

This file is part of the following work:

O'Brien, Paul (2022) *Host-microbe symbiosis and coevolution in coral reef invertebrates*. PhD Thesis, James Cook University.

Access to this file is available from:

<https://doi.org/10.25903/2r40%2Daa68>

Copyright © 2022 Paul O'Brien.

The author has certified to JCU that they have made a reasonable effort to gain permission and acknowledge the owners of any third party copyright material included in this document. If you believe that this is not the case, please email

researchonline@jcu.edu.au

**Host-microbe symbiosis and coevolution in coral
reef invertebrates**

Thesis submitted by

Paul O'Brien

4th May 2022

For the degree of Doctor of Philosophy

College of Science and Engineering

James Cook University

Acknowledgements

First and foremost, and I would like to give my sincere gratitude to my family, particularly my parents Marie McKay and Adrian O'Brien. Without your sacrifices and support I would not have had the opportunity to go to university and ultimately pursue a PhD and career in marine science.

I would like to send a huge thankyou to my advisory team; David Bourne, Nicole Webster and David Miller. I feel so fortunate to have worked with and learned from such a brilliant and amiable team. Specifically, David B taught me to keep pressing forward and making progress, even when things are difficult or not going as planned. Nicole showed me how important it is to lead with positivity and compassion and the importance of keeping up morale within a lab group. David M demonstrated the importance of attention to detail and how to be precise with your words and explanations.

I want to give a special mention to Pedro Frade. While you might not have been an official advisor, your mentoring and guidance was certainly aligned with a supervisory role, but more importantly developed into a life-long friendship.

None of this would have been possible without the support from AIMS@JCU. From day one, Libby Evans-Illidge and Lauren Gregory have always put the students first and their constant effort to assist with workshops and retreats has certainly improved the skills and productivity of those involved. Similarly, this thesis would not have been possible without the sequencing support from the Beijing Genome Institute, in particular Guojie Zhang and Shangjin Tan.

To all my friends in Townsville that I met along the way. Townsville would not be the place it is without the community of people that make it. It is because of all of you that I stayed

for many years longer than planned (for better or worse) and you all made my PhD experience one that I will look back on with fond memories.

Finally, I would like to thank my beautiful partner, Alyssa Budd. As fate would have it, my desk was assigned next to yours on the first day of my PhD. After many disagreements about lab protocols, I somehow didn't bug you too much and you are still here with me to this day.

Statement of the contribution of others

Nature of assistance	Contribution	Names and affiliations of co-contributors
Intellectual support	Editorial assistance - critically evaluated and provided feedback/edits on thesis chapters.	David G. Bourne (JCU), all chapters. Nicole S. Webster (AIMS), all chapters. David J. Miller (JCU), all chapters. Niko Andreakis (JCU), chapters 1-3. Pedro R. Frade (NHMW), chapters 2-4
	Project development – Contributed ideas and enhanced the scope of the project	David G. Bourne (JCU), Nicole S. Webster (AIMS), David J. Miller (JCU), Guojie Zhang (BGI), Pedro R. Frade (NHMW)
Data collection	Field assistance – provided assistance for sample collection and preservation.	David G. Bourne (JCU), Nicole S. Webster (AIMS), Pedro R. Frade (NHMW), Bettina Glasl (UV), Katharina Fabricius (AIMS), Heidi Luter (AIMS), Hillary Smith (JCU)
	Laboratory assistance – assisted with the development and optimisation of laboratory protocols. Assisted with DNA extractions	Pedro Frade (NHMW) & Kate Quigley (AIMS), metatranscriptomic protocol. Hillary Smith (JCU), DNA extractions (amplicon). Alyssa Budd (UWA), metagenomic protocol.
	Sequencing support – assisted with coordination of sequencing. Prepared libraries for genomic data	Shangjin Tan (BGI) and Qianyu Ye (BGI), library preparation. Guojie Zhang (BGI), sequencing coordination.
Data analysis	Phylogenetics – assisted with the reconstruction of host phylogenetic trees	Shangjin Tan (BGI), Niko Andreakis (JCU)

	Bioinformatics support – provided guidance on bioinformatic analysis pipelines and tools	Steven J. Robbins (UQ), Inka Vanwonderghem (UQ) and Patrick Laffy (AIMS), metatranscriptomics
Financial support	Sequencing costs	Beijing Genome Institute (BGI)
	Travel grants	AIMS@JCU
	Stipend	AIMS@JCU

Abstract

Marine invertebrates often host diverse microbial communities making it difficult to identify important symbionts and understand how these communities are structured. This complexity has also made it challenging to assign microbial functions and unravel the myriad of interactions among the microbiota. An enhanced understanding of the interactions between marine invertebrates and their microbial communities is urgently required as coral reefs face unprecedented local and global pressures and active restoration approaches, including manipulation of the microbiome, are proposed to improve the health and tolerance of reef species. In this thesis, I took a unique approach to address these knowledge gaps by studying the microbiome under concepts that relate to coevolution, i.e., the reciprocal evolution of interacting species. Specifically, I looked at the evidence for host-microbe coevolution in model systems and applied three research criteria for examining coevolution to complex marine invertebrate microbiomes; i) identifying a pattern of phylosymbiosis in diverse coral reef invertebrates, ii) assessing cophylogeny of host and microbe, and iii) confirming the intimate association between host and microbe based on microbial function and adaptation to the host environment. Since coevolution can occur through the interaction of a host and beneficial or parasitic symbionts, I hypothesised that studying the microbiome through this framework of coevolution would reveal critical insights into both microbiome assembly mechanisms and the functional attributes of those microbial taxa that contribute to host fitness.

If coevolution occurs between a host and its microbiota, the microbiota may show a degree of correlation with host phylogeny, an eco-evolutionary pattern known as phylosymbiosis. Using 16S rRNA gene sequencing to profile the microbiome, paired with COI and 18S rRNA host phylogenies, phylosymbiosis was investigated in four groups of coral reef invertebrates (scleractinian corals, octocorals, sponges and ascidians) (chapter 2). I tested three

commonly used metrics to evaluate the extent of phyllosymbiosis: a) intraspecific versus interspecific microbiome variation, b) topological comparisons between host phylogeny and hierarchical clustering (dendrogram) of host-associated microbial communities, and c) correlation of host phylogenetic distance with microbial community similarity. In all instances, intraspecific variation in microbiome composition was significantly lower than interspecific variation. Similarly, topological congruency between host phylogeny and the associated microbial dendrogram was more significant than would be expected by chance. Scleractinian corals, octocorals and sponges all showed a significant positive correlation between host phylogenetic distance and associated microbial dissimilarity. These findings provide new perspectives on the diverse nature of marine phyllosymbioses and the complex roles of the microbiome in the evolution of marine invertebrates.

