containing 2 µl of 10X DreamTaq buffer (already containing 2 mM $MgCl_2$), 200 µM of each dNTP, 0.025 $U.\mu l^{-1}$ DreamTaq polymerase (Fermentas, Thermo Scientific) and 1 µM of the each primer. The final concentration of the $MgCl_2$ was increased to 3 mM. The purified plasmid DNA was used as the template DNA. When the plasmid was not extracted from the clones, a colony PCR was performed according to the instructions of the CloneJET PCR Cloning Kit (Thermo Scientific).

The PCR reaction was performed by denaturing the DNA at 95 °C for 5 minutes. This was followed by 35 cycles where the template DNA was denatured at 95 °C for 30 seconds. The primers were annealed 65 °C for 30 seconds and then extended for 30 seconds at 72 °C. A final extension step was performed at 72 °C for 10 minutes.

After the PCR, the reaction mixture was analysed by agarose gel electrophoresis for the presence of PCR product (Section 2.4.3.2.).

2.4.4.6. Sequencing and identification of PCR products

After confirmation of transformation, the plasmids (containing the correct insert size) were extracted from the clones with the QIAprep miniprep spin column kit (Qiagen) according to the manufacturer's instructions.

Sanger sequencing was done by the Central Analytical Facility at the University of Stellenbosch with the forward and reverse primers for PCR products from isolates PE14-3, 11, 12, 13, 17, 25, 40, 58, 61, 92, 104, 126, 144, 145, 146. Only the forward primer was used for the sequencing of PCR products from isolates PE14-22, 55, 60, 63, 74.

Isolates PE14-7, 10, 20, 27, 28, 30, 33, 64, 97, 108 were selected and used by colleagues. The 16S rRNA genes were partially amplified and sequenced with only the forward primer at Macrogen (Macrogenlab Inc., Rockville, USA). The sequences were processed by removal of low quality sequences and base calling. The software, ChromasPro version 1.5 (Technelysium Pty. Ltd.), was used to assemble the processed sequences when the amplicons were sequenced with the forward and the reverse primers.

The sequences were compared to those available on the 16S ribosomal RNA sequences (Bacteria and Archaea) databases of Genbank with the BLAST algorithm (Altschul *et al.*, 1997) through the National Center for Biotechnology Information website (NCBI; http://www.ncbi.nlm.nih.gov). The closest related species of the isolates were then identified.

3. Results and discussion

3.1. Cultivation of sponge symbionts and bioactivity screening

The sponge, PE14 (Figure 19), was identified to be *H. esperioides* of the class Demospongiae (Parker-Nance, 2013). It has a tough leathery exterior layer (pinacocyte) that is highly pigmented. The sponge body (mesohyl) was rather thin and partially

encrusted with rock and coral material. This is an indigenous species to South Africa which has been described in 1886 by Ridley and Dendy (Ridley & Dendy, 1886). As stated earlier, no other research has been done on this sponge species (to the knowledge of the author) other than the description of the type species. For further information of the taxonomy and collection site of this sponge, refer to Appendix A Table 6.

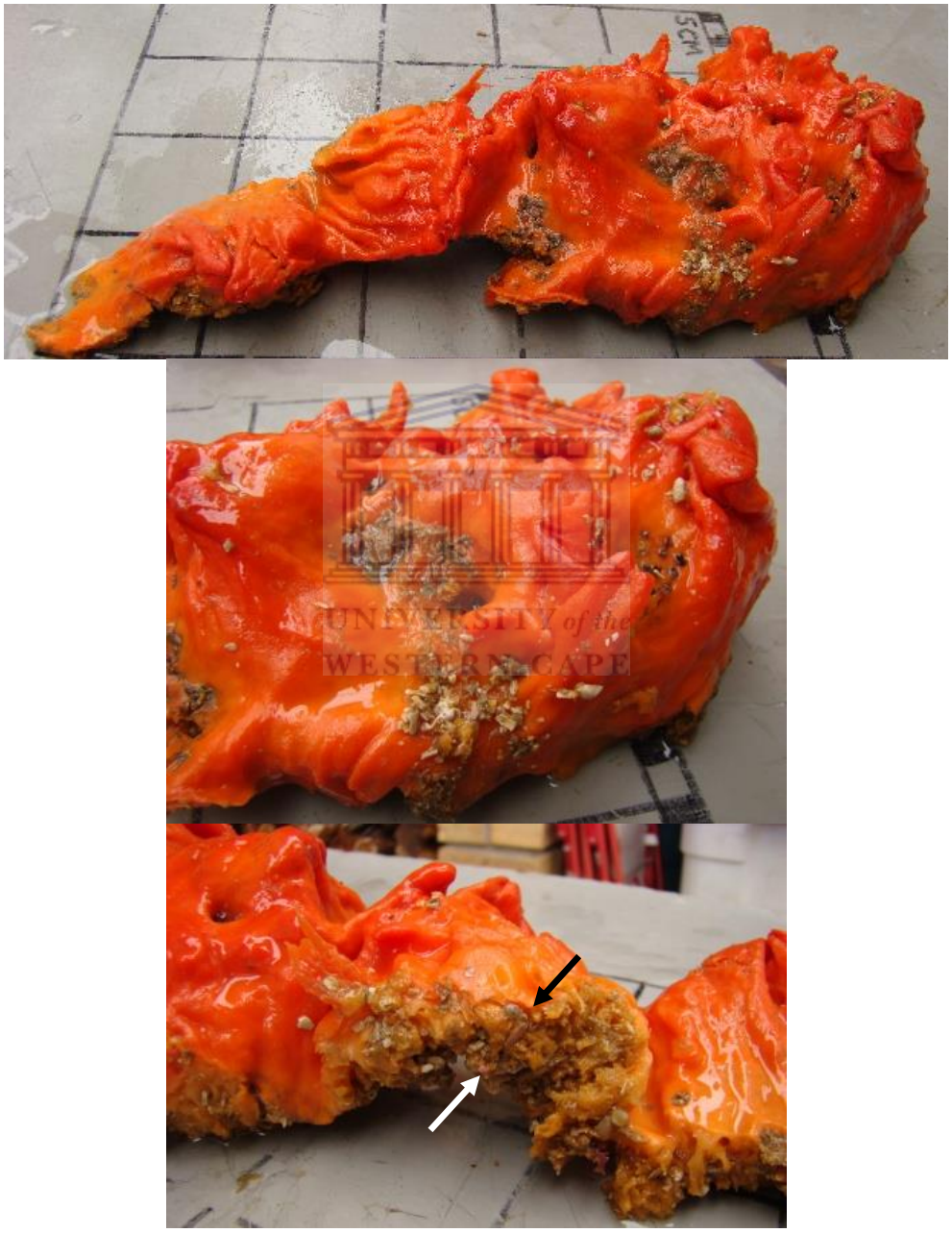


Figure 19 Sponge PE14 was identified as *H. esperioides*. It has a thin body, with little sponge material, which is encrusted with rock/coral debris (white arrow) and it has a thick leathery exterior layer (pinacocyte; black arrow). Photos are courtesy of Dr. Shirley Parker-Nance.

A wide variety of media was used to grow symbiotic bacteria from *H. esperioides*. Consequently, 150 isolates were obtained (Figure 20) and grown as pure cultures. Most of the isolates were cultured on the media favourable to aquatic bacteria (ZBA, ZBV, SWA, PMM and R2A) and firmicute selective media (TSA and SNA).

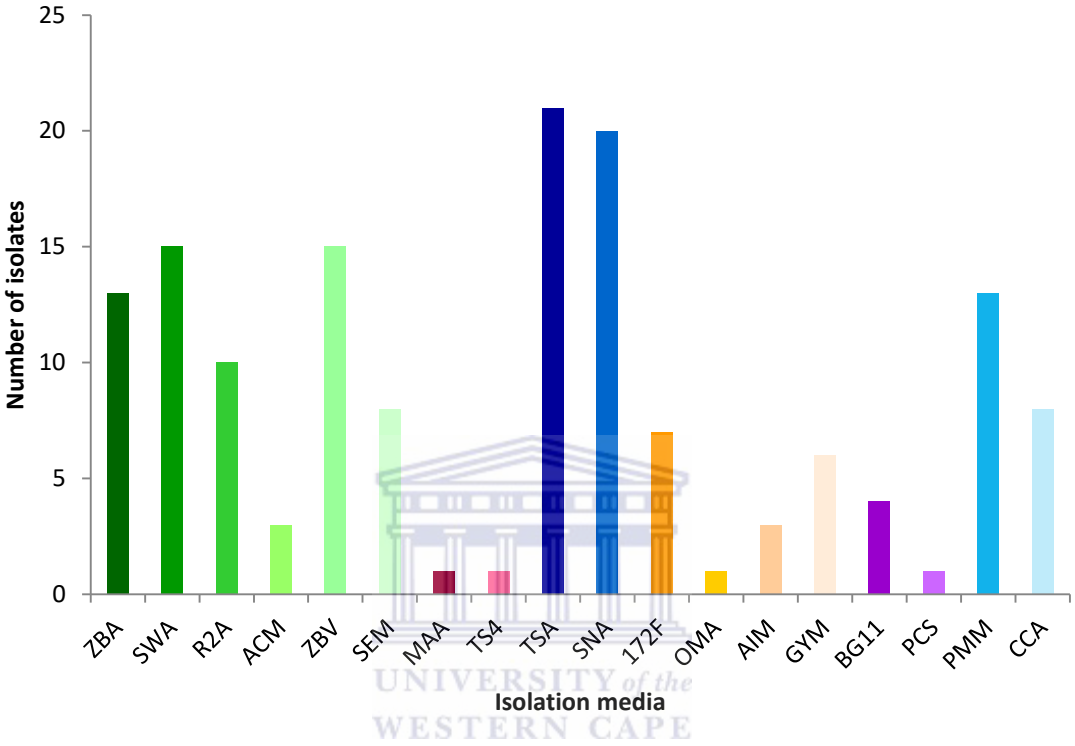


Figure 20 Number of isolates obtained from the different selective media. The media was grouped according to the selectivity as general (green), pH (pink), firmicute (dark blue), actinobacteria (orange), cyanobacteria (purple), other (light blue).

These isolates were screened for antimicrobial activity by an overlay assay with indicator strains. If antimicrobial activity is observed against one of these strains, it is an indication that the isolate that was overlaid with this organism is producing an antimicrobial secondary metabolite. The strains selected included Gram positive as well as Gram negative strains which are closely related to human pathogens (*E. coli* 1699, *B. cereus* ATCC10702, *S. epidermidis* ATCC14990, *M. smegmatis* LR222 *and P. putida* ATCC27853). The strain *E. coli* 1699 have been engineered by Cubist Pharmaceuticals to be resistant to 52 antibiotics. Its resistance covers a wide range of antibiotic classes with different mechanisms of action (Gross, 2011). This strain was selected as activity against it would indicate that a probable novel compound is produced by the isolate being tested.

For the antimicrobial screen, the isolates were grown on three different media before being overlaid as the production of secondary metabolites can be activated by changing the growth conditions of a culture (Sørensen & Sondergaard, 2014; Yoon & Nodwell, 2014; Pietra, 1990). Activity was observed by 27 of the isolates and 19 of these displayed activities against more than one strain (Table 6). Isolate PE14-74 was the only one to display antimicrobial activity against four of the indicator strains.

Table 6 Activity displayed against the indicator strains. The column total, summaries the number of indicator strains that were inhibited by the specific isolate.

PE14 ID	*E. coli* 1699	*M. smegmatis*	*S. epidermidis*	*B. cereus*	*P. putida*	Total
3	x	x	x			3
7		x		x		2
11		x		x		2
12		x		x		2
13		x	x	x		3
16			x			1
17		x	x	x		3
22		x	x	x		3
25		x	x	x		3
28		x	x			2
40	x			x		2
48			x	x		2
55		x	x	x		3
60		x	x	x		3
61		x	x	x		3
63		x	x			2
65			x			1
73			x			1
74	x	x	x	x		4
82				x		1
116		x	x			2
127		x				1
128		x				1
129		x	x	x		3
144			x			1
145		x				1
146		x	x			2

Green – isolates showing activity against three of the indicator strains.
Red – an isolate showing activity against four of the indicator strains.

Most of the antimicrobial activities observed were against *S. epidermidis* and *B. cereus* and when organisms were grown on GYM and ZBA (Figure 21). None of the bacterial isolates showed activity against *P. putida*, on any of the media.

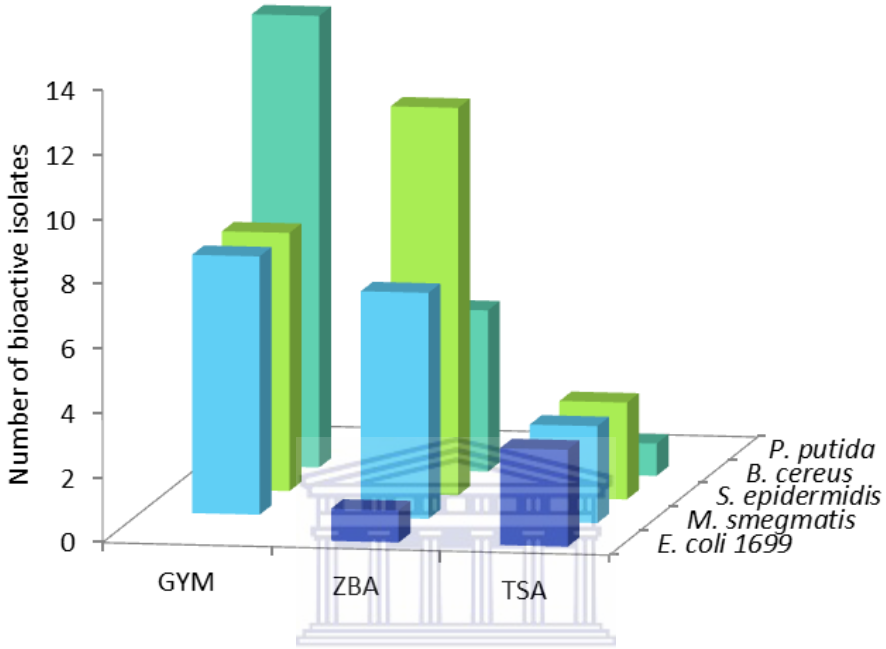


Figure 21 Number of isolates showing antimicrobial activity against the five indicator strains (*E. coli* 1699, *M. smegmatis*, *S. epidermidis*, *B. cereus* and *P. putida*) on different media (GYM, ZBA and TSA).

One of the interesting observations, when culturing the isolates for the antimicrobial screen, was that the colony morphology was irregular for some of the isolates when grown on GYM medium (Figure 22). This change in growth could be due to the change in the composition of the medium, as glucose and a high concentration of malt extract is included in the GYM medium. When isolates were grown on GYM, it was also noted that the $CaCO_3$ in the medium was utilised by 15 of the isolates. This was noted as a "zone of clearance" formed at the bottom of the plate where the $CaCO_3$ settled after the preparation of the agar (Figure 22).

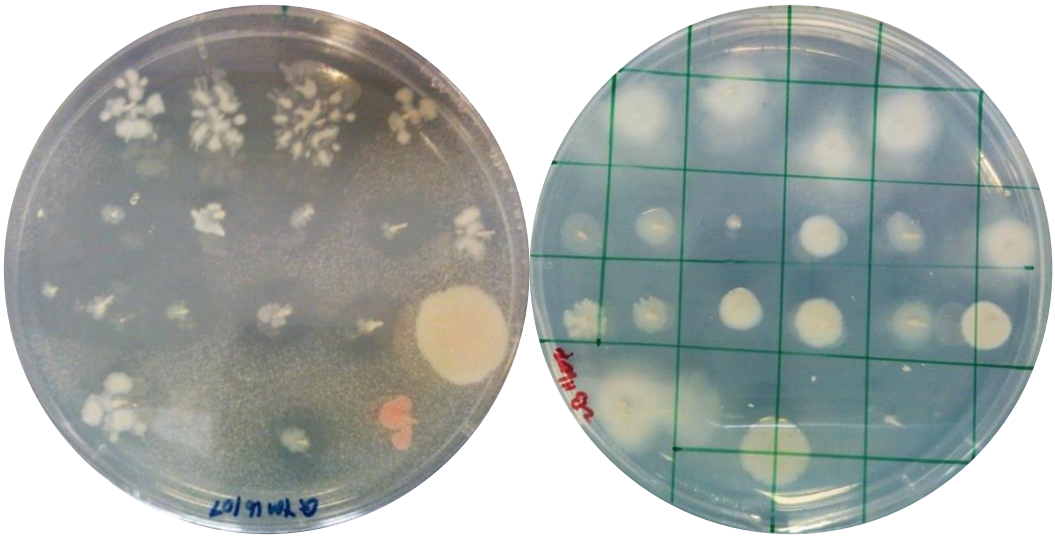


Figure 22 Comparison of the growth of the same isolates on GYM (left) and ZBA (right) medium. Irregular colony morphologies and $CaCO_3$ utilization (clear zones) was observed on GYM.

The production of the antimicrobial products may be more prevalent when the isolates were cultured on GYM, as it is rich in carbohydrates and proteins. This has been observed in several studies where it was found that small changes in growth conditions, such as different carbon and nitrogen sources, can influence the expression of secondary metabolites (Sørensen & Sondergaard, 2014; Yoon & Nodwell, 2014; Pietra, 1990). In a study where the concentration of glucose, peptone and minerals were varied it was found that glucose and peptone improved the production of antibiotics by the bacterium *Xenorhabdus nematophila* (Yoon & Nodwell, 2014). This approach of altering nutrient composition is also known as the one-strain-many-compounds (OSMAC) principle, and it is typically employed for drug discovery/screening (Bode *et al.*, 2002) as discussed in the literature review.

3.2. Screening liquid culture extracts by collaborators

This work was done by fellow colleagues in the laboratory and PharmaSea consortium partners. It is included to show the potential of the isolates for natural product screening and development for a range of biomedical applications.

Eleven of the 20 isolates screened by PharmaSea partners showed activity in the anti-fungal, anti-inflammatory and cell toxicity screens. Anti-fungal activity was detected by eight of the isolates and four isolates showed anti-inflammatory activity (Table 7) which may be of pharmaceutical importance. Extracts can be further purified to identify the

compound(s) with the desired bioactivity. These compounds can then be re-analysed to confirm whether they are cytotoxic.

Table 7 Activity displayed for the extracts from *H. esperioides* isolates that were screened by the PharmaSea partners.

PE14 ID	***A. fumigatus* ATCC 46645**	***C. albicans* MY155**	**Anti-inflammatory**
7	T, Z	T, Z	
12	Z	Z	
20	G, Z		
25	A		
34			Z
40	T		G, Z
48	A, Z		
55			Z
60			Z
104	Z	T, Z	
117	Z		

A = ACM, G = GYM, T = TSA, Z = ZBA.

3.3. Identification and dereplication of isolates

Some of the isolates (30) have been identified to their closest related species by amplifying and sequencing the 16S rRNA genes. A BLAST search of the sequences indicated to which species the isolates are closely related to. These isolates were found to be mostly from the genera *Vibrio* and *Pseudovibrio* (Table 8). A probable novel isolate, PE14-40, has an identity of 96% to the *Marinomonas dokdonensis* strain DSW10-10. This isolate is active against two of the indicator strains: *E. coli* 1699 and *B. cereus* (Table 6).

The diversity of bacteria found in sponges, include species from the phyla Acidobacteria, Actinobacteria, Bacteroidetes, Chloroflexi, Cyanobacteria, Deinococcus-Thermus, Firmicutes, Gemmatimonadetes, Nitrospira, Planctomycetes, Proteobacteria (Alpha-, Beta-, Delta-, and Gammaproteobacteria), Spirochaetes, and Verrucomicrobia (Taylor *et al.*, 2007). Marine species from the genera *Pseudovibrio* and *Vibrio* are regularly found to be sponge symbionts with culture-dependent and culture-independent methods and they are the dominant species in certain sponges (Flemer *et al.*, 2012;

Taylor *et al.*, 2007). These species have also shown antimicrobial activity against a broad range of strains (Flemer et al., 2012; O' Halloran et al., 2011; Kennedy et al., 2008; Muscholl-Silberhorn et al., 2008).

Table 8 Number of isolates identified according partial 16S rRNA sequences to genus level.

Phylum	**Genus**	**Number of isolates**
Alphaproteobacteria	*Pseudovibrio*	6
Gammaproteobacteria	*Vibrio*	15
Gammaproteobacteria	*Aliivibrio*	1
Gammaproteobacteria	*Enterovibrio*	1
Firmicutes	*Bacillus*	1
Gammaproteobacteria	*Marinomonas*	1
Gammaproteobacteria	*Shewanella*	3

Dereplication, of the number of organisms and bioactive compounds that are screened for drug discovery, is a routine procedure in the pharmaceutical industry to reduce the chance of repetitive work and to make drug discovery a high throughput process. Many studies dereplicate based on the 16S rRNA sequence. This is at a specified cut-off similarity percentage, typically 97 %, where organisms can be differentiated at species level (Stackebrandt & Goebel, 1994). However, in addition to PCR and sequencing errors which get introduced, it is well-known that taxonomic dereplication is not necessarily analogous to the metabolic capacity of organisms, particularly when it concerns secondary metabolite capabilities (Ridley *et al.*, 2005). Thus, the bioactivities of the isolates are compared together with identity of the 16S rRNA gene. For example, if more than one isolate is closely related to the same strain and they display the same bioactivities it is likely to be the same organism that was isolated as multiple isolates from the sponge.

In our study there were five isolates (PE14-3, 11, 12, 13, 20) closely related to *Pseudovibrio ascidiaceicola* strain F423 of which only PE14-11 and -12 displayed antimicrobial activity against the same indicator strains (Table 9, blue text). Similarly there were four isolates (PE14-17, 55, 60, 63) closely related to *Vibrio tasmaniensis* strain Carson D39 of which PE14-17, 55 and 60 displayed the same antimicrobial

activities (Table 9, orange text). If the bioactivities displayed in the screens from the PharmaSea partners are also considered (Table 7), only PE14-55 and -60 displayed the same bioactivities throughout this research (thus far) and could be considered to be the same strain. Even though isolates related to the same strain, such as PE14-11 and 12, showed similar antimicrobial activities it is difficult to dereplicate the isolates only on a few characteristics.

This was similarly observed by O'Halloran *et al.* (O' Halloran et al., 2011). The 72 strains of *Pseudovibrio* spp., isolated from three sponge species, were identified to belong to 33 types with molecular methods. For most of the types with multiple isolates the individual isolates had identical antimicrobial activity profiles. This suggests that the same isolates may have been selected during the initial isolation process. Opposing this, they also observed that an individual isolate had a different bioactivity profile than the rest of the isolates of the same type. This stress the importance of applying dereplication with great caution as isolates can be closely related but still have the ability to produce novel compounds and vice versa.

4. Summary

This study shows that the sponge *H. esperioides* has many bacterial symbionts which produce bioactive secondary metabolites, including those with antimicrobial and anti-inflammatory activities. This indicates that the respective compounds might be of importance for further application in the pharmaceutical industry.

Table 9 Identification of the PE14 isolates according to partial sequences of their 16S rRNA genes. The column total bp is the total base pairs included in the BLAST search.

Total bp	PE14 ID	Closest Related Species	% ID	Bioactivity
1329	3	*P. ascidiaceicola* strain F423	100	EMS
1330	11	*P. ascidiaceicola* strain F423	100	MB
1330	12	*P. ascidiaceicola* strain F423	100	MB
1329	13	*P. ascidiaceicola* strain F423	100	MSB
1359	17	*V. tasmaniensis* strain Carson D39	99	MSB
674	55	*V. tasmaniensis* strain Carson D39	99	MSB
712	60	*V. tasmaniensis* strain Carson D39	99	MSB
662	63	*V.tasmaniensis* strain Carson D39	99	MS
970	92	*Vibrio kanaloae* strain LMG 20539	100	E
744	74	*V. kanaloae* strain LMG 20539	99	EMSB
1217	61	*Vibrio lentus* strain CIP 107166	99	MSB
977	104	*Vibrio pomeroyi* strain CAIM 578	100	EM
1209	144	*Vibrio gigantis* strain LGP 13	100	S
614	22	*Vibrio splendidus* LGP32 strain LGP32	99	MSB
1383	58	*Shewanella sairae* strain SM2-1	99	S
1297	145	*S. sairae* strain SM2-1	100	M
956	146	*Shewanella pneumatophori* strain SCRC-2738	100	MS
1366	25	*Enterovibrio norvegicus* strain LMG 19839	98	MSB
1373	40	*M. dokdonensis* strain DSW10-10	96	EB
1393	126	*Staphylococcus warneri* strain SG1	100	-
799	7	*P. ascidiaceicola* strain NBRC 100514	99	MB
743	20	*P. ascidiaceicola* strain F423	100	E
435	10	*Vibrio gallaecicus* strain CECT 7244	98	E
360	97	*V. gallaecicus* strain CECT 7244	99	M
501	27	*V. atlanticus* strain VB 11.11	98	E
546	64	*V. atlanticus* strain VB 11.11	99	E
422	28	*V. splendidus* strain LGP32	98	EMS
320	30	*Aliivibrio wodanis* strain ATCC BAA-104	99	EM
670	33	*Photobacterium aplysiae* strain GMD509	98	E
934	108	*Bacillus anthracis* strain Ames	99	E

E = *E. coli* 1699, M = *M. smegmatis*, S = *S. epidermidis*, B = *B. cereus*, P = *P. putida*.
Green: isolates used and identified by colleagues.

Chapter 3 - Genomic Screening for RiPPs

1. Introduction

The class of natural product peptides that are synthesized by ribosomes and post-translationally modified into an active product are labelled as ribosomal peptides (RiPPs). The biosynthetic genes, arranged within an operon or cluster, play a role in the structure, modification, export and regulation of the RiPP biosynthesis. The structural gene, which becomes the precursor peptide, consists of a leader and core peptide. For many RiPPs the function of the leader peptide is not clear, but they seem to assist the post-translational process (Patton *et al.*, 2008; Yang & van der Donk, 2013). It is the core peptide that will be modified by the post-translational enzymes into the active natural product.

The interest in researching RiPPs as pharmaceutical drugs has increased over the last 15 years. One of the driving factors for this is the fact that RiPPs are structurally diverse, with more than 17 classes (Arnison *et al*., 2013). Some of the diversity is brought about by post-translational modifications that are so different to traditional ribosomal peptide synthesis that the active product is sometimes not recognised as a product of ribosomal synthesis (Kersten *et al.*, 2011). This is illustrated when the synthesis of polytheonamides was researched, as they were thought to be products of non-ribosomal synthesis. Recently the Piel laboratory discovered the ribosomal biosynthetic cluster of these peptides (Freeman *et al.*, 2012) within the metagenomic DNA library of the sponge, *Theonella swinhoei*. It was hypothesised that the peptides are of bacterial origin and further research by the team led to the discovery of the uncultivable (hitherto) producing organism, a species of the candidate genus, '*Entotheonella'* (Wilson *et al.*, 2014).

Since many uncultivable bacteria are potential sources of other bioactive secondary metabolites, cultivation-independent screening methods, such as metagenomics, have been developed. Metagenomics is ideally suited for the screening (also referred to as mining) of RiPPs, for a number of reasons. These include, but are not limited to, the conserved nature of the leader peptide (Schmidt, 2010) and the clustering of the biosynthetic genes. Traditionally, cultivation-independent screening methods require the construction of mDNA libraries that are screened either functionally or via sequence analysis. Examples of screening methods based on sequence analysis include

homology-based techniques, such as PCR, and those based on sequencing and bioinformatic analysis. The initial methods based on the sequencing of libraries proved to be unrefined, expensive and not widely available to researchers. However, advances in sequencing methods made the technology a financially viable option for more researchers. The improved sequence-based methods now enable one to screen mDNA without the construction of libraries, referred to as next-generation sequencing (NGS; Suenaga, 2012; Heidelberg *et al.*, 2010).

Homology-based screening methods such as PCR require the design of primers to conserved regions. The primers can be designed through two approaches. The first approach (Figure 23, A) is dependent on knowledge of the structure and biochemical characteristics of the compound. This knowledge can be used to deduce the sequence of the structural genes, which in the case of RiPPs, are the precursor genes. This approach is also referred to as retrobiosynthetic analysis (Piel, 2011; Huo *et al.*, 2012). This approach can be easily implemented for preliminary screens to determine if the RiPP of interest is present within a sample. The screening can then be enriched with complimentary approaches to reveal the entire biosynthetic pathway. This can be achieved by screening the metagenomic library or NGS of the mDNA.

The second approach relies on a characterised biosynthetic pathway of the compound (Figure 23, B). Primers can be designed either to bind to the precursor gene, or to key maturation enzymes within the biosynthetic pathway. Note that the second approach differs from the first approach, by not being dependent on prior knowledge of the compound.

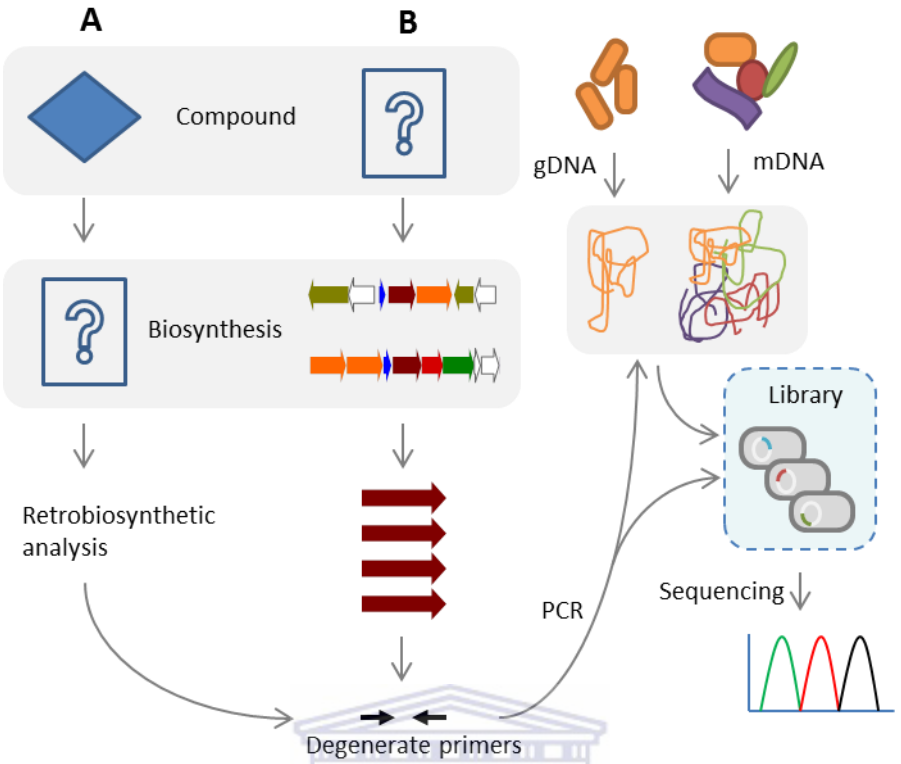


Figure 23 Sequence-based screening can be dependent on a characterised compound (A) or biosynthetic cluster (B). The construction of a DNA library for screening is optional.

Both approaches have been successfully used for the screening of RiPPs. Notable examples include polytheonamides and lasso peptides. The Piel laboratory used the first approach where structural and biochemical knowledge of the compound (polytheonamide) was used to infer the genomic sequence of the precursor gene (Freeman *et al.*, 2012). The mDNA library of *T. swinhoei* was screened with primers and the precursor gene was successfully amplified. Further research led to the characterisation of the entire biosynthetic pathway. The Link laboratory used a combination of both approaches where the structure of various lasso peptides, together with the conserved features within the known biosynthetic clusters, were used with bioinformatics to screen for novel peptides (Maksimov *et al.*, 2012; Maksimov & Link, 2014).

This chapter describes the development of a cultivation-independent screen for RiPPs. This entails the design and validation of primers, the extraction of mDNA and the screening of the extracted DNA with the primers.

2. Materials and methods

2.1. Sponge material

In this study, sponge material from PE14 was used to extract mDNA (Appendix B; Table 6). The mDNA extracted from the other sponges (Table 10) was kindly provided by colleagues at IMBM.

Table 10 Sponges and their associated mDNA sample names.

Sponge	Species	mDNA extracted	Sample name (mDNA)
PE03	*Guancha* sp. 001RSASPN	L. Black	PE03
PE05	*Waltherarndtia caliculatum*	L. Black	PE05-A
PE05	*W. caliculatum*	L. van Zyl	PE05-B
PE08	*Spongia* (Spongia) sp.001RSASPN	R. Matobole	PE08
PE14	*H. esperioides*	This study	PE14-A
PE14	*H. esperioides*	J. Navarro-Fernández	PE14-B

2.2. Bioinformatics and primer design

2.2.1. Analysis of the polytheonamide biosynthetic cluster

The BLAST Link (BLink) tool of NCBI, which displays the results of precomputed BLASTp searches of all the proteins against the non-redundant (nr) protein database, was used to analyse the homologues of the biosynthetic proteins of the polytheonamide gene cluster (accession: JX456532.1). The data was filtered as follows. Only hits from the RefSeq database, with a BLAST score of more than a 100, and only from the taxonomic groups: bacteria and archaea were included. Also, for proteins with identical sequences a representative protein was included.

