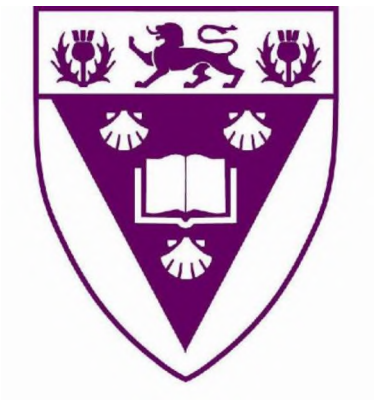
AN INVESTIGATION INTO THE BACTERIAL BIOSYNTHETIC ORIGINS OF BIOACTIVE NATURAL PRODUCTS ISOLATED FROM SOUTH AFRICAN LATRUNCULID SPONGES

A thesis submitted in fulfilment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY
of
RHODES UNIVERSITY



By
SAMANTHA CHE WATERWORTH
December 2017

Table of Contents

List of Figures v
List of Tables x
List of Abbreviations x
Acknowledgments xii
Abstract xv

Chapter 1 - Literature Review 1
1.1. Introduction 1
1.2. Bioactive Secondary Metabolites from Sponges 2
1.3. Sponge-associated Microbial Communities 4
1.4. Acquisition of Sponge-associated Bacteria 5
1.5. Acquisition of Biosynthetic Gene Clusters in Marine Sponge-associated Bacteria 7
1.6. Culture-dependent methods for isolation of bioactive secondary metabolites 8
1.7. Culture-independent methods for identification of bioactive secondary metabolites 11
1.8. Microbial biosynthesis of pyrroloquinoline alkaloids 13
1.8.1. Lymphostin 13
1.8.2. Ammosamides 15
1.9. Pyrroloiminoquinones isolated from Latrunculiidae sponges 16
1.9.1. Pyrroloiminoquinones isolated from sponges within the genus *Tsitsikamma* 17
1.9.2. Pyrroloiminoquinones isolated from sponges within the genus *Latrunculia, Cyclacanthia* and other Latrunculiidae genera 19
1.10. Pyrroloiminoquinones isolated from other organisms 21
1.11. Research Problem Identification 22

Chapter 2 – Identification of gene clusters producing antimicrobials in *Kocuria kristinae* isolated from *Tsitsikamma favus* 24
2.1. Introduction 24
2.2. Methods and Materials 26
2.2.1. Sponge collection and identification 26
2.2.2. Isolation of sponge-associated actinomycetes with antimicrobial activity 27

2.2.3. Bacterial culture conditions 27
2.2.4. Whole genome shotgun sequencing 28
2.2.5. Identification of biosynthetic gene clusters 28
2.2.6. Anti-microbial and siderophore activity assays 29
2.3. Results 30
2.3.1. Confirmation and extension of observed anti-microbial activity. 30
2.3.2. Genome sequence analysis of *K. kristinae* isolates RUTW2-3 and RUTW4-5 30
2.3.3. Characterization of the putative terpene biosynthetic gene cluster 31
2.3.4. Characterisation of a putative siderophore biosynthetic gene cluster 32
2.4. Discussion 35

Chapter 3 - Identification of candidate bacterial taxa as biosynthetic origins of tsitsikammamine production 38
3.1. Introduction 38
3.2. Methods and Materials 39
3.2.1. Sponge Collection 39
3.2.2. Extraction of Genomic DNA 40
3.2.3. Identification of Sponge Specimens 40
3.2.4. 454 Amplicon library preparation and pyrosequencing 41
3.2.5. 454 amplicon sequence data curations 42
3.2.6. Illumina amplicon preparation and sequencing 42
3.2.7. Illumina amplicon library data curation 42
3.3. Results 43
3.3.1. Sponge Collection and Identification 43
3.3.2. Characterization of Microbial Communities Associated with Sponges within the Family Latrunculiidae 45
3.3.3. Characterization of associated bacterial taxa across a broad range of sponge species 53
3.4. Discussion 56

Chapter 4 - Metagenomic Analysis of the *T. favus* Microbiome 60
4.1. Introduction 60
4.2. Methods and Materials 62
4.2.1. Metagenome sequencing of *T. favus* TIC2014-001 62

4.2.2. Metagenome sequencing of *T. favus* specimen TIC2016 - 050A 64
4.3. Results 65
4.3.1. Assessment of bacterial enrichment protocol 66
4.3.2. Phylogenetic analysis of *T. favus* metagenomic sequence data 67
4.3.3. Identification of potential gene function within *T. favus* metagenomic sequence data 69
4.3.4. Clustering of contigs into putative genome bins 70
4.4. Discussion 77
4.4.1. Pilot metagenomic study 77
4.4.2. Comparison of approaches for bacterial community assessment 79
4.4.3. Analysis of metabolic potential in *T. favus*-associated microbiome 80

Chapter 5 – Genome characterisation of a *T. favus*-associated Spirochaete species 83
5.1. Introduction 83
5.2. Methods and Materials 84
5.3. Results 85
5.3.1. Cluster A: Biosynthesis of a fatty acid 88
5.3.2. Cluster B: Putative O-antigen biosynthetic gene cluster 89
5.3.3. Cluster C: Potential production of capsular polysaccharide 90
5.4. Discussion 91

Chapter 6 - Secondary metabolite gene clusters detected in the *T. favus* metagenome 98
6.1. Introduction 98
6.2. Methods and Materials 101
6.3. Results 102
6.3.1. Characterization of Cluster 1 102
6.3.2. Characterization of Cluster 2 104
6.3.3. Characterization of Cluster 3 and Cluster 4 105
6.3.4. Characterization of Cluster 5 108
6.3.5. Characterization of Cluster 6 109
6.4. Discussion 112

Chapter 7 – General Discussion and Future Prospects 119
7.1. General Discussion 119
7.1.1. Conservation of sponge-associated microbiota 120
7.1.2. Potential for bioactive secondary metabolite production in *T. favus*-associated microbiota 123
7.2. Future Prospects 125
7.2.1. Culturing of Sponge-associated actinomycetes 125
7.2.2. Metagenomic analysis of *Tsitsikamma pedunculata* sponges 125
7.3. Concluding remarks 126
References 127
Appendix A – Sponge Collection Metadata 155
Appendix B – Media used 157
Appendix C – Publications 158

List of Figures

Figure 1.1: Chemical structure of namenamicin (**A**) and lomaiviticin (**B**) from marine *Streptomyces* sp. Structures were adapted from McDonald et al., 1996; Janso et al., 2014 and drawn using ChemSpider 10

Figure 1.2: Chemical structures of immunosuppressant lymphostin produced by *Streptomyces* sp. KY11783. Chemical structures were adapted from Aotani et al., 1997 and drawn using ChemSpider. 13

Figure 1.3: Schematic representation of the lymphostin (top) and ammosamide (bottom) biosynthetic gene clusters in *Salinispora* and *Streptomyces* bacteria respectively modified from Jordan and Moore, 2016. Genes are represented by coloured arrows orientated in the direction of transcription and are not drawn to scale. A coloured key is provided for functional annotation of genes. The conserved genes are highlighted with a striped orange region. ORFs annotated with an asterisk indicate genes vital to ammosamide production. Genes *amm*6 and *amm*7, which are hypothesised to form the pyrroloquinoline core are indicated with the appropriate number below. Abbreviations: RiPP: ribosomally synthesized, post-translationally modified peptide, NRPS: non-ribosomal peptide synthase, PKS: polyketide synthase. Modules *lym*A and *lym*B are situated within the hybrid NRPS/PKS encoding gene (lilac). 14

Figure 1.4: Chemical structure of ammosamide A (**A**) and ammosamide B (**B**) produced by *Streptomyces* sp. CNR-698. Chemical structures were adapted from Jordan and Moore, 2016 and drawn using ChemSpider. 15

Figure 1.5: Chemical structure of the pyrroloiminoquinone core (**A**), tsitsikammamines (**B**) and makaluvamines (**C**) Chemical structures were adapted from Antunes et al., 2004 and Antunes et al., 2005 and drawn using ChemSpider 17

Figure 1.6: Chemical structure of tsitsikammamine A (**A**) and tsitsikammamine B (**B**) isolated from *T. favus*, tsitsikammamine C (**C**) isolated from *Zyzzya* sp. and wakayin (**D**) isolated from a *Clavelina* tunicate. Chemical structures were adapted from Copp et al., 1991, Antunes et al., 2005 and Davies et al, 2012 and drawn with ChemSpider 18

Figure 1.7: Chemical structures of discorhabdin A (**A**) and discorhabdin N (**B**) isolated from South African *Latrunculia algoaensis*, makaluvamine A (**C**) and 1-methoxydiscorhabdin D (**D**) isolated from *Cyclacanthia bellae*, (−)-3-dihydrodiscorhabdin D (**E**) and (−)-discorhabdin Z (**F**) isolated from a Korean *Sceptrella* sponge, makaluvic acid C (**G**) N-1-β-d-ribofuranosylmakaluvic acid C (**H**) from *Strongylodesma aliwaliensis*. Chemical structures were adapted from Keyzers et al., 2004, Keyzers et al, 2005, Antunes et al., 2005 and Jeon et al., 2010 and drawn using ChemSpider. 20

Figure 2.2: Phylogeny of marine *K. kristinae* isolates relative to the Micrococcaceae family and rooted within the order Actinomycetales based on alignment of 16S rRNA gene sequence. Clustering of species within genera, family and order are indicated. Isolates identified in this study are boxed in red. The 16S rRNA gene sequences were aligned using ClustalW in MEGA6 using the Maximum Likelihood method with 500 bootstrap replicates. Bootstrap values are indicated on branch nodes. The scale bar indicates the average number of amino acid substitutions per site 31

Figure 2.3: Comparative gene cluster organization of putative terpene gene cluster identified in *K. kristinae* isolates RUTW2-3 and RUTW4-5 relative to previously identified homologous clusters. Transcriptional orientation of ORFs is represented by arrows coloured by putative function. Regions of homologous sequences spanning multiple aligned contigs that correspond to identified ORFs in RUTW4-5 of the proposed terpene pathway are represented by arrows described with dashed lines. The contig(s) on which each ORF is located, has been indicated. A scale bar is provided for relative size. Genbank Accession No.: Brachybacterium phenoliresistens: NZ_JDYK00000000.1; Brevibacterium linens: NZ_JTJZ00000000.1; Brevibacterium casei: NZ_AMSP00000000.1 32

Figure 2.4: Comparative gene cluster organization of putative NRPS gene clusters identified in *K. kristinae* isolates RUTW2-3 and RUTW4-5 relative to previously identified homologous clusters. Transcriptional orientation of ORFs is represented by arrows coloured by putative function. The contig on which each ORF is located, has been indicated. Regions of homologous sequences spanning multiple aligned contigs that correspond to identified ORFs in RUTW4-5 of the proposed terpene pathway are represented by arrows described with dashed lines. Scale bars are provided for relative contig and ORF size..... 33

Figure 2.5: Siderophore production in marine *K. kristinae* isolates. (**A**) The predicted structure generated in antiSMASH for the identified NRPS gene cluster following collinearity rules without tailoring. (**B**) Structure of enterobactin. The region matched to the predicted NRSP product has been indicated with red circle. (**C**) Siderophore activity observed from RUTW2-3 in CAS-agar assay. Blue agar indicates media in which the CAS dye is complexed with iron and pink regions indicate regions of complexed iron uptake by siderophore activity..... 34

Figure 2.6: Conservation of key gene fragments in identified biosynthetic clusters in *K. kristinae* isolates. PCR amplification products of the free-standing condensation domain of NRPS1 (1307 bp), the condensation domain of gene "F" of NRPS2 (930 bp) and a fragment of the phytoene synthase gene of the terpene (TER) pathway (426 bp) from four *K. kristinae* isolates and a sponge-associated Micrococcus sp.. Controls included as indicated (16S rRNA gene). 35

Figure 3.1: Phylogenetic relationship of latrunculid and related sponges based on alignment of 28S rRNA gene fragment sequences (627 nts) aligned using ClustalW and constructed using the maximum-likelihood method with 1500 bootstrap replicates in MEGA7. The scale bar represents an estimated 0.2% sequence substitution. Genbank accession numbers have been included where applicable. Several of the 28S rRNA gene sequences used in this analysis were generated by Dr. Tara Walmsley (Vaal University of Technology) and Ms Storm Hilliar (MSc Student, Marine Natural Products Laboratory, Rhodes University) 44

Figure 3.2: Phylogenetic classification of latrunculid sponge-associated bacterial communities using Naïve Bayesian classification with the Silva bacterial database (version 128) as the reference. A coloured key indicates phylogenetic classifications. Sponges have been grouped by taxa which is indicated by the label above each box with the following abbreviated names: *T. favus*: *Tsitsikamma favus*, *T. ng*: *Tsitsikamma nguni*, *T. mi*: *Tsitsikamma michaeli*, *T.* 5: *Tsitsikamma* sp. 005, *C. be*: *Cyclacanthia bellae*, *L. a*: *Latrunculia algoaensis*, *Myc*: *Mycale* sp., *T. rubra*: *Tethya rubra*..... 45

Figure 3.3: Statistical oordination of bacterial OTUs (clustered at distance of 0.03) in sponges using non-metric multidimensional scaling. Square root transformation was applied to abundance data prior to construction of Bray-Curtis resemblance matrices based on partial 16S rRNA gene sequences. A colour key is provided to distinguish sample types. 46

Figure 3.4: Diversity of dominant bacterial taxa associated with marine sponges from Algoa Bay and Rheeders Reef, South Africa, and the surrounding environment. Bacterial OTUs were clustered at a distance of 0.03 as an approximation to bacterial species. Relative abundance of bacterial reads was calculated as a percentage of the number of reads per OTU relative to the total number of reads per sample. Dominant bacterial OTUs are defined as having a relative abundance greater than 0.5%. OTUs discussed are indicated by a coloured key. 47

Figure 3.5: Phylogenetic relationship of betaproteobacteria identified in sponges collected from Algoa Bay and Rheeders Reef, South Africa, relative to other related, dominant betaproteobacteria associated with other sponge specimens. A 434bp region of the 16S rRNA gene was aligned using ClustalW and phylogeny was inferred using the Maximum-likelihood method in MEGA7 with 1000 bootstrap replicates. The Tamura-Nei substitution model was used with uniform rates across all sites with a strong branch swap filter. Boot strap values are indicated on branch nodes and a scale bar is provided which indicates the average number of nucleotide substitutions per site. 49

Figure 3.6: Diversity of dominant spirochetes associated with marine sponges from Algoa Bay and Rheeders Reef, South Africa, and the surrounding environment. Relative abundance of bacterial reads was calculated as a percentage of the number of reads per OTU relative to the total number of reads per sample. . Dominant bacterial OTUs are defined as having a relative abundance greater than 0.5%. OTUs discussed are indicated by a coloured key. 50

Figure 3.7: Phylogenetic diversity of dominant spirochetes associated with marine sponges from Algoa Bay and Rheeders Reef, South Africa. Dominant OTUs classified as Spirochetes are differentiated from general bacterial OTUs with a "S_" prefix. 51

Figure 3.8: Numerically and metabolically dominant bacteria identified in *T. favus* sponge specimens TIC2014-001 and TIC2012-057. Relative abundance of bacterial reads was calculated as a percentage of the number of reads per OTU relative to the total number of reads per sample. Dominant bacterial OTUs are defined as having a relative abundance greater than 0.5%. Classification of OTUs discussed here are identified as per Fig. 3.4 viz. a betaproteobacterium (OTU1) and three closely related spirochete strains (OTUs 2, 3 and 4). 52

Figure 3.9: Phylogenetic classification of latrunculid sponge-associated bacterial communities using Naïve Bayesian classification with the Silva bacterial database (version 128) as the reference. Phylogenetic classifications are indicated by a coloured key. Sponges have been grouped by taxa which is indicated by the label above each box with the following abbreviated names: *T. favus*: *Tsitsikamma favus*, *T. nguni*: *Tsitsikamma nguni*, *T. mich*, *Tsitsikamma michaeli*, *L. apic*: *Latrunculia apicalis*, *L**: *Latrunculia algoaensis*, *C. bellae*, *Cyclacanthia bellae*. Seawater amplicon sequence libraries were prepared and provided by Ms Danielle Vos and Ms Storm Hilliar. 54

Figure 3.10: Diversity of dominant bacterial species associated with marine sponges from Algoa Bay and Rheeders Reef (South Africa), Bouvet Island (Sub-Antarctica) and seawater collected from Algoa Bay and Rheeders Reef. Bacterial OTUs were clustered at a distance of 0.03 as an approximation to bacterial species. Relative abundance of bacterial reads was calculated as a percentage of the number of reads per OTU relative to the total number of reads per sample. Dominant bacterial OTUs are defined as having a relative abundance greater than 0.5%. OTUs discussed are indicated by a coloured key. 55

Figure 4.1: Dominant bacterial OTUs clustered at a distance of 0.00 present in the three *T. favus* TIC2014-001 sponge fractions created during bacterial enrichment. Read abundance is indicated as a

relative percentage of total bacterial reads analysed for each respective sponge sample. Coloured key is provided for identification of key conserved spirochaete and betaproteobacterial OTUs……..66

Figure 4.2: Comparative bacterial community composition of *T. favus* sponge specimens TIC2016-050A and TIC2014-001 generated from 16S rRNA amplicon and shotgun metagenomic sequence libraries. Read/Contig abundance is indicated as a relative percentage of total bacterial reads analysed for each respective sponge sample. Coloured key indicates phylogenetic classification…..68

Figure 4.3: Functional annotations assigned to predicted genes identified in metagenomic sequence data of *T. favus* sponges TIC2016-050A and TIC2014-001. Genes were annotated with reference to the KEGG orthologous database with a cut-off identity of 60% and a minimum length of 15 amino acids. Coloured key indicates predicted gene function annotation……70

Figure 4.4: Genomic bins of clustered contiguous sequences from *T. favus* specimen TIC2016-050A. Coloured keys indicated genomes designated 1 – 23 with genome numbers indicated on each cluster. 71

Figure 4.5: Distribution of COG classes in clustered bacterial genomes found in *T. favus* specimen TIC2016-050A. COG functional classes have been grouped into general functional categories indicated by boxed regions. COG functional classes are indicated by letter associations: D: Cell division and chromosome partitioning, M: Cell envelope biogenesis, outer membrane, N: Cell motility, O: Posttranslational modification, protein turnover, chaperone, T: Signal transduction mechanisms, J: Translation, ribosomal structure and biogenesis, K: Transcription, L: DNA replication, recombination and repair, C: Energy production and conversion, E: Amino acid transport and metabolism, F: Nucleotide transport and metabolism, G: Carbohydrate transport and metabolism, H: Coenzyme metabolism, I: Lipid metabolism, P: Inorganic ion transport and metabolism, Q: Secondary metabolites biosynthesis, transport, and catabolism, R: General function prediction only, S: COG of unknown function. Similarity in potential gene function annotation is inferred through a dendrogram generated by re-ordered clustering of the genome data matrix……..73

Figure 4.6: Distribution of secondary metabolite biosynthetic gene clusters in bacterial genomic bins from *T. favus* TIC2016-050A. Coloured keys indicate biosynthetic gene cluster classification with genome numbers indicated on each cluster. 75

Figure 5.1: Phylogenetic relationship of spirochaete taxa identified in Genome 3 relative to spirochete OTUs associated with *Tsitsikamma* and *Cyclacanthia* sponges (Chapter 3) and other spirochaete strains based on full length 16S rRNA gene sequence generated using MEGA7 software with the Maximum-likelihood method. Bootstrap values were calculated using 1000 replicates and are indicated at branch nodes. The number of substitutions per site, over 1468 nucleotides, is indicated with a scale bar. 16S rRNA gene sequences recovered from Genome 3 are highlighted with red boxes. 85

Figure 5.2: Contigs of Genome 3 aligned and re-ordered (bottom) relative to the genome of *Salinispira pacifica* L2-RPul-D2 (top) with the minimum weight of locally collinear blocks calculated to be 35 after 15 iterations. Regions of homology are indicated with blocks of matching colours with similarly coloured lines indicating a path between two homologous regions. Red lines indicate the borders of individual contigs in Genome 3. 86

Figure 5.3: Distribution of genes with predicted known gene function within Spirochaete Genome 3. Genes were classified into 26 different functional categories which are indicated on a coloured key. 87

Figure 5.4: Gene organization of Cluster A in Genome 3. The putative function of each ORF has been indicated. All ORFs are depicted to scale and relative size is indicated with a horizontal scale bar marked with size in number of base pairs. 88

Figure 5.5: Basic predicted chemical structure of compound produced from enzymes encoded on Cluster A. Rules of collinearity were assumed and tailoring reactions were not taken into account. Structure was drawn using ChemSpider. 89

Figure 5.6: Gene organization of Cluster B. ORFs have been indicated with coloured arrows. The putative function of each ORF has been indicated. All ORFs are depicted to scale and relative size is indicated with a horizontal scale bar marked with size in number of base pairs. 90

Figure 5.7: Gene organisation of Cluster C. The putative function of each ORF has been indicated. All ORFs are depicted to scale with coloured arrows and relative size is indicated with a horizontal scale bar marked with size in number of base pairs. Putative gene functions have been labelled using gene abbreviations; RimL: N-acetyltransferase, MATE-like: Multidrug and toxic drug-like extrusion protein, RfbF: Glucose-1-phosphate cytidylyltransferase, WcaG: CDP-glucose-4,6-dehydratase, WcaA: Glycosyltransferase Family II, NarL: transcriptional regulator, HisH: Imidazole glycerol phosphate synthase, HisF: Cyclase, WbuB: N-acetyl-L-fucosamine transferase, WecB: UDP-N-acetyleglucosamine-2-epimerase (non-hydrolyzing), WecC: UDP-N-acetyl-D-mannosamine dehydrogenase. 91

Figure 1.4: General mechanism of natural product formation in modular (**A**) NRPS gene clusters and (**B**) PKS gene clusters. C: Condensation domain, A: Adenylation domain, PCP: peptide carrier protein, KS: ketosynthase domain, AT: acetyltransferase domain, ACP: acyl-carrier protein..... 99

Figure 6.2: Gene organisation of the NRPS gene cluster identified in Genome 1. Modules are indicated with arrows orientated in the direction of transcription. Domains identified within each module are indicated using coloured circles and squares; the function of which can be identified using the key provided. A bar denoting the length of the gene cluster has been provided for relative size. Modules have been drawn to scale, domains have not. 103

Figure 6.3: Predicted linear structure of the hybrid non-ribosomal peptide/polyketide compound encoded by Cluster 1 when following the rule of collinearity, complemented with predicted substrates activated by all nine adenylation domains present within the cluster. Structure was drawn using ChemSpider 104

Figure 6.4: Gene organization of Type III-PKS gene clusters identified in clustered genomes extracted from metagenomic data from *T. favus* TIC2016-050A. Cluster 2 was identified in *Pseudovibrio* Genome 1 (**A**), Cluster 3 was identified in *Rhodopirellula* Genome 21 (**B**) and Cluster 4 was identified in *Rhodepirelulla* Genome 15 (**C**). Putative gene function of each ORF has been indicated. MDR: multidrug resistance protein (NorM), XBR: xenobiotic response transcriptional regulator, NUDIX: nucleotide diphosphate-linked moiety hydrolase, Rf1: peptide release factor H, RtcB: RNA ligases, PNDO: FAD-dependent pyridine nucleotide disulphide oxidoreductase, MarR: multiple antibiotic resistance transcriptional regulator, NCST: nucleobase cation-1-symporter, RimL: N-acetyltransferase, YgjV: inner membrane protein of unknown function, PKS: chalcone-like Type III polyketide synthase, ICMT: isoprenylcysteine carboxyl methyltransferase, ugpC: glycerol-phosphate transporter ATP binding protein, SBP: ABC transporter substrate binding protein. TEF: Translation elongation factor, RRF: Ribosome recycling factor, SelD: selenide water dikinase, SelA: L-seryl tRNA selenium transferase, Sel B: Selenocystein specific translation elongation factor, PKS: Type III

polyketide synthase, HMB synthase: Hydroxymethylbilane synthase. All ORFs are depicted to scale and relative size indicated with a horizontal bar demarked length denoted in number of base pairs. All ORFs are depicted to scale and relative size indicated with a horizontal bar demarked length denoted in number of base pairs.....107

Figure 6.5: Gene organisation of the hybrid NRPS/PKS Cluster 5 identified in Genome 15. ORFs are indicated with arrows orientated in the direction of transcription. Domains identified within each module are indicated using coloured shapes below each ORF; the function of which can be identified using the key provided. A bar denoting the length of the gene cluster has been provided for relative size. ORFs have been drawn to scale, domains have not.....108

Figure 6.6: Chemical structure of 4-methyl-2-morpholinone predicted to be produced from Cluster 5 found on Genome 15. Structure was drawn using ChemSpider.....109

Figure 6.7: Gene cluster organisation of Type I-PKS Cluster 6 identified on an unclustered contig. ORFs are indicated with arrows orientated in the direction of transcription. ARHD: Aromatic ring hydroxylating dioxygenase, FAAL: Fatty acyl-AMP ligase. ACP: acyl-carrier protein, Srf TE: Surfactin-like thioesterase, PKS: polyketide synthase, SAM-dep MT: SAM-dependent methyltransferase. Domains identified within each module are indicated using coloured shapes below each ORF; the function of which can be identified using the key provided. A bar denoting the length of the gene cluster has been provided for relative size. ORFs have been drawn to scale, domains have not. 110

Figure 6.8: Predicted chemical structure of polyketide produced by Cluster 6 in Genome 1 without tailoring reactions taken into consideration. Structure was drawn using ChemSpider.....111

Figure 6.9: Structural comparison of potential compound produced by Cluster 6 (**A**) and tsitsikammamine B (**B**); a pyrroloiminoquinone compound isolated from *T. favus* sponges. Predicted modifications by tailoring enzymes are indicated by colour; addition of oxygen molecules by aromatic ring hydroxylating dioxygenase (green), methylation by SAM-dependent methyltransferase (blue) and potential cyclization for pyrrole formation (pink). The corresponding region of tsitsikammamine B that the predicted compound matches is indicated within a dotted orange circle. Structures were drawn using ChemSpider.....112

List of Tables

Table 4.1: Metagenomic sequence data of TIC2016-050A.....64

Table 4.2. Characterisation of clustered genomic bins.....72

Table 6.1: Summary of broad characteristics of secondary metabolite biosynthetic gene clusters from the metagenome of *T. favus* specimen TIC2016-050A.....102

List of Abbreviations

A domain	Adenylation domain

ACP	Acyl carrier protein
antiSMASH	antibiotics & Secondary Metabolite Analysis Shell
AT	Adenyl-Transfer protein
ATP	Adenosine Triphosphate
BGC	Biosynthetic Gene Cluster
BLAST	Basic Local Alignment Sequence Tool
bp	Base pairs
C domain	Condensation domain
COG	Category of Orthologous Groups
COX 1	Cytochrome C Oxidase subunit I gene
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DNA	Deoxyribonucleic acid
FAAL	Fatty acyl-AMP ligase
gDNA	genomic DNA
IC_{50}	Half maximal inhibitory concentration
ITS	Internal Transcribed Spacer (ribosomal)
KS	Ketosynthase
LcK	Lymphocyte Tyrosine Kinase
MG-RAST	Metagenomic Rapid Annotations using Subsystems Technology
mRNA	messenger RNA
MRSA	Methicillin-resistant *Staphylococcus aureus*
mTOR	Mechanistic Target Of Rapamycin
NGS	Next Generation Sequencing
NRPS	Non-ribosomal peptide synthase
nts	Nucleotides
ORF	Open reading frame
OTU	Operational Taxonomic Unit
PCP	Peptide carrier protein
PCR	Polymerase Chain Reaction
PKS	Polyketide synthase

PVC	*Planctomyces*, *Verrucomicrobia* and *Chlamydia*
RAST	Rapid Annotations using Subsystems Technology.
RMS	Restriction modification system
RNA	Ribonucleic acid
ROV	Remote Operated Vehicle
rRNA	ribosomal RNA
SCUBA	Self-contained underwater breathing apparatus
SOP	Standard Operating Procedure
TE	Thioesterase
tRNA	transfer RNA

Acknowledgments

I would like to start by thanking the National Research Foundation and Rhodes University for funding my studies. Without the financial support provided I would not be in the fortunate position I am today.

To my supervisor, Prof Rosemary Dorrington, I will never be able to thank you enough. You took a lost and confused little girl under you wing and through your patience, guidance and sheer determination moulded me into the researcher I am today. We've had quite a few adventures together over the years and you've become far more than just a supervisor in my life, but also a mentor and friend.

And with Prof comes an epic group – Lab 417, aka "Team Awesome". Mez and Gwynn, my two big sisters – where do I even start? I will never be able to thank either of you enough for not only all the help in the lab - but your guidance, fun and laughter in my personal life as well. You've welcomed me into your lives and I couldn't be more stoked to call you my friends. I must make special mention of Dr Shirley Parker-Nance who has worked tirelessly alongside us to provide taxonomic classifications of the various sponge specimens. On a more personal level, Dr Parker-Nance has been a role-model for me over the years and I hope to enjoy many more adventures with you. Val, Janet, Mart, Holly, Sunet and Karin; you guys are real pro's and thank you for running such a tight ship. Big thank you to Rob and our lovely honours girls: Nasipi, Danielle, Ritoh and Rachel for making 417 such a fun place to work in. Jarmo, Xavier, Tarryn thank you for your friendship and incredible talent in chemistry.

I must also thank the Kwan Group, our collaborators at the University of Wisconsin, Madison. Dr Kwan, I visited your lab twice and both times I had a brilliant time. I learned more than I ever thought I could and I look forward to learning so much more from all of you. Ian in particular – you're an unbelievable human to learn from and you taught me more than I thought possible in my three weeks in Madison. So much of the work in this thesis is a result of your patience, intelligence, and brilliant take on leadership.

On a more personal note (and at the risk of sounding like I'm at the Oscars) I have an awful lot of wonderful people to thank, that helped me get to where I am today.

Mom, Dad, Tayla and Dylan (the Worthas); you guys have had my back since Day 1 and have seen me through some particularly rough times. Thank you for all the support, the love and genuinely believing in me. You're the most incredible family a girl could ask for and I love you all more than anyone else on this planet.

To my best friends: Page, Tam, Kyra, Storm, and Mollie. I am so spoilt to have all you girls in my life and to call you my best friends. You have all picked me up when I was down, encouraged me when I doubted myself and have brought more joy to my world than I could ever describe. Whether in person or over Skype, the various nights of giggles and wine have kept me going and I look forward to many more as we fumble our way through this amazing thing we call life.

I also have two wonderful gentlemen to thank. Eric and Nick – thank you for putting up with me and my nonsense. You've both been a great source of support and I'm incredibly lucky to continue to call you my friends. In particular, Eric, you're a brilliant researcher with an awesome mind and I'm looking forward on working on the stromatolites with you.

A final, big thank you needs to go to Sue Smailes. My debt of gratitude will never be repaid to you. This year would've been very tricky to navigate if it weren't for you, your relentless determination, and your gorgeous family.

Thank you everyone. Each of you influenced and changed my life in some way. I would never have gotten this far without each and every one of you.

Abstract

Several pyrroloiminoquinone alkaloids exhibiting cytotoxic, anti-tumour activity have been isolated from sponges within the Latrunculiidae family that are endemic to the South African coastline. Other, structurally similar pyrroloiminoquinone compounds have been isolated from geographically distant and phylogenetically distinct marine sponges, as well as terrestrial myxomycetes which suggested that sponge-associated bacteria may be the true biosynthetic origin of pyrroloiminoquinone compounds. Previous studies have shown that there is conservation of spirochete and betaproteobacterial species in the bacterial communities associated with South African Latrunculiidae sponges and it was proposed that these conserved bacteria represented candidate pyrroloiminoquinone-producers. This study aimed to confirm the conserved dominance of betaproteobacteria and spirochetes within bacterial communities associated with South African latrunculid sponges and employed a shotgun metagenomic approach to assess the functional and biosynthetic potential of associated microbiota in *Tsitsikamma favus* sponges. Clustering of assembled contigs revealed twenty-three putative bacterial genomes, of which, two were identified as representatives of the conserved betaproteobacteria and spirochete species previously identified in *Tsitsikamma* sponges. It was shown that the spirochete was most likely an obligate symbiont that benefitted the host sponge through possible defence against pathogenic bacteria and/or nutrient acquisition. The putative genome representing the conserved betaproteobacteria was found to be heavily contaminated and further sequencing is required to accurately resolve the genome for functional characterization. Several biosynthetic gene clusters were identified and demonstrated the bioactive potential of *Tsitsikamma favus*-associated bacteria. A biosynthetic gene cluster was identified on an unclustered contig that included several genetic features that were indicative of possible pyrroloiminoquinone production.

Chapter 1 - Literature Review

1.1. Introduction

The use of bioactive natural products dates back thousands of years with the employment of cedar oil, myrrh, and poppy juice to cure various ailments (Patridge et al., 2016). Using natural resources to cure disease continues to this day with a variety of natural products as prime candidates for the development of lead drugs (Li, 2015). Almost a third of new Food and Drug Administration (FDA) - approved molecular entities are derived from natural products (Amirkia and Heinrich, 2015; Patridge et al., 2016) with several new molecules in clinical development and many more in preclinical trials (Butler et al., 2014; Newman and Cragg, 2016). However, there has been a sharp decrease in the discovery of novel therapeutic compounds since 1988 (WHO Antimicrobial Resistance, 2014) with less than 10 new treatments approved per annum for the last decade (Newman and Cragg, 2016).

There has been increased interest in bioprospecting in the marine environment, which is a largely unexplored region of the planet that hosts a rich species diversity with unique adaptations. Marine organisms are therefore an attractive and important source of biologically active secondary metabolites (Gerwick and Moore, 2012). Compounds isolated from marine sources are typically structurally unique and include diverse core ring scaffolds, a high number of stereogenic centres and unusual proportions of heteroatoms (Barnes et al., 2016).

The potential of marine natural products for clinical use was first realised in 1951 when unusual nucleosides, spongothymidine and spongouridine, were isolated from the sponge *Tethya crypta* (Bergmann and Feeney, 1951). Both compounds were isolated from *T. crypta* specimens found in distinct geographic locations, Bahamas and Florida (Bergmann and Feeney, 1951), which indicated that the compounds may be specific to this sponge species. A third compound, spongosine, was isolated from *T. crypta* four years after the isolation of spongothymidine and spongouridine (Bergmann and Burke, 1956) and was successfully chemically synthesised (Bergmann and Stempien, 1957). Two synthetic analogues of spongosine have been developed (Walwick et al., 1959; Lee et al., 1960) and approved for clinical use (Glaser and Mayer, 2009).

Cytarabine is used as a chemotherapy drug in the treatment of leukaemia and non-Hodgkin's Lymphoma while another, Vidarabine, is an antiviral compound (Mayer et al., 2010).

Since the discovery of the novel nucleosides from the *T. crypta* sponge, approximately 20 000 structurally novel marine compounds have been discovered (Martins et al., 2014) but only eight have passed clinical trials with a further ten undergoing currently in the clinical testing stage (Li, 2015). Of these seven, only two are unaltered marine natural products *viz.* Ziconotide isolated from a cone snail, *Conus magus* and Trabectedin isolated from a tunicate, *Ecteinascidia turbinata*. The remaining five compounds are derivatives of natural products isolated from seaweed (carrageenans), fish (Lovaza), a mollusc (Brentuximab) and sponges (Cytarabine, Vidarabine and Eribulin) (Li, 2015).

Investigations by various groups have shown that species belonging to the Porifera and Cnidaria phyla are responsible for the majority (59.89%) of marine bioactive compounds (Hu et al., 2015; Blunt et al., 2015) with 4851 compounds recorded thus far from sponges (an average of 200 compounds reported annually). This number of compounds equate to approximately 30.51% of all marine natural products identified to date (Gerwick and Moore, 2012; Mehbub et al., 2014; Hu et al., 2015).

1.2. Bioactive Secondary Metabolites from Sponges

As sessile creatures, sponges are incapable of escaping predators or avoiding disease and fouling and instead, produce a variety of toxic secondary metabolites to ward off predators, compete with other sponges for living space and to protect against infection (Belarbi et al., 2003). These compounds are often extremely potent and stable, an evolutionary response to the large dilution effect of the ocean (Abad et al., 2011; Mehbub et al., 2014). For example, spongiostatin 1 isolated from *Spongia* and *Hyrtios* sponge species has exhibited cytotoxic effects against various human cancer cell lines at picomolar concentrations without significant damage to non-cancerous cells (Gomes et al., 2016).

Sponges produce a diversity of structurally and functionally varied metabolites that fall into clinically important chemical classes including quinones (Sunassee and Davies-Coleman, 2012; Jiao et al., 2014), alkaloids (Essack et al., 2011; Bailly, 2015), polyketides (Hirata and Uemara, 1986; Piel et al., 2006; Sala et al., 2014, Chianese et al., 2015; Esposito et al., 2015), terpenes (Chang et al., 2012; Gupta et al., 2012, Dattelbaum et al., 2015) and peptides (Mohammed et al., 2006; Zampella et al., 2008; Martin et al., 2014; Carstens et al., 2015). Peptide compounds account for the majority (40.85%) of bioactive products identified thus far (Hu et al., 2015). However, the number of novel compounds derived from sponges appears to be plateauing relative to previous years with only 291 novel sponge-derived compounds identified in 2015, the majority of which were terpenoids. Likewise, the number of novel cnidarian-derived compounds has decreased by 40% and is likewise dominated by terpenoid compounds (Blunt et al., 2017).

The apparent decrease in numbers of novel compounds from marine organisms could be due to the fact that multiple marine compounds are now being attributed to sponge/cnidarian-associated bacteria as the true producers of marine-derived compounds rather than the marine organism itself (Lackner et al., 2016). The hypothesis that host-associated microbial communities are responsible for the production of protective compounds was based on the observation that many structurally similar compounds are produced by unrelated species and/or geographically distant sponge specimens (Davis et al., 2012; Walmsley et al., 2012; Zou and Hamann, 2013; Botic et al., 2015). This initial observation, coupled with the isolation of polyketides and non-ribosomal peptides from sponges, strengthened the validity of the hypothesis of bacteria as true producers, as polyketide and non-ribosomal peptide compounds are exclusively produced by bacteria (Piel, 2006; Zhang et al., 2009). Advances in this regard have allowed identification of symbiotic bacterium *Candidatus Endoecteinascidia frumentensis* which is associated with a mangrove tunicate species, as the true producer of commercially available trabectedin (Yondelis©) used for cancer treatment (Schofield et al., 2015). Likewise, a *Vibrio harveyi* strain was identified as the biosynthetic origin of spongosine and other structurally-related compounds that were previously attributed to the sponge *T. crypta* (Bertin et al., 2015).

1.3. Sponge-associated Microbial Communities

Sponges feed on bacteria via filtration of large volumes of seawater and as a result are in direct contact with a diverse community of microorganisms. Some bacteria evade predation and are instead, recruited by the host sponge as symbionts (Thomas et al., 2016). The microbial communities of sponges are often specific to their host and provide a multitude of services such as carbon fixation (Freeman et al., 2013), removal or degradation of toxins (Santos-Gandelman et al., 2014), sequestration of phosphorous compounds (Zhang et al., 2015), nitrogen fixation (Mohamed et al., 2008), cycling of sulphur (Jensen et al., 2017) and protection from predation and infection via the production of bioactive secondary metabolites (Taylor et al., 2012; Horn et al., 2016).

Associated bacteria play an important role in survival of the host sponge (Hardoim et al, 2014; Thomas et al., 2016). The ability of the bacterium to perform this role is dependent on a maintained association with the host sponge and as such must be able to adapt to fluctuating conditions (Erwin et al., 2015). Sponge-associated microbial communities are distinct from the surrounding seawater and sediments (Pita et al., 2013; Horn et al., 2016; Thomas et al., 2016; Weigel and Erwin, 2016) and remain largely stable within a specific sponge host (Hardoim et al., 2012; Weigel and Erwin, 2016). A survey of global trends in sponge-associated bacterial communities has indicated that the sponge-host influences microbiome complexity rather than composition (Thomas et al., 2016) and that the conservation of associated bacteria remains largely stable despite changes in temperature, irradiance (Erwin et al., 2012), pollution levels (Gantt et al., 2017), seasons (Taylor et al., 2004; Thiel et al., 2007a) or geographical location (White et al., 2012; Weigel and Erwin, 2016).

The microbial communities associated with sponges are principally dominated by members of the Proteobacteria phylum, particularly of the Alpha- and Gamma-proteobacterial classes (Thomas et al., 2016) but many sponges harbour a candidate bacterial phylum, Poribacteria, which is found exclusively within sponge tissue and could not be detected in surrounding sediment, seawater, or neighbouring tunicates (Hug et al., 2016; Steinert et al., 2017). Poribacteria share greatest phylogenetic similarity to the Planctomyces, Verrucomicrobia and Chlamydia (PVC) phyla, with an

average 16S rRNA gene sequence identity of 74.6% (Fieseler et al., 2004). Like the members of the PVC phyla, poribacteria exhibit membrane-bound organelles (Fieseler et al., 2004; Jahn et al., 2016) and encode adhesions and eukaryotic-like proteins that may facilitate symbiotic interactions with its sponge host (Siegl et al., 2011; Kamke et al., 2014).

The metabolic role of poribacteria within sponge hosts was investigated through single-cell genomics (Siegl et al., 2011; Kamke et al., 2013). The poribacterial genome was found to harbour 30% of genes which have no known function (Siegl et al., 2011; Kamke et al., 2013) and several genes suggested to be involved in host interaction such as polyketide synthase (PKS)-derived methyl-branched fatty acids (Siegl et al., 2011), adhesins and adhesion-related protein gene sequences, ankyrin- and tetratricopeptide repeats (TPR) (Siegl et al., 2011; Kamke et al, 2014)

1.4. Acquisition of Sponge-associated Bacteria

Sponge-associated bacteria can comprise up to 35% of sponge biomass and can be found either in the mesohyl or intracellularly within sponge tissue (Webster and Thomas, 2016). Sponge symbionts contribute to the health and survival of their sponge host (Brinkmann et al., 2017b) and it has been proposed that sponge and associated microbiome be viewed as a holobiont (Webster and Thomas, 2016) as defined by Margulis and colleagues (1991). Nascent sponges/sponge larvae must therefore acquire the same bacterial consortia as that present in the tissues of its parent and this can be achieved through either vertical or horizontal transmission of microbiota, or through a combination of the two approaches (Webster, 2014).

Vertical transmission is the more dominant method of acquiring microbial symbionts (Bright and Bulgheresei, 2010) and is achieved when sponge symbionts are incorporated into developing sponge embryos. Sponge embryos are housed in the mesohyl thus limiting the distance between it and the associated bacteria which enables efficient acquisition of the symbionts by the embryos (Bright and Bulgheresei, 2010). If the symbiont is intracellular, it can be incorporated into the host embryo in several ways, including endocytosis (Ereskovsky et al., 2005), via a specialised

structure known as a "nurse" cell (Usher et al., 2001; Usher and Ereskovsky, 2005) or through a collagenous connection (Kaye, 1991).

Horizontal transfer is the acquisition of symbionts from the environment and generally occurs in addition to vertical transmission. Due to the challenges of host recognition, discrimination and establishment of environmental associations, horizontal transmission of symbionts remains poorly understood but is proposed to involve CRISPR and ELP enzymes as well as eukaryotic-like ankyrin repeat proteins (Fan et al., 2012; Nguyen, et al., 2013; Webster, 2014). A combination of horizontal and vertical transfer was observed in *Spongia lamella* sponges in which a "core" set of symbionts was vertically transferred but other bacterial symbionts were assimilated from the surrounding seawater in response to environmental stressors (Noyer et al., 2014). An investigation into the transmission of associated bacteria in *Tethya rubra* sponges collected from Algoa Bay, South Africa showed that conserved alphaproteobacterial and betaproteobacterial operational taxonomic units (OTUs) (clustered at a distance of 0.00) were inherited by sponge embryos from the parent sponge (Waterworth et al., 2017 - Appendix C). It was suggested that these bacteria produce bioactive compounds isolated from the *T. rubra* sponges and therefore, vertical transmission of the bacteria to nascent sponge embryos would enable continued production of these useful compounds in sponge progeny (Waterworth et al., 2017).

The structure of bacterial communities associated with marine sponges is influenced more by the phylogeny of their host sponge than the surrounding environmental factors (Reveillaud et al., 2014; Steinert et al., 2016; Astudillo-Garcia et al., 2017; Souza et al., 2017, Steinert et al., 2017) which is indicative of a shared evolutionary relationship (Webster and Thomas 2016; Matcher et al., 2017 - Appendix C). Co-evolution of sponges and their associated microbiota strengthens the hypothesis that the conserved, associated microorganisms play a key role in sponge health and are frequently implicated in the production of bioactive secondary metabolites (Thomas et al., 2010b; Florez et al., 2015; Blunt et al., 2017)

1.5. Acquisition of Biosynthetic Gene Clusters in Marine Sponge-associated Bacteria

There is mounting evidence to suggest that bioactive compounds isolated from sponges are produced by associated microorganisms to defend the host organism in return for a nutrient-rich habitat (Gurgui and Piel, 2010; Abdelmohsen et al., 2014a; Lackner et al., 2016; Blunt et al., 2017). Bacterial species within the order Actinomycetales are well known secondary metabolite producers and are responsible for the production of approximately two thirds of all known antibiotics (van der Meij et al., 2017). While novel compounds have traditionally been isolated from terrestrial actinomycetes, the marine-derived actinobacteria have become the focus of many research groups for the isolation of new bioactive compounds (Lam, 2006). The greatest proportion of marine actinomycetes (54%) have been identified in sediment, with marine sponges harbouring the second greatest number at 21%, (Abdelmohsen et al., 2014b). Sponge-associated actinomycetes include newly identified and rare genera *Actinokineospora, Actinomadura, Amycolatopsis, Demequina, Knoellia, Nonomuraea, Pseudonocardia, Saccharomonospora, Saccharopolyspora*, *Salinispora* and *Verrucosispora* (Bull and Stach, 2007; Abdelmohsen et al., 2014b).

Bacteria within the *Salinispora* genus are classified into three species *viz. S. tropica*, *S. arenicola* and *S. pacifica* (Jensen et al., 2007). *Salinispora* species isolated from different sponges have been shown to produce various bioactive compounds such as nazumamide, cycloaspeptide, and rifamycins B, SV and W (Kim et al., 2006; Bose et al., 2014; Singh et al., 2014). Despite close phylogenetic relatedness of *Salinispora* species, there has been no observed overlap of secondary metabolite production, yet some of the same compounds have been isolated from other actinomycete isolates (Jensen et al., 2007). A recent survey of 119 *Salinispora* genomes revealed that very few biosynthetic gene clusters were shared between species but were localized in concentrated genomic islands that provided “hotspots” for horizontal gene transfer (Letzel et al., 2017). It has been found that many of these biosynthetic gene clusters are specifically packaged for horizontal gene transfer and are often linked with mobile genetic elements (Jensen, 2016). It was suggested that the exchange of genes between closely related bacteria is favoured over transfer to more distantly related bacteria due to common GC content and codon usage (Jensen, 2016). The

subsequent genomic flexibility of these related strains enables rapid metabolic adaptation to various environmental niches (Jensen et al., 2017; Letzel et al., 2017).

1.6. Culture-dependent methods for isolation of bioactive secondary metabolites

Advances in sustainable drug development has led to increased focus on bacteria as producers of biologically active metabolites, as they could alleviate the bottle-neck of "supply problems" associated with large quantities of marine sponges being depleted for the retrieval of small quantities of bioactive products (Gomes et al., 2016). Several pharmaceutically important compounds produced by marine-associated bacteria have since been isolated and a few of the biosynthetic pathways thereof are discussed further here.

Sequencing and construction of the complete genome of the culturable, model species *Streptomyces coelicolor* A3(2), revealed numerous gene clusters with predicted biosynthetic potential, the products of which had never been isolated under laboratory conditions (Bentley et al., 2002). This discovery implied that many bacteria from which products were isolated may have further biosynthetic potential that had previously been overlooked (Gross et al., 2007; Winter et al., 2011). The number of culturable marine bacteria is estimated to be lower than 1%, but it is reported that dilution of inoculum, careful choice of media and long incubation times can result in a greater yield of culturable marine bacterial isolates (Perez and Fenical, 2017).

The production of bioactive secondary metabolites in bacteria is often triggered by environmental stressors such as depletion of nutrients or the presence of competing bacterial communities (Devine et al., 2017). These stressors can be mimicked using various media, plating techniques, pre-treatments and culture conditions (Purves et al., 2016). Cultivation of marine *Pseudovibrio* sp. isolated from *Arenosclera brasiliensis* sponges was dependent on the presence of XAD- resins in culture medium and led to the production of several bioactive bromo-tyrosine alkaloids (Nicacio et al., 2017) that have previously been shown to exhibit growth inhibitory activity against parasitic protozoa and various pathogenic bacteria (de Lira et al., 2006; Galeano et al., 2011).

Genome sequence analysis of bacterial isolates can be exploited to aid in the search for novel marine natural products in the form of orphan gene clusters. Genes encoding secondary metabolites are often grouped together in a cluster and controlled by a single regulator (Undabarrena et al., 2017). Orphan gene clusters are biosynthetic pathways for which the product has yet to be discovered as the molecular signal necessary for production is commonly absent in a laboratory setting (Chiang et al., 2011). The identification of biosynthetic gene clusters within a sequenced genome can be achieved using various bioinformatic tools tailored to the detection of bacterial secondary metabolite pathways (Offret et al., 2016). Subsequent adjustment to culture media/conditions etc regarding the requirements of the identified pathway may enable the expression and consequent isolation of the encoded compound (Maansson et al., 2016).

An illustrative case study of genome mining leading to the production of a novel compound encoded by an orphan gene cluster is that of culturable *Salinispora* bacteria. During experiments to identify the bacterial source of namenamicin (Fig. 1.1A) from bacteria associated with the marine ascidian *Polysyncraton lithostrotum* (McDonald et al., 1996), novel compounds lomaiviticin A and lomaiviticin B were identified. These compounds were produced by an associated *Micromonaspora lomaivitiensis* strain (later correctly identified as *Salinispora pacifica* (Janso et al., 2014)) and exhibited DNA-damaging activity, antibacterial activity, whilst lomaiviticin A (Fig. 1.1B) exhibited cytotoxicity against several cancer cell lines (He et al., 2001). A putative biosynthetic pathway encoding lomaiviticin was identified as an orphan pathway through genome mining of *S. tropica* CNB-440 (Kersten et al., 2013). The structure of lomaiviticin A (Fig.1.1B) included a dimeric diazofluorene core that was similar to the monomeric diazofluorene core found in kanamycins (Kersten et al., 2013). Once identified, it was revealed that the biosynthetic cluster encoding lomaiviticin consisted of 59 open reading frames (ORFs) which included all biosynthetic genes predicted to encode kanamycin (Kersten et al., 2013). A matching lomaiviticin biosynthetic pathway was subsequently identified as a polyketide synthase

(PKS) pathway approximately 70 kb in length on the genomes of *S. pacifica* strains DPJ-0016 and DPJ-0019 (Janso et al., 2014).

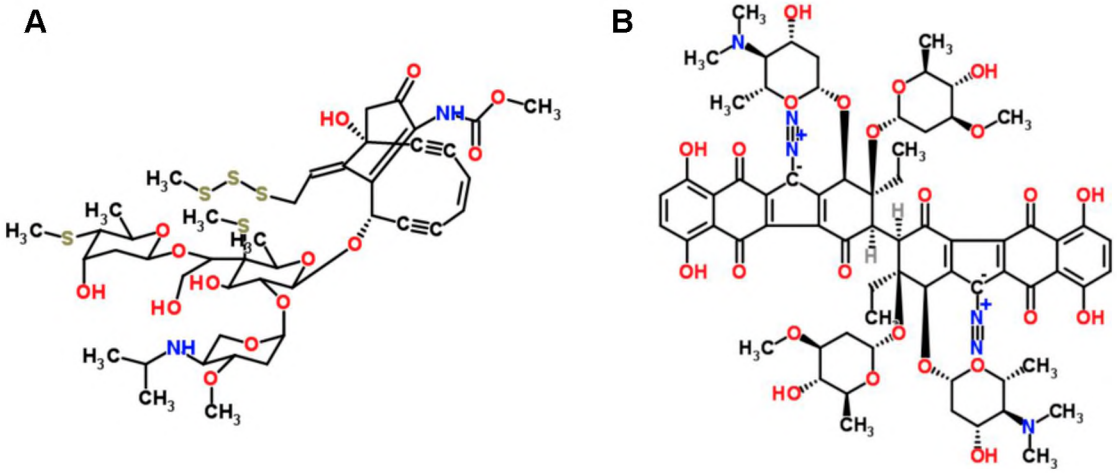


Figure 1.1: Chemical structure of namenamicin (**A**) and lomaiviticin (**B**) from marine *Streptomyces* sp. Structures were adapted from McDonald et al., 1996 and Janso et al., 2014 and drawn using ChemSpider.

The lomaiviticin pathway in *S. pacifica* DPJ-0016 was identified through screening and sequencing of a cosmid library, whilst the entire genome of DPJ-0019 was sequenced and analysed (Janso et al., 2014). Mutational disruption of a β-ketoacyl synthase gene within DPJ-0019 resulted in a total loss of lomaiviticin production and validated the gene cluster as the genetic origin of lomaiviticins (Janso et al.,2014). Subsequent *in silico* guided research of lomaiviticin biosynthesis has identified a new form of starter unit generation by a PKS system (Waldman and Balkus, 2014) as well as novel additional enzymatic mechanisms involved in peptide modification (Wang et al., 2014).

Despite advances in culturing approaches, isolation of sponge-associated bacteria remains a major challenge as the complexity of the marine environment cannot currently be accurately represented and therefore the sponge-microbiome and its associated chemistry may differ from nature (Behie et al., 2017). It has been proposed that these "unculturable" bacteria require particular nutritional requirements, are inhibited by factors inherent to culture media and/or rely on other symbiotic bacteria for the production of certain compounds and therefore cannot grow in isolation (Vartoukian, 2016). It has been proposed that future advances in culturing of fastidious marine bacteria will require assessment of culture-independent "meta-omics" for a

better understanding of physiological and biochemical requirements of uncultivated bacteria (Overmann et al., 2017).

1.7. Culture-independent methods for identification of bioactive secondary metabolites

Individual members within bacterial communities associated with marine sponges are potentially interacting with one another as well as the sponge host to effectively control the phenotype of the holobiont (Webster and Thomas, 2016). It has been proposed that the greatest limitation of culturing sponge-associated bacteria are the interactive networks between various bacterial species within the environment and their dependence on neighbouring bacteria for production of growth factors and other chemical compounds for successful proliferation (Vartoukian, 2016). The core microbiota of a sponge taxon, i.e. the bacterial isolates that are present in all given sponge specimens (Astudillo-Garcia et al., 2017), are hypothesised to fulfil various ecological roles within the host which facilitate its fitness in an environment (Schmitt et al., 2012). Furthermore, comparison of the sponge-associated and seawater-dwelling bacterial communities has shown that there are genomic differences between the two populations with several features characteristic of symbiosis observed in the sponge-associated consortia and absent in the free-living bacteria (Horn et al., 2016).

As discussed previously, many sponge-associated bacteria cannot be cultured using current culture media and technologies. Thus culture-independent methods need to be employed to assess the persistence, abundance, and potential contribution that a given bacterial taxon provides for its sponge host. Conservation of bacterial communities can be assessed through analysis of 16S rRNA amplicon libraries generated using Next-Generation Sequencing (NGS) platforms. However, care must be taken so as not to over-interpret results as primer choice can influence amplification bias which effectively over- or under-represents certain taxa in downstream analyses (Parada et al., 2015). Assessment of the metabolome of the core microbiota and consequently their potential role within the holobiont requires a shotgun metagenomic approach (Woyke et al., 2006).

Shotgun metagenomics is the untargeted sequencing of all genomic DNA in a given sample (reviewed in Quince et al., 2017). Preparation of sponge tissue for DNA extraction often includes an enrichment step to decrease the quantity of host tissue but as next-generation sequencing and downstream computational approaches advance, this treatment is becoming increasingly unnecessary (Wilson and Piel, 2013). However, with large sequence datasets comes the requirement of large computing power for accurate assembly of sequence reads into contiguous sequences (contigs) with the additional challenge of assembling non-ribosomal peptide synthase (NRPS) and PKS biosynthetic gene clusters which include highly conserved and repetitive regions (Wilson and Piel, 2013). Various computational assemblers are available for both online and stand-alone use, with metaSPAdes and MEGAHIT pipelines proving to be the most popular (Quince et al., 2017). The assembled contigs must be sorted and classified into taxonomic "bins" in a process called "binning" (Quince et al., 2017). Initially the extraction of bacterial genomes from metagenomic data was limited to samples in which there was very little diversity, but with advanced assembly, binning and annotation algorithms, this approach can be used for recovery of genomes from samples with complex communities (Bowers et al., 2017).

Various available bioinformatic tools are available for downstream analysis, such as open reading frame (ORF) calling, gene annotations, genome quality assessment and screening for biosynthetic gene clusters via identification of conserved regions (Quince et al., 2017). Metagenomic analysis is becoming increasingly popular (Mitchell et al., 2017) as it enables the unbiased, relatively rapid identification of both novel biosynthetic gene clusters, the relative abundance of these clusters across different bacterial communities as well as insight into the potential ecological functionalities of "unculturable" microbial symbionts (Wilson and Piel, 2013; Tracanna et al., 2017). This approach has been used to identify over 300 putative biosynthetic pathways in 100 bacterial genomes associated with *Hippospongia lachne* sponges (Miller et al., 2016), the identification of the mandelalide biosynthetic gene cluster in *Lissoclinum* tunicate-associated Verrucomicrobia species (Weyna et al., 2015) as well as clustering of thirty-seven putative genomes from a *Aplysina aerophoba* sponge which revealed various ecological gene functions of the associated bacteria (Slaby et al., 2017).

1.8. Microbial biosynthesis of pyrroloquinoline alkaloids

Pyrroloquinoline alkaloids have long been the subject of several investigations due to their novel structure and potent inhibitory activity, with most of these compounds isolated from marine sponges (Jordan and Moore, 2016). Biosynthesis of pyrroloquinolines has been shown to include a tryptophan molecule in its core scaffold (Miyanaga et al., 2011) which has long been postulated as the starter molecule in the biosynthesis of structurally related pyrroloiminoquinones (Discussed in detail in Section 1.9)

1.8.1. Lymphostin

Lymphostin (Fig. 1.2) is an orange aromatic alkaloid including a tricyclic pyrroloquinone core that is rare among natural product structures (Aotani et al., 1997; Nagata et al., 1997). First isolated from broth culture of soil-borne *Streptomyces* sp. KY11783, lymphostin exhibited immunosuppressant activity through inhibition of lymphocyte kinase and inhibition T-cell activation at an IC_{50} value of 0.05 µM and 0.009 µM respectively (Aotani et al., 1997; Nagata et al., 1997). It was later demonstrated that lymphostin exhibits a three-fold mechanism of inhibition: inhibition of the of lymphocyte tyrosine kinase (Lck) autophosphorylation, inhibition of phosphatidylinositol-3-kinase and inhibition of Lck-mediated phosphorylation of tyrosine (Nagata et al., 2002). Furthermore, it was found that lymphostin and various derivatives thereof exhibited potent inhibitory activity against the phosphatidylinositol 3-kinase-related mTOR enzyme within both nano- and picomolar ranges (Miyanaga et al., 2011).

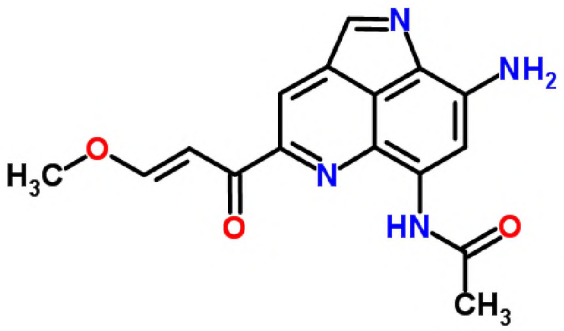


Figure 1.2: Chemical structures of immunosuppressant lymphostin produced by *Streptomyces* sp. KY11783 Structure was adapted from Aotani et al., 1997 and drawn using ChemSpider.