Host-microbe coevolution may lead to strong cophylogenetic patterns between microbial lineages and their respective hosts. To investigate this, I employed the Procrustean Approach to Cophylogeny (PACo) on 16S rRNA gene derived microbial community profiles paired with COI, 18S rRNA and ITS1 host phylogenies (chapter 3). Secondly, I undertook a network analysis to identify groups of microbes that were co-occurring within the host species. Across twelve coral, ten octocoral and five sponge species, I found that Bacteria and Archaea affiliated to *Endozoicomonadaceae*, *Microtrichaceae*, *Thermoanaerobaculaceae*, *Spirochaetaceae* and *Nitrosopumilaceae* had the strongest cophylogenetic signals. Further, four co-occurring sub-networks were identified, each of which was dominant in a different host group. *Endozoicomonadaceae* and *Spirochaetaceae* ASVs were abundant among the sub-networks, particularly one sub-network that was exclusively comprised of these two bacterial families and dominated the octocoral microbiota. These results disentangle key microbial interactions that occur within complex microbiomes and reveal long-standing, essential microbial symbioses in coral reef invertebrates.

Host-microbe coevolution may be facilitated through specific microbial functional pathways. In the case of sponges, carbon fixation, nitrogen metabolism, sulfur metabolism and supplementation of B-vitamins have all been proposed as central microbial functions based on genomic data. However, transcriptomic validation of the putative symbiont pathways are rarely explored. To evaluate metagenomic predictions, I sequenced the metagenomes of three common coral reef sponges; *Ircinia ramosa*, *Ircinia microconulosa* and *Phyllospongia foliascens* and conducted metatranscriptomic sequencing of *I. microconulosa* and *P. foliascens* (chapter 4). Expression of entire pathways for carbon fixation and multiple sulfur compound transformations were observed in both sponges. Gene expression of complete suites of genes involved in nitrification, denitrification and nitrate reduction were observed in *I. microconulosa*, however for *P. foliascens* expression of only some denitrification pathway genes was observed. Across both sponges, expression of the biosynthetic pathways for B-vitamins was common and spread across many phyla, however in some cases only the partial pathway was retrieved, and key microbial taxa were needed for complete biosynthesis. Overall, this highlights new microbial taxa that may play important roles within the metabolism of the sponge holobiont and identify metabolic differences between sponge species.

Symbionts can show evidence of evolution towards a host-associated lifestyle through adaptive traits and functions that may come as a cost or benefit to either the host or symbiont. To identify these within the sponge microbiome, I used gene enrichment patterns between sponge symbionts and closely related microbes recovered from other environments (chapter 5). I included five families of prokaryotes that had strong cophylogenetic patterns in marine invertebrates identified in chapter 3, suggesting they have tight associations with their host; *Endozoicomonadaceae*, *Nitrosopumiliaceae*, *Spirochaetaeaceae*, *Microtrichaeaceae* and *Thermoanaerobaculaceae*. Interestingly, the well-known symbionts *Endozoicomonadaceae* and *Nitrosopumiliaceae* did not fall into monophyletic sponge clades and show very little gene

enrichment. The remaining microbial families all showed monophyletic sponge clades and exhibited a diverse range of gene enrichment patterns. Patterns of enrichment typically considered to reflect adaptations to the symbiotic lifestyle, including genes encoding eukaryote-like proteins and restriction modification systems, were also observed in sponge symbionts. Further, I found enrichment patterns for genes that protect against super oxide damage and others encoding sulfatases and sulfotransferases, which may assist in remodelling sulfated polysaccharides used by the sponge for cell aggregation. Additionally, the transport and metabolism of urea was enriched in sponge-associated Spirochaetes compared to those from free-living environments. These results suggest mechanisms by which symbionts have adapted to living in association with sponges and show that these microbes have their own unique set of symbiont characteristics.

In summary, this thesis used a unique approach to study the complex web of interactions between marine invertebrates and their microbial communities, providing evidence for coevolution by i) investigating the entire microbial assemblage and ii) considering individual microbial lineages and their relationship to host evolutionary history. Further, exploring the metabolic pathways of these invertebrate microbes provided support for microbial evolution towards a host-associated lifestyle. Through the light of coevolution, this thesis has deepened our understanding of the structure, function and importance of the marine invertebrate microbiome.

Table of Contents

Acknowledgements	ii
Statement of the contribution of others	iv
Abstract	vi
Table of Contents	x
List of Tables	xvi
List of Figures	xvii
Chapter 1: General Introduction. Host-microbe coevolution: applying evidence from model systems to complex invertebrate holobionts.....	1
1.1. Summary	2
1.2. Introduction.....	2
1.3. Untangling patterns of host-microbial coevolution in a web of microbes	1
1.3.1. Phylosymbiosis and neutral theory: Identifying stochastic and deterministic components of the microbiome.....	1
1.3.2. Codivergence: microbial phylogeny and host phylogeny are congruent	4
1.3.3. Metabolic Collaboration: Intimate association between host and microbe..	7
1.3.4. Core microbiome and the potential of viruses	9
1.4. Challenges, further considerations and conclusions.....	11
1.5. Thesis aims and chapter structure	14
Chapter 2: Diverse coral reef invertebrates exhibit patterns of phylosymbiosis	16
2.1. Summary	17
2.2. Introduction.....	17
2.3. Materials and Methods	19
2.3.1. Sample collection	19
2.3.2. Sample processing and preservation	20
2.3.3. DNA extraction and sequencing	21