2.2.2. Design of primers

Degenerate primers were designed to the conserved residues of two genes (*poyB* and *poyC*) of the polytheonamide biosynthetic cluster and other homologous genes.

The PoyB and PoyC protein sequences (accession: AFS60637.1 and AFS60638.1) were obtained from the NCBI protein database. Sequences of the specific proteins to the

family model ribosomal peptide maturation radical SAM protein 1 (short name: rSAM_ocin_1; accession: TIGR03975) was accessed from the conserved domain database (CDD) of NCBI. These sequences together with PoyB and PoyC were aligned with the BioEdit Sequence Alignment Editor (Hall, 1999) and the Clustal W multiple alignment algorithm (Thompson *et al.,* 1994). A maximum parsimonious (MP) tree was constructed from this alignment with the MEGA5 software (Tamura *et al.,* 2011) to identify the proteins that are more closely related to PoyB and PoyC. All parameters were set to the default value. Bootstrap analysis (1000 replicates) was performed (Dunbar & Mitchell, 2013) to evaluate the support of the clades. Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates were collapsed. The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm (pg. 126 in reference: Knerr & van der Donk, 2012) with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). All positions containing gaps and missing data were eliminated.

Conserved residues were obtained by alignment of the selected proteins (Hall, 1999; Thompson *et al.*, 1994) and by setting the similarity threshold to 80 %. Degenerate primers were designed to these conserved residues by using the IUPAC code for incompletely specified bases in nucleic acid sequences (Cornish-Bowden, 1985). To reduce ambiguity the nucleotide sequences of the *poy* genes were compared to the reverse-translated sequences of the proteins. Primers were further analysed with OligoAnalyzer 3.1 (Owczarzy *et al.,* 2008) for any hairpin formation or homo- or hetero-dimer formation.

PoyA (precursor peptide) was aligned with the selective proteins of the family, TOMM_pelo (accession: TIGR03793), similar to the above-mentioned method.

2.2.3. *In silico* PCR

The amplification of designed primers was also tested by performing an *in silico* PCR with an online tool (http://insilico.ehu.es; Bikandi *et al.*, 2004). The parameters of the *in silico* PCR were varied to allow one or two mismatches, but not in the last nucleotides of the 3' end of the primers (Table 11). When available, plasmids were included as templates.

Table 11 Parameters used when the *in silico* PCR simulation was performed. A) The number of mismatches allowed. B) The number of nucleotides, at the 3' ends of the primers, where mismatches were not allowed.

A	B	A	B
1	2	2	2
1	3	2	3
1	5	2	5
1	10	2	10

2.3. Nucleic acid extraction

2.3.1. Genomic DNA extraction

2.3.1.1. PE14 isolates

Plate cultures (stored at 4 °C) of the isolates were used as inoculants for the liquid cultures from which gDNA was extracted as described in (Section 2.4.1; Chapter 2). The gDNA of isolates in Table 12 was extracted.

Table 12 Isolates (PE14) cultured for gDNA extraction.

	Isolates
Group 1	48, 64, 65, 144
Group 2	22, 28, 60, 92, 104
Group 3	30, 61, 63, 91, 107
Group 4	16, 17, 55, 74, 108W
No group	7, 11, 12, 13, 25, 27, 70, 97, 106

2.3.1.2. Sinorhizobium meliloti

A plate culture of *S. meliloti* 1021 was prepared from a glycerol stock stored at -80 °C. The plates were prepared with TY medium (5 $g.L^{-1}$ tryptone, 3 $g.L^{-1}$ yeast extract, 10 ml of sterile 1 M $CaCl_2$). The $CaCl_2$ stock solution was autoclaved separately and the appropriate amount was added before plates were poured or inoculation of liquid culture (Beringer, 1974). Cultures were incubated at 28 °C for 2 days. A single colony was picked from the culture, inoculated into 50 ml TY broth and incubated until the stationary phase of growth was reached. The cultures were used to extract the genomic

DNA of the organism with a large-scale preparation of the method of Wilson (2001). The DNA obtained was stored at 4 °C and used directly as template DNA in the PCR.

2.3.1.3. Extraction of plasmid DNA (*poyB* gene)

A construct with the *poy*B gene was kindly provided by the Piel laboratory (ETH Zurich, Department of Biology, Institute of Microbiology, Switzerland). The construct, Nhis-PoyB-pCDFDuet-1, was transformed into electrocompetent *E. coli* GeneHogs cells (Section 2.4.4.2; Chapter 2) and the QIAprep Miniprep kit (Qiagen) was used according to the manufactrer's instructions to extract the plasmid DNA. The DNA from the miniprep was diluted and stored at 4 °C. When needed the DNA (now referred to as *poyB*-plasmid DNA) was used directly as the template in the PCRs.

2.3.2. Metagenomic DNA extraction and analysis

Due to difficulty in generating high quality mDNA suitable for downstream applications, seven different extraction methods were used.

2.3.2.1. Modified CTAB lysis

Extraction was carried out at 4 °C by keeping samples on ice or setting the temperature on centrifuges, unless specified otherwise.

Frozen sponge material was grounded to a fine powder with a sterile pestle and mortar under liquid nitrogen. Two grams of the material was suspended in 8 ml lysis buffer (15 mM Tris, 15 mM EDTA, 1.4 M NaCl, 1 % CTAB, with 2 % SDS, pH 8.0) with lysozyme to a concentration of 200 µg.ml^{-1}. The mixture was blended, but to prevent shearing of the DNA the tube was only inverted once before and once during incubation at 37 °C for 2 hours. Proteinase K (100 µg.ml^{-1}) was added before inverting the tube once and incubating it at 65 °C for 2 hours. The sample was then centrifuged for 15 minutes at 4 500 xg. The supernatant was transferred to a clean tube and centrifuged as before until most of the sponge debris and polysaccharides were removed. The supernatant was then treated twice with an equal volume of a phenol:chloroform:isoamyl alcohol solution (PCI: 24:24:1 (v/v)) and centrifuge at 6 000 xg for 15 minutes. The DNA was precipitated with 0.7 times the volume of isopropanol and left at room temperature for 30 minutes. The DNA was pelleted by

centrifugation at 6 000 xg for 30 minutes. The supernatant was discarded and the pellet was washed three times with 70 % ethanol. After the pellet was air-dried at room temperature, it was dissolved in TE buffer (the volume depended on the size of DNA pellet) and stored at 4 °C.

To remove impurities from the extracted mDNA it was treated a second time with PCI and precipitated, washed, dried and resuspended in TE buffer.

The method was modified to enrich for the extraction of bacterial DNA. The sponge material was defrosted overnight at 4 °C and 60 g (wet weight) was homogenised in a blender with 250 ml TE buffer. The pellet obtained after enrichment (Section 2.3.2.1.1.) was transferred to a 15 ml Falcon tube and lysis was performed as described above, but without SDS in the lysis buffer. After lysis, SDS was added to a concentration of 2 % (v/v) together with the proteinase K. All of the centrifugation steps following lysis was performed at an increased speed of 10 000 xg. The supernatant was treated with PCI as described above until the aqueous layer was non-cloudy.

2.3.2.1.1. Enrichment for bacterial DNA

A modified enrichment step for the enrichment of bacterial cells (Ouyang *et al.*, 2010) was included where the pellet was resuspended in 25 ml TE buffer and filtered with a 70 µm cell strainer (BD Biosciences) by centrifugation at 10 000 xg for 20 minutes.

2.3.2.2. Agarose embedded cell lysis method

A method where cells are suspended in agarose gel plugs (Liles *et al.*, 2008) before extraction of mDNA was modified as follows. Ten grams of sponge tissue was homogenised in a sterile blender in 250 ml TE buffer. The homogenate was centrifuged twice at 300 xg for 2 minutes and the supernatant was transferred to a clean tube. The cells were then pelleted by centrifugation at 10 000 xg for 20 minutes. The cells were enriched for bacterial cells (Section 2.3.2.1.1.) and after filtration the pellet was resuspended in 1 ml TE buffer. The cell-suspension was heated to 45 °C in a water bath. Standard melting point (SMP) agarose (DNA typing grade; SeaKem LE Agarose, Lonza), prepared with TE buffer, was preheated to 65 °C. Equal volumes (0.5 ml) of the

agarose and the cells were mixed and set in a 1 ml syringe at 4 °C. The final concentration of the agarose was 1 % (w/v).

The plugs were incubated overnight at 37 °C in a modified lysis buffer (50 mM Tris, 50 mM EDTA, 1.4 M NaCl, 2 % CTAB, pH 8.0) with RNase A (3 $\mu g.ml^{-1}$) and lysozyme (5 $mg.ml^{-1}$). The buffer was replaced with and proteinase K (0.2 $mg.ml^{-1}$) was added before incubating the plugs for 24 hours at 55 °C. The plugs were washed three times for 10 minutes in TE buffer. The plugs were treated with the formamide and salt solution (Section 2.3.2.8.5.) to purify the DNA. Samples of the plugs were subjected to pulse field gel electrophoresis (Section 2.3.2.9.1.) to elute the DNA from the plugs.

2.3.2.3. Modified urea lysis

This method is described by Gurgui and Piel (Gurgui & Piel, 2010). Modifications to this method included the enrichment of bacterial cells (Section 2.3.2.1.1.). Also, TE buffer, used for homogenisation of the sponge material, was replaced with a Tris-NaCl buffer (10 mM Tris-HCl, 0.5 M NaCl, pH 8.0) before the extraction was done. Tris-NaCl buffer was added to the ratio of 10 to 20 ml per gram of sponge material. The cell pellet obtained was treated with the sponge lysis buffer (8 M urea, 2 % sarkosyl (Sigma–Aldrich Co.), 1 M NaCl, 50 mM EDTA, 50 mM Tris–HCl, pH 7.5) as instructed. The CTAB purification step was not performed.

Other extractions attempts with this method (Gurgui & Piel, 2010) included a modified bead-beating method of Hardoim *et al.* (Hardoim *et al.*, 2009). Shortly: the amount of sponge material used was increased to 10 g (wet weight). The sponge material was ground for 1,5 minutes with a mortar and pestle in 10 ml Tris-NaCl buffer. The material was transferred to an Erlenmeyer flask with 35 ml Tris-NaCl buffer and 5 ml sterile glass beads. The flask was then incubated for 20 minutes at RT on a rotator (170 rpm). The material was enriched for bacterial cells (Section 2.3.2.1.1.) and a washing step (Ouyang *et al.*, 2010) of the cell pellet with Tris-NaCl buffer was also included. The extraction procedure was then performed with the resulting cell pellet.

2.3.2.4. MoBio UltraClean Soil DNA Isolation kit

Sponge material was thawed at 4 °C. Samples were prepared for extraction with the MoBio UltraClean Soil DNA Isolation kit as follows. For the first sample was 0.25 g (wet weight) sponge material was used, since the manufacturer recommended that 0.25 to 1 g of soil should be used. For the other samples, 3 g sponge material (wet weight) was incubated for 5 minutes with 2 ml Tris-NaCl buffer. After incubation, the liquid was pressed out with a sterile hockey stick. The liquid was captured in a clean tube and centrifuged at 300 xg to obtain a pellet (sample two). The supernatant was transferred to a clean tube and centrifuged again at 5 000 xg (sample three). The supernatant was transferred as before and centrifuged at 10 000 xg (sample four). DNA was extracted from samples 1 to 4 as instructed by the manufacturer.

2.3.2.5. A combination of previous extraction methods

The initial objective was to extract HMW mDNA from the sponge PE14 to construct a library. Previously extracted mDNA was inadequate for further analysis or construction of a DNA library and the objective changed to extract DNA of sufficient quality for further analysis. A combination of some of the previous extraction methods was used to improve the quality and quantity of the mDNA.

Extraction was carried out at 4 °C, by keeping samples on ice or setting the temperature on centrifuges, unless stated otherwise.

Two different approaches were used to separate the bacterial cells from the sponge material. At first, sponge material (1.5 g; wet weight) was weighed and ground to a fine powder under liquid nitrogen. With the second approach, 1.5 g (wet weight) of sponge material was weighed in a 15 ml tube. A sterile hockey stick was used to press out the liquid from the sponge. The liquid was captured in a clean tube. The remaining sponge material (after the liquid was pressed out) was also used for extraction.

TE buffer (5 ml) and a few sterile glass beads (about 4 mm) were added to the ground sponge material and the material from which the liquid was pressed out. The samples were then placed within a MoBio Vortex Adapter and shaken for 30 minutes at room temperature (RT).

All samples were centrifuged at 300 xg for 5 minutes. The supernatants were then filtered to enrich for bacterial cells as described in Section 2.3.2.1.1. The volumes of the samples were adjusted with TE buffer to a total of 5 ml and not 25 ml as described in Section 2.3.2.1.1.

The cell pellet from the sponge material that was ground under liquid nitrogen was combined with the pellet from the cells of the liquid that was pressed out of the sponge material, because of the small size of the pellet of the liquid (now referred to as sample 1). The pellet that was obtained from the sponge material after the liquid was pressed out is now referred to as sample 2.

Plugs were made (Section 2.3.2.1.) with the samples and divided to be lysed with the urea lysis buffer (Section 2.3.2.3.) and the CTAB lysis buffer (Section 2.3.2.1.; Table 13).

Table 13 Lysis treatment of the plug samples.

Tube	Sample	Lysis buffer
A	1	Urea
B	2	Urea
C	1	CTAB
D	2	CTAB

After lysis the plugs were washed twice with TE buffer for an hour at RT. Plugs from tube A and B was treated with the formamide solution (Section 2.3.2.8.5.) for 8 hours at RT. The plugs were then washed three times for 15 minutes and overnight in TAE buffer before electrophoresis was performed to extract the DNA (Section 2.3.2.10.). Plugs from tube C and D was treated a second time with the formamide solution overnight. The plugs were then dialysed overnight with TE buffer.

2.3.2.6. Crude mechanical and enzymatic extraction

One gram of sponge material (wet weight) was cut into smaller pieces and 5 ml TE buffer was added with sterile glass beads to 15 ml tubes. The tubes were homogenised by incubation at RT for an hour on the MoBio vortex adaptor. Enrichment for bacterial cells was performed as described in Section 2.3.2.1.1.

The cell pellet was resuspended in 5 ml TE buffer and sonicated for a second with a SONOPULS ultrasonic homogenizer (Bandelin). Lysozyme was added (200 $\mu g.ml^{-1}$) to

the buffer and the sample was incubated for 30 minutes at 37 °C. After cell lysis, proteinase K was added (100 µg.ml^{-1}) and the sample was incubated overnight at 55 °C.

The sample was treated with PCI until the aqueous phase was clear and the DNA was precipitated with isopropanol (Section 2.3.2.1.). After evaporation of the isopropanol, the DNA was resuspended in 100 µl TE buffer. The dissolved DNA was pooled and reduced (by evaporation at 60 to 70 °C) to a volume of 100 µl.

2.3.2.7. Modified yeast DNA extraction protocol

Extraction was performed at 4 °C, by keeping samples on ice or setting the temperature on centrifuges, unless stated otherwise.

Sponge material was thawed at 4 °C. Three samples of 1 g (wet weight) sponge material were cut into smaller pieces before transferring it to clean 15 ml tubes. Sterile beads were added together with 10 ml of TE buffer. The samples were homogenised on a MoBio vortex adaptor for a total of 30 minutes. The sample was vortexed for 5 minutes intervals and in between vortexing the sample was cooled on ice for 5 minutes. The samples were centrifuged for 5 minutes at 300 xg. The supernatants of the three samples were pooled for enrichment of bacterial cells (Section 2.3.2.1.1.). The pellet was resuspended in 5 ml TE buffer and stored as 1 ml aliquots at -20 °C.

A tube of the prepared cells (stored at -20 °C) was thawed on ice and pelleted by centrifugation at 13 000 rpm for 5 minutes. The method for isolation of gDNA for southern blot analysis of Hoffman and Winston (Hoffman & Winston, 1987) was used. The following adjustments were made: TE buffer was not added before 100 % ethanol was added. After ethanol was added, the sample was incubated on ice for 10 minutes. The sample was also not treated with RNase.

2.3.2.8. Purification of mDNA

2.3.2.8.1. RNase treatment

RNase (0.1 mg.ml^{-1}) was added to the mDNA and incubated for one hour at 37 °C. The sample was treated with PCI (Section 2.3.2.1.) and precipitated by adding twice the volume of ice cold 100 % ethanol and one tenth the volume of 3 M sodium acetate (pH

7.0). The DNA was incubated for 2 hours at -20 °C and pelleted, washed, dried and resuspended as described before.

2.3.2.8.2. CTAB precipitation of polysaccharides

This purification method was described by Gurgui and Piel (2010). CTAB was added to a final concentration of 1 % (w/v).

2.3.2.8.3. PVPP columns

If amplification of the 16S rRNA genes was unsuccessful the mDNA was further purified with polyvinylpolypyrrolidone (PVPP) spin columns (Berthelet *et al.*, 1996). In brief, columns were prepared by stacking a cut, filtered-tip with PVPP (Sigma; P-6755). The columns were equilibrated with TE buffer and dried by centrifugation. The DNA was applied and eluted by centrifugation.

2.3.2.8.4. 2. Drop dialysis

In a sterile Petri dish, a 0.025 μm membrane (Millipore, VSWP01300 MF-Millipore) was floated on sterile water. The mDNA was placed (in aliquots of 10 μl) on the filter and incubated for at least 10 minutes. The dish was gently swirled a few times during incubation. After incubation, the mDNA was transferred to clean tube(s).

2.3.2.8.5. Formamide treatment

DNA was purified by following the method described by Newman *et al.* (2010). Plugs were slowly rotated while being incubated at room temperature.

2.3.2.9. Recovery of mDNA from plugs

2.3.2.9.1. Pulse field gel electrophoresis (PFGE)

Initially a 1 % (w/v) SMP agarose gel was prepared with 0.5 times TBE buffer. Electrophoresis was performed with the CHEF-DR II Pulsed Field Electrophoresis System (BioRad, South Africa) in 0.5 times TBE buffer with the following programme: the field strength was 6 $V.cm^{-1}$ (200 V) with the switch times set to be initially 1 second and then 6 seconds. After electrophoresis (approximately 22 hours), the gel was stained with ethidium bromide before visualisation. If electrophoresis of DNA into the gel was

observed, the experiment was repeated with a gel that was prepared with a low melting point agarose gel (LMP; 0.8 % (w/v)) section (Figure 24).

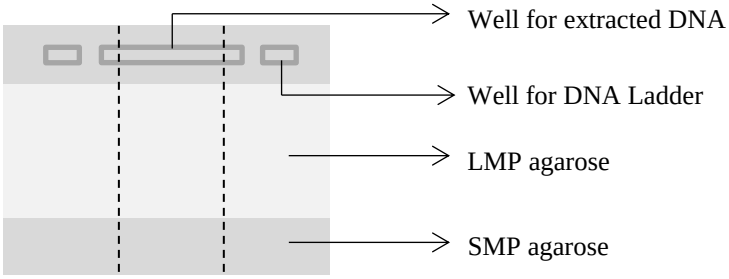


Figure 24 Electrophoresis setup for size selection of extracted DNA. The dashed lines indicate where the gel was cut after electrophoresis. A standard melting point (SMP) agarose gel was prepared and a part of it was replaced with low melting point agarose gel (LMP).

2.3.2.9.2. Size selection and recovery of mDNA from LMP

The method was followed as described by Gurgui and Piel (2010) with the following alterations. A SMP agarose gel was partially replaced with a LMP section (Figure 24). Overnight electrophoresis was performed at 10 V. The mDNA was recovered from the LMP gel with agarase (Thermo Scientific) according to the instructions of the manufacturer.

2.3.2.9.3. GeneJET Gel Extraction Kit

The GeneJET Gel Extraction Kit (ThermoScientific) was used to extract the mDNA from normal agarose plugs C and D (Section 2.3.2.5). With the first attempt 200 mg and with the second attempt 500 mg of the plugs were used according to the manufacturer's instructions.

2.3.2.10. Qualitative and quantitative analysis of mDNA

The quality and quantity of the extracted mDNA was determined with spectrophotometric or fluorometric analysis (Section 2.4.3.1.; Chapter 2), as well as agarose gel electrophoresis (Section 2.4.3.2.; Chapter 2). DNA was electrophoresed in a 0.8 to 1 % (w/v) agarose gel at 20 V. Ethidium bromide was not added during the preparation of the gels when the mDNA was required for further analysis.

2.4. PCR analysis

2.4.1. Validating purity of the mDNA

The amplification of the 16S rRNA genes within the mDNA was analysed with the universal eubacterial primers E9F and U1510R (Section 2.4.2; Chapter 2) to validate if the DNA was suitable for down-stream screening purposes. PCR reaction mixtures (20 µl) were prepared with DreamTaq polymerase (Fermentas, Thermo Scientific) according to the manufacturer's instructions.

The PCR reaction was performed by denaturing the DNA initially at 94 °C for 5 minutes. This was followed by 35 cycles where the template DNA was denatured at 94 °C for 30 seconds; the primers were annealed at 50 °C for 30 seconds and then extended for 1 minute 25 seconds at 72 °C. A final extension step of 10 minutes was performed at 72 °C.

The gDNA of *Dietzia maris*, obtained from a colleague (W. Mavengere), was a positive control to this PCR and a reaction with no template DNA was the negative control.

To determine if inhibitors were present in the mDNA sample, an internal control can be used (Dunbar & Mitchell, 2013). The internal control for this study, the gDNA of *D. maris*, was added in equal volumes to the mDNA and the resultant DNA, referred to as spiked-mDNA, was used as a template at various concentrations in the PCR.

2.4.2. PCR screening for '*Entotheonella'* spp. and RiPP genes

2.4.2.1. Ocin1 and '*Entotheonella'*-specific 16S rRNA primers and PCR setups

All primers (Table 14) were synthesized by Inqaba Biotech (Pretoria, South Africa) and purified on a reverse-phase cartridge.
These primers were used in the PCR setup (Table 15) for screening purposes.

Table 14 Primers used for screening of DNA.

Name	Sequence	T_m (°C)	Source	T_a (°C)
Ocin1F	ATKGGYGGHGCRAACTGYGARR	55.3/ 69.2	This study	58 to 49
Ocin1R	GTTSAGSCCRCARAARGTRCA	53.5/ 62.8	This study	58 to 49
Ento271F	GGGAAASGTTCGCBGGTCTG	58.4/ 61.2	(Wilson *et al.*, 2014)	60
Ento735F	GYATTAAGCCKYGGAAACKGT	50.4/ 61.0	(Wilson *et al.*, 2014)	60
Ento1290R	GCCCRGCWYVACCCGGTA	57.4/ 66.2	(Wilson *et al.*, 2014)	60

The letters in red are defined by the IUPAC ambiguity codes for nucleotides (Cornish-Bowden, 1985).
T_m: the melting temperature of the primer (min/ max) as calculated with OligoAnalyzer (Owczarzy *et al.*, 2008).
T_a: the annealing temperature used during PCR amplification.

2.4.2.2. '*Entotheonella*'-specific 16S rRNA PCR screen

The following primer pairs were used (Table 14) to screen for '*Entotheonella*' spp. Initially the 16S rRNA gene was partially amplified with the primers Ento271F and Ento1290R. A nested PCR was performed with products from the first PCR using the primers Ento735F and Ento1290R. DNA from an '*Entotheonella*' species was not available as a positive control for this screen and the gDNA of *D. maris* was used as a negative control along with a reaction without template DNA.

Table 15 PCR setup for the screening of rSAM_ocin_1 related proteins (primers: Ocin1F and Ocin1R) and *'Entotheonella'* species (primers: Ento271F, Ento735F, Ento1290R).

Primers	DNA polymerase	PCR reaction mixtures (20 µl)	PCR programme description	PCR programme			
				Reaction	Temperature (°C)	Duration	Cycles
Ocin1F Ocin1R	DreamTaq (Fermentas, Thermo Scientific)	1X DreamTaq buffer[1], 200 µM dNTP, 0.025 U.µl^{-1} polymerase, 1 µM of each primer, 50 to 100 ng template DNA. $MgCl_2$ was increased to 3 mM when specified.	Gradient: Annealing temperatures ranged from 55 to 65 °C across the 12 rows of the sample block.	Denature	95	5 m	1
				Denature	95	30 s	
				Anneal	55 to 65	30 s	35
				Extend	72	30 s	
				Extend	72	10 m	1
			Touchdown: The annealing temperatures were changed by -1 °C per cycle for 10 cycles. The annealing temperature of the next 25 cycles was kept constant at the lowest temperature reached after the first 10 cycles (49 or 51 or 53 °C).	Denature	95	5 m	1
				Denature	95	30 s	
				Anneal	58 or 60 or 62	30 s	10
				Extend	72	30 s	
				Denature	95	30 s	
				Anneal	49 or 51 or 53	30 s	25
				Extend	72	30 s	
				Extend	72	10 m	1
	Phusion High-Fidelity polymerase (Fermentas, Thermo Scientific)	1XPhusion HF buffer[2], 200 µM dNTP, 0.02 U.µl^{-1} polymerase, 1 µM of each primer, 50 to 100 ng template DNA.	Touchdown: See the description above.	Denature	98	1 m	1
				Denature	98	10 s	
				Anneal	58 or 60 or 62	15 s	10
				Extend	72	15 s	
				Denature	98	10 s	
				Anneal	49 or 51 or 53	15 s	25
				Extend	72	15 s	
				Extend	72	10 m	1
Ento271F Ento735F Ento1290R	Phusion High-Fidelity polymerase	See the reaction mixture above. DMSO: 3 % (v/v).		Denature	95	5 m	1
				Denature	95	15 s	
				Anneal	60	30 s	40
				Extend	72	15 s	
				Extend	72	1 m	1

[1] $MgCl_2$: 2 mM
[2] $MgCl_2$: 1.5 mM
m: minutes
s: seconds

2.4.2.3. Cloning and sequencing

Cloning and sequencing of selected PCR products was performed according to methods in Section 2.4.4.; Chapter 2. The nucleotide collection (nr/nt) and the non-redundant protein sequences (nr) databases were selected respectively for the BLASTn and BLASTx analysis of the sequences obtained from the Ocin1 screens.

3. Results and discussion

3.1. Analysis of the polytheonamide biosynthesis pathway and identification of putative pathways in sequenced genomes

Currently polytheonamides are the only known members of the RiPP family proteusins. They were first extracted from the sponge *T. swinhoei* in 1994 and became of interest when their anti-tumour properties were elucidated (Hamada *et al.*, 1994; Hamada *et al.*, 2005). It was thought that the peptides belonged to the class of non-ribsomal peptides and that the producing organism was a symbiotic bacterium of the sponge since other non-ribsomal peptides were shown to be produced by the symbionts of lithistid sponges (Hamada *et al.*, 2005). The biosynthetic cluster of polytheonamide was only discovered in 2012 within the metagenome of the *T. swinhoei* (Freeman *et al.*, 2012). Subsequently two closely related strains were isolated, '*E. factor* TSY1' and '*E. factor* TSY2', from the *T. swinhoei* (chemotype Y) sponge using fluorescence-assisted cell sorting and whole-genome amplification (Wilson *et al.*, 2014). The genomes were sequenced and revealed that the biosynthetic cluster of polytheonamide was located on the plasmid of TSY1 along with three additional putative proteusin gene clusters and another putative cluster from the genome of TSY2 (Figure 25; Wilson *et al.*, 2014).

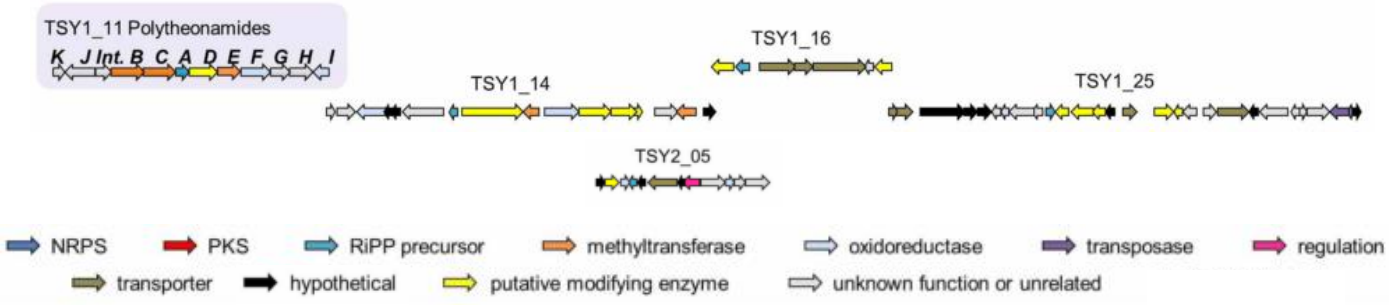


Figure 25 Polytheonamides biosynthetic cluster and putative proteusin pathways of '*Entotheonella factor* TSY1 and TSY2'. This figure is adapted from the Extended Data Figure 4 of Wilson *et al.* (2014).

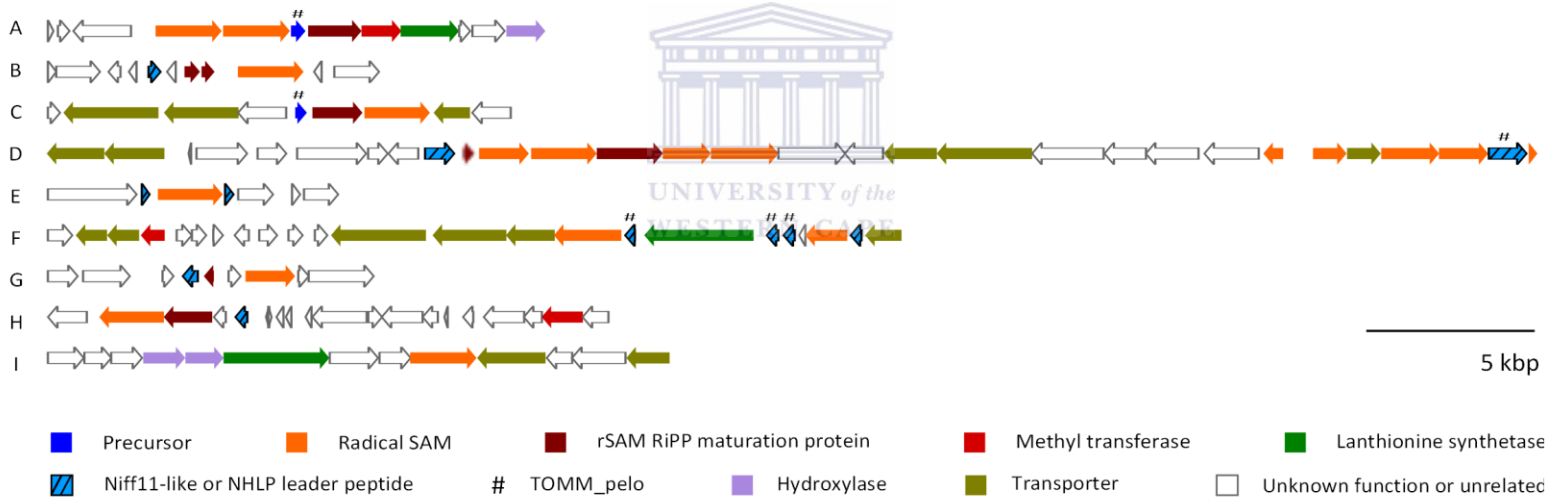


Figure 26 A map of the polytheonamide biosynthetic gene cluster (A) along with the hypothetical proteusin pathways on the genomes of other organisms (B-I). A) '*E. factor* TSY1' (KI932682.1), B) *Bradyrhizobium japonicum* USDA 6 (NC_017249.1), C) *Anabaena variabilis* ATCC 29413 (CP000117.1), D) *Desulfarculus baarsii* DSM 2075 (NC_014365.1), E) *Microcoleus* sp. PCC 7113 (NC_019738.1), F) *Nostoc* sp. PCC 7120 (NC_003272.1), G) *Rhizobium leguminosarum* bv. trifolii WSM2304 (NC_011368.1), H) *S. meliloti* SM11 (CP001830.1), I) *Stigmatella aurantiaca* DW43-1 (NC_014623.1). The open reading frames (arrows) are coded according to the predicted function of the product.