Lymphostin was successfully synthesised through six regio-selective oxidative reactions using tryptophan as a starting substrate inspired through mimicry of a predicted biological pathway (Tatsuta et al., 2004). The compound was also subsequently isolated from the marine actinomycete *Salinispora tropica* during fermentation experiments to produce salinosporamide A (Udwary et al., 2007). It was found that an identified hybrid NRPS/PKS cluster containing a modules *lym*A and *lym*B was responsible for the production of lymphostin (Fig. 1.3) (Udwary et al., 2007; Miyanaga et al., 2011). The *lym* gene cluster was also identified in the genomes of related *S. arenicola* CNS-205 which was isolated from tropical marine sediment (Penn et al., 2009) and *S. pacifica* strains DPJ-0024 and DPJ-0019 isolated from tunicate *Didemnum proliferum* (Miyanaga et al., 2011). It was proposed that the pathway had potentially been acquired from environmental actinomycetes via horizontal gene transfer (Penn et al., 2009).

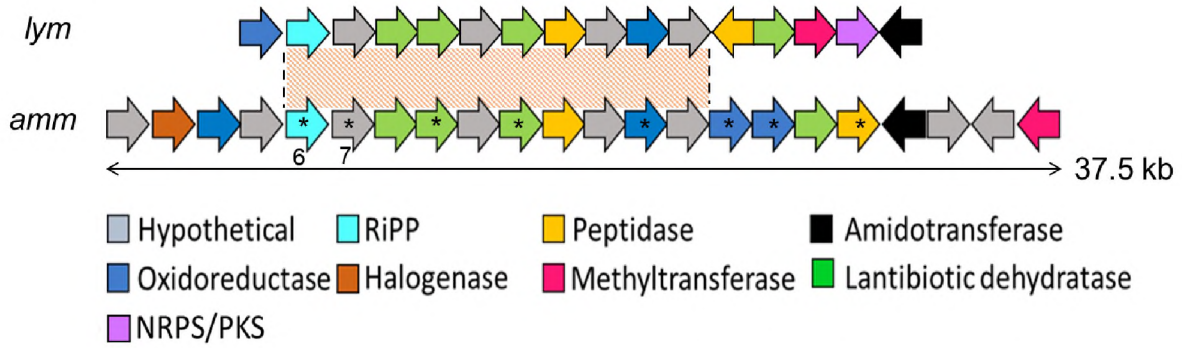


Figure 1.3: Schematic representation of the lymphostin (top) and ammosamide (bottom) biosynthetic gene clusters in *Salinispora* and *Streptomyces* bacteria respectively modified from Jordan and Moore, 2016. Genes are represented by coloured arrows orientated in the direction of transcription and are not drawn to scale. A coloured key is provided for functional annotation of genes. The conserved genes are highlighted with a striped orange region. ORFs annotated with an asterisk indicate genes vital to ammosamide production. Genes *amm*6 and *amm*7, which are hypothesised to form the pyrroloquinoline core are indicated with the appropriate number below. Abbreviations: RiPP: ribosomally synthesized, post-translationally modified peptide, NRPS: non-ribosomal peptide synthase, PKS: polyketide synthase. Modules *lym*A and *lym*B are situated within the hybrid NRPS/PKS encoding gene (lilac).

Computational analysis of the gene clusters suggested that tryptophan-derived starter unit that contains the pyrroloquinoline core was selectively bound to the acyl-carrier protein of *lym*A and underwent a single polyketide extension with malonate followed by N-acetylation and reductive release as an aldehyde which is trans-methylated by NRPS/PKS module *lym*B (Udwary et al., 2007; Miyanaga et al., 2011). This reported pathway for lymphostin production was validated using mutant strains of *Salinispora arenicola* and *Salinispora tropica* in which gene modules *lym*A and *lym*B were

inactivated and subsequently lymphostin production was lost (Miyanaga et al., 2011). Isotopic experiments confirmed that tryptophan and glutamine are involved in the biosynthesis of lymphostin but the enzymes responsible for the biosynthesis of the pyrroloquinoline core could not be identified (Miyanaga et al., 2011).

1.8.2. Ammosamides

Ammosamides A and B (Fig 1.4) are produced by marine *Streptomyces* sp. CNR-698 and along with lymphostin, are the only other compounds that harbour a pyrrolo[4,3,2-de]quinoline core (Reimer and Hughes, 2017). Ammosamides are biologically active and have been shown to bind myosin (Hughes et al., 2009), have moderate cytotoxic effects against human colorectal cancer cell line HCT-116, inhibit the growth of *Bacillus oceanisedimenis* (Reimer and Hughes, 2017) and have been fully synthesised (Hughes and Fenical, 2010).

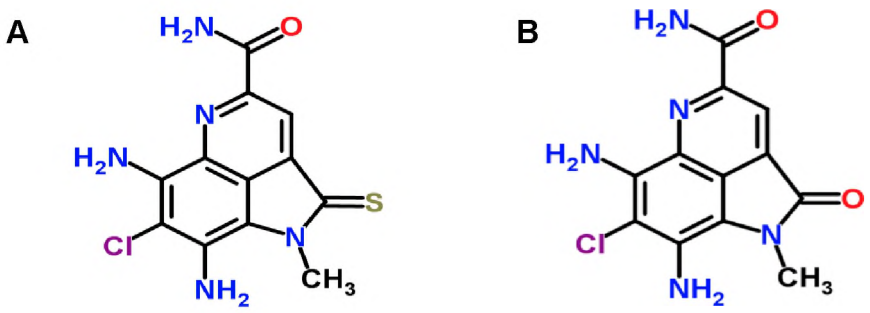


Figure 1.4: Chemical structure of ammosamide A (**A**) and ammosamide B (**B**) produced by *Streptomyces* sp. CNR-698. Chemical structures were adapted from Jordan and Moore, 2016 and drawn using ChemSpider.

The biosynthetic gene cluster responsible for producing ammosamides has recently been elucidated (Fig. 1.3) and it was found to share a conserved set of biosynthetic genes with that of the *lym* gene cluster (Fig. 1.3, orange region) (Jordan and Moore, 2016). Heterologous expression of the putative ammosamide gene cluster in *Streptomyces coelicolor* M512 resulted in high yields of ammosamides A, B and C, with ammosamide C being the predominant product (Jordan and Moore, 2016). Following several gene deletion experiments it was concluded that genes *amm*7, -9, -11, -14, -16, -17 and *amm*19 (Fig. 1.3. indicated with asterisks) are vital to the production of ammosamides and that genes *amm*6 and *amm*7 are likely involved in the production of the pyrroloquinoline core (Jordan and Moore, 2016).

1.9. Pyrroloiminoquinones isolated from Latrunculiidae sponges

The sponge family Latrunculiidae (Class Demospongiae; Order Poecilosclerida) consists of 7 accepted genera, *Bomba*, *Sceptrella*, *Strongylodesma*, *Cyclacanthia*, *Latrunculia*, *Latrunclava*, and *Tsitsikamma* (WoRMS Editorial Board, 2017). Latrunculid sponges were first described in 1922 (Topsent, 1922) and have since undergone several revisions and are still a matter of some controversy in taxonomic circles (Samaai et al., 2002; Samaai et al., 2003). Latrunculid sponges are characterized by the presence of discorhabd microscleres and have never been observed to possess chelae microscleres. Their gross morphology commonly includes a leathery texture, raised oscular fistules and areolate porefields (Samaai et al., 2002; Samaai et al., 2003). Latrunculid sponges are abundant in temperate and cold waters off the coasts of South Africa, Australia, New Zealand, Tasmania, and Antarctica in the Southern hemisphere as well as in Northern waters of British Columbia, the Aleutian Islands, the Gulf of Alaska, the Russian Sea of Okhotsk and the Pacific Northwest (Kelly et al., 2016). Sponges species belonging to *Strongylodesma*, *Cyclacanthia*, *Latrunculia*, and *Tsitsikamma* genera are endemic to South Africa.

Sponges of the family Latrunculiidae are well known producers of bioactive alkaloid compounds that possess a characteristic fused three-ring pyrroloiminoquinone core structure. This core is included in makaluvamines, discorhabdins, batzellines, damirones and tsitsikammamines (Fig. 1.5) (Antunes et al., 2005). Pyrroloiminoquinone compounds possess potent biological activities including anti-tumour, antibacterial, antifungal and anti-malarial activity through inhibition of topoisomerase, sortase A and indoleamine 2,3-dioxygenase (Dolusic et al., 2013; Mioso et al., 2017). Despite their promising bioactivity, no pyrroloiminoquinone compounds have yet to reach pre-clinical testing stages (Lin et al., 2017).

1.9.1. Pyrroloiminoquinones isolated from sponges within the genus *Tsitsikamma*

Sponges within the *Tsitsikamma* genus are characterised by a tough, leathery texture, large internal spicule tracts and are distributed exclusively in the warm, temperate waters off the south-east coast of South Africa (Samaai et al., 2002). There are currently only three officially recognised species within *Tsitsikamma* viz. *T. favus*, *T. pedunculata* and *T. scurra* (WoRMS database, 2017) but two new species, namely *T. nguni* and *T. michaeli*, have been identified in the course of this and other studies.

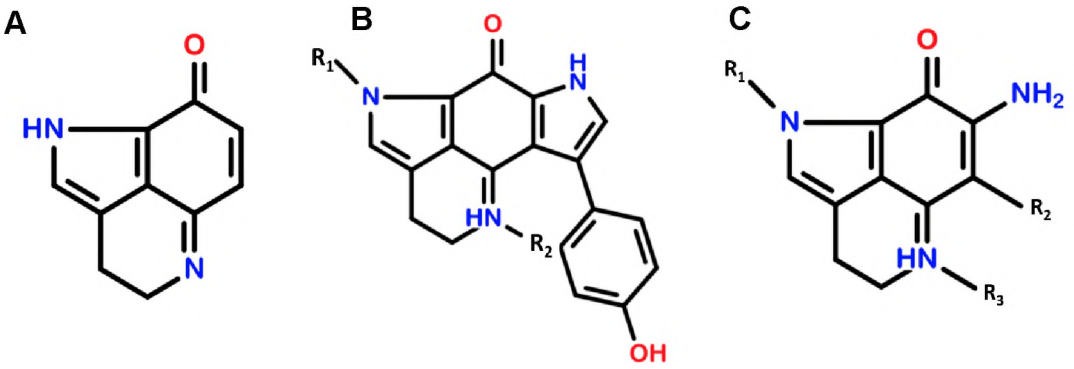


Figure 1.5: Chemical structure of the pyrroloiminoquinone core (**A**), tsitsikammamines (**B**) and makaluvamines (**C**). Chemical structures were adapted from Antunes et al., 2004 and Antunes et al., 2005 and drawn using ChemSpider.

Brominated derivatives of discorhabdin C, 14-bromodiscorhabdin C, 14-bromo-3-dihydrodiscorhabdin C, and 3-dihydrodiscorhabdin C have been isolated from *T. pedunculata* and *T. favus* sponges collected in Algoa Bay South Africa (Antunes et al., 2004). Novel discorhabdin derivatives 3-dihydro-7,8-dehydrodiscorhabdin C, 14-bromo-3-dihydro-7,8-dehydrodiscorhabdin C, discorhabdin V and 14-bromo-1-hydroxydiscorhabdin V were isolated from *T. pedunculata* and tsitsikammamines A and B (Fig 1.6 A-B) were isolated from *T. favus* as major metabolites (Antunes et al., 2004). It was noted by Antunes and colleagues (2004) that *T. pedunculata*, although phylogenetically closely related to *T. favus*, did not produce tsitsikammamines A and B. It was initially thought that *T. favus* was the exclusive source of tsitsikammamines (Antunes et al., 2005) until tsitsikammamine C was isolated from an Australian *Zyzzya* sponge species (Davies et al., 2012). Tsitsikammamines A and B were cytotoxic against human colon carcinoma HCT-116 cells (IC_{50} = 1.4 µM and 2.4 µM respectively) and inhibited topoisomerase I with DNA-binding efficiencies of 15 µM and 45 µM respectively (Antunes et al., 2004). Tsitsikammamine C exhibited antimalarial activity against *Plasmodium falciparum* at a concentration of 18 nM and was cytotoxic against

HEK-293 cells (IC_{50} = 3.6 μM) (Davies et al., 2012). Furthermore, an analogue of tsitsikammamine A has been shown to inhibit indoleamine 2,3-dioxygenase (IDO1) which forms part of the immune evasion system in human tumours (Dolusic and Frederick, 2013).

Tsitsikammamines A - C (Fig 1.6 A-C) are most structurally similar to tunicate-derived wakayin (Fig 1.6 D) (Copp et al., 1991) as they possess a unique tetracyclic bispyrroloiminoquinone ring moiety (Murugesan et al., 2009). Like tsitsikammamines, wakayin exhibits both cytotoxic and topoisomerase binding activities (Copp et al., 1991; Antunes et al., 2004; Antunes et al., 2005).

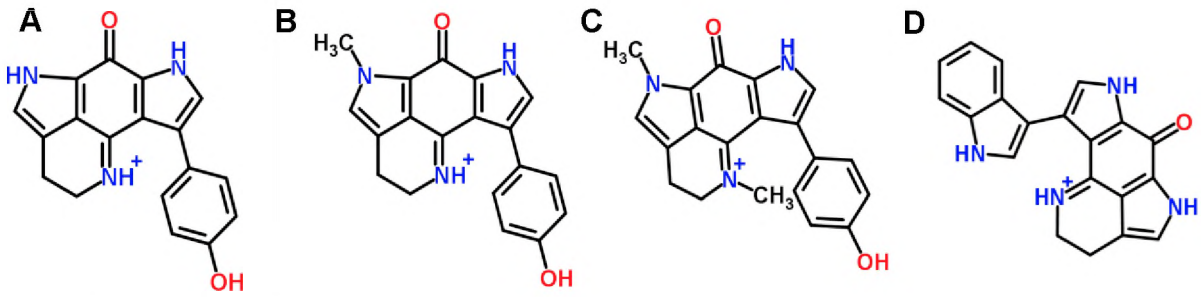


Figure 1.6: Chemical structure of tsitsikammamine A (**A**) and tsitsikammamine B (**B**) isolated from *T. favus*, tsitsikammamine C (**C**) isolated from *Zyzzya* sp. and wakayin (**D**) isolated from a *Clavelina* tunicate. Chemical structures were adapted from Copp et al., 1991, Antunes et al., 2005; and Davies et al, 2012 and drawn using ChemSpider.

Due to observed biological activity of tsitsikammamines and wakayin, but lack of ample supply, several attempts have been made to synthesise these compounds. The core bispyrroloiminoquinone ring structure was successfully synthesised in a single step from a 6-benzylamino pyrroloiminoquinone starter unit (Murugesan et al., 2009). Similarly, partial synthesis of wakayin was achieved using a tryptamine starter unit (Zhang et al., 1998). A series of tsitsikammamines and wakayin analogues were successfully synthesised and were found to exhibit some antitumour and topoisomerase inhibition activities but were not as effective as the natural compounds (Legentil et al., 2006).

Total synthesis for tsitsikammamine A was successfully achieved through a Michael reaction of two indolediones and 2′-amino-1-(4-methoxyphenyl)ethanol (Rives et al., 2009). Cytotoxicity assays demonstrated that the synthesised tsitsikammamine A exhibited potent antitumour activity against U373-GBM, A549-NSCLC and PC-3

prostate cancer cell lines (IC_{50} = 18.1 µM, 15.6 µM and 23 µM respectively) (Rives et al., 2010). Thus far there has been no successful total synthesis of wakayin nor tsitsikammamine B.

1.9.2. Pyrroloiminoquinones isolated from sponges within the genus *Latrunculia, Cyclacanthia* and other Latrunculiidae genera

Latrunculia sponges are prolific producers of bioactive pyrroloiminoquinone compounds (Kelly et al., 2016). Several discorhabdin derivatives have been isolated from an Alaskan *Latrunculia* sp. that exhibit antimalarial, antimicrobial and anti-viral activities (Na et al., 2010).

A suite of brominated pyrroloiminoquinones were isolated from an Antarctic *Latrunculia* sp. that showed unique reversible cholinesterase inhibition (Botic et al., 2017) similar to the inhibition of acetylcholinesterase by extracts isolated from Antarctic *Latrunculia bacogei* and *Latrunculia lendenfeldi* sponges (Turk et al., 2013). Epinadarins A – D were isolated from a *Latrunculia* sp. collected off the pre-Antarctic waters of Crozet Island (D' Ambrosia et al., 1996). A novel pyrroloiminoquinone, atkamine, was isolated from a *Latrunculia* sponge collected in the cold water off the Aleutian Islands (Zou and Hamann, 2013), although no biological activity has been reported.

Other pyrroloiminoquinone compounds isolated from *Latrunculia* species include cytotoxic discorhabdins A, C, G, R and W (Perry et al., 1986; Copp et al., 1994; Yang et al., 1995; Ford and Capon, 2000; Furrow et al., 2003; Lang et al., 2005). South African *L. lunaviridis* and *L. microacanthoxea* sponges have been shown to be sources of cytotoxic and anti-bacterial discorhabdins A and N (Fig. 1.7 A-B) (Antunes et al., 2004). Multiple NRPS-derived compounds including six novel antifungal peptides, callipeltins F-K, were isolated from a *Latrunculia* sp. collected off the shores of the island nation of Vanuatu located in the South Pacific (Sepe et al., 2006; D'Auria; 2007).

The genus *Cyclacanthia* is a recent addition to the Family Latrunculiidae consists of three species, *C. bellae* (previously *Latrunculia bellae*), *C. mzimayiensis* and *C. cloverlyae* (Samaai et al. 2004). The genus is endemic to the South African coastline with *C. mzimayiensis* and *C. cloverlyae* found on the subtropical east coast and *C. bellae* identified further south in the temperate waters of Algoa Bay (Samaai et al.

2004). Makaluvamine A (Fig. 1.7 C) and two novel discorhabdins - 1-methoxydiscorhabdin D (Fig. 1.7 D) and 1-aminodiscorhabdin D were isolated from *C. bellae* and exhibit potent cytotoxicity against human colon carcinoma HCT-116 cells (IC_{50} = 1.1 μM, 0.12 μM, 0.23 μM respectively) (Antunes et al., 2005). Other compounds isolated from *C. bellae* include damirone B, makaluvic acid A, makaluvamine C, and discorhabdins G and N (Antunes et al., 2004).

Sceptrella sponges were first identified in 1870 (Schmidt, 1870) are pre-dominantly found in deep-water and are distributed in the northern and western Atlantic Ocean (Keyzers et al., 2005). A total of fourteen pyrroloiminoquinone compounds, including two novel discorhabdin derivatives, were isolated from a dark green *Sceptrella* sp. collected off the shore of Gageodo, Korea, at a depth of 20m (Jeon et al., 2010). The two new discorhabdins (−)-3-dihydrodiscorhabdin D (Fig. 1.7 E) and (−)-discorhabdin Z (Fig. 1.7 F) exhibited moderate cytotoxicity against leukaemia K562 cells and were capable of inhibiting growth of several bacterial species, most likely through the inhibition of the sortase A enzyme (Jeon et al., 2010). A *Negombata* sp. was collected in Port Campbell, Australia, from which novel discorhabdin R was isolated, which showed inhibition activity against both Gram-positive and Gram-negative bacteria (Ford and Capon, 2000).

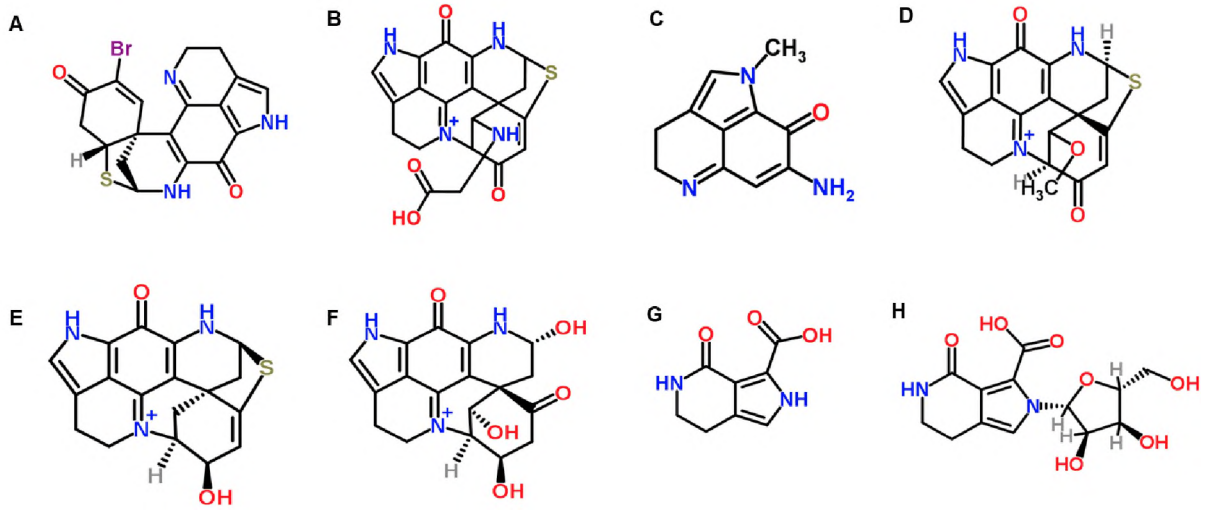


Figure 1.7: Chemical structures of discorhabdin A (**A**) and discorhabdin N (**B**) isolated from South African *Latrunculia algoaensis*, makaluvamine A (**C**) and 1-methoxydiscorhabdin D (**D**) isolated from *Cyclacanthia bellae*, (−)-3-dihydrodiscorhabdin D (**E**) and (−)-discorhabdin Z (**F**) isolated from a Korean *Sceptrella* sponge, makaluvic acid C (**G**) N-1-β-d-ribofuranosylmakaluvic acid C (**H**) from *Strongylodesma aliwaliensis*. Chemical structures were adapted from Keyzers et al., 2004, Keyzers et al, 2005, Antunes et al., 2005 and Jeon et al., 2010 and drawn using ChemSpider.

Strongylodesma aliwaliensis, a South African sponge species is the producer of four novel pyrroloiminoquinone metabolites, namely, makaluvic acid C (Fig 1.7 G) N-1-β-d-ribofuranosylmakaluvic acid C (Fig. 1.7 H), N-1-β-d-ribofuranosyldamirone C and N-1-β-d-ribofuranosylmakaluvamine I as well as three previously known makaluvamines (Keyzers et al., 2004; Keyzers et al, 2005). When tested against oesophageal cancer WHCO-1 cells, all novel compounds, save makaluvic acid C, showed moderate to strong cytotoxic activity (Keyzers et al., 2005). Both N-1-β-d-ribofuranosyldamirone C and N-1-β-d-ribofuranosylmakaluvamine I have since been successfully synthesised through oxidative cyclization of an aryl-methoxyamide precursor by a hypervalent iodine species (Inoue et al., 2010).

1.10. Pyrroloiminoquinones isolated from other organisms

Pyrroloiminoquinone compounds have also been isolated from marine sponges in the *Batzella*, *Prianos* and *Zyzzya* genera as well as ascidians, myxomycetes and actinomycetes (Davis et al., 2012; Copp and Ireland, 2001; Ishibashi et al., 2001; Hughes et al., 2009b; Davies et al., 2012; Lin et al., 2017). Makaluvamines A and B were isolated from a cultured *Didymium bahiense* myxomycete (Ishibashi et al., 2001) and makaluvamine I and damirone C have been isolated from closely related slime mold *Didymium iridis* (Ishibashi, 2005).

Makaluvamine G was first isolated from a *Histodermella* sponge collected in Indonesia (Carney and Scheuer, 1993) and later re-isolated from a *Zyzzya* sponge collected from Rodda Reef, Australia (Davies et al., 2012). Several makaluvamines have been isolated from *Zyzzya* sponges and were reported to be cytotoxic including, makaluvamine A, first isolated from a *Zyzzya* cf. *marsailis* sponge collected in Fiji (Radisky et al., 1993), makaluvamines H-M from *Zyzzya fuliginosa* collected from Pohnpei (Schmidt et al., 1995), makaluvamine N from *Z. fuliginosa* collected in the Philippines (Venables et al., 1997) and makaluvamine P from *Z. fuliginosa* collected off the island nation of Vanuatu (Casapullo et al., 2001). Structurally related pyrroloiminoquinone compounds within the batzelline and damirone classes, including novel compounds, batzelline D and isobatzelline E, were isolated from *Z. fuliginosa* and exhibited varying levels of cytotoxicity (Chang et al., 2002a; Dijoux et al., 2005).

The variation in cytotoxicity was hypothesised to be relative to the saturation levels and substitution patterns observed on the core ring system (Dijoux et al., 2005).

Sponges within the genus *Batzella* are also prolific producers of cytotoxic pyrroloiminoquinone compounds including discorhabdins S – U, secobatzellines A and B and isobatzellines A – C from various Caribbean *Batzella* sp. as well as batzellines A – C from a deep water *Batzella* sp. collected from the Bahamas (Sun et al., 1990; Gunasekera et al., 1999; Gunasekera et al., 2003; Sakemi et al., 1989).

1.11. Research Problem Identification

Suites of structurally related pyrroloiminoquinone compounds have been isolated from South African sponges within the Latrunculiidae family. Tsitsikammamines A and B are produced exclusively in South African *Tsitsikamma favus* and *Tsitsikamma nguni* sponges, whilst structurally similar tsitsikammamine C has been isolated from an Australian *Zyzzya* sponge. As these sponge species are both geographically and phylogenetically distinct, it is therefore proposed that these structurally similar bioactive products are produced by conserved bacterial communities within these sponge species. As mentioned previously, pyrroloquinoline compounds have been shown to include tryptophan during biosynthesis (Jordan and Moore, 2016) and this amino acid has been hypothesised as the potential starting unit for the biosynthesis of pyrroloiminoquinone compounds (Antunes et al., 2005). The mechanisms and gene clusters involved in the biosynthesis of pyrroloquinoline compounds have been inspected closely and may aid in the elucidation of biosynthetic gene clusters responsible for the biosynthesis of structurally related pyrroloiminoquinone compounds. The aim of the research presented in this thesis was to identify candidate biosynthetic pathways that potentially encode pyrroloiminoquinones, particularly tsitsikammamines, in *T. favus* sponges. Structurally similar pyrroloquinolines are biosynthesised by actinobacteria. Therefore, the first objective of this study was to culture *T. favus* associated actinomycetes and assay for antibacterial activity. Genomes of isolates that exhibit inhibition of test bacteria will be sequenced, assembled, annotated and inspected for potential biosynthetic gene clusters that may be responsible for the production of secondary metabolites. The bacterial

communities of latrunculid sponges will be assessed to identify patterns of conservation in the associated microbiota. Conservation of bacterial communities indicates that these populations may be of benefit to sponge health through production of defensive bioactive molecules and thus will be considered candidate bacteria for the production of pyrroloiminoquinone compounds in latrunculid sponges. Shotgun metagenomic sequencing of genomic DNA (gDNA) extracted from a *T. favus* sponge will provide insight into the various functional roles performed by *T. favus*-associated bacteria. Close inspection of clustered bacterial genomes from this metagenomic data will enable the identification of conserved biosynthetic motifs and aid in the elucidation of potential pyrroloiminoquinone-producing biosynthetic gene clusters.

Chapter 2 – Identification of gene clusters producing antimicrobials in *Kocuria kristinae* isolated from *Tsitsikamma favus*

2.1. Introduction

Algoa Bay is located off the coast of Port Elizabeth, South Africa and is considered one of the most important bays in Africa as both a marine biodiversity hotspot as well as a key region for tourism and socio-economic development (Dorrington et al., 2017). Marine sponges and tunicates found in this bay are present in abundance and have been identified as sources of several bioactive secondary metabolites with pharmaceutical potential (Antunes et al., 2004; Antunes et al., 2005). Latrunculid sponges are one such example, from which cytotoxic bis-pyrroloiminoquinone tsitsikammamines were isolated to found to be produced exclusively by *Tsitsikamma favus* sponges (Hooper et al., 1996; Antunes et al., 2004).

Actinomycetes are known to produce approximately two thirds of known antibiotics and marine actinobacteria have become increasingly prominent with the isolation of structurally novel metabolites (Valliappan et al. 2014) including abyssomicin C, marinopyrroles, manzamine A, 8-hydroxymanzamine A, salinsporamides and mandelalides (Riedlingera et al. 2004; Hughes et al. 2008; Peraud, 2006; Udwary et al. 2007; Sikorska et al. 2012). Due to this reputation of actinomycetes as prolific sources of novel bioactive metabolites efforts were made, in a previous study, to identify culturable actinomycetes from latrunculid sponges sourced from Algoa Bay (Walmsley, 2013, PhD thesis). A total of 13 actinomycete isolates were cultured from latrunculid sponges, *Tsitsikamma favus*, *Tsitsikamma scurra* and *Latrunculia algoaensis* and included 7 *Kocuria* species (Walmsley, 2013, PhD Thesis). *Kocuria kristinae* isolate RUSp02-3 (renamed RUTW2-3) isolated from a *T. favus* sponge showed complete inhibition of *Mycobacterium aurum* A+ (Walmsley, 2013, PhD Thesis). This was an unexpected result as *Kocuria* species are better known as medical pathogens than for their production of bioactive secondary metabolites (Horino et al., 2016; Kimura et al., 2017).

However, a novel thiazolyl cyclopeptide called kocurin has been isolated from both terrestrial and marine *Kocuria* species. Kocurin was first isolated from *Kocuria palustris* strain F-276,354, cultured from a *Xestospongia* sponge species (Mahajan et al., 2013; Palomo et al. 2013). The chemical structure of kocurin (Fig. 2.1) was elucidated using various chemical and spectroscopic methods and was found to match a previously identified compound, PM181104, which was isolated from a marine *Kocuria* sp. (Mahajan et al., 2013; Martin et al., 2013). Kocurin was later isolated from a *K. rosea* strain cultured from intertidal sediment (Linares-Otoya et al., 2017).

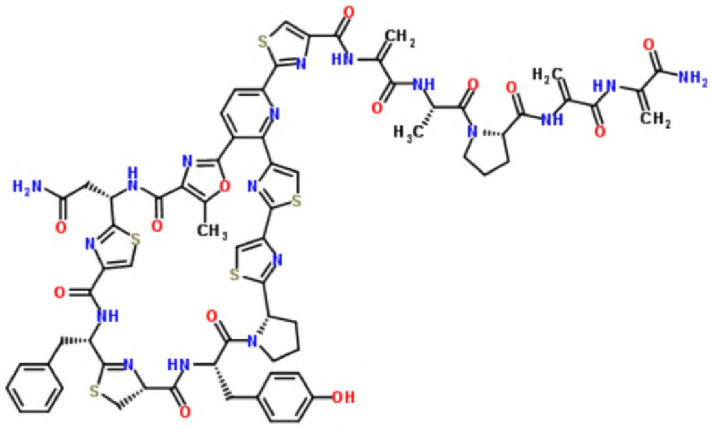


Figure 2.1: The chemical structure of kocurin produced by marine *Kocuria* species. Chemical structure adapted from Martin et al., 2013 and drawn using ChemSpider.

Kocurin was found to inhibit the growth of methicillin-resistant *Staphylococcus aureus* (MIC of 164 nM), *Bacillus subtilis* and *Enterococcus faecium* (Zones of clearance observed with kocurin at 5.27 µM) (Palomo et al., 2013; Martin et al., 2013). Its potent activity against methicillin-resistant *Staphylococcus aureus* prompted an investigation into the elucidation of the biosynthetic origin of kocurin. The gene cluster responsible for the production of kocurin was shown to be a 12kb lantipeptide pathway consisting of 9 ORFs and was successfully heterologously expressed in host bacterium *Streptomyces coelicolor* M1146 (Linares-Otoya et al., 2017). The *Kocuria* spp. from which kocurin was isolated were collected from different environments and geographically distinct locations (Peru, United States of America and Spain) which supports the hypothesis that horizontal gene transfer is most efficient and therefore favoured in closely related bacterial species (Jensen et al., 2016; Letzel et al., 2017).

As pyrroloquinoline compounds lymphostin and ammosamides are produced via a lantipeptide pathway, *Kocuria* spp. isolated from latrunculid sponges which exhibited bacterial growth inhibition activity warranted further investigation. The research described in this chapter aimed to develop genomic and metagenomic capacities for application in the identification of biosynthetic gene clusters in sponge-associated bacteria. Actinobacteria are prolific producers of a vast number of natural products and thus were selectively isolated from *T. favus* (TIC2009-001 and TIC2009-002) sponges and surrounding sediment and screened for anti-microbial activity. The genomes of two *Kocuria* isolates were sequenced, assembled, and annotated. Biosynthetic gene clusters encoding a putative siderophore and a putative terpene were identified through gene sequence homology and conservation. Siderophores are small molecules produced by bacteria and fungi that have the capacity to chelate ferric iron from the environment (Kohira et al., 2016) and some have been shown to exhibit antimicrobial activity (Shah et al., 1992; Wang et al., 2009). Terpenes are a class of volatile hydrocarbons most commonly isolated from plants but have been found in other kingdoms including prokaryotes and fungi (Yamada et al., 2014). Although several bacterial terpenes have been identified, their biological roles remain largely unknown but do include some observed antimicrobial activity (Tyc et al., 2017).

Data from the research described in this chapter has been submitted for publication in the journal Marine Biotechnology. The manuscript is currently under review.

Siderophore activity in a marine sponge-associated *Kocuria kristinae* isolate: a potential role in sponge iron assimilation? Waterworth, S.C., Walmsley, T.A., Matongo, T., Davies-Coleman, M.T., Dorrington, R.A. Marine Biotechnology, Submitted.

2.2. Methods and Materials

2.2.1. Sponge collection and identification

Sponges and sediment were collected in May 2009 by Dr Tara Walmsley (Former PhD student in the Dorrington laboratory, currently a Senior Lecturer in Biotechnology at Vaal University of Technology) at an approximate depth of 30m at Evan's Peak, Algoa Bay, South Africa. Sponge and sediment samples were stored in Ziploc bags with

surrounding seawater and kept at 4 °C for 3 hours before processing could take place. Sponge specimens TIC2009-001 and TIC2009-002 were morphologically identified as *Tsitsikamma favus* (Samaai and Kelly, 2002) by Dr Shirley Parker-Nance and by sequencing of a partial fragment of the 28S rRNA gene sequence (Walmsley et al, 2012) (Please see Appendix A – Sponge Collection Metadata)

2.2.2. Isolation of sponge-associated actinomycetes with antimicrobial activity

Twenty-four actinomycete isolates were cultured from *T. favus* sponges of which seven were identified as *Kocuria* species (Walmsley, 2013, PhD Thesis). Isolate RUTW1-3 was isolated from *T. favus* specimen TIC2009-001, isolates RUTW2-2 and RUTW2-3 were isolated from *T. favus* specimen TIC2009-002 and RUTW4-5 was isolated from the sediment. Isolates were grown from macerated sponge tissue on ISP2 agar plates for two weeks at 28 °C.

2.2.3. Bacterial culture conditions

Kocuria isolates were cultured for 48 hours at 28 °C in YEME broth media made up in artificial seawater (ASW) (See Appendix B- Media) and supplemented with nalidixic acid (10 µg/mL) and cycloheximide (10 µg/mL) for the suppression of bacteria with trailing growth and fungal growth respectively (Porter et al., 1959; Alferova et al., 1989). Genomic DNA (gDNA) was extracted using the ZR Fungal/Bacterial DNA MiniPrep kit (D6005, Zymo Research). The concentration of gDNA was quantified using a NanoDrop Lite Spectrophotometer (ThermoScientific)

Full length 16S rRNA gene fragments (1500bp) were PCR- amplified from each *Kocuria* isolate using the universal primer pair 8-27F (5‘-AGA GTT TGA TCM TGG CTC AG-3‘) and 1492R (5‘-TAC GGY TAC CTT GTT ACG ACT T-3‘). PCR amplification reactions consisted of 0.3 µM of each primer, 0.75 mM dNTPs, 1X Buffer with MgCl2, 1 U HiFi Taq polymerase (Kapa Biosystems) and 4ng of gDNA template. Thermal cycling parameters were as follows: 94 °C for 5 min, 25 cycles at 92 °C for 30 sec, 48 °C for 2 min, and 72 °C for 90 sec, followed by 72 °C for 5 min. Amplicons were ligated into the pGEM-T Easy vector and the insert sequences were determined using Sanger sequencing. The 16S rRNA gene fragment sequences were deposited in Genbank (Accession numbers HQ336741-3, KX055834).

Phylogenetic analysis of the isolates based on the 16S rRNA was performed using ClustalW in MEGA7 to align the sequences and phylogeny was inferred using the Maximum-likelihood method with 500 bootstrap replicates.

2.2.4. Whole genome shotgun sequencing

gDNA from *Kocuria* isolate RUTW2-3 was mechanically fragmented and sequenced using the Roche 454 GS pyrosequencing technology. A total of 139 123 reads, with an average length of 463.4 nucleotides, were assembled into contiguous sequences (contigs) using the GS *De Novo* Assembler (Version 3.0; Roche). Automated annotation of the contigs was performed using myRAST (Version 36; SEED Servers) and the online antiSMASH tool (Weber et al., 2015) was used to identify potential secondary metabolite biosynthetic pathways. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank (Accession No: NZ_LJBJ00000000.2).

gDNA extracted from *Kocuria* isolate RUTW4-5 was mechanically fragmented and sequenced using the Ion PGM Platform, generating a total of 240 153 reads, with an average length of 230.4 nucleotides, which were assembled into contigs using the GS Reference Mapper (Version 3.0; Roche) using the contigs of *Kocuria* RUTW2-3 as a reference. The data was deposited at DDBJ/ENA/GenBank (Accession No: LWGZ00000000).

2.2.5. Identification of biosynthetic gene clusters

Gene clusters identified in antiSMASH were manually inspected through individual analysis of open reading frames (ORFs). The BLASTp algorithm was used to infer putative gene functions and substrate specificity of NRPS adenylation domains were identified using NRPS Predictor2 (Rottig et al., 2011). To determine whether the identified gene clusters of interest were present in all *Kocuria* isolates PCR amplification of core biosynthetic genes was performed with the following parameters in combination with unique primers designed specifically for this study:

A fragment of the condensation domain of NRPS1 (1307 bp) was amplified using 0.3 mM of primer pair SW1F (5'-ATG CGT TTG ACC AAC ATC ACC-3') and SW1R (5'-TCA GGC GGC CAG CTC CTG ATG-3'), 1 U HiFi Taq polymerase (Kapa Biosystems)

and 10 ng of gDNA template with thermal cycling parameters: 98 °C for 5 min, 30 cycles at 98 °C for 30 sec, 58 °C for 45 sec and 72 °C for 90 sec, followed by 72 °C for 5 min.

A fragment of the condensation domain of NRPS2 (930 bp) was amplified using 0.3 mM of primer pair SW2F (5'-CGG ACA CGG CCA CCG CAG C -3') and SW2R (5'-CCC GAT CAC CGA GGC CGC CC-3'), 1 U HiFi Taq polymerase (Kapa Biosystems) and 10 ng of gDNA template with thermal cycling parameters: 98 °C for 5 min, 30 cycles at 98 °C for 45 sec, 64 °C for 30 sec and 72 °C for 1 min, followed by 72 °C for 5 min.

A fragment of the phytoene synthase gene from the terpene gene cluster was amplified using 0.4mM of primer pair SW4F (5' - AGT ACG CGC CGA GGA AGC CC - 3') and SW6R (5' - GAT GTG CCT GCG GAT CTT CGA G - 3'), 0.5 U Kapa Taq ReadyMix (Kapa Biosystems) and 8ng of gDNA template with thermal cycling parameters: 98 °C for 5 min, 30 cycles at 98 °C for 30 sec, 61 °C for 30 sec and 72 °C for 45 sec, followed by 72 °C for 5 min.

2.2.6. Anti-microbial and siderophore activity assays

Kocuria isolates were grown overnight in YEME-ASW broth supplemented with cycloheximide (10 µg/ml) and nalidixic acid (10 µg/ml) and incubated at 30 °C for two days shaking at 200rpm. A single line of each of the four isolates was streaked onto YEME-ASW agar and incubated at 30 °C for nine days.

Antimicrobial activity of the *Kocuria* isolates was assessed through sloppy agar (0.7%) overlays of 2YT agar supplemented with overnight culture of *Mycobacterium aurum* A+, and sloppy agar (0.7%) overlays of Luria agar supplemented with overnight cultures of *E. coli* K12, *Bacillus subtilis*, *Streptomyces albidoflavus*, *Staphylococcus aureus*. Overlays were incubated at 28 °C (*M. aurum* A+ overlays) and 37 °C (all other test isolates) overnight and inspected for zones of clearance the following morning. Siderophore activity was detected using a Chrome Azurol S (Alexander and Zuberer, 1991) modified sloppy agar (0.7%) overlay (Perez-Miranda et al. 2007) in triplicate. The presence of orange halos indicated siderophore activity.

To determine if observed antimicrobial activity of the *Kocuria* isolates was dependent on iron chelation (and by implication, siderophore production), *Kocuria* isolates were inoculated onto YEME-ASW agar as described above. Following nine days of growth at 30 °C, *Kocuria* isolates were overlaid with sloppy (0.7%) 2YT agar inoculated with overnight cultures of *M. aurum* A+ and supplemented with $FeCl_3$ at final concentrations ranging from 25 µM to 175 µM (adapted from Sayyed et al., 2005). Overlays were prepared in duplicate, incubated at 37 °C overnight and inspected for zones of clearance the following morning.

2.3. Results

2.3.1. Confirmation and extension of observed anti-microbial activity.

Kocuria isolates RUTW1-3, RUTW2-2, RUTW2-3 and RUTW4-5 were identified as *Kocuria kristinae* based on full length 16S rRNA gene sequence (Fig. 2.2), in agreement with identification reported previously (Walmsley 2013, PhD thesis). Isolates shared 99% sequence identity with one another indicating that they are likely within the same species. Furthermore, all four *K. kristinae* isolates displayed complete inhibition of *M. aurum* A+ growth, as well as inhibition of *E. coli* K12, *B. subtilis*, *S. albidoflavus* and *S. aureus* growth on agar plates.

2.3.2. Genome sequence analysis of *K. kristinae* isolates RUTW2-3 and RUTW4-5

To identify potential secondary metabolite biosynthetic pathways the genomes of *K. kristinae* isolates *viz.* RUTW2-3 (sponge-associated) and RUTW4-5 (sediment-associated), were determined by whole genome shotgun sequencing. A total of 64.47 Mb of data was obtained with approximately 10-fold coverage for strain RUTW2-3. The reads were assembled into forty-nine contigs generating a partial genome approximately 2.36 Mb in size with a GC content of 71.8%. A total of 1134 contigs covering approximately 2.1 Mb, were assembled for isolate RUTW4-5 using the contigs of RUTW2-3 as a reference. These genomes are comparable to other publicly available *K. kristinae* genome which range from 2.21 – 2.41 Mb in size with GC contents between 71.1 – 71.9 %. Automated annotation of the genome of RUTW2-3

using myRAST identified 2111 potential protein encoding genes, 47 tRNA genes and 5 rRNA genes. AntiSMASH analysis identified a putative terpene cluster on Contig13 and two NRPS gene clusters on Contig1 and Contig2.

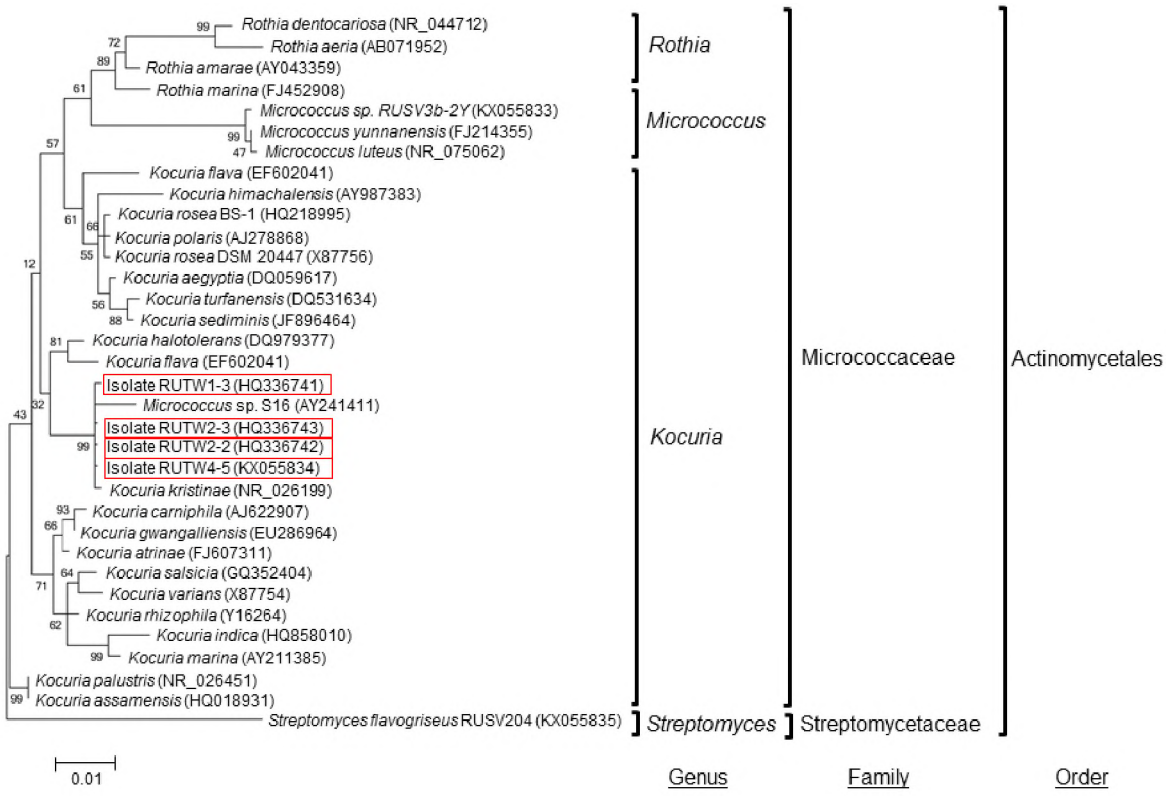


Figure 2.2: Phylogeny of marine *K. kristinae* isolates relative to the Micrococcaceae family and rooted within the order Actinomycetales based on alignment of 16S rRNA gene sequence. Clustering of species within genera, family and order are indicated. Isolates identified in this study are boxed in red. The 16S rRNA gene sequences were aligned using ClustalW in MEGA6 using the Maximum Likelihood method with 500 bootstrap replicates. Bootstrap values are indicated on branch nodes. The scale bar indicates the average number of amino acid substitutions per site.

2.3.3. Characterization of the putative terpene biosynthetic gene cluster

The putative terpene gene cluster, spanning approximately 6100 bp, comprises eight ORFs, six of which are clustered together on Contig13, with two additional ORFs found on Contig2 and Contig24 (Fig. 2.3). The gene cluster includes two consecutive lycopene cyclases, an isorenieratene synthase, a phytoene synthase, a phytoene dehydrogenase, cytochrome P450, a polyprenyl synthase and an isopentyl pyrophosphate isomerase (Fig. 2.3). The gene cluster shares greatest amino acid sequence identity (35%) with a carotenoid pathway present in *Brachybacterium phenoliresistant*, *Brevibacterium linens* and *Brevibacterium casei* strains (Fig. 2.3).

The terpene pathway was also present on the genome of *K. kristinae* RUTW4-5 (Fig. 2.3) with amino acid identity with RUTW2-3 ORFs ranging from 95 – 99%.

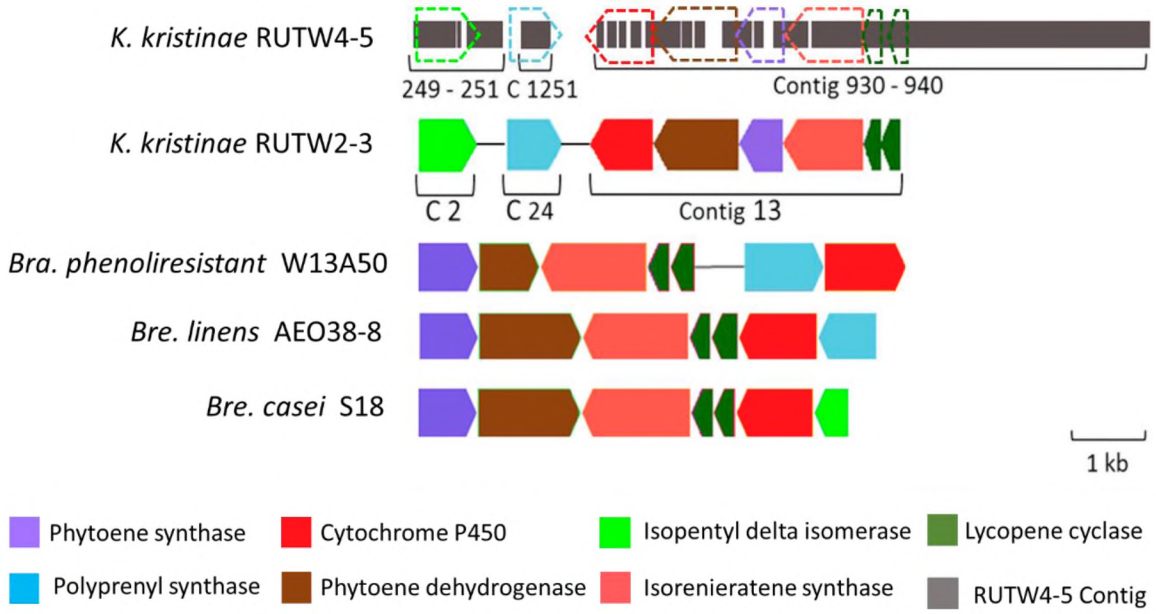


Figure 2.3: Comparative gene cluster organization of putative terpene gene cluster identified in *K. kristinae* isolates RUTW2-3 and RUTW4-5 relative to previously identified homologous clusters. Transcriptional orientation of ORFs is represented by arrows coloured by putative function. Regions of homologous sequences spanning multiple aligned contigs that correspond to identified ORFs in RUTW4-5 of the proposed terpene pathway are represented by arrows described with dashed lines. The contig(s) on which each ORF is located, has been indicated. A scale bar is provided for relative size. Genbank Accession No.: *Brachybacterium phenoliresistens*: NZ_JDYK00000000.1; *Brevibacterium linens*: NZ_JTJZ00000000.1; *Brevibacterium casei*: NZ_AMSP00000000.1

2.3.4. Characterisation of a putative siderophore biosynthetic gene cluster

A putative NRPS gene cluster (NRPS2) that consisted of 14 ORFs located on Contig0001 (3 ORFs) and Contig0002 (11 ORFs) and spanning approximately 3 Kb and 17 Kb, respectively, was identified in isolate RUTW2-3. A free-standing condensation domain (NRPS1) was identified upstream of the cluster, flanked by a hypothetical protein and an α/β hydrolase (Fig. 2.4). Sequence homology analyses revealed that the genes within the NRPS2 cluster share the greatest amino acid sequence identity (48% - 71%) with an orphan pathway identified in *Rothia mucilaginosa*. Furthermore, the NRPS2 gene cluster shared 21% amino acid identity with gene clusters present in Actinobacteria spp., *Actinosynnema mirum* and *Amycolatopsis* sp. AA4.

Further inspection of the RUTW2-3 genome sequence revealed the presence of direct repeats flanking a large section of the NRPS2 cluster, including ORFs “g” through to

"F" (Fig. 2.4). A single putative transposase was identified on Contig44, the closest homologue of which was an IS5 transposase identified in a *K. kristinae* strain isolated in China (MODZ01000004.1).

Analysis of the partial genome sequence of isolate RUTW4-5 revealed conservation of all the ORFs in the predicted NRPS2 gene cluster (encoded on Contigs217 – 233). The putative transposase and direct repeats were also conserved. Homologues of the free-standing condensation domain and flanking genes were also identified on Contigs 9 – 11 (Fig. 2.3) and the amino acid sequence identity between the NRPS2 gene clusters of RUTW2-3 and RUTW4-5 ranged from 98 – 100%.

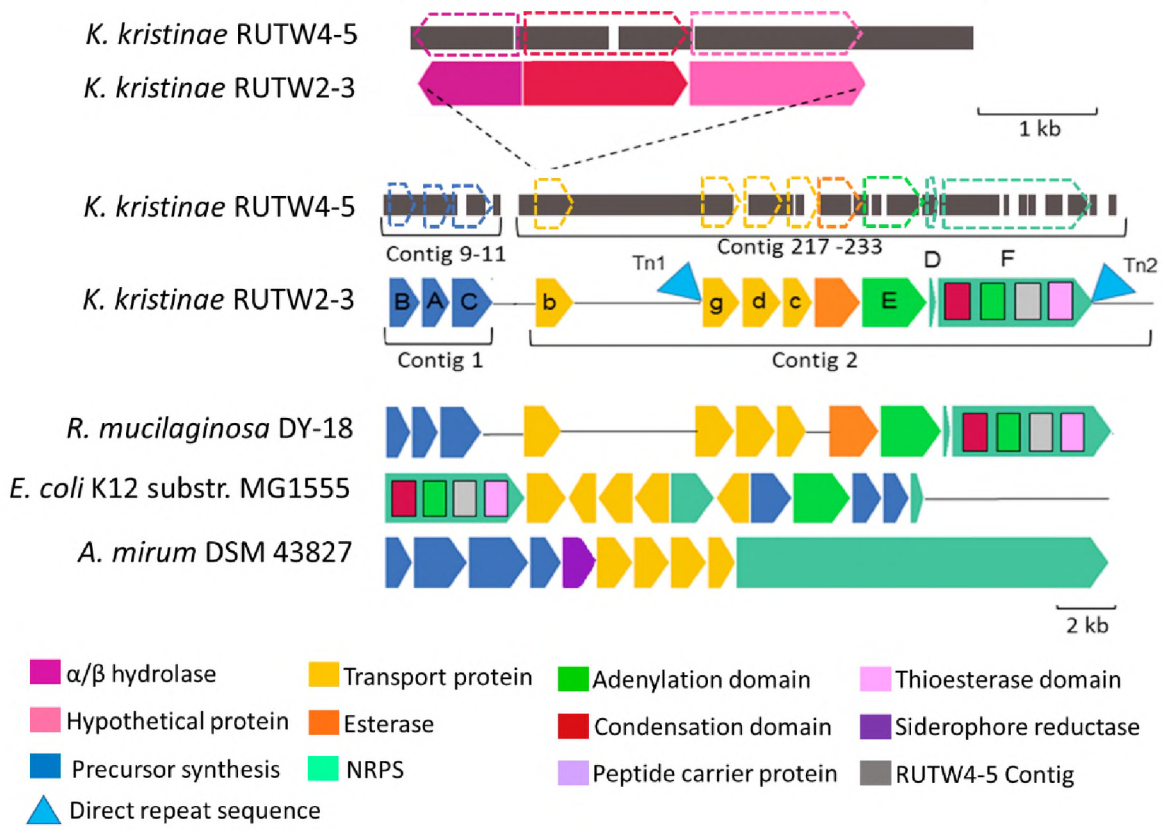


Figure 2.4: Comparative gene cluster organization of putative NRPS gene clusters identified in *K. kristinae* isolates RUTW2-3 and RUTW4-5 relative to previously identified homologous clusters. Transcriptional orientation of ORFs is represented by arrows coloured by putative function. The contig on which each ORF is located, has been indicated. Regions of homologous sequences spanning multiple aligned contigs that correspond to identified ORFs in RUTW4-5 of the proposed terpene pathway are represented by arrows described with dashed lines. Scale bars are provided for relative contig and ORF size.

Conserved adenylation domains were identified within NRPS2 genes "E" and "F". Homology analysis of conserved regions within the adenylation domain substrate recognition sites using NRPS Predictor2 (Rottig et al., 2011) predicted that genes "E" and "F" would recognize and activate dihydroxybenzoate (DHB) and serine monomers, respectively. Homologous adenylation domains in *E. coli* (Fig. 2.4) that also recognize DHB and serine, are involved in the biosynthesis of a siderophore, enterobactin. Co-linearity rules (Marahiel et al., 1997) enable prediction of a putative core structure for the NRPS product (Fig. 2.5 A), which matches one of three identical structural moieties of enterobactin (Fig. 2.5 B), providing further evidence that the putative NRPS2 cluster produces a siderophore-like compound.

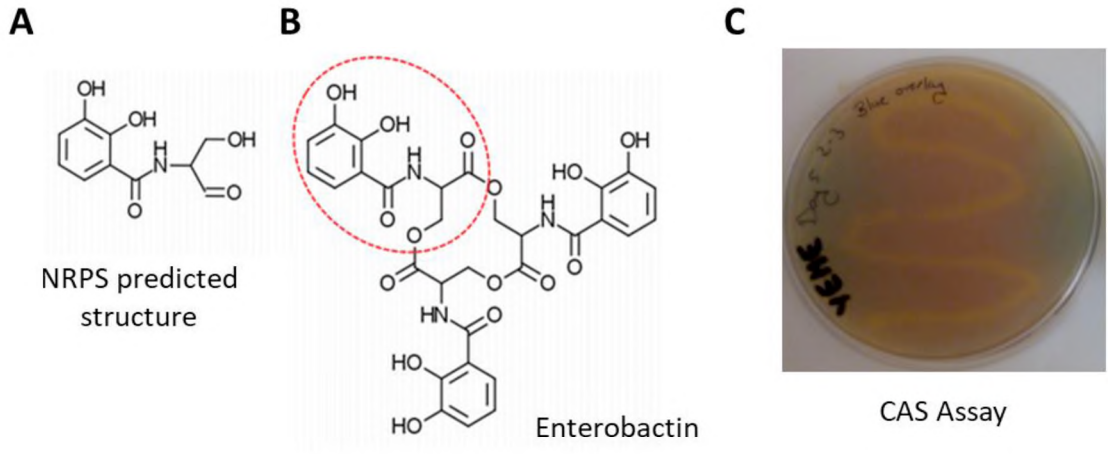


Figure 2.5: Siderophore production in marine *K. kristinae* isolates. (**A**) The predicted structure generated in antiSMASH for the identified NRPS gene cluster following collinearity rules without tailoring. (**B**) Structure of enterobactin. The region matched to the predicted NRSP product has been indicated with red circle. (**C**) Siderophore activity observed from RUTW2-3 in CAS-agar assay. Blue agar indicates media in which the CAS dye is complexed with iron and pink regions indicate regions of complexed iron uptake by siderophore activity.

Siderophore activity was confirmed in all four *K. kristinae* isolates, using CAS-agar overlay assays (Fig. 2.5C). To determine whether siderophore activity resulting in Fe^{3+} loss was responsible for anti-microbial activity, growth inhibition assays with $FeCl_3$ supplementation up to a final concentration of 175 mM, were conducted. No reduction in growth inhibition of test bacterial strains was observed.

Finally, PCR analysis was used to confirm the presence of two putative secondary metabolite biosynthetic pathway gene clusters in the remaining two *K. kristinae* isolates (RUTW1-3 and RUTW2-2) by amplification of key gene fragments using specific primers designed by the author for this study. A sponge-associated

Micrococcus sp. RUSV3b-2Y (See Fig. 2.2), which exhibits anti-microbial activity, but not siderophore activity, was included as a control. Gene fragments from all identified putative siderophore and terpene gene clusters were successfully amplified from the gDNA of all four *K. kristinae* isolates but were absent in the outlier organisms (Fig.2.6).

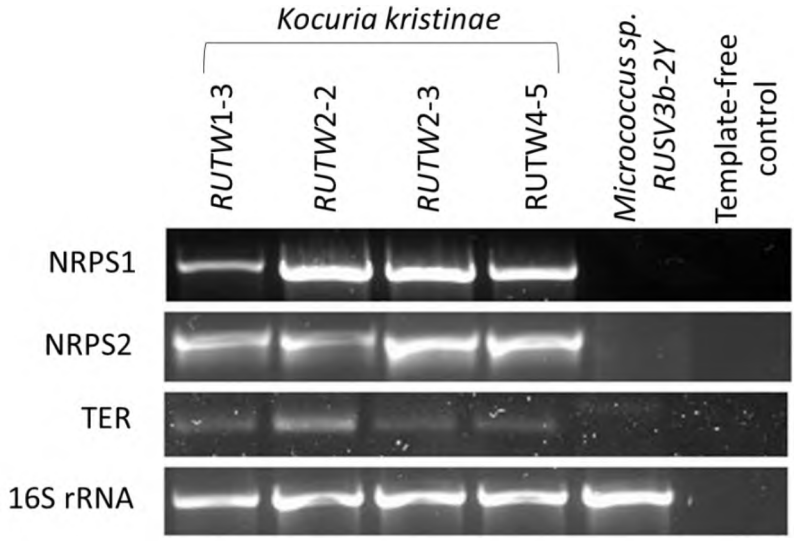


Figure 2.6: Conservation of key gene fragments in identified biosynthetic clusters in *K. kristinae* isolates. PCR amplification products of the free-standing condensation domain of NRPS1 (1307 bp), the condensation domain of gene "F" of NRPS2 (930 bp) and a fragment of the phytoene synthase gene of the terpene (TER) pathway (426 bp) from four *K. kristinae* isolates and a sponge-associated *Micrococcus* sp.. Controls included as indicated (16S rRNA gene).