2.3.4. 16S rRNA gene amplicon analysis.....	22
2.3.5. Characterisation of microbial diversity and composition.....	22
2.3.6. Host phylogenetic reconstructions	23
2.3.7. Phyllosymbiosis analysis	24
2.3.8. Data and code availability	25
2.4. Results	26
2.4.1. Sample collection and sequencing	26
2.4.2. Characterisation of microbial diversity and composition.....	26
2.4.3. Assessment of phyllosymbiosis among coral reef invertebrates	30
2.5. Discussion	35
2.5.1. Sponges demonstrated a strong signal of phyllosymbiosis	35
2.5.2. The signal of phyllosymbiosis was weaker in coral.....	36
2.5.3. The first direct evidence for phyllosymbiosis in octocorals.....	37
2.5.4. Inconsistent results for phyllosymbiosis in ascidians.....	38
2.5.5. Patterns of phyllosymbiosis occurred despite may uncontrolled factors	38
2.5.6. Conclusions.....	39
Chapter 3: Testing cophylogeny between coral reef invertebrates and their bacterial and archaeal symbionts.....	41
3.1. Summary.....	42
3.2. Introduction.....	42
3.3. Materials and Methods	45
3.3.1. Data collection and sequencing.....	45
3.3.2. Host sequence alignments, model selection and phylogenetic reconstructions	46
3.3.3. 16S rRNA data pre-processing and filtering.....	46
3.3.4. Cophylogenetic analysis using PACo and Moran's I.....	47

3.3.5. Co-occurrence network analysis	49
3.4. Results	49
3.4.1. Host sequence alignments, model selection and phylogenetic reconstructions	49
3.4.2. Coral reef invertebrates and their symbionts display a pattern of cophylogeny	50
3.4.3. Microbial groups with strong cophylogenetic signals.....	51
3.4.4. Distinct microbial co-occurrence sub-networks within each host group ...	55
3.5. Discussion	58
3.5.1. Cophylogenetic patterns suggest host-microbe codivergence and may contribute to patterns of phylosymbiosis.....	58
3.5.2. Cophylogenetic patterns reflect host generalist and host specialist microbial species	59
3.5.3. The strength of cophylogeny correlates with a host species tolerance to environmental stress.....	60
3.5.4. Five groups of microbes show a significant fit to cophylogeny when analysed independently.....	61
3.5.5. Limitations of cophylogenetic studies.....	63
3.5.6. Host groups are associated with unique sub-networks of co-occurring microbes	64
3.5.7. Conclusions.....	65
Chapter 4: Validation of key sponge symbiont pathways using genome-centric metatranscriptomics	66
4.1. Summary.....	67
4.2. Introduction.....	67
4.3. Materials and methods.....	70
4.3.1. Sample collection and preservation.....	70
4.3.2. Microbial cell enrichment by cell separation for metagenomics	71

4.3.3. Total DNA extraction for metagenomics	71
4.3.4. Total RNA extraction and microbial mRNA enrichment for metatranscriptomics	73
4.3.5. Library preparation and sequencing	74
4.3.6. Bioinformatics pipeline for metagenomic analysis	74
4.3.7. Bioinformatics pipeline for metatranscriptomic analysis	75
4.3.8. Metabolic pathways reconstruction	76
4.4. Results	77
4.4.1. Sequencing, metagenomic binning and taxonomy	77
4.4.2. Metatranscriptomic mapping results	79
4.4.3. Nutrient cycling in the sponge microbiome	81
4.4.4. Biosynthesis of B vitamins	84
4.5. Discussion	87
4.5.1. Autotrophic carbon fixation in sponge symbionts	87
4.5.2. Nitrogen metabolism is more common in the <i>Ircinia</i> sp. microbiome than <i>Phyllospongia foliascens</i>	88
4.5.3. Evidence of widespread sulphur metabolism in the sponge microbiome ..	90
4.5.4. Potential for supplementation of B-vitamins by sponge symbionts	91
4.5.5. Conclusions	92
Chapter 5: Comparative genomics identifies key adaptive traits of sponge associated microbial symbionts	95
5.1. Summary	96
5.2. Introduction	96
5.3. Materials & Methods	99
5.3.1. Laboratory methods and bioinformatics pre-processing	99
5.3.2. Genome curation, phylogeny and taxonomic classification	99

5.3.3. Enrichment analysis between sponge associated and non-sponge associated genomes	100
5.4. Results	101
5.4.1. Phylogeny and genome characteristics of sponge associated symbionts .	101
5.4.2. Overview of gene enrichment patterns in sponge associated and non-sponge associated microbial genomes	105
5.4.3. Avoiding the host immune response and infection from foreign DNA ...	106
5.4.4. Mechanisms of symbiont attachment and interaction with the host	108
5.4.5. Enrichment of genes related to the metabolism of carbohydrates	110
5.4.6. Additional enrichment patterns in genes of interest.....	112
5.5. Discussion	113
5.5.1. Sponge symbionts may escape phagocytosis using both eukaryote like proteins and superoxide dismutase	114
5.5.2. Sponge symbionts enriched in enzymes associated with the restriction modification system and CRISPR-cas system.....	116
5.5.3. Symbiont attachment through fibronectins and cadherins	117
5.5.4. Type II secretion systems may help degrade biopolymers in the extracellular matrix.....	118
5.5.5. Multienzyme degradation of carbohydrates including glycosyl hydrolases and sulfatases	118
5.5.6. Sponge associated Spirochaetaceae unique in their potential for urea degradation	120
5.5.7. High copy numbers of phytanoyl-CoA dioxygenase in sponge symbionts	121
5.5.8. Sponge symbionts investigated here likely represent a facultative symbiosis	121
5.5.9. Conclusions.....	122

Chapter 6: General discussion. Does the microbiome of coral reef invertebrates coevolve with their host?	124
6.1. Summary.....	125
6.2. Coral reef invertebrate microbiomes are a complex web of interactions	125
6.3. Future research priorities	128
6.4. Concluding remarks: Why it's important to understand whether coral reef invertebrates coevolve with their microbiome.....	130
References	133
Appendix A. Supporting tables and figures for chapter 2.....	178
Appendix B. Supporting tables and figures for chapter 3	193
Appendix C: Supporting tables and figures for chapter 4.....	204
Appendix D. Supporting tables and figures for chapter 5.....	223

List of Tables

Table 2.1. Normalised Robinson-Foulds (nRF) and mantel statistics across Bray-Curtis, weighted and unweighted UniFrac beta-diversity metrics. Non-significant values are highlighted in grey. 32

Table 3.1. Filtering and PACo model results for core ASVs associated with each host group and all samples combined. Results for PACo model are obtained from the filtered core data and mean p-value is calculated following a Benjamini-Hochberg adjustment for multiple comparisons. FI refers to results using the fragment-insertion phylogeny while DN refers to the de novo phylogeny. 50

Table 3.2 Microbial genera and family that had a significant fit to the cophylogenetic model using both phylogenetic methods. Microbial taxa are ordered by mean sums of squares, with a lower value indicating a stronger fit. Mean p-value represents mean adjusted p-value of the 100 data subsets. Taxonomic classification is based on the Silva database. 54