At the start of this study (2013), the polytheonamide biosynthetic cluster (Figure 26, A) was the only proteusin pathway available to design primers for the screening of metagenomic material for novel proteusin encoding pathways. However, it was reasoned that the polytheonamide genes might have homologues in other bacterial genomes, which if identified, could be used to design degenerate screening primers. Consequently, the Poy protein sequences were used to query for homologues in GenBank. Homologous proteins to at least four of the proteins within the biosynthetic cluster of polytheonamide were identified in ten organisms including *A. variabilis* ATCC 29413, *D. baarsii* DSM 2075, *Nostoc* sp. PCC 7120, *S. aurantiaca* DW4/3-1 and *Microcoleus* sp. PCC 7113 (Table 16; a more exhaustive summary of organisms with homologues is included in Appendix C; Table 8). During the analysis of the homologous proteins, it was observed that rSAM_ocin_1 related genes (accession: TIGR03975), that include *poy*B and *p*oyC, are positioned in close proximity to genes that encode TOMM_pelo superfamily proteins (Figure 26). The precursor of polytheonamide, PoyA, that is related to nitrile hydratase-related leader peptides (NHLPs), is described by the conserved domain model TOMM_pelo (Haft *et al.*, 2010). This could indicate that the organisms with homologues to the polytheonamide biosynthetic proteins harbour putative proteusin biosynthetic operons (Figure 26).

Table 16 A summary of the BLink analysis of proteins involved with polytheonamide biosynthesis (PoyB to I).

	PoyB		PoyC		PoyA		PoyD		PoyE		PoyF		PoyI	
	QC	ID	QC	ID	QC	ID	QC	ID	QC	ID	QC	ID	QC	ID
A. variabilis ATCC 29413	96	41	95	43	65	34	93	34			90	28		
D. baarsii DSM 2075	97	35	99	36	59	30	95	30						
Microcoleus sp. PCC 7113	96	43	96	43							89	29	74	31
Nostoc sp. PCC 7120	98	38	95	41	64	38			45	33	92	29		
S. meliloti SM11	99	40	95	38			94	31	64	33	89	28		
S. aurantiaca DW4/3-1	95	42	95	41							97	31	81	30
R. legumin bv, trifolii WSM2304	95	40	95	37			91	28			92	29		
B. japonicum USDA 6	97	39	99	37			30	29						
PoyC	99	47												

QC: query cover (%)
ID: identity (%)
Purple: Proteins not within the flanking regions of the other homologous proteins (Figure 26).
Orange: Proteins not identified with the BLink analysis, but observed in nearby regions (Figure 26).

3.2. Degenerate primer design

At the time this research was conducted, and to our knowledge, there were no degenerate primers available to screen for the biosynthetic genes of polytheonamides or related proteins.

The design of degenerate primers to the precursor of polytheonamide, PoyA, was considered. PoyA was aligned with the TOMM_pelo proteins. The alignment (with a threshold setting of 80 %, Figure 28) did not identify appropriate conserved regions within the leader to enable primer design.

Although PoyB and PoyC only have 47 % amino acid identity to each other (Table 16), both are C-methyltransferases (Freeman *et al.*, 2012), and the homologous bottromycin proteins (BmbB, BmbJ, BmbF and BstC, BstF, BstJ) perform a similar function. These proteins catalyse C-*β*-methylation (Figure 27) at un-activated carbons of polytheonamides, bottromycins and other RiPPs (Freeman *et al.*, 2012; Dunbar & Mitchell, 2013). These proteins belong to the rSAM_ocin_1 family (Figure 29) which in turn align with the Radical_SAM superfamily.

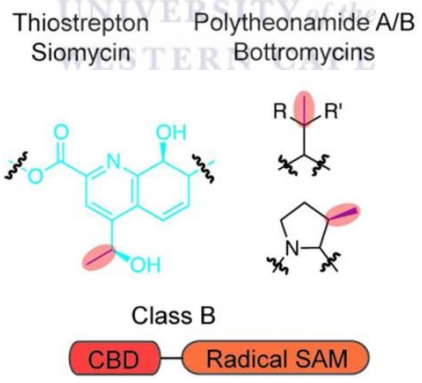


Figure 27 Class B radical SAM methyltransferases catalyse the C-β-methylation (highlighted in red) at un-activated carbons of RiPPs (Dunbar & Mitchell, 2013). These rSAM proteins include PoyB and PoyC with a cobalamin-binding domain (CBD).

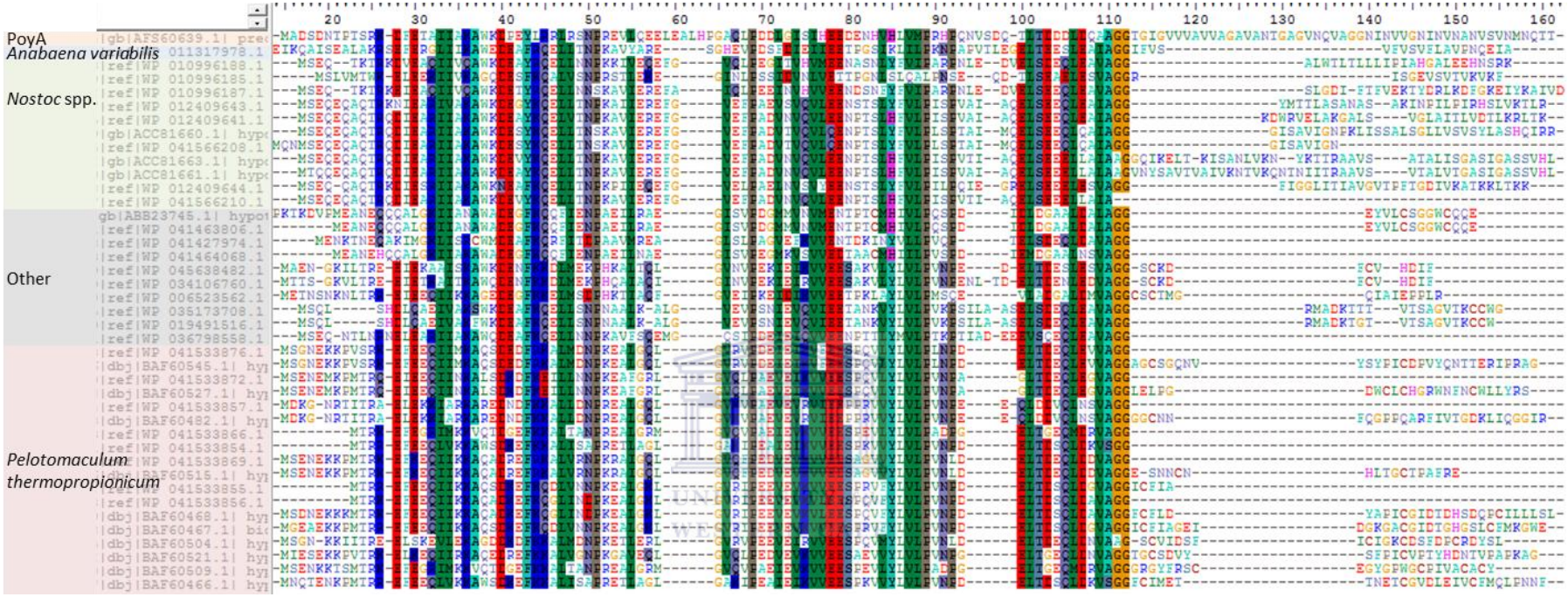


Figure 28 An alignment of PoyA with the representative proteins of the protein family, TOMM_pelo (accession: TIGR03793), that include proteins from *A. variabilis*, *Nostoc* spp., *Pelotomaculum thermopropionicum* and other species. The alignment was performed with the BioEdit Sequence Alignment Editor (Hall, 1999) and the Clustal W algorithm (Thompson *et al.*, 1994) and the similarity threshold was set to 80 % (highlighted residues).

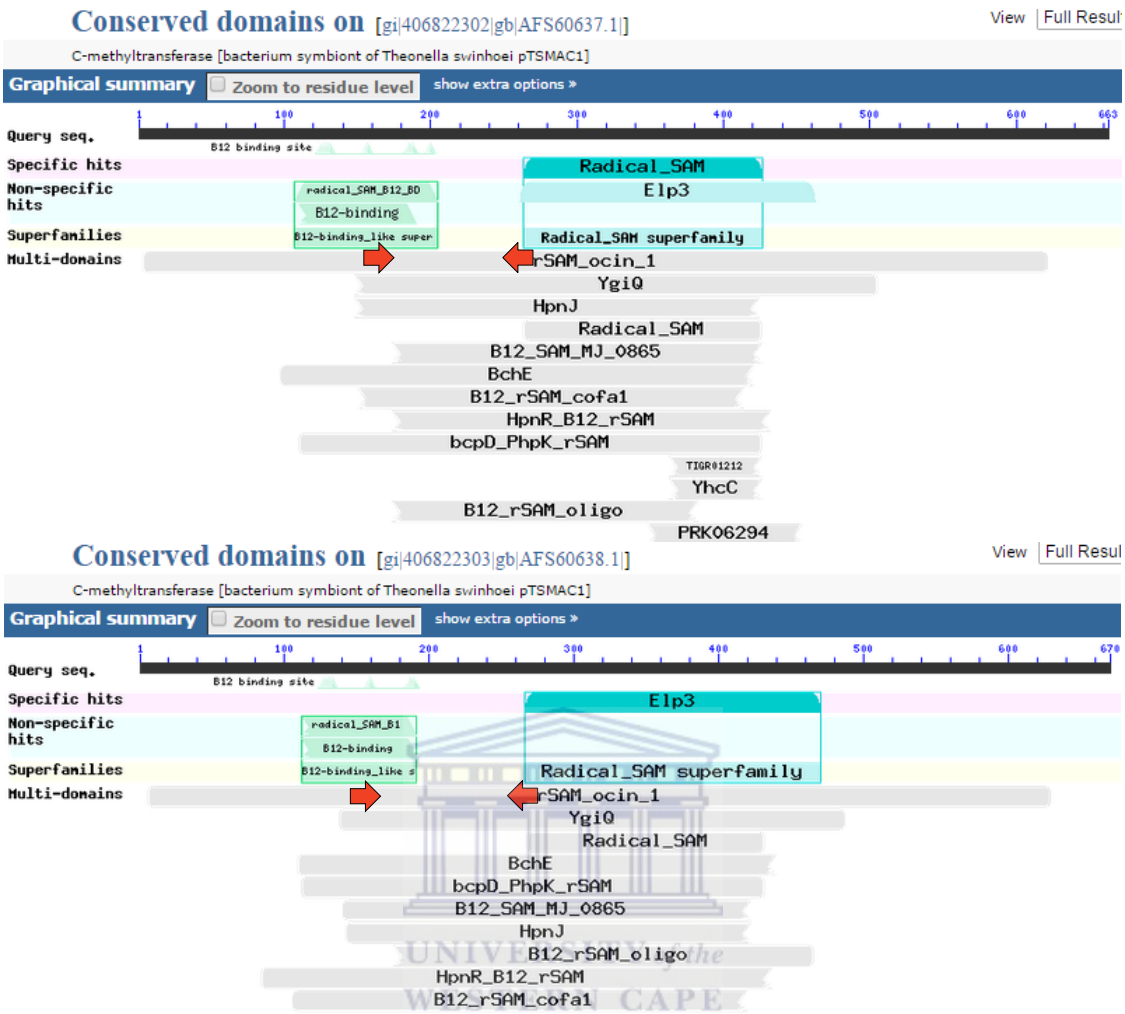


Figure 29 Conserved domain hits (from NCBI) of PoyB (top) and PoyC (bottom). The arrows in red indicate the regions to which the Ocin1 primers were designed. The jagged edges for some of the conserved domains indicate only a partial hit to the query.

The protein sequences of representative proteins of the rSAM_ocin_1 model were obtained (CDD; NCBI) and aligned (accession numbers are available in Appendix C; Table 9). A maximum parsimonious tree from the alignment showed seven proteins in the same clade as PoyB and PoyC (Figure 30, Clade B). These included proteins from *A. variabillis*, *Microcoleus* sp. PCC 7113, *Nostoc* sp. PCC 7120, *S. aurantiaca*, *Chondromyces apiculatis* and *Myxococcus* sp.

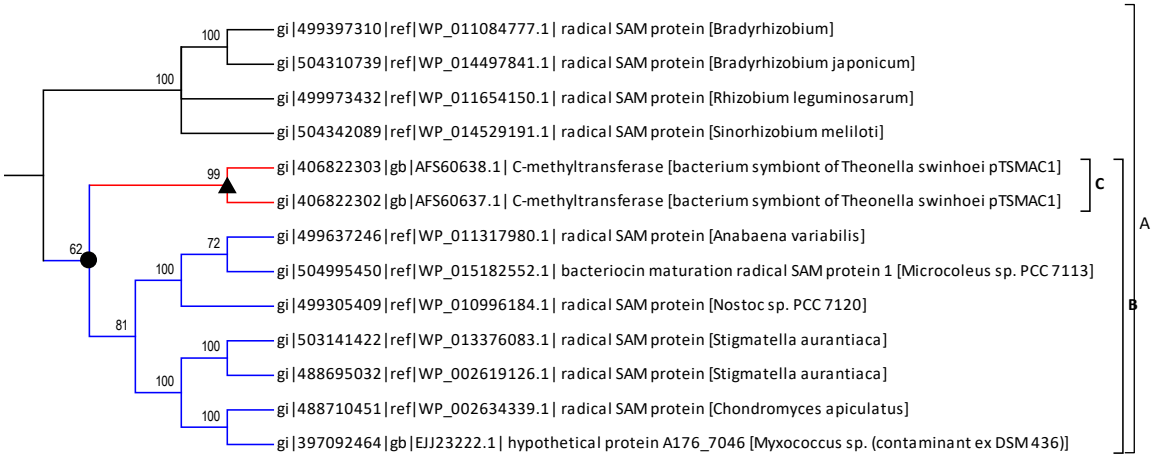


Figure 30 Clade A of the maximum parsimonious (MP) tree of the representative proteins of the rSAM_ocin_1 family (full tree is available in Appendix C; Figure 1). The presented clade is that of PoyB and PoyC (Clade C) and the seven closest related proteins (Clade B). Analyses were conducted in MEGA5 (Tamura *et al.*, 2011). The reference sequences (RefSeq; NCBI), denoted with the prefix 'WP_' in their accession numbers, were used as the representatives of identical sequences.

The proteins within Clade B and the homologous rSAM proteins associated with bottromycins were aligned to determine the conserved residues (Figure 31). Primers (Ocin1F and R) were designed to the conserved residues within the conserved domains. The designed primers (Table 17) were analysed (Section 2.2.2) and no hairpins or significant homo- or hetero-dimers formed. Taking the melting temperatures of the primers into account, the mean annealing temperature was 56.9 °C (OcinF1) and 53.2 °C (OcinR1).

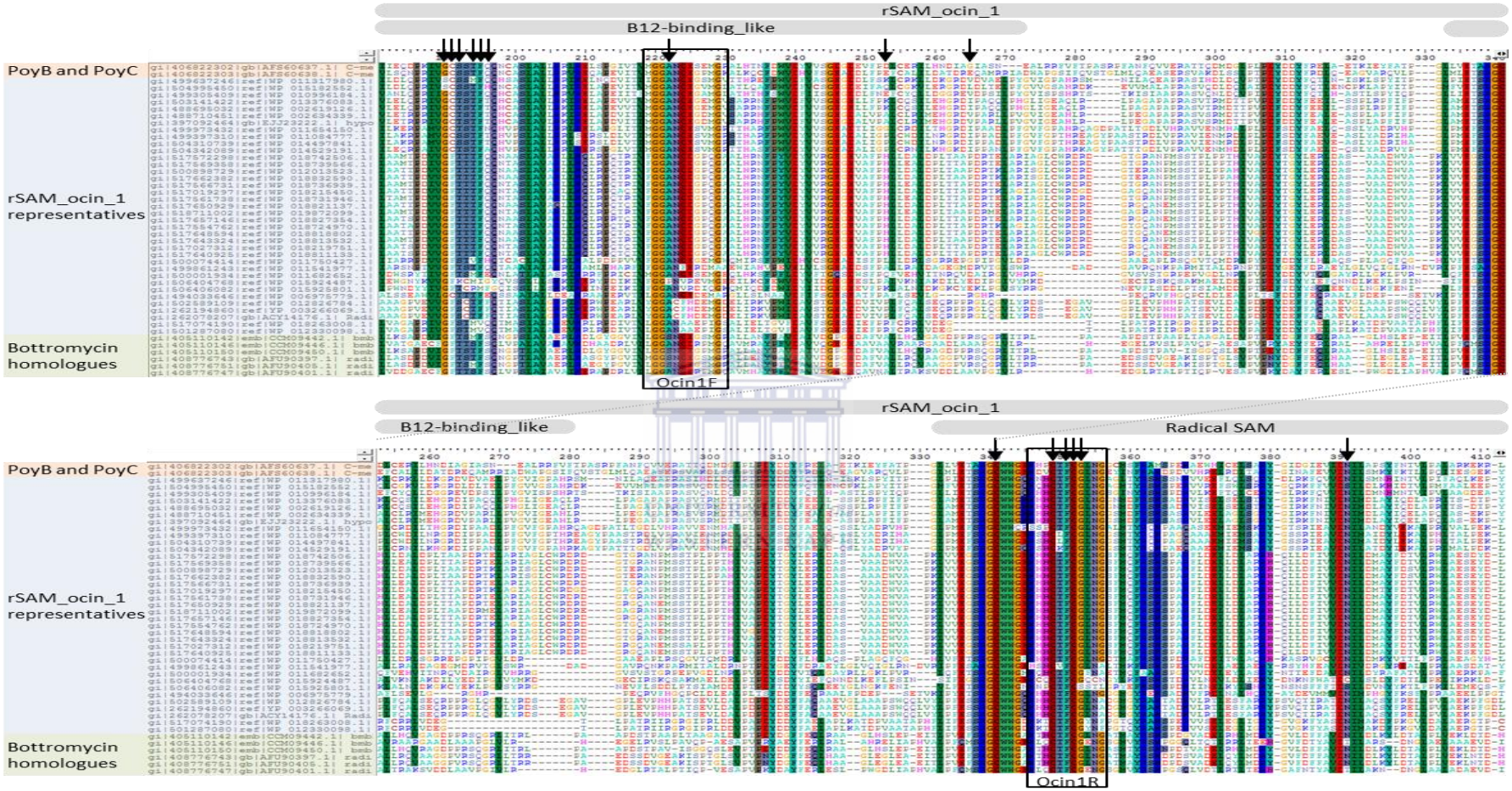


Figure 31 An alignment of the rSAM_ocin_1 proteins (accession: TIGR03975) that are closely related to PoyB and PoyC, together with the bottromycins' rSAM_ocin_1 proteins. The alignment was performed with the BioEdit Sequence Alignment Editor (Hall, 1999) and the Clustal W algorithm (Thompson *et al.*, 1994) and the similarity threshold was set to 80 % (highlighted residues). The conserved region in the black blocks corresponds to the region selected for the design of the primers (Ocin1F and R). The conserved domains for PoyB (grey bars) together with the B12-binding sites or FeS/SAM binding sites (black arrows) are indicated.

Table 17 Degenerate primers designed from conserved residues of rSAM_ocin_1 proteins.

Name	Sequence (5' - 3')	Size (bp)	Mean T_m (°C) [a]	Mean CG %	Degeneracy [b]
OcinF1	ATKGGYGGHGCRAACTGYGARR	22	61.9	56.1	192
OcinR1	GTTSAGSCCRCARAARGTRCA	21	58.2	52.4	64

[a] Calculated with OligoAnalyzer 3.1 (Owczarzy *et al.*, 2008)
[b] Degeneracy was calculated as the product of the number of possible nucleotides at each ambiguous position (red).

The rSAM_ocin_1 protein sequences were also aligned to other rSAM protein families, since the conserved domains, cobalamin (vitamin B12)-binding and rSAM, are present in other proteins. For example, when B12_rSAM_oligo proteins (Figure 32) were aligned with rSAM_ocin_1 protein less conserved residues were observed than between rSAM_ocin_1 proteins.

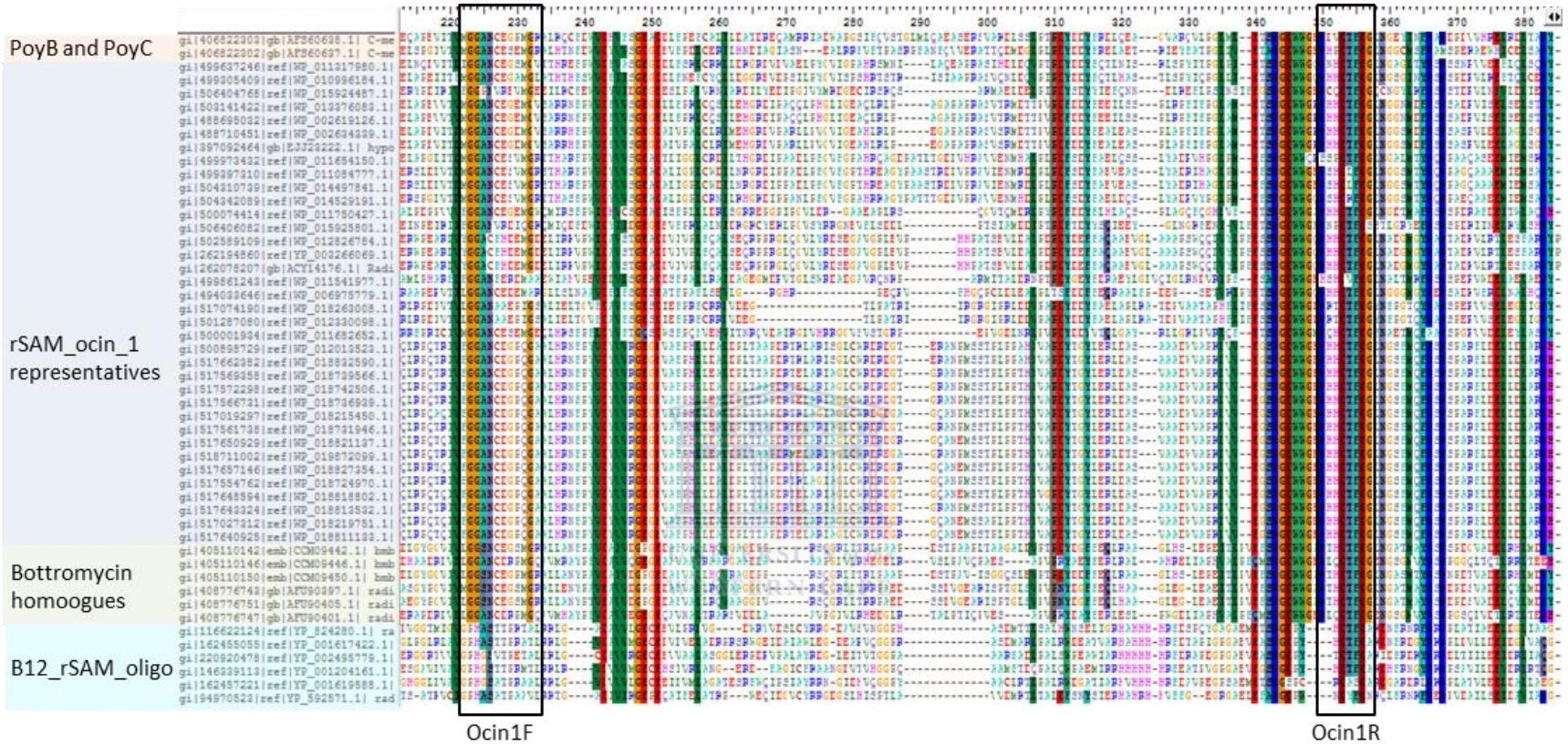


Figure 32 An alignment of the rSAM_ocin_1 proteins (accession: TIGR03975) that are closely related to PoyB and PoyC, bottromycins' rSAM_ocin_1 proteins and representative proteins of the B12_rSAM_oligo family. Alignment was performed with the BioEdit Sequence Alignment Editor (Hall, 1999) and the Clustal W algorithm (Thompson *et al.*, 1994) and the similarity threshold was set to 80 % (highlighted residues). The conserved region in the black blocks corresponds to the region selected for the design of the primers (Ocin1F and R), as in Figure 8.

3.3. *In silico* PCR screening with Ocin1 primers

An online tool was used for the simulation of a PCR to analyse the binding of the Ocin1 primers to *poyB* and *poyC*, as well as to the genomes available with putative PoyB and PoyC homologues. The species included were *S. meliloti*, *A. variabilis*, *Nostoc* sp. PCC 7120 and *S. aurantiaca*. *In silico* tools, such as this online PCR tool, is used to give some confidence of the designed primers.

The *in silico* amplification of the Ocin1 primers, using *poyB* as the template, resulted in the amplification of a 390 bp fragment of the gene (Table 18). Similarly, a 398 bp fragment was amplified from *poyC*. A 387 bp fragment of a homologous gene was amplified from the genome of *Nostoc* sp. PCC 7120. A BLAST analysis of the latter amplicon confirmed that the gene is a hypothetical protein and part of the rSAM_ocin_1 family of proteins. Similarly, a radical SAM domain-containing protein was amplified (327 bp) from *S. aurantiaca* DW4/3-1. The sequences of these fragments are reported in Appendix C; Figure 2.

Table 18 *In silico* PCR simulations with the Ocin1 primers.

Template sequence	**Strain**	**Accession number** [1]	**bp** [2]	**Parameters** [3]
poyB (2026 bp)		JX456532.1:8815-10806	390	(1, 2) (1, 3) (1, 5)
poyC (2047 bp)		JX456532.1:10848-12860	398	(2, 10)
S. meliloti	1021 [4]	PRJNA57603	NB	
S. meliloti	SM11	PRJNA159685	NB	
A. variabilis	ATCC 29413	PRJNA58043	NB	
Nostoc sp.	PCC 7120	PRJNA57803	387	(2, 10)
S. aurantiaca	DW4/3-1	PRJNA158509	327	(2, 10)

[1] Nucleotide sequences were obtained from NCBI.
[2] The number of base pairs amplified.
[3] The parameters given in brackets are those used when amplification was observed. The number of mismatches that were allowed and number of nucleotides in the 3' end where mismatches were not allowed.
[4] The strain used for *in vitro* PCR screening.
NB: The simulated amplification resulted in no bands.
Grey: strains identified with BLAST searches to have putative PoyB and/or PoyC homologues.

Although the BLink analysis showed that *A. variabilis* have putative proteins homologous to PoyB and PoyC, the *in silico* PCR did not produce any amplicons with the Ocin1 primers. Likewise, the *Sinorhizobium* strains did not produce amplicons (Table 18) even though strain SM11 was identified to have a homologous protein. Even

though the *in silico* PCR and BLAST algorithms did not produce the same set of results, *S. meliloti* 1021 was included in the *in vitro* experiments since the parameters of an *in vitro* PCR can be manipulated to change the stringency of the amplifications to obtain the desired amplicon.

3.4. *In vitro* PCR screening with Ocin1 primers

3.4.1. Verification of the amplification of the positive control genes

A requirement for screening with the Ocin1 primers was to confirm that the correct genes were amplified from the control DNA.

During the study of the biosynthetic gene cluster of polytheonamide by the Piel laboratory (Freeman *et al.*, 2012) the Nhis-*poyB*-pCDFDuet-1 plasmid, containing *poyB*, was constructed for expression experiments. The total size of Nhis-*poyB*-pCDFDuet-1 (now referred to as the *poyB*-plasmid) is 5550 bp. The *poyB* gene (1990 bp) in this plasmid was used as a positive control for the Ocin1 PCR screen. It was estimated that an amplicon of 360-390 bp would be obtained (Table 18). In order to compensate for variance in gene sizes, it was decided to focus on amplicons (obtained with the Ocin1 screens) within the size range of 350 to 400 bp.

S. meliloti 1021 was selected as another control because it is closely related to other rhizobacteria (*S. meliloti* SM11, *Bradyrhizobium* spp. and *R. leguminosarum*) with homologous genes to *poyB* and *poyC* (Figure 30, Clade A).

Amplicons generated from the gDNA of *S. meliloti* and the *poyB*-plasmid, that were within the expected size range of 350 to 400 bp, were sequenced (Appendix C; Figure 3). Analysis showed that the fragment amplified from *S. meliloti* was 99 % similar to the radical SAM domain protein of *S. meliloti* SM11 (total gene length: 1827 bp). This protein is part of a cluster that is conserved in rhizobacteria (PCLA 629334; protein cluster database, NCBI). The fragment amplified from the *poyB*-plasmid was 99 % similar to *poyB*.

Analysis of the polytheonamide biosynthetic cluster showed that *S. meliloti* SM11 has homologous proteins to (PoyB, C, D and F; Table 16, Figure 26 H). However, the *in silico* PCR with the Ocin1 primers detected no homologous genes within any of the

screened *S. meliloti* strains (1021 and SM11). *In silico* analysis is an indication of what could be found with *in vitro* experiments. The result of the *in silico* PCR is not as definitive as *in vitro* experiments. This was clearly demonstrated when a homologous gene (initially identified with BLAST searches) was amplified. This could be because the stringency of the *in silico* PCR is not as flexible as with *in vitro* experiments and thus the *in silico* PCR is not always an accurate representation.

3.4.2. Optimising the Ocin1 PCR parameters

Various parameters of the PCR reaction were altered to reduce the level of non-specific amplification obtained, as well as to establish the limits of the Ocin1 primers since the primers are degenerate and since metagenomic DNA (a complex mixture of DNA) was the target for screening. Different annealing temperatures, DNA polymerases and concentrations of $MgCl_2$ were used.

The melting temperature range for the Ocin1 primers was calculated to be between 53.5 °C (lowest for the reverse) and 69.2 °C (highest for the forward). A gradient PCR, with annealing temperatures ranging from 55 to 65 °C, was performed with the *poyB*-plasmid DNA. A DNA fragment, within the expected size range, was observed (Figure 33) across the entire gradient. Annealing temperatures above 61 °C reduced the binding capacity of some of the primers and resulted in a lower yield of the PCR product. Taking into account the amplification of the *poyB* gene at different annealing temperatures, together with the calculated melting temperatures, the optimal annealing temperature for the Ocin1 primers to the *poyB* gene was 58 °C and lower.

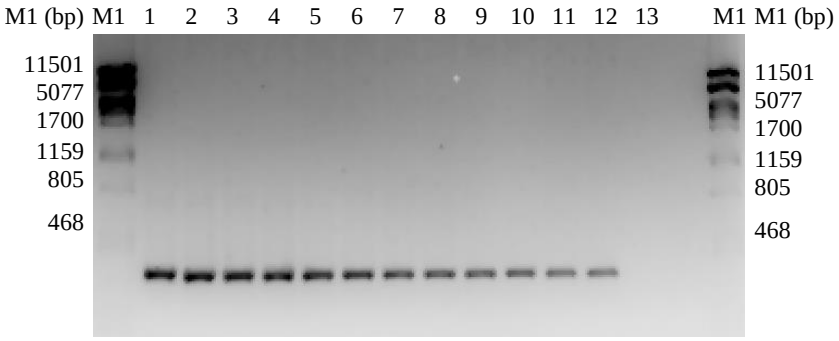


Figure 33 Gradient PCR of Ocin1 primers with *poyB*-plasmid DNA as the template. The annealing temperature for each lane was as follows: 1) 55.0 °C, 2) 55.6 °C, 3) 56.1 °C, 4) 57.4 °C, 5) 58.9 °C, 6) 60.3 °C, 7) 61.0 °C, 8) 62.3 °C, 9) 63.4 °C, 10) 64.4 °C, 11) 64.7 °C, 12) 65.0 °C. The negative control (lane 13), where no template DNA was added, had an annealing temperature of 60.3 °C. M1) λ phage DNA digested with the endonuclease PstI was used as the DNA ladder.

When screening with degenerate primers, where it is uncertain which conditions are the best suited for the binding of the primers to the template, it is recommended to use stringent annealing temperatures (Korbie & Mattick, 2008) for the first few cycles of the PCR. This reduces miss-priming and non-specific products (Don *et al.*, 1991). This is why a touchdown PCR was used for the Ocin1 screens to compensate for the variance in annealing temperatures of the degenerate primers. A touchdown PCR was performed with DreamTaq polymerase using three different annealing temperature ranges (Table 15). The temperature ranges were chosen around the optimal annealing temperature.

When the touchdown PCR at the lower annealing temperature range (58 to 49 °C) was used with the *S. meliloti* DNA there was non-specific amplification (Figure 34, lane 1), which was reduced as the annealing temperatures increased (60 to 51 °C and 62 to 53 °C). At higher annealing temperatures the control fragment from the more complex DNA (*S. meliloti*) was weakly amplified compared to the fragment from the *poyB*-plasmid. Therefore, it was decided not to use the higher temperature ranges for the screening of mDNA, as the target gene is not likely to be over-represented in the mDNA, as was the case with the *poyB*-plasmid, and as a consequence the amplification of the gene might be too weak to be detected.