2.4. Discussion

The primary aim of the research presented in this chapter was to determine the biosynthetic origin of antimicrobial activity in *K. kristinae* strains isolated from *T. favus* sponges and associated sediment. This study also was also important to develop the genomic capacity for identifying putative biosynthetic pathways that would be used for the metagenomics studies described later in this dissertation (Chapters 4 - 6).

K. kristinae is known in literature as a medical pathogen (Horino et al. 2016; Kimura et al. 2017) and this is the first report of *K. kristinae* isolated from a marine environment. Marine natural products have been isolated from other *Kocuria* species, including the novel thiozoyl peptide, kocurin, which exhibits inhibitory activity against methicillin-resistant *Staphylococcus aureus* (Martin et al., 2013; Palomo et al. 2013;

Calcutt and Foecking, 2015) and novel the polysaccharide, kocuran, which exhibits immunosuppressive activity (Kumar and Sujitha, 2014).

Two gene clusters potentially responsible for the production of an antimicrobial compound, were identified on the genomes of *K. kristinae* RUTW2-3 and RUTW4-5. The presence of these two gene clusters in two additional sponge-associated strains, RUTW1-3 and RUTW2-2, was confirmed by PCR amplification of core gene sequences. The NRPS2 gene cluster shared greatest sequence homology with known clusters identified in *Actinosynnema mirum* and *Amycolatopsis* sp. AA4. which are responsible for the production of siderophores mirubactin and amychelin, respectively (Giessen et al., 2012; Seyedsayamdost et al., 2011). This suggests that the putative NRPS2 gene cluster may produce a similar siderophore-like compound. Furthermore, the NRPS2 gene cluster was situated amongst a collection of genes encoding the large and small subunits of ribosomal proteins. The positioning, presence of direct repeats and a putative transposase supports a hypothesis that the NRPS2 gene cluster may have been acquired via transposition. Siderophore activity was confirmed in all four marine *K. kristinae* isolates but could not be definitively linked to the identified gene cluster. Conclusive evidence for this could be achieved using mutant strains from which the pathway has been deleted/disrupted or conversely efforts could be made to heterologously express the identified pathway to determine whether a siderophore is biosynthesised.

The observed anti-bacterial activity of *K. kristinae* isolates could be a result of expression of either the terpene or the siderophore biosynthetic pathways or, alternatively a different gene cluster that was not identified during this study due to the lack of genome completion. It was thought that *K. kristinae* isolates were inhibiting the growth of test bacteria by starving them of ferric iron through the action of the siderophore but increasing ferric iron concentrations in growth media did not affect the inhibitory ability of the *K. kristinae* isolates. A cephalosporin beta-lactam antibiotic with antimicrobial activity against Gram-negative bacteria (Kohira et al., 2015) has been shown to exploit ferric iron uptake mechanisms to enter bacterial cells (Ito et al., 2016). This may also be the case for the product of the NRPS2 gene cluster, but further chemical analysis to elucidate the structure of the compound would be required to confirm this. It is possible that the *K. kristinae* isolates increase the environmental iron availability through siderophore production and utilise the unknown terpene product,

like the cephalosporin antibiotic, to exploit the ferric iron uptake mechanisms in the test bacteria in order to inhibit their growth.

The conservation of the NRPS2 cluster in all four *K. kristinae* isolates suggests that it may play an important role in ensuring bacterial survival in the marine environment. Siderophores are produced in abundance by a wide variety of marine bacteria in response to limited free iron availability (Boiteau et al., 2016). It is thus possible that the *K. kristinae* isolates use the NRPS2 gene cluster to facilitate ferric iron uptake. An alternate source of the antimicrobial activity may be the product of the terpene gene cluster, which was also conserved in all four *K. kristinae* isolates. As with NRPS2, homologues of the terpene gene cluster could not be identified in any other *Kocuria* genomes currently available.

An association with *Kocuria* and *Micrococcus* sp. has been reported for a broad range of sponge species from different geographical locations (Beepat et al. 2016, Laport et al. 2016; Bramhachari et al. 2016, Cita et al. 2017). The *K. kristinae* isolates from *T. favus* identified during this study are not numerically dominant (Walmsley et al. 2012; Matcher et al. 2017, this study, Chapter 3) but they do appear to be ubiquitously associated with this sponge species since *K. kristinae* sequences were detected in metagenomics sequence data obtained from a *T. favus* specimen collected from the same reef in 2014 and 2016 (See Chapters 4 and 6). This ubiquitous presence, combined with the conservation of secondary metabolite gene clusters suggests that the *K. kristinae* isolates identified here confer a selective advantage to their sponge host, which includes the production of siderophore-like compounds and antimicrobial small molecules that support sponge health and resistance to bacterial pathogens. Due to the lack of either lantipeptide pathways reminiscent to that responsible for the production of structurally similar pyrroloquinoline compounds or otherwise promising gene clusters, it is unlikely that the *K. kristinae* isolates identified here are the bacterial origin of pyrroloiminoquinones.

Chapter 3 - Identification of candidate bacterial taxa as biosynthetic origins of tsitsikammamine production

3.1. Introduction

Previously, Walmsley et al., (2012) showed that bacterial communities associated with *T. favus* sponges remained spatially and temporally stable (Walmsley et al., 2012). Analysis of 16S rRNA gene clone libraries indicated Betaproteobacteria as the dominant phylum (75%) with Spirochaetes (10%), Gammaproteobacteria (7%) and Alphaproteobacteria (1%) also present. These results were confirmed with 454 pyrosequencing of amplicon libraries generated from the same sponges, Betaproteobacteria was the dominant phylum followed by Spirochaetes, Bacteroidetes, Planctomyces and Verrucomicrobia (Walmsley et al., 2012).

The dominant Betaproteobacterium in *T. favus* was most closely related to a *Nitrospira* sp. which is an ammonium-oxidising bacterium (Walmsley et al., 2012) similar to the dominant betaproteobacterial isolate, *AqS1*, in Australian coral reef sponge *Amphimedon queenslandica* which carries all necessary genes for ammonia assimilation (Gauthier et al., 2016). Dominant sponge-associated betaproteobacteria have also been observed in encrusting *Crambe crambe* sponges from the Mediterranean Sea (Croue et al., 2013), sympatric *Tethya rubra* sponges from Algoa Bay (Waterworth et al., 2017), and other *Tethya* species (Sipkema and Blanch, 2010; Thiel et al., 2007b; Simister et al., 2013). There was evidence from studies on both *C. crambe* and *A. queenslandica* suggesting that the betaproteobacteria were endosymbiotic (Croue et al., 2013; Gauthier et al., 2016). Analysis of the dominant Spirochaete population at strain level (OTUs clustered at a distance of 0.00) revealed that distinct spirochete OTUs were associated with *Tsitsikamma* sponge species (Matcher et al., 2017). A single spirochete strain, S_OTU2, was exclusive to *T. favus* and novel *Tsitsikamma nguni* sponges (Matcher et al., 2017).

The research described in this chapter aims to characterize the bacterial communities present in latrunculid sponges relative to their environment and sympatric sponges from a different family to identify conserved sponge-associated bacterial communities. To achieve this the study was focussed on the characterization of bacterial

communities in latrunculid sponges relative to sympatric *Tethya* sponge species were collected from both RIY Banks and Evans Peak, Algoa Bay collected in close proximity to *Tsitsikamma* sponge species. It was proposed that these *Tethya* sp. would function as a comparative control to account for geographic specificity of associated bacterial communities. Three *Latrunculia apicalis* sponges were collected from Bouvet Island in the Sub-Antarctic and were included in this study to assess the extent of sponge-host specificity of associated bacterial communities. It is thought that temporal conservation of bacteria indicates a functional role in host-sponge health and may enable the identification of a candidate bacterium as the biosynthetic origin of tsitsikammamines and/or other pyrroloiminoquinone compounds.

Data from this chapter has been published in two manuscripts (Appendix C - Publications) and in a manuscript submitted for publication.

Matcher, G.F., Waterworth, S.C., Walmsley, T.A., Matsatsa, T., Parker-Nance, S., Davies-Coleman, M.T., Dorrington, R.A. (2017). Keeping it in the family: Coevolution of latrunculid sponges and their dominant bacterial symbionts, Microbiology Open, 6(2), e00417. DOI: 10.1002/mbo3.417

Waterworth, S.C., Jiwaji, M., Kalinski, J.J., Parker-Nance-S., Dorrington, R.A. (2017). A Place to Call Home: An Analysis of the Bacterial Communities in Two *Tethya rubra* Samaai and Gibbons 2005 Populations in Algoa Bay, South Africa, Marine Drugs, 15(4). DOI: 10.3390/md15040095

Parker-Nance, S., Waterworth, S.C., Hilliar, S.H., Walmsley, T.A., Dorrington, R.A., New species in the genus *Tsitsikamma* (Poecilosclerida, Latrunculiidae) from South Africa, Zootaxa. Submitted.

3.2. Methods and Materials

3.2.1. Sponge Collection

Specimens were collected over several years by either SCUBA or ROV in Algoa Bay, Port Elizabeth, by SCUBA in Rheeders Reef in Tsitsikamma National Park and by Aggasiz net trawling around Bouvet Island, Antarctica. Specimens were stored at -

20°C until processing. Date and collection site for each sponge specimen used in this study are listed in collection metadata in Appendix A – Table A1.

3.2.2. Extraction of Genomic DNA

Small sections of each sponge were cut aseptically and ground with 2ml sterile artificial seawater (ASW) with a mortar and pestle. The resultant homogenate was centrifuged at 16000 rpm for 1 min to pellet the sponge cells. The *Mycale* sponge specimen TIC2012-030 was found as an encrusting sponge on novel *Tsitsikamma* sp. 005, was removed and processed separately. Total gDNA was extracted as described in Section 2.2.3. Seawater and sediment samples were collected in August 2014 along with *T. favus* sample TIC2014-001, and two further seawater samples were collected in November 2015 during collection of *T. rubra* specimens TIC2015-221 and TIC2015-222 and March 2016, during collection of *T. rubra* specimens TIC2016-001 and TIC2016-001. Marine bacteria were collected from 2 litres of seawater samples via filtration through a 0.2 µm polyethersulfone filter (Pall). Total gDNA was extracted from one-half of a water filter using the PowerWater DNA Isolation kit (Cat. No. 14900-S) (MoBio, San Diego, CA, USA). Total gDNA was extracted from the sediment samples using the ZR Soil Microbe DNA MiniPrep Kit (Zymo D6001).

3.2.3. Identification of Sponge Specimens

Sponge taxonomic identification was conducted by inspection the morphology and spicule structure by Dr Shirley Parker-Nance. Morphological identification was supported by Sanger sequencing of the D3-D5 region of the sponge 28S rRNA gene. The 28S rRNA gene fragment was PCR amplified using primer pair RD3a (5'-GAC CCG TCT TGA AAC ACG A-3') and RD5b2 (5'-ACA CAC TCC TTA GCG GA-3') (McCormack et al., 2002). PCR amplification reactions consisted of 0.3 µM of each primer, 0.75 mM dNTPs, 1X Buffer with $MgCl_2$, 0.5 U Kapa HiFi Taq Polymerase, and 50 ng of template DNA in a total volume of 25 µL. Cycling parameters were as follows: initial denaturation at 95 C for 5 min, followed by 35 cycles of 94 °C for 30 s, 45 °C for 20 s, 72 °C for 1 min, with a final extension of 72 °C for 10 min. Amplicons were ligated into the pGEM-T Easy vector and the insert sequences were determined using Sanger sequencing. At least 3 clones for each sponge specimen were aligned to correct for sequencing errors. Phylogenetic trees were constructed using the MEGA7 software

(Kumar et al., 2016) using ClustalW (Larkin et al., 2007; Goujoun et al., 2010) and Maximum-likelihood algorithms, with 1000 bootstrap replicates.

3.2.4. 454 Amplicon library preparation and pyrosequencing

Amplification of the 470 bp hypervariable region between V4 and V5 of the 16S rRNA gene was performed using primer pair E517F (5'-CAG CAG CCG CGG TAA-3') and E969-984 (5'-GTA AGG TTC YTC GCG T-3') (Wang et al., 2007; Wang et al., 2009) with different MID-tags for each respective sponge sample. PCR amplification of 16S rRNA gene fragments was performed using 0.3µM primers, 0.3 µM dNTPs, 1X Buffer with $MgCl_2$, and 1 U Kapa HiFi Taq polymerase with 30 ng gDNA template. Thermal cycling parameters employed were as follows: Initial denaturation at 95°C for 5 mins followed by amplification with 35 cycles at 94°C for 30s, 45°C for 20s, 72°C for 1 min, followed by a final elongation step of 72°C for 10 mins. Resultant amplicons were gel purified using the Isolate II PCR and Gel kit from Bioline (Cat. No. BIO-52060).

Sequence analysis of ribosomal RNA (rRNA) can provide insight into potentially metabolically active bacterial species in a given population as the ribosomes are more abundant in active bacterial cells than dormant cells (Klein et al., 2016). Sponge fragments stored in RNALater™ at -80°C were thawed gradually on ice. Sponge pieces were ground in 2ml RNAse-free water with a mortar and pestle. The resultant homogenate was centrifuged at 16000 rpm for 1 min and supernatant discarded. RNA was extracted using the ZR Tissue and Insect RNA MicroPrep from Zymo Research (Cat. No. R2030) as per the manufacturer's instructions. The RNA was reverse-transcribed to cDNA using the QuantiTect Reverse Transcription kit from Qiagen (Cat. No. 205311) as per instructions provided with universal primers 8-27F (5'-AGA GTT TGA TCM TGG CTC AG-3') and 1492R (5'-TAC GGY TAC CTT GTT ACG ACT T-3'). PCR amplification of 16S RNA was performed as above for DNA amplicon libraries using generated cDNA as template. Amplicon libraries were sequenced using 454 pyrosequencing by Dr Gwynneth Matcher at the Rhodes University Genome Sequencing Unit.

3.2.5. 454 amplicon sequence data curations

Amplicon library sequence datasets were curated using Mothur software v.1.39.5 (Schloss et al., 2009). Sequences shorter than 200 nts and containing ambiguous bases and/or homopolymeric runs greater than 7 nts were discarded. Sequence chimeras were identified using the VSEARCH algorithm (Rognes et al., 2016) and removed from the respective datasets. Sequences were classified using the Naïve Bayesian classifier against the Silva bacterial database (Version 128) and then scored as the relative percentage of reads per sample dataset. A distance matrix (cut-off of 0.05) was generated in Mothur and used to cluster the sequence reads into Operational Taxonomic Units (OTUs) at distance values of 0.03 (species level), and unique (strain level). OTUs with a relative percentage greater than 0.5% were included in the analysis of dominant bacteria. Phylogenetic trees were constructed as described Section 3.2.3. PRIMER7 (Quest Research Limited) was used to create a non-metric multidimensional scale (NMDS) plots with abundance data generated in Mothur. This data was standardised using square root transformation prior to Bray-Curtis resemblance and ANOSIM analyses.

3.2.6. Illumina amplicon preparation and sequencing

A broader range of sponge species became available at a time in which 454 pyrosequencing from Roche was discontinued. The preparation of 16S rRNA gene amplicon libraries was modified from the above protocol with the use of primers carrying tags suitable for Illumina sequencing using the MiSeq platform. All other aspects of the protocol described in Section 3.2.4. remained the same for Illumina amplicon library preparation. Amplicon libraries were sequenced at the Aquatics Genomics platform at the South African Institute for Aquatic Biodiversity.

3.2.7. Illumina amplicon library data curation

Amplicon libraries sequenced using the Illumina MiSeq platform were curated using Mothur software. Sequences were discarded from datasets if they were shorter than 150 nts, contained ambiguous bases, contained homopolymeric runs greater than 7 nts and/or had a quality score of 20 or less. Data curation from this point was performed as described in Section 3.2.5.

3.3. Results

3.3.1. Sponge Collection and Identification

A total of forty-one sponge specimens were collected over nine years from Algoa Bay, Tsitsikamma National Park, and Bouvet Island in the Sub-Antarctic (collection metadata in Appendix A – Table A1). Dr Shirley Parker-Nance and Ms Rachel Downey performed morphological taxonomic identification of the sponge specimens. Morphological classification of specimens identified two new *Tsitsikamma* species *viz*. *T. nguni* and *T michaeli* (Parker-Nance et al., submitted). To confirm morphological classification, the sponges were identified through amplification of the 28S rRNA gene (Fig. 3.1).

Of the forty-one sponge specimens investigated here, thirty-four were identified as members within the Latrunculiidae family. Within the genus *Tsitsikamma*, a total of twelve specimens were identified as *T. favus*, six were identified as novel species *T. nguni*, four were identified as *T. michaeli* and four were identified as *T. pedunculata* species. A total of four *Latrunculia* specimens, belonging to two species and four *Cyclacanthia* specimens were included in this study to assess the differences in bacterial communities associated with other genera within the Latrunculiidae family. Seawater samples from Rheeders Reef, Evans Peak and RIY Banks were included along with seven sympatric *Tethya* sponge species as representatives of the environment surrounding the latrunculid sponges. Full metadata of each sponge specimen has been detailed in Table A1 in Appendix A.

Sponges species *T. michaeli* and *T. pedunculata* are morphologically distinct and it was unexpected for these species to be phylogenetically clustered (Parker-Nance et al., submitted). Manual inspection of the sequence data revealed that these sponge species harbour identical D3-D5 regions of the 28S rRNA gene sequence. There was unexpected variability in the 28S rRNA gene sequences of species morphologically classified as *T. favus*. Manual inspection of the 627 nt long nucleotide sequence of the 28S rRNA gene fragment showed that the thymine residue at position 107 had been substituted for a guanine in TIC2016-050A and that the adenine residue at position 221 was substituted for a thymine residue and a cytosine residue replaced a guanosine in TIC2012-057. It was also noted that at position 535, TIC2012-057 and

TIC2009-002 had a guanosine residue whilst TIC2014-001 and TIC2016-050A had an adenosine residue. Novel species *T. nguni*, is morphologically very alike to *T. favus* (Parker-Nance et al., submitted) and this appears to be reflected in molecular similarities as *T. nguni* sequences clustered with sponges classified as *T. favus*.

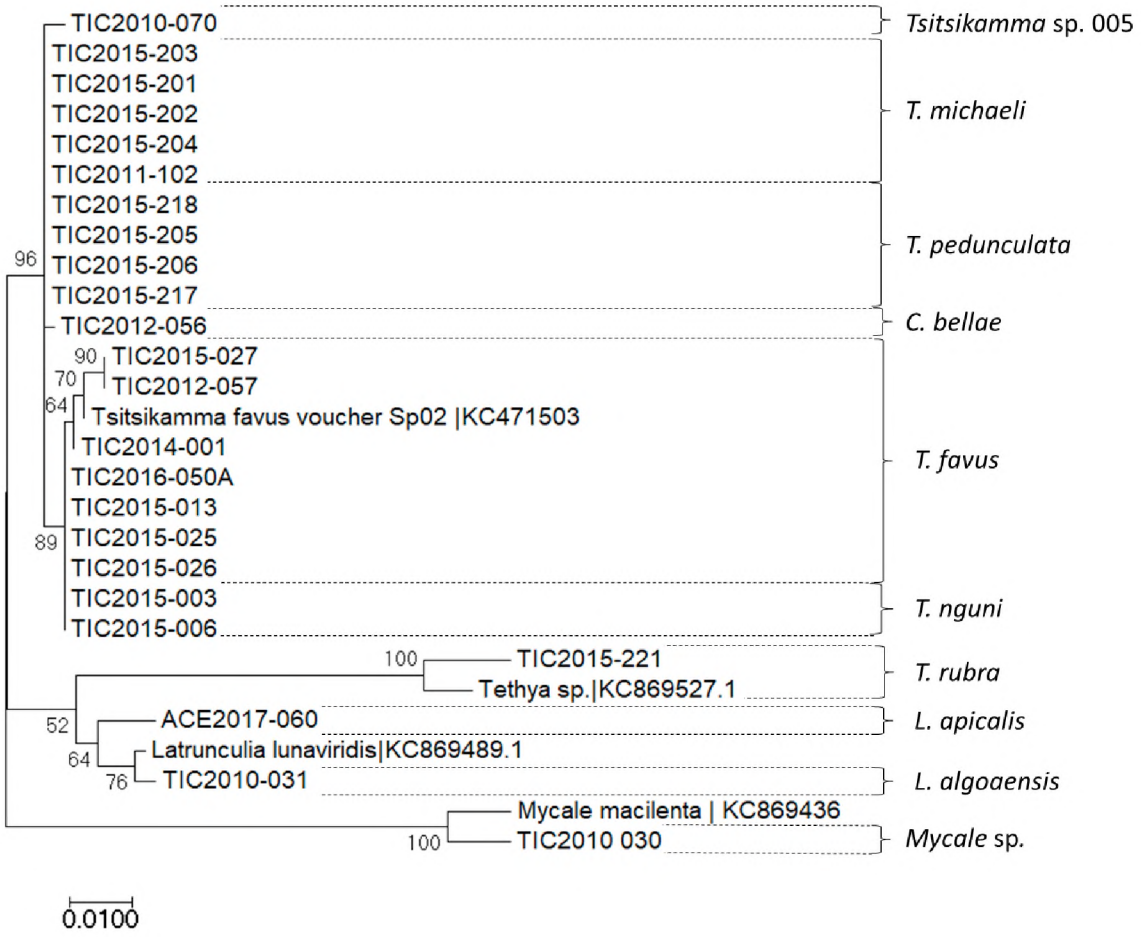


Figure 3.1: Phylogenetic relationship of latrunculid and related sponges based on alignment of 28S rRNA gene fragment sequences (627 nts) aligned using ClustalW and constructed using the maximum-likelihood method with 1500 bootstrap replicates in MEGA7. The scale bar represents an estimated 0.2% sequence substitution. Genbank accession numbers have been included where applicable. Several of the 28S rRNA gene sequences used in this analysis were generated by Dr. Tara Walmsley (Vaal University of Technology) and Ms Storm Hilliar (MSc Student, Marine Natural Products Laboratory, Rhodes University)

3.3.2. Characterization of Microbial Communities Associated with Sponges within the Family Latrunculiidae

Phylogenetic classification of the bacterial communities clearly illustrated the distinction of latrunculid sponges from their environment and *T. rubra* sponges collected in Algoa Bay (Fig. 3.2). A dominant betaproteobacterial phylum dominated the bacterial communities associated with all three Latrunculiidae genera as well as the bacterial community of the encrusting *Mycale* sp. sponge, accounting for 20% - 90% of bacterial sequence reads.

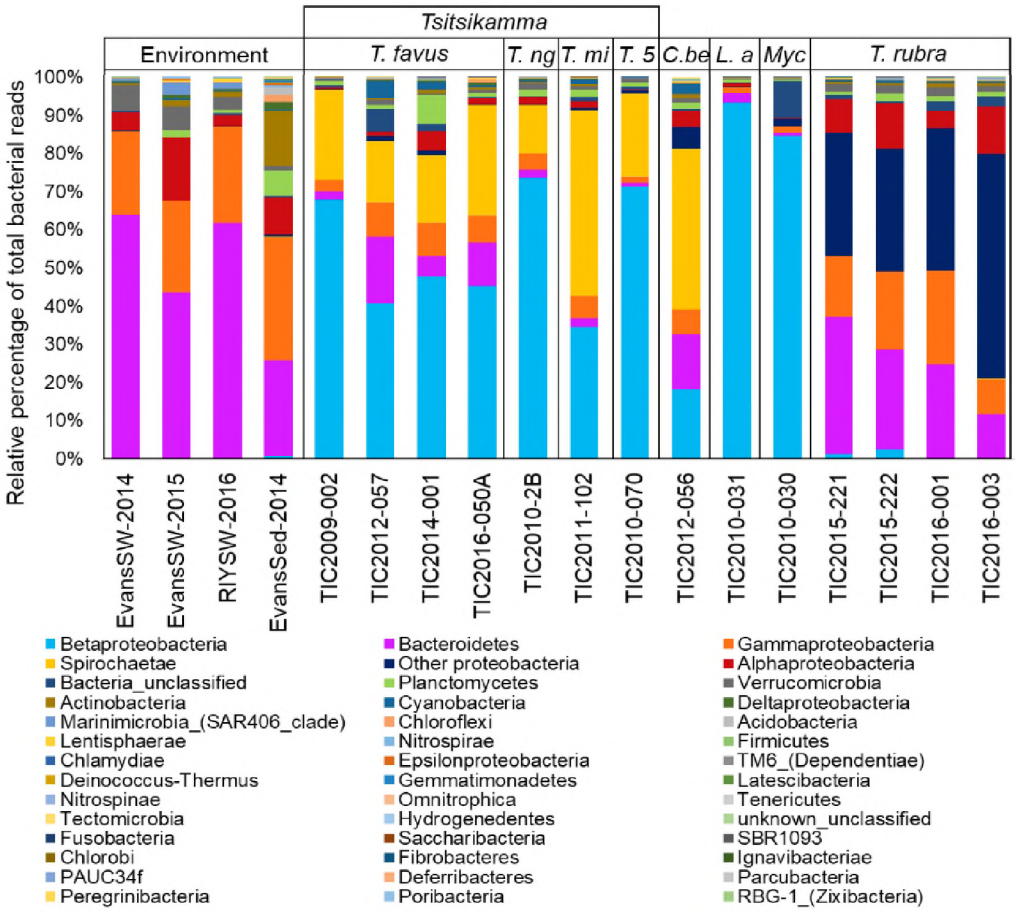


Figure 3.2: Phylogenetic classification of latrunculid sponge-associated bacterial communities using Naïve Bayesian classification with the Silva bacterial database (version 128) as the reference. A coloured key indicates phylogenetic classifications. Sponges have been grouped by taxa which is indicated by the label above each box with the following abbreviated names: *T. favus*: *Tsitsikamma favus*, *T. ng*: *Tsitsikamma nguni*, *T. mi*: *Tsitsikamma michaeli*, *T.* 5: *Tsitsikamma* sp. 005, *C. be*: *Cyclacanthia bellae*, *L. a*: *Latrunculia algoaensis*, *Myc*: *Mycale* sp., *T. rubra*: *Tethya rubra*.

Spirochaetes were also found as a dominant phylum with *Tsitsikamma* and *Cyclacanthia* sponges with the read abundance ranging between approximately 17% and 50%. In contrast, dominant communities of Bacteriodetes, Gammaproteobacteria,

Alphaproteobacteria and other proteobacteria sub-phyla were detected in sympatric *T. rubra* sponge species. Dominant communities of Bacteriodetes, Gammaproteobacteria and Alphaproteobacteria were also observed in the seawater and sediment samples and may account for the observed similarity between *Tethya* sponge species and the surrounding environment.

Alignment and phylogenetic classification of the V4-V5 region of the 16S rRNA gene was used to characterise the bacterial communities present in sponges, seawater and sediment. Statistical oordination and ANOSIM analysis of bacterial OTUs clustered at a distance of 0.03 (Fig. 3.3) revealed that the latrunculid sponges were significantly different from the surrounding environment (p = 0.005).

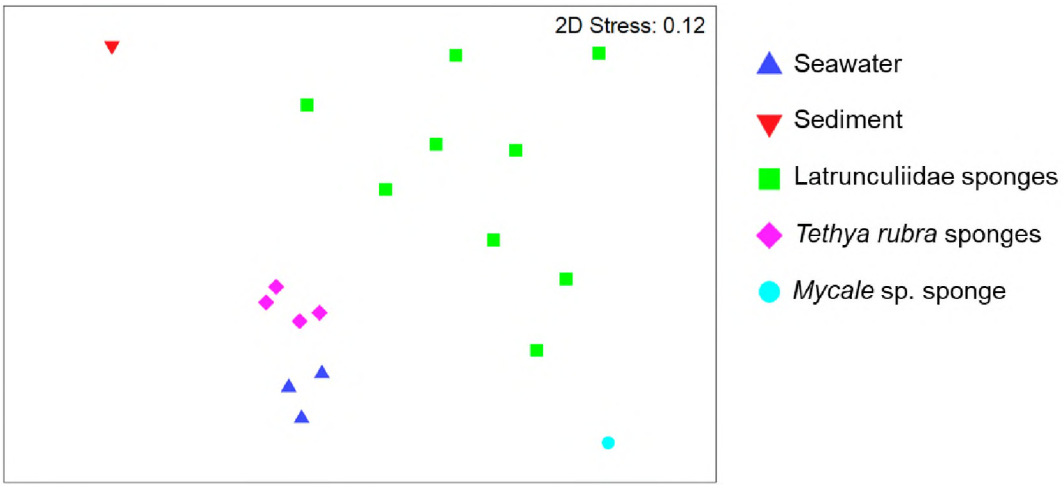


Figure 3.3: Statistical oordination of bacterial OTUs (clustered at distance of 0.03) in sponges using non-metric multidimensional scaling. Square root transformation was applied to abundance data prior to construction of Bray-Curtis resemblance matrices based on partial 16S rRNA gene sequences. A colour key is provided to distinguish sample types.

The *Mycale* sponge TIC2010-030 did not show any significant difference to the bacterial communities associated with latrunculid sponges nor that associated with *T. rubra* sponges nor the surrounding seawater. The analysis also showed that the bacterial communities associated with latrunculid sponges were significantly different to those found in *T. rubra* sponges (p = 0.001). The bacterial communities associated with *T. rubra* sponges were found not to show any significant difference to bacterial communities present in the seawater when clustered at a distance of 0.03 (Fig. 3.3). However, it was shown that the bacterial communities in *T. rubra* sponges differed

significantly to those in the seawater when analysed as unique OTUs clustered at a distance of 0.00 (Waterworth et al., 2017).

The diversity of bacterial species associated with sponge species was characterized through classification of dominant bacterial OTUs (clustered at a distance of 0.03). The seawater was observed to have relatively dominant communities of Flavobacteriacea and Gammaproteobacteria (OTU12 and OTU14), that were also detected in all *T. rubra* sponges but absent in latrunculid and *Mycale* sponges (Fig. 3.4).

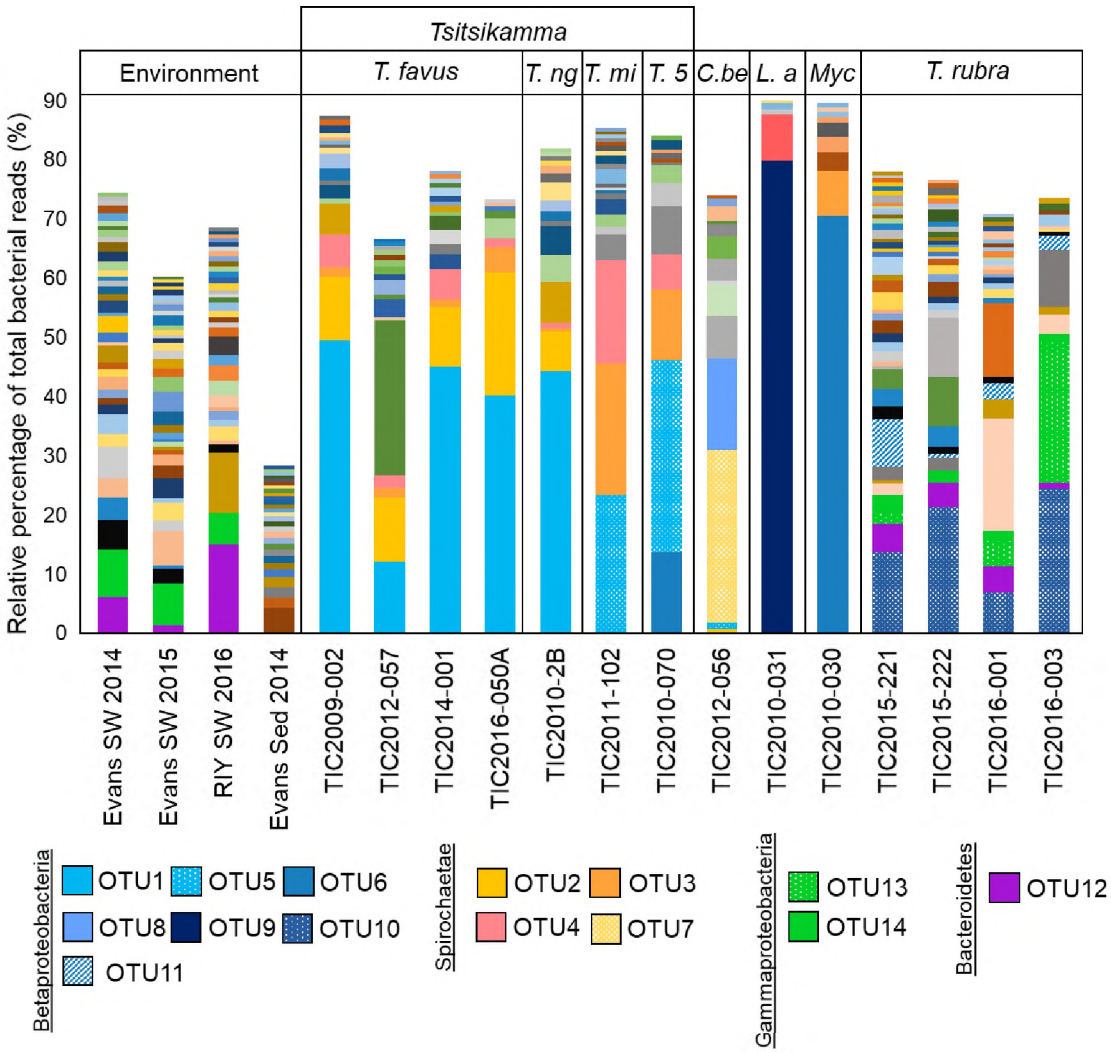


Figure 3.4: Diversity of dominant bacterial taxa associated with marine sponges from Algoa Bay and Rheeders Reef, South Africa, and the surrounding environment. Bacterial OTUs were clustered at a distance of 0.03 as an approximation to bacterial species. Relative abundance of bacterial reads was calculated as a percentage of the number of reads per OTU relative to the total number of reads per sample. Dominant bacterial OTUs are defined as having a relative abundance greater than 0.5%. OTUs discussed are indicated by a coloured key.

The latrunculid and *Mycale* sponge species were dominated by betaproteobacterial species that showed striking host specificity. *T. favus* sponges and the *T. nguni* sponge share a common taxon, OTU1, whilst *T. michaeli* and *Tsitsikamma* sp. 005 share OTU5 as a dominant betaproteobacterium. *Tsitsikamma* sp. 005 was associated with a second dominant betaproteobacterium, OTU8, which dominates the bacterial community of the *Mycale* sponge. Since the *Mycale* sp. was an encrusting sponge of *Tsitsikamma* sp. 005 it possible that the bacterial communities were cross-contaminated. *Tsitsikamma* sp. 005 and *T. michaeli* are associated with a common betaproteobacterial species, OTU5 whilst *C. bellae* and *L. algoaensis* each host distinctive betaproteobacterial OTUs, OTU8 and OTU9 respectively. Sympatric *T. rubra* is host to two conserved betaproteobacterial OTUs, OTU10 and OTU11. Phylogenetic analysis of the bacterial communities associated with *T. rubra* sponges did not appear to include betaproteobacteria. It is probable that because phylogenetic analysis was performed through comparison of reads to the SILVA database and the OTUs were classified using BLAST, that several reads were classified into "Other proteobacteria" during phylogenetic analysis and were later identified as betaproteobacteria during BLAST classification. It was found in a separate study that there was evidence to suggest that the conserved betaproteobacterium identified in the *T. rubra* sponges was vertically transferred from the parent sponge to its progeny (Waterworth et al., 2017). The betaproteobacterial OTU was consistently detected in all studied sponge embryos and not detected in the surrounding seawater and it was thought that this conserved taxon was responsible for the production of bioactive compounds isolated from these sponges (Waterworth et al., 2017).

A phylogenetic analysis of betaproteobacteria OTUs identified in marine demosponges showed that these bacteria are host specific and cluster relative to the phylogeny of their sponge host (Fig. 3.5). The dominant betaproteobacteria found in *T. favus* and *T. nguni* sponges cluster with uncultured bacteria clones previously isolated from *T. favus* specimens TIC2009-001 and TIC2009-002 (Walmsley et al., 2012). The betaproteobacteria associated with *T. michaeli*, *Tsitsikamma* sp. 005 and *C. bellae* more closely related to the betaproteobacteria associated with *T. favus* than the betaproteobacteria found in *L. algoaensis*. The phylogeny of the betaproteobacteria was more closely aligned to its host sponge lineage than of its geographic location as the betaproteobacteria identified in *T. rubra* collected from

Algoa Bay were more closely related to betaproteobacteria isolated from *T. californiana* sponges collected off the coast of California than to the betaproteobacteria found in sympatric latrunculid sponges. Likewise, the *L. apicalis* sponges collected from Bouvet Island in the Sub-Antarctic were host to a dominant betaproteobacterium (Fig. 3.10) which was closely related to the betaproteobacteria identified in *L. algoaensis* and other latrunculid sponges collected off the coast of South Africa.

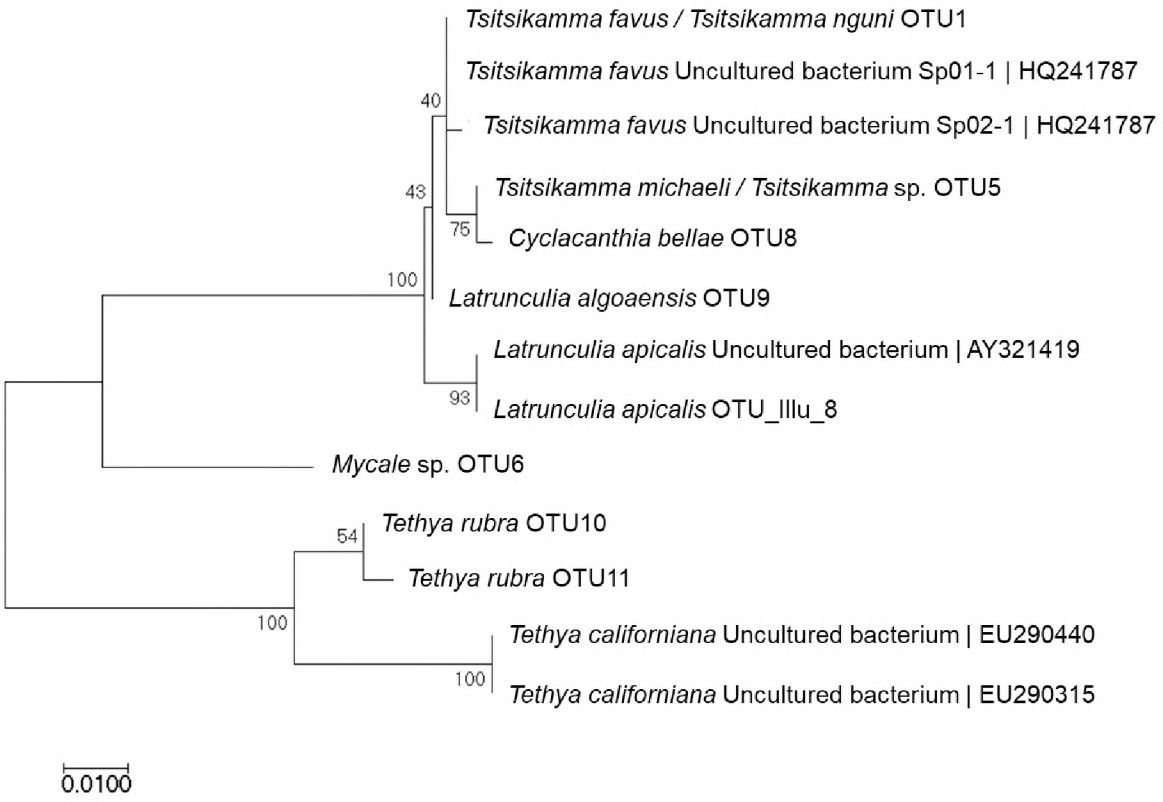


Figure 3.5: Phylogenetic relationship of betaproteobacteria identified in sponges collected from Algoa Bay and Rheeders Reef, South Africa, relative to other related, dominant betaproteobacteria associated with other sponge specimens. A 434bp region of the 16S rRNA gene was aligned using ClustalW and phylogeny was inferred using the Maximum-likelihood method in MEGA7 with 1000 bootstrap replicates. The Tamura-Nei substitution model was used with uniform rates across all sites with a strong branch swap filter. Boot strap values are indicated on branch nodes and a scale bar is provided which indicates the average number of nucleotide substitutions per site.

Dominant, conserved Spirochaete taxa were identified exclusively in *Tsitsikamma* sponge species and *C. bellae* (Fig. 3.4). Taxonomic classification of OTUs clustered at distance of 0.03 revealed that OTU2, OTU3, OUT4 and OTU7 were predicted to be species within the Spirochaete genus (Fig. 3.4). Analysis of OTUs clustered at a distance of 0.00 showed that distinct strains of spirochetes were present in the different sponge species (Fig. 3.6).

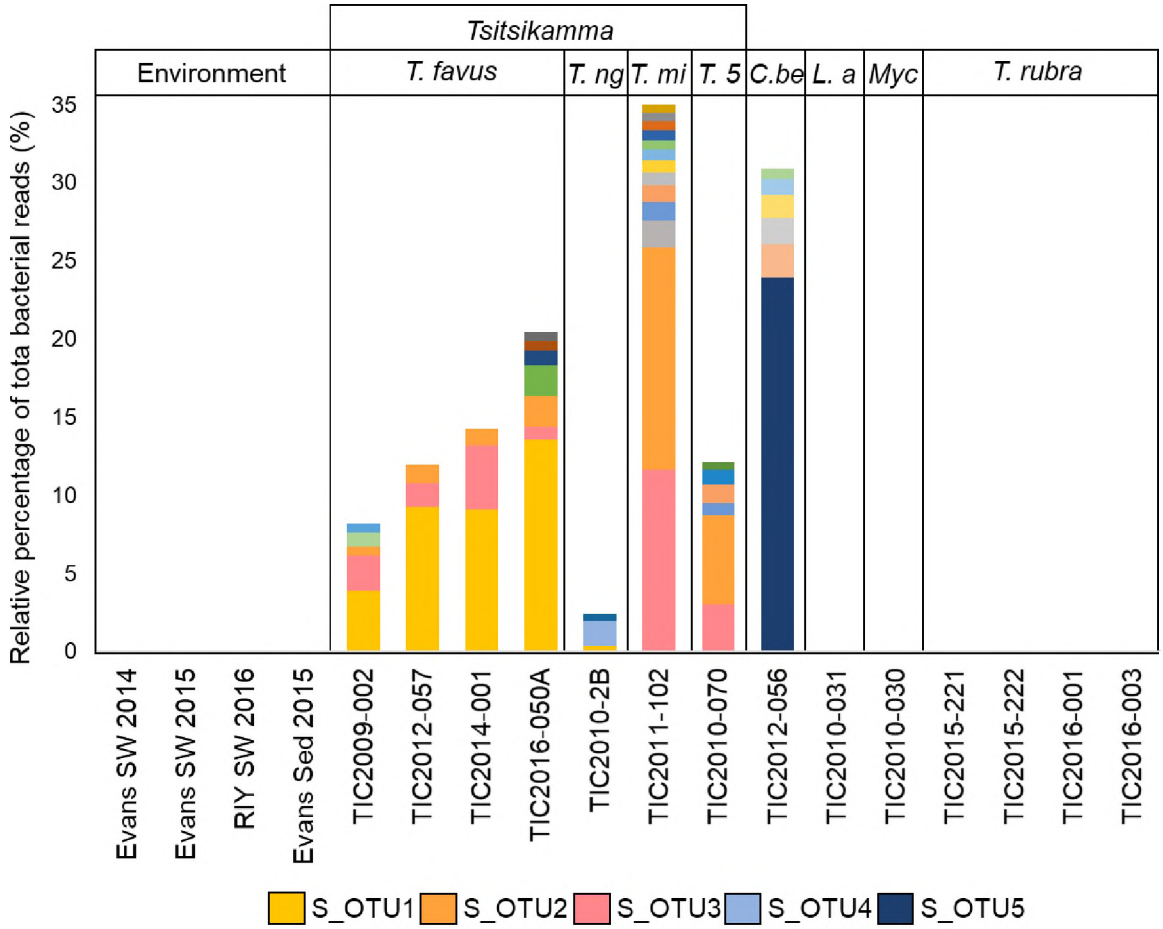


Figure 3.6: Diversity of dominant spirochetes associated with marine sponges from Algoa Bay and Rheeders Reef, South Africa, and the surrounding environment. Relative abundance of bacterial reads was calculated as a percentage of the number of reads per OTU relative to the total number of reads per sample. Dominant bacterial OTUs are defined as having a relative abundance greater than 0.5%. OTUs discussed are indicated by a coloured key.

Like the Betaproteobacteria, the distribution of Spirochaete bacteria appeared to be host specific and were not detected in the surrounding environment, nor in *T. rubra*, *Mycale* or *L. algoaensis* sponges. A dominant spirochete OTU, S_OTU5, was detected in *C. bellae* and accounted for almost 25% of the total bacterial reads in this sample. *T. michaeli* and *Tsitsikamma* sp. 005 were host to two spirochaete OTUs, S_OTU2 and S_OTU3, the latter of which appeared to be more dominant. These two OTUs were also detected in *T. favus* and *T. nguni*. The *T. nguni* sponge was host to S_OTU4, which was not detected in any other sample and may be specific to this sponge or a transient food bacterium.

Spirochete strain, S_OTU1, was found exclusively in *T. favus* and *T. nguni* and could not be detected in any other sponge samples nor in the surrounding environment. It was found through direct sequence comparison that OTU2, OTU3, OTU4 (Fig. 3.4)

corresponded to spirochete strains S_OTU1, S_OTU2 and S_OTU3 respectively and have been represented in appropriate matching colours in all graphs.

Phylogenetic analysis of the five spirochete strains (Fig. 3.7) showed that the bacteria identified within *Tsitsikamma* sponges were related and clustered with spirochete clones, Sp02-3 and Sp02-15, previously isolated from *T. favus* sponge TIC2009-002 (Walmsley et al., 2012), forming a distinct clade in comparison to known spirochetes. Spirochete S_OTU5 was dominant in *C. bellae* and was less phylogenetically related to *Tsitsikamma* sponge-associated spirochetes, although it was clustered more closely than known relatives. The closest known phylogenetic relative was a *Salinispira pacifica* strain that was isolated from hypersaline microbial mats (Hania et al., 2015).

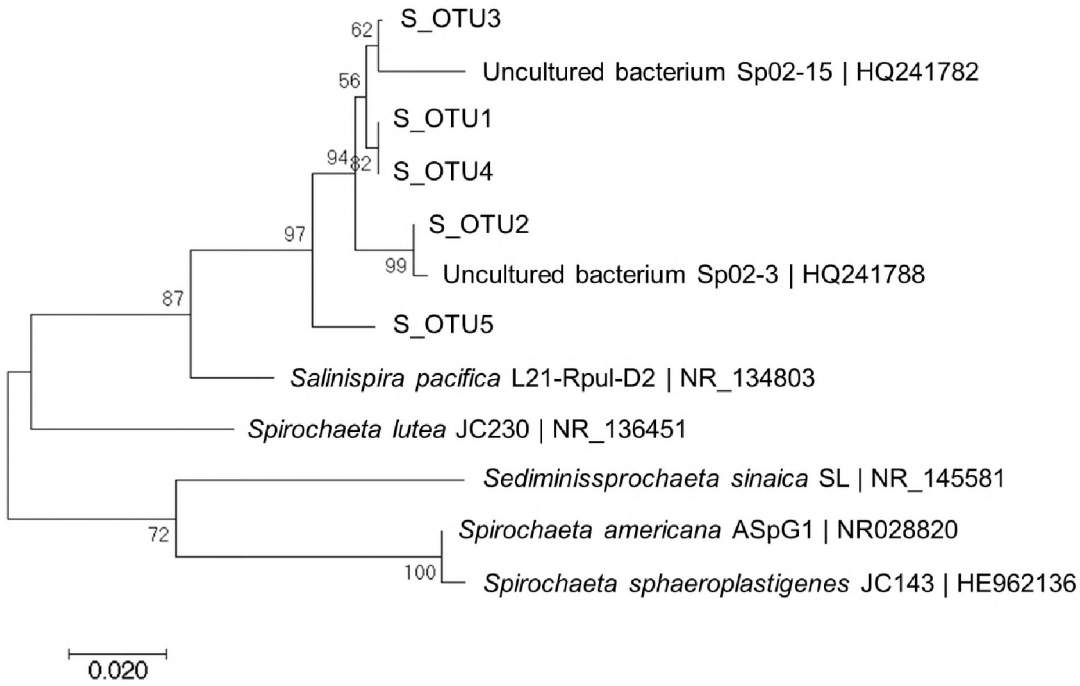


Figure 3.7: Phylogenetic diversity of dominant spirochetes associated with marine sponges from Algoa Bay and Rheeders Reef, South Africa. Dominant OTUs classified as Spirochetes are differentiated from general bacterial OTUs with a "S_" prefix.

Isolation and sequence analysis of ribosomal RNA for the estimation of metabolically active bacteria, within a given population, has been employed in over 100 peer-reviewed studies (Blazewicz et al., 2013) and has been demonstrated as a reasonable effective method to assess bacterial metabolic activity (Mills et al., 2012). This approach was employed to assess the metabolic potential of the bacterial communities associated with *T. favus* sponges *viz.* TIC2014-001 and TIC2012-057. Analysis of the

bacterial species (OTUs clustered at a distance of 0.03) revealed that there were significant differences between numerically and metabolically dominant bacteria in *T. favus* sponges (Fig. 3.8).

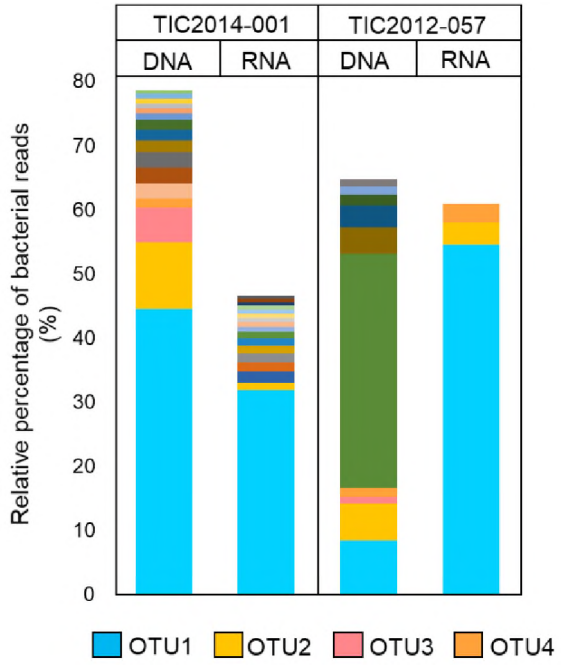


Figure 3.8: Numerically and metabolically dominant bacteria identified in *T. favus* sponge specimens TIC2014-001 and TIC2012-057. Relative abundance of bacterial reads was calculated as a percentage of the number of reads per OTU relative to the total number of reads per sample. Dominant bacterial OTUs are defined as having a relative abundance greater than 0.5%. Classification of OTUs discussed here are identified as per Fig. 3.4 *viz*. a betaproteobacterium (OTU1) and three closely related spirochete strains (OTUs 2, 3 and 4).

Betaproteobacteria species, OTU1, was both numerically and metabolically dominant within *T. favus* associated microbiota. However, the same was not true of the associated spirochete OTUs OTU2, OTU3 and OTU4. It appeared that OTU4 was not metabolically active in either TIC2014-001 nor TIC2012 and that an unconserved, numerically dominant betaproteobacterium (Fig. 3.8, dark green) present in TIC2012-057 did not appear to be metabolically active in the sponge specimen either. Spirochete species, OTU3, was only detected in the cDNA of TIC2012-057. It was found that only spirochete OTU2 was metabolically active in both sponge specimens but was represented at a lower relative abundance than observed for their numeric representation. However, genome size and abundance of ribosomes per cell in different bacteria may influence this result and caution needs to be taken with definitive interpretation this data.

3.3.3. Characterization of associated bacterial taxa across a broad range of sponge species

Several new sponge specimens were collected between 2015 and 2017 and included *Tsitsikamma* species *T. pedunculata* and several additional specimens of *T. nguni*, *T. favus*, *T. michaeli, C. bellae* and *Tethya* sp. from Algoa Bay, South Africa. A unique opportunity aboard the Akademik Tryoshnikov also afforded the collection of three Sub-Antarctic *Latrunculia apicalis* specimens. At this time 454 Pyrosequencing from Roche had been discontinued and sponge-associated bacterial communities were instead assessed through Illumina MiSeq sequencing of amplicon libraries.

Phylogenetic classification of bacterial reads showed that the bacterial communities in the seawater had dominant populations of Bacteriodetes, gammaproteobacteria and alphaproteobacteria consonant with results observed in Fig. 3.2 of this study. The bacterial communities observed in sponge samples were largely dominated by betaproteobacteria and gammaproteobacteria which contrasted earlier findings relative to *Tethya* sponge species which were previously observed to have very few betaproteobacteria. Spirochaetes were not detected in seawater and were observed in all latrunculid sponges except for sponges within the *Latrunculia* genus. A small proportion of Spirochaetes were observed in two of the three new *Tethya* sp. sponges but as these have not been formally barcoded yet, they may represent a distinct species from *T. rubra*.

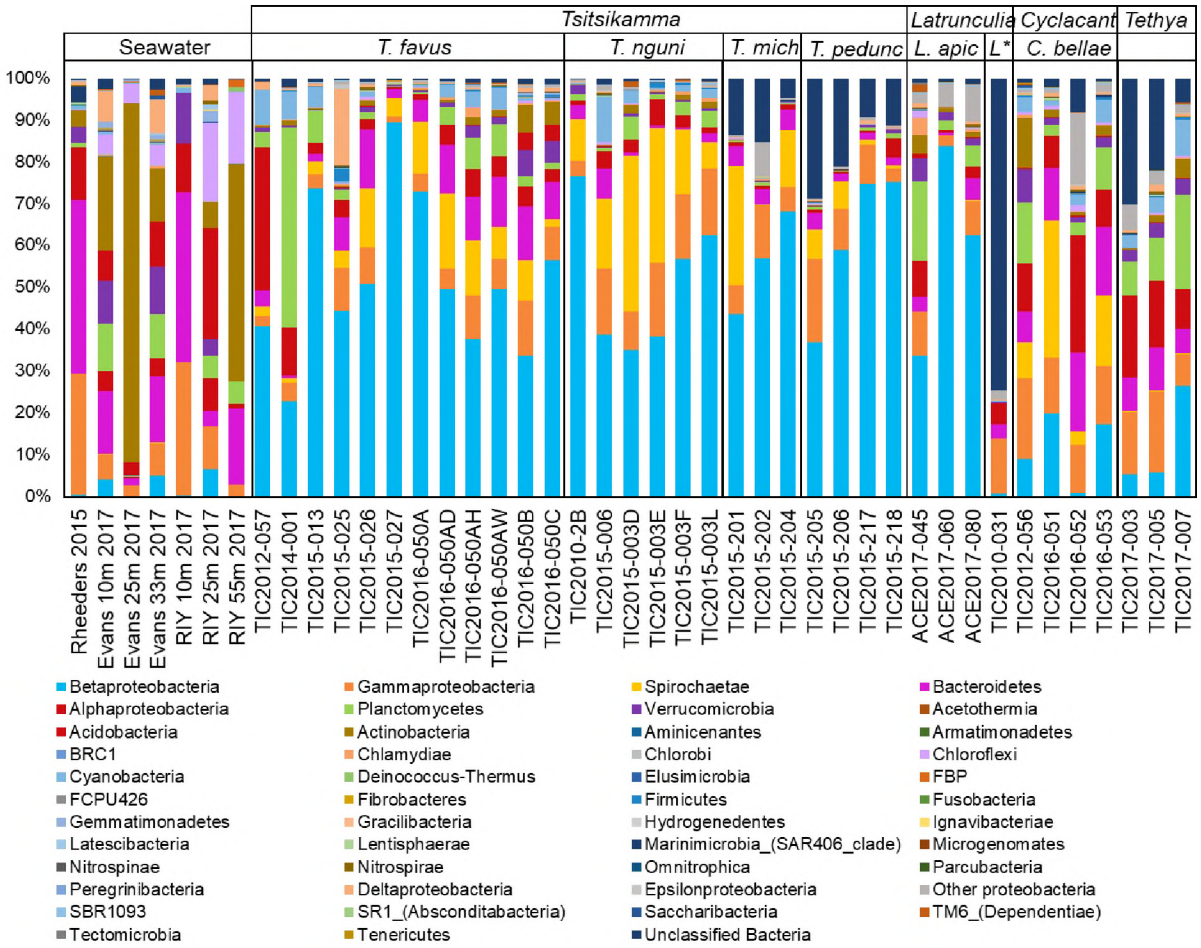


Figure 3.9: Phylogenetic classification of latrunculid sponge-associated bacterial communities using Naïve Bayesian classification with the Silva bacterial database (version 128) as the reference. Phylogenetic classifications are indicated by a coloured key. Sponges have been grouped by taxa which is indicated by the label above each box with the following abbreviated names: *T. favus*: *Tsitsikamma favus*, *T. nguni*: *Tsitsikamma nguni*, *T. mich*, *Tsitsikamma michaeli*, *L. apic*: *Latrunculia apicalis*, L*: *Latrunculia algoaensis*, *C. bellae, Cyclacanthia bellae*. Seawater amplicon sequence libraries were prepared and provided by Ms Danielle Vos and Ms Storm Hilliar.

Characterisation of bacterial species (OTUs clustered at 0.03) associated in the larger sponge collection revealed similar patterns observed in Fig. 3.4 of this study (Fig. 3.10). It was shown that *T. favus*, *T. nguni* and *T. pedunculata* share a conserved, dominant betaproteobacterial taxon (OTU_Illu_1) whilst *L. algoaensis* (OTU_Illu_17), *L. apicalis* (OTU_Illu_8) and *C. bellae* (OTU_Illu_12) each have distinct betaproteobacteria. *T. pedunculata* sponge species have two conserved gammaproteobacteria (OTU_Illu_4 and OTU_Illu_5) that were not identified in any of the other sponge species. *T. pedunculata* is not the focus of this project, however these bacteria do warrant further investigation.

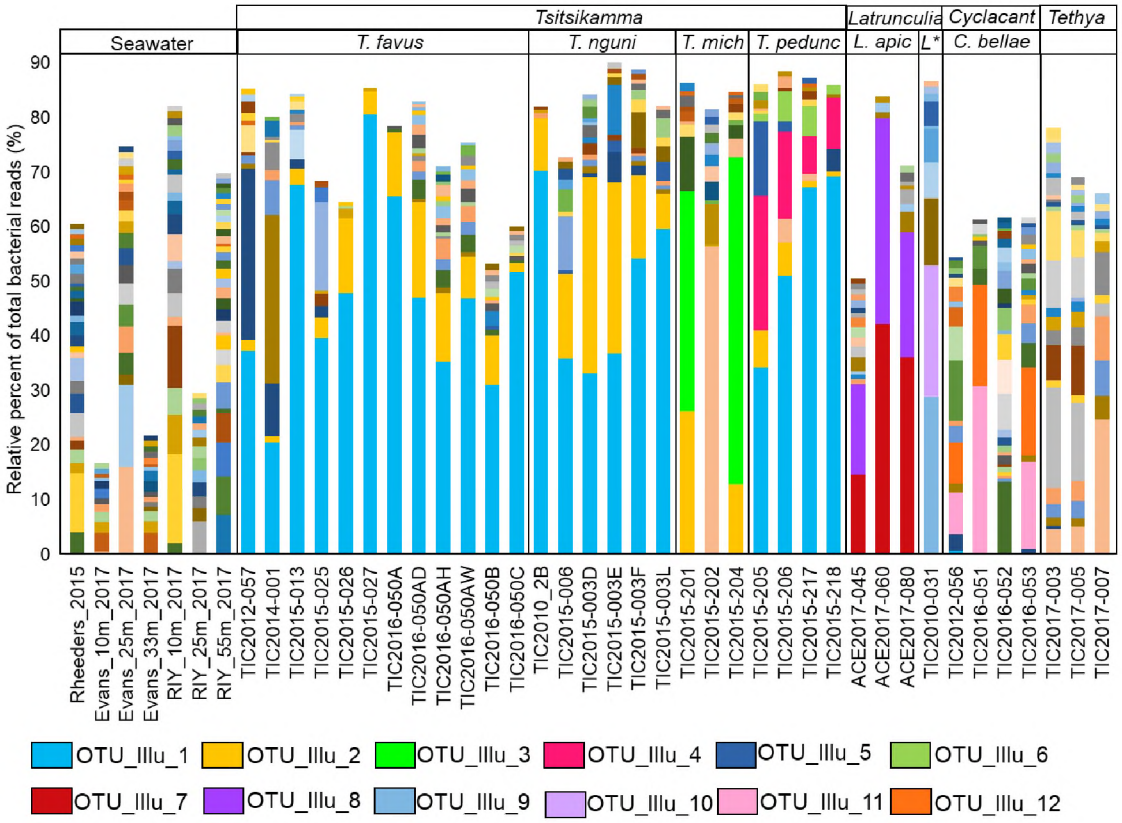


Figure 3.10: Diversity of dominant bacterial species associated with marine sponges from Algoa Bay and Rheeders Reef (South Africa), Bouvet Island (Sub-Antarctica) and seawater collected from Algoa Bay and Rheeders Reef. Bacterial OTUs were clustered at a distance of 0.03 as an approximation to bacterial species. Relative abundance of bacterial reads was calculated as a percentage of the number of reads per OTU relative to the total number of reads per sample. Dominant bacterial OTUs are defined as having a relative abundance greater than 0.5%. OTUs discussed are indicated by a coloured key.