Table 3.3. Summary of the number of ASVs in each sub-network community and the percentage of those ASVs that were found to be significantly related to host phylogeny..... 58

Table 5.1. Statistical summary for genome size and GC content for sponge associated and non-sponge associated genomes in all five microbial groups. Std er. = standard error of the mean, min = minimum value, max = maximum value. 105

Table 5.2. Overview of the total number of enriched genes for sponge associated and non-sponge associated genomes for all five microbial groups. Enriched genes are included if the number of genomes that contain the gene within each group are statistically different. CAZY = Carbohydrate active enzymes, KO = genes annotated using the Kegg Orthology database, PFAM = genes annotated using the Protein Families database..... 106

List of Figures

Figure 1.1. Spectrum of microbial diversity associated with different compartments of marine invertebrates. Microbial associations may involve a single symbiont in a specialised organ, or over 1000 operational taxonomic units (OTUs) associated with tissues. OTUs reported are the highest recorded in the referenced study for that species. Diversity may vary significantly within the same species across different studies. 6

Figure 1.2. Hypothetical scenario addressing three criteria for host-microbial coevolution in species A-D. **a)** Phylosymbiosis is shown through hierarchical clustering of the microbial community, resulting in a microbial dendrogram which mirrors host phylogeny. **b)** Neutral model shows the expected occurrence of microbes based on neutral population dynamics (blue line). As the relative abundance increases so too does the occurrence in host samples. *Bacteria spp. 1* is therefore more abundant than would be expected by chance and may indicate active selection, while *Bacteria spp. 2* is less abundant. **c)** Codivergence of *Bacteria spp. 1* with its hosts. *Bacteria spp. 1* is found within the microbial community of each host species and appears to be actively selected for. Its phylogeny indicates a host split at the strain level followed by diversification within each host species. Congruency between host and microbial lineages suggest important host-microbial interactions and warrant further investigation. **d)** Metabolic collaboration between Host *spp. A* and *Bacteria spp. 1*. Fluorescence *in-situ* hybridization (FISH) confirms *Bacteria spp. 1* is located within bacteriocyte cells in the tissues of Host *spp. A*. Genome and transcriptome data for each species suggest the amino acid cysteine is produced through a shared metabolic pathway between host and microbe. In corals from the genus *Acropora* for example, the genome is incomplete for biosynthesis of cysteine and presents a potential pathway for host and microbe to collaborate (Shinzato et al. 2011). Hypothetically, amino acids homocystesine and serine (potentially sourced from host diet and metabolism) are combined to form cystathionine through the enzyme Cystathionine V-synthase provided by the host's endosymbiont. The host enzyme cystathionine γ -lyase then breaks down cystathionine to form cysteine. 1

Figure 2.1. ASV richness (top panel) and Shannon-Wiener diversity index (bottom panel) for each invertebrate group and seawater. Letters indicate groups which are significantly different from each other. 27

Figure 2.2. Relative abundance of the top 25 prokaryotic families found across each invertebrate group as well as seawater and blank extractions.	29
Figure 2.3. Bray-Curtis dissimilarity based on microbial composition visualised using NMDS. Each symbol represents a sample where colour is the associated host and shape is reef zone where sample was collected.	30
Figure 2.4. Intraspecific and interspecific Bray-Curtis dissimilarity scores for each invertebrate group.	31
Figure 2.5. Host phylogeny inferred from COI, 18S rRNA and ITS1 sequences compared to a microbial dendrogram based on Bray-Curtis dissimilarity for microbial composition of each host species. <i>Cladiella sp.</i> was used as an outgroup for 5A) sponges, while <i>C. foliascens</i> was used as an outgroup for 5B) coral, 5C) octocoral and 5D) ascidians. Numbers at nodes reflect posterior probability for clade support in the host tree and jackknife support values in dendrograms. Branch tips are coloured to reflect clades in host phylogeny. Initials in brackets next to species names refer to collection site. <i>BR</i> = Broadhurst Reef, <i>DR</i> = Davies Reef, <i>OR</i> = Osprey Reef, <i>PI</i> = Palm Islands (Orpheus and Pelorus), <i>PR</i> = Pandora Reef, <i>RB</i> = Rib Reef, <i>RR</i> = Ribbon Reefs.	34
Figure 3.1. A presence-absence heatmap of each core ASV found across all host species samples. Rows are ordered by ASV phylogeny (y-axis dendrogram) and columns are order by host phylogeny (x-axis dendrogram). ASV phylogeny is coloured by phylum level classification while host phylogeny is coloured by phylum for sponges and subclass for corals and octocorals. Branches highlighted in red indicate ASVs that were significantly distributed by host phylogeny using local Moran’s <i>I</i> spatial autocorrelation.	52
Figure 3.2. Residual contributions of each host-microbe link to the overall fit of cophylogeny. Boxplots are grouped by host species and residuals are overlaid and coloured by microbe phylum. Where a species was collected from two locations, ‘RR’ indicates the replicate from the ribbon reefs.	53
Figure 3.3. Taxonomic classification of ASVs that were significantly distributed by host phylogeny using local Moran’s <i>I</i> measure of spatial autocorrelation. Occurrence refers to the number of host-ASV associations. <i>Endozoicomonadaceae</i> contains seven unique ASVs while the remaining families all contained one each.	55

Figure 3.4. Microbial cladogram depicting the topology of microbial families that show a highly significant fit to cophylogeny with significant Moran's *I* ASVs highlight in red. Bubble plot on the right side of the cladogram shows the relative abundance of each ASV grouped by host species including seawater controls and coloured by host group. Dendrogram on the x-axis depicts host species phylogeny (left side) and right side x-axis shows species names..... 56

Figure 3.5. Four sub-networks of positively co-occurring ASVs within the core microbiome across all species. A) Presence-absence co-occurrence network with node shape reflective of each sub-network, node colour reflecting the phylum of the ASV node, and node size indicating centrality betweenness, i.e., how often the shortest path between two nodes traverses the focal node. B) Presence of each sub-network within each host species. Circles represent the mean fraction of host species replicates the sub-network ASVs are found, lines represent the standard deviation. Where the same species has been collected from two locations, those collected from the Ribbon Reefs are indicated with 'RR'. 57

Figure 4.1. Phylogeny of 415 MAGs isolated from *I. ramosa*, *I. microconulosa* and *P. foliascens* illustrating the large diversity of microbial taxa found across the three sponge species. Tree clades are coloured by microbial phylum while outer heatmap indicates the relative abundance of each MAG in each sponge sample..... **Error! Bookmark not defined.**