The effect of different PCR reagents (DNA polymerase and $MgCl_2$) was also evaluated. There was less non-specific amplification from the *S. meliloti* DNA with the Phusion polymerase (Figure 34, lane 11) and although the desired fragment was better amplified

than with the DreamTaq polymerase, the opposite was true for the control fragment amplified from the *poyB*-plasmid. When an increased concentration of $MgCl_2$ with the DreamTaq reaction mixture (Figure 35) was used, the non-specificity of S. *meliloti* did not improve and the amplification of the 400 bp fragment of both controls was not negatively influenced. Therefore, it was decided to use the increased concentration of $MgCl_2$ for the Ocin1 screens, since the source of the DNA or the extraction methods used may result in increased chelators within the DNA sample which will sequester the $MgCl_2$ available in the PCR reaction mixture for the DNA polymerase (Schrader *et al.*, 2012).

The optimisation of the touchdown PCR with *S. meliloti* DNA gives an indication of what conditions could be used to screen other DNA for rSAM_ocin_1 homologous genes, but it should be remembered that the optimisation of the PCR is also template dependent. The conclusion of the optimisation procedure was to use DreamTaq polymerase with the annealing temperature range of 58 to 49 °C and an increased concentration of $MgCl_2$.

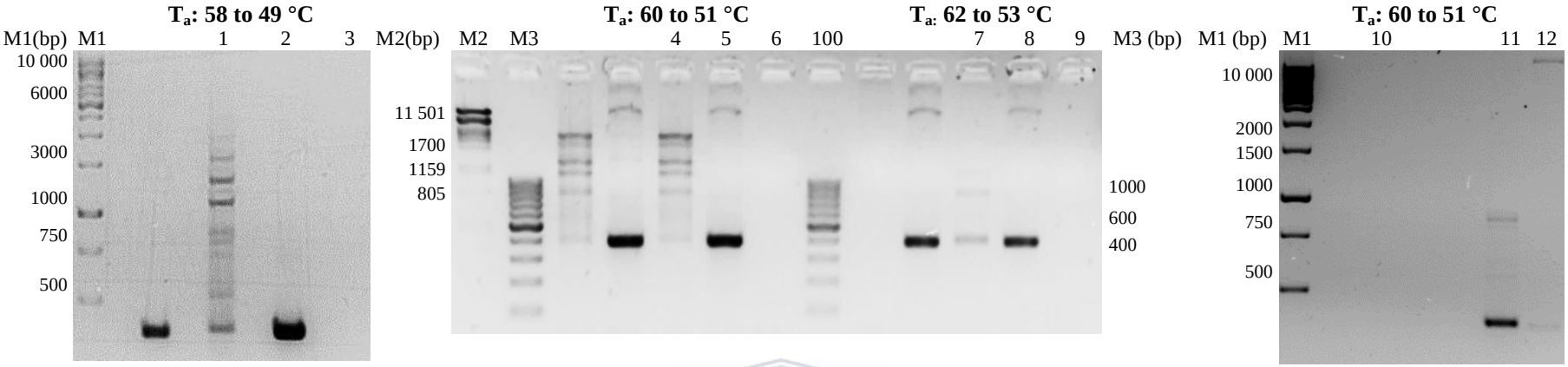


Figure 34 Touchdown PCRs with DreamTaq (Lanes 1-9) and Phusion (Lanes 10-12) polymerase and 2 mM $MgCl_2$ at annealing temperatures (T_a): 58 to 49 °C (Lanes 1-3), 60 to 51 °C (Lanes 4-6 and 10-12) and 62 to 53 °C (Lanes 7-9). The template DNA used for Lanes 1, 4, 7 and 11) *S. meliloti;* Lanes 2, 5, 8 and 12) *poyB*-plasmid and Lanes 3, 6, 9 and 10) no DNA. M1) 1 kb; M2) λ PstI and M3) 100 bp DNA markers.

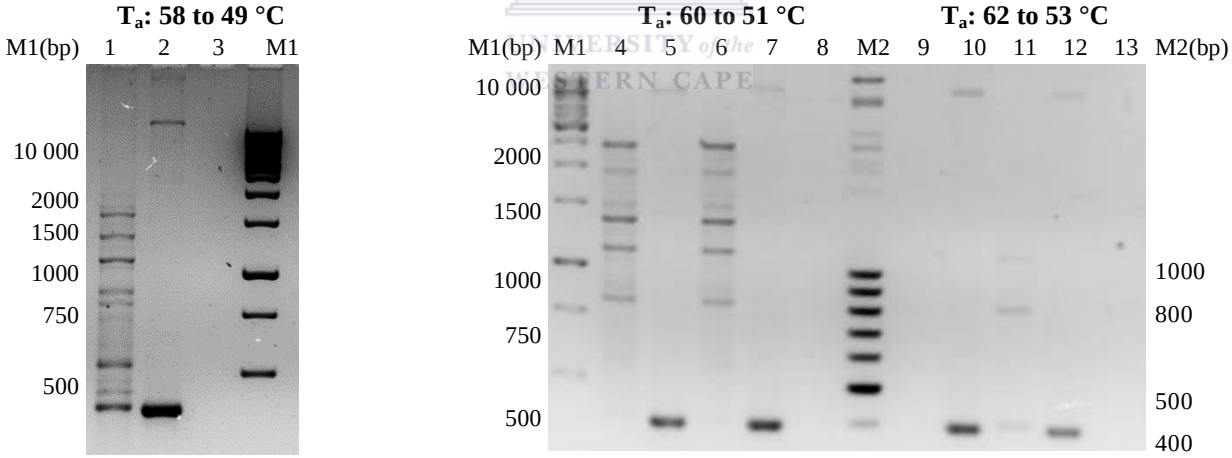


Figure 35 Touchdown PCRs with DreamTaq polymerase and 3 mM $MgCL_2$ at annealing temperature (T_a): 58 to 49 °C (Lanes 1-3), 60 to 51 °C (Lanes 4-8) and 62 to 53 °C (Lanes 9-13). Template DNA added: *S. meliloti* (Lanes 1, 4, 6, 10 and 12) and *poyB*-plasmid (Lanes 2, 5, 7, 9 and 11). The DNA ladders are as follows: M1) 1 kb; M2) λ PstI and 100 bp.

3.4.3. Extraction of mDNA from PE14

Initially the objective for this part of the research was to extract high molecular weight (HMW) metagenomic DNA to construct a library. The construction of a metagenomic library has many advantages for the screening of secondary products, such as RiPPs. The DNA library could be screened functionally or with sequence-based methods. Through functional screening, the genes are expressed heterologously. With the advances of next generation sequencing, sequence-based screening of DNA libraries has become more attractive. One of the attractive characteristics for bioprospecting secondary products through sequencing is that the biosynthetic genes are usually clustered together (Hoffmeister & Keller, 2007; Lee *et al.*, 2008). Thus, an entire pathway could be encoded in a single clone, depending on the size of the inserts. Another sequence-based method of screening is dependent on homology techniques such as PCR.

Although the conventional approach of metagenomics is to construct DNA libraries, various limitations experienced with library construction and difficulties with heterologous expression has driven homology-based screening of metagenomes (Charlop-Powers *et al.*, 2014) without the construction of DNA libraries (Więckowicz *et al.*, 2011; Woodhouse *et al.*, 2013; Ridley *et al.*, 2005).

In this study, seven different extraction methods were employed to extract HMW DNA (Table 19).

Table 19 Methods employed to extract HMW mDNA and the characteristics of the respective methods.

	Method	Characteristic of method
1	Modified CTAB lysis	CTAB is known to improve purification of DNA for further manipulation.
2	Agarose embedded cells	Embedding cells within agarose prevents shearing of DNA and help to obtain DNA that is more pure for downstream processes.
3	Modified urea lysis	Successfully used for HMW DNA libraries from sponges. The use of CTAB during extraction, which could influence the yield of DNA negatively, is optional.
4	MoBio Soil extraction kit	mDNA from soil is usually contaminated with humic acid and other compounds that inhibit molecular manipulation of DNA. This kit was used to see whether more DNA of better quality could be obtained.
5	Agarose plugs with CTAB or urea lysis	This was a combinatorial method of Newman et al (Newman *et al.*, 2010) with the lysis buffers of Method 1 and 3.
6	Crude extraction	
7	Yeast extraction method	

It is well known that it is a challenge to extract sufficient HMW mDNA from sponges that is pure enough for the construction of metagenomic libraries; for medium-insert (45 kb) or large-insert libraries (100 to 350 kb; Ouyang *et al.*, 2010; Salgado *et al.*, 2007; Gurgui & Piel, 2010; Charlop-Powers *et al.*, 2014). There were many difficulties experienced during the extraction of HMW mDNA, which included low concentration of DNA, shearing of DNA, co-extraction of inhibitors and loss of DNA after purification. Consequently, the objective changed rather to extract sufficient DNA that is of good enough quality to be screened without the construction of a library.

Some of the practical precautions applied to prevent shearing of DNA included trimming of pipette tips before sterilisation, to become wide bore tips, and gentle or minimal mixing of solutions during cell lysis or DNA precipitation.

Another practical method was to embed the cells within agarose plugs to prevent shearing (Liles *et al.*, 2008; Handelsman *et al.*, 2002; Table 20, Method 2). The plugs were subjected to PFGE to recover the DNA. Faint smears of DNA were observed after electrophoresis and the concentration of the DNA was too low for recovery after gel electrophoresis. Another problem encountered with the PFGE experiment was to maintain a low temperature during electrophoresis. The apparatus was used at room

temperature (and could not be moved to a cold room). This is acceptable and the buffer was circulated through tubes coiled through an ice water bath as recommended (BioRad), but low temperatures could not be maintained (below 30 °C; BioRad) and this could have influenced the resolution of DNA bands. Alternatively, the field strength could have been lowered to prevent overheating of the buffer. Another difficulty experienced with this method was the recovery of DNA from the plugs. The plugs were made with SMP agarose. This should not influence the recovery of the DNA after the plugs were subjected to PFGE in a LMP gel. However, if the plugs were made with LMP agarose the DNA could be recovered by enzymatic digestion of the agarose or by electroelution (Osoegawa *et al.*, 1998; Liles *et al.*, 2008).

Table 20 Metagenomic DNA extraction methods used and the result thereof. The amplification of 16S rRNA genes was evaluated. If there was no amplification from the mDNA it was further purified with various techniques.

	Extraction method	Section	Sponge (g)	DNA obtained	Initial concentration (ng/ul)	Purification	Concentration after purification (ng/ul)	16S rRNA	Conclusion
1	Modified CTAB lysis	2.3.2.1.	2	Yes	335.6	PCI	48.8	No	Too low concentration.
			50	Yes	754.1	RNase	7.6	No	Loss of DNA during purification.
2	Agarose embedded cells	2.3.2.2.	10	No	-	-	-	-	DNA not recovered from agarose with PFGE.
3	Modified urea lysis	2.3.2.3.	1	Yes	379.1	CTAB	5.2	No	Loss of DNA during purification.
			10	Yes	90.2	size selection PVPP and desalt	39.2 46.4	No	No amplification after purification.
			10	Yes	141.8	RNase; size selection; PVPP	48.8 (RNase); 127.5 (PVPP)	No	DNA not recovered after size selection.
			10	Yes	198.5	PVPP	180.2	Yes, after PVPP	Amplification when highly diluted, but concentration too low for screening.
4	MoBio Soil extraction kit	2.3.2.4.	3	Yes	-	-	-	-	Too low concentration.
5	Agarose plugs with CTAB lysis	2.3.2.5.	1.5 [1]	No	-	Formamide	-	-	DNA could not be recovered.
6	Agarose plugs and urea lysis	2.3.2.5.	1.5 [1]	No	-	Formamide	-	-	DNA could not be recovered.
7	Crude extraction	2.3.2.6.	1	Yes	113.0	PVPP	56.5	No	DNA too fragmented (< 500 bp).
8	Yeast extraction method	2.3.2.7.	1 ml [2]	Yes	68.6	PVPP	-	No	DNA not pure enough.

[1] See the methods section for a description of how the samples were prepared.
[2] The cells were separated from 3 g of sponge material and enriched for microbial cells. The cell-suspension (5 ml) was stored as 1 ml aliquotes.

Metagenomic DNA was obtained with various attempts of the modified urea lysis extraction method (Table 20, Method 3). This method seemed to yield better quality DNA than the modified CTAB method (Table 20, Method 1). Electrophoresis showed that most of the fragments obtained were about 23 to 48 kb and darker smears were observed below 2000 bp (Figure 36, lanes 1 to 4).

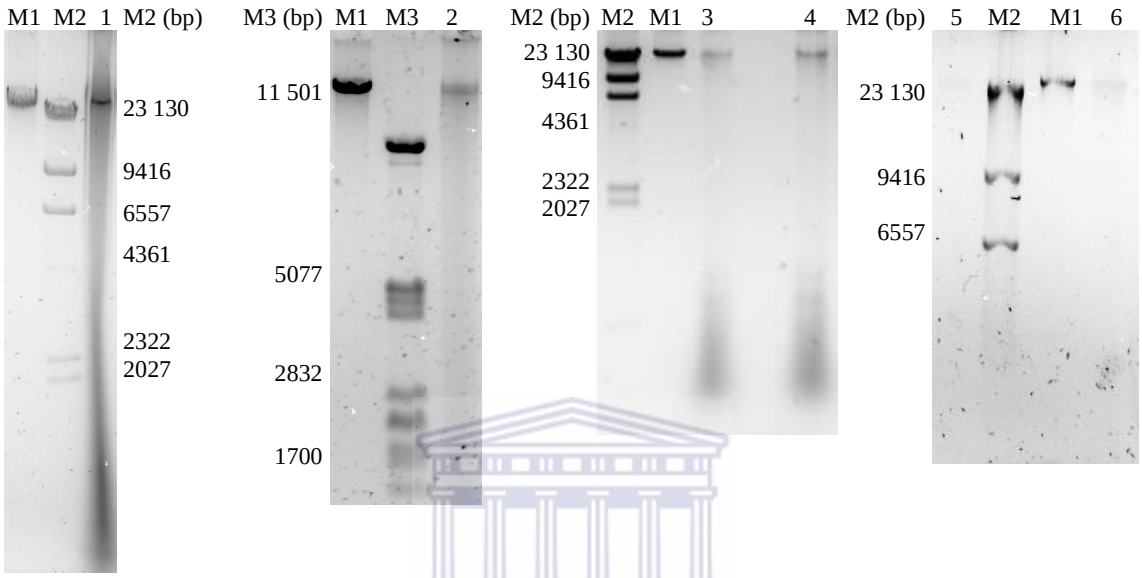


Figure 36 Different attempts of extracting mDNA with the modified urea lysis method. Lanes 1-4) the extracted mDNA. Lanes 5-6) the mDNA, from lanes 3 and 4, after purification with PVPP. The DNA markers are M1) λ DNA, about 48.5 kb; M2) λ Hind III and M3) λ PstI.

Sponges are known to contain phenolic compounds and high concentrations of polysaccharides (Gurgui & Piel, 2010; Salgado *et al.*, 2007) that are co-extracted with DNA. These are common PCR inhibitors that are also co-extracted from plant samples (Schrader *et al.*, 2012). Purification methods for DNA extracted from environmental material, with high concentrations of inhibitors for molecular methodologies, have been widely studied. The purification methods include the use of column chromatography, polyvinylpyrrolidone (PVP; Schrader *et al.*, 2012) and PVPP or precipitation of polysaccharides with a cationic detergent such as CTAB (Gurgui & Piel, 2010). Newman *et al.* (2010) found that the loss of DNA could be limited by embedding the DNA in agarose before purification. They incubated the agarose-plugs in a formamide and salt solution and effectively removed PCR inhibitors from samples, which were previously contaminated with PCR inhibitors.

Different purification methods (Section 2.3.2.8.) were used to remove contaminants and improve the quality of the mDNA. Less DNA was recovered after purification (Figure 36, lanes 5 and 6) and for some of the purification methods (CTAB and size selection) the concentration of the DNA was too low for further analysis.

The extraction of high quality mDNA (low or high molecular weight) mDNA from the sponge *H. esperioides* was proven difficult in our experience, even though numerous attempts were made. The yield of mDNA is dependent on the species, time and method of storage as well as the extraction method (Simister *et al.*, 2011). One of the factors that have been found to significantly influence the yield of DNA is the storage method of a sample. This have been analysed by various researchers (Simister *et al.*, 2011; Salgado *et al.*, 2007; Dunlap *et al.*, 2007). Generally, the preservation methods include freezing or the use of fixatives or buffered solutions, which are not always accessible in remote locations. For future studies we recommended to use fresh samples (Dunlap *et al.*, 2007) or samples that have only been stored for a short period of time.

It is not only the method of storage or method of extraction that influences the yield of DNA. As reported within previous chapters, the symbionts of sponges are important producers of natural products and thus metagenomic studies that screen for natural products focus on obtaining the DNA of the symbionts. Many of these organisms are endosymbionts and consequently it is difficult to obtain high quantities and quality DNA as the techniques employed for the separation of the cells from the host affects the yield and molecular weight of the DNA (Dunlap *et al.*, 2007).

3.4.4. DNA screens

3.4.4.1. Validating the amplification of mDNA

Before screening with the Ocin1 primers, the DNA was used in a validation PCR with universal 16S rRNA primers to determine whether the DNA is amplifiable with DreamTaq polymerase (Section 2.4.1.). The hypothesis was that if conserved genes, such as the 16S rRNA genes, are not amplified the concentration of the mDNA could be too low or PCR inhibitors were co-extracted with the mDNA.

To determine if inhibitors are present in the DNA sample, an internal control (Burkardt, 2000) was used by adding gDNA of *D. maris* to the mDNA (spiked-mDNA; Section

2.4.1.) The gDNA of *D. maris* has been successfully used in previous PCRs with the 16S rRNA primers. Therefore, amplification of a DNA fragment with the internal control should always be observed, but if amplification was not observed, there is a possibility of PCR inhibitors within the extracted mDNA sample. Inhibitors could be in solution or bound to the DNA. DNA could be purified from inhibitors that are in solution, but it is more difficult when the inhibitors are bound to the DNA. It should also be remembered that all of the DNA will not be fully recovered after purification (Burkardt, 2000) and the method of purification should be carefully selected.

In general, the extracted mDNA could not be manipulated in downstream steps. For all of the extraction attempts the 16S rRNA genes could not be amplified during the validation PCR. When the mDNA was spiked with gDNA of *D. maris* it could also not be amplified, suggesting the presence of inhibitors. For one of the validation PCRs (Figure 37) the mDNA and spiked-mDNA only amplified when it was highly diluted, but the DNA could not be used for screening purposes, as the DNA concentration would be too low for efficient PCR screening (Wintzingerode *et al.*, 1997).

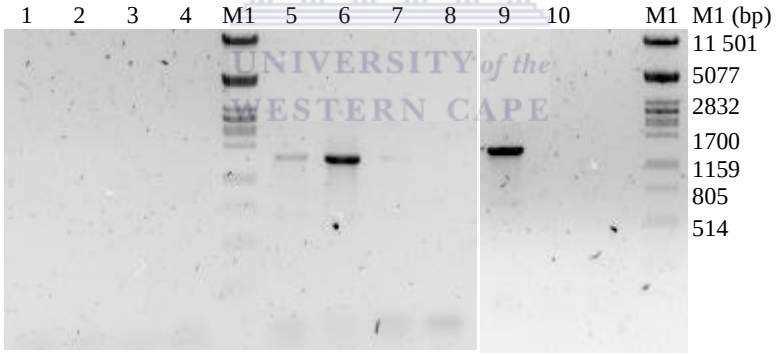


Figure 37 The validation PCR of the extracted mDNA with the universal 16S rRNA primers. Lanes 1, 3, 5 and 7) mDNA at concentrations: undiluted, 10 times diluted, 50 times diluted and 100 times diluted. Lanes 2, 4, 6 and 8) spiked-mDNA (mDNA with gDNA of *D. maris*). The spiked-mDNA was diluted the same as the mDNA. Lane 9) control reaction with gDNA of *D. maris* as template and lane 10) no template DNA. DNA marker λ PstI (M1).

Some of the extracted mDNA was purified by treatment with RNase, CTAB, PVPP, drop dialysis and formamide (Figure 38), but even after purification the DNA did not pass the validation PCR.

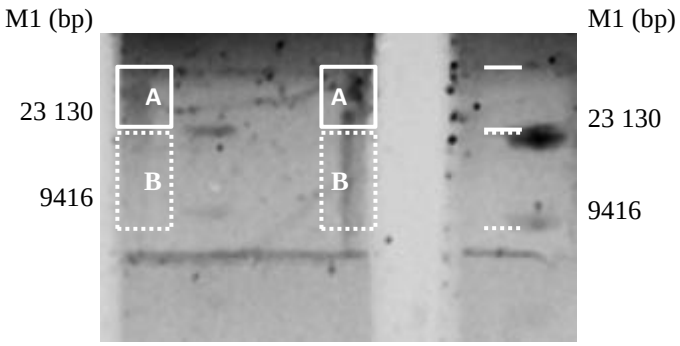


Figure 38 Size selection of the DNA purified with formamaide. The fragments greater than 23 kb in size was cleaned (A) as well as fragments between approximately 9 to 23 kb (B).

Metagenomic DNA, extracted by members of IMBM from the sponges PE03, PE05, PE08 and PE14 (Table 10) was also validated before screening it with the Ocin1 primers. The mDNA samples that passed the validation PCR with the DreamTaq DNA polymerase were PE03, PE05-B and PE14-B (Table 21). The validation of the latter two samples was confirmed by personal communication with the colleagues who extracted the DNA.

The mDNA screened for the presence of '*Entotheonella'* spp. was first validated with the universal 16S rRNA primers (E9F and U1510R) and the same PCR reaction mixture that was used for screening with the '*Entotheonella'*-specific primers (i.e. Phusion polymerase with 3 % DMSO; Table 15). Failure of amplification with the Phusion validation PCR led to the purification of the samples PE05-A, PE05-B, PE08, PE14-A and PE14-B with the PVPP method (Section 2.3.2.8.3.). Only sample PE05-A was amplifiable in the Phusion validation PCR after purification (1500 bp; Figure 39, Lane 1).

Table 21 Validation of the extracted mDNA with different DNA polymerases.

Species	Sample name (mDNA)	Validation	
		DreamTaq	Phusion
Guancha sp. 001RSASPN	PE03	Y	ND
W. caliculatum	PE05-A	N	Y
W. caliculatum	PE05-B	Y *	N
Spongia sp.001RSASPN	PE08	N	N
H. esperioides	PE14-A	N	N
H. esperioides	PE14-B	Y *	N

* Amplification was determined by the person that extracted the mDNA.
Y: yes, amplified.
N: no, did not amplify.
ND: not determined.

The discrepancy between the validation results when different polymerases were used, illustrates that the mDNA, validated with DreamTaq DNA polymerase, was still contaminated with inhibitors that inhibited the activity of the Phusion DNA polymerase (Table 21). Even though the manufacturer of the Phusion DNA polymerase states that it will tolerate numerous PCR inhibitors, our validation experiments showed that the Phusion DNA polymerase was sensitive to the unknown inhibitors still present in the mDNA samples. Although some PCR inhibitors have been identified for various samples, and methods to overcome the effects are available, many inhibitors and their mechanism of action remain unknown (Abu Al-Soud & Râdström, 1998; Wilson, 1997). It should also be noted that an increase in the fidelity of a polymerase usually affects some of the other characteristics such as resistance to inhibitors and efficiency. This is why manufacturers optimise DNA polymerases and their reaction mixtures to achieve the best amplification, but this is not a guarantee as each reaction is unique.

3.4.4.2. Screening for the presence of *'Entotheonella'* spp.

Concurrent to this study, the Piel laboratory determined that the organisms producing the proteusin, polytheonamide, belong to the candidate genus *'Entotheonella'* (Wilson *et al.*, 2014). In their study, they screened 37 sponge species (47 sponges in total), from various taxonomical groups and locations, with the *'Entotheonella'*-specific 16S rRNA gene primers they developed. These filamentous species were present in 40 of the sponges they have screened (Table 22).

Table 22 The number of sponges screened, from various locations, with the *'Entotheonella'*-specific 16S rRNA gene primers.[1]

Origin of sponges	Sponges screened	Sponges with *'Entotheonella'* spp.
Japan	29	24
Caribbean Sea	8	7
Croatia	3	3
Madagascar	1	0
Papua New Guinea	1	1
Saudi Arabia	3	3
USA	1	1
Vanuatu	1	1

[1] Data was adapted from Table S9 of (Wilson *et al.*, 2014).

It may seem that the presence of *'Entothenella'* is biased towards Japanese sponges; however, this might be due to a lack of bioprospecting efforts in other biodiversity hotspot regions. Leal *et al.* (2012) showed that the geographical trends of bioprospecting efforts for marine natural products from invertebrates (1990-2009) are primarily in Asian countries. The relatively increased focus on bioprospecting sponge and invertebrate natural products from the Pacific Ocean (Leal *et al.*, 2012) illustrates the great need for metagenomic studies of sponges from other geographical regions.

It was decided to do preliminary screens with the mDNA of South African sponges with the *'Entotheonella'*-specific 16S rRNA primers. The purpose of this screen was to determine whether the DNA of any *'Entotheonella'* spp. were present. The *'Entotheonella'* spp. are a rich source of different classes of bioactive secondary metabolites (Wilson *et al.*, 2014) and if they are detected in the South African sponges, it will be worthwhile to research which bioactive products they produce.

The mDNA used for this screen was extracted as described in Section 3.4.3. Only sample PE05-A was amplifiable with the Phusion validation PCR (Section 3.4.4.1.) and was screened with the *'Entotheonella*'-specific primers (Ento271F-Ento1290R; Table 14). DNA fragments of approximately 600, 1200 and 1500 bp were amplified (Figure 39, Lane 3). A nested PCR (with Ento735F-Ento1290R) was performed with the resultant PCR reaction mixture, but DNA fragments of incorrect sizes were obtained (Figure 39, Lane 2).

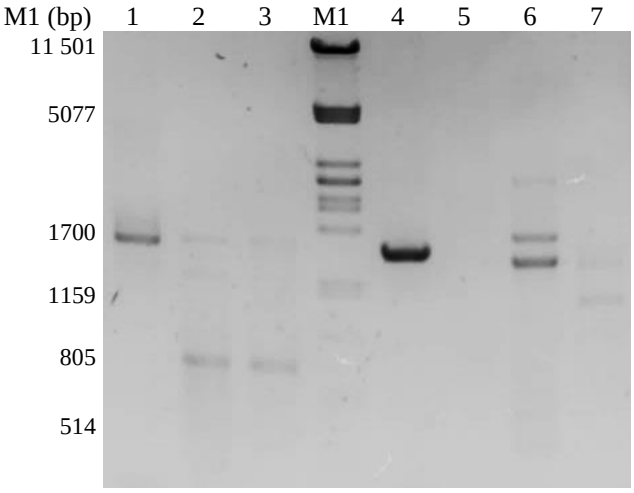


Figure 39 Phusion PCR with 16S rRNA primers: E9F with U1510R (lanes 1 and 4), Ento271F with Ento1290R (lanes 3 and 6) and Ento735F with Ento1290R (lanes 2, 5 and 7). Sample PE05-A was added as template to the reactions in lanes 1, 3 and 7. Lane 2) A nested PCR was performed with the products of a PCR reaction mixture that was the same as lane 3. Lanes 4-6) Control reactions with *D. maris* gDNA as template. DNA marker λ PstI (M1). The control PCR without template DNA was electrophoresed on another gel, and no amplicons were produced.

Screening with the *'Entotheonella'*-specific primers was not pursued any further, primarily because this was not a core objective of this study, moreover, an appropriate control for the screen was not available and the results obtained were inconclusive. Screening for polytheonamides or proteusin related proteins was pursued by screening the extracted mDNA with the designed Ocin1 primers.

3.4.4.3. Screening mDNA with Ocin1 primers

After numerous extraction and purification attempts, the metagenomic DNA that passed the validation PCR was screened with the Ocin1 primer pair. No DNA fragments were produced from the Ocin1 screens with samples PE03, PE05-B and PE14-B. This could be because non-optimal PCR conditions were used or contaminants were still within the sample. Due to the complexity of the mDNA, consisting of a consortium of genomes, it is difficult to find the correct set of PCR parameters for this type of template, as there is a multitude of possibilities for the parameters.

Another reason could be that the target gene is not within the sample, which represents a high probability. Analysing the BLink report (2014) of PoyB and PoyC (with the

parameters set as described within the methods section) only 489 and 386 species, respectively, have homologous proteins and generally one gene per species.

No microbial diversity studies have been done yet for the sponges used in this study and therefore it is not known whether the samples were from high or low microbial biomass sponges. Thus it was not known if the extracted DNA was biased towards the most dominant species within the sponge as this is often encountered with metagenomic DNA extraction (Suenaga, 2012).

It should also be remembered that because the DNA was screened directly it was diluted to a suitable concentration for the PCR, and thus may have resulted in the diluting out of genes represented in low abundance.

3.4.4.4. Screening sponge isolates with Ocin1 primers

It was decided rather to screen the gDNA of bioactive isolates from the *H. esperioides* (PE14; Appendix B; Table 6) sponge to determine if any of them have rSAM homologues and as a way to test the specificity of the primers.

Twenty-eight of the isolates from the sponge, which displayed antimicrobial activity, were selected for the Ocin1 PCR screen. DNA from these isolates was extracted as described in Section 2.4.1.; Chapter 2.

At first the DNA of some of the isolates was extracted and screened in groups (Table 12) with DreamTaq and an increased concentration of $MgCl_2$ with the annealing temperature range of 58 to 49 °C (Figure 40, lanes 1-4). When the same DNA was screened with the same annealing temperature range and Phusion polymerase, no amplicons were obtained (Figure 40, lanes 8-11). As we wanted to evaluate the specificity of the primers, it was decided to screen the gDNA of the isolates with the screening conditions chosen after optimisation with the gDNA of *S. meiloti* (DreamTaq polymerase and increased $MgCl_2$ at annealing temperatures of 58 to 49 °C; refer to Figure 35).

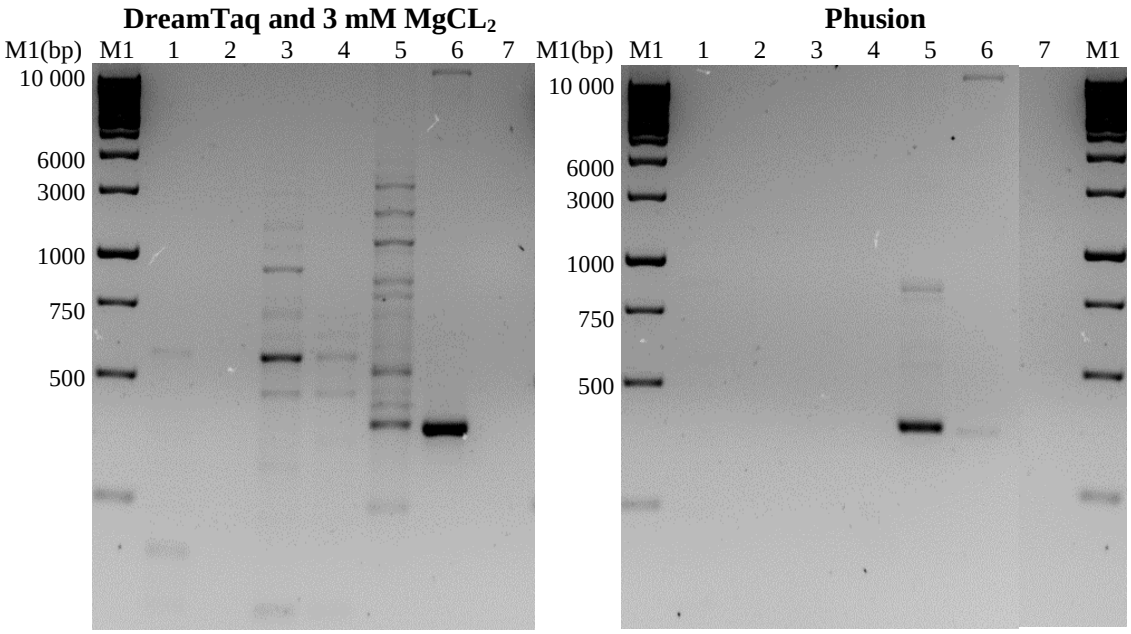


Figure 40 Touchdown PCRs with DreamTaq and 3 mM $MgCL_2$ and Phusion polymerase. Annealing temperature range used 58 to 49 °C. Template DNA added: Group 1 to 4 (lanes 1-4), *S. meliloti* (lane 5) and *poyB*-plasmid (Lane 6) and no DNA (lane 7). The 1 kb DNA ladder (M1) was used. Refer to Table 12 for the individual isolates included in the different groups.