OTU_Illu_2 was identified exclusively within sponges of the *Tsitsikamma* genus and was not detected in other sponge genera nor the surrounding seawater. Taxonomic classification showed that this OTU was most closely related to an uncultured spirochete clone Sp02-3 that was previously detected in *T. favus* TIC2009-002 and thus correlated with OTU2 in this study. The expansion of sponge species investigated confirmed the conservation of spirochetes within *Tsitsikamma* sponge species.

During the alignment of amplicons generated from 454 pyrosequencing it was found that the divergent nucleotides in the 16S rRNA consensus sequence of conserved, unique betaproteobacteria and spirochetes were located outside the 260nt region sequenced in the Illumina amplicon libraries. Therefore, analysis of bacterial OTUs clustered at a distance of 0.00 using the Illumina MiSeq amplicon libraries could not be performed reliably.

3.4. Discussion

The aim of this chapter was to investigate whether conservation of bacterial communities in *T. favus* sponges observed by Walmsley et al., (2012) was consistent in other latrunculid sponge species. The biosynthetic source of tsitsikammamines is hypothesised to be bacterial (Walmsley et al., 2012). Temporal conservation of a bacterial population unique to *T. favus* and *T. nguni* sponges would enable the identification of a candidate bacterium involved in the biosynthesis of tsitsikammamines. The data presented in this study reveal that the bacterial communities associated with latrunculid sponges are significantly different to their surrounding environments and remain relatively stable over time and geographic location.

Morphological identification of the sponge specimens led to the discovery of two new *Tsitsikamma* sponge species *T. nguni* and *T. michaeli*. Sponge species can be molecularly identified using various gene markers including mitochondrial cytochrome c oxidase subunit I gene (*cox*1), ribosomal internal transcribed spacer (*ITS*) regions and cytoplasmic ribosomal 28S and 18S rRNA gene fragments, (Yang et al., 2017). Previously, Walmsley (2013) found that *cox*1 genes were not readily amplified from latrunculid sponge samples and did not exhibit sufficient nucleotide substitutions to differentiate sponge specimens below genus level. It was found that amplification of the 28S rRNA gene was more successful and that distinction could be made between closely related *T. favus* and *T. scurra* species. This data was corroborated in a more recent study into the molecular identification of sponge species where it that found that the 28S rRNA gene was more variable than the *cox*1 gene and allowed for discrimination between specimens (Erpenbeck et al., 2016). Analysis of the D3-D5 region of the 28S rRNA gene sequence revealed there was no distinguishable difference between this sequence region of *T. pedunculata* and *T. michaeli*. The limitations of identification of metazoan species through short DNA "barcode" sequences have been discussed in Thacker et al., (2013). Incongruences between morphological and molecular classifications has been previously noted for sponges within the Axinellidae family (van Oppen et al., 2002) as well as for several taxa within sponges of the order Haplosclerida (McCormack et al., 2002). There are multiple copies of the 28S rRNA gene sequence present on eukaryotic genomes and

intragenomic variation of rRNA genes has been previously documented in other eukaryotic organisms *viz.* Chromalveolates, Opisthokonta, plants and fungi (Torres-Machorro et al., 2009; Reich and Labes, 2017). It is thus plausible that there is intragenomic variation of the 28S rRNA gene fragment in *T. favus* sponges. To overcome this complication of molecular classification, a multi-loci approach, with amplification of the 28S rRNA gene, the COI mitochondrial gene as well as amplification of the ITS region, is suggested as the combination thereof has provided greater accuracy in sponge specimen identification (Yang et al., 2017). This approach could be implemented to allow for greater molecular distinction between closely related latrunculid sponges.

It was also noted that sponges classified as *T. favus* contained three nucleotides within the D3-D5 sequence region which were not identical. This would imply that the four specimens classified as *T. favus* should not be classified within a single species. However, analysis of spicule structure (performed by Dr Parker-Nance) showed that all sponges in this collection, classified as *T. favus,* exhibited identical spicule structure that is congruent with the *T. favus* type specimen. It is possible that either *T. favus* has undergone cryptic speciation, similar to the *Scopalina lophyropoda, Hexadella dedritifera, Cliona celata* and various other sponges (Xavier et al., 2010) or the multiple copies 28S rRNA gene may be present in the *T. favus* genome and that mutations are present within these copies confounding the molecular classification of these sponges.

In this study, all investigated sponge species appear to host a conserved betaproteobacterium, the phylogeny of which appeared to be relative to that of its host sponge. Dominant betaproteobacteria have been reported in several other sponge species, including *Crambe crambe*, *Callyspongia* sp. and *Tethya bergquistae* (Croue et al., 2013; Cardenas et al., 2014; Steinert et al., 2016). Sponge associated betaproteobacteria have been reported as producers of bioactive compounds (Brinkmann et al., 2017b) and one of the most abundant, deep-branching, sponge-specific bacteria that have very few known relatives, making functional inference within the host sponge problematic (Simister et al., 2012). However, the advent of shotgun metagenomic analyses has enabled greater insight into the functional capabilities of sponge-associated bacteria.

A dominant betaproteobacterium identified in a *Sarcotragus foetidus* sponge was shown to encode several genes responsible for the production of proteins involved in CRISPR and restriction modification (RMS) defence systems (Horn et al., 2016) whilst a dominant betaproteobacterium in *A. queenslandica* is proposed to be an obligate symbiont with a reduced genome that may play a role in nutrient acquisition for the host sponge (Gauthier et al., 2016). The functional role of the conserved betaproteobacterium represented by OTU1 in this study therefore cannot be definitively speculated upon without further investigation. However, evidence for co-evolution with the host sponge presented here would suggest that it plays an important role in host-sponge health.

The results presented in this chapter differ slightly from that reported in Matcher et al., (2017) in which many of the same datasets were used. In Matcher et al., 2017, it was shown that all latrunculid sponges investigated were associated with a single betaproteobacterial species (OTUs clustered at 0.03) and a single spirochete specie was associated with all *Tsitsikamma* and *Cyclacanthia* sponges. In this study, it was shown that the OTUs previously described in Matcher et al., 2017 as unique strains were different species of betaproteobacteria and spirochetes. This may be because the analyses presented here were performed using an updated Silva Database (Version 128), a later release of Mothur v.1.39.5, chimeric sequences were identified using the VSEARCH algorithm rather than UCHIME and a different standard operating procedure SOP was followed than that described in Matcher et al., (2017). These different approaches in analyses may account for the minor discrepancies observed between results reported this study and those reported in Matcher et al., (2017).

Spirochete sponge symbionts have been reported in several sponge species such as *Xestospongia muta* and *Clathrina prolifera* (Fiore et al., 2013; Isaacs et al., 2009) and were a dominant taxon in the bacterial community associated with a *Clathrina clathrus* sponge (Neulinger et al., 2010). Dominant sponge-associated spirochetes appear to be rare but symbioses between spirochetes and termites have been reported in which they play a role in nitrogen fixation, acetogenesis and early stages of lignin degradation (Russell et al., 2014). The presence of a conserved, dominant spirochete is currently the only reported for sponge species in the family Latrunculiidae (Matcher et al., 2017). Phylogenetic analysis of unique spirochetes showed that strains detected in latrunculid sponges formed a distinct clade with Spirochaete clones previously

isolated from *T. favus* TIC2009-002 and that the spirochete associated with *C. bellae* was less closely related than those associated with *Tsitsikamma* sponges. This suggested that, like the betaproteobacteria described above, the spirochetes may be specialised symbionts that perform a specific function within the host sponge.

The closest known phylogenetic relative is *Salinispira pacifica*, the genome of which has been sequenced but no functional attributes have yet to be assigned to this bacterium in the context of in hypersaline microbial mat (Hania et al., 2015). Only one of the three spirochete strains identified in *T. favus* sponges, S_OTU1, was apparently metabolically active based on 16S rRNA relative abundance. This strain is conserved and found exclusively in *T. favus* and *T. nguni* sponge species. As *T. favus* and *T. nguni* are the sole producers of tsitsikammamines, S_OTU1 (OTU2) could represent a candidate bacterium for involvement in the biosynthesis of tsitsikammamines. However, no spirochetes were identified in the *Latrunculia* sponge species from which pyrroloiminoquinone alkaloids have been isolated. It was shown that *Tsitsikamma* and *Cyclacanthia* sponge species are more phylogenetically closely related to one another than to species within the *Latrunculia* genus and that the betaproteobacteria associated in these sponges showed a similar level of phylogenetic relatedness. The spirochete identified in the *Cyclacanthia* sponge was more distantly related to those detected in *Tsitsikamma* sponges indicating that this symbiont may have been acquired more recently, possibly before *Tsitsikamma* and *Cyclacanthia* sponges diverged. These spirochetes may be involved in the tailoring of pyrroloiminoquinone alkaloids resulting in tsitsikammamines or perform a different ecological role required in *Tsitsikamma* and *Cyclacanthia* sponges that isn't needed in *Latrunculia* species.

Chapter 4 - Metagenomic Analysis of the *T. favus* Microbiome

4.1. Introduction

Amplicon libraries of 16S rRNA gene fragments have been useful for characterisation of sponge-associated bacterial communities but do not provide insight into the functional roles of these associated microbes in the holobiont. Advances have been made in the development of cultivation conditions for the isolation of sponge-associated bacteria (Esteves et al., 2016). However, the cultivation and isolation of sponge-associated bacteria remains challenging due to the inability to mimic the sponge environment, varying levels of bacterial cell viability after laboratory processing of samples and long growth periods. Thus, cultivation and isolation of pure sponge-associated bacteria is not an efficient route for identification of biosynthetic gene clusters in symbiotic sponge isolates (Christensen and Martin, 2017).

Metagenomic approaches have led to a revolution in the identification of biosynthetic gene clusters in the largely underexplored unculturable "microbial dark matter" (Rinke et al., 2013). This has already led to the discovery of many novel biosynthetic gene clusters encoding several bioactive compounds from a diverse set of host organisms (Owen et al., 2012; Wilson et al., 2014; Lackner et al., 2016; Agarwal et al., 2017). Shotgun metagenomic techniques rely on the direct extraction, sequencing, and analysis of total gDNA from all organisms present in an environmental sample (Handelsman, 2004). This culture-independent approach provides insight into the functional and biosynthetic potential of several bacterial associated with the environmental sample and includes reduced biases related to amplification and 16S rRNA gene copy numbers associated with amplicon libraries (Clooney et al., 2016).

Despite the advantages, there are still several limitations to consider when adopting a metagenomics approach to characterising sponge microbiota. One problem is the volume of sponge-host DNA relative to sponge-associated bacterial DNA (Sharpton et al., 2014). This limitation can be overcome by enriching for the bacterial consortia by manual separation from the host tissue prior to metagenomic sequencing (Schirmer et al., 2005; Fieseler et al., 2006; Thomas et al., 2010b) or through computational

separation of host and bacterial DNA sequences via binning (Wang and Jia, 2016). Other limitations to this approach include cost of sequencing, large and complex computational requirements, non-uniform coverage of distinct bacterial genomes and biases in functional gene annotation (Quince et al., 2017). However, some of these limitations can be overcome by careful study design and the choice of appropriate computational algorithms.

The use of shotgun metagenomics was first employed by Edward F. DeLong where large DNA fragments were sequenced to characterize genes from uncultivated archaebacterial species from coastal water in order to infer potential function (DeLong, 1992). It later employed to investigate and compare the microbial communities associated with six *Arenosclera brasiliensis* sponge specimens as well as two associated seawater samples using 454 pyrosequencing (Trinidade-Silva et al., 2012). A comparative analysis between sponge and seawater metagenomic data revealed a clear distinction in associated bacterial community composition between the sponge and seawater samples. This analysis also revealed the composition of the core microbiome of *A. brasiliensis* sponges. Taxa belonging to fourteen conserved bacterial phyla were identified within the core sponge microbiome. Annotation and identification of bacterial genes involved in metabolism revealed that sponge-associated bacteria carried multiple genes involved in membrane transport, carbon metabolism and degradation of secondary metabolites which were suggested to aid bacterial survival in the presence of toxic arenosclerin compounds produced by the host sponge. In contrast, the metagenome of free-living bacteria in the seawater samples carried genes for virulence, defence, phages and transposable elements (Trinidade-Silva et al., 2012). This comparison showed that sponge-associated bacteria were specialised symbionts that maintained distinct functional genes relative to free-living marine bacteria. Although this early study did not include the assembly of sequence reads into contigs, it did demonstrate that potential functional roles of sponge-associated bacteria could be inferred from metagenomic data.

With rapid advances in technologies and the reduction in sequencing costs, metagenomic approaches have been developed and optimised for high throughput sequence data (Gao, 2018; Porter and Hajibabaei, 2018). Recent comparative studies between available sequencing platforms found that the HiSeq platform from Illumina was most suitable for shotgun metagenomic sequencing (Clooney et al., 2016). Ion

Torrent PGM and MiSeq platforms sequence reads are approximately 250 nt in length which are longer than those generated from the HiSeq platform, but the assembly of these longer reads into contigs is restricted by read number. The HiSeq platform has a comparatively lower read length (~90bp) but consisted of a 10-fold greater number of reads. This much larger raw dataset enables the assembly of longer contigs than that of MiSeq or Ion Torrent PGM reads. The contigs assembled from MiSeq sequence data were reported to be the most reliable in terms of potential gene function annotation as the low read length in HiSeq libraries led to fewer predicted complete genes (Clooney et al., 2016). However, the choice of genome binning software was found to play a greater role in reliable metagenomic sequence analysis than choice of sequencing platform.

A collaboration with Dr Jason Kwan at the University of Wisconsin, Madison provided access to a binning software called Autometa. Autometa is a new algorithm developed by the Kwan group for automatic clustering of contigs into genomic bins that employs a combination of gene marker recognition and machine learning methods that classify contigs according to k-mer frequency, coverage data and gene homology and is reported to outperform current genome binning packages (Miller et al., 2017, submitted).

The research reported in this chapter includes the analysis of two metagenome data sets. The first, a pilot study, was generated in 2015 using 454 pyrosequencing and the second was generated in 2017 using three Ion Torrent sequence libraries. The aim of the research in this chapter was to sequence the metagenome of *T. favus*, assemble sequence contigs, cluster the contigs into putative genomic bins and annotate the resultant bins for the identification of biosynthetic pathways.

4.2. Methods and Materials

4.2.1. Metagenome sequencing of *T. favus* TIC2014-001

T. favus sponge TIC2014-001 (collection metadata in Appendix A – Table A1) was manually cleaned of seaweed, shells, debris and other invertebrates (e.g. starfish, tube worms etc). The sponge was placed in a conical flask, covered with calcium and

magnesium free seawater (CMFSW) buffer (See Appendix B - Media) and placed on a shaker set to 200 rpm for 5 minutes to remove transient/food bacteria. The buffer was removed, and the wash step repeated with fresh CMFSW buffer. The sponge tissue was ground with 50 ml CMFSW buffer using a mortar and pestle. The following steps taken to separate the bacterial cells from the sponge host tissue was modified from the protocol set out in Thomas et al., 2010b. Collagenase D was added (500 µg/ml) and the sample was agitated on ice for 90 minutes. The sample was filtered through a 150 µm sieve and flow-through was centrifuged at 100 x g for 15 minutes at 4°C. The sponge cell pellet was removed and stored at -20°C. Supernatant was vacuum-filtered through Whatman® Grade 1 filters and the filtrate was centrifuged at 8800 x g for 20 minutes at 4°C. The supernatant was discarded and the pellet containing bacterial cells was re-suspended in 50 ml resuspension buffer (RB). Resuspension was centrifuged at 10000 x g for 20 minutes at 4°C. The supernatant was discarded, and the resuspension step repeated. The pellet was re-suspended in 5 ml RB, divided into 1 ml aliquots which were centrifuged at 13000 rpm for 5 minutes. The supernatant was discarded, and the pellet stored at -20°C.

Extraction of total gDNA from the bacterial pellet, remaining sponge cell pellet as well as from cleaned, whole *T. favus* TIC2014-001 was performed using the ZR Bacterial/Fungal DNA MiniPrep kit from Zymo Research (Cat. No. D6005) as per instructions provided. To determine the efficiency of the separation, amplicon sequence libraries of the V4-V5 region of the 16S rRNA gene from each fraction of TIC2014-001 viz. total sponge, the extracted bacterial pellet (BE) and the remaining sponge tissue after the bacterial extraction (RS) were generated by 454 pyrosequencing and curated using the Mothur software as described previously (Section 3.2.5).

Total gDNA isolated from the enriched bacterial pellet was sheared mechanically into fragments approximately 500 nt in size. Reads were sequenced using the GS Junior Titanium Sequencing platform at the Genome Sequencing Unit at Rhodes University. Attempts to assemble sequence reads into contigs proved unsuccessful. Therefore, the raw dataset was analysed using the online MG-RAST tool (V. 4.0.2) (Meyer et al., 2008) using default settings. A total of 31% of the sequence reads were identified as artificial duplicate reads and removed from the dataset and the remaining 71 968 reads were classified as bacterial using Genbank as the reference database. Gene

annotation predicted 73 853 protein-coding sequences and 277 rRNA features. Alignment of these features against the M5NR database resulted in the identification of 19 675 known protein features and 65 known rRNA features.

4.2.2. Metagenome sequencing of *T. favus* specimen TIC2016 - 050A

T. favus sponge specimen TIC2016-050A (collection metadata in Appendix A – Table A1) was manually cleaned as described above. Bacterial enrichment was not performed on this specimen. Extraction of genomic DNA was performed as described previously in Section 3.2.2. Metagenomic sequence libraries were generated using Ion Torrent sequencing platforms at the Central Analytical Facilities at Stellenbosch University (Table 4.1.).

Table 4.1: Metagenomic sequence data of TIC2016-050A

Sequencing platform	Average read length (bp)	No. of reads generated	Total sequence generated
S5 - 530	550	8 836 492	9.2 Gb
530 (A)	329	14 537 738	16. 5 Gb
530 (B)	363	12 710 484	15.7 Gb

Sequence reads were assembled into contigs using SPAdes (version 3.10.1) (Nurk et al., 2013) with the --iontorrent option. The debug mode was turned off and PHRED offset parameter set to auto-detect. K-mer parameters were set to 21, 33 and 55. Repeat resolution was enabled whilst mismatch careful mode, mismatch corrector and coverage cut-off were disabled. The scaffolds generated from SPAdes were used as input into the Autometa pipeline (Miller et al., 2017). Contigs with lengths less than 3000 nt were removed from the dataset. Contigs with single copy taxonomic marker genes were identified as “seeds” and categorised in Kingdom bins. Subsequent machine learning methods recruited contigs with similar k-mer frequency, nucleotide composition and coverage information as the “seeds” into genomic bins.

A total of 23 genomic bins were generated. FASTA files of contigs for each genomic bin were extracted for quality assessment using CheckM (Parks et al., 2015) in which

completeness and contamination levels are predicted based on presence and positioning of a collection of 104 of single-copy gene markers grouped into 58 collocated sets to estimate genome completeness and contamination. Potential genome contamination is calculated through the identification of multiple copies of single marker genes in a binned genome. Gene-calling and annotation was performed using the RAST Annotation Server (Brettin et al., 2015) as well as subsequent phylogenetic classification of genomic bin. This is achieved through identification and comparison of a set of conserved universal proteins and 200 single-copy protein encoding genes with genomes stored in the SEED database. All contigs taxonomically classified within the Eubacteria kingdom were annotated using the online tool BASys (Van Domselaar et al., 2005). Categorization of genes into "Clusters of Orthologous Groups" (COGs) and contigs annotated with secondary metabolic gene clusters were visualised in R using the "gplots", "RColorBrewer" and "ggplot2" packages. Identification of potential biosynthetic gene clusters (BGCs) was achieved through the use of antiSMASH bacterial version (version 4) (Weber et al., 2015).

4.3. Results

Metagenomic DNA was sequenced from a bacterial pellet of *T. favus* TIC2014-001 and from total gDNA isolated from *T. favus* TIC2016-050A using 454 pyrosequencing and IonTorrent sequencing respectively. The initial 454 library was constructed during a pilot project that aimed to develop metagenomic sequence analysis capacity in the Marine Natural Products Laboratory at Rhodes University. A total of approximately 160 000 reads were generated and reduced to 103 348 reads following quality control. Approximately 45 Mb of data was generated with an average read length of 435 nt. This pilot project provided understanding of which DNA preparation approaches, software packages and analysis pipelines were best suited to analysis of sponge-associated microbiota. During this time, it became apparent that too few reads were generated to facilitate the construction of contigs that were large enough to accommodate a biosynthetic gene cluster. A second set of three libraries were prepared and sequenced using the IonTorrent platform to generate a cumulative sequence library with a greater number of metagenomic sequence reads.

4.3.1. Assessment of bacterial enrichment protocol

To reduce the abundance of host-sponge gDNA, sponge tissue of TIC2014-001 was separated from bacterial cells. Comparison of 16S rRNA gene amplicon libraries was used to assess whether the bacterial enrichment protocol had been successful. Characterization of unique bacterial OTUs (clustered at a distance of 0.00) showed

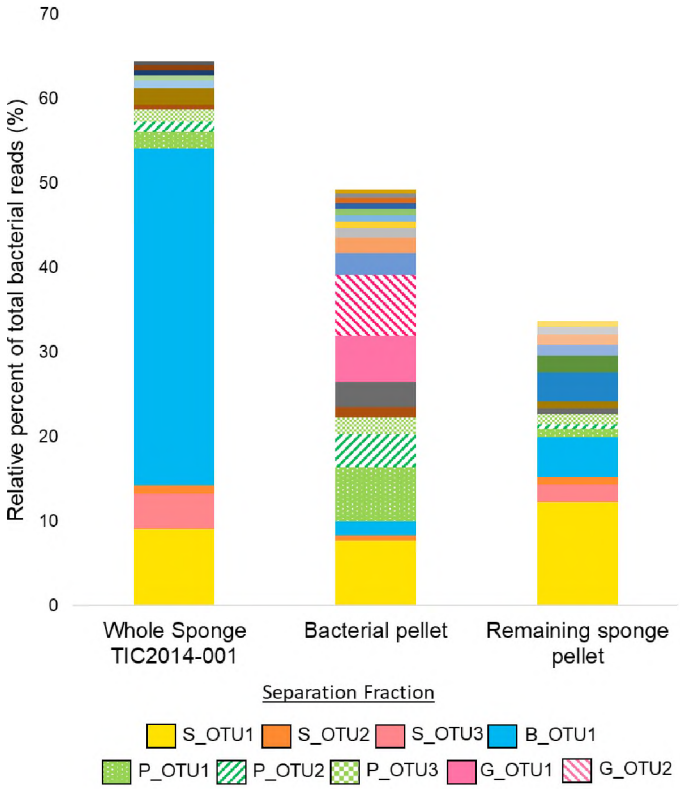


Figure 4.1: Dominant bacterial OTUs clustered at a distance of 0.00 present in the three *T. favus* TIC2014-001 sponge fractions created during bacterial enrichment. Read abundance is indicated as a relative percentage of total bacterial reads analysed for each respective sponge sample. Coloured key is provided for identification of key conserved spirochaete and betaproteobacterial OTUs.

that there was a significant difference between the bacterial communities' compositions in the different sponge fractions (Fig. 4.1). As expected, the bacterial diversity was higher in the bacterial pellet relative to the whole sponge from which it was extracted, whilst the remaining sponge pellet comprised of dominant populations of three identified spirochete OTUs as well as the dominant betaproteobacterial OTU.

The abundance of the dominant, conserved betaproteobacterial OTU, B_OTU1, was greatly reduced in the bacterial pellet fraction but relatively abundant in the remaining sponge pellet yet at a lower relative abundance than that observed for the whole sponge sample. This suggested that this betaproteobacterium may be intracellular or

tightly associated with sponge cells and as a result was not easily dislodged from sponge tissue. The conserved spirochaete OTUs, S_OTU1 and S_OTU3 (Refer to Fig. 3.4) were identified in all three sponge fractions but S_OTU2 was not detected in the bacterial pellet. These data suggest that the spirochetes may be tightly associated with sponge cells and that S_OTU2 may, like B_OTU1, be an intracellular symbiont. Due to its exclusive presence within tsitsikammamine-producing *T. favus* and *T. nguni* sponge species, bacterial strain S_OTU1 (Fig. 3.6) could possibly be involved in the production of tsitsikammamines and was detected at a relative abundance comparable to that observed in the whole sponge. Three planctomycete OTUs, P_OTU1, P_OTU_2 and P_OTU3 were detected in all three sponge fractions but were enriched in the bacterial cell pellet fraction. Similarly, two gammaproteobacteria, G_OTU1 and G_OTU_2 were detected in relatively high abundance in the bacterial pellet but were present in relative abundances less than 0.5% in the whole sponge. As only a limited number of reads could be generated using 454 pyrosequencing data and the aim of the study was to identify potential bacterial biosynthetic gene clusters involved in the production of pyrroloiminoquinone compounds, it was decided that gDNA from the enriched bacterial pellet would be sequenced for pilot project metagenomic analysis and that isolation of conserved dominant betaproteobacterium genome would be re-visited at a later period.

4.3.2. Phylogenetic analysis of *T. favus* metagenomic sequence data

Metagenomic analysis of the bacterial pellet extracted from *T. favus* TIC2014-001 using MG-RAST showed that aproximately 18% of the metagenomic sequences were classified as eukaryotic. A total of 22.6% of the eukaryotic sequences were classified within the Porifera phylum. Other phlya that contributed to the total number of eukaryotic sequences included Chlorophyta (18.43%), Streptophyta (12.57%), Chordata (11.10 %), and Bacillariophyta (11.86%). Viral sequence accounted for 8.09% of reads whilst bacterial read abundance was predicted to account for approximately 72.48 % of the dataset. Analysis of the bacterial sequence reads (Fig. 4.2) showed that Alphaproteobacteria (24.50%), Planctomycetacia (23.13%) and Gammaproteobacteria (10.60%) accounted for a large precentage of classified reads whilst bacterial reads classified as Betaproteobacteria and Spirochaetes accounted for only 2.96% and 0.64% respectively. Analysis of 16S rRNA gene amplicon libraries

showed an enrichment of Gammaproteobacteria and Planctomycete bacteria in the bacterial cell pellet (Fig. 4.1) and it is possible that these bacteria were not tightly associated with the host sponge tissue, which could account for the increased abundance of these bacteria in the metagenomic dataset.

Of the 73 853 predicted protein sequences, only 42 were classified within the Secondary Metabolites subsystem category. It was found that 24 of these genes were predicted to be sponge derived arylsulfatases and 18 were involved in the production of auxins. The aim of this study was to identify potential gene clusters involved in the biosynthesis of tsitsikammamines. The small dataset generated from 454 pyrosequencing prevented reliable assembly of reads into contigs and subsequent identification of gene clusters and the number of reads obtained from target bacterial, the spirochetes and betaproteobacteria were low. Therefore, a second shotgun metagenomic library was prepared from *T. favus* TIC2016-050A.

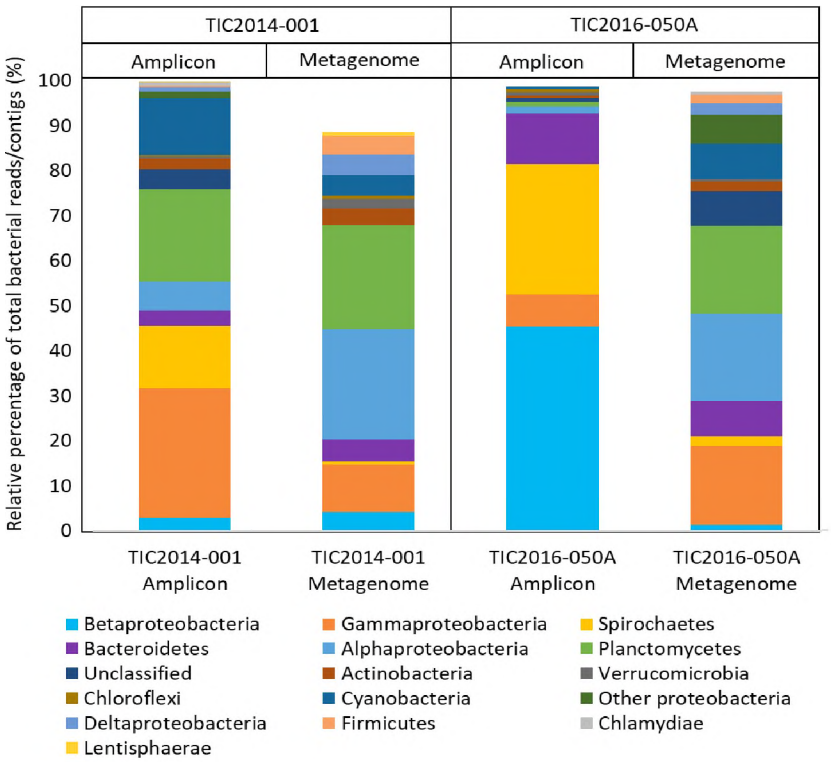


Figure 4.2: Comparative bacterial community composition of *T. favus* sponge specimens TIC2016-050A and TIC2014-001 generated from 16S rRNA amplicon and shotgun metagenomic sequence libraries. Read/Contig abundance is indicated as a relative percentage of total bacterial reads analysed for each respective sponge sample. Coloured key indicates phylogenetic classification.

The metagenome sequence library was generated using the Ion Torrent PGM platform which produced a far larger number of metagenomic sequence reads. Of the 23 107 contigs assembled from the combined TIC2016-050A metagenomic sequence libraries 43% were classified as being of bacterial origin, 28% were unclassified, 18% were classified as viral, 10% were classified as eukaryotic and only 0.5% were classified as archaeal. MG-RAST analysis showed that the distribution of bacteria in the metagenomic dataset of TIC2016-050A was dominated by taxa classified as Planctomycetes (19.65%), Alphaproteobacteria (19.25%) and Gammaproteobacteria (17.40%). Betaproteobacteria and Spirochaetes were present in abundances of 1.33% and 2.21% respectively (Fig. 4.2). An enrichment of Planctomycetes, Alpha- and Gammaproteobacteria was consistent with data obtained from the metagenomic dataset of TIC2014-001.

4.3.3. Identification of potential gene function within *T. favus* metagenomic sequence data

The metabolic potential of the TIC2016-050A and TIC2014-001 metagenomes was investigated using the KEGG orthologous gene database as a reference. The analysis revealed that the distribution of genes with predicted known function were consistent across the two metagenomic libraries and the majority of identified bacterial genes in both specimens were involved in amino acid, nucleotide and carbohydrate metabolism (Fig. 4.3). The abundance of secondary metabolite associated genes was predicted to to account for only 1% of the total number of identified genes in both metagenomic datasets. Of this 1%, the largest proportion of genes were predicted to be associated with the production of streptomycin (78.2%) as well as phenylpropanoid, penicillin and cephalosporin biosynthesis. A single gene involved in pyridine alkaloid biosynthesis was identified that was predicted to be present in the genome of a *Rhodopseudomonas* bacterium (Alphaproteobacterium) in the metagenomic library from TIC2016-050A.

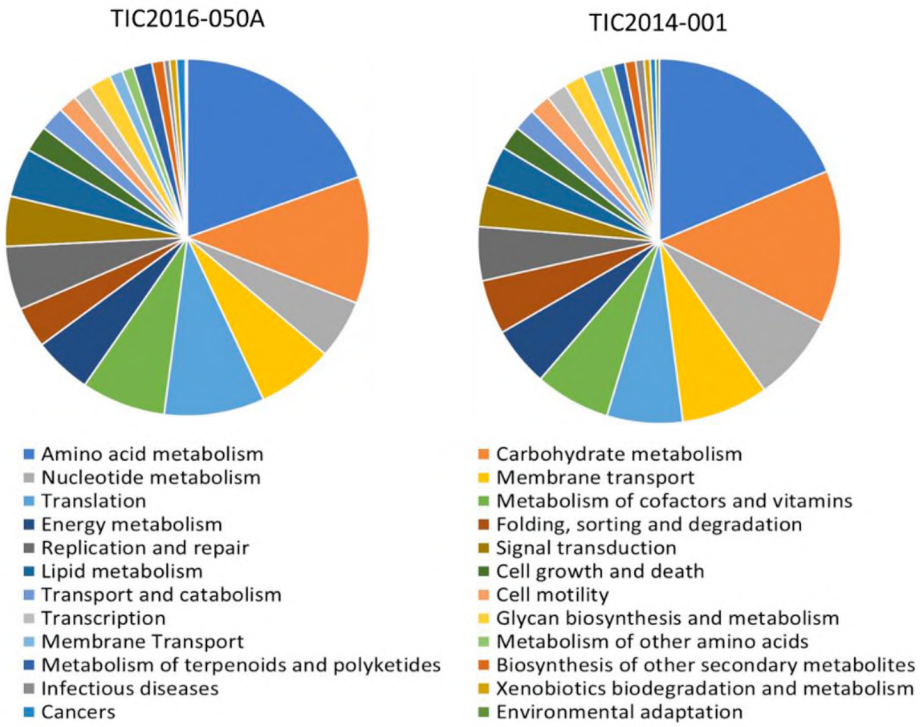


Figure 4.3: Functional annotations assigned to predicted genes identified in metagenomic sequence data of *T. favus* sponges TIC2016-050A and TIC2014-001. Genes were annotated with reference to the KEGG orthologous database with a cut-off identity of 60% and a minimum length of 15 amino acids. Coloured key indicates predicted gene function annotation.

4.3.4. Clustering of contigs into putative genome bins

Assembled contigs from the metagenomic library of TIC2016-050A were clustered into putative genomic bins using Autometa. Contigs carrying single marker genes were identified and used as "seed" contigs. Contigs from the remaining pool were then clustered with "seed" contigs if they shared similar GC content, pentamer frequencies and coverage data. Machine learning was implemented over several iterations to recruit new contigs into genome bins before re-assessment of the genome characteristics (Miller et al., 2017). A total of 23 genomic bins were generated (Fig. 4.4.) with 3745 of 9937 bacterial contigs classified as "unclustered". Phylogenetic classification and quality assessment of individual genomes revealed that Genomes 1 and 3 were the most complete of the binned clusters with predicted genome completeness of 98% and 95% respectively (Table 4.2).

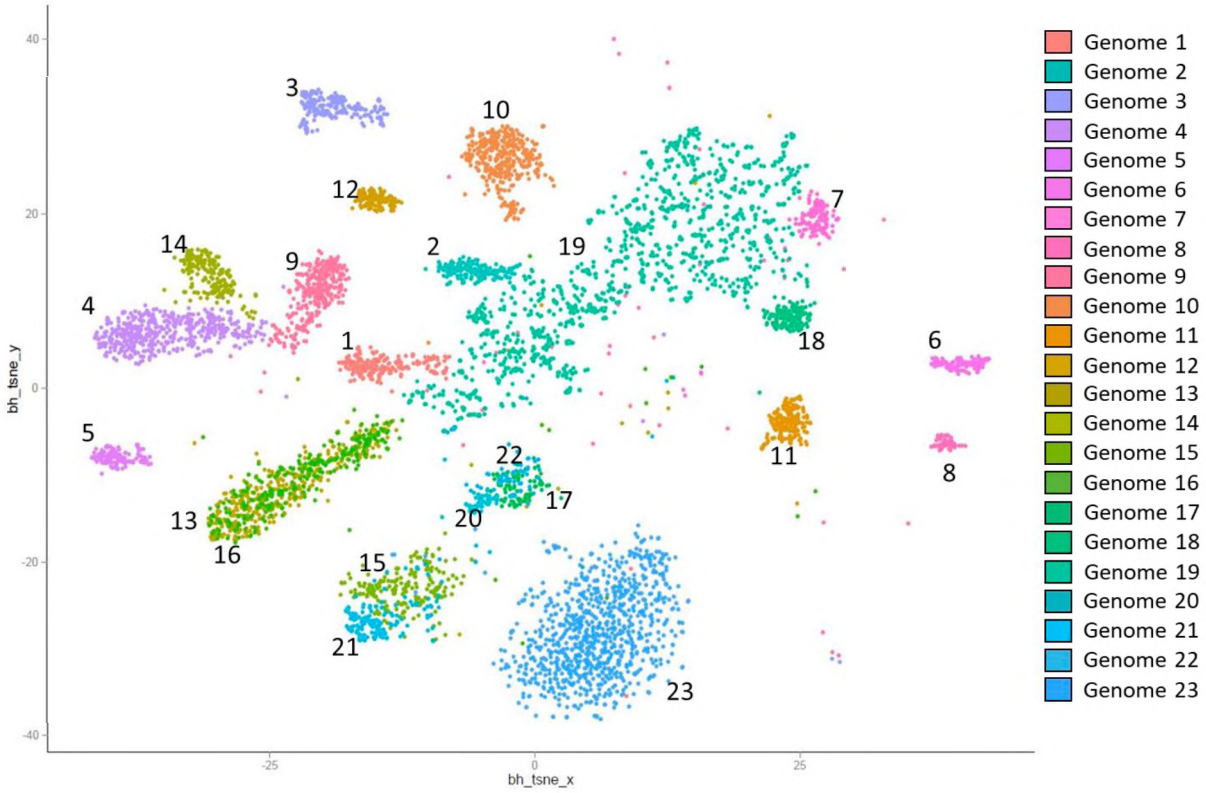


Figure 4.4: Genomic bins of clustered contiguous sequences from *T. favus* specimen TIC2016-050A. Coloured keys indicated genomes designated 1 – 23 with genome numbers indicated on each cluster.

Table 4.2: Characterisation of clustered genomic bins

Genome	Taxonomic classification (level)	Approximate size	Completeness (%)	Contamination (%)	No. of BGCs
Genome 1	*Pseudovibrio* sp. (genus)	6.25 Mb	98.33	0.88	33
Genome 2	Bacteriodetes (phylum)	1.27 Mb	60.80	1.99	2
Genome 3	*Salinispira* sp. (genus)	2.88 Mb	95.15	5.73	6
Genome 4	*Ruegeria* sp. (genus)	5.07 Mb	81.31	6.62	11
Genome 5	Actinobacteria (class)	0.63 Mb	18.12	0.00	0
Genome 6	Deltaproteobacteria (class)	0.55 Mb	23.00	0.00	0
Genome 7	Candidatus *Pelagibacter* (genus)	1.36 Mb	64.45	3.40	2
Genome 8	Chlamydiales (order)	1.03 Mb	61.39	0.38	1
Genome 9	Sphingomonadales (order)	3.16 Mb	82.81	2.15	8
Genome 10	Moraxellaceae (family)	2.98 Mb	81.45	0.81	4
Genome 11	Gammaproteobacterium (class)	1.21 Mb	58.52	0.82	0
Genome 12	Gammaproteobacterium (class)	0.63 Mb	30.80	1.88	0
Genome 13	Cyanobacterium (phylum)	1.69 Mb	39.26	11.10	0
Genome 14	Rhodobacteraceae (family)	0.80 Mb	15.67	0.50	0
Genome 15	*Rhodopirellula* sp. (genus)	7.18 Mb	80.12	2.94	33
Genome 16	*Synechococcus* sp. (genus)	1.58 Mb	25.01	1.54	0
Genome 17	Planctomycetaceae (family)	1.45 Mb	23.41	0.00	10
Genome 18	Rickettsiaceae (family)	0.92 Mb	58.01	1.42	0
Genome 19	Proteobacterium (phylum)	10.65 Mb	13.43	38.29	10
Genome 20	Planctomycetaceae (family)	0.71 Mb	29.46	0.00	1
Genome 21	*Rhodopirellula* sp. (genus)	3.74 Mb	40.34	2.65	10
Genome 22	*Pirellula* sp. (genus)	0.92 Mb	32.06	1.18	2
Genome 23	Planctomycetaceae (family)	8.52 Mb	69.31	30.29	22

Categorization of gene into functional COG classes revealed that Genomes 7 and 11 have several genes involved in energy production/conversion and amino acid transport and metabolism (Fig. 4.5) and may account for the majority of the carbohydrate and amino acid metabolism genes identified within the *T. favus* metagenome (Fig. 4.3). Genomes 7 and 11 were taxonomically identified as Alpha- and Gamma-proteobacteria (Table 4.2), with the former identified as a species within the ubiquitous marine genus *Pelagibacter*.

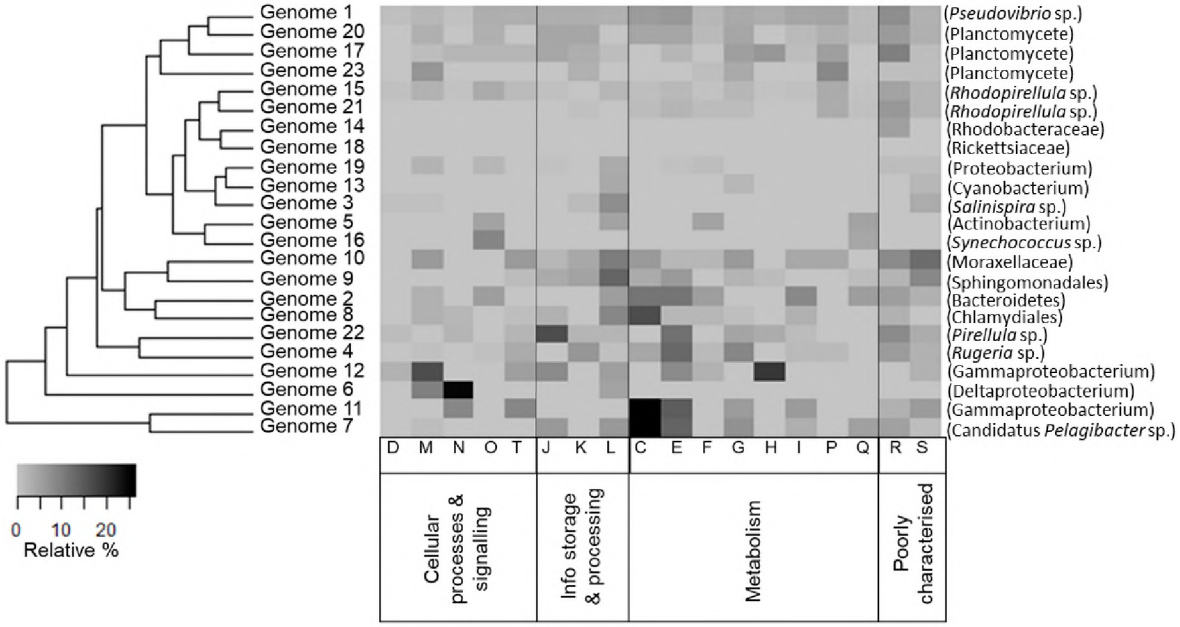


Figure 4.5: Distribution of COG classes in clustered bacterial genomes found in *T. favus* specimen TIC2016-050A. COG functional classes have been grouped into general functional categories indicated by boxed regions. COG functional classes are indicated by letter associations: D: Cell division and chromosome partitioning, M: Cell envelope biogenesis, outer membrane, N: Cell motility, O: Posttranslational modification, protein turnover, chaperone, T: Signal transduction mechanisms, J: Translation, ribosomal structure and biogenesis, K: Transcription, L: DNA replication, recombination and repair, C: Energy production and conversion, E: Amino acid transport and metabolism, F: Nucleotide transport and metabolism, G: Carbohydrate transport and metabolism, H: Coenzyme metabolism, I: Lipid metabolism, P: Inorganic ion transport and metabolism, Q: Secondary metabolites biosynthesis, transport, and catabolism, R: General function prediction only, S: COG of unknown function. Similarity in potential gene function annotation is inferred through a dendrogram generated by re-ordered clustering of the genome data matrix.

The *Pelagibacter* sp. genome detected in the metagenome of *T. favus* is predicted to be 1.36 Mb in size which is consistent with reports that this bacterium carries only the most necessary genes for survival (Giovannoni et al., 2005), however, as the genome is estimated to be only 64% complete suggesting that it may be larger. Closer inspection of the genome revealed the presence of six genes encoding proteorhodopsin, which is a retinylidene protein that functions as light-driven proton

pumps to generate ATP. *Pelagibacter* sp. accounted for 0.3% - 2.6% of the total bacterial sequence reads identified in seawater samples, whilst none were detected in the marine sediment. Relative abundances of *Pelagibacter* sp. present in *T. favus* sponges TIC2014-001 and TIC2016-050A were 0.09% and 0.47% respectively. Since sponges filter large quantities of water, it is unsurprising that these bacteria were detected in the metagenome of TIC2016-050A and they are likely present as transient bacteria rather than associated sponge microbiota as they are not conserved in all *T. favus* sponge specimens.

Genome 10 was identified as a gammaproteobacterium, of the family Moraxellaceae, and was predicted to be approximately 82% complete with only 0.81% contamination. Within CheckM, a single-copy marker gene is defined as a gene that is only present once in within more than 97% of validated, complete genomes. Only four biosynthetic gene clusters were identified in this genome of which three were classified as putative and one was predicted to produce a saccharide. Closer inspection of Genome 10 revealed the presence of several genes encoding proteins potentially involved in the production of fatty acids, isoprenoids, and lipids, as well as genes involved in the metabolism of sulphur and phosphorus. Notably, there were many genes detected that were predicted to be involved in resistance to toxic and antibiotic compounds, particularly those encoding multidrug-resistance efflux pumps. These efflux pumps indicate that the gammaproteobacterium represented by Genome 10 is adapted to live in the toxic environment of a *T. favus* sponge and thus may exist in symbiosis with the sponge and play a role in sulphur and/phosphorus cycling within the host sponge.

Analysis of bacterial contigs using the antiSMASH server revealed the presence of 175 biosynthetic gene clusters (Fig. 4.6). Most of the clusters classified as “Putative” (56.06%/), fatty acids (15.6%) and terpenes (4.64%). Gene clusters predicted to encode NRPS, PKS or hybrid biosynthetic machinery had a cumulative total of 4.04% and were identified in Genomes 1, 15, 17 and 21.

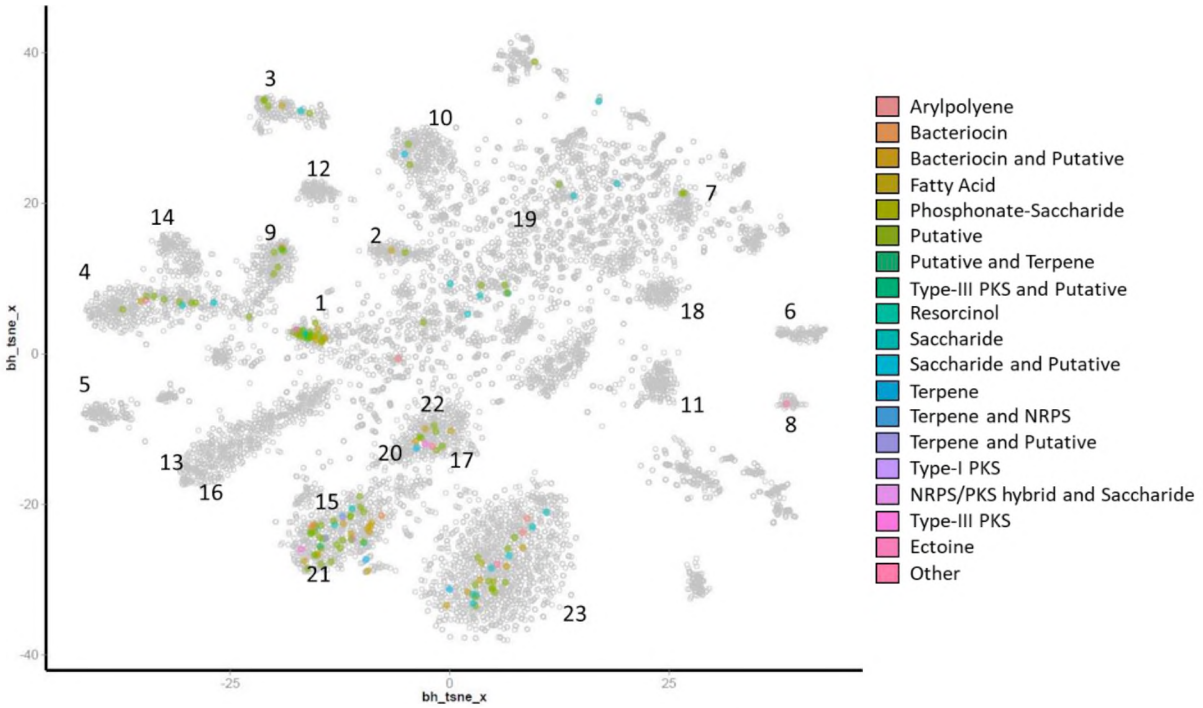


Figure 4.6: Distribution of secondary metabolite biosynthetic gene clusters in bacterial genomic bins from *T. favus* TIC2016-050A. Coloured keys indicate biosynthetic gene cluster classification with genome numbers indicated on each cluster.

Genome 1 was classified as a *Pseudovibrio* sp. (Table 4.2). Genome 1 was approximately 6 Mb in size with less than 1% predicted contamination (Table 4.2.). Inspection of the annotated genome revealed that Genome 1 devoted a large proportion of its genome to the production of membrane transporters as well as a total of 85 genes predicted to confer resistance to toxins and antibiotics, particularly within the class of multidrug efflux pumps. A total of thirty-three biosynthetic gene clusters were detected in Genome 1 with twenty-three classified as putative, four were predicted to produce fatty acids, two were predicted to produce saccharides and the remaining four encoded a bacteriocin, a terpene, an NRPS and a Type-III PKS. These pathways are discussed in detail in Chapter 6.

Genomes 15, 17 and 21 were all classified within the Planctomycetaceae family of bacteria. Genome 15 was estimated to be approximately 80% complete and included several genes encoding genes predicted to be involved in both sulphur and phosphorus metabolism. The genome of this bacterium was also predicted to include several genes encoding multidrug efflux pumps as well as efflux systems which specifically export macrolides and acriflavins. Notably, four genes were predicted to encode internalin-like proteins. Internalin is a surface protein of *Listeria*

monocytogenes that enables its invasion into non-phagocytic cells of a multicellular host (de Silva et al., 2012). A total of 25 proteins containing tetracopeptide repeat domains (TPRs) as well as 9 proteins including ankyrin repeats were identified within this genome. TPRs and ankyrin repeats are commonly identified within bacterial sponge symbionts (Reynolds and Thomas, 2016). The cumulation of these genes would suggest that Planctomycete Genome 15 represented a potential symbiont of *T. favus* that, like the gammaproteobacterium represented by Genome 10, is resistant to various toxins that may be present in the host *T. favus* sponge and may be involved in the metabolism of sulphur or phosphorus in the sponge host.

Genome 19 was predicted to be over 10 Mb in size but only 13.43% complete. Annotation using the RAST server revealed that the closest phylogenetic relative of Genome 19 was *Ralstonia solanacearum*, a betaproteobacterium isolated from soil that is a pathogenic to plants. This genome could potentially represent the dominant B_OTU1 strain identified in *T. favus, T. nguni* and *T. pedunculata* sponges. A total of ten biosynthetic gene clusters were detected in Genome 19, five of which were classified as putative, four were classified as producing saccharides and one was predicted to be responsible for the production of an arylpolyene. Also present was a single lantithionine biosynthesis protein was identified within Genome 19 that shared greatest amino acid sequence identity (30 – 34%) with several alphaproteobacterial and cyanobacterial species, but not with the lantipeptide genes identified in *Streptomyces* sp. and *S. pacifica* sp. that produce pyrroloquinoline compounds. Strikingly, over 31% of the genes identified within this genome were classified as originating from phages and prophages which would suggest extensive lateral gene transfer events occurring within this putative genome bin. However, the percentage of contamination was predicted to be over 38% and this would suggest that Genome 19 consists of multiple unresolved genomes and that the biosynthetic gene clusters identified upon the clustered contigs may well belong to several different bacterial species within the proteobacterial phylum and only one among them may represent the conserved betaproteobacterium B_OTU1.

Genome 3 was identified as a *Salinispira*, a species within the Spirochaetaceae family with contamination levels predicted to be approximately 5%. It was previously shown through 16S rRNA amplicon sequencing (Chapter 3), that three conserved spirochete OTUs were associated with *T. favus* sponges *viz*. S_OTU1, S_OTU2 and S_OTU3

(Fig. 3.6) which were most closely phylogenetically related to *Salinispira pacifica* (Fig. 3.7). It was found that 5% of detected single-copy marker genes in this bin were present in multiple copies which suggested that Genome 3 may represent a mix of the three, conserved spirochete OTUs associated with *T. favus* sponges.

4.4. Discussion

The aim of the research described in this chapter was to investigate the potential of metagenomic sequencing to identify sponge-associated bacterial symbionts that might be responsible for the biosynthesis of tsitsikammamines or other pyrroloiminoquinone compounds, in *T. favus* sponges.

4.4.1. Pilot metagenomic study

The first objective was to conduct a pilot metagenomic study of *T. favus* to evaluate the efficiency of separating bacterial cells from whole sponge tissue, to assess the depth of sequencing required to perform a satisfactory metagenomic sequence analysis and to select optimal software packages for downstream analysis. Enrichment of bacterial cells from host cell tissue was reasonably successful, in that the majority of sponge gDNA was removed from the sample. Notably, it was found that a large proportion of the conserved, dominant betaproteobacterium and spirochete S_OTU1, remained in the sponge pellet and was not well represented in the enriched bacterial cell fraction. These results suggest that these bacteria may be located within the sponge cells (intracellular). Attempts to localize these bacteria using confocal imagery coupled with fluorescent labelling of specific bacterial clades proved difficult due to inherent autofluorescence of the sponge tissue. Various efforts to quench autofluorescence either proved to be inadequate or damaged the tissue to severely. Inspection of the sponge cells using electron microscopy could potentially be employed to determine the localisation of conserved, dominant betaproteobacterial and spirochete OTUs of interest.

Over 100 000 sequence reads were generated from the gDNA of the bacterial cell pellet and assembly of this data into contigs proved inefficient indicating that a larger dataset would be required to assemble sufficiently large bacterial contigs for the

identification of biosynthetic gene clusters. Thereafter, a large scale metagenomic sequencing effort was undertaken to provide sufficient data for the assembly of numerically dominant bacteria into genome bins that could be used to identify putative biosynthetic gene clusters.

Separation of bacterial cells from sponge tissue was performed with *T. favus* specimen TIC2014-001 and analysis of amplicon sequence data from each of the fractions showed that dominant bacterial betaproteobacterial phylum was greatly reduced in the bacterial pellet relative to the whole sponge. It was shown in Chapter 3 that the conserved betaproteobacteria are not only numerically dominant but metabolically dominant as well, suggesting that this phylum may provide essential functions within the *T. favus* sponge. The observed reduction of betaproteobacteria in the extracted bacterial cell pellet could indicate that they are intracellular endosymbionts found within the sponge cells rather than within the mesohyl, such as the observations made regarding potentially intracellular alphaproteobacterial strain in *Rhopaloides odorabile* (Webster and Hill, 2001) and intracellular calcifying bacteria identified in *Hemimycale columella* (Garate et al., 2017). Furthermore, a conserved, dominant betaproteobacterium was identified in an *A. queenslandica* sponge that carried a reduced genome (Gauthier et al., 2016) which is a common feature among endosymbiotic bacteria (McCutcheon and Moran, 2011; O'Malley et al., 2016).

It was further noted the spirochaete OTU, S_OTU2, was absent from the enriched bacterial pellet. Like the betaproteobacterium discussed prior, this bacterium may have remained in the sponge pellet if it were an intracellular symbiont and therefore could not be manually separated from the host tissue through centrifugation or filtration steps. Endosymbiotic spirochetes have been documented within protists associated with termite guts and were found to perform a majority of reductive acetogenesis and nitrogen fixation processes in this environment, as well as aid in the movement of the host-protist (Ohkuma et al., 2015; Tu et al., 2015). Likewise, symbiotic spirochetes have been identified in wood-eating cockroaches (Sravanthi et al., 2016) and a marine red coral (van der Water et al., 2016) which would indicate that the observed conserved spirochete could be either an intracellular symbiont associated with either the sponge tissue or an associated protist harboured within the sponge. Genome 3 was identified as a Spirochaete species, most closely related to novel spirochete *Salinispira pacifica* which was isolated from a hypersaline microbial mat (Hania et al.,

2015). It was thought that this genome represented the conserved spirochetes identified in *T. favus* which may be involved in the production of tsitsikammamines. This genome is discussed in detail in the following chapter.

Conversely, it was shown that the abundance of three planctomycete strains and two gammaproteobacteria were enriched in the bacterial pellet. Very few studies, save the study conducted by Webster and Hill (2001), in which bacterial cells were separated from sponge tissue include an analysis of the bacterial communities present in the remaining sponge pellet nor a comparative study to the environmental sample as a whole. This finding therefore demonstrates that particular care must be taken when interpreting the results of previous studies in which there has been manual separation of bacterial cells from host sponge tissue. Within the context of this study, however, these planctomycete and gammaproteobacterial strains do not appear to be conserved within Tsitsikamma sponges nor within any of the Latrunculid sponges (Fig 3.9). This result would suggest that these bacteria are loosely associated with the sponge host. However, the manual separation was performed on a single sponge specimen and this process would need to be repeated with several more sponge samples to determine if this is a true trend during bacterial enrichment.

4.4.2. Comparison of approaches for bacterial community assessment

A large discrepancy was observed between the distribution of bacteria predicted through analysis of the 16S rRNA amplicon library and the phylogenetic classification of assembled shotgun metagenomic contigs. Lack of correlation between bacterial distribution predicted via the two methods employed here has been observed in other studies and can be attributed to differences in read coverage, bacterial DNA GC content, genome size, target gene copy number variation as well as PCR or sequencing errors (Chen et al., 2017; Delforno et al., 2017).

Discrepancies between amplicon and metagenomic sequencing approaches have been observed in other studies including assessment of microbial communities of stromatolites (Casaburi et al., 2016) and waste-water treatment reactors (Delforno et al., 2017). Differences in apparent abundances can be skewed by read coverage, GC content, target gene copy number variation as well as PCR or sequencing errors (Chen et al., 2017; Delforno et al., 2017). Various algorithms have been put forward to overcome these problems, such as, identification of species-specific taxonomic

markers (Chen et al., 2017) and identification of 16S rRNA sequences within shotgun metagenomic data (Guo et al., 2015), but each approach has various constraints and downfalls.

A comparative study was recently carried out to determine any biases the different sequencing and subsequent bioinformatic approaches have on a given bacterial community (Jovel et al., 2016). The results of this study revealed that neither 16S amplicon libraries nor metagenomic sequence analysis could accurately reflect a given community structure, with primer bias playing a large role in 16S rRNA amplicon library inaccuracies and regions of high similarity obscuring the precision of taxonomic assignment of metagenomic data (Jovel et al., 2016). It was suggested that an integration of genomics, metatranscriptomics, proteomics and metabolomics be used for accurate models of microbiomes (Jovel et al., 2016).