Figure 4.2. Relative abundance of metatranscriptomic reads that successfully aligned to the reference MAGs illustrating a larger diversity of microbes are active within the *I. microconulosa* microbiota compared to *P. foliascens*. Reads are grouped by microbial phylum and the size of the circle is proportional to the relative abundance. *I. microconulosa* is represented in blue while *P. foliascens* is represented in yellow..... 81

Figure 4.3. Overview of complete and near complete (>80%) gene expression for metabolic pathways found within the microbiomes of *Ircinia microconulosa* and *Phyllospongia foliascens*. The microbiota of both sponges showed expression for multiple transformations of sulfur along with carbon fixation and B-vitamin synthesis. However, only the *I. microconulosa* microbiota showed expression of nitrogen cycling. Arrows indicate proposed transfer of nutrients within the holobiont. A '+' symbol is used to indicate where microbes from multiple taxonomic classifications are needed to complete a pathway. B-vitamin synthesis pathways are numerically labelled adjacent to the corresponding microbe and those

labelled with an asterisk indicate pathways only found in *I. microconulosa* while open circles indicate pathways only found in *P. foliascens*..... 86

Figure 5.1. Phylogenomic tree of all sponge and non-sponge associated microbial genomes. Sponge-associated clades can be seen for *Microtrichaceae*, *Spirochaetaceae* and *Thermoanaerobaculaceae* while *Endozoicomonadaceae* and *Nitrosopumiliaceae* show multiple sponge-associated lineages. Branch tips in blue indicate sponge associated genomes while branch tips in black indicate non-sponge associated genomes. Clades are coloured by microbial family..... 103

Figure 5.2. Genome size and GC content for all sponge associated and non-sponge associated genomes for each microbial group. In general, sponge associated symbionts do not have smaller genomes than free-living relatives, indicating a facultative symbiont lifestyle. Last panel shows genome completeness for sponge associated and non-sponge associated genomes for each microbial group..... 104

Figure 5.3. Phylogenomic relationships and gene enrichment patterns in sponge associated and non-sponge associated Spirochaetaceae genomes. Branches with blue labels indicate sponge associated genomes, branches with black labels indicate non-sponge associated genomes while grey labels indicate outgroups and weren't included in the enrichment analysis. Numbers at nodes indicate branch support values using the Shimodaira-Hasegawa test. Heatmap indicates copy numbers for a selection of genes in key symbiotic signatures. Data that has been square root transformed for illustration is indicated with (sqrt). Abbreviations: ANK = Ankyrin repeat proteins, R-M = restriction modification system enzymes, CAZy = carbohydrate active enzymes, Sulfur = sulfatases and sulfatransferases. 107

Figure 5.4. Phylogenomic relationships and gene enrichment patterns in sponge associated and non-sponge associated Thermoanaerobaculaceae genomes. Branches with blue labels indicate sponge associated genomes, branches with black labels indicate non-sponge associated genomes while grey labels indicate outgroups and weren't included in the enrichment analysis. Numbers at nodes indicate support values. Heatmap indicates copy numbers for a selection of genes in key symbiotic signatures. Abbreviations: ANK = Ankyrin repeat proteins, R-M = restriction modification system enzymes, CAZy = carbohydrate active enzymes, Sulfur = sulfatases and sulfatransferases 109

Figure 5.5. Phylogenomic relationships and gene enrichment patterns in sponge associated and non-sponge associated Microtrichaceae genomes. Branches with blue labels

indicate sponge associated genomes, branches with black labels indicate non-sponge associated genomes while grey labels indicate outgroups and weren't included in the enrichment analysis. Numbers at nodes indicate support values. Heatmap indicates copy numbers for a selection of genes in key symbiotic signatures. Abbreviations: ANK = Ankyrin repeat proteins, R-M = restriction modification system enzymes, Sulfur = sulfatases and sulfatransferases, BCAA = branch chain amino acid..... 111

Figure 6.1. Overview of the three major lines of evidence that support host-microbe coevolution in sponges (adapted from Figure 1.2). The microbiome of the host is reflected in the host phylogeny (phylosymbiosis), a subset of those microbes show evidence of codivergence (for example, the *Spirochaetaceae*), the genomes of these microbes suggest an adaptation to the host with unique functions (for example, urea metabolism)..... 131

**Chapter 1: General Introduction. Host-microbe coevolution:
applying evidence from model systems to complex invertebrate
holobionts**

This chapter is published as:

O'Brien, P. A., Webster, N. S., Miller, D. J., & Bourne, D. G. (2019). Host-microbe coevolution: applying evidence from model systems to complex marine invertebrate holobionts. *MBio*, *10*(1), e02241-18

1.1. Summary

This introductory chapter reviews the literature on host-microbe coevolution and in doing so, provides a framework for exploring host-microbe coevolution in hosts with complex and diverse microbiomes. It outlines three main themes to focus on, where each theme addresses coevolution at a different level of the microbial community; 1) the overall microbial community, 2) the individual microbial lineage and 3) the interaction between host and microbe. Firstly, phylosymbiosis investigates coevolution considering the associated microbial community as a unit of selection. Secondly, codivergence focuses on individual microbial lineages that parallel the host phylogeny. Thirdly, metabolic collaboration focuses on the interaction between host and microbe. By studying the microbiome under this framework of coevolution, this thesis applies a unique approach to understand how the microbiota of marine invertebrates are structured, identifies microbial functions that are critical to the host and uncovers how coevolved symbionts have adapted to the host.

1.2. Introduction

Coevolution theory dates back to the 19th century (Box 1), and is referred to as the reciprocal evolution of one lineage in response to another (Zaneveld et al., 2008). This definition encompasses a broad range of interactions such as predator-prey, host- symbiont or host-parasite, or a community of organisms such as a host and its associated microbiome (Van den Abbeele et al., 2011; Zaneveld et al., 2008). In the case of host-microbial associations, this has produced some of the most remarkable evolutionary outcomes that have shaped life on Earth, such as the eukaryotic cell, multicellularity and the development of organ systems (Archibald, 2015; McFall-Ngai et al., 2013). It is now recognised that microbial associations with a multicellular host are the rule rather than the exception (McFall-Ngai et al., 2013), but in complex associations of this kind, the extent to which coevolution operates is often unclear.