Experiment A, D, B and E (Figure 41) showed the difference in amplified fragments with more stringent annealing temperatures, compared to fragments obtained with annealing temperatures of 58 to 49 °C. Fewer non-specific amplicons were obtained from the sponge isolates when more stringent annealing temperatures were used.

Each isolate was screened at least twice and PCR reactions were relatively reproducible as different reactions with the same template DNA produced similar profiles of amplified DNA fragments (Figure 41). For all of the screens the control DNA fragment (400 bp) was amplified from the *poyB*-plasmid and gDNA from *S. meliloti*. Some of the prominent DNA fragment sizes amplified were about 150 bp and 500 to 600 bp. Fragments between 300 to 500 bp were extracted and sequenced (as described in Section 2.4.4.1.; Chapter 2).

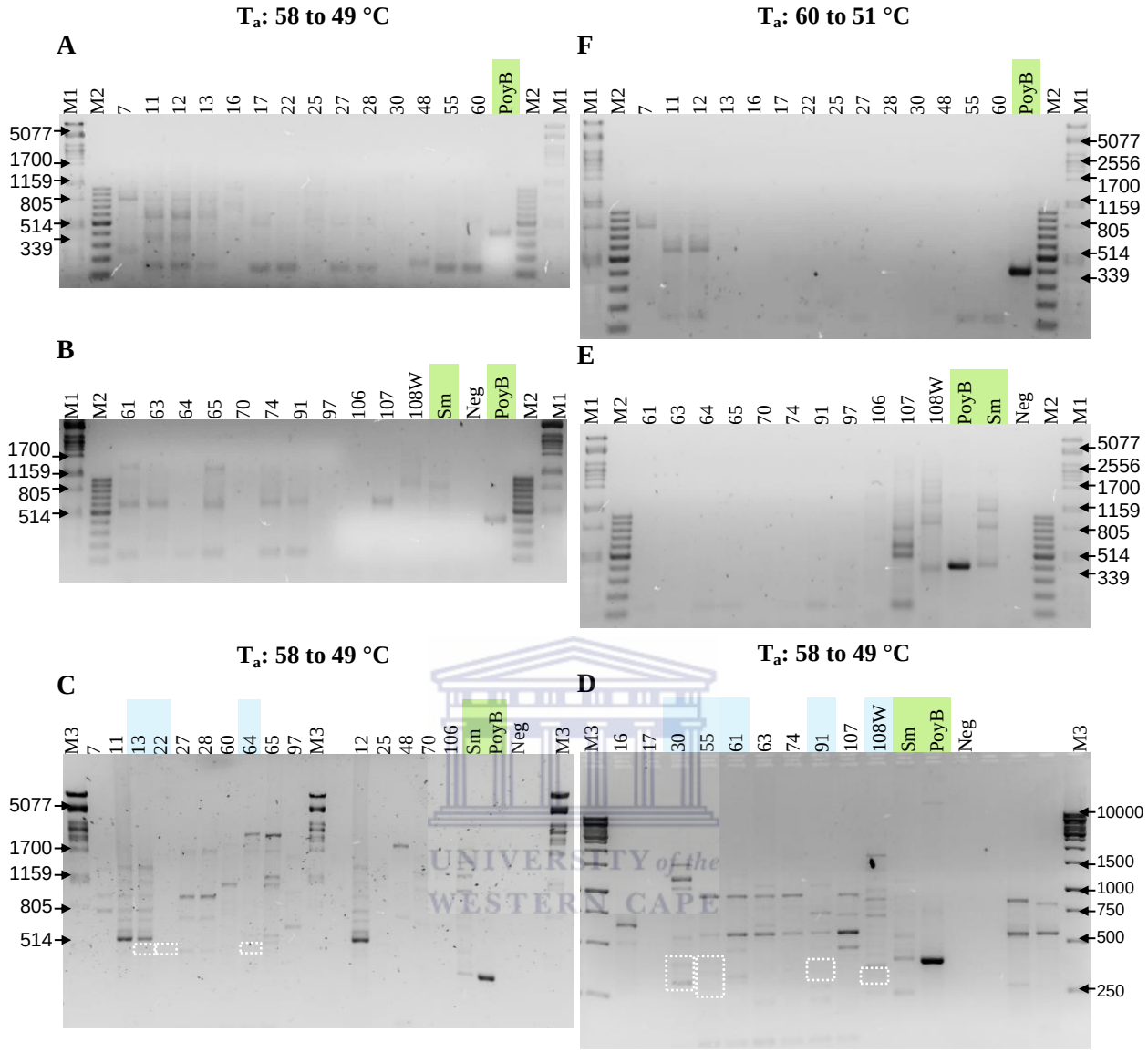


Figure 41 DreamTaq touchdown PCR with annealing temperature ranges 58 to 49 °C (A-D) and 62 to 53°C (E and F). The number above each lane is the PE14-isolate number. The control reactions (green) with *poyB*-plasmid (PoyB) and gDNA from *S. meliloti* (Sm) where the 400 bp fragment represents the partially amplified rSAM_ocin_1 homologous gene. No template DNA added (Neg). DNA markers included M1) λ PstI; M2) 100 bp and M3) 1 kb. The negative control in B is the control for experiment A and B. Similarly, the negative control in E was for experiment E and F. The region between 250 to 600 bp (C and D) was cleaned and used in a second PCR with Ocin1 primers and selected fragments (white blocks) were sequenced and analysed. Unfortunately the loading dye used for some of the samples in A and B masked some of the DNA bands (bright lines at 1 500 and 250 bp region).

The sequences of the selected DNA fragments were analysed with BLASTn and BLASTx. The closest hit, according to the expectation value (E-value) to the query sequence, was recorded (Table 23). None of the amplified genes were related to the target genes. Interestingly homologous genes were amplified from different isolates:

formate dehydrogenase subunit alpha from PE14-13 and PE14-22 and C4-dicarboxylate ABC transporter from PE14-61 and PE14-91.

Alignment of the 16S rRNA sequences of PE14-13 (*P. ascidiaceicola* strain F423) and PE14-22 (*V. splendidus* strain LGP32) showed that these isolates are not the same organisms. The sequences had a similarity of only 82 % (BLAST; NCBI). Although, both organisms showed activity against the same indicator strains (*M. smegmatis*, S. *epidermidis and B. cereus*; Chapter 2), another anti-microbial screen that was performed by colleagues in the laboratory showed that PE14-13 also inhibited the growth of *E. coli* 1699 (Appendix B; Table 7). From this we can conclude that these isolates are most probably not the same strain.

The bioactivity profiles for isolates P14-61 and PE14-91 were not the same (Appendix B; Table 7). They appear to be different strains, but since the 16S rRNA identity of both strains is not known, we cannot definitely conclude whether they are the same strain.

The conserved domains of *poy*B and C, the cobalamin-binding and rSAM domain, were not present in any of the amplified genes. Further research of the function and conserved domains of the top hit (Table 24) indicated that two of the proteins have iron-sulphur binding sites: formate dehydrogenase (FDH, WP_017111042.1) subunit alpha in *V. tasmaniensis* and ribonucleotide-diphosphate reductase (RDR, WP_011261821.1) in *A. fischeri*. The reverse primer was designed to the N-terminus region of the Radical_SAM domain (Figure 31), that contains binding sites for the co-factor SAM and an iron-sulphur cluster. The amplification of non-specific genes with protein domains featuring an iron-sulphur binding site indicates that the reverse primer target the correct sequence, but it is not specific enough for the purpose of screening for rSAM_ocin_1 related genes. This was supported when the iron-sulphur binding sites (protein and gene sequences) of these proteins were aligned with the reverse primer (Figure 42). The alignment showed that the Ocin1R primer could bind to the iron-sulphur binding sequence of other unrelated proteins, for example FDH (Figure 42, A and B). That said, even though there is an iron-sulphur binding sequence in RDR, a low binding consensus was obtained from the alignment of this protein with the Ocin1R primer (Figure 42, D).

Table 23 Top BLASTx hit with the nr database, according to the E-value, of the selected sequenced DNA fragments. The column Apx, indicates the approximate size (after gel electrophoresis) of the DNA fragment that was sequenced. The conserved domains found on the queried sequence was also noted. Different amplicons from the same isolate (column PE14) are denoted as "a" or "b". Proteins of the same colour are homologous to each other.

PE14	Apx (bp)	Description	Organism	Conserved domains (query)	Query cover (%)	E value	ID (%)	Accession
13	500	formate dehydrogenase subunit alpha	*V. tasmaniensis*	MCP_like superfamily; Fer4_7; Fer4_10;	99	5.00E-75	98	WP_017111042.1
22	500	formate dehydrogenase subunit alpha	*V. tasmaniensis*	MCP_like superfamily; Fer4_7; Fer4_10	99	5.00E-75	98	WP_017111042.1
30a	350	ammonium transporter	*Vibrio* sp. JCM 19236	Ammonium_transp superfamily	89	1.00E-49	90	GAM67232.1
30b	350	ribonucleotide-diphosphate reductase	*Al. fischeri*	RNR_PFL superfamily; PRK09103	91	6.00E-45	95	WP_011261821.1
55a	350	phosphoenolpyruvate synthase	*V. tasmaniensis*	PRK06464; Pyruvate_Kinase super family	92	4.00E-54	99	WP_017110707.1
55b	350	4-hydroxyphenylpyruvate dioxygenase	*Vibrio* sp. J2-17	AP_endonuc_2; TIM_phosphate_binding superfamily	92	4.00E-55	93	CDU12663.1
55c	300	putative transcriptional regulator, LysR family protein	*Vibrio* sp. MED222	No putative conserved domains detected	72	1.00E-22	87	EAQ52045.1
61a	500	C4-dicarboxylate ABC transporter	MULTISPECIES: *Vibrio*	DcuA_DcuB superfamily; RE_SinI	68	5.00E-55	94	WP_009844713.1
61b	420	C4-dicarboxylate ABC transporter	*Vibrio cyclitrophicus*	DcuA_DcuB superfamily	98	5.00E-46	93	WP_010435396.1
64	400	flagella basal-body protein	*V. tasmaniensis*	PHA02989	91	3.00E-61	93	WP_017106340.1
91a	500	Conserved hypothetical protein; putative hemolysin-type calcium-binding region	*Vibrio* sp. J2-15	T1SS_rpt_143	98	5.00E-66	93	CDT76620.1
91b	480	C4-dicarboxylate ABC transporter	MULTISPECIES: *Vibrio*	DcuA_DcuB superfamily	95	1.00E-62	96	WP_009844713.1
108 W	500	RND family efflux transporter MFP subunit	*P. putida*	8a0102; HlyD3	93	2.00E-68	93	WP_009395914.1

Table 24 Properties of the protein that was the top hit to the sequenced DNA fragments. The first two columns are a repeat of Table 13 for referencing purposes. Superfamilies of conserved domains were determined with the conserved domain database. The protein cluster, the taxonomical groups in which the protein cluster is conserved (NCBI databases) and the function of the protein according to the clusters of orthologous groups (COGs) were recorded. Proteins of the same colour are homologous to each other.

PE14	Top hit accession #	Superfamily of conserved domains	Protein cluster	PCLA #	Conserved in	COG functional categories
13	WP_017111042.1	fer2 superfamily; HCP_like superfamily; Molybdop_Fe4S4 superfamily; Molybdopterin-Binding superfamily; MopB_CT superfamily; NADB_Rossmann superfamily	formate dehydrogenase	906042	Gammaproteo-bacteria	Amino acid transport and metabolism; Energy production and conversion; General function prediction only
22	WP_017111042.1	fer2 superfamily; HCP_like superfamily; Molybdop_Fe4S4 superfamily; Molybdopterin-Binding superfamily; MopB_CT superfamily; NADB_Rossmann superfamily	formate dehydrogenase	906042	Gammaproteo-bacteria	Amino acid transport and metabolism; Energy production and conversion; General function prediction only
30a	GAM67232.1	Ammonium_transp superfamily	-	-	-	-
30b	WP_011261821.1	ATP-cone superfamily; Ribonuc_red_lgN superfamily; RNR_PFL superfamily	ribonucleotide-diphosphate reductase subunit alpha	419463	Bacteria	Function unknown
55a	WP_017110707.1	Pyruvate_Kinase superfamily	phosphoenolpyruvate synthase	747200	Vibrionales	-
55b	CDU12663.1	Glo_EDI_BRP_like superfamily; TIM_phosphate_binding superfamily	-	-	-	-
55c	EAQ52045.1	HTH superfamily; Periplasmic_Binding_Protein_Type_2 superfamily	-	-	-	-
61a	WP_009844713.1	DcuA_DcuB superfamily	C4-dicarboxylate ABC transporter	2553644	Bacteria	General function prediction only
61b	WP_010435396.1	DcuA_DcuB superfamily	C4-dicarboxylate ABC transporter	2553644	Bacteria	General function prediction only
64	WP_017106340.1		flagellar basal-body protein	874712	Vibrionales	-
91a	CDT76620.1	T1SS_rpt_143 superfamily	-	-	-	-
91b	WP_009844713.1	DcuA_DcuB superfamily	C4-dicarboxylate ABC transporter	2553644	Bacteria	General function prediction only
108W	WP_009395914.1	Biotinyl_lipoyl_domains superfamily; MT superfamily; OEP superfamily; PHA00671 superfamily	cytochrome C peroxidase	908411	Bacteria	-

Information that was not available is indicated by a dash.
PCLA #: Protein cluster accession number.

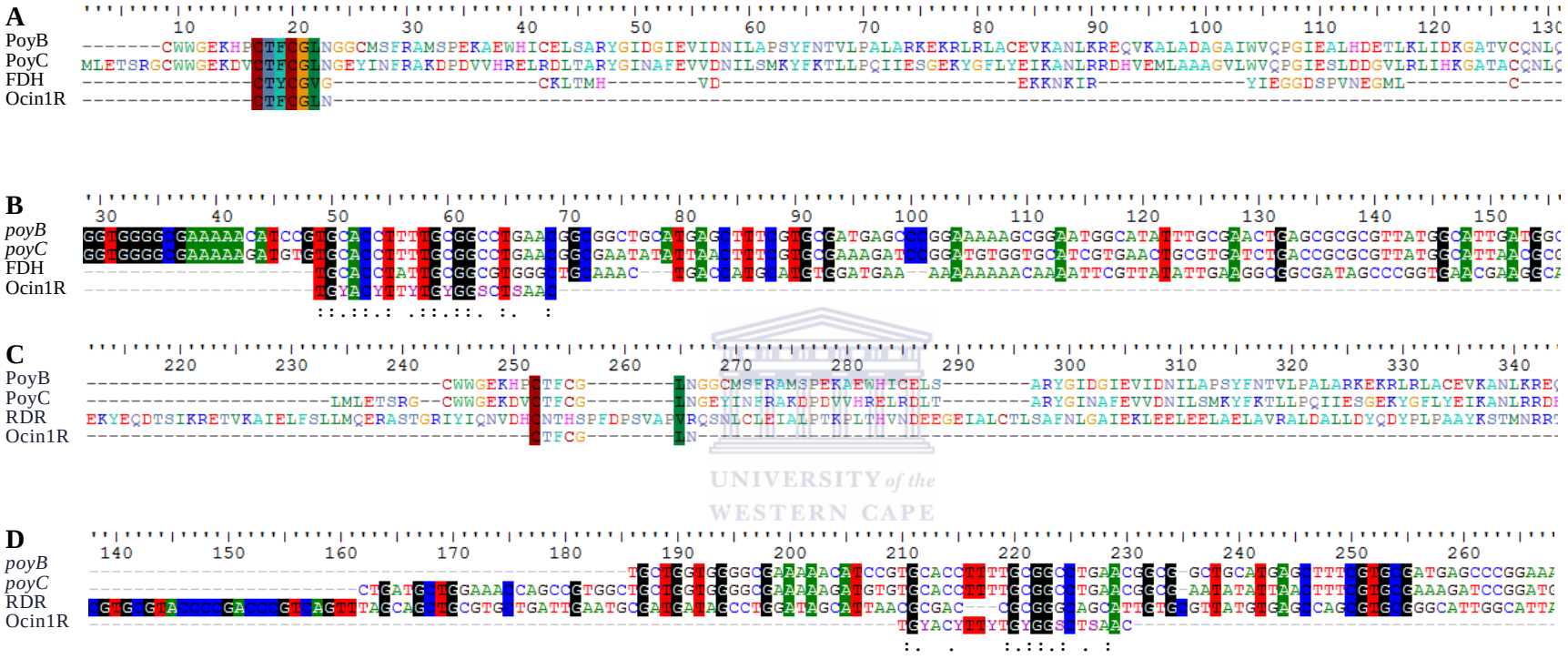


Figure 42 The iron-sulphur binding regions of poyB, poyC, (A and B) formate dehydrogenase subunit alpha (FDH; WP_017111042.1) and (C and D) ribonucleotide-diphosphate reductase (RDR; WP_011261821.1). The proteins and gene was aligned with the reverse-compliment sequence of primer Ocin1R. Alignment was performed with the BioEdit Sequence Alignment Editor (Hall, 1999) and the Clustal W algorithm (Thompson *et al.*, 1994) and the similarity threshold was set to 80 % (highlighted residues). Below the alignment of the nucleotides, the binding consensus is illustrated with a colon for an exact match and a period for a match with an ambiguous sequence.

To determine the binding efficiency of the primers (Table 25) the amplicon sequences were aligned with the ClustalW algorithm (Figure 43) to the original gene sequence of the closest BLAST hit (as listed in Table 24). The primers were then manually aligned to the appropriate regions. This analysis showed that amplification occurred even though there were mismatches scattered across the length of the primer and within the 3' region of the primer. There are indications that the binding of primer Ocin1R was better than the binding of primer Ocin1F. For amplicon 55c, amplification was initiated since the binding of the 3' region of one of the primers was strong, even though the overall binding of the primers was poor (Table 25).

Table 25 Comparison of the binding efficiency of the primers to the Ocin1 amplicons.

Ocin1 amplicon	**BLAST hit accession number (see Table 23)**	**Primer1**		**3' match (last 5 bp)**	**Primer2**		**3' match (last 5 bp)**
13 and 22	WP_017111042.1	NA [1]		NA [1]	NA [1]		NA [1]
30a	GAM67232.1	R	poor	2	R (RC)	strong	5
30b	WP_011261821.1	F	good	5	R (RC)	strong	4
55a	WP_017110707.1	R	good	4	F (RC)	poor	1
55b	CDU12663.1	R	good	5	F (RC)	strong	5
55c	EAQ52045.1	R	poor	1	F (RC)	poor	4
61a and 91b	WP_009844713.1	F	poor	1	R (RC)	strong	5
61b	WP_010435396.1	NA [2]		NA [2]	NA [2]		NA [2]
64	WP_017106340.1	R	strong	4	F (RC)	good	4
91a	CDT76620.1	R	good	4	F (RC)	good	4
108W	WP_009395914.1	R	good	4	F (RC)	good	2

Binding was defined as poor, good or strong according to the number of matches across all three sequences (gene, amplicon and primer). Poor <50%, good =50 to 70%, strong >70%.

[1] The amplicons obtained were sequenced directly and not cloned before sequencing.

[2] Primer binding site could not be identified, even though the amplicon could be aligned to the hit gene, due to poor sequence quality.

RC: The reverse compliment of the sequence was used.

Figure 43 Alignment of the sequenced amplicons with the DNA sequence of the top hit (accession numbers are as in Table 23). The binding sites of the primers are also indicated. Alignment was performed with the BioEdit Sequence Alignment Editor (Hall, 1999) and the Clustal W algorithm (Thompson *et al.*, 1994). Matches in the gene, amplicon and primers are highlighted in yellow. RC indicates that the reverse compliment of a sequence was used.

The amplification of non-specific products with the Ocin1 primers might be related to a few factors. One such factor is the regions selected for the design of primers. These regions are located within conserved domains, specifically the binding site of cobalamin (Ocin1F) and the iron-sulphur cluster (Ocin1R). That being said, during the design of the primers, proteins from the rSAM_ocin_1 family were aligned with proteins from the B12_rSAM_oligo family that are also part of the Radical_SAM superfamily. In this case, the regions selected for the primers were not conserved between the two families (Figure 32). This topic requires more research and analysis can focus on comparing the conserved domains of rSAM_ocin_1 proteins to more protein families with similar domain architectures (a cobalamin binding domain followed by an rSAM domain). From the analysis, regions that are more specific to the rSAM_ocin_1 family can be identified.

4. Summary

In this chapter degenerate primers were designed to screen for a novel RiPP by targeting the biosynthetic genes of the known proteusin, polytheonamide. The first control was to amplify the target gene of the primers, *poyB*, from a construct. Bioinformatic analysis of homologous proteins to the polytheonamide biosynthetic proteins, PoyB and PoyC, showed that *S. meliloti* contains a homologous gene. This was confirmed through *in vitro* amplification with the Ocin1 primers. Subsequently, the mDNA of South African marine sponges was screened with the primers for putative biosynthetic genes of polytheonamide or proteusin related peptides. No amplification was observed from any of the samples, mainly because the mDNA used was not of good enough quality and quantity for PCR screening purposes. There is also a high probability that the gene of interest is not present in the samples. The biosynthetic genes of secondary metabolites constitute less than 2 % of sequenced genomes and therefore it can be concluded that the biosynthetic genes of a natural product is rarely found within mDNA (Charlop-Powers *et al.*, 2014). Our study found that, in the sequenced genomes of prokaryotes available on the NCBI database, 489 and 386 species have homologous proteins to PoyB and PoyC respectively. This further substantiates the claim that the target genes might not be present in the mDNA that was screened.

Consequently, it was decided to focus rather on determining the specificity of the primers. This is why the gDNA of the isolates from the sponge PE-14, *H. esperioides*, was extracted (to represent a fraction of the metagenome of the sponge) and screened with the designed primers. Numerous non-specific genes were amplified, although some of the genes have

similar conserved domains as the target genes. To reduce the non-specificity of the primers, the PCR parameters of the screen with the designed primers can be altered for future screens. Alternatively, the specificity of the primer sequence can be improved or different primers can be designed to regions that do not contain conserved binding sites within the sequences.

Chapter 4

1. Natural product discovery

Natural products have been developed into suitable drugs for the treatment of a wide array of diseases (Newman & Cragg, 2012). Many of the pharmaceuticals available on the market are natural products produced by terrestrial organisms. Marine natural products have gained popularity as highly effective compounds with unprecedented chemical structures (Cuevas & Francesch, 2009; Nakao *et al.*, 1995; Hamada *et al.*, 2005). Marine invertebrates, such as sponges are rich sources of natural products and it has been established that the microbial symbionts are the true producers of some of the known natural products. Examples of these include patellamides, polytheonamides, bryostatins and discodermins, as discussed in Chapter 1. The discovery and development of marine natural products is vital as there is a need for pharmaceuticals such as antimicrobials, anti-tumour, anti-inflammatory, neuro-active, antiviral and antimalarial drugs.

The methods used for the discovery of microbial natural products are classified as cultivation-dependent and -independent. Cultivation-dependent methods were favoured before the turn of the century, whereas cultivation-independent methods have been employed since advances were made with techniques such as sequencing and metagenomics. There are advantages and limitations experienced with cultivation-dependent and -independent methods and this is why a few studies have incorporated both approaches to screen the symbiotic bacteria of invertebrates (Kennedy *et al.*, 2008; Engelhardt *et al.*, 2010; O'Halloran *et al.*, 2011; Öztürk *et al.*, 2013; Bondarev *et al.*, 2013). By employing both approaches some of the limitations are circumvented. With cultivation-dependent methods the organisms producing bioactive compounds are identified, whereas natural product biosynthetic pathways can be elucidated with cultivation-independent methods. The combination of the methods could contribute to the ultimate goal of associating the genotype, chemotype and phenotype of a natural product and exploiting the knowledge gained for the development and production of therapeutic drugs.

In this study the bacterial symbionts of the South African marine sponge, *H. esperioides*, have been researched for the first time. Cultivation-dependent and -independent methods were employed to screen for natural products.

2. Biodiversity and bioactivity of microbial communities in South African marine sponges

2.1. The phylogenetic diversity of the symbiotic communities in sponges

Note that this discussion focusses on the prokaryotic (bacterial and archaeal) symbiotic species of sponges as the eukaryotic symbionts have not been widely researched (Rodríguez-Marconi *et al.*, 2015). The composition of the symbiotic communities of marine invertebrates is dependent on biotic factors such as host-species, and abiotic factors such as the climate and location. The composition of the microbial communities associated with sponges has been widely investigated around the world. Although not absolutely true for all sponge species, the communities within the same sponge species seem to be temporally stable (Taylor *et al.*, 2007; Blanquer *et al.*, 2013). Generally there is no core community of symbionts amongst different sponge species, even within the same location (Taylor *et al.*, 2007; Schmitt *et al.*, 2012; Blanquer *et al.*, 2013). However, the sponge symbionts are more closely related to each other when compared to species found in other marine environments (Schmitt *et al.*, 2012; Blanquer *et al.*, 2013). This is supported by the monophyletic clustering of sponge-specific bacterial species (Schmitt *et al.*, 2012).

South Africa offers a unique environment as it has the cold, mineral rich Benguela current on its Western coast and the warm Agulhas current on the eastern coast. These currents influence the diversity of organisms greatly. Although the biodiversity of South African marine fauna and flora is well known, limited literature is available about the microbial biodiversity (Cowan *et al.*, 2013). Recently the microbial diversity of the South African sponge *Tsitsikamma favus* was published using culture-independent analyses. The bioactive tsitsikammamines extracted from *T. favus* is believed to be produced by the symbiotic microorganisms (Walmsley *et al.*, 2012). Tsitsikammamine B has anti-fungal, antimicrobial and cytotoxic activities (Walmsley *et al.*, 2012). The prevalent species in *T. favus* were within the phylum Proteobacteria, specifically a unique Betaproteobacterium. This is different from what was found in other sponges as this is the first time that a Betaproteobacterium has been reported as the dominant taxon.

T. favus was collected in the Algoa Bay area, as was the *H. esperioides* material used in this study. The scope of this thesis was not to determine the diversity of the symbionts of *H. esperioides*, however, phylogenetic analysis was performed for some of the isolates. These isolates were primarily Proteobacteria of the genera *Pseudovibrio* and *Vibrio* (Alphaproteobacteria and Gammaproteobacteria respectively, Chapter 2). Studies that have used similar culture-dependent and culture-independent methods have found that *Pseudovibrio* and *Vibrio* are regularly found within sponges and that they produce antimicrobial compounds (Flemer *et al.*, 2012; Taylor *et al.*, 2007; O'Halloran *et al.*, 2011; Kennedy *et al.*, 2008; Muscholl-Silberhorn *et al.*, 2008). Species from other genera: *Aliivibrio*, *Enterovibrio*, *Marinomonas*, *Shewanella* and *Bacillus* were isolated from *H. esperioides*, which are also known to be ubiquitous in other sponges (Taylor *et al.*, 2007). Most of the species identified showed 99% similarity to known type cultures. However, a probable novel isolate, PE14-40, was only 96% similar to the *M. dokdonensis* strain DSW10-10 (Gammaproteobacteria) at the 16S rRNA level (discussed further later in this chapter).

In future, a phylogenetic diversity study of the symbionts of *H. esperioides* can be performed to expand our current understanding of the biodiversity and novelty of South African marine microorganisms.

2.2. The effect of climate change on sponges and their symbiotic communities

The effect of climate change has been observed worldwide in all biomes. The fauna and flora of the ocean is also greatly affected. According to the World Wide Fund for Nature (WWF) some of the major effects of climate change are (coral) reef bleaching and acidification of the ocean. Coral reefs are unique ecosystems inhabited by many other macroorganisms such as sponges. The effect of coral bleaching is important as all of the inhabitants of that environment (macro and micro) are influenced.

The effect of climate change on sponges and their symbionts have not been studied extensively, however, Fan *et al.* (2013) recently monitored the response of a Great Barrier Reef sponge and its associated microorganisms when the sponge was subjected to controlled thermal stress. Elevated temperatures induced sponge necrosis. Changes in

the expression of stress-related genes were observed and the symbiotic relationship between the sponge and its associated microorganisms also changed (Fan *et al.*, 2013). The diversity of the microorganisms changed and species with lower host specificity and high growth rates were prevalent. These microorganisms were proven to be nutrient-scavenging bacteria that preyed on the decaying sponge material (Fan *et al.*, 2013). This example of the change in sponge-microbe symbiosis due to thermal stress indicates that environmental changes, such as climate change, greatly affect the survival of marine species.

Bleaching was observed at the high-latitude marginal reefs on the East coast of South Africa. This is a result of, among other things, elevated sea surface temperatures. The health of an encrusting sponge was also deteriorating as a result of bleaching (Celliers & Schleyer, 2002). Although bleaching is observed, the effect of climate change on the symbiotic microorganisms of marine sponges in South Africa has not been significantly researched. This is of great concern, because if we are losing marine species (macro and micro), we are losing a large source of natural products and this could have dire consequences on ecosystem function and on marine industries. Although this thesis was focused more on a drug discovery angle, what has been presented in this piece of work and that of *T. favus* represents important insights of the microbial biodiversity of this unique environment. Since symbiotic microorganisms are considered bio-indicators of changes in environmental conditions (Papapetropoulou & Rodopoulou, 1994; Kefalas *et al.*, 2003), these studies could form the baseline for future monitoring studies to help us understand the effect of climate change on these communities and how they respond to it. Microbial communities or key indicator species (Gordon & Leggat, 2010; Motti, 2012) could be monitored, with metagenomics and metabolomics. Further analysis of the symbiotic relationship of marine organisms and the natural products that they produce could provide us with an understanding of the ecological role of the microbial symbionts.

2.3. Potential bioactive compounds

This study generated a large collection of symbionts of *H. esperioides* that produce bioactive secondary metabolites. The bioassays performed showed that these secondary metabolites have antimicrobial and anti-inflammatory properties and the activities were

primarily against *M. smegmatis* and *S. epidermidis*. Almost a third of the isolates showed antimicrobial activity with the overlay assays that were performed. More than half of the active isolates showed activity against more than one of the tested strains. These activities demonstrate that this collection could harbour a range of important natural products suitable for future screening and development.

The bioactive compounds of this study have not been elucidated and therefore future investigations will include the purification, identification and characterisation of these natural products. The compounds could be purified by sequential bioactivity-guided fractionation of the extracts with chromatographic techniques (Van Middlesworth & Cannell, 2005). Once a pure compound is obtained the structure can be determined with spectroscopic methods and the compound can finally be identified. Dereplication of the bioactive compounds early on in the study is important to prevent replicate work. However, this should be implemented carefully. For instance, if a compound is related to, or identified as a previously extracted compound, the compound can be subjected to different screens to determine what other bioactivities it may exhibit. Also, if the related compound can be supplied more efficiently and exhibit less side effects it might be a better candidate for drug development (Sipkema *et al.*, 2005).

Crude extractions of liquid cultures from 20 isolates were screened by some of the PharmaSea consortium partners. From this, eleven of the extracts showed either anti-fungal or anti-inflammatory activity. The extracts that showed bioactivity can be purified to identify the compound and further analysis can be performed to determine its drug-like properties. Further investigations of the bioactivity capabilities of the isolates could include other pathogenic screens or even non-pharmaceutical related screens.

An isolate that is of interest for future research is the above-mentioned novel isolate PE14-40. Cultivation-dependent screens showed that it produces either a wide range of bioactive compounds or a singular compound with antimicrobial activity, anti-fungal activity as well as anti-inflammatory activity. Sequencing the genome (personal communication; unpublished data) of this organism further revealed its potential to produce novel secondary metabolites. Three NRPSs, a butyrolactone and a bacteriocin biosynthetic pathway was identified. Based on average nucleotide identity analysis of the sequenced genome; PE14-40 represents a novel species, and possibly a novel genus.

The discoveries made from the symbionts of *H. esperioides* demonstrate that traditional cultivation-dependent screening methods are important as novel species that produce potential natural products can be isolated. The research of *T. favus* (Walmsley *et al.*, 2012) and sponges within our laboratory is the foundation for future cultivation-dependent and -independent studies on South African sponges. With this type of research more novel organisms and bioactive compounds from South African marine sponges can be discovered.

3. Discovering RiPPs and understanding their evolutionary mechanisms

3.1. Precursor genes and maturation enzymes

It was originally thought that there is little similarity between the proteins of different RiPP classes (Arnison *et al.*, 2013) and that each class has its own defined precursor and biosynthetic genes. However, recent research shows that the evolution of RiPPs is based on the promiscuity of the maturation enzymes. Various precursors from distantly related genes can be modified with distantly related enzymes. The following paragraphs will discuss recent findings concerning the evolution of RiPPs, with a focus on the proteusins.