The distribution of assignment of predicted gene function appeared consistent in both *T. favus* specimens TIC2014-001 and TIC2016-050A with the largest proportion of genes with predicted known function involved in the metabolism of amino acids and carbohydrates (Fig. 4.3). Heterotrophic marine bacteria are believed to be responsible for a large proportion of inorganic compound oxidation and play an important role in carbon and nitrogen cycles within the ocean environment (Edwards et al., 2010; Li et al., 2014). An investigation into the metabolic capabilities of bacteria associated with the deep-sea sponge *Neamphius huxleyi* and shallow-water sponge *Theonella swinhoei* (Li et al., 2014; Li et al., 2016a) revealed similar results to that observed in *T. favus* sponges, that the majority of bacterial genes were involved in carbohydrate and amino acid metabolism and originated from proteobacterial species. However, an earlier study of six different sponge species found that sponges hosted distinct microbial consortia that achieved equivalent nutrient cycling using different enzymes (Fan et al., 2012) and thus bacterial community dominance cannot be strictly correlated to nutrient cycling abilities.

4.4.3. Analysis of metabolic potential in *T. favus*-associated microbiome

Contigs from TIC2016-050A were clustered into 23 genomic bins using the Autometa algorithm with levels of completeness ranging from 13% to 98%. The greatest number of genes linked to carbohydrate and amino acid metabolism were identified in Genome

7 (*Pelagibacter* sp.). *Pelagibacter ubique* is an alphaproteobacterium that is the major microbe in seawater and may be the most abundant bacterial species on Earth (Morris et al., 2002). *P. ubique* has a tiny genome that is purported to carry the minimum number of genes to sustain life and encodes no genes for secondary metabolic compounds (Perez and Fenical, 2017). Genome 7 was most likely acquired as a food bacterium from the surrounding seawater and thus no longer considered a genome of interest.

Genome 19 was identified as a proteobacterium using the CheckM algorithm and was estimated to be only 13% complete. Full genome comparison of Genome 19 using the RAST annotation pipeline revealed that it is most closely related to a betaproteobacterium. There was a predicted 38% level of contamination which indicated that single-copies genes were found in several replicates and that Genome 19 does not represent a single bacterium. This evidence in combination with the broad spread of contigs (Fig. 4.4) would suggest that Genome 19 could consist of multiple, unresolved genomes. Additional sequence data coupled with more stringent limitations during clustering of contigs into genomes may reduce the level of contamination observed here and provide more accurate insight into this genome and whether it represents the conserved, dominant bacterium identified using 16S rRNA amplicon sequence data.

Genomes 10, 11 and 12 were classified as gammaproteobacteria. Genome 10 was found to include several genes involved in the metabolism of sulphur and phosphorus. Recent studies have shown a symbiotic gammaproteobacterium performs a pivotal role in the life-cycle of *A. queenslandica* sponges and is capable of sulphur oxidation (Gauthier et al., 2016; Fieth et al., 2016). It is tempting to speculate that gammaproteobacterial bacteria represented by Genome 10 may be important to the health of *T. favus* sponges.

Genomes 1 and 15 included the greatest number of biosynthetic gene clusters including Type 3 PKSs, NRPS and hybrid gene clusters. Both bacteria represented by these genomes were predicted to encode various proteins, such as multidrug efflux pumps, that allowed them to survive within a toxic environment and provide potential nutrient and defensive benefits to the host *T. favus* sponge. Type 3 PKS gene clusters are similar to PKS gene clusters found in plants and are responsible for the assembly

of small aromatic metabolites (Moore and Hopke, 2001). Type 3 PKSs assemble a linear polyketide intermediate through a series of condensation steps which is then cyclized to form the final product (Shimizu et al., 2016). Pyrroloiminoquinone compounds are cyclic in nature and may be produced through Type 3 PKSs. Analysis of the four identified Type 3 PKSs, as well as other secondary metabolite gene clusters will be addressed in Chapter 6.

Chapter 5 – Genome characterisation of a *T. favus*-associated Spirochaete species

5.1. Introduction

In this and previous studies (Chapter 3, Walmsley et al., 2012; Matcher et al., 2017) *T. favus* and *T. nguni* sponges have been shown to be associated with a unique spirochete, S_OTU1. The conservation and host-specificity of this OTU with tsitsikammamine-producing sponges indicated that it may be involved in the biosynthetic modification or production of these pyrroloiminoquinone alkaloids. Metagenomic sequencing of *T. favus* specimen TIC2016-050A and subsequent clustering of assembled contigs into genomic bins resulted in the generation of 23 putative bacterial genomes. Genome 3 was taxonomically classified within the *Salinispira* genus. Since 16S rRNA gene amplicon sequence analysis had indicated that there were potentially three Spirochaete strains present in *T. favus* sponges, it was suggested that the observed contamination in Genome 3 may have been due to the genomic content of the three strains being clustered into a single genome (Chapter 4).

Salinispira was proposed as a new genus in 2015 when the complete genome sequence of novel spirochaete, *Salinispira pacifica*, was elucidated (Hania et al., 2015). The genome of *S. pacifica* was approximately 3.78 Mb in size, with a GC content of 51.9% and 3450 protein-encoding genes of which 1.13% encoded genes were predicted to be involved in secondary metabolite biosynthesis (Hania et al., 2015). *S. pacifica* was isolated from a hypersaline microbial mat (Hania et al., 2015) but spirochetes have been identified in marine sponges where several distinct morphotypes have been observed to inhabit a single sponge (Neulinger et al., 2010). To date, there is no information on the putative roles that these bacteria may play within the host sponge. Dominant populations of spirochetes have been observed in other marine organisms, including corals, arthropods, molluscs and oligochaetes where it is hypothesised that they play a role in host physiology (van der Water et al., 2016).

The aim of the research presented in this chapter was to annotate the draft *Spirochaetes* sp. genome (Genome 3) identified in *T. favus* TIC2016-050A and to elucidate the potential role of these bacteria within the host sponge.

5.2. Methods and Materials

The 207 contigs that were binned into Genome 3 were uploaded to the RAST server (Overbeek et al., 2014) and annotated using the Classic annotation option using the RAST gene calling algorithm and the FIGfams V7.0 protein dataset. Frameshift correction and identification of missing genes was enabled. Resultant features of the genome were downloaded, quantified, and plotted (http://rast.nmpdr.org/ : Accessed Oct-Dec 2017)

Full length 23S and two 16S rRNA gene sequences were recovered from the RNA subset of RAST-annotated genes and compared with the NCBI Genbank database using BLAST. The 16S rRNA gene sequences were aligned against closest relatives retrieved from the BLAST analysis using ClustalW algorithm and phylogenetic analysis was performed using a maximum-likelihood tree using the Tamura-Nei model with 1000 bootstrap replicates using MEGA7 software.

Genome 3 was analysed against the publicly available *Salinispira pacifica* strain L21-RPul-D2 genome (CP006939.1) for genome-to-genome distance using the online Genome-to-Genome Distance Calculator (GGDC 2.1) (http://ggdc.dsmz.de/distcalc2.php; Meier-Kolthoff et al., 2013) and the average nucleotide identity level using the online EzGenome ANI tool (https://www.ezbiocloud.net/tools/ani; Yoon et al., 2017). Assembled contigs in Genome 3 were aligned against the genome of *S. pacifica* strain L21-RPul-D2 using progressiveMauve with default parameters in Mauve version 20150226 build 10 (Darling et al., 2010) for contig reordering.

Biosynthetic gene clusters were detected using the online antiSMASH server (Bacterial version, Version 4) where Cluster Finder, BGC border prediction and ClusterBlast were enabled. Functions of individual ORFs were assigned using the NCBI BLASTp algorithm (https://blast.ncbi.nlm.nih.gov/Blast.cgi) searching against

the Non-redundant protein sequences (nr) database in conjunction with the NCBI Conserved Domains Database (https://www.ncbi.nlm.nih.gov/cdd).

5.3. Results

RAST annotation and analysis showed that Genome 3 consisted of 3024 coding regions, 2976 of which were protein-encoding whilst the remaining 48 features were identified as rRNA and tRNA sequences. The genome was approximately 2.8 Kb in size, with a GC content of 48.3% and was predicted to be potentially missing 41 genes. The annotation led to the recovery of two full-length 16S rRNA sequences (designated Genome 3A and 3B) that shared 90% nucleotide sequence identity with each other and shared greatest sequence homology with phylogenetically distinct spirochete clones previously isolated from *T. favus* specimen TIC2009-002, collected in Algoa Bay, South Africa (Fig. 5.1).

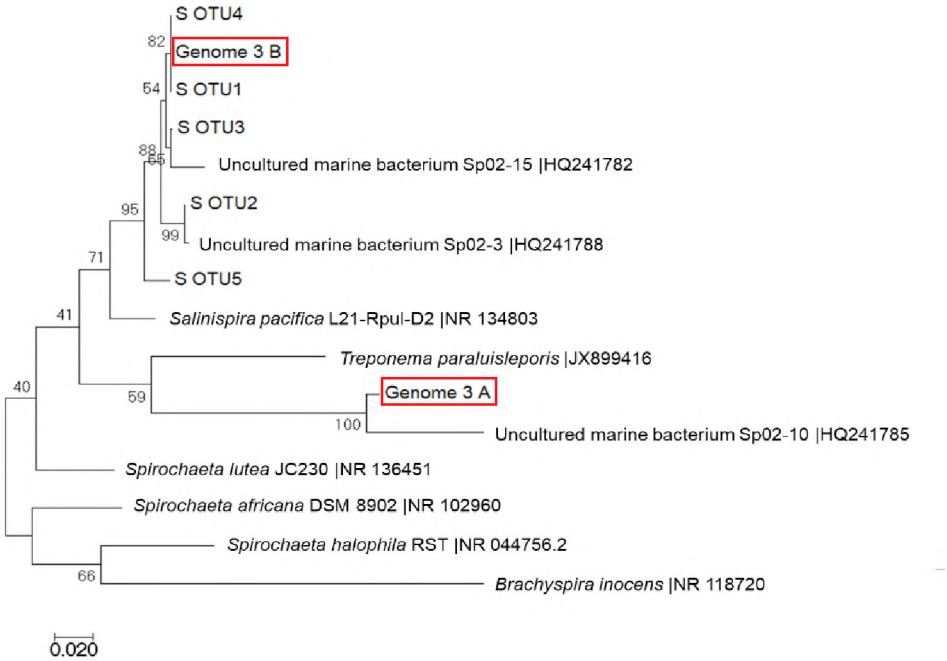


Figure 5.1: Phylogenetic relationship of spirochaete taxa identified in Genome 3 relative to spirochete OTUs associated with *Tsitsikamma* and *Cyclacanthia* sponges (Chapter 3) and other spirochaete strains based on full length 16S rRNA gene sequence generated using MEGA7 software with the Maximum-likelihood method. Bootstrap values were calculated using 1000 replicates and are indicated at branch nodes. The number of substitutions per site, over 1468 nucleotides, is indicated with a scale bar. 16S rRNA gene sequences recovered from Genome 3 are highlighted with red boxes.

It was shown that the 16S rRNA sequence of Genome 3B clustered with spirochete OTUs associated with *T. favus* and *T. nguni* sponges as well as *T. favus* associated spirochete clones Sp02-3 and Sp02-15 (Walmsley et al., 2012). The closest known relative of the spirochete represented by Genome 3B is *Salinispira pacifica* L2-RPul-D2 with which only 93% sequence identity is shared. The spirochete represented by Genome 3A shares only 90% sequence identity with the 16S rRNA of Genome 3B and is most closely related to a spirochete clone previously isolated from *T. favus* TIC2009-002 and shares only 83% sequence identity with its closest known relative *Treponema paraluisleporis*. The genome furthermore contained a single copy of the large 23S rRNA gene. Comparison of the 23S rRNA sequence using BLAST revealed that the greatest level of sequence homology (90%) was shared with *S. pacifica* strain L2-RPul-D2.

Whole genome comparison of Genome 3 and the publicly available genome sequence of *S. pacifica* showed that the bacterium represented by Genome 3 was a distinct species to *S. pacifica*. The average shared nucleotide identity was determined to be 68% between the two genomes using the online ANI Calculator. Reordering of contigs in Genome 3 revealed that 69% of contigs could be successful mapped to the genome of *S. pacifica* (Fig. 5.2). Inspection of the predicted function of the 655 genes encoded by the 64 contigs that were not mapped to the *S. pacifica* genome showed that over half of the encoded proteins were hypothetical. The contigs including the 16S rRNA genes were successfully mapped to *S. pacifica* and therefore it cannot be determined whether the unmapped contigs belong to either Genome 3A or Genome 3B.

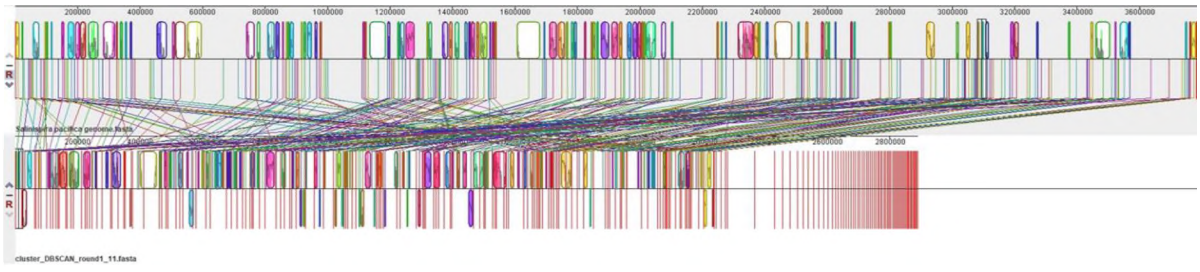


Figure 5.2: Contigs of Genome 3 aligned and re-ordered (bottom) relative to the genome of *Salinispira pacifica* L2-RPul-D2 (top) with the minimum weight of locally collinear blocks calculated to be 35 after 15 iterations. Regions of homology are indicated with blocks of matching colours with similarly coloured lines indicating a path between two homologous regions. Red lines indicate the borders of individual contigs in Genome 3.

Automated annotation using RAST indicated that the greatest number of putative genes (185) present in Genome 3 were involved in carbohydrate metabolism (Fig. 5.3), which on closer inspection included genes encoding enzymes involved in the degradation of pullulan, chitin, fructose, lactose, maltose, xylose, D-ribose, mannose, mannitol and glycerol amongst several others. The second highest proportion of genes (141) were predicted to be involved in protein metabolism, which consisted largely of genes involved in protein synthesis. Only 11 of these genes were predicted to aid in protein folding, which included two copies of *dnaJ*. Numerous genes were implicated in the production of amino-acids, as well as DNA and RNA metabolism.

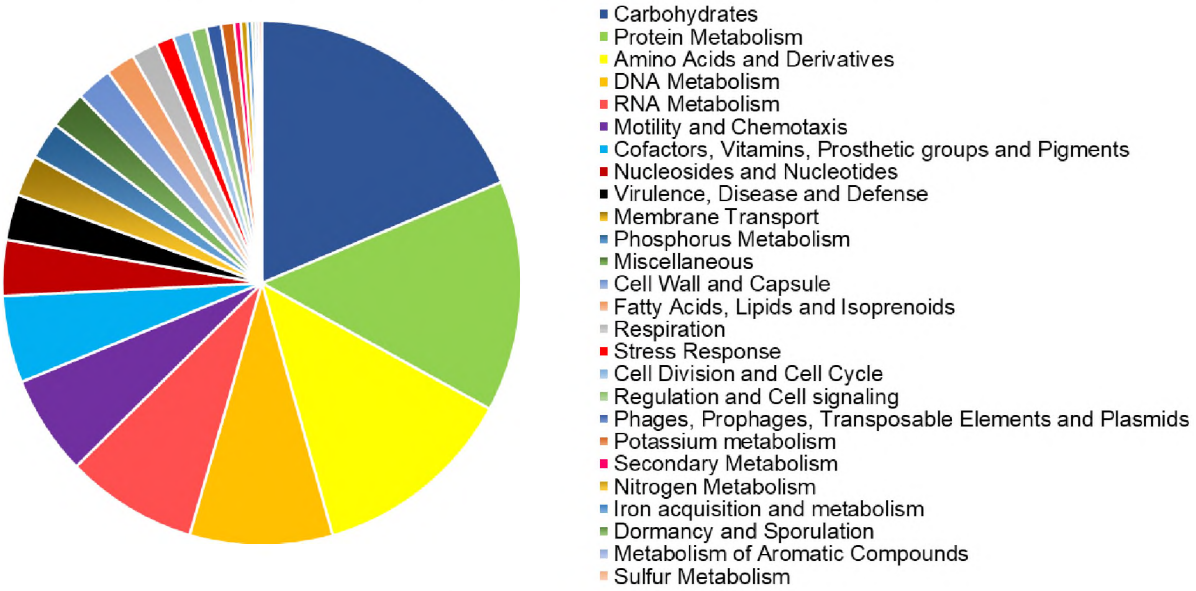


Figure 5.3: Distribution of genes with predicted known gene function within Spirochaete Genome 3. Genes were classified into 26 different functional categories which are indicated on a coloured key.

A total of 86 genes whose encoded products are predicted to be involved in restriction-modifications systems were also identified. A total of 28 genes predicted to encode proteins involved in virulence, disease and defence and the majority of these genes were predicted to be a part of a virulence operon similar to that found in *Mycobacterium tuberculosis*.

Several genes encoding enzymes that confer antibiotic resistance and are predicted to be involved in the production multidrug efflux pumps were also identified in Genome 3 and were predicted to provide protection against fluoroquinolones, various heavy metals, and beta-lactam antibiotics. A total 74 genes were found in Genome 3

encoding tetratricopeptide repeat domains that have previously been implicated in aiding bacteria-host interactions.

Several carbohydrate ABC transporter genes were identified that allowed the entry of various substrates including maltose, mannitol, chitin, D-ribose, and N-acetylglucosamine. Comparison of these genes to those identified in *S. pacifica* revealed that the Genome 3 spirochete was capable of similar carbohydrate metabolism. However, *S. pacifica* does not encode the genes necessary for metabolism of N-glucosamine, mannitol or inositol. Nearly all genes encoding proteins involved in the citric acid cycle with the exception of fumarase were identified in Genome 3 for the central metabolism of carbon.

RAST annotation revealed only six putative genes involved in secondary metabolite production in Genome 3, which were also were identified using antiSMASH. Two of these pathways were most likely fragments of larger pathways that could not be further resolved, with one likely to be responsible for peptidoglycan biosynthesis. Of the remaining three pathways, Cluster A was predicted to produce a fatty acid, Cluster B was classified as a saccharide-producing pathway whilst Cluster C was categorized as a "putative" gene cluster.

5.3.1. Cluster A: Biosynthesis of a fatty acid

Biosynthetic Cluster A includes 17 ORFs that span a region approximately 19 Kb in size (Fig. 5.4). Cluster A shares 24% of genes with similar clusters found on the genomes of marine myxomycete isolates *Enhygromyxa salina* and *Plesiocycstis pacifica.* A GNAT-N-acetyltransferase-encoding gene was identified at the one end of the cluster, which is followed by six genes that are predicted to be involved in fatty acid biosynthesis.

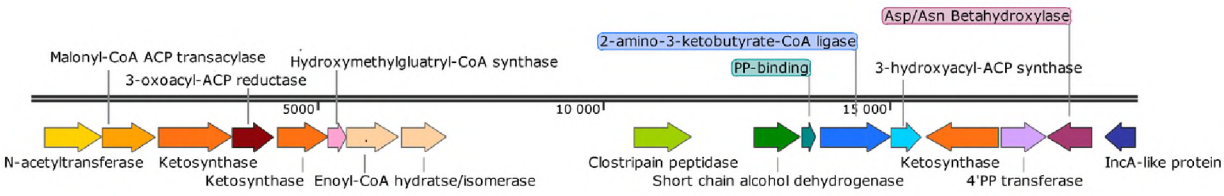


Figure 5.4: Gene organization of Cluster A in Genome 3. The putative function of each ORF has been indicated. All ORFs are depicted to scale and relative size is indicated with a horizontal scale bar marked with size in number of base pairs.

It is possible that the encoded GNAT N-acetyltransferase acts in tandem with the ACP-malonyl transacylase as an acetyl-CoA starter unit, like that observed in curacin A biosynthesis in marine cyanobacterium *Lyngbya majuscula* (renamed *Moorea producens*) (Gu et al., 2007). The starter unit could undergo an elongation step using the predicted acetoacetyl-ACP produced by the beta-ketoacyl ACP synthase and condensation protein. This molecule could potentially undergo a further elongation step by incorporating another acetyl group through the encoded hydroxymethylglutaryl-CoA synthase, after which the downstream enoyl-CoA hydratase/isomerase could further tailor the ketide chain. The growing polyketide product could potentially be loaded onto the 4'phosphopantethiene (PP) domain by the downstream encoded 4'-PP transferase and undergo further elongation with a glycine moiety via the 2-amino-3-ketobutyrate-CoA ligase. Other enzymes in this clusters such as 3-hydroxyacl-ACP synthase, dehydrogenase and hydroxylase enzymes could further modify the predicted ketide chain (Fig. 5.5). The presence of an *incA* gene would suggest the modification of the produced fatty acid may include the IncA-like serine/threonine phosphoprotein in its structure.

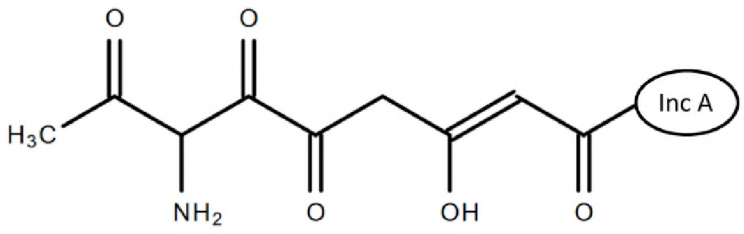


Figure 5.5: Basic predicted chemical structure of compound produced from enzymes encoded on Cluster A. Rules of collinearity were assumed and tailoring reactions were not taken into account.

5.3.2. Cluster B: Putative O-antigen biosynthetic gene cluster

Cluster B consists of 12 ORFs that span a region of approximately 15 Kb (Fig. 5.6) that is predicted to encode enzymes that produce a polysaccharide. Gene organization of this cluster most closely resembled that of gene clusters encoding O-antigens in *Pseudomonas aeruginosa* and *Vibrio cholera* (Belanger et al., 1999; Chatterjee and Chaudhuri, 2004). Cluster B includes two different glycosyl families, namely, three ORFs encode putative Family 4 glycosyltransferases and two Family 1 glycosyltransferases. Family 4 glycosyltransferases in bacteria are phospho-N-acetylmuramoyl-pentapeptide-transferases that are responsible for some of the first steps in biosynthesis of cell wall components whilst Family 1 are most commonly

associated with the production of capsular polysaccharides. Similarly, five glycosyl transferases were detected in the gene cluster of immunogenic glycoproteins in *Campylobacter jejuni*, along with genes involved in sugar biosynthesis as well as homologues of an amino transferase and a dehydratase.

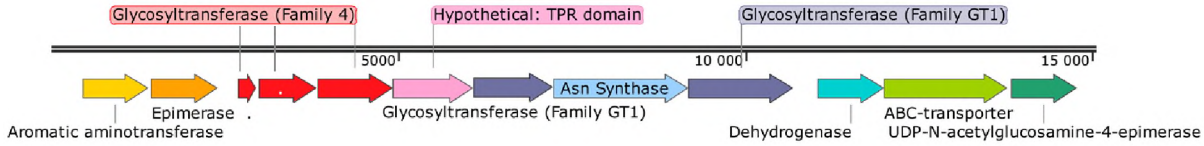


Figure 5.6: Gene organization of Cluster B. ORFs have been indicated with coloured arrows. The putative function of each ORF has been indicated. All ORFs are depicted to scale and relative size is indicated with a horizontal scale bar marked with size in number of base pairs.

5.3.3. Cluster C: Potential production of capsular polysaccharide

Cluster C consists of 17 ORFs and spans approximately 21 Kb (Fig. 5.7) and can be divided into three regions. The first region spans from the gene encoding a surface carbohydrate biosynthesis protein to the dehydrogenase gene. The enzymes encoded by the putative genes in this region are involved in the biosynthesis of a CDP-bound sugar that may incorporate an ammonia moiety via the N-acetyl-sugar amidotransferase. The second region spanning from the *wbuB* to the *wecC* genes could produce a structure that may incorporate a UDP-N-acetyl-D-mannosaminouronate molecule linked to a N-acetyl-L-fucosamine molecule. Heparinase activity involves the cleavage of the link between a hexosamine, such as mannosamine, and a uronic acid. It is though that the encoded heparinase cleaves this region in the UDP-N-acetyl-D-mannosaminouronate and allows the transfer of the CDP-bound sugar by the glycosyltransferase. The third region comprises genes encoding RimL, an aldo/ketoreductase and a MATE-like efflux protein. Both N-acetyltransferases and efflux proteins provide the host bacterium with resistance to aminoglycosides (Magnet et al., 2001) and this region could potentially confer immunity to the aminoglycoside produced by the downstream genes in regions 1 and 2.

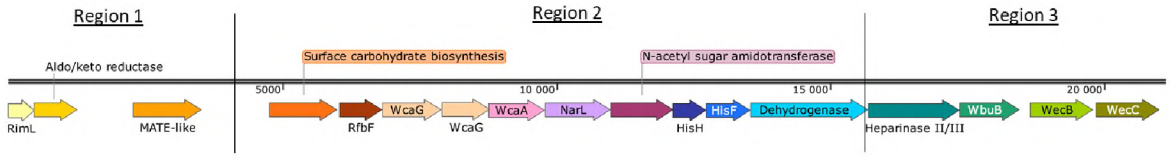


Figure 5.7: Gene organisation of Cluster C. The putative function of each ORF has been indicated. All ORFs are depicted to scale with coloured arrows and relative size is indicated with a horizontal scale bar marked with size in number of base pairs. Putative gene functions have been labelled using gene abbreviations; RimL: N-acetyltransferase, MATE-like: Multidrug and toxic drug-like extrusion protein, RfbF: Glucose-1-phosphate cytidylyltransferase, WcaG: CDP-glucose-4,6-dehydratase, WcaA: Glycosyltransferase Family II, NarL: transcriptional regulator, HisH: Imidazole glycerol phosphate synthase, HisF: Cyclase, WbuB: N-acetyl-L-fucosamine transferase, WecB: UDP-N-acetyleglucosamine-2-epimerase (non-hydrolyzing), WecC: UDP-N-acetyl-D-mannosamine dehydrogenase.

5.4. Discussion

Genome 3 was initially taxonomically classified using CheckM (Parks et al., 2015) as a species within the *Salinispira* genus. However, full genome comparison with *S. pacifica* strain L21-RPul-D2 found that Genome 3 represents a distinct spirochete with approximately 70% shared genes with predicted known function. Taxonomic classification and the functional role of a bacterial species in a particular community often do not correlate well, due to gene transfer events and differential gene expression (Lau et al., 2014) and thus these two genomes likely represent different functionally adapted bacterial species. Recovery of two full length, distinct 16S rRNA genes showed that Genome 3 is most likely a mixed genomic bin of two distantly related species within the Spirochaete phylum. Based on the phylogenetic analysis of the two 16S rRNA sequences, Genome 3A and Genome 3B share 90% sequence identity indicating that these 16S rRNA sequences represent distinct, and likely, novel bacterial families. Both Genomes 3A and B 16S rRNA genes were most closely phylogenetically related to spirochetes previously identified in *T. favus* TIC2009-002 (Walmsley et al., 2012). The closest phylogenetic relative of Genome 3A is not numerically dominant which would suggest that this bacterium is not a numerically dominant spirochete associated with *Tsitsikamma* and *Cyclacanthia* sponges. Instead, the spirochete represented by Genome 3B is most closely related to S_OTU1 which is a numerically dominant spirochete in *T. favus* and *T. nguni* sponges and thus the assembled genome likely represents this numerically dominant bacterium. Separation of these two genomes would require further sequencing and application of more

stringent parameters during the binning process. At this time, however, Genome 3 could not be further deconvoluted.

The largest proportion of genes present on Genome 3 are involved in the metabolism of carbohydrates and proteins. Various genes encoding ABC carbohydrate transporters were identified as well as numerous genes for the utilization of chitin, N-acetylglucosamine, xylose and D-ribose, indicating that carbohydrate metabolism in Genome 3 is versatile and adaptable. There are particularly high number of genes predicted to be involved in the production of D-fructose-6-phosphate from chitin. Many marine bacteria are capable of degrading chitin to saccharides that can then be used by various other organisms (Li et al., 2016b). Chitin has been identified as in the skeletons of several demosponges (Ehrlich et al., 2007; Mutsenko et al., 2017; Ehrlich et al., 2017; Pertrenko et al., 2017). Only a single gene encoding chitinase was located on the genome of *S. pacifica* and it is thus possible that the *Spirochaete* spp. represented by Genome 3 converts chitin in the host sponge to a soluble, more simple form that can be used by other bacterial symbionts. It was also noted that nearly all genes required for a complete Krebs cycle were present, save for a gene predicted to encode fumarase. The absence of fumarase would suggest an incomplete Krebs cycle. It has been proposed that succinate dehydrogenases could instead generate the proton motive force required in the absence of fumarase (Golyshina et al., 2016). The presence of a fumarase-encoding gene in the *S. pacifica* genome suggests that unlike the Genome 3 spirochete, *S. pacifica* would not have to rely on succinate dehydrogenase to complete the citric acid cycle.

Most genes encoding proteins involved in protein metabolism were predicted to be tRNA aminoacylation enzymes and evidence for potential biosynthesis of sixteen proteinogenic amino acids which could potentially also be used by the host sponge. Genes involved in the biosynthesis of tryptophan, a key molecule in the pyrroloiminoquinone core scaffold, were not identified. Genes encoding proteins involved in DNA metabolism were abundant in Genome 3. Many of these genes were predicted to encode enzymes involved in restriction modification systems, a feature that has been previously observed in sponge metagenomes and have been implicated in defence against phage infection and persistence of toxin-antitoxin systems which influence genetic uptake from conjugation events (Horn et al., 2016; Sitaraman, 2016; Slaby et al., 2017). Restriction modification (RM) systems are primarily used

defensively against foreign DNA, such as that associated with viruses, but are also thought to play an important role in mediating symbiosis between the bacterium and its sponge host through rearrangement in the symbiont genome effectively accelerating its adaptive evolution relative to its host (Zheng et al., 2016; Slaby et al., 2017). Interestingly, the defensive role of the restriction-modification system may be countered by the presence of sixteen mycobacterium-like genes that were predicted to be involved in bacterial virulence but provide support for the potential role of RM systems in symbiont adaptation. This was notable because six genes encoding different conjugative transfer proteins were identified within Genome 3 suggesting that the virulence operon may have been acquired via horizontal gene transfer. Comparison of genes with predicted known function revealed that these conjugative genes were not present in the genome of close phylogenetic relative *S. pacifica*. The enzymes encoded by these genes are required for the invasion and intracellular persistence of *Mycobacterium* sp. in macrophages (Gao et al., 2006). It is tempting to speculate that this cluster may aid the spirochete in colonizing the sponge host. Alternatively, it was predicted that the mycobacterium-like included genes that encode necessary machinery for the production of quinolate. Quinolate has been reported to inhibit the invasion and spread of pathogenic *Shigella* spp. (Prunier et al., 2007) and is proposed to perform a similar function for the inhibition of Group A *Streptococcus* spp. (Sorci et al., 2013). It is possible that the spirochete symbiont acquired this operon and can subsequently defend the host sponge against pathogenic bacteria through the production of quinolate. Support for the hypothesis of a symbiotic lifestyle was further provided by the detection of numerous genes encoding tetratricopeptide repeat domains which have been found to improve attachment of the bacterium to eukaryotic host cells (Alex and Antunes, 2015). Only one gene encoding a multidrug efflux pump and nineteen genes encoding TPR proteins were identified in the genome of *S. pacifica* suggesting that this bacterium may not encounter many toxic compounds in its natural environment within a hypersaline microbial mat. Moreover, the lower number of TPR proteins indicates that *S. pacifica* is potentially not as well suited as Genome 3 to exist as a sponge symbiont, which is unsurprising, considering the locality of its initial isolation in a hypersaline microbial mat.

With respect to secondary metabolite biosynthesis, three putative gene clusters were identified in Genome 3. Cluster A provided further evidence supporting the hypothesis

that the conserved spirochete is an obligate symbiont of *T. favus* host sponge. It is possible that Cluster A produces a fatty acid with an inclusion protein on the terminal end, encoded by an *incA* gene. The IncA protein is used by pathogenic *Chlamydia* bacteria in their inclusion membranes, which facilitates fusion with vesicles extruded from the trans-Golgi body in the host cell cytoplasm. This allows for the incorporation of host proteins in the inclusion membrane thereby conferring a level of "camouflage" for the bacterium against host defence systems (Wyrick, 2000). It is possible that the conserved spirochete, represented by Genome 3, has developed a similar system within the mesohyl of host *Tsitsikamma* sponges to evade predation.

Both Clusters B and C are predicted to produce antigenic molecules. Cluster B is predicted to encode enzymes that produce a lipopolysaccharide (LPS) O-antigen whilst Cluster C is thought to include genes that encode a capsular polysaccharide. Cluster B shared several genes involved with sugar biosynthesis with *Campylobacter jejuni*. The cluster in *C. jejuni* includes a UDP-N-acetylglucosamine and an ABC-transporter ATPase. All these features were observed within spirochete Genome 3, Cluster B. Homologues to the *C. jejuni* cluster have been found in *Neisseria meningitides* and a *Campylobacter* sp. that are both thought to produce O-linked glycans in the membrane of the pili and flagella of these bacteria (Schoenhofen et al., 2006; Weerapana and Imperiali, 2006). The point of glycosylation of these proteins is an asparagine residue in both mammalian and bacterial systems (Weerapana and Imperiali, 2006). There is an asparagine synthase in the centre of Cluster B that likely fulfils the same function.

It was recently shown that the LPS O-antigen produced by *Burkholderia* and *Bradyrhizobia* bacteria aided in mediating the symbiotic relationship between themselves and their *Riptortus pedestris* bean bug and *Aeschynomene indica* legume host species respectively (Kim et al., 2016; Li et al., 2017). Similar mechanisms of symbiosis have been observed in bacteria associated with sponges. It was found that LPS structures on *Endoziocomonas* sp. and *Pseudoalteromonas* sp. inhabiting a *Suberites domuncula* sponge did not elicit a response from defensive macrophages in the sponge (Garderes et al., 2015). Similarly, a more recent study showed that the fatty acid composition of a *Pseudoalteromonas* sp. inhabiting the *Suberites domuncula* sponge differed to a closely related opportunistic, pathogenic strain of *Pseudoalteromonas* in three ways: The presence of bis-phosphorylated lipid A species

in place of a mono-phosphorylated form, an acyloxyamide in place of an acyloxyacyl ester in the secondary fatty acid and unsaturated chains as secondary O-acyl substitutions. However, it was concluded that although there were differences in the fatty acid make-up in the commensal and pathogenic forms of *Pseudoalteromonas* sp. further investigation was warranted to confirm the role of LPS structure in interactions with the host sponge (Lorenzo et al., 2017).

It has been proposed by Webster and Thomas (2016) that marine sponges are able to differentiate between bacterial "friend" and "foe" using LPS binding-like proteins and other receptors. Since the *T. favus*-associated spirochetes have been identified as a numerically dominant species within *Tsitsikamma* and *Cyclacanthia* sponges over a period of nine years and in different geographic locations (Chapter 3, Matcher et al., 2017), it is unlikely that these spirochetes are opportunistic pathogens of the host sponge. Rather, they may be employing LPS molecules encoded by Cluster B as a means of differentiating themselves from other bacteria in the consortia as symbionts rather than food for the host sponge. This hypothesis requires biochemical validation.

Capsular polysaccharides (K-antigens) and lipopolysaccharides (O-antigens) are two forms of bacterial surface antigens. O-antigens are anchored in the outer membrane and encapsulated by the K-antigen layer (Sarkar et al., 2014). Cluster C in Genome 3, likely encodes a capsular polysaccharide that incorporates one or more N-acetyl-D-mannosamine moieties in its structure. It was also shown that Genome 3 differs from *S. pacifica* in that it can utilize N-acetylglucosamine. Capsular polysaccharides containing N-acetylglucosamine have been shown to aid *Staphylococcus* spp. in adhering to surfaces and evading phagocytosis (Laverty et al., 2013). A capsular polysaccharide consisting of homopolymers of N-acetyl-mannosamine, like that predicted to be incorporated by the enzymes encoded on Cluster C, is present on the cell surface of bacterium *Neisseria meningitidis* Serogroup A. *N. meningitidis* is responsible for large outbreaks of meningitis in sub-Saharan Africa and it is thought that the capsular polysaccharides aid in bacterial evasion of the immune system (Tzeng et al., 2016). It is possible that like the capsular polysaccharides on the cell surface of *Staphylococcus* spp. and *N. meningitidis* that aid in the bacteria evading eukaryotic host responses, the capsular polysaccharide produce via Cluster C may help the spirochete evade phagocytosis in the host *T. favus* sponge.

The capsular polysaccharide produced by Cluster C is also thought to include a N-acetyl-fucosamine molecule as well as the N-acetyl-D-mannosamine in the capsular polysaccharide structure. Studies into the structure of *Staphylococcus aureus* capsular polysaccharides have found that capsular polysaccharide Type 5 consists of trisaccahride repeats of N-acetyl mannosaminuronic acid, N-acetyl L-fucosamine, and N-acetyl D-fucosamine (Jones, 2005) and that this Type 5 polysaccharide allows the bacterium to hide from host defence systems and avoid phagocytosis (Watts et al., 2005). It is possible that the conserved spirochetes utilize the capsular polysaccharide produced by Cluster C to similarly avoid phagocytosis in the *Tsitsikamma* host sponge and further solidify its role as symbiont.

Bioinformatic analysis has revealed that, unlike its phylogenetic relative *S. pacifica*, Genome 3 includes several features that potentially enable the two represented spirochetes to exist symbiotically with the host *Tsitsikamma* sponge. However, until further sequence data is accumulated, it cannot be confirmed whether both spirochetes binned in this genome are true *T. favus* symbionts. As the 16S rRNA sequence that represent Genome 3B clustered most closely with conserved, dominant spirochetes previously identified in *T. favus* sponges, it is tempting to speculate that Genome 3B represents a specialized symbiont of *T. favus* and the genes identified in this study have allowed it to establish a strong symbiosis with the host which accounts for its numeric abundance.

No evidence was found to suggest the production of pyrroloiminoquinone compounds by these spirochetes. It is hypothesised that the compounds are produced via a lantipeptide gene cluster similar to that identified for structurally related lymphostin and ammosamides (Jordan and Moore, 2016). No lantipeptide pathway nor fragment thereof was identified within Genome 3, nor was there any evidence of an alternative pathway identified. Furthermore, Genome 3 does not appear to include the biosynthetic machinery required to produce tryptophan, a predicted key molecule in pyrroloiminoquinone biosynthesis. Therefore, these conserved spirochetes do not produce tsitsikammamines but rather exist as persistent symbionts within the sponge and likely fulfil a different role such as defence or nutrient acquisition. This conclusion is consistent with the presence of a closely related spirochete strain in *C. bellae* sponges which do not produce tsitsikammamines (Dorrington et al., 2016). The presence of several resistance genes encoding multidrug efflux pumps and

fluoroquinolones would indicate however, that it adapted to survive in a toxic environment where various pyrroloiminoquinone compounds are produced.

Chapter 6 - Secondary metabolite gene clusters detected in the *T. favus* metagenome

6.1. Introduction

The structural diversity of microbial secondary metabolites can largely be attributed to the activity of two classes of large, multimodular enzymes: polyketide synthases (PKSs) and non-ribosomal peptide synthases (NRPSs) (Ziemert et al., 2014). Many of these secondary metabolites mediate both microbe-microbe and microbe-host interactions as well as play important ecological roles such as nutrient acquisition and defensive chemistry (Fischbach et al., 2008; Wietz et al., 2013). The genes encoding NRPS and PKS enzymes are commonly found in clusters encoding biosynthetic, regulatory and transport genes, arranged in the order of the successive steps required to generate the respective product (reviewed in Osborne, 2010 and Medema and Fischbach, 2015). Rapid evolution of PKS/NRPS gene clusters through mutations, domain rearrangements, module duplication and horizontal gene transfer has led to a large structural and physiological diversity of secondary metabolites which can be exploited for pharmaceutical benefit (Fischbach et al., 2008; Huang and Lin, 2017).

Genes encoding NRPS enzymes are characterized by the presence of four essential domains *viz.* adenylation, condensation, peptide carrier and thioesterase proteins which are organized into modules, with each module generally responsible for the incorporation of one monomer (Caboche et al., 2010; Strieker et al., 2010). An NRPS gene cluster can consist of several modules containing essential domains as well as domains encoding tailoring enzymes such as epimerases, hydroxylases and methyltransferases (Caboche et al., 2010). Adenylation domains (A-domains) recognize and activate a substrate monomer. The protein carrier protein (PCP) is activated by a 4'phosphopantetheine cofactor, binds the activated monomer and is responsible for the propagation of the nascent peptide chain. The condensation domain (C-domain) catalyses a Claisen condensation reaction to create peptide bonds between substrates bound to sequential PCP's (Strieker et al., 2010). The general mechanism utilized by NRPS machinery is outlined in Fig. 6.1 A. The thioesterase domain (TE domain) is responsible for the cleavage and release of the final peptide

from the NRPS and variants of TE domains can allow for cyclization of the peptide chain into a cyclopeptide (Caboche et al., 2010; Strieker et al., 2010). Monomers recognised and bound by NRPS domains are not limited to the twenty proteinogenic amino acids but can include over 500 different monomers including fatty acids, dihydroxybenzoic acid, D-amino acids, modified amino acids and non-proteinogenic amino acids (Caboche et al., 2010; Strieker et al., 2010) which allows for diversification of protein structure.

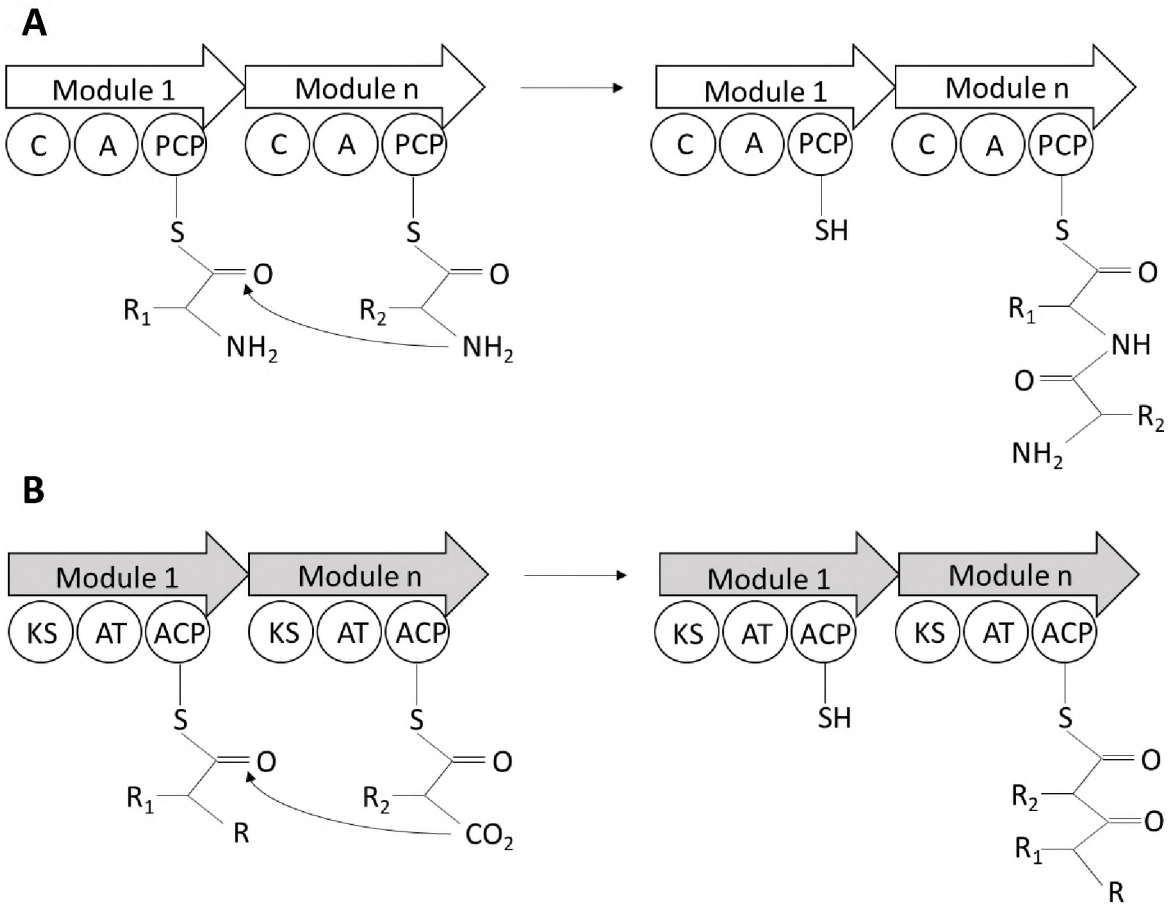


Figure 6.1: General mechanism of natural product formation in modular (**A**) NRPS gene clusters and (**B**) PKS gene clusters. C: Condensation domain, A: Adenylation domain, PCP: peptide carrier protein, KS: ketosynthase domain, AT: acetyltransferase domain, ACP: acyl-carrier protein.

PKSs function in a similar manner to both NRPS gene clusters and fatty acid biosynthesis (Fig. 6.1 B). They consist of several modular domains that incorporate acyl-CoA monomers in sequential via Claisen condensation reactions (Hwang et al., 2014). The core domains in each module include acyltransferase domain (AT domain), acyl carrier protein (ACP) and a ketosynthase (KS) domain (Hwang et al., 2014). The ACP acts in a similar manner to the PCP by binding the monomers and the nascent

peptide chain. The AT domain activates the ACP domain by acetylating the next monomer and binding it to the phosphopanteinyl group of the ACP. The KS domain acts in much the same way as an NRPS C-domain by catalysing a condensation reaction between the nascent chain and the next sequential ACP-bound monomer (Robbins et al., 2016). The elongation of the polyketide is terminated, and the peptide chain is released by a thioesterase domain (Hwang et al., 2014).

There are three types of PKS gene clusters that are differentiated by their structure and function. Type I PKS gene clusters consist of several modules and generally produce macrolide structures. Type I PKSs are further categorised into *cis*- and *trans*-AT classes, where trans-AT PKS complexes encode a discrete acetyltransferase that is separate from the rest of the PKS cluster. Type II PKS systems are iterative and produce aromatic polyketides, which incorporate nearly exclusively acetyl- and malonyl-CoA monomers that in turn are modified by tailoring enzymes encoded in the gene cluster. Type III PKS gene clusters are largely found in plants, but microbial forms do exist. Finally, Type III PKS complexes exist as ketoreductase homodimers that are devoid of ACPs and instead iteratively condensates acyl-CoA monomers to extend the nascent carbon chain (Hwang et al., 2014; Bertin et al., 2016). There are some exceptions in the classification of PKS gene clusters, including iterative Type I PKS gene clusters that produce aromatic polyketides in bacteria (Gassier et al., 1997; Ahlert et al., 2002), Type I PKS systems that require a separate AT domain (Cheng et al., 2003) and non-iterative Type III PKS clusters that incorporate C-O bonds as well as the traditional C-C bonds in their biosynthetic products (Shen, 2003).

Due to the structural and catalytic similarities between PKS and NRPS genes, several clusters have been identified as hybrid systems incorporating both PKS and NRPS domains within single modules (Du et al., 2001; Ansari et al., 2004). Examples of marine natural products produced by hybrid NRPS-PKS gene clusters include a chlorinated lipopeptide called barbamide from marine cyanobacterium *Lyngbya majuscula* (Chang et al., 2002b), haliamide from marine myxobacterium *Haliangium ochraceum* (Sun et al., 2016) and have been implicated in the tailoring of pyrroloquinoline compound lymphostin in marine *Salinispora* species (Miyanaga et al., 2011).

Advances in sequencing technologies have prompted the rapid development of computational tools that enable users to analyse large datasets for the discovery of novel secondary metabolite gene clusters and effectively bypassing the "grind and find" approach of drug discovery through manipulation of culture conditions (Weber and Kim, 2016). Various tools are available for analysis of metagenomic datasets including algorithms for the identification and annotation of ORFs such as Prokka (Seemann, 2014) or myRAST (Aziz et al., 2008). Comparison of whole genome data can be achieved using MG-RAST online servers (Meyer et al., 2008) and identification of putative secondary metabolite biosynthetic gene clusters, and subsequent prediction of potential natural bioactive compound structure can be achieved with antiSMASH (Weber et al., 2015). Predicted structures can be compared against the Norine database (Pupin et al., 2016) or against the ChemSpider database for identification of structurally similar compounds (Pence and Williams, 2010).

Research described in Chapter 4 included the identification of one NRPS cluster, one hybrid NRPS/PKS cluster, one Type I-PKS cluster and three Type III PKS clusters in the metagenome sequence dataset obtained from *T. favus* TIC2016-050A. The aim of the research described in this chapter was to further characterise these NRPS and PKS gene clusters and to predict potential compounds produced by these clusters.

6.2. Methods and Materials

Biosynthetic gene clusters identified as NRPS, PKS or hybrids thereof in antiSMASH (as described in Section 5.2) were downloaded and individual ORFs were studied using BLAST analysis for functional inference. ORFs were aligned against the Non-redundant protein sequence database using the BLASTp algorithm. Sequence identity above 75% was considered to represent a homologous gene product. Compound structures were predicted following the co-linearity rule. The collinearity rule dictates that the structure of the NRPS or PKS product is determined by the ordered, sequential addition of amino-acid encoded on the gene cluster (Donadio et al., 1991; Walsh, 2004). Subsequent hypothetical structures were analysed using both Norine (Pupin et al., 2016) and ChemSpider (Pence and Williams, 2010) for identification of known compounds of similar structure.

6.3. Results

Six biosynthetic pathways classified within the class of NRPS, PKS or a hybrid of the two gene cluster types were identified in the metagenome of *T. favus* specimen TIC2016-050A (Table 6.1). Genome 1, a *Pseudovibrio* sp., was found to encode an NRPS and a Type III-PKS cluster that shared similarity with genes in biosynthetic clusters identified in a *Pseudomonas* sp. and a *Pseudovibrio* sp. respectively (Table 6.1). A further two Type III-PKS gene clusters were identified in Genomes 15 and 21, both of which were taxonomically classified within the *Rhodopirelulla* genus. Genome 15, furthermore, carried a hybrid NRPS/PKS gene cluster that shared 28% of genes identified in the microsclerodermin biosynthesis gene cluster found in the genome of *Jahnella* sp. MSr9139. A Type I-PKS cluster was identified on an unclustered contig that was shared greatest sequence identity with a cluster identified on the genome of the cyanobacterium *Lyngbya majuscala*.

Table 6.1: Summary of broad characteristics of secondary metabolite biosynthetic gene clusters from the metagenome of *T. favus* specimen TIC2016-050A.

Cluster	Type	Assigned genome	Taxonomy	Gene Cluster match	Accession no.	Shared genes
1	NRPS	1	*Pseudovibrio*	*Pseudomonas fuscovaginae*	JSYZ01000011_c1	19 %
2	Type III-PKS	1	*Pseudovibrio*	*Pseudovibrio* sp.	DS996811_c1	50 %
3	Type III-PKS	21	*Rhodopirellula*	*Kocuria rhizophila*	AP009152_c1	7 %
4	Type III-PKS	15	*Rhodopirellula*	*Kocuria marina*	JROM01000055_c1	11 %
5	Hybrid NRPS-PKS	15	*Rhodopirellula*	*Jahnella* sp.	KF657739_c1	28 %
6	Type I-PKS	Unclustered	N/A	*Lyngbya majuscula*	AY522504_c1	23 %

*Assigned genomes refer to the genomic bin number assigned to individual clusters as per Figure 4.4 and Table 4.2.

6.3.1. Characterization of Cluster 1

Cluster 1 was identified in antiSMASH as an NRPS gene cluster, but inspection of individual ORFs within the cluster revealed the presence of two ketosynthase domains indicating that the cluster is a NRPS/PKS hybrid. The cluster is predicted to be complete and consists of 11 modules that span a region of approximately 34 Kb (Fig. 6.2).

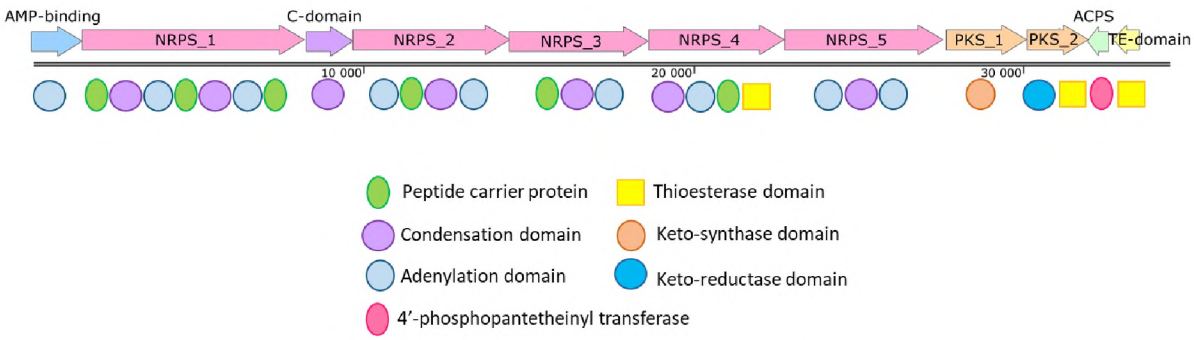


Figure 6.2: Gene organisation of the NRPS gene cluster identified in Genome 1. Modules are indicated with arrows orientated in the direction of transcription. Domains identified within each module are indicated using coloured circles and squares; the function of which can be identified using the key provided. A bar denoting the length of the gene cluster has been provided for relative size. Modules have been drawn to scale, domains have not.

Computational prediction based on conserved regions within the substrate activation sites of the NRPS adenylation domains was used to identify potentially activated monomers. If the rule of collinearity is followed, then the sequential order of the nine monomers activated within Cluster 1 is predicted to be Tyr|Thr|Gly|Arg|Pro|Cys|X|X|Pro|, (where "X" denotes an unknown substrate), to generate a compound with a hypothetical structure shown in Figure 6.3. The presence of three thioesterase domains within the gene cluster, as well as ketoreductase tailoring domains (Fig. 6.2) suggests that the structure of the putative compound is more complex than the proposed hypothetical linear structure (Fig. 6.3). Additional thioesterases have been implicated in proofreading and other editing roles during peptide assembly (Chang et al, 2002).

A comparison of the peptide predicted to be produced by Cluster 1 with structures on the Norine database identified two structurally similar, putative NRPS products, cyclonellin and phakellistatin 7. Both compounds are large, cyclopeptides consisting of 8 to 10 monomers that were isolated from *Axinella* and *Phakellia* sponge species respectively. Protein homology of individual genes using the BLASTp algorithm showed that various *Pseudovibrio* strains cultured from an Irish *Axinella dissimilis* sponge had the closest sequence identity matches to genes identified in Cluster 1 (O'Halloran et al., 2011). A comparison between the annotated genes with predicted known function included in the genomes of Genome 1 and *Pseudovibrio* sp. Ad5 isolated from an *A. dissimilis* sponge revealed that 96% of the genes were common to both genomes and sequence comparison showed that 78% of all proteins shared over

90% homology. Calculation of the average nucleotide identity between these two bacteria showed that that the overall sequence identity was 93.5%. All these data indicate that these two bacterial isolates are likely very similar.

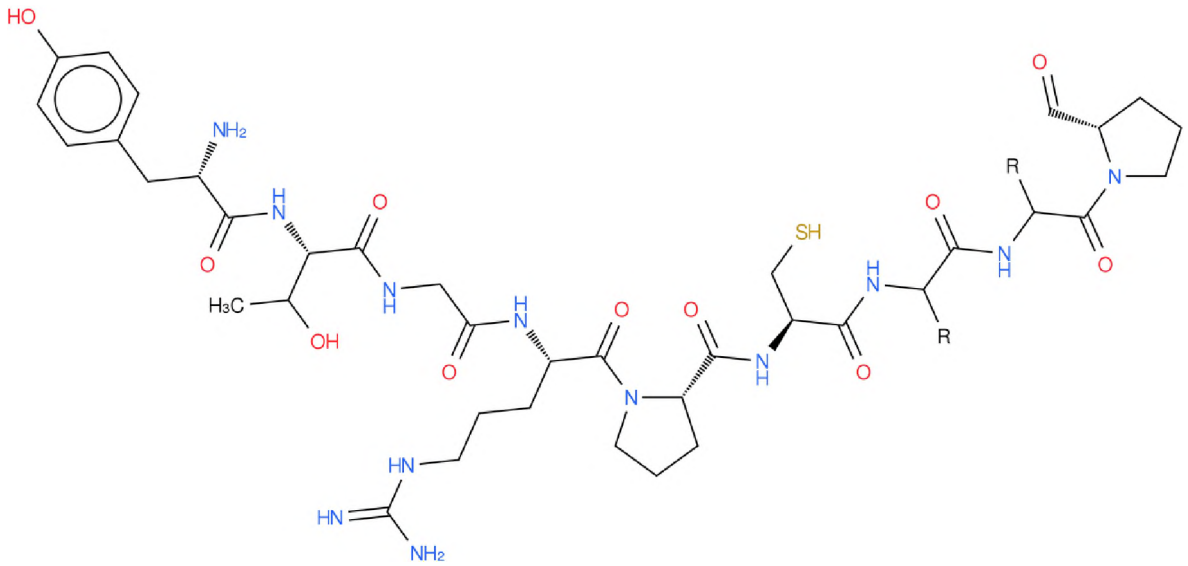


Figure 6.3: Predicted linear structure of the hybrid non-ribosomal peptide/polyketide compound encoded by Cluster 1 when following the rule of collinearity, complemented with predicted substrates activated by all nine adenylation domains present within the cluster. Chemical structure was drawn using ChemSpider.

6.3.2. Characterization of Cluster 2

Biosynthetic Cluster 2 was identified as a Type III-PKS gene cluster in *Pseudovibrio* Genome 1. The cluster consists of 27 ORFs and spans approximately 28 kb (Fig. 6.4). Cluster 2 shared 50% of genes in a cluster found on a contig of *Pseudovibrio* sp. JE062 (WP_008551233.1) and 48% genes present on a plasmid in *Pseudovibrio* sp. FO-BEG1 (WP_041769282.1). These bacteria were isolated from the sponge *Mycale laxissama* and a black band diseased coral respectively (Bondarev et al., 2013). Regions matching Cluster 2 were identified in *Pseudovibrio* sp. JE062 and *Pseudovibrio* sp. FO-BEG1, directly upstream of a larger Type III PKS. Cluster 2 consists of three regions. Within the first region, ORFs encoding several genes predicted to be involved in the production of proteins that confer resistance to toxic compounds including sodium driven multidrug resistance protein NorM and multiple antibiotic transcriptional regulator MarR were identified (Fig 6.4 A). The second region begins with an ORF that includes a domain for a nucleobase cation-1-symporter, which is followed by ORFs encoding an N-acetyltransferase (RimL family), an inner

membrane protein of unknown function (YgjV family), a chalcone-like Type III PKS, a isoprenylcysteine carboxyl methyltransferase homologue and an aldehyde dehydrogenase. The third region identified gene cluster appears to consist largely of genes encoding various enzymes involved in carbohydrate metabolism. The combination of these genes suggests that this cluster may encode various defences against toxic compounds. This would be advantageous for a bacterium that is associated with *T. favus* from which several highly cytotoxic pyrroloiminoquinone compounds have been isolated, including tsitsikammamine B which inhibits the growth of both Gram-positive and Gram-negative bacteria (Kalinski, Jiwaji, Waterworth et al., unpublished data).

6.3.3. Characterization of Cluster 3 and Cluster 4

Clusters 3 and 4 were identified in Genomes 21 and 15 respectively, which were both taxonomically classified within the *Rhodopirellula* genus. Both clusters comprise 15 ORFs with Cluster 3 spanning approximately 21 Kb (Fig. 6.4B) and Cluster 4 approximately 30 Kb (Fig. 6.4C). The gene organisation of these two clusters appeared to be largely similar with the presence of at least one ORF encoding a ribosomal recycling factor, a lipolytic protein, a leucine tRNA ligase, an adenylate cyclase, a dehydrogenase, selenoproteins SelA, SelB and SelD, ferritin, a Type III polyketide synthase and a sulfatase.

Cluster 3 was identified on a smaller contig than Cluster 4 but consisted of two ORFs encoding leucine-tRNA ligases, adenylate cyclases and sulfatases where Cluster 4 included only a single copy of each of these ORFs. Additionally, an ORF encoding a SelD-fusion protein was identified in Cluster 3 that was absent in Cluster 4. Cluster 4 included ORFs encoding a translation elongation factor, a UMP kinase, a hydroxymethylbilane synthase as well as a serine kinase but as these ORFs were identified in regions that were beyond the full length of the contig on which Cluster 3 was found and thus it cannot be definitively confirmed whether these ORFs were absent in Cluster 3.