Box 1. A brief history of coevolution

Charles Darwin once explained the sudden and rapid diversification of flowering plants as an “abominable mystery”, since it could not be explained by traditional views of evolution alone (Friedman, 2009). While his correspondent, Gaston de Saporta speculated that a biological interaction between flowering plants and insects might be the cause of the phenomenon, it wasn’t until nearly 100 years later that the concept of coevolution developed. In a pioneering study, Ehrlich & Raven (Ehrlich & Raven, 1964) had observed that related groups of butterflies were feeding on related groups of plants, and speculated this was due to a process they coined “coevolution”. Using butterflies, they argued that plants had evolved mechanisms to overcome predation from herbivores, which have in turn evolved new ways to prey on plants. Decades on, the introduction of phylogenetics has shown that plants had evolved in the absence of butterflies, which colonised the diverse group of plants after their chemical defences were already in place (Janz & Nylin, 1998). Nevertheless, the theory of coevolution was endorsed, and two important points came to light. Firstly, care must be taken when inferring coevolution from seemingly parallel lines of evolution, and where possible, divergence times and common ancestry should be included. Secondly, coevolution can occur between communities of organisms (‘guild’ coevolution) as observed in the case of flowering plants where predation and pollination from a wide variety of insects likely influenced the diversification of angiosperms (Ryan & Byrne, 1988).

Since coevolution can occur across multiple levels of interactions, multiple theories have also developed. The Red Queen theory is based on the concept of antagonistic coevolution, and assumes that an adaptation that increases the fitness of one species will come at the cost to the fitness of another (Van Valen, 1974). This type of coevolution has been most pronounced in host-parasite interactions, where the antagonistic interactions are closely coupled (Paterson et al., 2010). However, coevolutionary patterns may also arise in the case of mutualistic symbioses, which require reciprocal adaptations to the benefit of each partner (Herre et al., 1999). Mutualistic coevolution is associated with a number of key traits that will be discussed further in this chapter, such as obligate symbiosis, vertical inheritance and metabolic collaboration. Lastly, coevolution has also recently been placed in context of the hologenome theory (Theis et al., 2016), which suggests that the holobiont can act as a unit of selection (but not necessarily the primary unit) since the combined genomes influence host phenotype on which selection may operate (Bordenstein & Theis, 2015; Zilber-Rosenberg & Rosenberg, 2008). However, hologenome theory also acknowledges that selection acts on each component of the holobiont individually as well as in combination with others (including the host). Thus, the entity that is the hologenome may be formed, in part, through coevolution of interacting holobiont compartments, in addition to neutral processes (Theis et al., 2016).

With the ubiquitous nature of host-microbial associations and the huge metabolic potential microorganisms provide, it is not surprising that evidence of host-microbial coevolution is emerging. Model representatives of both simple and complex associations are being used to study coevolution, allowing researchers to look for specific traits, signals and patterns (Wilson & Duncan, 2015; Zaneveld et al., 2008). A well-known model system is the pea-aphid and its endosymbiotic bacteria in the genus *Buchnera*. This insect has evolved specialized cells known as bacteriocytes to host its endosymbionts, which in turn synthesise and translocate amino acids that are missing from the pea-aphids diet (Baumann et al., 1997).

Amino acid synthesis occurs through intimate cooperation between host and symbiont, with some pathways missing from the host and some from the symbiont, so this relationship is obligate to the extent that one organism cannot survive without the other (Russell et al., 2013). In complex systems, the human gut microbiome has been extensively studied and shown to be intimately associated with human health. Through metabolic processes, such as microbial regulation of the essential amino acid tryptophan, gut microbes are now linked with human behaviour and development (Collins et al., 2012; Kennedy et al., 2017). The human microbiome contains around 150-fold more non-redundant genes compared to the human genome (Qin et al., 2010), and the metabolic capacity of microbes residing in the intestine is believed to have been a driving evolutionary force in the host-microbial coevolution of humans (Van den Abbeele et al., 2011). In these examples, and many others (Brune & Dietrich, 2015; Fenn & Blaxter, 2004; D. Wu et al., 2006), host and symbiont have both evolved to maintain and facilitate the symbiosis. Furthermore, phylogenies of host and symbiont in these systems are often mirrored, indicating that host and symbiont are diverging in parallel (Baumann et al., 1997; Clark et al., 2000; Moeller et al., 2016), a phenomenon known as codivergence (Moran, 2006).

In the marine environment, invertebrates can host microbial communities as simple and stable as that of the pea-aphid, or as complex and dynamic as that of the human gut (Figure 1.1). The Hawaiian bobtail squid for example, maintains an exclusive symbiosis with a single bacterial symbiont which it hosts within a specialised light organ (McFall-Ngai, 2008). On the other hand, corals host enormously diverse microbial communities, comprising thousands of species-level operational taxonomic units (OTUs), which are often influenced by season, location, host health and host genotype (C. Chen et al., 2011; Gil-Agudelo et al., 2006; Koren & Rosenberg, 2006; Littman et al., 2009). Marine sponges also host complex microbial communities with diversity comparable to corals (Thomas et al., 2016), but with associations

that are generally far more stable in space and time (Webster & Thomas, 2016). Less diverse microbial communities are found in the sea anemone *Aiptasia*, where the number of OTUs is generally in the low hundreds (Röthig et al., 2016). Due to *Aiptasia*'s close taxonomic relationship with coral and its comparatively simple microbial community, it has been proposed as a model organism for studying coral microbiology and symbiosis (Röthig et al., 2016). Some marine invertebrates also include species along a continuum of microbial diversities. Ascidians, for example, have been shown to host below 10 (*Polycarpa aurata*) or close to 500 (*Didemnum sp.*) microbial OTUs within their inner tunic (Erwin et al., 2014). Furthermore, species with low microbial diversity such as *P. aurata* can exhibit high intraspecific variation, with as few as 8% of OTUs shared among individuals of the same species (Erwin et al., 2014). Taken together, these studies highlight the vast spectrum of associations that marine invertebrates form with microbial communities in terms of diversity, composition and stability (Figure 1.1).