The only characterised proteusin is polytheonamide B. The polytheonamide precursor, PoyA, is clustered with RiPP maturation enzymes such as the LanM-like PoyF and the rSAM-related proteins PoyB, PoyC, PoyD and PoyE (Freeman *et al.*, 2012). PoyA, is part of the TOMM_pelo protein family (Haft *et al.*, 2010). The precursor peptides of different RiPP classes including lanthipeptides and TOMMs also have a conserved TOMM_pelo domain (CDD; NCBI). The leader peptides of TOMM_pelo proteins are classified as nitrile hydratase-related leader peptides (NHLP) since they are related to the alpha subunit of nitrile hydratases (NHases). As a result Haft *et al.* (2010) concluded that NHLP proteins and NHases share a common ancestor (Haft *et al.*, 2010; Yang & van der Donk, 2013). NHases convert nitriles into their corresponding amides, but the TOMM_pelo proteins do not function as NHases since they lack the catalytic metal-binding site (Haft *et al.*, 2010).

Another protein family, Nif11 nitrogen-fixing proteins (N11P), could serve as RiPP precursors (Haft *et al.*, 2010; Zhang *et al.*, 2014). These proteins are found in some cyanobacteria and nitrogen-fixing bacteria and even though they play a role in nitrogen metabolism, their exact function is unknown (Jacobson *et al.*, 1989). The NHLP and N11P proteins were found to cluster with known RiPP maturation enzymes and transporter proteins (Haft *et al.*, 2010). The insights Haft *et al.* (2010) obtained from their research led to their conclusion that Nif11 proteins and NHases have been adapted for secondary metabolism by assisting allosteric regulation (Yang & van der Donk, 2013). Since NHLP- and N11P-containing proteins might be precursors to natural products, Haft *et al.* (2010) also cautioned that the classification of small proteins should be considered carefully, even if it is a homologue of a functionalised protein (as for instance NHases). This is supported by the fact that classic gene-identification algorithms do not always recognise the biosynthetic genes of natural products (Haft, 2009; de Jong *et al.*, 2010).

During the maturation of polytheonamide the enzyme PoyF, homologous to LanM-like proteins, catalyse a dehydration reaction (Freeman *et al.*, 2012). LanM-like proteins have been found within the biosynthesis pathways of various RiPPs, even though LanM enzymes were first identified to be involved in the synthesis of lanthipeptides. Hence the mnemonic locus code, Lan, from lanthipeptides. LanM enzymes have two functional domains for dehydration and cyclisation (Zhang *et al.*, 2012), which is key for the formation of the thioether bridges and the resultant rings of lanthipeptides.

Zhang *et al.* (2014) showed that putative precursor peptides of cyanobacterial lanthipeptides (collectively termed LanA) with N11P and NHLP leader peptides are clustered with proteins related to the maturation enzymes of lanthipeptides, LanM-like proteins. Examples include the LanM-like proteins, ProcM and NpnM, of *Prochlorococcus* MIT9313 and *N. punctiforme* PCC 73102. The putative precursor peptide genes associated with these LanM-like synthetases appear to have different evolutionary origins to LanA peptides, since their leader peptides are related to the abovementioned NHases and Niff11 proteins. Some of the cyanobacteria with these precursors do not have other NHase- and Niff11-related genes, and thus it seems that they were obtained by horizontal gene transfer. While analysing the putative LanA

precursors that are clustered with the LanM-like proteins, they also found that some lacked cysteine residues. Analysis of the modification of these precursors by these LanM-like proteins (ProcM and NpnM) resulted in the core peptide to be only dehydrated and not cyclised as for lanthipeptides with cysteine-containing precursors (Zhang *et al.*, 2014). Since the precursor peptides do not require cyclisation Zhang *et al.* (2014) suggested that the lanthipeptide synthetases (ProcM and NpnM) co-evolved with their associated precursors as only the dehydration domain remained functional. This shows that the putative RiPP precursor peptides lacking cysteine residues that are associated with lanthipeptide synthetases are not lanthipeptides, however, their function is not known (Zhang *et al.*, 2014).

Similar to these putative RiPPs, the precursor peptide of polytheonamide, PoyA, has a NHLP leader peptide and also does not have a cysteine residue in the core peptide. Consequently the core peptide is dehydrated by the LanM-like protein, PoyF (Freeman *et al.*, 2012). It seems that LanA precursors with NHLP- and N11P-leader peptides that lack cysteine residues might be more closely related to other RiPP precursors, such as TOMM_pelo precursors, but further phylogenetic analysis is needed to determine the relationship of these precursors.

The different subfamilies of radical SAM proteins perform various enzymatic reactions including the post-translational modification of peptides into natural products (Haft, 2011). The *in silico* analyses presented in this thesis showed that rSAM_ocin_1 related genes, such as *poy*B and *poy*C, are positioned in close proximity to TOMM_pelo family proteins in various bacteria (Chapter 3, Figure 26). Similar to the recent findings of Morinaka *et al.* (2014) our analysis also showed that homologues of the rSAM protein PoyD are also found in several bacteria including some cyanobacteria and nitrogen fixing bacteria, such as *A. variabilis* ATCC 29413, *D. baarsii* DSM 2075 and *S. meliloti* SM11 (Chapter 3, Figure 26). The PoyD homologues clustered with NHLP genes and other putative related biosynthetic genes. Morinaka *et al.* (2014) showed that one of the cyanobacteria strains, *Pleurocapsa sp.* PCC 7319, has N11P-type precursors that could be modified by PlpD, a homologue of PoyD. From their analysis Morinaka *et al.* (2014) concluded that the putative cyanobacterial proteusin genes could produce peptides with

diverse structures. This is due to the difference in the sequence of the precursor peptides as well as the combination of genes involved in biosynthesis.

As seen in the above-mentioned studies we are beginning to understand the evolutionary mechanisms of RiPPs. This indicates the need for detailed analysis of RiPP biosynthetic genes. By conducting studies similar to the ones discussed above we can determine whether the biosynthetic genes of RiPPs from different classes share ancestral origins. This might identify the commonalities and evolutionary mechanisms of RiPPs.

3.2. Focussing on nitrogen metabolising bacteria

It is well known that cyanobacteria produce a plethora of natural products of various chemical classes, including RiPPs (Wang *et al.*, 2011). From this study (Chapter3) and the discussion above, we speculate that many nitrogen metabolising bacteria (diazotrophs), such as some cyanobacteria and rhizobacteria, may contain putative proteusin biosynthetic gene clusters that were probably adapted from their nitrogen metabolising genes, namely Nif11 proteins and NHases. There seems to be an evolutionary relationship between the biosynthetic genes of polytheonamides and other proteusins discovered in *'Entotheonella'* spp. and nitrogen-metabolising genes. It is thus intriguing to consider whether *'Entotheonella'* spp. are nitrogen metabolisers. There is no direct evidence that *'Entotheonella'* spp. are nitrogen metabolising bacteria as they have only been sequenced recently and they have not, despite several efforts, been cultured. Dos Santos *et al.* (Dos Santos *et al.*, 2012) proposed that the presence of a minimum set of six genes (*nifHDK and nifENB*) is indicative of a diazotroph. A preliminary scan of the *'Entotheonella'* genome indicated that proteins related to nitrogen metabolism (Appendix C; Table 11) are present; however a nitrogen fixation operon was not identified. The origin of the nitrogen fixation genes may thus be due to horizontal gene transfer as was proposed by Zhang *et al.* (2014) for cyanobacteria with NHLP- and N11P RiPP precursors that do not have other NHase or Nif11 related genes.

Since some of the biosynthetic genes of proteusins and other RiPPs are related to nitrogen metabolising genes it will be interesting to determine how or why these primary metabolising genes have evolved into secondary metabolism products.

Final Conclusions

As we have seen, the search for novel natural products, driven by the need for new and effective antimicrobials, is achieved by combining cultivation-dependent and –independent methods. These methods were used in this study to screen the bacterial symbionts of *H. esperiordes.* We showed that various pharmaceutical-relevant natural products are produced by these symbionts and future studies will show us their true potential.

There is also great potential in the screening of RiPPs, especially proteusins, as only polytheonamides have been isolated so far and other (non-*'Entotheonella'*) species have been identified that contain putative proteusin pathways. It will be exciting to see what future studies might reveal, as we have suggested that nitrogen-metabolising bacteria can be targeted for the screening of RiPPs and proteusins.

Appendices

Appendix A

Artificial sea salts (Table 1) were added to the media, unless specified, to have a similar salinity to sea water. If a medium was used in a solid state 15 $g.L^{-1}$ of bacteriological agar was added before sterilisation, except for PMM that was made with 12 $g.L^{-1}$.

Table 1 Solid media used to cultivate the symbiotic bacteria isolated from the sponge PE14.

Medium	pH	Composition	Per 1 L	Source
Artificial sea salts	NA	NaCl $MgCl_2$ KCl $CaCl_2$	18.0 g 2.0 g 0.525 g 0.075 g	
TSA	7.5	Tryptic soya broth	3.0 g	Merck
TS4	4.5	Tryptic soya broth	3.0 g	
SNA	7.5	Peptone Beef Extract	5.0 g 3.0 g	
ZBA	7.5	Yeast Extract Peptone	1.25 g 3.75 g	
ZBV	7.5	ZBA (see above) 0.1% Crystal Violet Solution [1]	 1.0 ml	
SWA	6.5	Sea Water Wolf's Mineral Solution [2] Wolf's Vitamin Solution [2]	250.0 ml 5.0 ml 10.0 ml	
CCA	7.5	Cellulose Xylan Chitin NH_4Cl	0.5 g 0.5 g 0.5 g 0.5 g	
SEM	NA	Sponge Extract [3]	10.0 g	
GYM [5, 6]	7.2	D-Glucose Yeast Extract Malt Extract $CaCO_3$ [4] NaCl $MgCl_2$ • 6H2O KCl $CaCl_2$	4.0 g 4.0 g 10.0 g 2.0 g 24.0 g 5.3 g 0.7 g 0.1 g	
172 [6]	7.5	Glucose Yeast Extract Starch Tryptone	10.0 g 5.0 g 10.0 g 5.0 g	

		$MgSO_4 \cdot 7H_2O$ $CaSO_4 \cdot 7H_2O$	2.0 g 2.0 g	
OMA [6]	NA	Jungle Oats Oatso Easy (No sugar)	40.0 g	Tiger Brands, South Africa
AIM	8.1	Actinomycete Isolation Agar	22.0 g	Sigma-Aldrich
R2A [5]	NA	Reasoner´s 2A agar	15.2 g	Sigma-Aldrich
ACM	7.0	HEPES Sodium pyruvate Yeast Extract Soybean peptone $NaNO_3$ $KHPO_4$ $MgSO_4 \cdot 7H_2O$ Activated charcoal [4]	2.38 g 3.00 g 0.10 g 3.00 g 0.34 g 0.10 g 0.15 g 3.00 g	
MAA [5,6]	10.0-10.5	Glucose Peptone Yeast Extract KH_2PO_4 $MgSO_4 \cdot 7H_2O$ NaCl Na_2CO_3	10.0 g 5.0 g 5.0 g 1.0 g 0.1 g 20.0 g 10.0 g	
PCS	NA	KNO_3 $(NH_4)2HPO_4$ $MgSO_4 \cdot 7H_2O$ $CaCl_2 \cdot 2H_2O$ $FeCl_2$	0.20 g 0.02 g 0.01 g 5.00 mg 0.50 mg	
PMM	NA	Glucose $(NH_4)2SO_4$ Peptone Yeast Extract	1.00 g 0.25 g 0.15 g 0.15 g	

NA: pH was not adjusted.

[1] Crystal violet solution: 2.3% (w/v) crystal violet; 0.1% (w/v) ammonium oxalate; 20% (v/v) ethanol.

[2] Wolfe's solution: refer to Table 2.

[3] Sponge extract was prepared by homogenising 1 part Redbait with 3 parts marine sponge. The solid material was removed by centrifugation and supernatant was freeze dried before addition to the medium.

[4] This was omitted in liquid medium.

[5] Artificial sea water salts were not included.

[6] Cycloheximide and nalidixic acid were added at 100 µg.ml^{-1}and 50 µg.ml^{-1}, respectively.

Table 2 Wolfe's supplementary solutions for sea water agar. The solutions were filter sterilised through a 0.22 µm filter and stored at 4°C. The mineral solution was prepared by dissolving the nitrilotriacetic acid in water and adjusting the pH to 6.5 with KOH before adding the other reagents.

Wolfe's mineral solution	**g.L^{-1}**	**Wolfe's vitamin solution**	**mg.L^{-1}**
Nitrilotriacetic acid	1.5	Biotin	2.0
$MgSO_4$ • 7H2O	3.0	Folic Acid	2.0
$MnSO_4$ • H2O	0.5	Pyridoxine hydrochloride	10.0
NaCl	1.0	Thiamine-HCl	5.0
$FeSO_4$ • 7H2O	0.1	Riboflavin	5.0
$CaCl_2$	0.1	Nicotinic acid	5.0
$CoCl_2$ • 6H2O	0.1	Calcium D-(+)-pantothenate	5.0
$ZnSO_4$ • 7H2O	0.1	Vitamin B12	0.1
$CuSO_4$ • $5H_2O$	0.01	p-Aminobenzoic acid	5.0
$AlK(SO)_4$ • $12H_2O$	0.01	Thioctic acid	5.0
H_3BO_3 (Boric Acid)	0.01		
Na_2MoO_4 • $2H_2O$	0.01		

Table 3 BG11 liquid medium **(Rippka et al., 1979)** for the isolation of cyanobacteria.

Stock Solution	**Reagent**	**g.L^{-1}**	**BG11 (ml.L^{-1})**
1	$NaNO_3$	15.00	100
2	K_2HPO_4	2.00	10
3	$MgSO_4$ • $7H_2O$	3.75	10
4	$CaCl_2$ • $2H_2O$	1.80	10
5	Citric Acid	0.30	10
6	Ammonium ferric citrate	0.30	10
7	Na_2 EDTA	0.05	10
8	Na_2CO_3	1.00	10
Wolf's mineral solution [a]			1

[a] The composition of the solution is presented in Table 2.

Table 4 General media used in this study.

Medium	Composition (per litre or specified otherwise)	Ref.
Lysogeny broth (LB)	10 g Tryptone, 5 g Yeast Extract, 5 g NaCl	(Bertani, 2004)
TY	5 g Tryptone, 3 g Yeast Extract 10ml of sterile 1 M $CaCl_2$ ($CaCl_2$ stock made and autoclaved separately and added before pouring of plates or inoculation of liquid culture.)	(Beringer, 1974)

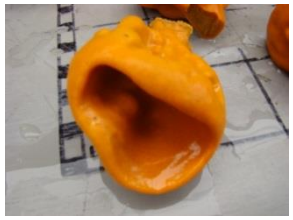
Table 5 Buffers and antibiotics used within this study. The composition and concentration is also listed.

Buffer	Abbr.	Composition	pH
Tris-EDTA buffer	TE	10 mM Tris-HCl, 1 mM EDTA	8.0
Tris-acetate-EDTA buffer	TAE	40 mM Tris-base, 20 mM acetic acid, 1 mM EDTA	8.4
Tris-NaCl buffer	-	10 mM Tris-HCl, 0.5 M NaCl	8.0
Tris-borate-EDTA buffer	TBE	89 mM Tris-base, 89 mM Boric acid, 2 mM EDTA	8.0

Antibiotic	Abbr.	Working concentration ($ug.ml^{-1}$)	Media
Kanamycin	Kan	100	LB
Ampicillin	Amp	50	LB
Spectinomycin	Spc	50	LB

Appendix B

Table 6 Details of the sponge material used. All collections and identifications were done by Dr. Shirley Parker-Nance (2013).

	Collection details		Taxonomy and description notes		Photo
PE03	Date collected:	23 January 2013	Phylum:	Porifera	 
	Site collected:	Algoa Bay, White Sands Reef	Class:	Calcarea	
			Subclass:	Calcinea	
	Depth:	23-25m depth	Order:	Clathrinida	
	Longitude:	34°00.366S	Family:	Clathrinidae	
	Latitude:	25°43.209E	Genus:	*Guancha*, Miklucho-Maclay, 1868	
	Habitat notes:	Low profile rocky reef	Species:	*Guancha* sp. 001RSASPN	
	Also:				
	Date collected:	23 January 2013	Close affinities to *Guancha pulcherrima* (Dendy, 1891) and *G. stiptata* (Dendy, 1891) both from Australia and both species very small with differences in spicules shape, size, growth form and colour of the organism. Not *Clathrina blanca* or C. *blanca* forma *pulcherrima* as previously identifies in literature.		
	Site collected:	Algoa Bay, Phillips Reef			
	Depth:	10-13m depth			
	Longitude:	33°58.452S			
	Latitude:	25°40.807E			
	Habitat notes:	Medium profile rocky reef			
PE05	Date collected:	23 January 2013	Phylum:	Porifera	 
	Site collected:	Algoa Bay, White Sands Reef	Class:	Demospongiae	
			Order:	Poecilosclerida	
	Depth:	23-25m depth	Suborder:	Microcionina	
	Longitude:	34°00.366S	Family:	Raspailiidae	
	Latitude:	25°43.209E	Subfamily:	Cyamoninae	
	Habitat notes:	Low profile rocky reef	Genus:	*Waltherarndtia*	
			Species:	*Waltherarndtia caliculatum*, Kirkpatrick, 1903	
			Common and characteristic sponge.		

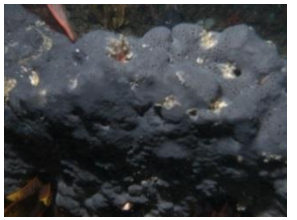
PE08	Date collected: Site collected: Reef Depth: Longitude: Latitude: Habitat notes:	23 January 2013 Algoa Bay, White Sands 23-25m depth 34°00.366S 25°43.209E Low profile rocky reef	Phylum: Class: Order: Family: Genus: Subgenus: Species:	Porifera Demospongia Dictyoceratida Spongiidae Spongia *Spongia* (Spongia), Linnaeus, 1759 *Spongia*(Spongia) sp.001RSASPN		
PE14	Date collected: Site collected: Reef Depth: Longitude: Latitude: Habitat notes:	4 April 2013 Algoa Bay, White Sands 23-25m depth 34°00.406S 25°43.117E Low profile rocky reef	Phylum: Class: Order: Suborder: Family: Genus: Subgenus: Species:	Porifera Demospongia Poecilosclerida Mycalina Hamacanthidae Hamacantha *Hamacantha* (Vomerula), Schmidt, 1880 *Hamacantha* (Vomerula) *esperioides*, Ridley and Dendy, 1886		

Table 7 Isolates from the sponge PE14 which showed antimicrobial activity with the overlay assay. This is a more comprehensive table (compared to Table 6; Chapter 2) of the antimicrobial assays. The assays were performed by the author of this thesis as well as colleagues at IMBM.

PE14 ID	*E. coli* 1699	*M. smegmatis*	*S. epidermidis*	*B. cereus*	*P. putida*	Number of organisms active against
3	x	x	x			3
7		x		x		2
10	x					1
11	x	x		x		3
12		x		x		2
13	x	x	x	x		4
16			x			1
17		x	x	x		3
20	x					1
22		x	x	x		3
25		x	x	x		3
27	x					1
28	x	x	x			3
30	x	x				2
33	x					1
40	x			x		2
48			x	x		2
55		x	x	x		3
58			x			1
60		x	x	x		3
61		x	x	x		3
62	x					1
63		x	x			2
64	x					1
65			x			1
73			x			1
74	x	x	x	x		4
82				x		1
91		x	x			2
92	x					1
96	x					1

97		x				1
104	x	x				2
106	x					1
108	x					1
116		x	x			2
117			x			1
127		x				1
128		x				1
129		x	x	x		3
144			x			1
145		x				1
146		x	x			2
Total	18	24	22	15	0	43

Appendix C

Table 8 The structural gene, *poyA*, is found along with at least two of the maturation enzymes involved in post-translational modification of PoyA, within three known organisms (dark grey). Key maturation enzymes, as well as the organism which have homologous genes to *poyB* or *poyC*, are highlighted in light grey.

	PoyK	PoyJ	PoyB&C	PoyA	PoyD	PoyE	PoyF	PoyG	PoyH	PoyI	Number homologous proteins[a]
Cystobacter violaceus Cb vi76			B				x			x	3
Desulfarculus baarsii DSM 2075			x	x	x						3
Enhygromyxa salina			x				x			x	3
Fischerella sp. PCC 9605				x			x			x	3
Kamptonema				x	x					x	3
***Microcoleus* sp. PCC 7113**	**x**		**x**				**x**			**x**	**4**
Micromonospora sp. ATCC 39149			x					x	x		3
Microvirgula aerodenitrificans		x				x	x				3
Nostoc sp. PCC 7120			x	x			x				3
Pleurocapsa sp. PCC 7319				x	x		x				3
Sandaracinus amylolyticus		x	C						x		3
Scytonema hofmanni				**x**	**x**		**x**			**x**	**4**
Streptomyces		**x**	**x**			**x**	**x**		**x**		**5**
Streptomyces acidiscabies			x			x			x		3
Streptomyces sp. NTK 937		x	x				x				3
unclassified *Verrucomicrobia* (miscellaneous)			x	x	x						3

[a] Homologous proteins to the biosynthetic genes of polytheonamide from the bacterium symbiont of *Theonella swinhoei* pTSMAC1

Table 9 Accession numbers of the rSAM_ocin_1 proteins used for the construction of the maximum parsimonious tree (Figure 1).

AFS60638.1	WP_015925801.1	WP_019872099.1
AFS60637.1	WP_015182552.1	WP_011541977.1
CCM09442.1	WP_012826784.1	WP_018827354.1
CCM09446.1	YP_003266069.1	WP_018724970.1
CCM09450.1	ACY14176.1	WP_018818802.1
AFU90397.1	WP_002634339.1	WP_018813532.1
AFU90405.1	EJJ23222.1	WP_018219751.1
AFU90401.1	WP_011654150.1	WP_018811133.1
WP_010996184.1	WP_018739566.1	WP_011084777.1
WP_011317980.1	WP_012013523.1	WP_014497841.1
WP_011682652.1	WP_018832590.1	WP_018742506.1
WP_015924487.1	WP_018736939.1	WP_014529191.1
WP_013376083.1	WP_018215450.1	WP_018263008.1
WP_002619126.1	WP_018731946.1	WP_012330098.1
WP_011750427.1	WP_018821137.1	WP_011750427.1

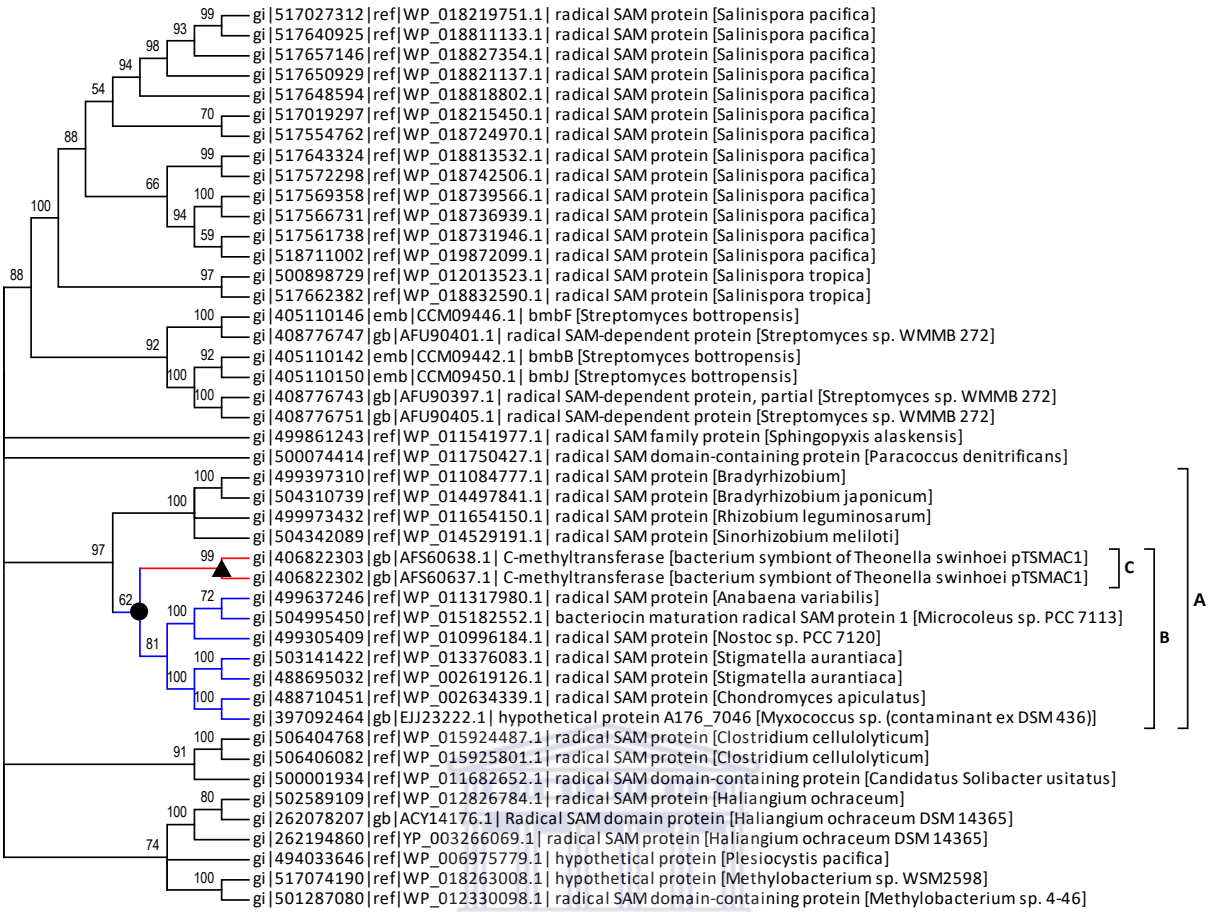


Figure 1 The maximum parsimonious (MP) tree of the representative proteins of the rSAM_ocin_1 family. Clade A contains the proteins: PoyB and PoyC (Clade C) and the seven closest related proteins (Clade B). Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates were collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown above the branches (Dunbar & Mitchell, 2013). The MP tree was obtained using the SPR algorithm (pg. 126 in ref. (Knerr & van der Donk, 2012)) with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). The analysis involved 45 amino acid sequences. All positions containing gaps and missing data were eliminated. There were 459 positions in the final dataset. Analyses were conducted in MEGA5 (Tamura *et al.*, 2011). The reference sequences (RefSeq; NCBI), denoted with the prefix 'WP_' in their accession numbers, were used as the representatives of identical sequences.

Table 10 The degeneracy of different regions were compared when designing the primers. The IUPAC ambiguity code for incompletely specified bases in nucleic acid sequences was used (Cornish-Bowden, 1985).

Primer1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**
Consensus (MEGA5)		M			G			G			A			N			C			E			G			P			M			G			K	
IUPAC	A	T	G	G	G	N	G	G	N	G	C	N	A	A	Y	T	G	Y	G	A	R	G	G	N	C	C	N	A	T	G	G	G	N	A	A	R
Conserved r-transl degenerate	**A**	**T**	**K**	**G**	**G**	**Y**	**G**	**G**	**H**	**G**	**C**	**R**	**A**	**A**	**C**	**T**	**G**	**Y**	**G**	**A**	**R**	**R**	**G**	**C**	**S**	**M**	**R**	**A**	**T**	**G**	**G**	**G**	**K**	**S**	**T**	**G**
Degeneracy			**2**			**2**			**3**			**2**						**2**			**2**	**2**			**2**	**2**	**2**						**2**	**2**		
	96																																			
Conserved at 100%	**A**	**T**	**K**	**G**	**G**	**Y**	**G**	**G**	**H**	**G**	**C**	**R**	**A**	**A**	**C**	**T**	**G**	**Y**	**G**	**A**	**R**	**R**	**G**	**C**												
	192																																			
										128																										
										128																										
										256																										
PoyB_nt	A	T	G	G	G	C	G	G	T	G	C	G	A	A	C	T	G	T	G	A	G	G	G	G	C	C	A	A	T	G	G	G	T	A	A	G
PoC_nt	A	T	G	G	G	T	G	G	A	G	C	A	A	A	T	T	G	C	G	A	A	A	G	C	G	A	G	A	T	G	G	G	G	C	G	G
		M			G			G			A			N			C			E			G			P			M			G			K	
PoyB Primer1 219-230	A	T	G	G	G	C	G	G	C	G	C	G	A	A	C	T	G	C	G	A	A	G	G	C	C	C	G	A	T	G	G	G	C	A	A	A
		M			G			G			A			N			C			E			S			E			M			G			R	
PoyC Primer1 219-230	A	T	G	G	G	C	G	G	C	G	C	G	A	A	C	T	G	C	G	A	A	A	G	C	G	A	A	A	T	G	G	G	C	C	G	T
		M			G			G			A			N			C			E			G			A			M			G			L	
Nostoc homolog Primer1 219—230	A	T	G	G	G	C	G	G	C	G	C	G	A	A	C	T	G	C	G	A	A	G	G	C	G	C	G	A	T	G	G	G	C	C	T	G
		I			G			G			A			N			C			E			G			S			M			G			V	
Anabeana homolog Primer1 219—230	A	T	T	G	G	C	G	G	C	G	C	G	A	A	C	T	G	C	G	A	A	G	G	C	A	G	C	A	T	G	G	G	C	G	T	G
		M			G			G			A			N			C			E			G			E			M			G			V	
Stigmatella homolog Primer1 219-230	A	T	G	G	G	C	G	G	C	G	C	G	A	A	C	T	G	C	G	A	A	G	G	C	G	A	A	A	T	G	G	G	C	G	T	G
		M			G			G			A			N			C			E			G			V			M			G			L	
Microcoleus homolog Primer1 219-230	A	T	G	G	G	C	G	G	C	G	C	G	A	A	C	T	G	C	G	A	A	G	G	C	G	T	G	A	T	G	G	G	C	C	T	G
		M			G			G			A			N			C			E			G			D			M			G			V	
Chondnomyces homolog Primer1 219-230	A	T	G	G	G	C	G	G	C	G	C	G	A	A	C	T	G	C	G	A	A	G	G	C	G	A	T	A	T	G	G	G	C	G	T	G
		M			G			G			A			N			C			E			G			D			M			G			V	
Myxococcus homolog Primer1 219-230	A	T	G	G	G	C	G	G	C	G	C	G	A	A	C	T	G	C	G	A	A	G	G	C	G	A	T	A	T	G	G	G	C	G	T	G
		L			G			G			S			N			C			E			G			S			M			G			R	
bmbBJ Primer1 219-230	C	T	G	G	G	C	G	G	C	A	G	C	A	A	C	T	G	C	G	A	A	G	G	C	A	G	C	A	T	G	G	G	C	C	G	T
		M			G			G			S			N			C			E			G			S			M			G			R	
bstCF Primer1 219-230	A	T	G	G	G	C	G	G	C	A	G	C	A	A	C	T	G	C	G	A	A	G	G	C	A	G	C	A	T	G	G	G	C	C	G	T
		L			G			G			A			N			C			D			R			P			M			G			Q	
bmbF bstJ Primer1 219-230	C	T	G	G	G	C	G	G	C	G	C	G	A	A	C	T	G	C	G	A	T	C	G	T	C	C	G	A	T	G	G	G	C	C	A	G