Clusters 3 and 4 share several genes present in similar clusters on the genomes of *Kocuria marina* S09-6, *Kocuria rhizophila* DC2201 and *Kocuria rhizophila* TPW45. Highest sequence homology was identified between the ORFs encoding

selenoproteins SelB, SelD, SelA and the Type III-PKS. The presence of a selenocysteine-specific translation elongation factor (SelB), an L-seryl-tRNA selenium transferase (SelA) and a selenide water dikinase (SelD) directly downstream of PKS genes in both Cluster 3 and Cluster 4 suggests that both *Rhodopirellula* Genomes 15 and 21 produce selenoproteins. Selenoproteins are relatively rare compounds characterised by the presence of at least one selenium-containing amino acid. Selenoproteins have been reported to have diverse biological capabilities, with many functions still unknown (Rezanka and Sigler, 2007; Peng et al., 2016). Selenoproteins have also been reported to accumulate intracellular selenium which conferred increased protection against antibiotics through the scavenging of free radicals (Li, et al., 2017).

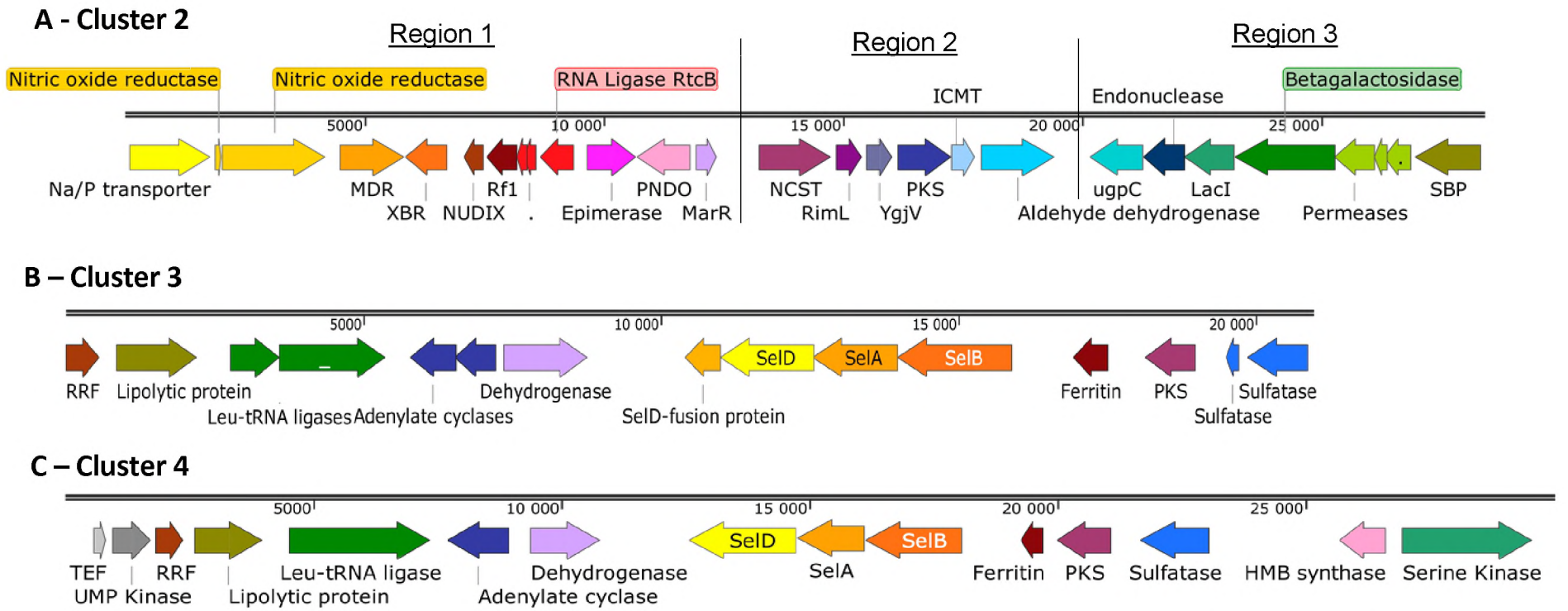


Figure 6.4: Gene organization of Type III-PKS gene clusters identified in clustered genomes extracted from metagenomic data from *T. favus* TIC2016-050A. Cluster 2 was identified in *Pseudovibrio* Genome 1 (**A**), Cluster 3 was identified in *Rhodopirellula* Genome 21 (**B**) and Cluster 4 was identified in *Rhodopirelulla* Genome 15 (**C**). Putative gene function of each ORF has been indicated. MDR: multidrug resistance protein (NorM), XBR: xenobiotic response transcriptional regulator, NUDIX: nucleotide diphosphate-linked moiety hydrolase, Rf1: peptide release factor H, RtcB: RNA ligases, PNDO: FAD-dependent pyridine nucleotide disulphide oxidoreductase, MarR: multiple antibiotic resistance transcriptional regulator, NCST: nucleobase cation-1-symporter, RimL: N-acetyltransferase, YgjV: inner membrane protein of unknown function, PKS: chalcone-like Type III polyketide synthase, ICMT: isoprenylcysteine carboxyl methyltransferase, ugpC: glycerol-phosphate transporter ATP binding protein, SBP: ABC transporter substrate binding protein. TEF: Translation elongation factor, RRF: Ribosome recycling factor, SelD: selenide water dikinase, SelA: L-seryl tRNA selenium transferase, Sel B: Selenocysteine specific translation elongation factor, PKS: Type III polyketide synthase, HMB synthase: Hydroxymethylbilane synthase. All ORFs are depicted to scale and relative size indicated with a horizontal bar demarked length denoted in number of base pairs. All ORFs are depicted to scale and relative size indicated with a horizontal bar demarked length denoted in number of base pairs.

6.3.4. Characterization of Cluster 5

Cluster 5, the second biosynthetic gene cluster identified on Genome 15, was characterised as a hybrid NRPS/Type I-PKS gene cluster. Cluster 5 consists of eight ORFs and spans approximately 21 Kb. In Cluster 5 the ORF encoding a Type-I PKS was identified and included an NRPS-like thioesterase domain, as well as an NRPS-like adenylation domain (Fig. 6.5). The PKS-acetyltransferase is predicted to activate a malonyl-CoA and the NRPS-adenylation domain is predicted to activate either a glycine or a modified glycine moiety. A BLAST analysis of the PKS ORF revealed that it shared 44% amino acid sequence identity with a gene found in *Pedosphaeara parvula* which is a Verrucomicrobium isolated from soil (Kant et al., 2011).

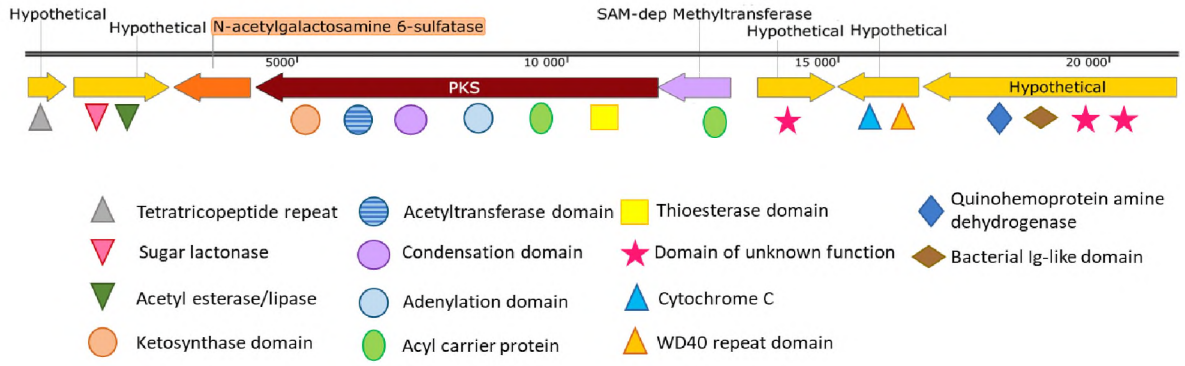


Figure 6.5: Gene organisation of the hybrid NRPS/PKS Cluster 5 identified in Genome 15. ORFs are indicated with arrows orientated in the direction of transcription. Domains identified within each module are indicated using coloured shapes below each ORF; the function of which can be identified using the key provided. A bar denoting the length of the gene cluster has been provided for relative size. ORFs have been drawn to scale, domains have not.

Studies analysing phenotypic microarrays have suggested that carbon sources are stimulants for secondary metabolite production via polyketide synthase enzymes (Fuerst, 2017). A N-glucosamine sulfatase was identified upstream of the PKS ORF in Cluster 5, the presence of which is a defining characteristic of many marine planctomycetes (Fuerst, 2017). N-glucosamine sulfatases may function to facilitate the metabolism of sulphated carbohydrates by commensal marine planctomycete bacteria. If the collinearity rule is applied to this gene cluster and it assumed that the NRPS TE domain catalyses a cyclization reaction, the predicted structure of the produced compound would match that of 4-Methyl-2-morpholinone (Fig. 6.6) which is currently used for curing epoxy resins (US Patent 5093458 A).

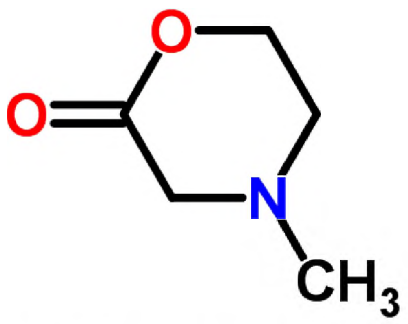


Figure 6.6: Chemical structure of 4-methyl-2-morpholinone predicted to be produced from Cluster 5 found on Genome 15. Chemical structure was drawn using ChemSpider.

Genome sequencing of marine planctomycete *Rhodopirellula baltica* revealed a large number of proteins (approximately 63%) of unknown function (Glockner et al., 2003; Studholm et al., 2004). Planctomycetes are unique among bacteria in that they have cell compartmentalisation and peptidoglycan-less proteinaceous cell walls and it was hypothesised that many of these proteins of unknown function were responsible for membrane transport (Studholm et al., 2004). Five of the ORFs identified in Cluster 5 were predicted to encode hypothetical proteins with no assigned function. It is possible that these encoded proteins may aid in the transport of the produced 4-methyl-2-morpholinone-like structure produced by Cluster 5.

6.3.5. Characterization of Cluster 6

Cluster 6 was identified as a Type I-PKS cluster on an unclustered contig and consists of six ORFs that span 12.4 Kb. The ORFs encoded an aromatic ring hydroxylating dioxygenase, a fatty acyl-AMP ligase with an additional acyl-carrier domain within it. A surfactin-like thioesterase encoding gene was upstream of two PKS modules, with the first consisting of a ketosynthase and a dehydratase domain and the second consisting of a quinone enoyl reductase domain positioned between two sub-domains of a ketosynthase-fatty acid complex. An ORF predicted to encode a S-adenosyl-methionine(SAM)-dependant methyltransferase was identified at the end of the cluster (Fig. 6.7).

Fatty acyl-AMP ligases (FAALs) differ from traditional fatty-acyl CoA ligases in that they activate fatty acids and transfer them directly to the acyl-carrier domain without the need of a CoA co-factor and facilitates the production of unusual acyl chains (Mohanty et al., 2011). The substrate specificity of this domain, predicted via profile hidden Markov models (pHMM) in antiSMASH, was predicted to be phenylalanine (Prieto et al., 2011; Weber et al., 2015). The presence of an ORF encoding a probable phenylpropionate dioxygenase (PP-dioxygenase) directly upstream of the FAAL indicates that the ring structure of phenylalanine may become hydroxylated.

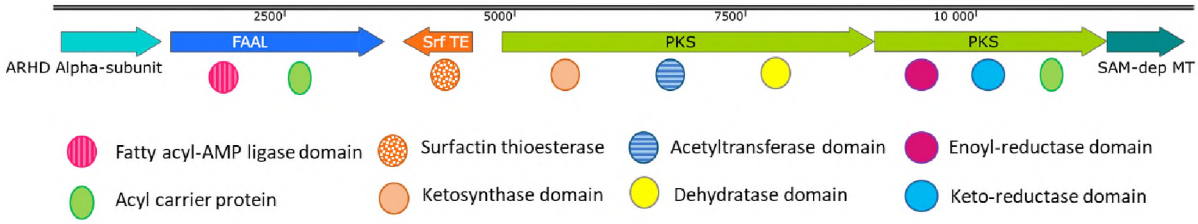


Figure 6.7: Gene cluster organisation of Type I-PKS Cluster 6 identified on an unclustered contig. ORFs are indicated with arrows orientated in the direction of transcription. ARHD: Aromatic ring hydroxylating dioxygenase, FAAL: Fatty acyl-AMP ligase. ACP: acyl-carrier protein, Srf TE: Surfactin-like thioesterase, PKS: polyketide synthase, SAM-dep MT: SAM-dependent methyltransferase. Domains identified within each module are indicated using coloured shapes below each ORF; the function of which can be identified using the key provided. A bar denoting the length of the gene cluster has been provided for relative size. ORFs have been drawn to scale, domains have not.

The acetyl-transferase domain found on the first PKS module is predicted to activate and bind a malonyl-CoA group to the acyl-carrier protein. The ketosynthase domain will catalyse a Claisen condensation reaction and bind the malonyl and phenylalanine in a single chain. The dehydratase domain found on the first PKS module may result in the hydroxyl groups present on the ring structure being converted to ketone groups. Finally, the presence of a SAM-dependent O-methyltransferase indicates that some of the oxygen moieties present within the structure may become methylated. The ORF encoding the surfactin-like thioesterase domain is positioned within the gene cluster rather than at the 3' terminal end. Internal thioesterase domains have been identified in other clusters and appear to function as both a thioesterase as well as an acetyltransferase (Steller et al., 2004). The surfactin-like thioesterase domain identified within Cluster 6 may thus function in a similar manner.

BLAST analysis of each ORF indicated that the bacterium from which Cluster 6 originated was most likely within the Planctomycetaceae family as four of the six ORFs match genes identified within species of this family. However, the ORF encoding a surfactin-like thioesterase shared greatest amino acid identity with an ORF from a *Streptomyces* sp. isolated from the root of a sugar cane plant (WP_079312436.1). Inspection of this ORF within the host *Streptomyces* sp. GKU 895 revealed that it was positioned amongst ORFs encoding several hypothetical proteins and two transposases (NZ_MWJO01000137.1). Most Type I-PKS systems in bacteria are generally modular and produce nascent polyketides in a sequential, assembly line manner (Xu et al., 2013). If it is accepted that PKS ORFs present in Cluster 6 are modular rather than iterative, an approximate prediction of the compound structure produced can be made (Fig. 6.8).

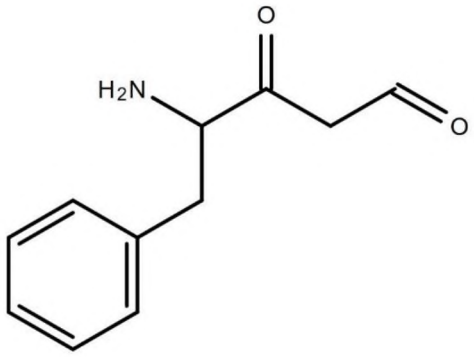


Figure 6.8: Predicted chemical structure of polyketide produced by Cluster 6 in Genome 1 without tailoring reactions taken into consideration. Chemical structure was drawn using ChemSpider.

The FAAL domain in Cluster 6 was predicted to recognize and activate phenylalanine. However, the substrate specificity of FAAL domains is influenced by the characteristic insertion motif within the protein and is reportedly "promiscuous" which effectively facilitates the production of compounds with varying chemical structures (Arora et al., 2009; Kuhn et al., 2016). It is thus proposed that the phenylalanine substrate predicted to be activated by the FAAL domain could be replaced by a different non-polar, aromatic amino acid such as tryptophan. This proposed substrate substitution, combined with the tailoring effects of the SAM-dependant methyltransferase and the aromatic ring hydroxylating dioxygenase, could produce a compound (Fig. 6.9 A) that is structurally similar to makaluvamines and would require the addition of a single tyrosine residue to create tsitsikammamine B (Fig. 6.9 B).

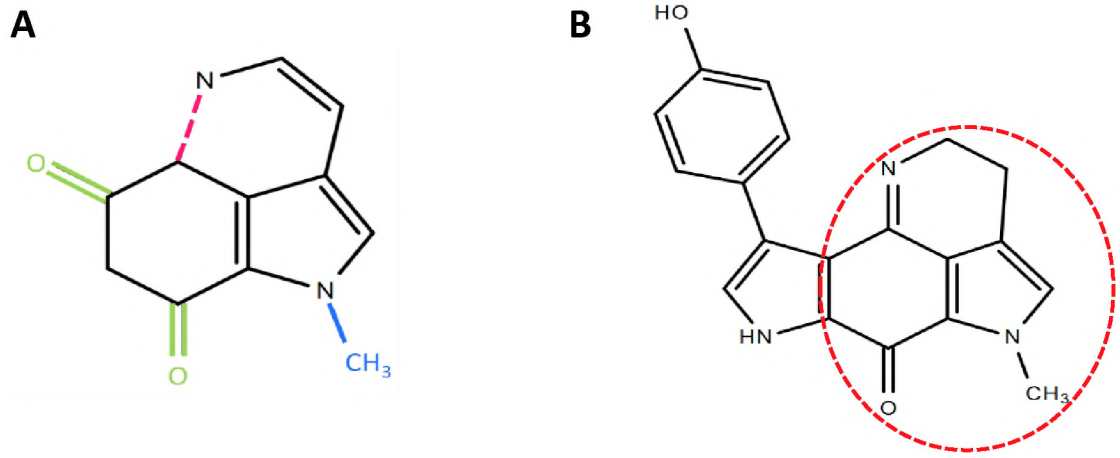


Figure 6.9: Structural comparison of potential compound produced by Cluster 6 (**A**) and tsitsikammamine B (**B**); a pyrroloiminoquinone compound isolated from *T. favus* sponges. Predicted modifications by tailoring enzymes are indicated by colour; addition of oxygen molecules by aromatic ring hydroxylating dioxygenase (green), methylation by SAM-dependent methyltransferase (blue) and potential cyclization for pyrrole formation (pink). The corresponding region of tsitsikammamine B that the predicted compound matches is indicated within a dotted orange circle. Chemical structures were drawn using ChemSpider.

6.4. Discussion

The aim of the research described in this chapter was to characterise the six biosynthetic gene clusters predicted to produce secondary metabolites within the bacterial metagenomic data of *T. favus* TIC2016-050. Furthermore, computational approaches were used to predict the core chemical structures of the compounds produced by some of the investigated pathways.

Genome 1, a *Pseudovibrio* sp., has two potential secondary metabolite biosynthetic gene clusters, comprising an NRPS and a Type-III PKS system. Sponge-associated *Pseudovibrio* species have found to be one of the most abundant culturable marine bacteria and as such have received increased attention as producers of structurally and functionally diverse PKS and NRPS bioactive compounds (Naughton et al., 2017). Clusters 1 and 2 shared the greatest amino acid identity with biosynthetic gene clusters identified in *Pseudomonas* and *Pseudovibrio* isolates, respectively and homologues of both Cluster 1 and Cluster 2 have been identified in sponge-associated *Pseudovibrio* species.

Several studies have previously demonstrated sponge-associated *Pseudovibrio* species host a large number of PKS and NRPS related genes and exhibit high levels of antimicrobial activity (Kennedy et al., 2008; Flemer et al., 2011; O'Halloran et al., 2011; Graca et al., 2013). An orphan NRPS gene cluster was identified in the genomes of *Pseudovibrio* Ad5 and *Pseudovibrio* Ad46 strains, isolated from *Axinella dissimilis* sponges, which shared high amino acid identity per ORF and conserved gene cluster organisation to Cluster 1. The cluster identified within *Axinella*-associated *Pseudovibrio* isolates included genes encoding a polyketide cyclase and an enoyl-CoA hydratase domain at the beginning of the gene cluster (Naughton et al., 2017) which were not identified in Cluster 1. Both *Pseudovibrio* Ad5 and *Pseudovibrio* Ad46 have been successfully cultured and exhibited antimicrobial activity against *E. coli* and *Salmonella typhimurium* (Naughton et al., 2017). *Pseudovibrio* Ad5 and Genome 1 were shown to share 96% of their genes with predicted known function and therefore, it would be worthwhile attempting to culture the *Pseudovibrio* sp. (Genome 1) from *T. favus* sponges using similar culture conditions.

The predicted peptide for Cluster 1 was found to share greatest structural similarity with phakellistatin 7 and cyclonellin, both of which are cyclopeptides isolated from marine sponges within the Axinellidae family (Napolitano et al., 2005; Milanowski et al, 2004; Berer et al., 2004). Phakellistatin 7 has been shown to inhibit cancer cell growth (Pettit et al., 1995; Meli et al., 2017), whilst cyclonellin exhibits no anti-tumour activity (Milanowski et al., 2004). Three thioesterase (TE) domains were identified within Cluster 1. In NRPS clusters, TE domains are responsible for the macrocyclization of nascent NRPS peptides (Challis and Naismith, 2004). It is possible that the first TE domain performs this function, passes the cyclized protein to the following adenylation domain where two further monomers are added, after which the second TE domain releases the peptide through either hydrolysis or a second cyclization reaction. The third, stand-alone TE domain is hypothesised to remove either incorrectly loaded monomers from PCP domains or removes inactivating acetyl groups from the PCP domains, effectively regenerating the free thiol groups (Schwazer et al., 2002; Challis and Naismith, 2004). These proposed series of reactions could result in a large cyclopeptide, structurally related to cyclonellin and phakellistatin 7.

Taken together, Cluster 1 shares greatest amino acid identity with a *Pseudovibrio* sp. identified within an *Axinella* sponge and the proposed compound produced by Cluster 1 was predicted to be structurally similar to compounds isolated from Axinellidae sponges. Several sponged from the family Axinellidae have been found alongside latrunculid sponges on rocky reefs in Algoa Bay (Parker-Nance, Dorrington, unpublished data). Triterpenoid sodwanones and yardenones, which inhibit the hypoxia-inducible factor 1 in both breast and prostate tumour cell lines have been isolated from *Axinella* sponges collected off the coast of South Africa (Dai et al., 2006; Whibley et al., 2006). *Pseudovibrio* sp. are numerically abundant in culturable bacterial fractions from marine sponges (Naughton et al., 2017) and thus plans to culture the *Pseudovibrio* isolate from *Tsitsikamma* and *Axinella* sponge species collected in Algoa Bay are underway in the Dorrington Laboratory.

Cluster 2 was also identified within Genome 1 and appears to consist of several genes encoding putative proteins dedicated to facilitating drug resistance and efflux. *norM* and *norB* genes were identified in putative (Type III-PKS) Cluster 2. Both NorB and NorM facilitate the removal of toxic, fused ring compounds from bacterial cells through motive force mediated by a sodium pump, where an increase in *norB* transcripts resulted in a 4-fold increase in quinolone resistance in *S. aureus* bacteria (Truong-Bolduc et al., 2011; Kumar et al., 2013). The expression of *norB* can be inhibited by phenolic compounds such as chalcone (Ayaz et al., 2017) which was predicted to be produced by the PKS system in Cluster 2. NorM efflux pumps are found in Gram-positive bacteria are generally involved in the efflux of fluoroquinolones, toxic aminoglycosides and other cationic compounds (Tauch et al., 2005; Piddock, 2006). *Pseudovibrio* sp. are Gram-negative bacteria and it is tempting to speculate that the genes encoding the efflux pumps were acquired via horizontal gene transfer. The NorM protein was also found in Gram-negative *Erwinia amylovora* bacteria and was proposed to aid in colonization of its host plant through resistance to toxic compounds produced by co-colonizing epiphytic bacteria (Piddock, 2006). It is possible that, like *E. amylovora*, the *Pseudovibrio* (Genome 1) employs these multidrug efflux pumps to survive within *T. favus* sponges by removing toxic pyrroloiminoquinones that may enter the cell of the *Pseudovibrio* sp.

Two putative transcriptional regulators were also identified in Cluster 2 encoding a HipB-like XRE family transcriptional regulator and an MarR-like transcriptional

regulator. HipB is an antitoxin that binds the HipA toxin to form a dimer that can subsequently bind DNA and prevent further transcription of the *hipA* toxin encoding gene (Schumacher et al., 2009). The HipA-HipB toxin-antitoxin system plays an important role in the formation of persister cells in *E. coli*, which are phenotypically altered to slow protein production and survive antibiotic treatments through (Wen et al., 2014). Activation of the HipA toxin mediates the activation of RNase II toxins which leads to persistence of the bacterial host cell (Page and Peti, 2016). The *Pseudovibrio* sp. (Genome 1) may employ this retardation of cell proliferation to survive within the toxic environment of the *T. favus* sponge and could account for its absence as a dominant *T. favus* associated bacterial species.

A gene encoding a MarR transcriptional regulator was identified upstream of the *norB/M* and *hipA/B* genes but was orientated in the reverse direction. The *marR* gene is generally found orientated in the opposing direction to the operon on which it acts which allows a repressor to bind to the intergenic region between these genes and effectively inhibit the transcription of both (Grove, 2013). MarR transcriptional regulators are genes involved in regulation of the degradation of environmental toxins and are often controlled by the presence of small phenolic compounds such as stilbenes and chalcones (Grove, 2013). A gene encoding a chalcone-like synthase is located directly downstream of these two drug resistance loci and the translation products of these genes may be involved in mediating the activity of the drug resistant proteins.

The significant number of multidrug resistance proteins identified within Cluster 2 would suggest that that the *Pseudovibrio* bacterium from which Genome 1 is derived, is specifically adapted to survive environments in which antibiotics with fused ring structures are abundant such as within a *T. favus* sponge from which several pyrroloiminoquinone structures have been isolated. The *Pseudovibrio* bacterium investigated here would possess an ecological advantage in that it can survive the toxic chemical environment of the sponge host and effectively out compete bacteria devoid of this ability for a nutrient-rich, stable place to live.

Clusters 3 and 4 are very similar in terms of gene organization and predicted gene functions. Both clusters were found on Genomes 15 and 21 which were taxonomically classified as *Rhodopirellula* species. Both clusters include ORFs that encode enzymes

that enable the inclusion of selenocysteine into peptides, which are subsequently referred to as selenoproteins. Selenocysteine is considered the "21st naturally occurring amino acid" in which a selenium atom replaces the sulphur present in a cysteine residue (Cheng and Arner, 2017). Selenoproteins have been identified in all three domains of life and has been proposed as an ancient trait (Peng et al., 2016). However, the gene encoding the SelD (selenium utilization protein) has only been detected in a few bacterial genomes, suggesting that many bacteria have lost the ability to utilize selenium for functions such as redox homeostasis (Peng et al., 2016). Selenoproteins have been most commonly identified in marine host-associated bacteria but reasons underlying this trend have not yet been elucidated (Peng et al., 2016). The sulfatase present directly upstream of the *sel* genes cluster may be responsible for the hydrolysis of the sulphur residue from cysteine, enabling the addition of a selenium atom on the residue. Selenoproteins have been implicated in diverse biological functions (Labunskyy et al., 2014) and several bacterial selenoproteins have been found to play a role in defence against xenobiotic compounds through oxidoreductase activity (Dudkiewicz et al., 2012). Numerous genes encoding selenoproteins were identified in three marine Planctomycetes and were proposed to aid in adaptation to environmental stressors (Kim et al., 2016b).

The unusual co-localization of the *sel* genes within a single cluster has previously been observed in a *Rhodopirellula* isolate "P2", which encoded formylmethanofuran dehydrogenase which is an iron-sulphur selenoprotein (Vorholt et al., 1997; Kim et al., 2016b). Further inspection revealed that genes encoding formylmethanofuran dehydrogenase subunits A, B and D were co-localised in Genomes 15 and 21. Formylmethanofuran dehydrogenase performs the first reduction step of carbon dioxide fixation in methanogenic bacteria and indicates that the *Rhodopirellula* sp. identified here may possess the biosynthetic machinery required for methanogenesis. Clusters 3 and 4 share significant overall similarity to selenoprotein-encoding ORFs identified in *Kocuria* isolates *K. marina* S09-6 and *K. rhizophila* DC2201.

Research carried out by Kim and colleagues (2016b) suggested that the gene clusters encoding the selenoproteins found in *Rhodopirellula* isolate P2 were acquired via horizontal gene transfer to adapt to environmental pressures. It is tempting to speculate that the *Rhodopirellula* bacterial isolates from *T. favus*, viz. Genome 15 and Genome 21, may have acquired genes from the *Kocuria kristinae* isolates that are also

associated with *T. favus* (See Chapter 2). Further inspection of the draft genome of *K. kristinae* isolate RUTW02-3 revealed the presence of a single ORF predicted to encode a selenoprotein on Contig 4 which shared 37% amino acid identity with the ORF predicted to encode a selenoprotein in Genome 15. Furthermore, a leucine-tRNA synthase is present in Cluster 3 and Cluster 4 and studies have shown (Kim et al., 2016b) that this enzyme may enable the more efficient translation of newly acquired gene clusters such as the co-localized selenoprotein genes identified here.

Cluster 5, also present on *Rhodopirellula* sp. Genome 15, was identified as a hybrid NRPS/Type I-PKS gene cluster that shared greatest similarity with a gene cluster on the genome of terrestrial myxomycete *Pedosphaera parvula* (NZ_ABOX00000000.2). Regrettably, there have been no studies investigating the secondary metabolome of *P. parvula*. The acetyl-transferase and adenylation domains recognize and activate malonyl-CoA and glycine respectively. No tailoring domains were identified within the putative PKS gene but ORFs encoding a methyltransferase and a N-acetylgalatosamine-6-sulfatase flanked the putative PKS gene. Furthermore, sugar lactonase and acetyl esterase domains were identified within a hypothetical ORF and could provide further modifications to the nascent hybrid molecule. The predicted structure from Cluster 5 is predicted to match that of 4-Methyl-2-morpholinone. A review of the current literature showed that no biological activity has been attributed to this molecule although synthesised morpholinone derivatives have been shown to inhibit Tumour necrosis factor alpha (TNF-α) and interleukin 6 (IL-6) in microglial cells (Singh et al., 2017). The presence of tailoring enzymes such as the sugar lactonases, sulfatase and esterase/lipase indicate that further modifications could be made to the molecule that would provide ecological bioactivity.

Unlike Clusters 1 - 5, Cluster 6 was located on an “unclustered” contig as a Type I-PKS that binds N-acetyl-L-phenylalanine and malonyl monomers. BLAST analysis of individual ORFs revealed that the majority of putative genes shared greatest amino acid identity with genes found within Planctomycete species. The predicted structure of the Cluster 6 metabolite contained multiple features present within tsitsikammamine B. From the outset of this study it was hypothesised that the biosynthetic pathway responsible for the production of tsitsikammamines would resemble lantipeptide pathways found in *Streptomyces* sp. that produce structurally related pyrroloquinolines lymphostin and ammosamides (Jordan and Moore, 2016). However, no lantipeptide

pathways have thus far been identified within the metagenome of *T. favus* TIC2016-050A. This indicates that either, the pathway had not been sequenced/assembled and that more sequence data is required and/or that the tsitsikammamines are produced using an alternative approach that may not include a lantipeptide pathway.

The reported promiscuity of FAAL substrate recognition, combined with the presence of methyltransferase and oxygenase tailoring enzymes support an alternative proposal for pyrroloiminoquinone biosynthesis, in which tryptophan could replace the predicted FAAL substrate phenylalanine. This hypothesis motivates for an alternative route by which pyrroloiminoquinones could be produced by planctomycetes associated with *T. favus*. However, more metagenome sequence data is required to extend the contigs encoding Cluster 6 and subsequent clustering of a draft genome bin for taxonomic classification. Biochemical experiments, possibly using labelled isotopes (tryptophan and tyrosine) would be required to identify chemical intermediates consistent with this alternative biosynthetic pathway. Furthermore, heterologous expression of the gene in Cluster 6 could potentially confirm the predicted product structure. Attempts to culture both the Pseudovibrio and the Planctomycete identified here are currently underway.

In conclusion, metagenomic data analysis of the microbiome associated with *T. favus* sponges has shown that there is significant potential for secondary metabolite production. A total of 175 biosynthetic gene clusters were identified within this relatively small dataset indicating that the information garnered thus far is likely scratching the surface of the bioactive potential of the *T. favus* associated microbiome. Cluster 1 and Cluster 6 are pathways of greatest interest since the predicted cyclopeptide produced by Cluster 1 may have pharmaceutical application, whilst further investigation of Cluster 6 will determine if it is involved in the production of pyrroloiminoquinone compounds.

Chapter 7 – General Discussion and Future Prospects

7.1. General Discussion

Pyrroloiminoquinone compounds are pharmaceutically important due to their cytotoxic, antimicrobial, antimalarial and antifungal properties (Davis et al., 2012; Vanicat et al., 2016). Pyrroloiminoquinones have been exclusively isolated from sponges from the Acarnidae and the Latrunculiidae families, as well as a single tunicate within the *Clavelina* genus (Vanicat et al., 2016; Lin et al., 2017). To date, the biosynthetic pathways encoding these compounds have yet to be elucidated and synthetic forms of bioactive pyrroloiminoquinones remain the primary source of these compounds.

The gene clusters encoding structurally similar pyrroloquinolines, lymphostin and ammosamide, have been elucidated and potentially provide clues to the pyrroloiminoquinone biosynthetic pathway. Both pathways were discovered in culturable bacteria within the order Actinomycetales. The gene cluster encoding lymphostin was identified in three strains within the *Salinispora* genus, whilst the cluster responsible for the production of ammosamides was found in *Streptomyces* sp. CNR-698 (Jordan and Moore, 2016).

The research described in this thesis provided support for the hypothesis that pyrroloiminoquinones isolated from South African latrunculid sponges may be produced by bacterial symbionts (Antunes et al., 2005; Walmsley et al., 2012; Matcher et al., 2017). Tsitsikammamines A and B are found exclusively within *T. favus* and novel *T. nguni* sponge species (Antunes et al.,2005; Parker-Nance et al., submitted). The isolation of structurally closely related tsitsikammamine C from a geographically distant and phylogenetically distinct *Zyzzya* sponge species (Davis et al., 2012) provides further evidence that the biosynthetic origin of tsitsikammamines may be bacterial as there is no geographic nor phylogenetic similarity between the *Zyzzya* and *Tsitsikamma* sponge species. Previous studies that were focussed on the bacterial communities associated with latrunculid sponges showed that there were conserved betaproteobacteria and spirochete populations associated with latrunculid sponges (Walmsley et al., 2012). It was later shown that the phylogeny of conserved

betaproteobacteria present within latrunculid sponges followed that of the sponge species with which they were associated. The conserved spirochete was postulated as the potential producer of bioactive compounds isolated from latrunculid sponges (Matcher et al., 2017). This study focussed on expanding upon these previous studies and using both genomic and metagenomic approaches to identify candidate bacterial biosynthetic gene clusters for the production of pyrroloiminoquinone compounds, particularly tsitsikammamines, in South African *T. favus* sponges.

7.1.1. Conservation of sponge-associated microbiota

The bacterial communities associated with latrunculid sponges have shown to be conserved over a period of nine years, with conserved, numerically dominant betaproteobacteria and spirochetes. Furthermore, the bacterial communities associated with latrunculid sponges of the same genera collected in geographically distinct locations (Algoa Bay and Tsitsikamma National Park) were shown to be conserved. An encrusting *Mycale* sp. was shown to have a distinct microbiota to that of the *Tsitsikamma* sp. 005 on which it was living. This data indicates that latrunculid sponge-associated bacterial communities are host specific. Conserved betaproteobacterial OTUs were also associated with *L. apicalis* sponges collected from Bouvet Island (Sub-Antarctic) as well as with sympatric *T. rubra* sponges collected in Algoa Bay. Phylogenetic analysis of these OTUs revealed that betaproteobacteria associated with Algoa Bay sponges are more closely related to betaproteobacteria associated with geographically distant sponges of the same genera, than to one another. I.e. Betaproteobacteria associated with Algoa Bay latrunculid sponges are more closely related to betaproteobacteria associated with latrunculid sponges from the Southern Ocean than to those betaproteobacteria harboured in sympatric *Tethya* sponges in Algoa Bay. This provided evidence that conserved betaproteobacteria formed an ancient symbiosis with these sponges and appear to have co-evolved with their sponge host. Dominant, conserved betaproteobacteria have also been identified in *Crambe crambe*, *Amiphemedon queenslandica* and *Rhopaloeides odorabile* sponges (Webster et al., 2001; Croue et al., 2013; Gauthier et al., 2016). It has been proposed that the dominant betaproteobacterium associated with *C. crambe* had adapted to survive the toxic compounds produced by the host sponge which was partially validated by a

metatranscriptomics study revealing that the microbiota in *C. crambe* produced several proteins that conferred resistance to beta-lactam antibiotics (Croue et al., 2013; Versluis et al., 2015). The betaproteobacterium associated with *R. odorabile* was identified as an intracellular symbiont and the betaproteobacterium associated with *A. queenslandica* was shown to possess a reduced genome (Webster et al., 2001; Gauthier et al., 2016).

In this study, experiments to enrich bacterial cells from *T. favus* host sponge tissue showed that the conserved betaproteobacterium remained primarily in the remaining sponge pellet suggesting that it may be an endosymbiont like that observed in *R. odorabile.* The dominant betaproteobacterial bacterium in the *A. queenslandica* sponge, "AqS2", possessed a reduced genome that was complemented by the larger genomic repertoire of a dominant gammaproteobacterial symbiont (Gauthier et al., 2016). It is plausible that a similar relationship is present between the spirochetes and betaproteobacteria identified in *T. favus*. It is proposed that the genome of the betaproteobacterium is reduced (which would accommodate an intracellular lifestyle) whilst the larger genome of the conserved spirochete could provide metabolic complementation in order for the bacteria to work synergistically and thereby provide essential functions for the *T. favus* sponge host.

The genome of the dominant spirochete in *T. favus* (Genome 3) was successfully recovered with minor contamination but the genome recovered for a betaproteobacterium associated with *T. favus* (Genome 19) was predicted to have 38% contaminating sequence. Therefore, this data could not reliably be used to draw any conclusions from the genomic content of this bin. Thus, a comparative analysis investigation the potential for co-complementation between these two *T. favus* symbionts was not possible. More sequence data will be acquired in the coming year and once combined and assembled with the current data, it should enable the recovery of more accurate genome representatives from metagenomic data.

Conserved spirochete OTUs were identified exclusively in *Tsitsikamma* and *Cyclacanthia* sponges suggesting that this symbiosis may have been established more recently than that between the latrunculid sponges and the conserved betaproteobacteria. A single spirochete strain was shared between *T. favus* and *T. nguni* sponges (S_OTU1). Two further spirochetes, S_OTU2 and S_OTU3 were

detected in all *Tsitsikamma* sponges except *T. nguni*, whilst *C. bellae* was associated with a distinct spirochete strain (S_OTU5). Like the conserved betaproteobacterial taxa associated with latrunculid and *Tethya* sponges, the phylogeny of conserved spirochetes appeared to follow that of their host sponge.

To date, the conservation of a dominant spirochete symbiont has only been reported in latrunculid sponges, suggesting that these bacteria perform a specialised function within these closely related sponge genera. It was shown that the draft spirochete genome recovered from *T. favus* metagenomic data (Genome 3) consisted of two distinct spirochetes which appeared to represent novel genera. Due to phylogenetic relatedness, it is likely that Genome 3B represents the numerically dominant conserved spirochete present in *T. favus* and *T. nguni* sponges (S_OTU1) which was found to be the only metabolically active spirochete OTU in *T. favus* sponges. Genome 3A likely represents a spirochete present in the sponge at a lower abundance.

Inspection of Genome 3 revealed several genetic features that are purportedly involved in symbiosis including tetracopeptide repeat proteins, lipopolysaccharides, and mechanisms to evade host predation (Wyrick, 2000; Garderes et al., 2015; Webster and Thomas, 2016; Lorenzo et al., 2017). This study provides evidence to support the proposal that this *T. favus*-associated spirochete is likely an obligate symbiont that has adapted to survive in the toxic environment of the *T. favus* sponge through the biosynthesis of various multi-drug efflux pumps and proteins conferring resistance to toxins and antibiotics.

The conserved spirochete may play a role in defence of the host sponge against pathogenic bacteria through the production of quinolate and/or aid in the acquisition of carbohydrates through the metabolism of diverse carbon sources such as pullulan, maltose, and mannitol. Additional sequencing will aid in resolving more accurate draft genomes of both the conserved betaproteobacteria and spirochete. Comparative analysis of these genomes will aid in predicting whether any co-complementary roles are present that may further benefit the host sponge or influence the lifecycle of the *T. favus* sponge, as observed in the dominant symbionts of the *A. queenslandica* sponge (Gauthier et al., 2016).

7.1.2. Potential for bioactive secondary metabolite production in *T. favus*-associated microbiota

This study has not definitively identified a gene cluster that may be responsible for the production of tsitsikammamine nor of any other pyrroloiminoquinone compound. Cluster 6 was located on an unclustered contig and was predicted to encode a Type-I PKS system that could produce a compound similar to the makaluvamine-like moiety of tsitsikammamine B. Amino acid homology suggested that the contig encoding Cluster 6 may originate from a planctomycete genome. However, there were also the presence of genes that encoded proteins related to enzymes originating from a *Streptomyces* sp.. Several secondary metabolite biosynthetic gene clusters were reported to be located on plasmids (Le et al., 2013; Gonzales-Escalona et al., 2014; Wang et al., 2015) including a Type-III PKS gene cluster responsible for the production violapyrone B in the marine actinomycete *Streptomyces somaliensis* (Huang et al., 2016). It is possible that Cluster 6 may likewise be located on a plasmid because it does not appear to be included in a specific genome. If this were the case, then the production of the PKS product encoded on Cluster 6 could originate from various bacterial species within the sponge. Continuation off this study would require additional metagenomic sequence data. In this study, sequence reads were assembled into contigs using the SPAdes genome assembler. However, a new assembler has recently been released called metaSPAdes, the underlying algorithms of which are optimized for metagenomic datasets (Nurk et al., 2017). metaSPAdes currently only supports sequence data that consists of paired-end sequences and thus the Ion Torrent data generated in this study could not be assembled using this optimized assembler. It would therefore be suggested that any new sequence data be generated using paired-end Illumina sequence platforms such as MiSeq or HiSeq which include the added benefits of a 100-fold greater number of bases per run than Ion Torrent and has the lowest cost per Gb of data generated amongst commercially available sequencing platforms (Garrido-Cardenas et al., 2017). An optimised assembly of sequence reads, coupled with a greater number of raw sequence reads may result in the accurate recovery of both the genome of the conserved, dominant betaproteobacterial taxon as well as extension and binning of Cluster 6, proposed as the tsitsikammamine biosynthetic pathway. Furthermore, a new collaborator Dr Rachel Sipler from Memorial University, Canada, will conduct stable isotope analysis of tryptophan and tyrosine incorporation in pyrroloiminoquinone compounds in *T. favus*

tissue and extracted bacteria to identify structural intermediates that might correlate with the biosynthetic model proposed from Cluster 6.

Shotgun metagenomic analysis of microbiota from diverse habitats has resulted in the discovery of many novel genes which have proven to be of pharmaceutical or industrial benefit (Roumpeka et al., 2017). The biosynthetic potential of the *Pseudovibrio* bacterium represented by Genome 1 is of interest due the presence of an NRPS gene cluster that is predicted to encode a cyclopeptide which shares predicted structural features with an anti-tumour compound isolated from an Axinellidae sponge (Pettit et al., 1995; Meli et al., 2017). *Pseudovibrio* isolates exhibiting antibacterial activity have previously been cultured from an Irish *Axinella* sponge and have exhibited antimicrobial activity against a broad spectrum of microorganisms (O' Halloran et al., 2011; Naughton et al., 2017). This suggests that the isolation and subsequent assessment of bioactive secondary metabolites produced by the *Pseudovibrio* described in this study may be possible. The presence of a multidrug efflux pump in this putative *Pseudovibrio* sp. was proposed to enable its survival within the *T. favus* sponge within which toxic pyrroloiminoquinone compounds are produced. This protective mechanism can potentially be exploited during culturing efforts through the supplementation of culture media with purified pyrroloiminoquinone compounds, allowing the targeted *Pseudovibrio* sp. to outcompete lesser adapter sponge-associated bacteria. Furthermore, there are several *Axinella* sponge species that have been collected from Algoa Bay and application of the same culture strategy described above could provide a compelling potential comparison of sponge-associated *Pseudovibrio* isolates and their biosynthetic potential. The capacity to assay crude extracts from *T. favus*- and *Axinella* sp.- associated *Pseudovibrio* sp. for antiviral, anti-tumour and antimicrobial activity is available in the Dorrington Laboratory. Potential drug lead compounds can be purified and characterised through High Performance Liquid Chromatography (HPLC) and Liquid Chromatography – Mass Spectrometry (LC-MS).

7.2. Future Prospects

7.2.1. Culturing of Sponge-associated actinomycetes

Culturing and shotgun genome sequencing of marine actinomycete *K. kristinae* strains RUTW2-3 and RUTW4-5 revealed the presence of biosynthetic gene clusters encoding a putative siderophore and a putative terpene which were anticipated to aid in the survival of these bacteria in a marine environment. However, no gene clusters encoding novel bioactive compounds were recovered. Culturing of *Tsitsikamma* sponge associated bacteria was continued and resulted in the isolation of *Streptomyces* sp. RUSV204 from *T. michaeli* sponge TIC2015-204. Antibacterial activity assays revealed that *Streptomyces* sp. RUSV204 can inhibit the growth of *B. subtilis* and furthermore was found to inhibit the growth of *E. coli* after heat shock treatment. This bacterial growth inhibition activity, prompted the sequencing of *Streptomyces* sp. RUSV204 genome which showed that several biosynthetic gene clusters were present, including two NRPS gene clusters, a Type-III PKS gene cluster and a lantipeptide gene cluster. However, the N_{50} value of the genome was 3045 nts, which is low relative to recent bacterial genome assemblies and therefore indicates that there are few contigs of biologically meaningful size. Crude chemical extracts were obtained from cultured *Streptomyces* sp. VS204 and have been processed by HPLC and are currently being analysed by LC-MS-MS. Should the data appear promising then the crude extract will be fractionated, and fractions assayed for biological activity. Additional sequence data will be obtained for a more accurate assessment of the biosynthetic gene clusters in the *Streptomyces* sp. VS204 isolate to identify the genes responsible for the production of the compound eliciting the observed anti-bacterial activity.

7.2.2. Metagenomic analysis of *Tsitsikamma pedunculata* sponges

During analysis of sponge-associated bacterial communities, it was observed that *T. pedunculata* sponge species harboured two conserved gammaproteobacterial OTUs, OUT_Illu_4 and OUT_Illu_6, that were not detected in any other sponge species nor in the surrounding environment (Fig. 3.11). *T. pedunculata* sponges produce a suite of secondary metabolites that are distinct from those produced by other *Tsitsikamma*

sponge species including several brominated discorhabdin compounds. Gammaproteobacterial isolates from Australian *Candidaspongia flabellata* and *Rhopaloeides odorabile* sponges have been shown to exhibit antimicrobial activity (Brinkmann et al., 2017a) and gammaproteobacterium, *Vibrio harveyi*, is responsible for the production of spongosine in *Tectitethya crypta* sponges (Bertin et al., 2015). Metagenomic analysis and potential recovery of gammaproteobacterial genomes from *T. pedunculata* may reveal bacterial candidates that are involved in the production of the brominated pyrroloiminoquinones isolated from this sponge species.

7.3. Concluding remarks

A large number of bioactive secondary metabolites have been isolated from South African marine species collected from Algoa Bay which includes cytotoxic pyrroloiminoquinones from marine sponges (Antunes et al., 2004; Antunes et al., 2005) as well as extremely potent anti-tumour cephalostatins from marine tube worm *Cephalodiscus gilchristi* (Pettit et al., 1987) and cytotoxic mandelalides from a *Lissoclinum* tunicate (Sikorska et al., 2012). The potential for the discovery of novel bioactive products with unique and unusual structures is undoubtedly high and with the diversity present in Algoa Bay, it is clear that we have merely scratched the surface of the pharmaceutical potential hidden in these reefs. Further research into sponges, tunicates and other marine invertebrates located in Algoa Bay and other South African waters is evidently warranted.

References

Abad, M. J., Bedoya, L. M., & Bermejo, P. (2011). Marine Compounds and their Antimicrobial Activities. In Science against Microbial Pathogens: Communicating Current Research and Technological Advances (pp. 1293– 1306). Retrieved from http://www.formatex.info/microbiology3/book/1293-1306.pdf

Abdelmohsen, U.R., Bayer, K., Hentschel, U. (2014a). Diversity, abundance and natural products of marine sponge-associated actinomycetes, Natural Product Reports, 31, 381 – 399. DOI: 10.1039/C3NP70111E

Abdelmohsen, U.R., Yang, C., Horn, H., Hajjar, D., Ravasi, T., Hentschel, U. (2014b) Actinomycetes from Red Sea Sponges: Sources for Chemical and Phylogenetic Diversity, Mar Drugs, 12, 2771–2789. DOI: 10.3390/md12052771

Agarwal, V., Blanton, J.M., Podell, S., Taton, A., Schorn, M.A., Busch, J., Lin, Z., Schmidt, E.W., Jensen, P.R., Biggs, J.S., Golden, J.W., Allen, E.E., Moore, B.S. (2017). Metagenomic discovery of polybrominated diphenyl ether biosynthesis by marine sponges, Nature Chemical Biology, 13, 537 – 543. DOI: 10.1038/nchembio.2330

Ahlert, J., Shepard, E., Lomovskaya, N., Zazopoulos, E., Staffa, A., Bachmann, B.O., Huang, K., Fonstein, L., Czisny, A., Whitwam, R.E., Farnet, C.M., Thorson, J.S. (2002). The Calicheamicin Gene Cluster and Its Iterative Type I Enediyne PKS, Science, 297(5584), 1173 – 1176. DOI: 10.1126/science.1072105

Alex, A., Antunes, A. (2015). Whole Genome Sequencing of the Symbiont *Pseudovibrio* sp. from the Intertidal Marine Sponge *Polymastia penicillus* Revealed a Gene Repertoire for Host-Switching Permissive Lifestyle, Genome Biology and Evolution, 7(11), 3022 – 3032. DOI: 10.1093/gbe/evv199

Alexander, D. B., Zuberer, D.A. (1991) Use of Chrome Azurol S Reagents to Evaluate Siderophore Production by Rhizosphere Bacteria, Biology and Fertility of Soils, 12(1), 39 – 45. DOI: 10.1007/BF00369386

Alferova, I.V., Terekhova, L.P., Prauzer, K.H. (1989). Antibiotiki I khimioterapiia, 34(5), 344 – 348. Translated from Russian.

Amirkia, V., Heinrich, M. (2015) Natural products and drug discovery: A survey of stakeholders in industry and academia, Frontiers in Pharmacology, 6, 237. https://doi.org/10.3389/fphar.2015.00237

Ansari, M.Z., Yadav, G., Gokhale, R.S., Mohanty, D. (2004). NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases, Nucleic Acids Research, 32(2), 405 – 413. DOI: 10.1093/nar/gkh359

Antunes, E.M., Beukes, D.R., Kelly, M., Samaai, T., Barrows, L.R., Marshall, K.M., Sincich, C., Davies-Coleman, M.T. (2004). Cytotoxic pyrroloiminoquinones from four new species of South African latrunculid sponges, Journal of Natural Products, 67, 1268-1276. DOI: 10.1021/np034084b

Antunes, E.M., Copp, B.R., Davies-Coleman, M.T., Samaai, T. (2005). Pyrroloiminoquinone and related metabolites from marine sponges. Natural Products Reports, 22(1), 62-72. DOI: 10.1039/b407299p

Aotani, Y., Nagata, H., Yoshida, M. (1997). Lymphostin (LK6-A), a Novel Immunosuppressant from *Streptomyces* sp. KY11783: Structural Elucidation, The Journal of Antibiotics, 50(7), 543 – 545.

Arora, P., Goyal, A., Natarajan, V.T., Rajakumara, E., Verma, P., Gupta, R., Yousuf, M., Trivedi, O.A., Mohanty, D., Tyagi, A., Sankaranarayanan, R., Gokhale, R.S. (2009). Mechanistic and functional insights into fatty acid activation in *Mycobacterium tuberculosis*, Nature Chemical Biology, 5(3), 166 – 173. DOI: 10.1038/nchembio.143

Astudillo-Garcia, C., Bell, J.J., Webster, N.S., Glasl, B., Jompa, J., Montoya, J.M., Taylor, M.W. (2017) Evaluating the core microbiota in complex communities: A systematic investigation, Environmental Microbiology, 19, 1450 – 1462. DOI: 10.1111/1462-2920.13647

Ayaz, M., Subhan, F., Sadiq, A., Ullah, F., Ahmed, J., Sewell, R.D.E. (2017). Cellular efflux transporters and the potential role of natural products in combating efflux mediated drug resistance, Frontiers in Bioscience, 22, 732 – 756. DOI: 10.2741/4513

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O. (2008). The RAST server: Rapid annotations using subsystems technology, BMC Genomics, 9, 75. DOI: 10.1186/1471-2164-9-75

Bailly, C. (2015). Anticancer Properties of Lamellarins, Marine Drugs, 13(3), 1105–1123. DOI:10.3390/md13031105

Barnes, E.C., Kumar, R., Davis, R.A. (2016). The use of isolated natural products as scaffolds for the generation of chemically diverse screening libraries for drug discovery, Natural Products Reports, 33, 372 – 381. DOI: 10.1039/C5NP00121H

Beepat, S.S., Appadoo, C., Marie, D.E.P., Sadally, S.B., Paula, J.P.M., Sivakumar, K., Rao, R.R., Salah, M (2016) First Record of Sponge-associated Actinomycetes from Two Coastal Sponges from Mauritius, Western Indian ocean Journal of Marine Science., 15(1), 31 – 38.

Behie, S.W., Bonet, B., Zacharia, V.M., McClung, D.J., Traxler, M.F. (2017). Molecules to Ecosystems: Actinomycete Natural Products In situ, Frontiers in Microbiology, 7. DOI: 10.3389/fmicb.2016.02149

Belanger, M., Burrows, L.L., Lam, J.S. (1999). Functional analysis of genes responsible for the synthesis of the B-band O antigen of *Pseudomonas aeruginosa* serotype O6 lipopolysaccharide, Microbiology, 145, 3505 – 3521. DOI: 10.1099/00221287-145-12-3505

Belarbi, E. H., Contreras Gómez, A., Chisti, Y., García Camacho, F., Molina Grima, E. (2003). Producing drugs from marine sponges. Biotechnology Advances, 21(7), 585–598. DOI:10.1016/S0734-9750(03)00100- 9

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., Rabbinoswitch, 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(6885), 141–147. DOI: 10.

Berer, N., Rudi, A., Goldberg, I., Benayahu, Y., Kashman, Y. (2004). Callynormine A, a New Marine Cyclic Peptide of a Novel Class, Organic Letters, 6(15), 2543 – 2545. DOI: 10.1021/ol0491787

Bergmann, W., Burke, D. C. (1956). Contributions to the Study of Marine Products. XL. The Nucleosides of Sponges. IV. Spongosine. Journal of Organic Chemistry, 21(1343), 226 – 228. DOI:10.1021/jo01108a020

Bergmann, W., Feeney, R. J. (1951). Contributions to the study of marine products. XXXII. The nucleosides of sponges. The Journal of Organic Chemistry, 16, 981–987. DOI:10.1021/jo01146a023

Bergmann, W., Stempien, M. F. (1957). Contributions to the Study of Marine Products. XLIII. The Nucleosides of Sponges. V. The Synthesis of Spongosine'. Journal of Organic Chemistry, 22(December), 1575– 1577. DOI:10.1021/jo01363a009

Bertin, M.J., Vulpanovici, A., Monroe, E.A., Korobeynikov, A., Sherman, D.H., Gerwick, L., Gerwick, W.H. (2016). The Phormidolide Biosynthetic Gene Cluster: A trans-AT PKS Pathway Encoding a Toxic Macrocyclic Polyketide, ChemBioChem, 17, 164 – 173. DOI: 10.1002/cbic.201500467

Bertin, M.J.; Schwartz, S.L.; Lee, J.; Korobeynikov, A.; Dorrestein, P.C.; Gerwick, L.; Gerwick,W.H. (2015). Spongosine Production by a *Vibrio harveyi* Strain Associated with the Sponge *Tectitethya crypta*. Journal of Natural Products, 78(3), 493–499. DOI: 10.1021/np5009762

Blazewicz, S. J., Barnard, R. L., Daly, R., Firestone, M. K. (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. The ISME Journal, 7(11), 2061–2068.

Blunt, J.W., Copp, B.R., Keyzers, R.A., Munro, M.H.G., Prinsep, M.R. (2017). Marine natural products, Natural Product Reports, 32, 116 – 211. DOI: 10.1039/c4np00144c

Blunt, J.W., Copp, B.R., Keyzers, R.A., Munro, M.H.G., Prinsep, M.R. (2017). Marine natural products, Natural Products Reports, 34, 235 – 294. DOI: 10.1039/C6NP00124F

Boiteau, R.M., Mende, D.R., Hawco, N.J., McIlvin, M.R., Fitzsimmons, J.N., Saito, M.A., Sedwick, P.N., DeLong, E.F., Repeta, D.J. (2016). Siderophore-based microbial adaptations to iron scarcity across the eastern Pacific Ocean, Proceedings of the National Academy of Sciences of the United States of America, 113(50), 14237 – 14242. DOI: 10.1073/pnas.1608594113

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, 2095 – 2113. DOI: 10.1111/1462-2920.12123

Bose, U., Hodson, M.P., Shaw, P.N., Fuerst, J.A., Hewavitharana, A.K. (2014). Two peptides, cycloaspeptide A and nazumamide A from a sponge associated marine actinobacterium *Salinispora* sp., Natural product Communications, 9(4), 454 – 546.

Botic, T., Cor, D., Anesi, A., Guella, G., Sepcic, K., Janussen, D., Kerksen, D., Knez, Z. (2015). Fatty acid composition and antioxidant activity of Antarctic marine sponges of the genus *Latrunculia*, Polar Biology, 38, 1605 – 1612. DOI 10.1007/s00300-015-1722-z

Botic, T., Defant, A., Zanini, P., Zuzek, M.C., Fragez, R., Janussen, D., Kerksen, D., Knez, Z., Mancini, I., Sepcic, K. (2017). Discorhabdin alkaloids from Antarctic *Latrunculia* spp. sponges as a new class of cholinesterase inhibitors, European Journal of Medicinal Chemistry, 136(18), 294 – 304. DOI: 10.1016/j.ejmech.2017.05.019

Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., Schulz, F., Jarett, J., Rivers, A.R., Eloe-Fadrosh, E.A.E., Tringe, S.G., Ivanova, N.N., Copeland, A., Clum, A., Becraft, E.D., Malstrom, R.R., Birren, B., Podar, M., Bork, P., Weinstock, G.M., Garrity, G.M., Dodsworth, J.A., Yooseph, S., Sutton, G., Glockner, F.O., Gilbert, J.A., Nelson, W.C., Hallam, S.J., Jungbluth, S.P., Ettema, T.J.G., Tighe, S., Konstantinidis, K.T., Liu, W., Baker, B.J., Rattei, T., Eisen, J.A., Hedlund, B., McMahon, K.D., Fierer, N., Knight, R., Finn, R., Cochrane, G., Karsch-Mizrachi, I., Tyson, G.W., Rinke, C., Lapidus, A., Meyer, F., Yilmaz, P., Parks, D.H., Eren, A.M., Schriml, L., Banfield, J.F., Hugenholtz, P., Woyke, T. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea, Nature Biotechnology, 35, 725 – 731. DOI: 10.1038/nbt.3893

Bramhachari, P.V., Mutyala, S., Bhatnagar, I., Pallela, R. (2016) In: Pallela R. and Ehrlich H. (ed.) Marine Sponges: Chemicobiological and Biomedical Applications, Springer, India.

Brettin, T., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Olsen, G.J., Olson, R., Overbeek, R., Parrello, B., Pusch, G.D., Shukla, M., Thomason, J.A., Stevens, R., Vonstein, V., Wattam, A.R., Xia, F. (2015). RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes, Scientific Reports, 5, 8365. DOI:10.1038/srep08365.