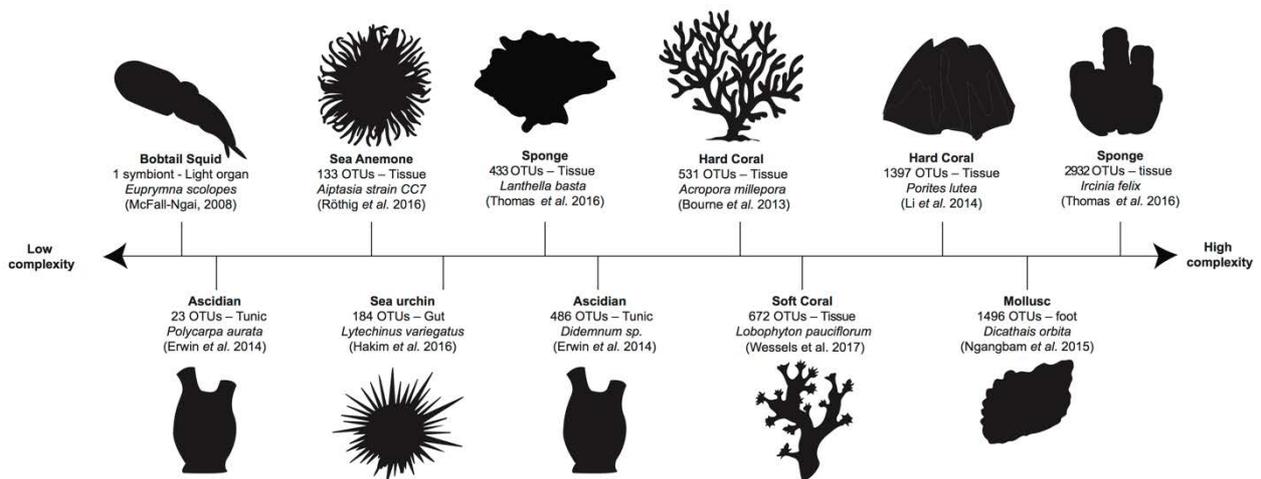


Figure 1.1. Spectrum of microbial diversity associated with different compartments of marine invertebrates. Microbial associations may involve a single symbiont in a specialised organ, or over 1000 operational taxonomic units (OTUs) associated with tissues. OTUs reported are the highest recorded in

the referenced study for that species. Diversity may vary significantly within the same species across different studies.

While previous research has provided a good understanding of the composition of marine invertebrate microbiomes, our understanding of how the microbiome interacts with the host, and the potential to coevolve, is far more limited. Moreover, the increasing number of studies generating tremendous volumes of host-associated microbiome sequence data requires theoretical development to interpret these relationships. Coevolved microbial symbionts are presumed to be intimately linked with host fitness and metabolism (Moran & Baumann, 2000), therefore understanding these relationships in marine invertebrates will have direct implications for health and disease processes in these animals. Three research criteria arise for examining coevolution in marine invertebrates; i) identifying stochastic and deterministic microbial components of the microbiome, ii) assessing codivergence of host and microbe, and iii) confirming an intimate association between host and microbe related to shared metabolic function (metabolic collaboration). While each of these criteria may be fulfilled without the involvement of coevolution (Douglas & Werren, 2016; Moran, 2006; Moran & Sloan, 2015), evidence of their existence in combination provides a strong basis for establishing coevolution patterns (Figure 1.2.). This chapter positions these three criteria in coevolution as a complementary approach to study complex marine invertebrate microbiomes by drawing from examples of model systems. Focussing on keystone coral reef invertebrates, this chapter also evaluates the current evidence for each criterion. Finally, while parasites and pathogens also contribute to host coevolution, the focus of this chapter is mutualistic symbionts and thus pathogens and parasitism will not be discussed.

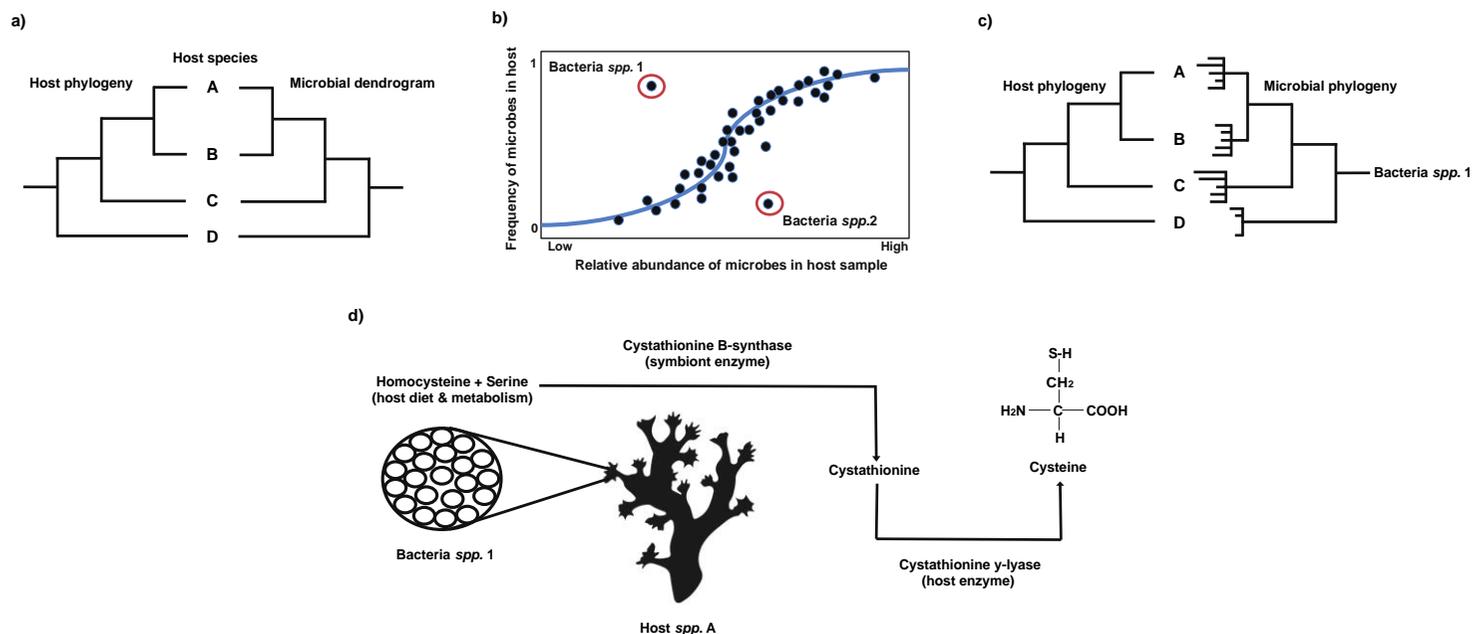


Figure 1.2. Hypothetical scenario addressing three criteria for host-microbial coevolution in species A-D. **a)** Phyllosymbiosis is shown through hierarchical clustering of the microbial community, resulting in a microbial dendrogram which mirrors host phylogeny. **b)** Neutral model shows the expected occurrence of microbes based on neutral population dynamics (blue