Primer2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**
Consensus (MEGA5)		K			H			H			C			T			F			C			G			L			N			G	
IUPAC	A	A	R	C	A	Y	C	A	Y	T	G	Y	C	A	N	T	T	Y	T	G	Y	G	G	N	Y	T	N	A	A	Y	G	G	N
Conserved r-transl degenerate	**A**	**A**	**A**	**C**	**A**	**Y**	**C**	**A**	**T**	**T**	**G**	**C**	**A**	**C**	**C**	**T**	**T**	**T**	**T**	**G**	**C**	**G**	**G**	**C**	**C**	**T**	**G**	**A**	**A**	**C**	**G**	**G**	**C**
Degeneracy						**2**			**2**																								
Consensus poyBC_nt_r-transl	**A**	**A**	**A**	**C**	**A**	**Y**	**S**	**A**	**G**	**T**	**G**	**C**	**A**	**C**	**C**	**T**	**T**	**T**	**T**	**G**	**C**	**G**	**G**	**C**	**C**	**T**	**G**	**A**	**A**	**C**	**G**	**G**	**C**
						2	**2**																										
Consensus r-transl IUPAC poyBC	**A**	**A**	**A**	**C**	**A**	**B**	**C**	**H**	**K**	**T**	**G**	**C**	**A**	**C**	**C**	**T**	**T**	**T**	**T**	**G**	**C**	**G**	**G**	**C**	**C**	**T**	**G**	**A**	**A**	**C**	**G**	**G**	**C**
Degeneracy						**3**		**3**	**2**																								
Conserved at 100%	**18**																																
Conserved at 90%							**6**																										
							6																										
	18																																
										0																							
										T	**G**	**C**	**A**	**C**	**C**	**T**	**T**	**T**	**T**	**G**	**C**	**G**	**G**	**C**	**C**	**T**	**G**	**A**	**A**	**C**			
										0																							
more degenerate										T	G	Y	A	C	Y	T	T	Y	T	G	Y	G	G	S	C	T	S	A	A	C			
										64																							
PoyB_nt	A	A	A	G	A	C	G	T	C	T	G	T	A	C	C	T	T	C	T	G	T	G	G	G	T	T	G	A	A	T	G	G	C
PoC_nt	A	A	A	C	A	C	C	C	G	T	G	C	A	C	T	T	T	T	T	G	C	G	G	C	C	T	C	A	A	C	G	G	T
		K			D			V			C			T			F			C			G			L			N			G	
PoyB Primer2 344—354	A	A	A	G	A	T	G	T	G	T	G	C	A	C	C	T	T	T	T	G	C	G	G	C	C	T	G	A	A	C	G	G	C
		K			H			P			C			T			F			C			G			L			N			G	
Poyc Primer2 344—354	A	A	A	C	A	T	C	C	G	T	G	C	A	C	C	T	T	T	T	G	C	G	G	C	C	T	G	A	A	C	G	G	C
		K			H			H			C			T			F			C			G			L			N			G	
Nostoc homolog Primer2 344—354	A	A	A	C	A	T	C	A	T	T	G	C	A	C	C	T	T	T	T	G	C	G	G	C	C	T	G	A	A	C	G	G	C
		K			N			H			C			T			F			C			G			L			N			G	
Anabeana homolog Primer2 344—354	A	A	A	A	A	C	C	A	T	T	G	C	A	C	C	T	T	T	T	G	C	G	G	C	C	T	G	A	A	C	G	G	C
		K			H			H			C			T			F			C			G			L			N			G	
Stigmatella homolog Primer2 344-354	A	A	A	C	A	T	C	A	T	T	G	C	A	C	C	T	T	T	T	G	C	G	G	C	C	T	G	A	A	C	G	G	C
		K			Q			H			C			T			F			C			G			L			N			G	
Microcoleus homolog Primer2 344-354	A	A	A	C	A	G	C	A	T	T	G	C	A	C	C	T	T	T	T	G	C	G	G	C	C	T	G	A	A	C	G	G	C
		K			H			H			C			T			F			C			G			L			N			G	
Chondromyces homolog Primer2 344-354	A	A	A	C	A	T	C	A	T	T	G	C	A	C	C	T	T	T	T	G	C	G	G	C	C	T	G	A	A	C	G	G	C
		K			H			H			C			T			F			C			G			L			N			G	
Myxococcus homolog Primer2 344-354	A	A	A	C	A	T	C	A	T	T	G	C	A	C	C	T	T	T	T	G	C	G	G	C	C	T	G	A	A	C	G	G	C
		K			T			H			C			T			F			C			G			L			N			G	
bmbBJ Primer2 344-354	A	A	A	A	C	C	C	A	T	T	G	C	A	C	C	T	T	T	T	G	C	G	G	C	C	T	G	A	A	C	G	G	C
		K			T			H			C			T			F			C			G			L			N			G	
bstCF Primer2 344—354	A	A	A	A	C	C	C	A	T	T	G	C	A	C	C	T	T	T	T	G	C	G	G	C	C	T	G	A	A	C	G	G	C
		K			L			Q			C			T			F			C			G			E			N			G	
bmbF bstJ Primer2 344-354	A	A	A	C	T	G	C	A	G	T	G	C	A	C	C	T	T	T	T	G	C	G	G	C	G	A	A	A	A	C	G	G	C

A ***poyB*** **(JX456532.1)**

GTTGAGGCCGCAAAAAGTGCACGGGTGTTTTTCCCCCCACCAGCAACCGCGCGCGGTCTCGACGAGAAGCCCAGGTGTCGCAA
AGTATTCGATTTTTGTCTCTGCCAATTCGCGGAAGTAATCGTCGTAGTCCGGCAGGGGTAATCCGTCCATGTCCTGAGTCGTG
GCGCGCTCGACCACCTGAAAATTAGCGAACGGTCGGGACGCGGGGGTGAATACGAAACGGCGCAGCGCCTCGTTAGACGCGAT
GCCCGCAATGTCGTTGTGAAGAATCCGCTCGCAGAGGTCGGGGAACACCTCGTCCCCATCACCCGAGACGACGTAATCTACCC
AGTCGAAGTTCGTATGGAGAGCCCGCCCCATCTCGCTTTCGCAATTTGCTCCACCCAT

390 bp

B ***poyC*** **(JX456532.1)**

TTCAACCCACAGAAGGTACAGACGTCTTTCTCTCCCCACCAGCAGCCCCGCGAGGTCTCAAGCATGAGCCCCGGCAGAACCTG
CCGGGCGACCCCGGCTTCCTGAAGGTCGCGGAAGTAGTCGTCGTAATCGGGGGTTGGGAGAGAACTGAGATCCTTAGCCACAG
AGCGCTCGCTCGCTTCCGCCTGCAGCATGAGCCCTGTTGAGACCTGAAAGATCGACCCAGGCGCCCAGTCAGCGATGCGGCGC
ATGGCTTGCTCCCGGTCAGTTGCGTCCAGGATCAATGCACAGAACTCCGGAAAAAGGTCGTCCGCCTCGCCTGAGAACACGTG
ATCCACCCAGTCGAAACACTGCTTGAGCGCCTTACCCATTGGCCCCTCACAGTTCGCACCGCCCAT

398 bp

C **Nostoc sp. PCC 7120 (NC_003272)**

GTTCAATCCACAAAAAGTGCAATGATGTTTTTGCCCCCACCAACATCCACGAGAGGTTTCAATCAAAAGTCCAGGTTGAATGT
AGGGCGAAAGTGTAGAAGCATGGAGGGTTTGAAAGTAATCATCGTAGTCAGGAATTGGTACTTGATCTAAGTTCTGTACGGAA
GCTCTTGGCGCGGCTATGCTGATTTTGGTAGAAGTCCGATGAGATGGACTAATTACACCATAAGGCAATATTGATGGATCAAC
CTCTCGACCACCATCCAAAAGCTGATAGCAAAACTCATTAAATAAATCGTCACCCTCTCCCGAAATGATAAAATCTACCCAAG
AAAAATGTGTGTGAGTAGCTAACCCCATAGCCCCTTCACAGTTAGCACCTCCCAT

387 bp

D ***Stigmatella aurantiaca*** **DW4/3-1 (NC_014623)**

CGTTGAGGCCACAGAAGGTGCAATGCGCCTTGGCACCCCACCAGCAACCGCGAGAGGACTCAGCGGTCAGGTTGGCGCGCTTG
GCCATCGGCATGTCCTTGATGGCCTGGAAGTAGTGATCGAAATCCGGCATCGGCAGCGAGTCCATGTTCTCGATCTGAGGCGC
CGCCACATAGCGCGGCTGCGGCCGGTTCGTCCCAGGGTTCAGCACCCCCTTGACGAGATCGAGGATGACCCCCTCCCCCTCTC
CGGAGACCACATGATCCAGGAAGGGGAAGTTGTCCGAGATCGCCTTGCCCATGTCCGCTTCGCAGTTGGCGCCCCCCA

327 bp

Figure 2 A - D) Amplicons obtained with the *in silico* PCR analysis. The accession number of the genome that was analysed is given in brackets.

A ***poyB***

CGCAGAAGGTGCACGGGTGTTTTTCCCCCCACCAGCAACCGCGCGCGGTCTCGACGAGAAGCCCAGGTGTCGCAAAGTATTCG
ATTTTTGTCTCTGCCAATTCGCGGAAGTAATCGTCGTAGTCCGGCAGGGGTAATCCGTCCATGTCCTGAGTCGTGGCGCGCTC
GACCACCTGAAAATTAGCGAACGGTCGGGACGCGGGGGTGAATACGAAACGGCGCAGCGCCTCGTTAGACGCGATGCCCGCAA
TGTCGTTGTGAAGAATCCGCTCGCAGAGGTCGGGGAACACCTCGTCCCCATCACCCGAGACGACGTAATCTACCCAGTCGAAG
TTCGTATGGAGAGCCCGCCCCATCTCGCTTTCACAGTTCGC

373 bp

B ***In vitro* amplicon *S. meliloti*:**

TGTTCAGGCCGCAGAAGGTGCAGTGGCTGCGCTGCCCCCACCAGCAGCCGCGAGAGAACTCCATCGGCAGCCCGGGATAGACG
CGGTCGGCGTAGAGCGACTGATCGAGTTCGGCAAAATAGTCCGAGAAATCGGGCAGCGGCAGGTCATGCATATTCTCGACCAC
TGCGCGGGGCACCAAGTCGCCCGTTGTGGTCGCAGGATAGCCGGCCTTGCGATGTGCTGGACCAAAGACGCCGAATGGTAGGT
TGGCCGGCGGAATATCTTTACCGTGCTTTAAAATGTTGCGACAAAGCGGGCCGATCAGTGCGTCGGCCTCGCCCGACACGACG
TAGTCGACCCACGGGAAACTCGCGTGAGTGGTACGTCCCATCACCG

378 bp

C

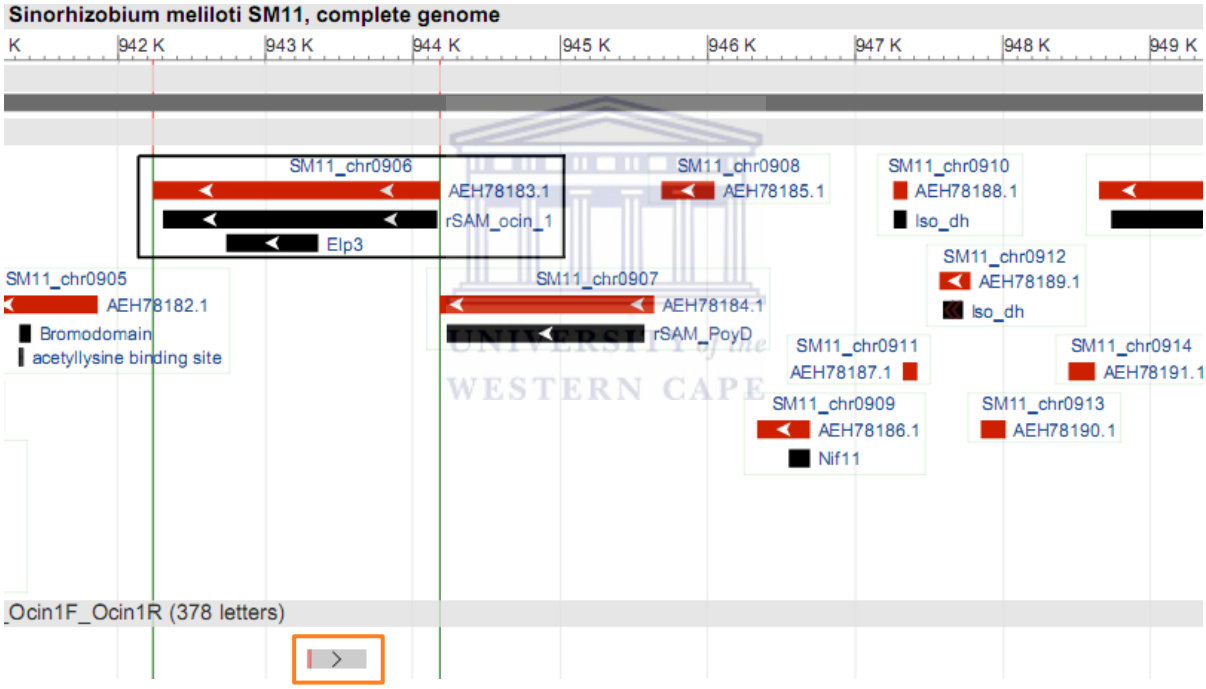


Figure 3 The *in vitro* PCR product of the control DNA: poyB (A) and the PCR product obtained from the gDNA of *S. meliloti* 1021 (B). C) BLAST analysis showed that the amplicon (from B; orange box) was a hit to the rSAM_ocin_1 protein (black box) in *S. meliloti* SM11.

Table 11 Examples of the nitrogen fixation related genes in Entotheonella spp. Nitrogen fixation genes (defined by their conserved domains) were used as the query for BLAST analysis. The analysis was restricted by selecting Candidatus Entotheonella (taxid: 93171) in the "Organism" field of the BLAST search set. All other settings were set as default.

Conserved domain/ gene	**BLASTp (locus_tag)**	**BLASTn**
nifD	NA	No significant similarity found
nifH	NA	No significant similarity found
nifK	NA	No significant similarity found
nifE	NA	No significant similarity found
nifN	NA	No significant similarity found
nifB	NA	No significant similarity found
NifA	ETSY1_04610 ETSY2_36840	
Nif11	ETSY2_03690 ETSY1_03980 ETSY1_15720 ETSY_26035	
NifM	ETSY2_32570	

NA: not available

References

Abu Al-Soud, W. & Râdström, P. 1998. Capacity of nine thermostable DNA polymerases to mediate DNA amplification in the presence of PCR-inhibiting samples. *Applied and Environmental Microbiology*, 64(10): 3748–3753.

Ah Tow, L. 2003. Non-specificity of Staphylococcus generic primers. *Microbiology*, 149(7): 1605–1607.

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

Ang, K.K.H., Holmes, M.J., Higa, T., Hamann, M.T. & Kara, U. a K. 2000. In vivo antimalarial activity of the β-carboline alkaloid manzamine A. *Antimicrobial Agents and Chemotherapy*, 44(6): 1645–1649.

Arnison, P.G., Bibb, M.J., Bierbaum, G., Bowers, A. a, Bugni, T.S., Bulaj, G., Camarero, J. a, Campopiano, D.J., Challis, G.L., Clardy, J., Cotter, P.D., Craik, D.J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P.C., Entian, K.-D., Fischbach, M. a, Garavelli, J.S., Göransson, U., Gruber, C.W., Haft, D.H., Hemscheidt, T.K., Hertweck, C., Hill, C., Horswill, A.R., Jaspars, M., Kelly, W.L., Klinman, J.P., Kuipers, O.P., Link, a J., Liu, W., Marahiel, M. a, Mitchell, D. a, Moll, G.N., Moore, B.S., Müller, R., Nair, S.K., Nes, I.F., Norris, G.E., Olivera, B.M., Onaka, H., Patchett, M.L., Piel, J., Reaney, M.J.T., Rebuffat, S., Ross, R.P., Sahl, H.-G., Schmidt, E.W., Selsted, M.E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Süssmuth, R.D., Tagg, J.R., Tang, G.-L., Truman, A.W., Vederas, J.C., Walsh, C.T., Walton, J.D., Wenzel, S.C., Willey, J.M. & van der Donk, W. a. 2013. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Natural Product Reports*, 30(1): 108–160.

Begley, M., Cotter, P.D., Hill, C. & Ross, R.P. 2009. Identification of a novel two-peptide lantibiotic, lichenicidin, following rational genome mining for LanM proteins. *Applied and Environmental Microbiology*, 75(17): 5451–5460.

Béjà, O., Suzuki, M.T., Koonin, E. V, Aravind, L., Hadd, A., Nguyen, L.P., Villacorta, R., Amjadi, M., Garrigues, C., Jovanovich, S.B., Feldman, R.A. & DeLong, E.F. 2000. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environmental Microbiology*, 2(5): 516–529.

Bentley, S., Chater, K., Cerdeño-Tárraga, A.-M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.-H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M.-A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J. & Hopwood, D.A. 2002. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature*, 417(2): 141–147.

Beringer, J.E. 1974. R Factor Transfer in Rhizobium leguminosarum. Microbiology, 84(1): 188–198.

Bernan, V.S., Greenstein, M. & Maiese, W.M. 1997. Marine microorganisms as a source of new natural products. In S. L. Neidleman & A. I. Laskin, eds. *Advances in applied microbiology*. New York: Academic Press: 57–90.

Bertani, G. 2004. Lysogeny at Mid-Twentieth Century: P1, P2, and Other Experimental Systems. *Journal of Bacteriology*, 186(3): 595–600.

Berthelet, M., Whyte, L.G. & Greer, C.W. 1996. Rapid, direct extraction of DNA from soils for PCR analysis using polyvinylpolypyrrolidone spin columns. *FEMS Microbiology Letters*, 138(1): 17–22.

Bikandi, J., Millán, R.S., Rementeria, A. & Garaizar, J. 2004. In silico analysis of complete bacterial genomes: PCR, AFLP-PCR and endonuclease restriction. *Bioinformatics*, 20(5): 798–799.

Birnboim, H.C. & Doly, J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*, 7(6): 1513–1523.

Biver, S. 2013. Bacillus subtilis as a tool for screening soil metagenomic libraries for antimicrobial activities. *Journal of Microbiology and Biotechnology*, 23(6): 850–855.

Blanquer, A., Uriz, M.J. & Galand, P.E. 2013. Removing environmental sources of variation to gain insight on symbionts vs. transient microbes in high and low microbial abundance sponges. *Environmental Microbiology*, 15(11): 3008–3019.

Blin, K., Kazempour, D., Wohlleben, W. & Weber, T. 2014. Improved lanthipeptide detection and prediction for antiSMASH. *PLoS ONE*, 9(2): e89420.

Bode, H.B., Bethe, B., Höfs, R. & Zeeck, A. 2002. Big effects from small changes: possible ways to explore nature's chemical diversity. *ChemBioChem*, 3(7): 619–627.

Bondarev, V., Richter, M., Romano, S., Piel, J., Schwedt, A. & Schulz-Vogt, H.N. 2013. The genus Pseudovibrio contains metabolically versatile bacteria adapted for symbiosis. *Environmental Microbiology*, 15(7): 2095–2113.

Brady, S.F., Chao, C.J., Handelsman, J. & Clardy, J. 2001. Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. *Organic Letters*, 3(13): 1981–1984.

Brück, W.M., Sennett, S.H., Pomponi, S.A., Willenz, P. & McCarthy, P.J. 2008. Identification of the bacterial symbiont Entotheonella sp. in the mesohyl of the marine sponge Discodermia sp. *The ISME Journal*, 2(3): 335–339.

Burkardt, H.J. 2000. Standardization and quality control of PCR analyses. *Clinical chemistry and laboratory medicine: CCLM / FESCC*, 38(2): 87–91.

Caboche, S., Leclère, 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*, 192(19): 5143–5150.

Caboche, S., Pupin, M., Leclere, V., Fontaine, A., Jacques, P. & Kucherov, G. 2007. NORINE: a database of nonribosomal peptides. *Nucleic Acids Research*, 36(Database): D326–D331.

Cane, D.E. 1997. Introduction: polyketide and nonribosomal polypeptide biosynthesis. From Collie to Coli. *Chemical Reviews*, 97(7): 2463–2464.

Cane, D.E. & Walsh, C.T. 1999. The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. *Chemistry and Biology*, 6: 319–325.

Celliers, L. & Schleyer, M.H. 2002. Coral bleaching on high-latitude marginal reefs at Sodwana Bay, South Africa. *Marine Pollution Bulletin*, 44(12): 1380–1387.

Charlop-Powers, Z., Banik, J.J., Owen, J.G., Craig, J.W. & Brady, S.F. 2013. Selective enrichment of environmental DNA libraries for genes encoding nonribosomal peptides and polyketides by phosphopantetheine transferase-dependent complementation of siderophore biosynthesis. *ACS Chemical Biology*, 8(1): 138–143.

Charlop-Powers, Z., Milshteyn, A. & Brady, S.F. 2014. Metagenomic small molecule discovery methods. *Current Opinion in Microbiology*, 19: 70–75.

Claesen, J. & Bibb, M.J. 2011. Biosynthesis and regulation of grisemycin, a new member of the linaridin family of ribosomally synthesized peptides produced by Streptomyces griseus IFO 13350. *Journal of Bacteriology*, 193(10): 2510–2516.

Claesen, J. & Bibb, M.J. 2010. Genome mining and genetic analysis of cypemycin biosynthesis reveal an unusual class of posttranslationally modified peptides. *Roceedings of the National Academy of Sciences of the United States of America*, 107(37): 16297–16302.

Connon, S.A. & Giovannoni, S.J. 2002. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Applied and Environmental Microbiology*, 68(8): 3878–3885.

Cornish-Bowden, A. 1985. Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. *Nucleic Acids Research*, 13(9): 3021–3030.

Courtois, S., Cappellano, C.M., Ball, M., Francou, F.-X., Normand, P., Helynck, G., Martinez, A., Kolvek, S.J., Hopke, J., Osburne, M.S., August, P.R., Nalin, R., Guerineau, M., Jeannin, P., Simonet, P. & Pernodet, J.-L. 2003. Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Applied and Environmental Microbiology*, 69(1): 49–55.

Cowan, D.A., Rybicki, E.P., Tuffin, M.I., Valverde, A. & Wingfield, M.J. 2013. Biodiversity: so much more than legs and leaves. *South African Journal of Science*, 109(11/12): 1–9.

Cuevas, C. & Francesch, A. 2009. Development of Yondelis (trabectedin, ET-743). A semisynthetic process solves the supply problem. *Natural Product Reports*, 26(3): 322–337.

Davidson, S.K., Allen, S.W., Lim, G.E., Anderson, C.M. & Haygood, M.G. 2001. Evidence for the biosynthesis of bryostatins by the bacterial symbiont "Candidatus Endobugula sertula" of the bryozoan Bugula neritina. *Applied and Environmental Microbiology*, 67(10): 4531–4537.

Davies-Coleman, M.T. 2005. Bioactive natural products from southern African marine invertebrates. In A.- Rahman, ed. *Studies in Natural Products Chemistry: Bioactive Natural Products (Part L)*. Studies in Natural Products Chemistry. Gulf Professional Publishing: 61–107.

de Jong, A., van Heel, A.J., Kok, J. & Kuipers, O.P. 2010. BAGEL2: mining for bacteriocins in genomic data. *Nucleic Acids Research*, 38(Web Server): W647–W651.

de Jong, A., van Hijum, S.A.F.T., Bijlsma, J.J.E., Kok, J. & Kuipers, O.P. 2006. BAGEL: a web-based bacteriocin genome mining tool. *Nucleic Acids Research*, 34(Web Server): W273–W279.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K. & Mattick, J.S. 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research*, 19(14): 4008–4008.

Donia, M.S., Ravel, J. & Schmidt, E.W. 2008. A global assembly line for cyanobactins. *Nature Chemical Biology*, 4(6): 341–343.

Donia, M.S. & Schmidt, E.W. 2011. Linking chemistry and genetics in the growing cyanobactin natural products family. *Chemistry and Biology*, 18(4): 508–519.

Dos Santos, P.C., Fang, Z., Mason, S.W., Setubal, J.C. & Dixon, R. 2012. Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics*, 13(1): 162.

Dower, W.J., Miller, J.F. & Ragsdale, C.W. 1988. High efficiency transformation of E.coli by high voltage electroporation. *Nucleic Acids Research*, 16(13): 6127–6145.

Dunbar, K.L. & Mitchell, D. a. 2013. Revealing nature's synthetic potential through the study of ribosomal natural product biosynthesis. *ACS Chemical Biology*, 8(3): 473–487.

Dunlap, W.C., Battershill, C.N., Liptrot, C.H., Cobb, R.E., Bourne, D.G., Jaspars, M., Long, P.F. & Newman, D.J. 2007. Biomedicinals from the phytosymbionts of marine invertebrates: a molecular approach. *Methods*, 42(4): 358–376.

Duquesne, S., Destoumieux-Garzón, D., Peduzzi, J. & Rebuffat, S. 2007. Microcins, gene-encoded antibacterial peptides from enterobacteria. *Natural Product Reports*, 24(4): 708–734.

Engelhardt, K., Degnes, K.F., Kemmler, M., Bredholt, H., Fjaervik, E., Klinkenberg, G., Sletta, H., Ellingsen, T.E. & Zotchev, S.B. 2010. Production of a new thiopeptide antibiotic, TP-1161, by a marine Nocardiopsis species. *Applied and Environmental Microbiology*, 76(15): 4969–4976.

Engelhardt, K., Degnes, K.F. & Zotchev, S.B. 2010. Isolation and characterization of the gene cluster for biosynthesis of the thiopeptide antibiotic TP-1161. *Applied and Environmental Microbiology*, 76(21): 7093–7101.

Eppelmann, K., Doekel, S. & Marahiel, M. a. 2001. Engineered biosynthesis of the peptide antibiotic bacitracin in the surrogate host Bacillus subtilis. *Journal of Biological Chemistry*, 276(37): 34824–34831.

Evans, D.A., Carter, P.H., Carreira, E.M., Charette, A.B., Prunet, J.A. & Lautens, M. 1999. Total Synthesis of Bryostatin 2. *Journal of the American Chemical Society*, 121(33): 7540–7552.

Fan, L., Liu, M., Simister, R., Webster, N.S. & Thomas, T. 2013. Marine microbial symbiosis heats up: the phylogenetic and functional response of a sponge holobiont to thermal stress. *The ISME Journal*, 7(5): 991–1002.

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*, 61(7): 2798–2801.

Faulkner, D.J. 2000. Highlights of marine natural products chemistry (1972-1999). *Natural Product Reports*, 17(1): 1–6.

Fedorova, N.D., Moktali, V. & Medema, M.H. 2012. Bioinformatics approaches and software for detection of secondary metabolic gene clusters. In N. P. Keller & G. Turner, eds. *Fungal Secondary Metabolism: Methods and Protocols*. New York: Humana Press: 23–45.

Férir, G., Petrova, M.I., Andrei, G., Huskens, D., Hoorelbeke, B., Snoeck, R., Vanderleyden, J., Balzarini, J., Bartoschek, S., Brönstrup, M., Süssmuth, R.D. & Schols, D. 2013. The lantibiotic peptide labyrinthopeptin A1 demonstrates broad anti-HIV and anti-HSV activity with potential for microbicidal applications. *PLoS ONE*, 8(5): e64010.

Finking, R. & Marahiel, M.A. 2004. Biosynthesis of Nonribosomal Peptides 1. *Annual Review of Microbiology*, 58(1): 453–488.

Fischbach, M.A. & Walsh, C.T. 2009. Antibiotics for emerging pathogens. *Science*, 325(5944): 1089–1093.

Flemer, B., Kennedy, J., Margassery, L.M., Morrissey, J.P., O'Gara, F. & Dobson, A.D.W. 2012. Diversity and antimicrobial activities of microbes from two Irish marine sponges, Suberites carnosus and Leucosolenia sp. *Journal of Applied Microbiology*, 112(2): 289–301.

Floss, H.G. 2006. Combinatorial biosynthesis – Potential and problems. *Journal of Biotechnology*, 124(1): 242–257.

Flühe, L. & Marahiel, M.A. 2013. Radical S-adenosylmethionine enzyme catalyzed thioether bond formation in sactipeptide biosynthesis. *Current Opinion in Chemical Biology*, 17(4): 605–612.

Freeman, M.F., Gurgui, C., Helf, M.J., Morinaka, B.I., Uria, A.R., Oldham, N.J., Sahl, H.-G., Matsunaga, S. & Piel, J. 2012. Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. *Science* (New York, N.Y.), 338(6105): 387–390.

Fusetani, N. & Matsunaga, S. 1993. Bioactive sponge peptides. *Chemical Reviews*, 93(5): 1793–1806.

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

Gomez-Escribano, J.P., Song, L., Bibb, M.J. & Challis, G.L. 2012. Posttranslational β-methylation and macrolactamidination in the biosynthesis of the bottromycin complex of ribosomal peptide antibiotics. *Chemical Science*, 3(12): 3522–3525.

Gordon, B.R. & Leggat, W. 2010. Symbiodinium - Invertebrate symbioses and the role of metabolomics. *Marine Drugs*, 8(10): 2546–2568.

Gross, R. 2011. Diversity and catalytic potential of Actinobacteria collected from fynbos soils of the Western Cape of South Africa. Internal Institute for Microbial Biotechnology and Metagenomics report. Unpublished.

Guex, N. & Peitsch, M.C. 1997. SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis*, 18(15): 2714–2723.

Gurgui, C. & Piel, J. 2010. Metagenomic approaches to identify and isolate bioactive natural products from microbiota of marine sponges. *Methods in Molecular Biology*, 668: 247–264.

Haft, D.H. 2009. A strain-variable bacteriocin in Bacillus anthracis and Bacillus cereus with repeated Cys-Xaa-Xaa motifs. *Biology Direct*, 4(1): 15.

Haft, D.H. 2011. Bioinformatic evidence for a widely distributed, ribosomally produced electron carrier precursor, its maturation proteins, and its nicotinoprotein redox partners. *BMC Genomics*, 12(1): 21.

Haft, D.H., Basu, M. & Mitchell, D.A. 2010. Expansion of ribosomally produced natural products: a nitrile hydratase- and Nif11-related precursor family. *BMC Biology*, 8(1): 70.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program fro Windows 95/98/NT. *Nucleic Acids Symposiom Series*, 41: 95–98.

Hamada, T., Matsunaga, S., Yano, G. & Fusetani, N. 2005. Polytheonamides A and B, s, Theonella s winhoei. *Journal of the American Chemical Society*, 127(1): 110–118.

Hamada, T., Sugawara, T., Matsunaga, S. & Fusetani, N. 1994. Polytheonamides, unprecedented highly cytotoxic polypeptides from the marine sponge Theonella swinhoei 2. Structure elucidation. *Tetrahedron Letters*, 35(4): 609–612.

Hammami, R., Zouhir, A., Ben Hamida, J. & Fliss, I. 2007. BACTIBASE: a new web-accessible database for bacteriocin characterization. *BMC Microbiology*, 7(1): 89.

Hammami, R., Zouhir, A., Le Lay, C., Ben Hamida, J. & Fliss, I. 2010. BACTIBASE second release: a database and tool platform for bacteriocin characterization. *BMC Microbiology*, 10(1): 22.

Handelsman, J., Liles, M.R. & Mann, D. 2002. Cloning the metagenome: culture-independent access to thediversity and functions of the uncultivated microbial world. In *Methods in Microbiology*. Elsevier Science Ltd: 241–255.

Hardoim, C.C.P., Costa, R., Araújo, F. V, Hajdu, E., Peixoto, R., Lins, U., Rosado, A.S., van Elsas, J.D. & Elsas, J.D. Van. 2009. Diversity of bacteria in the marine sponge Aplysina fulva in Brazilian coastal waters. *Applied and Environmental Microbiology*, 75(10): 3331–3343.

Harvey, A.L. 2008. Natural products in drug discovery. *Drug Discovery Today*, 13(19-20): 894–901.

Harvey, A.L. 2000. Strategies for discovering drugs from previously unexplored natural products. *Drug Discovery Today*, 5(7): 294–300.

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*, 1(1): 33–43.

Heidelberg, K.B., Gilbert, J.A. & Joint, I. 2010. Marine genomics: at the interface of marine microbial ecology and biodiscovery. *Microbial Biotechnology*, 3(5): 531–543.

Hoffman, C.S. & Winston, F. 1987. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of Escherichia coli. *Gene*, 57: 267–272.

Hoffmeister, D. & Keller, N.P. 2007. Natural products of filamentous fungi: enzymes, genes, and their regulation. *Natural Product Reports*, 24(2): 393–416.

Hou, Y., Tianero, M.D.B., Kwan, J.C., Wyche, T.P., Michel, C.R., Ellis, G. a, Vazquez-Rivera, E., Braun, D.R., Rose, W.E., Schmidt, E.W. & Bugni, T.S. 2012. Structure and biosynthesis of the antibiotic bottromycin D. *Organic Letters*, 14(19): 5050–5053.

Huo, L., Rachid, S., Stadler, M., Wenzel, S.C. & Müller, R. 2012. Synthetic biotechnology to study and engineer ribosomal bottromycin biosynthesis. *Chemistry & Biology*, 19(10): 1278–1287.

Hussain, S., Fareed, S., Ansari, S. & Sajid, M. 2012. Marine natural products : a lead for anti-cancer. *Indian Journal of Geo-Marine Sciences*, 41(February): 27–39.

Ianora, A., Miralto, A. & Romano, G. 2012. *Handbook of marine natural products*. E. Fattorusso, W. H. Gerwick, & O. Taglialatela-Scafati, eds. Dordrecht: Springer Netherlands.

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*, 41(Database issue): D408–D414.

Ikeda, H., Shin-ya, K. & Omura, S. 2014. Genome mining of the Streptomyces avermitilis genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. *Journal of the Society for Industrial Microbiology and Biotechnology*, 41(2): 233–250.

Imhoff, J.F., Labes, A. & Wiese, J. 2011. Bio-mining the microbial treasures of the ocean: new natural products. *Biotechnology Advances*, 29(5): 468–482.