Bright, M., Bulgheresi, S. (2010). A complex journey: transmission of microbial symbionts, Nature Reviews Microbiology, 8, 218-230. DOI: 10.1038/nrmicro2262

Brinkmann, C.M., Kearns, P.S., Evans-Illidge, E., Kurtboke, I. (2017a). Diversity and Bioactivity of Marine Bacteria Associated with the Sponges *Candidaspongia flabellata* and *Rhopaloeides odorabile* from the Great Barrier Reef in Australia, Diversity, 9(3), 39. DOI: 10.3390/d9030039

Brinkmann, C.M., Marker, A., Kurtboke, D.I. (2017b). An Overview on Marine Sponge-Symbiotic Bacteria as Unexhausted Sources for Natural Product Discovery, Diversity, 9(4), 40. DOI: 10.3390/d9040040

Bull, A. T., Stach, J. E. M. (2007). Marine actinobacteria: new opportunities for natural product search and discovery. Trends in Microbiology, 15(11), 491–499. DOI: 10.1016/j.tim.2007.10.004

Butler, M.S., Robertson, A.A., Cooper, M.A. (2014). Natural product and natural product derived drugs in clinical trials, Natural products Reports, 31(11), 1612 – 1661. DOI: 10.1039/C4NP00064A

Caboche, S., Leclere, V., Pupin, M., Kucherov, G., Jacques, P. (2010). Diversity of Monomers in Nonribosomal Peptides: towards the Prediction of Origin and Biological Activity, Journal of Bacteriology, 192(19), 5143 – 5150. DOI: 10.1128/JB.00315-10

Calcutt, M.J., Foecking, M.F. (2015). Complete Genome Sequence of *Kocuria palustris* MU14/1, Genome Announcements, 3(5). DOI: 10.1128/genomeA.01195-15.

Cardenas, C.A., Bell, J.J., Davy, S.K., Hoggard, M., Taylor, M.W. (2014). Influence of environmental variation on symbiotic bacterial communities of two temperate sponges, FEMS Microbiology Ecology, 88(3), 516 - 527. DOI: 10.1111/1574-6941.12317

Carney, J.R., Scheuer, P.J. (1993). Makaluvamine G, a cytotoxic pigment from an Indonesian sponge *Histodermella* sp., Tetrahedron, 49(38), 8483 – 8486. DOI: 10.1016/S0040-4020(01)96256-8

Carstens, B. B., Rosengren, K. J., Gunasekera, S., Schempp, S., Bohlin, L., Dahlström, M., Clark, R.J., Göransson, U. (2015). Isolation, Characterization, and Synthesis of the Barrettides: Disulfide-Containing Peptides from the Marine Sponge *Geodia barrette*, Journal of Natural Products, 150729101204002. DOI: 10.1021/acs.jnatprod.5b00210

Casburi, G., Duscher, A.A., Reid, R.P., Foster, J.S. (2016). Characterization of the stromatolite microbiome from Little Darby Island, The Bahamas using predictive and whole shotgun metagenomic analysis, Environmental Microbiology, 18(5), 1452 – 1469. DOI: 10.1111/1462-2920.13094

Cassapullo, A., Cutignano, A., Bruno, I., Bifulco, G., Debitus, C., Gomez-Paloma, L., Riccio, R. (2001). Makaluvamine P, a New Cytotoxic Pyrroloiminoquinone from *Zyzzya cf. fuliginosa*, Journal of Natural Products, 64(10), 1354 – 1356. DOI: 10.1021/np010053+

Challis, G.L., Naismith, J.H. (2004). Structural aspects of non-ribosomal peptide biosynthesis, Current Opinion in Structural Biology, 14(6), 748 – 756. DOI: 10.1016/j.sbi.2004.10.005

Chang, L.C., Otero – Quintero, S., Hooper, J.N.A., Bewley, C.A. (2002a). Batzelline D and Isobatzelline E from the Indopacific Sponge *Zyzzya fuliginosa*, Journal of Natural Products, 65(5), 776 – 778. DOI: 10.1021/np010581l

Chang, Y. C., Tseng, S. W., Liu, L. L., Chou, Y., Ho, Y. S., Lu, M. C., Su, J. H. (2012). Cytotoxic sesterterpenoids from a sponge *Hippospongia* sp. Marine Drugs, 10(5), 987–997. DOI:10.3390/md10050987

Chang, Z., Flatt, P., Gerwick, W.H., Nguyen, V., Willis, C.L., Sherman, D.H. (2002b). The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit, Gene, 296, 235 – 247. DOI: 10.1016/s0378-1119(02)00860-0

Chatterjee, S.N., Chaudhuri, K. (2004). Lipopolysaccharides of *Vibrio cholerae*: II. Genetics of biosynthesis, Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 1690(2), 93 – 109. DOI: 10.1016/j.bbadis.2004.06.006

Chen, E.Z., Bushman, F.D., Li, H. (2017). A Model-Based Approach for Species Abundance Quantification Based on Shotgun Metagenomic Data, Statistics in Bioscience, 9(1), 13 – 27. DOI: 10.1007/s12561-016-9148

Cheng, Q., Arner, E.S.J. (2017). Selenocysteine insertion at a predefined UAG codon in a release factor 1 (RF1) depleted *Escherichia coli* host strain bypasses species barriers in recombinant selenoprotein translation, Journal of Biological Chemistry, 292, 5476 – 5487. DOI: 10.1074/jbc.M117.776310

Cheng, Y., Tang, G., Shen, B. (2003). Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis, Proceedings of the National Academy of Sciences of the United States of America, 100(6), 3149 – 3154. DOI: 10.1073/pnas.0537286100

Chianese, G., Gu, B.-B., Yang, F., Jiao, W.-H., Guo, Y.-W., Lin, H.-W., Taglialatela-Scafati, O. (2015). Spiroplakortone, an unprecedented spiroketal lactone from the Chinese sponge *Plakortis simplex*. RSC Adv., 5(78), 63372–63376. DOI:10.1039/C5RA09840H

Chiang, Y., Chang, S., Oakley, B.R., Wang, C.C.C. (2011). Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms, Current Opinion in Chemical Biology, 15, 137 – 143. DOI: 10.1016/j.cbpa.2010.10.011

Christensen, A., Martin, G.D.A. (2017). Identification and bioactive potential of marine microorganisms from selected Florida coastal areas, Microbiology Open, 6(4), 00448. DOI: 10.1002/mbo3.448

Cita, Y.P., Suhermanto, A., Radjasa, O.K., Sudharmono, P. (2017) Antibacterial activity of marine bacteria isolated from sponge *Xestospongia testudinaria* from Sorong, Papua, Asian Pacific Journal of Tropical Biomedicine, 7(5), 450 – 454.DOI: 10.1016/j.apjtb.2017.01.024

Clooney, A.G., Fouhy, F., Sleator, R.D., O'Driscoll, A., Stanton, C., Cotter, P.D., Claesson, M.J. (2016). Comparing Apples and Oranges?: Next Generation Sequencing and Its Impact on Microbiome Analysis, PLoS ONE, 11(2), e0148028. DOI: 10.1371/journal.pone.0148028

Copp, B.R., Fulton, K.F., Perry, N.B., Blunt, J.W., Munro, M.H.G. (1994). Natural and Synthetic Derivatives of Discorhabdin C, a Cytotoxic Pigment from the New Zealand Sponge *Latrunculia* cf. *bocagei*, The Journal of Organic Chemistry, 59(26), 8233 – 8238. DOI: 10.1021/jo00105a047

Copp, B.R., Ireland, C.M. (1991) Wakayin: A novel cytotoxic pyrroloiminoquinone alkaloid from the ascidian *Clavelina* species, Journal of Organic Chemistry, 56(15), 4596 – 4597. DOI: 10.1021/jo00015a005

Croue, J., West, N.J., Escande, M., Intertaglia, L., Lebaron, P., Suzuki, M.T. (2013). A single betaproteobacterium dominates the microbial community of the crambescidine-containing sponge *Crambe crambe*, Scientific Reports, 3, DOI: 10.1038/srep02583

D'Ambrosio, M., Guerriero, A., Chiasera, G., Pietra, F., Tato, M. (1996). Epinardins A-D, new pyrroloiminoquinone alkaloids of undetermined deep-water green demosponges from pre-Antarctic Indian ocean, Tetrahedron, 52(26), 8899 – 8906. DOI: 10.1016/0040-4020(96)00438-3

D'Auria, M.V., Sepe, V., D'Orsi, R., Bellotta, F., Debitus, C., Zampella, A. (2007). Isolation and structural elucidation of callipeltins J–M: antifungal peptides from the marine sponge *Latrunculia* sp., Tetrahedron, 63(1), 131 – 140. DOI: 10.1016/j.tet.2006.10.032

Dai, J., Fishback, J.A., Zhou, Y., Nagle, D.G. (2006). Sodwanone and Yardenone Triterpenes from a South African Species of the Marine Sponge *Axinella* Inhibit Hypoxia-Inducible Factor-1 (HIF-1) Activation in both Breast and Prostate Tumor Cells, Journal of Natural Products, 69(12), 1715 – 1720. DOI: 10.1021/np060278q

Darling, A.E., Mau, B., Perna, N.T. (2010). progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement, PLoSOne, 5(6), e11147. DOI: 10.1371/journal.pone.0011147

Dattelbaum, J. D., Singh, A. J., Field, J. J., Miller, J. H., Northcote, P. T. (2015). The Nitrogenous Hamigerans: Unusual Amino Acid-Derivatized Aromatic Diterpenoid Metabolites from the New Zealand Marine Sponge *Hamigera tarangaensis*, Journal of Organic Chemistry, 80, 304–312. DOI: 10.1021/jo502370b

Davis, R. A., Buchanan, M. S., Duffy, S., Avery, V. M., Charman, S. A., Charman, W. N., White, K.L., Shackleford, D.M., Edstein, M.D., Andrews, K.T., Camp, D., Quinn, R. J. (2012). Antimalarial activity of pyrroloiminoquinones from the Australian marine sponge *Zyzzya* sp., Journal of Medicinal Chemistry, 55(12), 5851–5858. DOI:10.1021/jm3002795

de Lira, T.O., Berlinck, R.G.S., Nascimento, G.G.F., Hajdu, E. (2006). Further dibromotyrosine-derived metabolites from the marine sponge *Aplysina caissara*, Journal of the Brazilian Chemical Society, 17(7). DOI: 10.1590/S0103-50532006000700007

Delforno, T.P., Junior, G.V.L., Noronha, M.F., Sakamoto, I.K., Varesche, M.B.A., Olivera, V.M. (2017). Microbial diversity of a full-scale UASB reactor applied to poultry slaughterhouse wastewater treatment:

integration of 16S rRNA gene amplicon and shotgun metagenomic sequencing, MicrobiologyOpen, 6(3), 00443. DOI: 10.1002/mbo3.443

De Long, E.F. (1992). Archaea in coastal marine environments, Proceedings of the National Academy of Science USA, 89(12), 5685 – 5689.

De Silva, C.V., Cruz, L., Araujo, N., Angeloni, M.B., Fonseca, B.B., Goomes, A., Carvalho, F., Goncalves, L.R., Barbosa, B. (2012). A glance at *Listeria* and *Salmonella* cell invasion: Different strategies to promote host actin polymerization, International Journal of Medical Microbiology, 302(1), 19 – 32. DOI: 10.1016/j.ijmm.2011.05.003

Devine, R., Hutchings, M.I., Holmes, N.A. (2017). Future directions for the discovery of antibiotics from actinomycete bacteria, Emerging Topic in Life Sciences, 1(1), 1 – 12. DOI: 10.1042/ETLS20160014

Dijoux, M., Schnabel, P.C., Hallock, Y.F., Boswell, J.L., Johnson, T.R., Wilson, J.A., Ireland, C.M., van Soest, R., Boyd, M.R., Barrows, L.R., Cardellina, J.H. (2005). Antitumor activity and distribution of pyrroloiminoquinones in the sponge genus *Zyzzya*, Bioorganic and Medicinal Chemistry, 13(21), 6035 – 6044. DOI: 10.1016/j.bmc.2005.06.019

Dolusic, E., Larrieu, P., Meinguet, C., Colette, D., Rives, A., Blanc, S., Ferain, T., Pilotte, L., Stroobant, V., Wouters, J., van den Eynde, B., Masereel, B., Delfourne, E., Frederick, R. (2013). Indoleamine 2,3-dioxygenase inhibitory activity of derivatives of marine alkaloid tsitsikammamine A, Bioorganic and Medicinal Chemistry Letters, 23(1), 47 - 54. DOI: 10.1016/j.bmcl.2012.11.036

Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J., Katz, L. (1991). Modular Organization of Genes required for Complex Polyketide Biosynthesis, Science, 252, 675 -679. DOI: 10.1126/science.2024119

Dorrington, R.A., Hilliar, S.H., Kalinski, J.C.J., Krause, R.W.M., McPhail, K.L., Parker-Nance, S., Walmsley, T.A., Waterworth, S.C. (2016). Latrunculid sponges, their microbial communities and secondary metabolites: connecting conserved bacterial symbionts to pyrroloiminoquinone production, Planta Medica, 82 (S 01). S1-S381. DOI: 10.1055/s-0036-1596655

Dorrington, R.A., Lombard, A.T., Bornman, T.G., Adams, J.B., Cawthra, H.C., Deyzel, S.H.P., Goschen, W.S., Liu, K., Mahler-Coetzee, J., Matcher, G.F., McQuaid, C., Parker-Nance, S., Paterson, A., Perissinotto, R., Porri, F., Roberts, M., Snow, B., Vrancken, P. (2017). Working together for our oceans: A Marine Spatial Plan for Algoa Bay, South Africa. South African Journal of Science. In the Press

Du, L., Sanchez, C., Shen, B. (2001). Hybrid Peptide-Polyketide Natural Products: Biosynthesis and Prospects toward Engineering Novel Molecules, Metabolic Engineering, 3, 78 – 95. DOI: 10.1006/mben.2000.017

Dudkiewicz, M., Szczepinska, T., Grynberg, M., Pawlowski, K. (2012). A Novel Protein Kinase-Like Domain in a Selenoprotein, Widespread in the Tree of Life, PLoS One, 7(2), e32138. DOI: 10.1371/journal.pone.0032138

Edwards, J.L., Smith, D.L., Connolly, J., McDonald, J.E., Cox, M.J., Joint, I., Edwards, C., McCarthy, A.J. (2010). Identification of Carbohydrate Metabolism Genes in the Metagenome of a Marine Biofilm Community Shown to Be Dominated by Gammaproteobacteria, Bacteroidetes, Genes (Basal), 1(3), 371 – 384. DOI: 10.3390/genes1030371

Ehrlich, H., Bazhenov, V.V., Debitus, C., de Voogd, N., Galli, R., Tsurkan, M.V., Wysokowski, M., Meissner, H., Bulut, E., Kaya, M., Jesionowski, T. (2017). Isolation and identification of chitin from heavy mineralized skeleton of *Suberea clavata* (Verongida: Demospongiae: Porifera) marine demosponge, International Journal of Biological Macromolecules, 104, 1706 – 1712.

Ehrlich, H., Maldonado, M., Spindler, K., Eckert, C., Hanke, T., Born, R., Goebel, C., Simon, P., Heinemann, S., Worch, H. (2007). First evidence of chitin as a component of the skeletal fibers of marine sponges. Part I. Verongidae (demospongia: Porifera), Jez-B Molecular and Developmental Evolution, 308B(4), 347 – 356. DOI: 10.1002/jez.b.21156

Ereskovsky, A.V., Gonobobleva, E., Vishnayakov, A. (2005). Morphological evidence for vertical transmission of symbiotic bacteria in the viviparous sponge *Haliscara dujardini* Johnston (Porifera, Demospongiae, Halisarcida), Marine Biology, 146, 869-875. DOI 10.1007/s00227-004-1489-1

Erpenbeck, D., Voigt, O., Al-Aidaroos, A.M., Berumen, M.L., Buttner, G., Catania, D., Guirguis, A.N., Paulay, G., Schatzle, S., Worheide, G. (2016). Molecular biodiversity of Red Sea demosponges, Marine Pollution Bulletin, 105(2), 507 - 514. DOI: 10.1016/j.marpolbul.2015.12.004

Erwin, P. M., Coma, R., Lopez-Sendino, P., Serrano, E., Ribes, M. (2015). Stable symbionts across the HMA-LMA dichotomy: low seasonal and interannual variation in sponge-associated bacteria from taxonomically diverse hosts, FEMS Microbial Ecology, 91(10), doi:10.1093/femsec/fiv115

Esposito, G., Teta, R., Miceli, R., Ceccarelli, L., Della Sala, G., Camerlingo, R., Irollo, E., Mangoni, A., Pirozzi, G., Costantino, V. (2015). Isolation and Assessment of the in Vitro Anti-Tumor Activity of Smenothiazole A and B, Chlorinated Thiazole-Containing Peptide/Polyketides from the Caribbean Sponge, *Smenospongia aurea*. Marine Drugs, 13(1), 444–459. DOI:10.3390/md13010444

Essack, M., Bajic, V. B., Archer, J. A. C. (2011). Recently confirmed apoptosis-inducing lead compounds isolated from marine sponge of potential relevance in cancer treatment. Marine Drugs, 9(9), 1580–1606. DOI:10.3390/md9091580

Esteves, A.I.S., Amer, N., Nguyen, M., Thomas, T. (2016). Sample Processing Impacts the Viability and Cultivability of the Sponge Microbiome, Frontiers in Microbiology, 7, 499. DOI: 10.3389/fmicb.2016.00499

Fan, L., Reynolds, D., Liu, M., Stark, M., Kjelleberg, S., Webster, N. S., Thomas, T. (2012). Functional Equivalence and Evolutionary Convergence in Complex Communities of Microbial Sponge Symbionts. Proceedings of the National Academy of Sciences, 109(27), 1878–1887. http://doi.org/10.1073/pnas.1203287109

Fieseler, L., Horn, M., Wagner, M., Hentschel, U. (2004). Discovery of the Novel Candidate Phylum Poribacteria in Marine Sponges Discovery of the Novel Candidate Phylum “Poribacteria” in Marine Sponges. Applied and Environmental Microbiology, 70(6), 3724–3732. DOI:10.1128/AEM.70.6.3724

Fieseler, L., Quaiser, A., Schleper, C., Hentschel, U. (2006). Analysis of the first genome fragment from the marine sponge-associated, novel candidate phylum Poribacteria by environmental genomics. Environmental Microbiology, 8(4), 612–624. DOI: 10.1111/j.1462-2920.2005.00937

Fieth, R.A., Gauthier, M.A., Bayes, J., Green, K.M., Degnan, S.M. (2016). Ontogenetic Changes in the Bacterial Symbiont Community of the Tropical Demosponge *Amphimedon queenslandica*: Metamorphosis Is a New Beginning, Frontiers in Marine Science, 3, 228. https://doi.org/10.3389/fmars.2016.00228

Fiore, C.L., Jarett, J.K., Lesser, M.P. (2013). Symbiotic prokaryotic communities from different populations of the giant barrel sponge, *Xestospongia muta*, Microbiology Open, 2(6), 938 - 952. DOI: 10.1002/mbo3.135

Fischbach, M.A., Walsh, C.T., Clardy, J. (2008). The evolution of gene collectives: How natural selection drives chemical innovation, Proceedings of the National Academy of Sciences of the United States of America, 105(12), 4601 – 4608. DOI: 10.1073/pnas.0709132105

Flemer, B., Kennedy, J., Margassery, L.M., Morrissey, J.P., O’Gara, F., Dobson, A.D. (2011). Diversity and antimicrobial activities of microbes from two Irish marine sponges, *Suberites carnosus* and *Leucosolenia* sp. Journal of Applied Microbiology, 112(2), 289–301. DOI: 10.1111/j.1365-2672.2011.05211.x

Flissi, A., Dufresne, Y., Michalik, J., Tonon, L., Janot, S., Noe, L., Jacques, P., Leclere, V., Pupin, M. (2016). Norine, the knowledgebase dedicated to non-ribosomal peptides, is now open to crowdsourcing, Nucleic Acids Research, 44, 1113 – 1118. DOI: 10.1093/nar/gkv1143

Florez, L.V., Biederman, P.H.W., Engl, T., Kaltenpoth, M. (2015). Defensive symbioses of animals with prokaryotic and eukaryotic microorganisms, Natural Products Reports, 32(7), 904 – 936. DOI: 10.1039/C5NP00010F

Ford, J., Capon, R.J. (2000). Discorhabdin R: A New Antibacterial Pyrroloiminoquinone from Two Latrunculiid Marine Sponges, *Latrunculia* sp. and *Negombata* sp., Journal of Natural Prodducts, 63(11), 1527 – 1528. DOI: 10.1021/np000220q

Freeman, C. J., Thacker, R. W., Baker, D. M., Fogel, M. L. (2013). Quality or quantity: is nutrient transfer driven more by symbiont identity and productivity than by symbiont abundance? The International Society for Microbial Ecology Journal, 7(6), 1116-1125, http:// DOI: 10.1038/ismej.2013.7.

Fuerst, J.A. (2017). Chapter 1: Planctomycetes – New models for microbial cells and activities In Microbial Resources: From Functional Existence in Nature to Applications, eds. Kurtboke, I. Academic Press, London.

Furrow, F.B., Amsler, C.D., McClintock, J.B., Baker, B.J. (2003). Surface sequestration of chemical feeding deterrents in the Antarctic sponge *Latrunculia apicalis* as an optimal defense against sea star spongivory, Marine Biology, 143(3), 443 – 449. DOI: 10.1007/s00227-003-1109-5

Galeano, E., Thomas, O.P., Robledo, S., Munoz, D., Martinez, A. (2011). Antiparasitic Bromotyrosine Derivatives from the Marine Sponge *Verongula rigida*, Marine Drugs, 9(10), 1902 – 1913. DOI: 10.3390/md9101902

Gantt, S.E., Lopez-Legentil, S., Erwin, P.M. (2017). Stable microbial communities in the sponge *Crambe crambe* from inside and outside a polluted Mediterranean harbour, FEMS Microbiology Letters, 364(11), https://doi.org/10.1093/femsle/fnx105

Gao, F. (2018). Recent developments of software and database in microbial genomics and functional genomics, Briefings in Bioinformatics, bby013. DOI: 10.1093/bib/bby013

Gao, L., Pak, M., Kajihara, K., Brown, E.J. (2006). A Mycobacterial Operon Essential for Virulence In Vivo and Invasion and Intracellular Persistence in Macrophages, Infection and Immunity, 74(3), 1757 – 1767. DOI: 10.1128/IAI.74.3.1757-1767.2006

Garate, L., Sureda, J., Agell, G., Uriz, M.J. (2017). Endosymbiotic calcifying bacteria across sponge species and oceans, Scientific Report, 7, 43674. DOI: 10.1038/srep43674

Garderes, J., Bedoux, G., Koutsouveli, V., Crequer, S., Desriac, F., Le Pennec, G. (2015). Lipopolysaccharides from Commensal and Opportunistic Bacteria: Characterization and Response of the Immune System of the Host Sponge *Suberites domuncula*, Marine Drugs, 13(8), 4985 – 5006. DOI: 10.3390/md13084985

Garrido-Cardenas, J.A., Garcia-Maroto, F., Alvarez-Bermejo, J.A., Manzano-Agugliaro, F. (2017). DNA Sequencing Sensors: An Overview, 17(3), 588. DOI: 10.3390/s17030588

Gassier, S., Trefzer, A., Stockert, S., Kirschning, A., Bechthold, A. (1997). Cloning of an avilamycin biosynthetic gene cluster from *Streptomyces viridochromogenes* Tü57, Journal of Bacteriology, 179(20), 6271 – 6278. DOI: 10.1128/jb.179.20.6271-6278.1997

Gauthier, M.A., Watson, J.R., Degnan, S.M. (2016). Draft Genomes Shed Light on the Dual Bacterial Symbiosis that Dominates the Microbiome of the Coral Reef Sponge *Amphimedon queenslandica*, Frontiers in Marine Science. DOI: 10.3389/fmars.2016.00196

Gerwick, W.H., Moore, B.S. (2012). Lessons from the past and charting the future of marine natural products drug discovery and chemical biology, Chemistry and Biology, 19(1), 85 – 98. https://doi.org/10.1016/j.chembiol.2011.12.014

Giessen, T., Franke, K. B., Knappe, T. A., Kraas, F. I., Bosello, M., Xie, X., Linne, U., Marahiel, M. A. (2012). Isolation, Structure Elucidation, and Biosynthesis of an Unusual Hydroxamic Acid Ester-Containing Siderophore from *Actinosynnema mirum*. Journal of Natural Products, 75, 905–914. DOI: 10.1021/np300046k

Giovannoni, S.J., Tripp, H.J., Givean, S., Podar, M., Vergin, K.L., Baptista, D., Bibbs, L., Eads, J., Richardson, T.H., Noordewier, Rappe, M.S., Short, J.M., Carrington, J.C., Mathur, E.J. (2005). Genome Streamlining in a Cosmopolitan Oceanic Bacterium, Science, 309(5738), 1242 – 1245. DOI: 10.1126/science.1114057

Glaser, K.B., Mayer, A.M.S. (2009). A renaissance in marine pharmacology: From preclinical curiosity to clinical reality, Biochemical pharmacology, 78(5), 440 - 448. DOI: 10.1016/j.bcp.2009.04.015

Glockner, F.O., Kube, M., Bauer, M., Teeling, H., Lombardot, T., Ludwig, W., Gade, D., Beck, A., Borzym, K., Heitmann, K., Rabus, R., Schlesner, H., Amann, R., Reinhardt, R. (2003). Complete

genome sequence of the marine planctomycete *Pirellula* sp. strain 1, Proceedings of the National Academy of Sciences of the United States of America, 100(14), 8298 – 8303. DOI: 10.1073/pnas.1431443100

Golyshina, O.V., Kublanov, I.V., Tran H., Korzhenkov, A.A., Lunsdorf, H., Nechitaylo, T.Y., Gavrilov, S.N., Toshchakov, S.V., Golyshin, P.N. (2016). Biology of archaea from a novel family Cuniculiplasmataceae (Thermoplasmata) ubiquitous in hyperacidic environments, Scientific Reports, 6, 39034. DOI: 10.1038/srep39034

Gomes, N.G.M., Dasari, R., Chandra, S., Kiss, R., Kornienko, A. (2016). Marine Invertebrate Metabolites with Anticancer Activities: Solutions to the "Supply Problem", Marine Drugs, 14(5), 98. DOI: 10.3390/md14050098

Gonzalez-Escalona, N., Timme, R., Raphael, B.H., Zink, D., Sharma, S.K. (2014). Whole-Genome Single-Nucleotide-Polymorphism Analysis for Discrimination of *Clostridium botulinum* Group I Strains, Applied and Enviromental Microbiology, 80(7), 2125 - 2132. DOI: 10.1128/AEM.03934-13

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI, Nucleic Acids Research, W695-9. DOI: 10.1093/nar/gkq313

Graca, A.P., Bondoso, J., Gaspar, H., Xavier, J.R., Monteiro, M.C., de la Cruz, M., Oves-Costales, D., Vicente, F., Lage, O.M. (2013). Antimicrobial Activity of Heterotrophic Bacterial Communities from the Marine Sponge *Erylus discophorus* (Astrophorida, Geodiidae). PLoS One, 8, e78992. DOI: 10.1371/journal.pone.0078992

Gross, H., Stockwell, V. O., Henkels, M. D., Nowak-Thompson, B., Loper, J. E., Gerwick, W. H. (2007). The Genomisotopic Approach: A Systematic Method to Isolate Products of Orphan Biosynthetic Gene Clusters, Chemistry and Biology, 14(1), 53–63. DOI: 10.1016/j.chembiol.2006.11.007

Grove, A. (2013). MarR transcriptional family regulators, Current Biology, 23(4), 142 – 143. DOI: 10.1016/j.cub.2013.01.013

Gu, L., Geders, T.W., Wang, B., Gerwick, W.H., Hakansson, K., Smith, J.L., Sherman, D.H. (2007). GNAT-like strategy for polyketide chain initiation, Science, 9(318), 970 – 974. DOI: 10.1126/science.1148790

Gunasekera, S.P., McCarthy, P.J., Longley, R.E., Pomponi, S.A., Wright, A.E. (1999). Secobatzellines A and B, Two New Enzyme Inhibitors from a Deep-Water Caribbean Sponge of the Genus *Batzella*, Journal of Natural Products, 62(8), 1208 – 1211. DOI: 10.1021/np990177a

Gunasekera, S.P., Zuleta, I.A., Longley, R.E., Wright, A.E., Pomponi, S.A. (2003). Discorhabdins S, T, and U, New Cytotoxic Pyrroloiminoquinones from a Deep-Water Caribbean Sponge of the Genus *Batzella*, Journal of Natural Products, 66(12), 1615 – 1617. DOI: 10.1021/np030292s

Guo, J., Cole, J.R., Zhang, Q., Brown, C.T., Tiedje, J.M. (2015). Microbial Community Analysis with Ribosomal Gene Fragments from Shotgun Metagenomes, Applied and Environmental Mircobiology, 82(1), 157 – 166. DOI: 10.1128/AEM.02772-15

Gupta, P., Sharma, U., Schulz, T. C., McLean, A. B., Robins, A. J., West, L. M. (2012). Bicyclic C21 terpenoids from the marine sponge *Clathria compressa*, Journal of Natural Products, 75(6), 1223–1227. DOI: 10.1021/np300265p

Gurgui C., Piel J. (2010). Metagenomic Approaches to Identify and Isolate Bioactive Natural Products from Microbiota of Marine Sponges. In: Streit W., Daniel R. (eds) Metagenomics. Methods in Molecular Biology (Methods and Protocols), vol 668. Humana Press, Totowa, NJ

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms, Microbiology and Molecular Biology Reviews, 68(4), 669 – 685. DOI: 10.1128/MMBR.68.4.669-685.2004

Hania, B.W., Joseph, M., Schumann, P., Bunk, B., Fiebig, A., Sproer, C., Klenk, H.P., Fardeau, M.L., Spring, S. (2015). Complete genome sequence and description of *Salinispira pacifica* gen. nov., sp.

nov., a novel spirochaete isolated form a hypersaline microbial mat, Standards in Genomic Science, 10(7). DOI: 10.1186/1944-3277-10-7

Hardoim, C. C. P., Cardinale, M., Cacio, A. C. B., Esteves, A. I. S., Berg, G., Xavier, J. R., Cox, C.J., Costa, R. (2014). Effects of sample handling and cultivation bias on the specificity of bacterial communities in keratose marine sponges. Frontiers in Microbiology, 1–15. DOI:10.3389/fmicb.2014.00611

Hardoim, C. C. P., Esteves, A. I. S., Pires, F. R., Gonçalves, J. M. S., Cox, C. J., Xavier, J. R., & Costa, R. (2012). Phylogenetically and Spatially Close Marine Sponges Harbour Divergent Bacterial Communities. PLoS ONE, 7(12). DOI: 10.1371/journal.pone.0053029

Harvey, A.L. (2008). Natural Products in Drug Discovery, Drug Discovery Today, 13(20), 894 – 901. DOI: 10.1016/j.drudis.2008.07.004

He, H., Ding, W. D., Bernan, V. S., Richardson, a. D., Ireland, C. M., Greenstein, M., Ellestad, G.A., Carter, G. T. (2001). Lomaiviticins A and B, potent antitumor antibiotics from *Micromonospora lomaivitiensis*, Journal of the American Chemical Society, 123(22), 5362–5363. DOI: 10.1021/ja010129o

Hirata, Y., Uemura, D. (1986). Halichondrins - antitumor polyether macrolides from a marine sponge. Pure and Applied Chemistry, 58(5), 701–710. DOI:10.1351/pac198658050701

Hooper, G.J., Davies-Coleman, M.T., Kelly-Borges, M., Coetzee, P.S. (1996). New Alkaloids from a South African Latrunculid Sponge, Tetrahedron Letters, 37(39), 7135 – 7138. DOI: 10.1016/0040-4039(96)01560-2

Horino, T., Shimamura, Y., Ogata, K., Inoue, K., Erada, Y. (2016). *Kocuria kristinae* septic arthritis associated with infectious endocarditis in a hemodialysis patient with diabetes mellitus: a case report and literature review, Renal Replacement Therapy, 2(32). DOI: 10.1186/s41100-016-0041-3

Horn, H., Slaby, B.M., Jahn, M.T., Bayer, K., Moitinho-Silva, L., Forster, F., Abdelmohsen, U.R., Hentschel, U. (2016). An Enrichment of CRISPR and Other Defense-Related Features in Marine Sponge-Associated Microbial Metagenomes, Frontiers in Microbiology, 7, 1751. DOI: 10.3389/fmicb.2016.01751

Hu, Y., Chen, J., Hu, G., Yu, J., Zhu, X., Lin, Y., Chen, S., Yuan, J. (2015). Statistical Research on the Bioactivity of New Marine Natural Products Discovered during the 28 Years from 1985 to 2012. Marine Drugs, 13(1), 202–221. DOI:10.3390/md13010202

Huang, H., Hou, L., Li, H., Ju, J., Li, W. (2016). Activation of a plasmid-situated type III PKS gene cluster by deletion of a wbl gene in deepsea-derived *Streptomyces somaliensis* SCSIO ZH66, Microbial Cell Factories, 15, 116. DOI: 10.1186/s12934-016-0515-6

Huang, T., Lin, S. (2017). Microbial Natural Products: A Promising Source for Drug Discovery, Journal of Applied Microbiology and Biochemistry, 1, 2:5.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hernsdorf, A.W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D.A., Finstad, K.M., Amundson, R., Thomas, B.C., Banfield, J.F. (2016). A new view on the tree of life, Nature Microbiology, 1, 16048. DOI: 10.1038/nmicrobiol.2016.48

Hughes, C. C., Prieto-Davo, A., Jensen, P. R., Fenical, W. (2008). The marinopyrroles, antibiotics of an unprecedented structure class from a marine *Streptomyces* sp., Organic Letters, 10(4), 629–631. http://doi.org/10.1021/ol702952n

Hughes, C.C., Fenical, W. (2010). Total synthesis of ammosamides, Journal of the American Chemical Society, 132(8), 2528 – 2529. DOI: 10.1021/ja9106572

Hughes, C.C., MacMillan, J.B., Gaudencio, S.P., Fenical, W., La Clair, J.J. (2009a) Ammosamides A and B target myosin, Angewandte Chemie, 48(4), 728 – 732. DOI: 10.1002/anie.200804107

Hughes, C.C., MacMillan, J.B., Gaudencio, S.P., Jensen, P.R., Fenical, W. (2009). The Ammosamides: Structures of Cell Cycle Modulators from a Marine-Derived *Streptomyces* Species, Angewandte Chemie International Edition, 48(4), 725. DOI: 10.1002/anie.200804890

Hwang, K., Kim, H.U., Charusanti, P., Palsson, B.O., Lee, S.Y. (2014). Systems biology and biotechnology of *Streptomyces* species for the production of secondary metabolites, Biotechnology Advances, 32, 255 – 268. DOI: 10.1016/j.biotechadv.2013.10.008

Inoue, K., Ishikawa, Y., Nishiyama, S. (2010) Synthesis of Tetrahydropyrroloiminoquinone Alkaloids Based on Electrochemically Generated Hypervalent Iodine Oxidative Cyclization, Organic Letters, 12(3), 436 – 439. DOI: 10.1021/ol902566p

Isaacs, L.T., Kan, J., Nguyen, L., Videau, P., Anderson, M.A., Wright, T.L., Hill, R.T. (2009). Comparison of the Bacterial Communities of Wild and Captive Sponge *Clathria prolifera* from the Chesapeake Bay, Marine Biotechnology, 11, 758 - 770. DOI: 0.1007/s10126-009-9192-3"

Ishibashi M, Iwasaki T, Imai S, Sakamoto S, Yamaguchi K, Ito, A. (2001). Laboratory culture of the myxomycetes: Formation of fruiting bodies of *Didymium bahiense* and its plasmodial production of makaluvamine A. Journal of Natural Products, 64(1), 108-110. DOI: 10.1021/np000382m

Ishibashi, M. (2005). Isolation of Bioactive Natural Products from Myxomycetes, Medicinal Chemistry, 1(6), 575 – 590. DOI: 10.2174/157340605774598135

Ito, H., Kurokawa, H., Hirayama, A., Indo, H.P., Majima, H.J., Matsui, H. (2016). Cancer cell-specific mitochondrial reactive oxygen species promote non-heme iron uptake and enhance the proliferation of gastric epithelial cancer cell, Journal of Clinical Biochemistry and Nutrition, 61(3), 183 – 188. DOI: 10.3164/jcbn.17-8

Jahn, M.T., Markert, S.M., Ryu, T., Ravasi, T., Stigloher, C., Hentschel, U., Moitinho-Silva, L. (2016). Shedding light on cell compartmentation in the candidate phylum Poribacteria by high resolution visualisation and transcriptional profiling, Scientific Reports, 6, DOI: 10.1038/srep35860

Janso, J. E., Haltli, B. A., Eustáquio, A. S., Kulowski, K., Waldman, A. J., Zha, L. Nakamura, H., Bernan, V.S., He, H., Carter, G.T., Koehn, F.E., Balskus, E. P. (2014). Discovery of the lomaiviticin biosynthetic gene cluster in *Salinispora pacifica*, Tetrahedron, 70(27-28), 4156–4164. DOI: 10.1016/j.tet.2014.03.009

Jensen, P. R., Williams, P. G., Oh, D. C., Zeigler, L., Fenical, W. (2007). Species-specific secondary metabolite production in marine actinomycetes of the genus *Salinispora*. Applied and Environmental Microbiology, 73(4), 1146–1152. DOI:10.1128/AEM.01891-06

Jensen, P.R. (2016). Natural Products and the Gene Cluster Revolution, Trends in Microbiology, 24(12), 968 - 977. DOI: 10.1016/j.tim.2016.07.006

Jensen, S., Fortunato, S.A.V., Hoffmann, F., Hoem, S., Rapp, H.T., Ovreas, L., Torsvik, V.L. (2017). The Relative Abundance and Transcriptional Activity of Marine Sponge-Associated Microorganisms Emphasizing Groups Involved in Sulfur Cycle, Microbial Ecology, 73(3), 668 – 676. DOI:10.1007/s00248-016-0836-3

Jeon, J., Na, Z., Jung M., Lee, H., Sim, C.J., Nahm, K., Oh, K., Shin, J. (2010). Discorhabdins from the Korean Marine Sponge *Sceptrella* sp., Journal of Natural Products, 73(2), 258 – 262. DOI: 10.1021/np9005629

Jiao, W. H., Xu, T. T., Yu, H. B., Chen, G. D., Huang, X. J., Yang, F., Li, Y.S., Han, B.n>, Liu, X.Y., Lin, H. W. (2014). Dysideanones A-C, unusual sesquiterpene quinones from the South China Sea sponge *Dysidea avara*. Journal of Natural Products, 77(2), 346–350. DOI:10.1021/np4009392

Jones, C. (2005). Revised structures for the capsular polysaccharides from *Staphylococcus aureus* Types 5 and 8, components of novel glycoconjugate vaccines, Carbohydrate Research, 340(6), 1097 – 1106. DOI: 10.1016/j.carres.2005.02.001

Jordan, P.A., Moore, B.S. (2016). Biosynthetic Pathway Connects Cryptic Ribosomally Synthesized Posttranslationally Modified Peptide Genes with Pyrroloquinoline Alkaloids, Cell Chemical Biology, 23, 1 – 11. DOI: 10.1016/j.chembiol.2016.10.009

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., Perry, T., Kao, D., Mason, A.L., Madsen, K.L., Wong, G.K.S. (2016). Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics, Frontiers in Microbiology, 7, 459. DOI: 10.3389/fmicb.2016.00459

Kamke, J., Rinke, C., Schwientek, P., Mavromatis, K., Ivanova, N., Sczyrba, A., Woyke, T., Hentschel, U. (2014). The candidate phylum Poribacteria by single-cell genomics: New insights into phylogeny, cell-compartmentation, eukaryote-like repeat proteins, and other genomic features. PLoS ONE, 9(1). DOI: 10.1371/journal.pone.0087353

Kamke, J., Sczyrba, A., Ivanova, N., Schwientek, P., Rinke, C., Mavromatis, K., Woyke, T., Hentschel, U. (2013). Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges, International Society for Microbial Ecology, 7, 2287 – 2300. DOI: 10.1038/ismej.2013.111

Kant, R., van Passel, M.W., Sangwan, P., Palva, A., Lucas, S., Copeland, A., Lapidus, A., Glavina-del Rio, T., Dalin, E., Tice, H., Bruce, D., Goodwin, L., Pitluck, S., Chertkov, O., Larimer, F.W., Land, M.L., Hauser, L., Brettin, T.S., Detter, J.C., Han, S., de Vos, W.M., Janssen, P.H., Smidt, H. (2011). Genome sequence of "Pedosphaera parvula" Ellin514, an aerobic Verrucomicrobial isolate from pasture soil, Journal of Bacteriology, 193(11), 2900 – 2901. DOI: 10.1128/JB.00299-11

Kaye, H. R. (1991). Sexual reproduction in 4 Caribbean commercial sponges .2. Oogenesis and transfer of bacterial symbionts. Invertebrate Reproduction and Development 19(1), 13-24. DOI: 10.1080/07924259.1991.9672152

Kelly, M., Sim-Smith, C., Stone, R., Samaai, T., Reiswig, H., Austin, W. (2016). New taxa and arrangements within the family Latrunculiidae (Demospongiae, Poecilosclerida), Zootaxa, 4121(1), 1 – 48. DOI: 10.11646/zootaxa.4121.1.1

Kennedy, J., Codling, C.E., Jones, B.V., Dobson, A.D., 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. DOI: 10.1111/j.1462-2920.2008.01614.x

Kersten, R.D., Lane, A.L., Nett, M., Richter, T.K.S., Duggan, B.M., Dorrestein, P.C., Moore, B.S. (2013). Bioactivity-Guided Genome Mining Reveals the Lomaiviticin Biosynthetic Gene Cluster in *Salinispora tropica*, Chembiochem, 14(8), 955 – 962. DOI: 10.1002/cbic.201300147

Keyzer, R.A., Arendse, C.E., Hendricks, D.T., Samaai, T., Davies-Coleman, M.T. (2005). Makaluvic Acids from the South African Latrunculid Sponge *Strongylodesma aliwaliensis*, Journal of Natural Products, 68(4), 506 – 510. DOI: 10.1021/np049589w

Keyzer, R.A., Samaai, T., Davies-Coleman, M.T. (2004). Novel pyrroloquinoline ribosides from the South African latrunculid sponge *Strongylodesma aliwaliensis*, Tetrahedron Letters, 45(51), 9415 – 9418. DOI: 10.1016/j.tetlet.2004.10.106

Kim, J.K., Park, H.Y., Lee, B.L. (2016). The symbiotic role of O-antigen of *Burkholderia* symbiont in association with host *Riptortus pedestris*, Development and Comparative Immunology, 60, 202 – 208. DOI: 10.1016/j.dci.2016.02.009

Kim, J.W., Brawley, S.H., Prochnik, S., Chovatia, M., Grimwood, J., Jenkins, J., La Butti, K., Mavromatis, K., Nolan, M., Zane, M., Schmutz, J., Stiller, J.W., Grossman, A.R. (2016) Genome Analysis of Planctomycetes Inhabiting Blades of the Red Alga *Porphyra umbilicalis*, PLoS ONE, 11(3), e0151883. DOI: 10.1371/journal.pone.0151883

Kim, T.K., Hewavitharana, A.K., Shaw, N., Fuerst, J.A. (2006) Discovery of a New Source of Rifamycin Antibiotics in Marine Sponge Actinobacteria by Phylogenetic Prediction, Applied and Environmental Microbiology, 72(3), 2118 – 2125. DOI: 10.1128/AEM.72.3.2118-2125.2006

Kimura, M., Kawai, E., Yaoita, H., Ichinoi, N., Sakamoto, O., Kure, S. (2017). Central Venous Catheter-related Bloodstream Infection with *Kocuria kristinae* in a Patient with Propionic Acidemia, Infectious Diseases, vol. 2017, Article ID 1254175. DOI:10.1155/2017/1254175

Klein, A.M., Bohannan, B.J.M., Jaffe, D.A., Levin, D.A., Green, J.L., Molecular Evidence for Metabolically Active Bacteria in the Atmosphere, Frontiers in Microbiology, 7, 772. DOI: 10.3389/fmicb.2016.00772

Kohira, N., West, J., Ito, A., Ito-Horiyama, T., Nakamura, R., Sato, T., Rittenhouse, S., Tsuji, M., Yamano, Y. (2015). In Vitro Antimicrobial Activity of a Siderophore Cephalosporin, S-649266, against

Enterobacteriaceae Clinical Isolates, Including Carbapenem-Resistant Strains, Antimicrobial Agents and Chemotherapy, 60(2), 729 – 734. DOI: 10.1128/AAC.01695-15

Kuhn, M.L., Alexander, E., Minasov, G., Page, H.J. Warwrzak, Z., Shuvalova, L., Flores, K.J., Wilson, D.J., Shi, C., Aldrich, C.C., Anderson, W.F. (2016). Structure of the Essential Mtb FadD32 Enzyme: A Promising Drug Target for Treating Tuberculosis, ACS Infectious Diseases, 2(8), 579 – 591. DOI: 10.1021/acsinfecdis.6b00082

Kumar, C.G., Sujitha, P. (2014). Kocuran, an exopolysaccharide isolated from *Kocuria rosea* strain BS-1 and evaluation of its in vitro immunosuppression activities, Enzyme Microb. Technol., 5(55), 113 – 120, DOI: 10.1016/j.enzmictec.2013.10.007

Kumar, S., Mukherjee, M.M., Varela, M.F. (2013). Modulation of Bacterial Multidrug Resistance Efflux Pumps of the Major Facilitator Superfamily, International Journal of Bacteriology, Article ID: 204141, DOI: 10.1155/2013/204141

Kumar, S., Stecher, G., Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets, Molecular Biology and Evolution, 33(7), 1870 – 1874. DOI: 10.1093/molbev/msw054

Labunskyy, V.M., Hatfield, D.L., Gladyshev, V.N. (2014). Selenoproteins: Molecular Pathways and Physiological Roles, Physiological Reviews, 94(3), 739 - 777. DOI: 10.1152/physrev.00039.2013

Lackner, G., Peters, E.E., Helfrich, E.J.N., Piel, J. (2016). Insights into the lifestyle of uncultured bacterial natural product factories associated with marine sponges, Proceeding of the National Academy of Sciences of the United States of America, 114(3), 347 – 356. DOI: 10.1073/pnas.1616234114

Lam, K. S. (2006). Discovery of novel metabolites from marine actinomycetes. Current Opinion in Microbiology, 9(3), 245–251. DOI: 10.1016/j.mib.2006.03.004

Lang, G., Pinkert, A., Blunt, J.W., Munro, M.H.G. (2005). Discorhabdin W, the First Dimeric Discorhabdin, Journal of Natural Products, 68(12), 1796 – 1798. DOI: 10.1021/np050333f

Laport, M.S., Santos-Gandelman, J.F., Muricy, G., Giambiagi-deMarval, M., George, I. (2016). Antagonistic Interactions among Bacteria Isolated from either the Same or from Different Sponges Native to the Brazilian Coast, Marine Science: Research and Development, 6, 2. DOI: 0.4172/2155-9910.1000185

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G. (2007). ClustalW and ClustalX version 2, Bioinformatics, 23(21), 2947 – 2948. DOI: 10.1093/bioinformatics/btm404

Lau, M.C.Y., Cameron, C., Magnabosco, C., Brown, C.T., Schilkey, F., Grim, S., Hendrickson, S., Pullin, M., Sherwood-Lollar, B., van Heerden, E., Kieft, T.L., Onstott, T.C. (2014). Phylogeny and phylogeography of functional genes shared among seven terrestrial subsurface metagenomes reveal N-cycling and microbial evolutionary relationships, Frntiers in Microbiology, 5, 531. DOI: 10.3389/fmicb.2014.00531

Laverty, G., Gorman, S.P., Gilmore, B.F. (2013). Biomolecular mechanisms of staphylococcal biofilm formation, Future Microbiology, 8(4), 509 -524. DOI: 10.2217/fmb.13.7

Le, D.T.L., Tran, T., Duviau, M., Meyrand, M., Guerardel, Y., Castelain, M., Loubiere, P., Chapot-Charter, M., Dague, E., Mercier-Bonin, M. (2013). Unraveling the Role of Surface Mucus-Binding Protein and Pili in Muco-Adhesion of *Lactococcus lactis*, PloS ONE, 8(11), 79850. DOI: 10.1371/journal.pone.0079850

Lee, W.W.; Benitez, A.; Goodman, L.; Baker, B.R. (1960). Potential anticancer agents. 1 xl. synthesis of the-anomer of 9-(d-arabinofuranosyl)-adenine. Journal of the American Chemical Society. 82, 2648–2649.

Legentil, L., Benel, L., Bertrand, V., Lesur, B., Delfourne, E. (2006). Synthesis and Antitumor Characterization of Pyrazolic Analogues of the Marine Pyrroloquinoline Alkaloids: Wakayin and Tsitsikammamines, Journal of Medicinal Chemistry, 49, 2979 – 2988. DOI: 10.1021/jm051247f

Letzel, A., Li, J., Amos, G.C.A., Millan-Aguinaga, N., Ginigini, J., Abdelmohsen, U.R., Guadencio, S.P., Ziemert, N., Moore, B.S., Jensen, P.R. (2017). Genomic insights into specialized metabolism in the marine actinomycete *Salinispora*, Environemntal Microbiology, 19(9), 3660 - 3673. DOI: 10.1111/1462-2920.13867

Li, R. (2015). Marinopyrroles: Unique Drug Discoveries Based on Marine Natural Products, Medicinal Research Reviews, 36(1), 169 - 189. DOI: 10.1002/med.21359

Li, W., Silipo, A., Gersby, L.B.A., Newman, M., Molinaro, A., Yu, B. (2017). Synthesis of Bradyrhizose Oligosaccahrides Relevant to the *Bradyrhizobium* O-antigen, Angewandte Chemie, 129, 2124 – 2128. DOI: 10.1002/anie.201610680

Li, Y., Lei, X., Zhu, H., Zhang, H., Guan, C., Chen, Z., Zheng, W., Fu, L., Zheng, T. (2016b). Chitinase producing bacteria with direct algicidal activity on marine diatoms, Scientific Reports, 6, 21984. DOI: 10.1038/srep21984

Li, Z., Wang, Y., He, L., Zheng, H. (2014). Metabolic profiles of prokaryotic and eukaryotic communities in deep-sea sponge *Neamphius huxleyi* indicated by metagenomics, Scientific Reports, 4, 3895. DOI: 10.1038/srep03895

Li, Z., Wang, Y., Li, J., Liu, F., He, L., Wang, S. (2016a). Metagenomic Analysis of Genes Encoding Nutrient Cycling Pathways in the Microbiota of Deep-Sea and Shallow-Water Sponges, Marine Biotechnology, 18(6), 659 – 671. DOI: 10.1007/s10126-016-9725-5

Lin, S., McCauley, E.P., Lorig-Roach, N., Tenney, K., Naphen, C.N., Yang, A., Johnson, T.A., Hernadez, T., Rattan, R., Valeriote. F.A., Crews, P. (2017). Another Look at Pyrroloiminoquinone Alkaloids—Perspectives on Their Therapeutic Potential from Known Structures and Semisynthetic Analogues, Marine Drugs, 15, 98. DOI: 10.3390/md15040098

Linares-Otya, L., Linares-Otya, V., Armas-Mantilla, L., Blanco-Olano, C., Crusemann, M., Ganoza-Yupanqui, M.L., Campos-Florian, J., Konig, G.M., Schaberle, T.F. (2017). Identification and heterologous expression of the kocurin biosynthetic gene cluster, Microbiology. DOI: 10.1099/mic.0.000538

Lorenzo, F.D. (2017). The lipopolysaccharide lipid A structure from the marine sponge-associated bacterium *Pseudoalteromonas* sp. 2A, Antonie van Leeuwenhoek, 110(11), 1401 – 1412. DOI: 10.1007/s10482-017-0865-z

Maansson, M., Vuynne, N.G., Klitgaard, A., Nybo, J.L., Melchiorsen, J., Nguyen, D.D., Sanchez, L.M., Ziemert, N., Dorrestein, P.C., Andersen, M.R., Gram, L. (2016). An Integrated Metabolomic and Genomic Mining Workflow To Uncover the Biosynthetic Potential of Bacteria, American Society for Microbiology, mSystems 1(3), e00028-15. DOI:10.1128/mSystems.00028-15

Magnet, S., Courvalin, P., Lambert, T. (2001). Resistance-Nodulation-Cell Division-Type Efflux Pump Involved in Aminoglycoside Resistance in *Acinetobacter baumannii* Strain BM4454, Antimicrobial Agents and Chemotherapy, 45(12), 3375 – 3380. DOI: 10.1128/AAC.45.12.3375-3380.2001

Mahajan, G., Thomas, B., Parab, R., Patel, Z.E., Kuldharan, S., Yemparala, V., Mishra, P.D., Ranadive, P., D'Souza, L., Pari, K., Sivaramkrishnan, H. (2013). In Vitro and In Vivo Activities of Antibiotic PM181104, Antimicrobial Agents and Chemotherapy, 57(11), 5315 – 5319. DOI: 10.1128/AAC.01059-13

Marahiel, M.A., Stachelhaus, T., Mootz, H.D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis, Chemical Reviews, 97(7), 2651 – 2674. DOI: 10.1021/cr960029e

Margulis, L., Fester, R. (1991). Symbiosis as a Source of Evolutionary Innovation: Speciation and Morphogenesis, MIT Press: Boston.

Martin, J., Sousa, T., Crespo, G., Palomo, S., Gonzalez, I., Tormo, J., de la Cruz, M., Anderson, M., Hill, R. T., 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, 387–398. DOI: 10.3390/md11020387

Martin, M. J., Rodriguez-Acebes, R., Garcia-Ramos, Y., Martinez, V., Murcia, C., Digon, I., Marco, I., Pelay-Gimeno, M., Fernandez, R., Reyes, F., Francesch, A.M., Munt, S., Tulla-Puche, J., Albericio, F., Cuevas, C. (2014). Stellatolides, a new cyclodepsipeptide family from the sponge *Ecionemia acervus*: Isolation, solid-phase total synthesis, and full structural assignment of stellatolide a. Journal of the American Chemical Society, 136(18), 6754–6762. DOI:10.1021/ja502744a

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. DOI:10.3390/md12021066

Matcher, G.F., Waterworth, S.C., Walmsley, T.A., Matsatsa, T., Parker-Nance, S., Davies-Coleman, M.T., Dorrington, R.A. (2017). Keeping it in the family: Coevolution of latrunculid sponges and their dominant bacterial symbionts, Microbiology Open, 6(2), e00417. DOI: 10.1002/mbo3.417

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 Pharmocological Sciences, 31(6), 255 - 265. DOI: 10.1016/j.tips.2010.02.005

McCormack, G.P., Erpenbeck, D., Van Soest, R.W.M. (2002). Major discrepancy between phylogenetic hypotheses based on molecular and morphological criteria within the Order Haplosclerida (Phylum Porifera: Class Demospongiae), Journal of Zoological Systematics and Evolutionary Research, 40(4), 237 – 240. DOI: 10.1046/j.1439-0469.2002.00204.x

McCutcheon, J.P., Moran, N.A. (2011). Extreme genome reduction in symbiotic bacteria, Nature Reviews Microbiology, 10(1), 13 -26. DOI: 10.1038/nrmicro2670

McDonald, L. A., Capson, T. L., Krishnamurthy, G., Ding, W. D., Ellestad, G. A., Bernan, V. S., ... Ireland, C. M. (1996). Namenamicin, a new enediyne antitumor antibiotic from the marine ascidian *Polysyncraton lithostrotum*. Journal of the American Chemical Society, 118(44), 10898–10899. DOI:10.1021/ja961122n

McKenzie, J.L., Robson, J., Berney, M., Smith, T.C., Ruthe, A., Gardner, P.P., Arcus, V.L., Cook, G.M. (2012). A VapBC Toxin-Antitoxin Module Is a Posttranscriptional Regulator of Metabolic Flux in Mycobacteria, Journal of Bacteriology, 194(9), 2189 – 2204. DOI: 10.1128/JB.06790-11

Medema, M. H., Fischbach, M. A. (2015). Computational approaches to natural product discovery. Nature Chemical Biology, 11(9), 639–648. DOI: 10.1038/nchembio.1884

Mehbub, M., Lei, J., Franco, C., Zhang, W. (2014). Marine Sponge Derived Natural Products between 2001 and 2010: Trends and Opportunities for Discovery of Bioactives. Marine Drugs, 12(8), 4539–4577. DOI:10.3390/md12084539

Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.P., Goker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics, 14(60). DOI: 10.1186/1471-2105-14-60

Meli, A., Tedesco, C., Sala, G.D., Schettini, R., Albericio, F., De Riccardis, F., Izzo, I. (2017). Phakellistatins: An Underwater Unsolved Puzzle, Marine Drugs, 15(3), 78. DOI: 10.3390/md15030078

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R.A. (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes, BMC Bioinformatics, 9(386). DOI: 10.1186/1471-2105-9-386

Milanowski, D.J., Rashid, M.A., Gustafson, K.R., O'Keefe, B.R., Nawrocki, J.P., Pannell, L.K., Boyd, M.R. (2004). Cyclonellin, a New Cyclic Octapeptide from the Marine Sponge *Axinella carteri*, Journal of Natural Products, 67(3), 441 – 444. DOI: 10.1021/np030336x

Miller, I.J., Lopera, J.G., Montgomery, K., Rose, W., Kwan, J.C. (2016). A comparative metagenomics approach to marine natural product drug discovery in *Hippospongia lachne*, Planta Medica, 82. DOI: 10.1055/s-0036-1596166

Miller, I.J., Rees, E., Miller, I., Ross, J., Baxa, J., Kerby, R.L., Rey, F.E., Kwan, J.C. (2017). Fully-automated and scalable metagenomic binning through unsupervised clustering and decision-tree classification, SUBMITTED.

Mills, H. J., Reese, B. K., Shepard, A. K., Riedinger, N., Dowd, S. E., Morono, Y., & Inagaki, F. (2012). Characterization of Metabolically Active Bacterial Populations in Subseafloor Nankai Trough Sediments above, within, and below the Sulfate–Methane Transition Zone. Frontiers in Microbiology, 3(April), 1–12. http://doi.org/10.3389/fmicb.2012.00113

Mioso, R., Marante, F.J.T., Bezerra, R.S., Borges, F.V.P., Santos, B.V.O., de Laguna, I.H.B. (2017). Cytotoxic Compounds Derived from Marine Sponges. A Review (2010–2012), Molecules, 22(2), 208. DOI: 10.3390/molecules22020208

Mitchell, A.L., Scheremetjew, M., Denise, H., Potter, S., Tarkowska, A., Qureshi, M., Salazar, G.A., Pesseat, S., Boland, M.A., Hunter Petra ten Hoopen, F.M.I., Alako, B., Amid, C., Wilkinson, D.J., Curtis, T.P., Cochrane, G., Finn, R.D. (2017). EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies, Nucleic Acids Research, gkx967. DOI: 10.1093/nar/gkx967

Miyanaga, A., Janso, J.E., Mcdonald, L., He, M., Liu, H., Barbieri, Eustaìquio, A.S., Fielding, E.N., Carter, G.T., Jensen, P.R., Feng, X., Leighton, M., Koehn, F.E., Moore, B.S. (2011). Discovery and assembly-line biosynthesis of the lymphostin pyrroloquinoline alkaloid family of mTOR inhibitors in *Salinispora* bacteria. Journal of the American Chemical Society, 133(34), 13311-13313. DOI: 10.1021/ja205655w

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, http:// i:10.1111/j.1462-2920.2008.01704.x

Mohammed, R., Peng, J., Kelly, M., Hamann, M. T. (2006). Cyclic heptapeptides from the Jamaican sponge *Stylissa caribica*. Journal of Natural Products, 69(12), 1739–1744. DOI:10.1021/np060006n

Mohanty, D., Sankaranarayanan, R., Gokhale, R.S. (2011). Fatty acyl-AMP ligases and polyketide synthases are unique enzymes of lipid biosynthetic machinery in *Mycobacterium tuberculosis*, Tuberculosis, 91(5), 448 – 455. DOI: 10.1016/j.tube.2011.04.006

Moore, B.S., Hopke, J.N. (2001). Discovery of a New Bacterial Polyketide Biosynthetic Pathway, ChemBioChem, 2(1), 35 – 38. DOI: 10.1002/1439-7633(20010105)2:1<35::AID-CBIC35>3.0.CO;2-1

Morris, R.M., Rappe, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A., Giovannoni, S.J. (2002). SAR11 clade dominates ocean surface bacterioplankton communities, Nature, 420(6917), 806 – 810. DOI: 10.1038/nature01240

Murugesan, S., Nadkarni, D.H., Velu, S.E. (2009). A facile synthesis of bispyrroloquinone and bispyrroloiminoquinone ring system of marine alkaloids, Tetrahedron Letters, 50(25), 3074 – 3076. DOI: 10.1016/j.tetlet.2009.04.021

Mutsenko, V.V., Gryshkov, O., Lauterboeck, L., Rogulska, O., Tarusin, D.N., Bazhenov, V.V., Schutz, K., Bruggemeier, S., Gossla, E., Akkeineni, A.R., Meibner, H., Lode, A., Meschke, S., Fromont, J., Stelling, A.L., Tabachnik, K.R., Gelinsky, M., Nikulin, S., Rodin, S., Tonevitsky, A.G., Petrenko, A.Y., Glasmacher, B., Schupp, P.J., Ehrlich, H. (2017). Novel chitin scaffolds derived from marine sponge Ianthella basta for tissue engineering approaches based on human mesenchymal stromal cells: Biocompatibility and cryopreservation, International Journal of Biological Macromolecules, 104B, 1955 – 1965. DOI: 10.1016/j.ijbiomac.2017.03.161

Na, M., Ding, Y., Wang, B., Tekwani, B.L., Schinazi, R.F., Franzblau, S., Kelly, M., Stone, R., Li, X., Ferreira, D., Hamann, M.T. (2010). Anti-infective Discorhabdins from a Deep-Water Alaskan Sponge of the Genus *Latrunculia*, Journal of Natural Products, 73(3), 383 – 387. DOI: 10.1021/np900281r

Nagata, H., Ochiai, K., Aotani, Y., Ando, K., Yoshida, M., Takahashi, I., Tamaoki, T. (1997). Lymphostin (LK6-A), a Novel Immunosuppressant from *Streptomyces* sp. KY11783: Taxonomy of the Producing Organism, Fermentation, Isolation and Biological Activities, The Journal of Antibiotics, 50(7), 537 – 542.

Nagata, H., Yano, H., Sasaki, K., Sato, S., Nakanishi, S., Takahashi, I., Tamaoki, T. (2002). Inhibition of Lymphocyte Kinase Lck and Phosphatidylinositol 3-kinase by a Novel Immunosuppressant, Lymphostin, Bioscience, Biotechnology, and Biochemistry, 66(3), 501 – 507. DOI: 10.1271/bbb.66.501

Napolitano, A., Bruno, I., Riccio, R., Gomez-Paloma, L. (2005). Synthesis, structure, and biological aspects of cyclopeptides related to marine phakellistatins 7–9, Tetrahedron, 61(28), 6808 – 6815. DOI: 10.1016/j.tet.2005.04.067

Naughton, L.M., Romano, S., O'Gara, F., Dobson, A.D.W. (2017). Identification of Secondary Metabolite Gene Clusters in the Pseudovibrio Genus Reveals Encouraging Biosynthetic Potential toward the Production of Novel Bioactive Compounds, Frontiers in Microbiology, 8, 1494. DOI: 10.3389/fmicb.2017.01494

Neulinger, S.C., Stohr, R., Thiel, V., Schmaljohann, R., Imhoff, J.F. (2010). New phylogenetic lineages of the Spirochaetes phylum associated with *Clathrina* species (Porifera), The Journal of Microbiology, 48(4), 411 – 418. DOI: 10.1007/s12275-010-0017-x

Newman, D.J., Cragg, G.M. (2016). Natural Products as Sources of New Drugs from 1981 to 2014, Journal of Natural Products, 79(3), 629 – 661. DOI: 10.1021/acs.jnatprod.5b01055

Nguyen, M.T.H.D., Liu, M., Thomas, T. (2013). Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis, Molecular Ecology, 23(6) 1635 – 1645. DOI: 10.1111/mec.12384

Nicacio, K.J., Ioac, L.P., Froes, A.M., Leomil, L., Appolinario, L.R., Thompson, C.C., Thompson, F.L., Ferreira, A.G., Williams, D.E., Andersen, R.J., Eustaquio, A.S., Berlinck, R.G.S. (2017). Cultures of the Marine Bacterium *Pseudovibrio denitrificans* Ab134 Produce Bromotyrosine-Derived Alkaloids Previously Only Isolated from Marine Sponges, Journal of Natural Products, 80, 235 – 240. DOI: 10.1021/acs.jnatprod.6b00838

Noyer, C., Casamayor, E.O., Becerro, M.A. (2014). Environmental Heterogeneity and Microbial Inheritance Influence Sponge-Associated Bacterial Composition of *Spongia lamella*, Microbial Ecology, 68(3), 611 – 620. DOI: 10.1007/s00248-014-0428-z

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A., Korobeynikov, A., Lapidus, A., Prjibelsky, A., Pyshkin, A., Sirotkin, A., Sirotkin, Y., Stepanuakus, R., McLean, J., Lasken, R., Cligenpeel, S.R., Woyke, T., Tesler, G., Alekseyev, M.A., Pevzner, P.A. (2013). Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads, Research in Computational Molecular Biology in Annual International Conference on Research in Computational Molecular Biology, eds. M. Deng et al. pp.158 – 170. Springer-Verlag, Berlin.

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, 1495 – 1508. DOI: 10.1111/j.1365-2672.2011.05008.x

Ohkuma, M., Noda, S., Hattori, S., Iida, T., Yuki, M., Starns, D., Inoue, J., Darby, A.C., Hongoh, Y. (2015). Acetogenesis from H2 plus CO2 and nitrogen fixation by an endosymbiotic spirochete of a termite-gut cellulolytic protist, PNAS, 112(33), 10224 – 10230. DOI: 10.1073/pnas.1423979112

Offret, C., Desriac, F., Le Chevalier, P., Mounier, J., Jegou, C., Fleury, Y. (2016). Spotlight on Antimicrobial Metabolites from the Marine Bacteria *Pseudoalteromonas*: Chemodiversity and Ecological Significance, Marine Drugs, 14(7), 129. DOI: 10.3390/md14070129

O'Malley, M.A., Wideman, J.G., Ruiz-Trillo, I. (2016). Losing Complexity: The Role of Simplification in Macroevolution, Trends in Ecology and Evolution, 31(8), 608 - 621. DOI: 10.1016/j.tree.2016.04.004

Osbourn, A. (2010). Secondary metabolic gene clusters: Evolutionary toolkits for chemical innovation. Trends in Genetics, 26(10), 449–457. DOI: 10.1016/j.tig.2010.07.001

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F., Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST), Nucleic Acids Research, 42, 206 – 214. DOI: 10.1093/nar/gkt1226

Overmann, J., Abt, B., Sikorski, J. (2017). Present and Future of Culturing Bacteria, Anuual Review of Microbiology, 71, 711 - 730. DOI: 10.1146/annurev-micro-090816-093449

Owen, J.G., Reddy, B.V.B., Ternei, M.A., Charlop-Powers, Z., Calle, P.Y., Kim, J.H., Brady, S.F. (2012). Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products, Proceeding of the National Academy of Sciences of the United States of America, 110(29), 11797 – 11802. DOI: 10.1073/pnas.1222159110

Page, R., Peti, W. (2016). Toxin-antitoxin systems in bacterial growth arrest and persistence, Nature Chemical Biology, 12, 208 - 214. DOI: 10.1038/nchembio.2044

Palomo, S., González, I., de la Cruz, M., Martin, J., Tormo, J. R., Anderson, M., Hill, R. T., Vicente, F., Reyes, F., Genilloud, O. (2013). Sponge-Derived *Kocuria* and *Micrococcus* spp. as Sources of the New Thiazolyl Peptide Antibiotic Kocurin. Marine Drugs, 11, 1071–1086. DOI: .3390/md1104107116

Parada, A.E., Needham, D.M., Fuhrman, J.A. (2015). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples, Environmental Microbiology, 18, 1403 – 1414. DOI: 10.1111/1462-2920.13023

Parker-Nance, S., Waterworth, S., Hilliar, S., Walmsley, T., Dorrington, R.A. (2017) New species in the genus *Tsitsikamma* (Poecilosclerida, Latrunculiidae) from South Africa, Zootaxa. Submitted.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes, Genome Research, 25(7), 1043 – 1055. DOI: 10.1101/gr.186072.114

Patridge, E., Gareiss, P., Kinch, M.S., Hoyer, D. (2016). An analysis of FDA-approved drugs: Natural Products and their Derivatives, Drug Discovery Today, 21(2), 204 – 207. https://doi.org/10.1016/j.drudis.2015.01.009

Pence, H.E., Williams, A. (2010). ChemSpider: An Online Chemical Information Resource, Journal of Chemical Education, 87(11), 1123 – 1124. DOI: 10.1021/ed100697w

Peng, T., Lin, J., Xu, Y., Zhang, Y. (2016). Comparative genomics reveals new evolutionary and ecological patterns of selenium utilization in bacteria, The ISMA Journal, 10(8), 2048 – 2059. DOI: 10.1038/ismej.2015.246

Penn, K., Jenkins, C., Nett, M., Udwary, D.W., Gontang, E.A., McGlinchey, R.P., Foster, B., Lapidus, A., Podell, S., Allen, E.E., Moore, B.S., Jensen, P.R. (2009). Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria, The ISME Journal, 3, 1193 – 1203. DOI: 10.1038/ismej.2009.58

Peraud, O. (2006) Isolation and Characterization of a Sponge-Associated Actinomycete that Produces Manzamines. PhD. University of Maryland.

Perez, L.B., Fenical, W. (2017). Chapter 8: Accessing Marine Microbial Diversity for Drug Discovery, in Microbial Resources: From Functional Existence in Nature to Applications, eds. Kurtboke, I. pp. 169 – 189. Academic Press: London.

Pérez-Miranda, S., Cabirol, N., George-Téllez, R., Zamudio-Rivera, L. S., Fernández, F. J. (2007). O-CAS, A Fast and Universal Method for Siderophore Detection, Journal of Microbiological Methods, 70, 127–131. DOI: 10.1016/j.mimet.2007.03.023

Perry, N.B., Blunt, J.W., McCombs, J.D., Munro, M.H.G. (1986). Discorhabdin C, a highly cytotoxic pigment from a sponge of the genus *Latrunculia*, The Journal of Organic Chemistry, 51(26), 5476 – 5478. DOI: 10.1021/jo00376a096

Petrenko, I., Bazhenov, V.V., Galli, R., Wysokowski, M., Fromont, J., Schupp, P.J., Stelling, A.J., Niederschlag, E., Stoker, H., Kutsova, V.Z., Jesionowski, T., Ehrlich, H. (2017). Chitin of poriferan origin and the bioelectrometallurgy of copper/copper oxide, International Journal of Biological Macromolecules, 104B, 1626 – 1632. DOI: 10.1016/j.ijbiomac.2017.01.084

Pettit, G.R., Inoue, M., Kamano, Y., Herald, D.L., Arm, C., Dufresne, C., Christie, N.D., Schmidt, J.M., Doubek, D.L., Krupa, T.S. (1987). Isolation and Structure of Powerful Cell Growth Inhibitor

Cephalostatin 1, Journal of the American Chemical Society, 110(6)), 2006 - 2007. DOI: 10.1021/ja00214a078

Pettit, G.R., Xu, J., Dorsaz, A., Williams, M.D., Boyd, M.R., Cerny, R.L. (1995). Isolation and structure of the human cancer cell growth inhibitory cyclic decapeptides phakellistatins 7, 8 and 9, Bioorganic and Medicinal Chemsitry Letters, 5(13), 1339 – 1344. DOI: 10.1016/0960-894X(95)00219-J

Piddock, L.J.V. (2006). Multidrug-resistance efflux pumps? Not just for resistance, Nature Reviews Microbiology, 4, 629 - 636. DOI: 10.1038/nrmicro1464

Piel, J. (2006). Bacterial symbionts: prospects for the sustainable production of invertebrate-derived pharmaceuticals. Current Medicinal Chemistry, 13(1), 39–50. DOI:10.2174/092986706775197944

Pita, L., Turon, X., López-Legentil, S., Erwin, P. M. (2013). Host rules: Spatial stability of bacterial communities associated with marine sponges (*Ircinia* spp.) in the western Mediterranean Sea. FEMS Microbiology Ecology, 86(2), 268–276. DOI:10.1111/1574-6941.12159

Porter, J.N., Wilhem, J.J., Tresner, H.D. (1960). Method for the Preferential Isolation of Actinomycetes from Soils, Applied Microbiology, 8(3), 174 – 178.

Porter, T.M., Hajibabaei, M. (2018). Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis, Molecular Ecology, 00:1–26. DOI: 10.1111/mec.14478

Prunier, A.L., Schuch, R., Fernndez, R.E., Mumy, K.L., Kohler, H., McCormick, B.A., Maurelli, A.T. (2007). nadA and nadB of *Shigella flexneri* 5a are antivirulence loci responsible for the synthesis of quinolinate, a small molecule inhibitor of *Shigella* pathogenicity, Microbiology, 153(7), 2363 - 2372. DOI: 10.1099/mic.0.2007/006916-0

Pupin, M., Esmaeel, Q., Flissi, A., Dufresne, Y., Jacques, P., Leclere, V. (2016). Norine: A powerful resource for novel nonribosomal peptide discovery, Synthetic and Systems Biology, 1(2), 89 – 94. DOI: 10.1016/j.synbio.2015.11.001

Purves, K., Macintyre, L., Brennan, D., Hreggvidsson, G.O., Kuttner, E., Asgeirsdottir, M.E., Young, L.C., Green, D.H., Edrada-Ebel, R., Duncan, K.R. (2016). Using Molecular Networking for Microbial Secondary Metabolite Bioprospecting, Metabolites, 6(1), 2. DOI: 10.3390/metabo6010002

Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J., Segata, M. (2017). Shotgun metagenomics, from sampling to analysis, Nature Biotechnology, 35(9). DOI: 10.1038/nbt.3935

Radisky, D.C., Radisky, E.S., Barrows, L.R., Copp, B.R., Kramer, R.A., Ireland, C.M. (1993). Novel cytotoxic topoisomerase II inhibiting pyrroloiminoquinones from Fijian sponges of the genus *Zyzzya*, Journal of the American Chemical Society, 115(5), 1632 – 1638. DOI: 10.1021/ja00058a003

Reich, M., Labes, A. (2017). How to boost marine fungal research: A first step towards a multidisciplinary approach by combining molecular fungal ecology and natural products chemistry, Marine Genomics, pii: S1874-7787(17)30066-1. DOI: 10.1016/j.margen.2017.09.007.

Reimer, D., Hughes, C.C. (2017). Thiol-Based Probe for Electrophilic Natural Products Reveals That Most of the Ammosamides Are Artifacts, Journal of Natural Products, 80(1), 126 -133. DOI: 10.1021/acs.jnatprod.6b00773

Reveillaud, J., Maignien, L., Eren, A.M., Huber, J.A., Apprill, A., Sogin, M.L., Vanreusel, A. (2014). Host-specificity among abundant and rare taxa in the sponge microbiome, The ISME Journal, 8, 1198 – 1209. DOI: 10.1038/ismej.2013.227

Reynolds, D., Thomas, T. (2016). Evolution and function of eukaryotic-like proteins from sponge symbionts, Molecular Ecology, 25(20), 5242 – 5253. DOI: 10.1111/mec.13812

Rezanka, T., Sigler, K. (2008). Biological active compounds of semi-metals, Phytochemistry, 69, 585 - 606. DOI: 10.1016/j.phytochem.2007.09.018

Riedlingera, J., Reike, A., Zahner, H., Krismer, B., Bull, A.T., Maldonado, L.A., Ward, A.C., Goodfellow, M., Bister, B., Bischoff, D., Suusmuth, R.D., Fiedler, H.P. (2004). Abyssomicins, inhibitors of the para-aminobenzoic acid pathway produced by the marine *Verrucosispora* strain AB-18-032, The Journal of Antibiotics, 57(4), 271 – 279. DOI: 10.7164/antibiotics.57.271

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J., Darling, A., Malfatti, S., Swan, B.K., Gies, E.A., Dodsworth, J.A., Hedlund, B.P., Tsiamis, G., Sievert, S.M., Liu, W., Eisen, J.A., Hallam, S.J., Kyrpides, N.C., Stepanauskas, R., Rubin, E.M., Hugenholtz, P., Woyke, T. (2013). Insights into the phylogeny and coding potential of microbial dark matter, Nature, 499, 431 – 437. DOI: 10.1038/nature12352

Rives, A., Delaine, T., Legentil, L., Delfourne, E. (2009). Total synthesis of the marine pyrroloiminoquinone alkaloid tsitsikammamine A. Tetrahedron Letters, 50(10), 1128–1130. DOI: 10.1016/j.tetlet.2008.12.078

Rives, A., Le Calvé, B., Delaine, T., Legentil, L., Kiss, R., Delfourne, E. (2010). Synthesis and antitumor evaluation of analogues of the marine pyrroloiminoquinone tsitsikammamines. European Journal of Medicinal Chemistry, 45(1), 343–51. DOI: 10.1016/j.ejmech.2009.10.019

Robbins, T., Liu, Y., Cane, D.E., Khosla, C. (2016). Structure and mechanism of assembly line polyketide synthases, Current Opinion in Structural Biology, 41, 10 – 18. DOI: 10.1016/j.sbi.2016.05.009

Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahe, F. (2016). VSEARCH: a versatile open source tool for metagenomics, PeerJ, 4, e2584. DOI: 10.7717/peerj.2584

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, 362-367. DOI: 10.1093/nar/gkr323

Roumpeka, D.D., Wallace, R.J., Escalettes, F., Fotheringham, I., Watson, M. (2017). A Review of Bioinformatics Tools for Bio-Prospecting from Metagenomic Sequence Data, Fronteirs in Genetics, 8, 23. DOI: 10.3389/fgene.2017.00023

Russell, J.A., Dublier, N., Rudgers, J.A. (2014). Nature's Microbiome: Introduction, Molecular Ecology, 23, 1225 - 1237. DOI: 10.1111/mec.12676

Sakar, S., Ulett, G.C., Totsika, M., Phan, M., Schembri, M.A. (2014). Role of Capsule and O Antigen in the Virulence of Uropathogenic *Escherichia coli*, PLoS ONE 9(4): e94786. DOI: 10.1371/journal.pone.0094786

Sakemi, S., Sun, H.H., Jefford, C.W., Bernardinelli, G. (1989). Batzellines A, B, and C. Novel pyrroloquinoline alkaloids from the sponge *Batzella* sp., tetrahedron Letters, 30(19), 2517 – 2520. DOI: 10.1016/S0040-4039(01)80439-1

Sala, G.D., Hochmuth, T., Teta, R., Constantino, V., Mangoni, A. (2014). Polyketide Synthases in the Microbiome of the Marine Sponge *Plakortis halichondrioides*: A Metagenomic Update, Marine Drugs, 12(11), 5425 – 5440. DOI: 10.3390/md12115425

Samaai, T., Gibbons, M.J., Kelly, M., Davies-Coleman, M. (2003). South African Latrunculiidae (Porifera: Demospongiae: Poecilosclerida): descriptions of new species of *Latrunculia* du Bocage, *Strongylodesma* Lévi, and *Tsitsikamma* Samaai and Kelly. Zootaxa 371, 1–26. DOI: 10.11646/zootaxa.371.1.1

Samaai, T., Govender, V., Kelly, M. (2004). *Cyclacanthia* n.g. (Demospongiae: Poecilosclerida: Latrunculiidae incertea sedis), a new genus of marine sponges from South African waters, and description of two new species, Zootaxa, 725, 1- 18. DOI: 10.11646/zootaxa.725.1.1

Samaai, T., Kelly, M. (2002). Family Latrunculiidae Topsent, 1922. In: Hooper JNA, van Soest RWM (eds) Systema Porifera: guide to the classification of sponges. Kluwer, New York, pp 708–719

Santos-Gandelman, J. F., Giambiagi-deMarval, M., Muricy, G., Barkay, T., Laport, M. S. (2014) Mercury and methylmercury detoxification potential by sponge-associated bacteria, Antonie van Leewenhoek, 106(3), 585- 590, http://doi: 10.1007/s10482-014-0224-2

Sarkar, S., Ulett, G.C., Totsika, M., Phan, M., Schembri, M.A. (2014). Role of Capsule and O Antigen in the Virulence of Uropathogenic Escherichia coli, PLoS One, 9(4), e94786. DOI: 10.1371/journal.pone.0094786

Sayyed, R. Z., Badgujar, M. D., Sonawane, H. M., Mhaske, M. M., Chincholkar, S. B. (2005). Production of Microbial Iron Chelators (siderophores) by Fluorescent Pseudomonads, Indian Journal of Biotechnology, 4, 484–490.

Schirmer, A., Gadkari, R., Reeves, C. D., Ibrahim, F., Delong, E. F., Richard, C., & Hutchinson, C. R. (2005). Metagenomic Analysis Reveals Diverse Polyketide Synthase Gene Clusters in Microorganisms Associated with the Marine Sponge *Discodermia dissolute*, Applied and Environmental Microbiology, 71(8), 4840– 4849. DOI: 10.1128/AEM.71.8.4840

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology, 75, 7537–7541. DOI: 10.1128/AEM.01541-09

Schmidt, O. (1870). Grundzüge einer Spongien-Fauna des atlantischen Gebietes. (Wilhelm Engelmann: Leipzig): iii-iv, 1-88, pls I-VI.

Schmidt, E.W., Harper, M.K., Faulkner, D.J. (1995). Makaluvamines H-M and damirone C from the Pohnpeian sponge Zyzzya fuliginosa, 58(12), 1861 – 1867. DOI: 10.1021/np50126a008

Schmitt, S., Tsai, P., Bell, J., Fromont, J., Illan, 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. DOI: 10.1038/ismej.2011.116

Schoenhofen, I.C., McNally, D.J., Vinogradov, E., Whitfield, D., Young, N.M., Dick, S., Wakarchuk, W.W., Brisson, J., Logan, S.M. (2006). Functional characterization of dehydratase/aminotransferase pairs from *Helicobacter* and *Campylobacter*: enzymes distinguishing the pseudaminic acid and bacillosamine biosynthetic pathways, The Journal of Biological Chemistry, 281(2), 723 – 732. DOI: 10.1074/jbc.M511021200

Schofield, M.M., Jain, S., Porat, D., Dick, G.J., Sherman, D.H. (2015). Identification and analysis of the bacterial endosymbiont specialized for production of the chemotherapeutic natural product ET-743, Environmental Microbiology, 17(10), 3964 – 3975. DOI: 10.1111/1462-2920.12908

Schumacher, M.A., Piro, K.M., Xu, W., Hansen, S., Lewis, K., Brennan, R.G. (2009). Molecular Mechanisms of HipA-Mediated Multidrug Tolerance and Its Neutralization by HipB, Science, 323(5912), 396 – 401. DOI: 10.1126/science.1163806

Schwartsmann, G., da Rocha, A. B., Berlinck, R. G., Jimeno, J. (2001). Marine organisms as a source of new anticancer agents. The Lancet Oncology, 2(4), 221–225. DOI:10.1016/S1470-2045(00)00292-8

Schwazer, D., Mootz, H.D., Linne, U., Marahiel, M.A. (2002). Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases, Proceedings of the National Academy of Sciences of the United States of America, 99(22), 14083 – 14088. DOI: 10.1073/pnas.212382199

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation, Bioinformatics, 30(14), 2068 – 2069. DOI: 10.1093/bioinformatics/btu153

Sepe, V., D'Orsi, R., Borbone, N., D'Auria, M.V., Bifulco, G., Monti, M.C., Catania, A., Zampella, A. (2006). Callipeltins F–I: new antifungal peptides from the marine sponge *Latrunculia* sp., Tetrahedron, 62(5), 833 – 840. DOI: 10.1016/j.tet.2005.10.058

Seyedsayamdost, M. R., Traxler, M. F., Zheng, S., Kolter, R., Clardy, J. (2011). Structure and Biosynthesis of Amychelin, an Unusual Mixed-Ligand Siderophore from *Amycolatopsis* sp. AA4. Journal of the American Chemical Society, 133, 11434–11437. DOI: 10.1021/ja203577e

Shah, S., Karkhanis, V., Desai, A. (1992). Isolation and characterization of siderophore, with antimicrobial activity, fromAzospirillum lipoferum M, Current Microbiology, 25(6), 347 – 351. DOI: 10.1007/BF01577233

Sharp, K.H., Eam, B., Faulkner, D.J., Haygood, M.G. (2007). Vertical Transmission of Diverse Microbes in the Tropical Sponge *Corticium* sp., Applied and Environmental Microbiology, 73(2), 622 – 629. DOI: 10.1128/AEM.01493-06

Sharpton, T. J. (2014). An introduction to the analysis of shotgun metagenomic data. Frontiers in Plant Science, 5, 1–14. DOI: 10.3389/fpls.2014.00209

Shen, B. (2003). Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms, Current Opinion, 7, 285 – 295. DOI: 10.1016/s1367-5931(03)00020-6

Shimizu, Y., Ogata, H., Goto, S. (2016). Type III Polyketide Synthases: Functional Classification and Phylogenomics, ChemBioChem, 18(1), 50 – 65. DOI: 10.1002/cbic.201600522

Siegl, A., Kamke, J., Hochmuth, T., Piel, J., Richter, M., Liang, C., Dandekar, T., Hentschel, U. (2011). Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges, The ISME Journal, 5, 61 - 70. DOI: 10.1038/ismej.2010.95

Sikorska, J., Hau, A. M., Anlklin, C., Parker-Nance, S., Davies-Coleman, M. T., Ishmael, J. E., McPhail, K. L. (2012). Mandelalides A-D, Cytotoxic Macrolides from a New *Lissoclinum* Species of South African Tunicate, Journal of Organic Chemistry, 29(6), 997–1003. DOI: 10.1016/j.biotechadv.2011.08.021.

Simister, R., Taylor, M. W., Rogers, K. M., Schupp, P. J., Deines, P. (2013). Temporal Molecular and Isotopic Analysis of Active Bacterial Communities in Two New Zealand sponges. FEMS Microbiology Ecology, 85(1), 195–205. http://doi.org/10.1111/1574-6941.12109

Simister, R.L., Deines, P., Botte, E.S., Webster, N.S., Taylor, M.W. (2012). Sponge-specific clusters revisited: A comprehensive phylogeny of sponge-associated microorganisms, Environmental Microbiology, 14(2), 517 -524. DOI: 10.1111/j.1462-2920.2011.02664.x"

Singh, P., Kaur, S., Sharma, A., Kaur, G., Bhatti, R. (2017). TNF-α and IL-6 inhibitors: Conjugates of N-substituted indole and aminophenylmorpholin-3-one as anti-inflammatory agents, European Journal of Medicinal Chemistry, 140(10), 92 - 103. DOI: 10.1016/j.ejmech.2017.09.003

Singh, S., Prasad, P., Subramani, R., Aalbersberg, W. (2014). Production and purification of a bioactive substance against multi-drug resistant human pathogens from the marine-sponge-derived *Salinispora* sp. Asian Pacific Journal of Tropical Biomedicine, 4(10), 825–831. DOI:10.12980/APJTB.4.2014C1154

Sipkema, D., Blanch, H.W. (2010). Spatial distribution of bacteria associated with the marine sponge *Tethya californiana*, Marine Biology, 157(3), 627 – 638. DOI: 10.1007/s00227-009-1347-2

Sitaraman, R. (2016). The Role of DNA Restriction-Modification Systems in the Biology of *Bacillus anthracis*, Frontiers in Microbiology, 7(11). DOI: 10.3389/fmicb.2016.00011

Slaby, B.M., Hackl, T., Horn, H., Bayer, K., Hentschel, U. (2017). Metagenomic binning of a marine sponge microbiome reveals unity in defense but metabolic specialization, The ISME Journal, 11, 2465 - 2478. DOI: 10.1038/ismej.2017.101

Sorci, L., Blaby, I.K., Rodionova, I.A., De Ingeniis, J., Tkachenko, S., de Crecy-Lagard, V., Osterman, A.L. (2013). Quinolinate Salvage and Insights for Targeting NAD Biosynthesis in Group A *Streptococci*, Journal of Bacteriology, 195(4), 726 - 732. DOI: 10.1128/JB.02002-12

Souza, D.T., Genuario, D.B., Silva, F.S.P., Pansa, C.C., Kavamura, V.N., Moraes, F.C., Taketani, R.G., Melo, I.S. (2017). Analysis of bacterial composition in marine sponges reveals the influence of host phylogeny and environment, FEMS Microbiology Ecology, 93(1). DOI:10.1093/femsec/fiw204

Sravanthi, T., Tushar, L., Sasikala, C., Ramana, C.V. (2016). *Alkalispirochaeta cellulosivorans* gen. nov., sp. nov., a cellulose-hydrolysing, alkaliphilic, halotolerant bacterium isolated from the gut of a wood-eating cockroach (*Cryptocercus punctulatus*), and reclassification of four species of *Spirochaeta* as new combinations within *Alkalispirochaeta* gen. nov., International Journal of Systematic and Evolutionary Microbiology, 66, 1612 – 1619. DOI: 10.1099/ijsem.0.000865

Steinert, G., Rohde, S., Janussen, D., Blaurock, C., Schupp, P.J. (2017). Host-specific assembly of sponge-associated prokaryotes at high taxonomic ranks, Scientiic Reports, 7, 2542. DOI: 10.1038/s41598-017-02656-6

Steinert, G., Rohde, S., Janussen, D., Blaurock, C., Schupp, P.J. (2017). Host-specific assembly of sponge-associated prokaryotes at high taxonomic ranks, Scientific Reports, 7, 2542. DOI: 10.1038/s41598-017-02656-6

Steinert, G., Taylor, M.W., Deines, P., Simister, R.L., de Voogd, N.J., Hoggard, M., Schupp, P.J. (2016). In four shallow and mesophotic tropical reef sponges from Guam the microbial community largely depends on host identity, PeerJ, 4, https://doi.org/10.7717/peerj.1936

Steller, S., Sokoll, A., Wilde, C., Bernhard, F., Franke, P., Vater, J. (2004). Initiation of Surfactin Biosynthesis and the Role of the SrfD-Thioesterase Protein, Biochemistry, 43, 11331 – 11343. DOI: 10.1021/bi0493416

Strieker, M., Tanovic, A., Marahiel, M.A. (2010). Nonribosomal peptide synthetases: structures and dynamics, Current Opinion in Structural Biology, 20, 234 – 240. DOI: 10.1016/j.sbi.2010.01.009

Studholm, D.J., Fuerst, J.A., Bateman, A. (2004). Novel protein domains and motifs in the marine planctomycete *Rhodopirellula baltica*, FEMS Microbiology Letters, 236(2), 333 – 340. DOI: 10.1111/j.1574-6968.2004.tb09666.x

Sun, H.H., Sakemi, S., Burres, N., McCarthy, P. (1990). Isobatzellines A, B, C and D. Cytotoxic and antifungal pyrrolquinoline alkaloids from the marine sponge *Batzella* sp., Journal of Organic Chemistry, 55(16), 4964 – 4966. DOI: 10.1021/jo00303a043

Sun, Y., Tomura, T., Sato, J., Iizuka, T., Fudou, R., Ojika, M. (2016). Isolation and Biosynthetic Analysis of Haliamide, a New PKS-NRPS Hybrid Metabolite from the Marine Myxobacterium *Haliangium ochraceum*, Molecules, 21(1), 59. DOI: 10.3390/molecules21010059

Sunassee, S. N., Davies-Coleman, M. T. (2012). Cytotoxic and antioxidant marine prenylated quinones and hydroquinones. Natural Product Reports, 29(5), 513. DOI:10.1039/c2np00086e

Tatsuta, K., Imamura, K., Itoh, S., Kasai, S. (2004). The first total synthesis of lymphostin, Tetrahedron letters, 45(13), 2847 – 2850. DOI: 10.1016/j.tetlet.2004.01.140

Tauch, A., Kaiser, O., Hain, T., Goesmann, A., Weisshaar, B., Albersmeier, A., Bekel, T., Bischoff, N., Brune, I., Chakraborty, T., Kalinowski, J., Meyer, F., Rupp, O., Schneiker, S., Viehoever, P., Puhler, A. (2005). Complete Genome Sequence and Analysis of the Multiresistant Nosocomial Pathogen Corynebacterium jeikeium K411, a Lipid-Requiring Bacterium of the Human Skin Flora, Journal of Bacteriology, 187(13), 4671 – 4682. DOI: 10.1128/JB.187.13.4671-4682.2005

Taylor, M. W., Schupp, P. J., Dahllöf, I., Kjelleberg, S., Steinberg, P. D. (2004). Host specificity in marine sponge-associated bacteria, and potential implications for marine microbial diversity. Environmental Microbiology, 6(2), 121–130. DOI:10.1046/j.1462-2920.2003.00545.x

Taylor, M. W., Tsai, P., Simister, R. L., Deines, P., Botte, E., Ericson, G., Schmitt, S., Webster, N. S. (2012). "Sponge- specific" bacteria are widespread (but rare) in diverse marine environments. The ISME Journal, 438–443. DOI:10.1038/ismej.2012.111

Thacker, R.W., Hill, A.L., Hill, M.S., Redmond, N.E., Collins, A.G., Morrow, C.C., Spicer, L., Carmack, C.A., Zappe, M.E., Pohlmann, D., Hall, C., Diaz, M.C., Bangalore, P.V. (2013). Nearly Complete 28S rRNA Gene Sequences Confirm New Hypotheses of Sponge Evolution, Integrative and Comparative Biology, 53(3), 373 – 387. DOI: 10.1093/icb/ict071

Thiel, V., Leininger, S., Schmaljohann, R., Brümmer, F., Imhoff, J. F. (2007a). Sponge-specific bacterial associations of the Mediterranean sponge *Chondrilla nucula* (Demospongiae, Tetractinomorpha). Microbial Ecology, 54(1), 101–111. DOI:10.1007/s00248-006-9177-y

Thiel, V., Neulinger, S.C., Staufenberger, T., Schmaljohann, R., Imhoff, J.F. (2007b). Spatial distribution of sponge-associated bacteria in the Mediterranean sponge *Tethya aurantium*, FEMS Microbial Ecology, 59(1), 47 – 63. DOI: 10.1111/j.1574-6941.2006.00217.x

Thogersen, M.S., Delpin, M.W., Melchiorsen, J., Kilstrup, M., Mansson, M., Bunk, B., Sproer, C., Overmann, J., Nieelsen, K.F., Gram, L. (2016). Production of the Bioactive Compounds Violacein and Indolmycin Is Conditional in a maeA Mutant of *Pseudoalteromonas luteoviolacea* S4054 Lacking the Malic Enzyme, Frontiers in Microbiology,7, 1461. DOI: 10.3389/fmicb.2016.01461

Thomas, T., Moitinho-Silva, L., Lurgi, M., Björk, J.R., Easson, C., Astudillo-García, C., Olson, J.B., Erwin, P.M., López-Legentil, S., Luter, H., Chaves-Fonnegra, A., Costa, R., Schupp, P.J., Steindler, L., Erpenbeck, D., Gilbert, J., Knight, R., Ackermann, G., Lopez, V. J., Taylor, M.W., Thacker, R.W., Montoya, J.M., Hentschel, U., Webster, NS. (2016). Diversity, structure and convergent evolution of the global sponge microbiome, Nature Communications, 16(7). DOI: 10.1038/ncomms11870.

Thomas, T., Rusch, D., DeMaere, M. Z., Yung, P. Y., Lewis, M., Halpern, A., Kjelleberg, S. (2010a). Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. The ISME Journal, 4(12), 1557–1567.DOI:10.1038/ismej.2010.74

Thomas, T.R.A., Kavlekar, D.P., LokaBharathi, P.A. (2010b). Marine Drugs from Sponge-Microbe Association—A Review, Marine Drugs, 8(4), 1417 – 1468. DOI: 10.3390/md8041417

Topsent, E. (1922). Les mégasclères polytylotes des Monaxonides et la parenté des Latrunculiines. Bulletin de l'Institut océanographique Monaco. 415: 1-8.

Torres-Machorro, A.L., Hernandez, R., Cevallos, A.M., Lopez-Villasenor, I. (2009). Ribosomal RNA genes in eukaryotic microorganisms: witnesses of phylogeny?, FEMS Microbioology Reviews, 34(1), 59 – 86. DOI: 10.1111/j.1574-6976.2009.00196.x

Tracanna, V., de Jong, A., Medema, M.H., Kuipers, O.P. (2017). Mining prokaryotes for antimicrobial compounds: from diversity to function, FEMS Microbiology Reviews, 41(3), 417 - 429. DOI: 10.1093/femsre/fux014

Trindade-Silva, A.E., Rua, C., Silva, G.G.Z., Dutilh, B.E., Moreira, A.P.B., Edwards, R.A., Hajdu, E., Lobo-Hajdu, G., Vasconcelos, A.T., Berlinck, R.G.S., Thompson, F.L. (2012). Taxonomic and Functional Microbial Signatures of the Endemic Marine Sponge *Arenosclera brasiliensis*, PloS ONE, 7(7), e39905. DOI: 10.1371/journal.pone.0039905

Truong-Buldoc, Q.C., Dunman, P.M., Eidem, T., Hooper, D.C. (2011). Transcriptional Profiling Analysis of the Global Regulator NorG, a GntR-Like Protein of *Staphylococcus aureus*, Journal of Bacteriology, 193(22), 6207 – 6214. DOI: 10.1128/JB.05847-11

Tu, J.H., Arcak, M., Maharbiz, M.M. (2015). Decoupling translational and rotational effects on the phase synchronization of rotating helices, Physical Review, 91, 023018. DOI: 10.1103/physreve.91.023018

Turk, T., Avgustin, J.A., Batista, U., Strugar, G., Kosmina, R., Civovic, S., Janussen, D., Kauferstein, S., Mebs, D., Sepcic, K. (2013). Biological Activities of Ethanolic Extracts from Deep-Sea Antarctic Marine Sponges, Marine Drugs, 11, 1126-1139. DOI: 10.3390/md11041126

Tyc, O., Song, C., Dickschat, J.S., Vos, M., Garbeva, P. (2017). The Ecological Role of Volatile and Soluble Secondary Metabolites Produced by Soil Bacteria, Trends in Microbiology, 25(4), 280 – 292. DOI: 10.1016/j.tim.2016.12.002

Tzeng, Y., Thomas, J., Stephens, D.S. (2016). Regulation of capsule in *Neisseria meningitidis*, Critical Reviews in Microbiology, 42(5), 759 – 772. DOI: 10.3109/1040841X.2015.1022507

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. DOI: 10.1073/pnas.0700962104

Undabarrena, A., Ugalde, J.A., Seeger, M., Camara, B. (2017). Genomic data mining of the marine actinobacteria *Streptomyces* sp. H-KF8 unveils insights into multi-stress related genes and metabolic pathways involved in antimicrobial synthesis, PeerJ, 5: e2912 https://doi.org/10.7717/peerj.2912

Uriz, M.J., Turon, X. (2006)

Usher, K. M., Kuo, J., Fromont, J., Sutton, D.C. (2001). Vertical transmission of cyanobacterial symbionts in the marine sponge *Chondrilla australiensis* (Demospongiae), Hydrobiologia, 461, 15-23.

Usher, K.M., Erekovsky, A.V. (2005). Larval development, ultrastructure and metamorphosis in *Chondrilla australiensis* Carter, 1873 (Demospongiae, Chondrosida, Chondrillidae), Invertebrate Reproduction and Development, 47(1), 51 – 62. http://dx.doi.org/10.1080/07924259.2005.9652146

Valliappan, K., Sun, W. and Li, Z. (2014) Marine actinobacteria associated with marine organisms and their potentials in producing pharmaceutical natural products, Applied Microbiology and Biotechnology, 98(17), 7365-7377. DOI: 10.1007/s00253-014-5954-6.

Van der Meij, A., Worsley, S.F., Hutchings, M.I., van Wezel, G.P. (2017). Chemical ecology of antibiotic production by actinomycetes, FEMS Microbiology Reviews, 41, 392 - 416. DOI: 10.1093/femsre/fux005

Van der Water, J.A.J.M., Melkonian, R., Junca, H., Voolstra, C.R., Reynaud, S., Allemand, D., Ferrier-Pages, C. (2016). Spirochaetes dominate the microbial community associated with the red coral *Corallium rubrum* on a broad geographic scale, Scientific Reports, 6, e27277. DOI: 10.1038/srep27277

Van Domselaar, G.H., Stothard, P., Shrivastava, S., Cruz, J.A., Guo, A., Dong, X., Lu, P., Szafron, D., Greiner, R., Wishart, D.S. (2005). BASys: A web server for automated bacterial genome annotation, Nucleic Acids Research, 1(33), W455-459. DOI: 10.1093/nar/gki593

Vanicat, A., Andre-Barres, C., Delfourne, E. (2017). New reaction of formation of the fused tricyclic bispyrroloquinone ring system, Tetrhedron Letters, 58(4), 342 – 345. DOI: 10.1016/j.tetlet.2016.12.030

Vartoukian, S.R. (2016). Cultivation strategies for growth of uncultivated bacteria, Journal of Oral Biosciences, 58(4), 143 - 149. DOI: 10.1016/j.job.2016.08.001

Venables, D.A., Concepcion, G.P., Matsumoto, S.S., Barrows, L.R., Ireland, C.M. (1997). Makaluvamine N: A New Pyrroloiminoquinone from *Zyzzya fuliginosa*, Journal of Natural Products, 60(4), 408 – 410. DOI: 10.1021/np9607262

Versluis, D., D'Andrea, M.M., Garcia, J.R., Leimena, M.M., Hugenholtz, F., Zhang, J., Ozturk, B., Nyland, L., Sipkema, D., van Schaik, de Vos, W.M., Kleerebezem, M., Smidt, H., van Passel, M.W.J. (2015). Mining microbial metatranscriptomes for expression of antibiotic resistance genes under natural conditions, Scientific Reports, 5, 11981. DOI: 10.1038/srep11981

Vorholt, J.A., Vaupel, M., Thauer, R.K. (1997). A selenium-dependent and a selenium-independent formylmethanofuran dehydrogenase and their transcriptional regulation in the hyperthermophilic *Methanopyrus kandleri*, Molecular Microbiology, 23(5), 1033 – 1042. DOI: 10.1046/j.1365-2958.1997.2931653.x

Waldman, A. J., Balskus, E. P. (2014). Lomaiviticin biosynthesis employs a new strategy for starter unit generation. Organic Letters, 16(2), 640–643. DOI: 10.1021/ol403714g

Walmsley, T.A. (2013). An Investigation into the Bacterial Diversity Associated with South African Latrunculid Sponges that Produce Bioactive Secondary Metabolites, Retrieved from: http://hdl.handle.net/10962/d1012943

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. DOI: 10.1007/s10126-012-9430-y

Walsh, C.T. (2004). Polyketide and Nonribosomal Peptide Antibiotics: Modularity and Versatility, Science, 303(5665), 1805–1810. DOI: 10.1126/science.1094318

Walwick, E.A.; Dekker, C.A.; Roberts, W.K. (1959). Cyclization during Phosphorylation of Uridine and Cytidine by Polyphosphoric Acid—A New Route to the 0-2-2' Cyclonucleosides. Proceedings of the Chemical Society, 3, 84.

Wang, H., Fewer, D.P., Holm, L., Rouhiainen, L., Sivonen, K. (2014). Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes, Proceedings of the National Academy of Sciences of the United States of America, 111(25), 9259 – 9264. DOI: 10.1073/pnas.1401734111

Wang, J., Jia, H. (2016). Metagenome-wide association studies: fine-mining the microbiome, Nature Reviews Microbiology, 14, 508 – 522. DOI: 10.1038/nrmicro.2016.83

Wang, J., Stephan, R., Zurfluh, K., Hachler, H., Fanning, S. (2015). Characterization of the genetic environment of *bla*ESBL genes, integrons and toxin-antitoxin systems identified on large transferrable

plasmids in multi-drug resistant *Escherichia coli*, Frontiers in Microbiology, 5, 716. DOI: 10.3389/fmicb.2014.00716

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy, Applied and Environmental Microbiology, 73(16), 5261 – 5267. DOI: 10.1128/AEM.00062-07

Wang, W.L., Chi, Z.M., Chi, Z., Li, J., Wang, X.H. (2009). Siderophore production by the marine-derived *Aureobasidium pullulans* and its antimicrobial activity, Bioresource Technology, 100(9), 2639 – 2641. DOI: 10.1016/j.biortech.2008.12.010

Wang, Y., Qian, P.Y. (2009) Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. PLoS ONE, 4, e7401. DOI: 10.1371/journal.pone.0007401

Waterworth, S.C., Jiwaji, M., Kalinski, J.J., Parker-Nance-S., Dorrington, R.A. (2017). A Place to Call Home: An Analysis of the Bacterial Communities in Two *Tethya rubra* Samaai and Gibbons 2005 Populations in Algoa Bay, South Africa, Marine Drugs, 15(4). DOI: 10.3390/md15040095

Watts, A., Ke, D., Wang, Q., Pillay, A., Nicholson-Weller, A., Lee, J.C. (2005). *Staphylococcus aureus* strains that express serotype 5 or serotype 8 capsular polysaccharides differ in virulence, Infection and Immunity, 73(6), 3502 – 3511. DOI: 10.1128/IAI.73.6.3502-3511.2005

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Bruccoleri, R., Lee, S.Y., Fischbach, M.A., Muller, R., Wohlleben, W., Breitling, R., Takano, E., Medema, M.H. (2015). antiSMASH 3.0 — a comprehensive resource for the genome mining of biosynthetic gene clusters, Nucleic Acids Research, 43, 237 – 243. DOI: 10.1093/nar/gkv437

Weber, T., Kim, H.U. (2016). The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production, Synthetic and Systems Biotechnology, 1(2), 69 – 79. DOI: 10.1016/j.synbio.2015.12.002

Webster, N.S. (2014). Cooperation, communication, and co-evolution: grand challenges in microbial symbiosis research, Frontiers in Microbiology, 5(164). DOI: 10.3389/fmicb.2014.00164

Webster, N.S., Thomas, T. (2016). The Sponge Hologenome, mBio, 7(2), e00135 -16. DOI: 10.1128/mBio.00135-16

Webster, N.S., Wilson, K.J., Blackall, L.L., Hill, R.T. (2001). Phylogenetic Diversity of Bacteria Associated with the Marine Sponge *Rhopaloeides odorabile*, Applied and Environmental Microbiology, 67(1), 434 – 444. DOI: 10.1128/AEM.67.1.434-444.2001

Weerapana, E., Imperiali, B. (2006). Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems, Glycobiology, 16(6), 91 – 101. DOI: 10.1093/glycob/cwj099

Weigel, B.L., Erwin, P.M. (2016). Intraspecific Variation in Microbial Symbiont Communities of the Sun Sponge, *Hymeniacidon heliophila*, from Intertidal and Subtidal Habitats, Applied and Environmental Microbiology, 82(2), 650 – 658. DOI: 10.1128/AEM.02980-15

Wen, Y., Behiels, E., Devreese, B. (2014). Toxin–Antitoxin systems: their role in persistence, biofilm formation, and pathogenicity, Pathogens and Disease, 70(3), 240 - 249. DOI:/10.1111/2049-632X.12145

Weyna, T.R., Barkkei, J.J., Miller, I.J., McPhail, K.L., Kwan, J.C. (2015). The mandelalides are produced by a symbiont in the phylum *Verrucomicrobia* in the tunicate *Lissoclinum* sp., Planta Medica, 81, PF4. DOI: 10.1055/s-0035-1556230

Whibly, C.E., Keyzers, R.A., Soper, A.G., Davies-Coleman, M.T., Samaai, T., Hendrick, D.T. (2006). Antiesophageal cancer activity from Southern African marine organisms, Annals of the New York Academy of Sciences, 1056, 405 - 412. DOI: 10.1196/annals.1352.031

White, J. R., Patel, J., Ottesen, A., Arce, G., Blackwelder, P., Lopez, J. V. (2012). Pyrosequencing of bacterial symbionts within *Axinella corrugata* sponges: Diversity and seasonal variability. PLoS ONE, 7(6). DOI:10.1371/journal.pone.0038204

Wietz, M., Duncan, K., Patin, N.V., Jensen, P.R. (2013). Antagonistic Interactions Mediated by Marine Bacteria: The Role of Small Molecules, Journal of Chemical Ecology, 39(7), 879 – 891. DOI: 10.1007/s10886-013-0316-x

Wilson, M.C., Mori, T., Ruckert, 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., Crusemann, M., Hentschel, U., Abe, I., Matsunaga, S., Kalinowski, J., Takeyma, H., Piel, J. (2014). An environmental bacterial taxon with a large and distinct metabolic repertoire, Nature, 506, 58 – 62. DOI: 10.1038/nature12959

Wilson, M.C., Piel, J. (2013). Metagenomic Approaches for Exploiting Uncultivated Bacteria as a Resource for Novel Biosynthetic Enzymology, Chemistry and Biology, 20(5), 636 – 647. DOI: 10.1016/j.chembiol.2013.04.011

Winter, J. M., Behnken, S., Hertweck, C. (2011). Genomics-inspired Discovery of Natural Products, Current Opinion in Chemical Biology, 15(1), 22–31. DOI: 10.1016/j.cbpa.2010.10.020

WoRMS Editorial Board (2017). World Register of Marine Species. Available from http://www.marinespecies.org at VLIZ. Accessed 2017-08-17. doi:10.14284/170

Woyke, T., Teeling, H., Ivanova, N.N., Huntemann, M., Richter, M., Gloekner, F.O., Boffelli, D., Anderson, I.J., Barry, K.W., Shapiro, H.J., Szeto, E., Kyrpides, N.C., Mussmann, M., Amann, R., Bergin, C., Ruehland, C., Rubin, E.M., Dublier, N. (2006). Symbiosis insights through metagenomic analysis of a microbial consortium, Nature, 443, 950 – 955. DOI: 10.1038/nature05192

Wyrick, P.B. (2000). Intracellular survival by *Chlamydia*, Cellular Microbiology, 2(4), 275 – 282. DOI: 10.1046/j.1462-5822.2000.00059.x

Xavier, J.R., Rachello-Dolmen, P.G., Parra-Velandia, F., Schonberg, C.H.L., Breeuwer, J.A.J., van Soest, R.W.M. (2010). Molecular evidence of cryptic speciation in the “cosmopolitan” excavating sponge *Cliona celata* (Porifera, Clionaidae), Molecular Phylogenetics and Evolution, 56(1), 13 – 20. DOI: 10.1016/j.ympev.2010.03.030

Yang, A., Baker, B.J., Grimwade, J., Leonard, A., McClintock, J.B. (1995). Discorhabdin alkaloids from the Antarctic sponge *Latrunculia apicalis*, Journal of Natural Products, 58(10), 1596 – 1599. DOI: 10.1021/np50124a020

Yang, Q., Farnco, C.M.M., Sorokin, S.J., Zhang, W. (2017). Development of a multilocus-based approach for sponge (phylum Porifera) identification: refinement and limitations, Scientific Reports, 7, 41422. DOI: 10.1038/srep41422

Yoon, S.H., Ha, S.M., Lim, J.M., Kwon, S.J., Chun, J. (2017). A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie van Leeuwenhoek, 110, 1281–1286. DOI: 10.1007/s10482-017-0844-4

Zampella, A., Sepe, V., Luciano, P., Bellotta, F., Monti, M. C., D’Auria, M. V., Jepsen, T., Petek, S., Adeline, M., Laprevote, O., Aubertin, A., Debitus, C., Poupat, C., Ahond, A. (2008). Homophymine A, an anti-HIV cyclodepsipeptide from the sponge *Homophymia* sp. Journal of Organic Chemistry, 73(14), 5319–5327. DOI:10.1021/jo800583b

Zhang, F., Blasiak, L. C., Karolin, J. O., Powel, R. J., Geddes, C. D., Hill, R. T. (2015). Phosphorous sequestration in the form of polyphosphate by microbial symbionts in marine sponges, Proceedings of the National Academy of Sciences of the United States of America, 112(14), http:// www.pnas.org/cgi/doi/10.1073/pnas

Zhang, L., Cava, M.P., Rogers, R.D., Rogers, L.M. (1998). Synthesis of a wakayin model compound: Oxidative formation of a new pyrrole ring in the indol-3-yl-indoloquinone system, Tetrahedron Letters, 39(42), 7677 – 7678. DOI: 10.1016/S0040-4039(98)01718-3

Zhang, W., Zhang, F., Li, Z., Miao, X., Meng, Q., & Zhang, X. (2009). Investigation of bacteria with polyketide synthase genes and antimicrobial activity isolated from South China Sea sponges. Journal of Applied Microbiology, 107(2), 567–575. DOI:10.1111/j.1365-2672.2009.04241.x

Zheng, H., Dietrich, C., Hongoh, Y., Brune, A. (2016). Restriction-Modification Systems as Mobile Genetic Elements in the Evolution of an Intracellular Symbiont, Molecular Biology and Evolution, 33(3), 721 – 725. DOI: 10.1093/molbev/msv264

Ziemart, N., Lechner, A., Wietz, M., Millan-Aguinaga, N., Chavarria, K.L., Jensen, P.R. (2014). Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*, Proceedings of the National Academy of Sciences of the United States of America, 111(12), 1130 – 1139. DOI: 10.1073/pnas.1324161111

Zou, Y., Hamann, M.T. (2013). Atkamine: a new pyrroloiminoquinone scaffold from the cold water Aleutian Islands *Latrunculia* sponge, Organic Letters, 15(7), 1516- 1519. DOI: 10.1021/ol400294v

Appendix A – Sponge Collection Metadata

Table A.1. Sponge Collection Metadata

Specimen ID	Taxonomy	Collection date	Collection site	Depth
TIC2009 – 001	*T. favus*	May 2009	Evans Peak – Algoa Bay	30 m
TIC2009 - 002	*T. favus*	May 2009	Evans Peak – Algoa Bay	30 m
TIC2010 – 2B	*Tsitsikamma nguni*	May 2010	Evans Peak – Algoa Bay	30 m
TIC2010 – 030	*Mycale* sp.	May 2010	Evans Peak – Algoa Bay	30 m
TIC2010 – 031	*Latrunculia algoaensis*	May 2010	Evans Peak – Algoa Bay	30 m
TIC2010 - 070	*Tsitsikamma* sp. 005	May 2010	Evans Peak – Algoa Bay	30 m
TIC2011 - 102	*Tsitsikamma michaeli*	April 2011	Evans Peak – Algoa Bay	30 m
TIC2012 – 056	*Cyclacanthia bellae*	Dec 2012	RIY Banks – Algoa Bay	30 m
TIC2012 – 057	*Tsitsikamma favus*	Dec 2012	Evans Peak – Algoa Bay	30 m
TIC2014 – 001	*Tsitsikamma favus*	Aug 2014	Evans Peak – Algoa Bay	30 m
Evans Seawater 2014	N/A	Aug 2014	Evans Peak – Algoa Bay	30 m
Evans Sediment 2014	N/A	Aug 2014	Evans Peak – Algoa Bay	39 m
TIC2015 -003D	*Tsitsikamma nguni*	Jun 2015	Rheeders Reef -Tsitsikamma National Park	20 m
TIC2015 -003E	*Tsitsikamma nguni*	Jun 2015	Rheeders Reef -Tsitsikamma National Park	20 m
TIC2015 -003F	*Tsitsikamma nguni*	Jun 2015	Rheeders Reef -Tsitsikamma National Park	20 m
TIC2015 -003L	*Tsitsikamma nguni*	Jun 2015	Rheeders Reef -Tsitsikamma National Park	20 m
TIC2015 -006	*Tsitsikamma nguni*	Jun 2015	Rheeders Reef -Tsitsikamma National Park	20 m
Rheeders Seawater 2015	N/A	Jun 2015	Rheeders Reef -Tsitsikamma National Park	20 m
TIC2015 – 013	*Tsitsikamma favus*	Sep 2015	Evans Peak – Algoa Bay	30 m
TIC2015 – 025	*Tsitsikamma favus*	Sep 2015	Evans Peak – Algoa Bay	30 m
TIC2015 – 026	*Tsitsikamma favus*	Sep 2015	Evans Peak – Algoa Bay	30 m
TIC2015 – 027	*Tsitsikamma favus*	Sep 2015	Evans Peak – Algoa Bay	30 m
TIC2015 – 201	*Tsitsikamma michaeli*	Nov 2015	Evans Peak – Algoa Bay	30 m
TIC2015 – 202	*Tsitsikamma michaeli*	Nov 2015	Evans Peak – Algoa Bay	30 m
TIC2015 – 204	*Tsitsikamma michaeli*	Nov 2015	Evans Peak – Algoa Bay	30 m

TIC2015 – 205	*Tsitsikamma pedunculata*	Nov 2015	Evans Peak – Algoa Bay	30 m
TIC2015 – 206	*Tsitsikamma pedunculata*	Nov 2015	Evans Peak – Algoa Bay	30 m
TIC2015 – 217	*Tsitsikamma pedunculata*	Nov 2015	Evans Peak – Algoa Bay	30 m
TIC2015 – 218	*Tsitsikamma pedunculata*	Nov 2015	Evans Peak – Algoa Bay	30 m
TIC2015 – 221	*Tethya rubra*	Nov 2015	Evans Peak – Algoa Bay	30 m
TIC2015 - 222	*Tethya rubra*	Nov 2015	Evans Peak – Algoa Bay	30 m
Evans Seawater 2015	N/A	Nov 2015	Evans Peak – Algoa Bay	30 m
TIC2016 - 001	*Tethya rubra*	Mar 2016	RIY Banks – Algoa Bay	50 m
TIC2016 – 003	*Tethya rubra*	Mar 2016	RIY Banks – Algoa Bay	50 m
RIY Seawater 2016	N/A	Mar 2016	RIY Banks – Algoa Bay	50 m
TIC2016 – 050A	*Tsitsikamma favus*	Mar 2016	Evans Peak – Algoa Bay	20 m
TIC2016 – 050B	*Tsitsikamma favus*	Mar 2016	Evans Peak – Algoa Bay	20 m
TIC2016 – 050C	*Tsitsikamma favus*	Mar 2016	Evans Peak – Algoa Bay	20 m
TIC2016 – 050AD	*Tsitsikamma favus*	Mar 2016	Evans Peak – Algoa Bay	20 m
TIC2016 – 050AH	*Tsitsikamma favus*	Mar 2016	Evans Peak – Algoa Bay	20 m
TIC2016 – 050AW	*Tsitsikamma favus*	Mar 2016	Evans Peak – Algoa Bay	20 m
TIC2016 – 051	*Cyclacanthia bellae*	June 2016	Evans Peak – Algoa Bay	20 m
TIC2016 – 052	*Cyclacanthia bellae*	June 2016	Evans Peak – Algoa Bay	20 m
TIC2016 – 053	*Cyclacanthia bellae*	June 2016	Brenton Island – Algoa Bay	20 m
ACE2017 – 045	*Latrunculia apicalis*	Mar 2017	Bouvet Island – Southern Ocean	300 m
ACE2017 – 060	*Latrunculia apicalis*	Mar 2017	Bouvet Island – Southern Ocean	300 m
ACE2017 – 080	*Latrunculia apicalis*	Mar 2017	Bouvet Island – Southern Ocean	300 m
TIC2017 – 002	*Tethya* sp.	Aug 2017	RIY Banks – Algoa Bay	50 m
TIC2017 – 003	*Tethya* sp.	Aug 2017	RIY Banks – Algoa Bay	50 m
TIC2017 – 005	*Tethya* sp.	Aug 2017	RIY Banks – Algoa Bay	50 m
Evans Seawater 2017	N/A	Aug 2017	Evans Peak – Algoa Bay	10m, 25m and 33m
RIY Seawater 2017	N/A	Aug 2017	RIY Banks – Algoa Bay	10m, 25m and 55m

Appendix B – Media used

Artificial seawater:

24.6 g NaCl, 0.67 g KCl, 1.36 g $CaCl_2.2H_2O$, 6.29 g $MgSO_4.7H_2O$, 4.66 g $MgCl_2.6H_2O$ and 0.18 g $NaHCO_3$, distilled H_2O to 1 L

Yeast extract/Malt Extract agar:

4 g yeast extract, 10 g malt extract, 4 g glucose, 20 g agar; distilled H_2O to 1 L

Calcium and Magnesium Free Seawater:

25 g NaCl, 0.8 g KCl, 1 g Na_2SO_4, 0.04 g $NaHCO_3$, distilled H_2O to 1 L

Resuspension Buffer:

0.5 M NaCl, 100 mM EDTA, 10 mM Tris (pH8)

2YT media:

16 g tryptone, 10 g yeast extract, 5 g NaCl; distilled H_2O to 1 L

Appendix C – Publications