Inoue, M., Shinohara, N., Tanabe, S., Takahashi, T., Okura, K., Itoh, H., Mizoguchi, Y., Iida, M., Lee, N. & Matsuoka, S. 2010. Total synthesis of the large non-ribosomal peptide polytheonamide B. *Nature Chemistry*, 2(4): 280–285.

Iwamoto, M., Shimizu, H., Muramatsu, I. & Oiki, S. 2010. A cytotoxic peptide from a marine sponge exhibits ion channel activity through vectorial-insertion into the membrane. *FEBS Letters*, 584(18): 3995–3999.

Jack, R.W., Tagg, J.R. & Ray, B. 1995. Bacteriocins of gram-positive bacteria. *Microbiological Reviews*, 59(2): 171–200.

Jacobson, M.R., Brigle, K.E., Bennett, L.T., Setterquist, R.A., Wilson, M.S., Cash, V.L., Beynon, J., Newton, W.E. & Dean, D.R. 1989. Physical and genetic map of the major nif gene cluster from Azotobacter vinelandii. *Journal of Bacteriology*, 171(2): 1017–1027.

Jensen, P.R. & Fenical, W. 1996. Marine bacterial diversity as a resource for novel microbial products. *Journal of Industrial Microbiolog*, 17: 346–351.

Kawulka, K., Sprules, T., McKay, R.T., Mercier, P., Diaper, C.M., Zuber, P. & Vederas, J.C. 2003. Structure of subtilosin A, an antimicrobial peptide from Bacillus subtilis with unusual posttranslational modifications linking cysteine sulfurs to α-carbons of phenylalanine and threonine. *Journal of the American Chemical Society*, 125(16): 4726–4727.

Kefalas, E., Castritsi-Catharios, J. & Miliou, H. 2003. Bacteria associated with the sponge Spongia officinalis as indicators of contamination. *Ecological Indicators*, 2(4): 339–343.

Kelly, W.L., Pan, L. & Li, C. 2009. Thiostrepton biosynthesis: prototype for a new family of bacteriocins. *Journal of the American Chemical Society*, 131(12): 4327–4334.

Kennedy, J., Codling, C.E., Jones, B. V, Dobson, A.D.W. & Marchesi, J.R. 2008. Diversity of microbes associated with the marine sponge, Haliclona simulans, isolated from Irish waters and identification of polyketide synthase genes from the sponge metagenome. *Environmental Microbiology*, 10(7): 1888–1902.

Kennedy, J., Marchesi, J.R. & Dobson, A.D.W. 2007. Metagenomic approaches to exploit the biotechnological potential of the microbial consortia of marine sponges. *Applied Microbiology and Biotechnology*, 75(1): 11–20.

Kersten, R.D., Yang, Y.-L., Xu, Y., Cimermancic, P., Nam, S.-J., Fenical, W., Fischbach, M.A., Moore, B.S. & Dorrestein, P.C. 2011. A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nature Chemical Biology*, 7(11): 794–802.

Ketchem, R., Hu, W. & Cross, T. 1993. High-resolution conformation of gramicidin A in a lipid bilayer by solid-state NMR. *Science*, 261(5127): 1457–1460.

Knappe, T.A., Linne, U., Zirah, S., Rebuffat, S., Xie, X. & Marahiel, M.A. 2008. Isolation and Structural Characterization of Capistruin, a Lasso Peptide Predicted from the Genome Sequence of Burkholderia thailandensis E264. *Journal of the American Chemical Society*, 130(34): 11446–11454.

Knerr, P.J. & van der Donk, W.A. 2012. Discovery, biosynthesis, and engineering of lantipeptides. *Annual Review of Biochemistry*, 81: 479–505.

Kokel, D., Bryan, J., Laggner, C., White, R., Cheung, C.Y.J., Mateus, R., Healey, D., Kim, S., Werdich, A. a, Haggarty, S.J., MacRae, C.A., Shoichet, B. & Peterson, R.T. 2010. Rapid behavior—based identification of neuroactive small molecules in the zebrafish. *Nature Chemical Biology*, 6(3): 231–237.

Korbie, D.J. & Mattick, J.S. 2008. Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nature Protocols*, 3(9): 1452–1456.

Kornprobst, J.-M. 2014. *Encyclopedia of Marine Natural Products*. 2nd ed. Weinheim, Germany: Wiley-VCH.

Laport, M.S., Santos, O.C.S. & Muricy, G. 2009. Marine sponges: potential sources of new antimicrobial drugs. *Current Pharmaceutical Biotechnology*, 10(1): 86–105.

Leal, M.C., Puga, J., Serôdio, J., Gomes, N.C.M. & 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*, 7(1): e30580.

Lee, S.W., Mitchell, D. a, Markley, A.L., Hensler, M.E., Gonzalez, D., Wohlrab, A., Dorrestein, P.C., Nizet, V. & Dixon, J.E. 2008. Discovery of a widely distributed toxin biosynthetic gene cluster. *Roceedings of the National Academy of Sciences of the United States of America*, 105(15): 5879–5884.

Lee, Y., Lee, J. & Lee, H. 2001. Microbial symbiosis in marine sponges. *The Journal of Microbiology*, 39(4): 254–264.

Leikoski, N., Fewer, D.P., Jokela, J., Wahlsten, M., Rouhiainen, L. & Sivonen, K. 2010. Highly diverse cyanobactins in strains of the genus Anabaena. *Applied and Environmental Microbiology*, 76(3): 701–709.

Li, J.W.-H. & Vederas, J.C. 2009. Drug discovery and natural products: end of an era or an endless frontier? *Science*, 325(5937): 161–165.

Li, M.H., Ung, P.M., Zajkowski, J., Garneau-Tsodikova, S. & Sherman, D.H. 2009. Automated genome mining for natural products. *BMC Bioinformatics*, 10(1): 185.

Li, X. & Qin, L. 2005. Metagenomics-based drug discovery and marine microbial diversity. *Trends in Biotechnology*, 23(11): 539–543.

Liao, R., Duan, L., Lei, C., Pan, H., Ding, Y., Zhang, Q., Chen, D., Shen, B., Yu, Y. & Liu, W. 2009. Thiopeptide biosynthesis featuring ribosomally synthesized precursor peptides and conserved posttranslational modifications. *Chemistry and Biology*, 16(2): 141–147.

Light, D.W. 2009. Global drug discovery: Europe is ahead. *Health affairs*, 28(5): w969–w977.

Liles, M.R., Williamson, L.L., Rodbumrer, J., Torsvik, V., Goodman, R.M. & Handelsman, J. 2008. Recovery, purification, and cloning of high-molecular-weight DNA from soil microorganisms. *Applied and Environmental Microbiology*, 74(10): 3302–3305.

Maksimov, M.O. & Link, A.J. 2014. Prospecting genomes for lasso peptides. *Journal of Industrial Microbiology and Biotechnology*, 41(2): 333–344.

Maksimov, M.O., Pelczer, I. & Link, a J. 2012. Precursor-centric genome-mining approach for lasso peptide discovery. *Proceedings of the National Academy of Sciences of the United States of America*, 109(38): 15223–15228.

Martín, J., da S. Sousa, T., Crespo, G., Palomo, S., González, I., Tormo, J., de la Cruz, M., Anderson, M., Hill, R., Vicente, F., Genilloud, O. & Reyes, F. 2013. Kocurin, the true structure of PM181104, an anti-methicillin-resistant Staphylococcus aureus (MRSA) thiazolyl peptide from the marine-derived bacterium Kocuria palustris. *Marine Drugs*, 11(2): 387–398

Martins, A., Vieira, H., Gaspar, H. & Santos, S. 2014. Marketed marine natural products in the pharmaceutical and cosmeceutical industries: tips for success. *Marine Drugs*, 12(2): 1066–1101.

Martin-Visscher, L. a, van Belkum, M.J., Garneau-Tsodikova, S., Whittal, R.M., Zheng, J., McMullen, L.M. & Vederas, J.C. 2008. Isolation and characterization of carnocyclin A, a novel circular bacteriocin produced by Carnobacterium maltaromaticum UAL307. *Applied and Environmental Microbiology*, 74(15): 4756–4763.

Matsunaga, S., Fusetani, N. & Konosu, S. 1985. Bioactive marine metabolites, IV. Isolation and the amino acid composition of discodermin A, an antimicrobial peptide, from the marine sponge Discodermia kiiensis. *Journal of Natural Products*, 48(2): 236–41.

Mavromatis, K., Land, M.L., Brettin, T.S., Quest, D.J., Copeland, A., Clum, A., Goodwin, L., Woyke, T., Lapidus, A., Klenk, H.P., Cottingham, R.W. & Kyrpides, N.C. 2012. The fast changing landscape of sequencing technologies and their impact on microbial genome assemblies and annotation. *PLoS ONE*, 7(12): 1–6.

Mayer, A.M.S., Glaser, K.B., Cuevas, C., Jacobs, R.S., Kem, W., Little, R.D., McIntosh, J.M., Newman, D.J., Potts, B.C. & Shuster, D.E. 2010. The odyssey of marine pharmaceuticals: a current pipeline perspective. *Trends in Pharmacological Sciences*, 31(6): 255–265.

McIntosh, J.A., Donia, M.S. & Schmidt, E.W. 2009. Ribosomal peptide natural products: bridging the ribosomal and nonribosomal worlds. *Natural Product Reports*, 26(4): 537–559.

Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. a, Weber, T., Takano, E. & Breitling, R. 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research*, 39: W339–W346.

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*, 10(9): e1003822.

Melby, J.O., Nard, N.J. & Mitchell, D. a. 2011. Thiazole/oxazole-modified microcins: complex natural products from ribosomal templates. *Current Opinion in Chemical Biology*, 15(3): 369–378.

Mendola, D. 2003. Aquaculture of three phyla of marine invertebrates to yield bioactive metabolites: process developments and economics. *Biomolecular Engineering*, 20(4-6): 441–458.

Mohamed, N.M., Colman, A.S., Tal, Y. & Hill, R.T. 2008. Diversity and expression of nitrogen fixation genes in bacterial symbionts of marine sponges. *Environmental Microbiology*, 10(11): 2910–2921.

Mohimani, H., Kersten, R.D., Liu, W.-T., Wang, M., Purvine, S.O., Wu, S., Brewer, H.M., Pasa-Tolic, L., Bandeira, N., Moore, B.S., Pevzner, P. a & Dorrestein, P.C. 2014. Automated genome mining of ribosomal peptide natural products. *ACS Chemical Biology*, 9(7): 1545–1551.

Monteiro, M.C., de la Cruz, M., Cantizani, J., Moreno, C., Tormo, J.R., Mellado, E., De Lucas, J.R., Asensio, F., Valiante, V., Brakhage, A.A., Latge, J.-P., Genilloud, O. & Vicente, F. 2012. A new approach to drug discovery: high-throughput screening of microbial natural extracts against Aspergillus fumigatus using resazurin. *Journal of Biomolecular Screening*, 17(4): 542–549.

Mori, T., Kokubo, H., Shimizu, H., Iwamoto, M., Oiki, S. & Okamoto, Y. 2007. Normal mode analysis of polytheonamide B. *Journal of the Physical Society of Japan*, 76(9): 094801.

Morinaka, B.I., Vagstad, A.L., Helf, M.J., Gugger, M., Kegler, C., Freeman, M.F., Bode, H.B. & Piel, J. 2014. Radical S -adenosyl methionine epimerases: regioselective introduction of diverse D -amino acid patterns into peptide natural products. *Angewandte Chemie*, 126(32): 8643–8647.

Morris, R.P., Leeds, J.A., Naegeli, H.U., Oberer, L., Memmert, K., Weber, E., LaMarche, M.J., Parker, C.N., Burrer, N., Esterow, S., Hein, A.E., Schmitt, E.K. & Krastel, P. 2009. Ribosomally synthesized thiopeptide antibiotics targeting elongation factor Tu. *Journal of the American Chemical Society*, 131(16): 5946–5955.

Motti, C. 2012. Environmental marine metabolomics: from whole organism system biology to ecosystem management. Journal of Marine Science Research and Development, 2(3): 2–3.

Murakami, M., Sun, Q., Ishida, K. & Matsuda, H. 1997. Microviridins, elastase inhibitors from the cyanobacterium Nostoc minutum(NIES-26). *Phytochemistry*, 45(6): 1197–1202.

Murphy, K., O'Sullivan, O., Rea, M.C., Cotter, P.D., Ross, R.P. & Hill, C. 2011. Genome mining for radical SAM protein determinants reveals multiple sactibiotic-like gene clusters. *PLoS ONE*, 6(7): e20852.

Muscholl-Silberhorn, A., Thiel, V. & Imhoff, J.F. 2008. Abundance and bioactivity of cultured sponge-associated bacteria from the Mediterranean sea. *Microbial Ecology*, 55(1): 94–106.

Nakao, Y., Matsunaga, S. & Fusetani, N. 1995. Three more cyclotheonamides, C, D, and E, potent thrombin inhibitors from the marine sponge Theonella swinhoei. *Bioorganic and Medicinal Chemistry*, 3(8): 1115–1122.

Newman, D. & Cragg, G. 2012. Natural products as sources of new drugs over the 30 years. *Journal of Natural Products*, 75(3): 311–335.

Newman, D.J. & Cragg, G.M. 2007. Natural products as sources of new drugs over the last 25 years. *Journal of Natural Products*, 70(3): 461–477.

Newman, M.M., Feminella, J.W. & Liles, M.R. 2010. Purification of genomic DNA extracted from environmental sources for use in a polymerase chain reaction. Cold Spring Harbor Protocols, 2010(2): 5383–5383.

O' Halloran, J.A., Barbosa, T.M., Morrissey, J.P., Kennedy, J., O' Gara, F. & Dobson, A.D.W. 2011. Diversity and antimicrobial activity of Pseudovibrio spp. from Irish marine sponges. *Journal of Applied Microbiology*, 110(6): 1495–1508.

Oman, T.J. & van der Donk, W.A. 2010. Follow the leader: the use of leader peptides to guide natural product biosynthesis. *Nature Chemical Biology*, 6(1): 9–18.

Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. & de Jong, P.J. 1998. An improved approach for construction of bacterial artificial chromosome libraries. *Genomics*, 52(1): 1–8.

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

Owczarzy, R., Tataurov, A. V, Wu, Y., Manthey, J.A., McQuisten, K.A., Almabrazi, H.G., Pedersen, K.F., Lin, Y., Garretson, J., McEntaggart, N.O., Sailor, C.A., Dawson, R.B. & Peek, A.S. 2008. IDT SciTools: a suite for analysis and design of nucleic acid oligomers. *Nucleic Acids Research*, 36(Web Server): W163–W169.

Öztürk, B., de Jaeger, L., Smidt, H. & Sipkema, D. 2013. Culture-dependent and independent approaches for identifying novel halogenases encoded by Crambe crambe (marine sponge) microbiota. *Scientific Reports*, 3: 2780.

Papapetropoulou, M. & Rodopoulou, G. 1994. Occurrence of enteric and non-enteric indicators in coastal waters of southern Greece. *Bulletin of Marine Science*, 54(1): 63–70.

Parker-Nance, S. 2013. Porifera and Ascidiacea of Algoa Bay, South Africa. Internal Nelson Mandela Metropolitan University report. Unpublished.

Patton, G.C., Paul, M., Cooper, L.E., Chatterjee, C. & van der Donk, W.A. 2008. The importance of the leader sequence for directing lanthionine formation in lacticin 481. *Biochemistry*, 47(28): 7342–7351.

Pettit, G.R., Herald, C.L. & Hogan, F. 2002. Biosynthetic products for anticancer drug design and treatment: the bryostatins. In B. C. Baguley & D. J. Kerr, eds. *Anticancer Drug Development*. Elsevier Inc.: 203–235.

Piel, J. 2011. Approaches to capturing and designing biologically active small molecules produced by uncultured microbes. *Annual Review of Microbiology*, 65(1): 431–453.

Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N. & Matsunaga, S. 2004. Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge Theonella swinhoei. *Proceedings of the National Academy of Sciences of the United States of America*, 101(46): 16222–16227.

Pietra, F. 1990. *A Secret World Natural Products of Marine Life*. F. Pietra, ed. Basel: Birkhäuser.

Poptsova, M.S. & Gogarten, J.P. 2010. Using comparative genome analysis to identify problems in annotated microbial genomes. *Microbiology*, 156(7): 1909–1917.

Proksch, P., Edrada, R.A. & Ebel, R. 2002. Drugs from the seas - current status and microbiological implications. *Applied Microbiology and Biotechnology*, 59(2-3): 125–134.

Qu, X., Lei, C. & Liu, W. 2011. Transcriptome mining of active biosynthetic pathways and their associated products in Streptomyces flaveolus. *Angewandte Chemie International Edition*, 50(41): 9651–9654.

Querellou, J., Børresen, T., Boyen, C. & Dobson, A. 2010. *Marine Biotechnology: A New Vision and Strategy for Europe*. The Marine Board of the European Science Foundation: Belgium.

Reysenbach, A.-L. & Pace, N.R. 1995. Reliable amplification of hyperthermophilic archaeal 16S rRNA genes by the polymerase chain reaction. In F. T. Robb & A. R. Place, eds. *Archaea: a laboratory manual – Thermophiles.* New York: Cold Spring Harbor: 101–105.

Ridley, C.P., Bergquist, P.R., Harper, M.K., Faulkner, D.J., Hooper, J.N.A. & Haygood, M.G. 2005. Speciation and biosynthetic variation in four dictyoceratid sponges and their cyanobacterial symbiont, Oscillatoria spongeliae. *Chemistry and Biology*, 12(3): 397–406.

Ridley, C.P. & Khosla, C. 2009. Polyketides. In M. Schaechter, ed. *Encyclopedia of Microbiology*. Oxford: Academic Press: 472–481.

Ridley, M.S.O. & Dendy, A. 1886. Preliminary report on the Monaxonida collected by H.M.S. 'Challenger'. *The annals and magazine of natural history*, 18(107): 326–493.

Rippka, R., Deruelles, J., Waterbury, J.B., Herdman, M. & Stanier, R.Y. 1979. Generic assignments, strain histories and properties of pure cultures of Cyanobacteria. *Journal of General Microbiology*, 111(1): 1–61.

Robinson, J.A. 1991. Polyketide synthase complexes: their structure and function in antibiotic biosynthesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 332(1263): 107–114.

Rodríguez-Marconi, S., De La Iglesia, R., Díez, B., Fonseca, C.A., Hajdu, E., Trefault, N. & Webster, N. 2015. Characterization of bacterial, archaeal and eukaryote symbionts from antarctic sponges reveals a high diversity at a three-domain level and a particular signature for this ecosystem. *PLoS ONE*, 10(9): 1–19.

Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R., Loiacono, K.A., Lynch, B.A., MacNeil, I.A., Minor, C., Tiong, C.L., Gilman, M., Osburne, M.S., Clardy, J., Handelsman, J. & Goodman, R.M. 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied and Environmental Microbiology*, 66(6): 2541–2547.

Rottig, 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*, 39(Web Server issue): W362–W367.

Salgado, A., Vieiralves, T., Lamarão, F.R.M., Assumpção, L.L.M., Gomes, D., Jascone, L., Valadão, A.L., Albano, R.M. & Lôbo-Hajdu, G. 2007. Field preservation and optimization of a DNA extraction method for Porifera. In M. R. Custodio, G. Lôbo-Hajdu, E. Hajdu, & G. Muricy, eds. *Porifera Research: Biodiversity, Innovation and Sustainability*. Rio de Janeiro: Museu Nacional: 555–560.

Santos, O.C.S., Pontes, P.V.M.L., Santos, J.F.M., Muricy, G., Giambiagi-deMarval, M. & Laport, M.S. 2010. Isolation, characterization and phylogeny of sponge-associated bacteria with antimicrobial activities from Brazil. *Research in Microbiology*, 161(7): 604–612.

Schleper, C., DeLong, E.F., Preston, C.M., Feldman, R.A., Wu, K.Y. & Swanson, R. V. 1998. Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon Cenarchaeum symbiosum. *Journal of Bacteriology*, 180: 5003–5009.

Schmidt, E.W. 2010. The hidden diversity of ribosomal peptide natural products. *BMC Biology*, 8(1): 83.

Schmidt, E.W. & Donia, M.S. 2009. Cyanobactin Ribosomally Synthesized Peptides—A Case of Deep Metagenome Mining. In Schmidt, E.W. & Donia, M.S. 2009. Cyanobactin ribosomally synthesized peptides—a case of deep metagenome mining. In D. A. Hopwood, ed. *Methods in enzymology*. Complex Enzymes in Microbial Natural Product Biosynthesis, Part A: Overview Articles and Peptides. Academic Press: 575–596.

Schmidt, E.W., Nelson, J.T., Rasko, D.A., Sudek, S., Eisen, J.A., Haygood, M.G. & Ravel, J. 2005. Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. *Proceedings of the National Academy of Sciences of the United States of America*, 102(20): 7315–7320.

Schmidt, E.W., Obraztsova, A.Y., Davidson, S.K., Faulkner, D.J. & Haygood, M.G. 2000. Identification of the antifungal peptide-containing symbiont of the marine sponge Theonella swinhoei as a novel δ-proteobacterium, ‘ Candidatus Entotheonella palauensis’. *Marine Biology*, 136(6): 969–977.

Schmitt, S., Tsai, P., Bell, J., Fromont, J., Ilan, M., Lindquist, N., Perez, T., Rodrigo, A., Schupp, P.J., Vacelet, J., Webster, N., Hentschel, U. & Taylor, M.W. 2012. Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *The ISME Journal*, 6(3): 564–576.

Schrader, C., Schielke, A., Ellerbroek, L. & Johne, R. 2012. PCR inhibitors - occurrence, properties and removal. *Journal of Applied Microbiology*, 113(5): 1014–1026.

Shen, Layer & McCabe. 2000. Conopeptides: from deadly venoms to novel therapeutics. *Drug Discovery Today*, 5(3): 98–106.

Shimamura, H., Gouda, H., Nagai, K., Hirose, T., Ichioka, M., Furuya, Y., Kobayashi, Y., Hirono, S., Sunazuka, T. & Ōmura, S. 2009. Structure determination and total synthesis of bottromycin A2 : a potent antibiotic against MRSA and VRE. *Angewandte Chemie International Edition*, 48(5): 914–917.

Silakowski, B., Schairer, H.U., Ehret, H., Kunze, B., Weinig, S., Nordsiek, G., Brandt, P., Blo, H., Ho, G., Beyer, S. & Mu, R. 1999. New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of Stigmatella aurantiaca DW4/3-1. *The Journal of Bological Chemistry*, 274(52): 37391–37399.

Silver, L.L. 2011. Challenges of antibacterial discovery. *Clinical Microbiology Reviews*, 24(1): 71–109.

Simister, R.L., Schmitt, S. & Taylor, M.W. 2011. Evaluating methods for the preservation and extraction of DNA and RNA for analysis of microbial communities in marine sponges. *Journal of Experimental Marine Biology and Ecology*, 397(1): 38–43.

Simmons, T.L., Andrianasolo, E., McPhail, K., Flatt, P. & Gerwick, W.H. 2005. Marine natural products as anticancer drugs. *Molecular Cancer Therapeutics*, 4(2): 333–342.

Simon, C. & Daniel, R. 2011. Metagenomic analyses: past and future trends. *Applied and Environmental Microbiology*, 77(4): 1153–1161.

Sipkema, D., Franssen, M.C.R., Osinga, R., Tramper, J. & Wijffels, R.H. 2005. Marine sponges as pharmacy. *Marine Biotechnology*, 7(3): 142–162.

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*, 90(2): 201–222.

Sørensen, J.L. & Sondergaard, T.E. 2014. The effects of different yeast extracts on secondary metabolite production in Fusarium. *International Journal of Food Microbiology*, 170: 55–60.

Stackebrandt, E. & Goebel, B.M. 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic Bacteriology*, 44(4): 846–849.

Steele, H.L. & Streit, W.R. 2005. Metagenomics: advances in ecology and biotechnology. *FEMS Microbiology Letters*, 247(2): 105–111.

Sudek, S., Lopanik, N.B., Waggoner, L.E., Hildebrand, M., Anderson, C., Liu, H., Patel, A., Sherman, D.H. & Haygood, M.G. 2007. Identification of the putative bryostatin polyketide synthase gene cluster from ‘Candidatus Endobugula sertula’, the uncultivated microbial symbiont of the marine bryozoan Bugula neritina. *Journal of Natural Products*, 70(1): 67–74.

Suenaga, H. 2012. Targeted metagenomics: a high-resolution metagenomics approach for specific gene clusters in complex microbial communities. *Environmental Microbiology*, 14(1): 13–22.

Svenson, J. 2013. MabCent: Arctic marine bioprospecting in Norway. *Phytochemistry Reviews*, 12(3): 567–578.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, 28(10): 2731–2739.

Taylor, M.W., Hill, R.T., Piel, J., Thacker, R.W. & Hentschel, U. 2007. Soaking it up: the complex lives of marine sponges and their microbial associates. *The ISME Journal*, 1(3): 187–190.

Taylor, M.W., Radax, R., Steger, D. & Wagner, M. 2007. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiology and Molecular Biology Reviews*, 71(2): 295–347.

Thompson, J.J.D., Higgins, D.G.D. & Gibson, T.T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix. *Nucleic Acids Research*, 22(22): 4673–4680.

Udwary, D.W., Zeigler, L., Asolkar, R.N., Singan, V., Lapidus, a., Fenical, W., Jensen, P.R. & Moore, B.S. 2007. Genome sequencing reveals complex secondary metabolome in the marine actinomycete Salinispora tropica. *Proceedings of the National Academy of Sciences of the United States of America*, 104(25): 10376–10381.

van Belkum, M.J., Martin-Visscher, L. a & Vederas, J.C. 2011. Structure and genetics of circular bacteriocins. *Trends in Microbiology*, 19(8): 411–418.

van Heel, A.J., de Jong, A., Montalban-Lopez, M., Kok, J. & Kuipers, O.P. 2013. BAGEL3: automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Research*, 41(W1): W448–W453.

Van Middlesworth, F. & Cannell, R.J.P. 2005. *Methods in Biotechnology, Vol. 20, Natural Products Isolation*. 2nd ed. S. D. Sarker, Z. Latif, & A. I. Gray, eds. Totowa, NJ: Humana Press.

Waisvisz, J.M., van der Hoeven, M.G., van Peppen, J. & Zwennis, W.C.M. 1957. Bottromycin. I. A new sulfur-containing antibiotic. *Journal of the American Chemical Society*, 79(16): 4520–4521.

Walmsley, T.A., Matcher, G.F., Zhang, F., Hill, R.T., Davies-Coleman, M.T. & Dorrington, R. a. 2012. Diversity of bacterial communities associated with the Indian Ocean sponge Tsitsikamma favus that contains the bioactive pyrroloiminoquinones, tsitsikammamine A and B. *Marine Biotechnology*, 14(6): 681–691.

Walsh, C. 2000. Molecular mechanisms that confer antibacterial drug resistance. *Nature*, 406(6797): 775–781.

Walsh, C.T. 2004. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science*, 303(5665): 1805–1810.

Walsh, C.T. & Fischbach, M.A. 2010. Natural products version 2.0: connecting genes to molecules. *Journal of the American Chemical Society*, 132(8): 2469–2493.

Wang, H. & van der Donk, W.A. 2011. Substrate Selectivity of the Sublancin S-Glycosyltransferase. *Journal of the American Chemical Society*, 133(41): 16394–16397.

Wang, H., Fewer, D.P. & Sivonen, K. 2011. Genome Mining Demonstrates the Widespread Occurrence of Gene Clusters Encoding Bacteriocins in Cyanobacteria. *PLoS ONE*, 6(7): 1–10.

Wang, J., Yu, Y.Y., Tang, K., Liu, W., He, X., Huang, X.X. & Deng, Z. 2010. Identification and analysis of the biosynthetic gene cluster encoding the thiopeptide antibiotic cyclothiazomycin in Streptomyces hygroscopicus 10-22. *Applied and Environmental Microbiology*, 76(7): 2335–2344.

Ward, D.M., Weller, R. & Bateson, M.M. 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature*, 345(6270): 63–65.

Weber, T. 2014. In silico tools for the analysis of antibiotic biosynthetic pathways. *International Journal of Medical Microbiology*, 304(3-4): 230–235.

Weber, T., Rausch, C., Lopez, P., Hoof, I., Gaykova, V., Huson, D.H. & Wohlleben, W. 2009. CLUSEAN: a computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters. *Journal of Biotechnology*, 140(1-2): 13–17.

Wei, M., Deng, J., Wang, S., Liu, N. & Chen, Y. 2011. A simple reverse genetics approach to elucidating the biosynthetic pathway of nocathiacin. *Biotechnology Letters*, 33(3): 585–591.

Weiz, A.R., Ishida, K., Makower, K., Ziemert, N., Hertweck, C. & Dittmann, E. 2011. Leader peptide and a membrane protein scaffold guide the biosynthesis of the tricyclic peptide microviridin. *Chemistry and Biology*, 18(11): 1413–1421.

Wender, P.A., De Brabander, J., Harran, P.G., Jimenez, J.-M., Koehler, M.F.T., Lippa, B., Park, C.-M. & Shiozaki, M. 1998. Synthesis of the First Members of a New Class of Biologically Active Bryostatin Analogues. *Journal of the American Chemical Society*, 120(18): 4534–4535.

Więckowicz, M., Schmidt, M., Sip, A. & Grajek, W. 2011. Development of a PCR-based assay for rapid detection of class IIa bacteriocin genes. *Letters in Applied Microbiology*, 52(3): 281–289.

Wieland Brown, L.C., Acker, M.G., Clardy, J., Walsh, C.T. & Fischbach, M. a. 2009. Thirteen posttranslational modifications convert a 14-residue peptide into the antibiotic thiocillin. *Proceedings of the National Academy of Sciences of the United States of America*, 106(8): 2549–2553.

Wilson, I.G. 1997. Inhibition and facilitation of nucleic acid amplification. *Applied and Environmental Microbiology*, 63(10): 3741–3751.

Wilson, K. 2001. Preparation of Genomic DNA from Bacteria. In *Current Protocols in Molecular Biology*. Hoboken, USA: John Wiley & Sons, Inc.: 2.4.1–2.4.5.

Wilson, M.C., Mori, T., Rückert, C., Uria, A.R., Helf, M.J., Takada, K., Gernert, C., Steffens, U.A.E., Heycke, N., Schmitt, S., Rinke, C., Helfrich, E.J.N., Brachmann, A.O., Gurgui, C., Wakimoto, T., Kracht, M., Crüsemann, M., Hentschel, U., Abe, I., Matsunaga, S., Kalinowski, J., Takeyama, H. & Piel, J. 2014. An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature*, 506(7486): 58–62.

Wintzingerode, F. V., Göbel, U.B. & Stackebrandt, E. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews*, 21(3): 213–229.

Woodhouse, J.N., Fan, L., Brown, M. V, Thomas, T. & Neilan, B.A. 2013. Deep sequencing of non-ribosomal peptide synthetases and polyketide synthases from the microbiomes of Australian marine sponges. *The ISME Journal*, 7(9): 1842–1851.

Wright, A.E., Forleo, D.A., Gunawardana, G.P., Gunasekera, S.P., Koehn, F.E. & McConnell, O.J. 1990. Antitumor tetrahydroisoquinoline alkaloids from the colonial ascidian Ecteinascidia turbinata. *The Journal of Organic Chemistry*, 55(15): 4508–4512.

Yang, X. & van der Donk, W.A. 2013. Ribosomally synthesized and post-translationally modified peptide natural products: new insights into the role of leader and core peptides during biosynthesis. Chemistry - A European Journal, 19(24): 7662–7677.

Yoon, V. & Nodwell, J.R. 2014. Activating secondary metabolism with stress and chemicals. *Journal of the Society for Industrial Microbiology and Biotechnology*, 41(2): 415–424.

Zhang, Q., Yang, X., Wang, H. & Van Der Donk, W.A. 2014. High divergence of the precursor peptides in combinatorial lanthipeptide biosynthesis. *ACS Chemical Biology*, 9(11): 2686–2694.

Zhang, Q., Yu, Y., Vélasquez, J.E. & van der Donk, W.A. 2012. Evolution of lanthipeptide synthetases. *Proceedings of the National Academy of Sciences* of the United States of America, 109(45): 18361–18366.

Ziemert, N., Ishida, K., Liaimer, A., Hertweck, C. & Dittmann, E. 2008. Ribosomal synthesis of tricyclic depsipeptides in bloom-forming cyanobacteria. *Angewandte Chemie International Edition*, 47(40): 7756–4459.

Ziemert, N., Ishida, K., Weiz, A., Hertweck, C. & Dittmann, E. 2010. Exploiting the natural diversity of microviridin gene clusters for discovery of novel tricyclic depsipeptides. *Applied and Environmental Microbiology*, 76(11): 3568–3574.