line). As the relative abundance increases so too does the occurrence in host samples. *Bacteria spp. 1* is therefore more abundant than would be expected by chance and may indicate active selection, while *Bacteria spp. 2* is less abundant. **c)** Codivergence of *Bacteria spp. 1* with its hosts. *Bacteria spp. 1* is found within the microbial community of each host species and appears to be actively selected for. Its phylogeny indicates a host split at the strain level followed by diversification within each host species. Congruency between host and microbial lineages suggest important host-microbial interactions and warrant further investigation. **d)** Metabolic collaboration between *Host spp. A* and *Bacteria spp. 1*. Fluorescence *in-situ* hybridization (FISH) confirms *Bacteria spp. 1* is located within bacteriocyte cells in the tissues of *Host spp. A*. Genome and transcriptome data for each species suggest the amino acid cysteine is produced through a shared metabolic pathway between host and microbe. In corals from the genus *Acropora* for example, the genome is incomplete for biosynthesis of cysteine and presents a potential pathway for host and microbe to collaborate (Shinzato et al. 2011). Hypothetically, amino acids homocysteine and serine (potentially sourced from host diet and metabolism) are combined to form cystathionine through the enzyme Cystathionine V-synthase provided by the host's endosymbiont. The host enzyme cystathionine γ-lyase then breaks down cystathionine to form cysteine.

1.3. Untangling patterns of host-microbial coevolution in a web of microbes

1.3.1. Phyllosymbiosis and neutral theory: Identifying stochastic and deterministic components of the microbiome

Host-microbial coevolution may occur to some degree at the level of the hologenome, i.e., reciprocal evolution of the host genome and microbiome. Therefore, it is necessary to understand microbial community structure and population dynamics within the host environment. This may illustrate that a) the microbiome associated with a host is structured through phylogenetically related host traits and may therefore retain a host phylogenetic signal (phyllosymbiosis), and b) certain microbes deviate from the expected patterns of neutral population dynamics, i.e., stochastic births, deaths and immigration. It is likely that phyllosymbiosis and neutral population dynamics are linked, therefore their potential to contribute to coevolution is discussed together.

The term phyllosymbiosis is not intended to imply coevolution (Douglas & Werren, 2016; Theis et al., 2016), however coevolution of a host and microbiome may reinforce patterns of phyllosymbiosis. There are many host traits that correlate with host phylogeny, some of which can act as environmental filters, preventing the establishment of microbes in the host environment. Thus, neutral population dynamics, with host traits acting as an ecological filter to microbial immigration, may be sufficient to result in phyllosymbiotic patterns (Mazel et al., 2018; M. Sieber et al., 2018). However, host traits are not static, and thus the evolution of these microbial niches may further drive the radiation of the microbes that reside within them. In turn, the continuous colonisation over many generations of a microbial community likely adds to the selective pressure on host traits. Therefore, ecological filtering of microbes through host traits and coevolution of a host and microbiome need not be mutually exclusive in the appearance of phyllosymbiosis (Mazel et al., 2018). Moreover, assessing patterns of

phylosymbiosis and neutral population dynamics also allows for the detection of microbes that deviate from these patterns, and may identify important microbial species that are actively selected for (or against) by the host. In this context, neutral models can simulate expected microbial abundance allowing for easier detection of microbes that do not fit these patterns (Sloan et al., 2006). This reasoning justifies why phylosymbiosis and microbial population dynamics should be considered when assessing coevolution in complex holobionts.

Patterns of phylosymbiosis are frequently detected in complex holobionts. One particular study tested for phylosymbiosis across 24 species of terrestrial animals from 4 groups that included *Peromyscus* deer mice, *Drosophila* flies, mosquitos and *Nasonia* wasps, and an additional data set of 7 hominid species (Brooks et al., 2016). Since these animals could be reared under controlled laboratory conditions (with the exception of hominids), environmental influences could be eliminated, leaving the host as the sole factor influencing the microbial community. Under these conditions, phylosymbiotic patterns were clearly observed for all five groups, with phylogenetically related taxa sharing similar microbial communities, and microbial dendrograms mirroring host phylogenies. Similar patterns of phylosymbiosis have been observed in a growing number of terrestrial systems including all five gut regions in rodents (Kohl et al., 2018), the skin of ungulates (Ross et al., 2018), the distal gut in hominids (Ochman et al., 2010a) and roots of multiple plant phyla (Yeoh et al., 2017), providing evidence that such patterns are common among host-associated microbiomes.

In the marine environment, two major studies have provided the most convincing examples of phylosymbiosis; one involving 236 colonies across 32 genera of scleractinian coral collected from the east and west coasts of Australia (Pollock et al., 2018), and the other involving 804 samples of 81 sponge species collected from the Atlantic, Pacific and Indian oceans and the Mediterranean and Red Seas (Thomas et al., 2016). Both studies found a

significant evolutionary signal of the host on microbial diversity and composition. Specifically, mantel tests were used to delineate that closely related corals and sponges hosted more similar microbial communities in terms of composition than would be expected by chance. In the case of corals, this was most pronounced in the skeleton and to a lesser extent in the tissue microbiome, while the mucus microbiome was most influenced by the surrounding environment (Pollock et al., 2018). However, both studies found that host species was the strongest factor in explaining dissimilarity among microbial communities. Additional studies on both cold water and tropical sponges have found similar phylogenetic patterns within the microbiome of the host species (Easson & Thacker, 2014; Schöttner et al., 2013). Together, these results suggest that host phylogeny (or associated traits) has a significant role in structuring associated microbial communities, although there are additional factors related to host identity (and unrelated to phylogeny) that also likely play a major role.

Most studies to date have focussed on the microbes that adhere to these patterns of phylosymbiosis, though arguably more useful information may also lie in the microbes that do not. Since phylosymbiosis is a pattern that shows correlation between the microbiome dissimilarity and host phylogeny, it does not indicate active microbial selection, nor does it indicate co-speciation (Douglas & Werren, 2016), and the species that deviate from these patterns would be interesting targets for codivergence and metabolic collaboration (see below). Neutral models have been applied to three species of sponges, a jellyfish and a sea anemone, and while neutral models fit well to the expectation of microbial abundance in sponges (which also show phylosymbiosis), jellyfish and sea anemone microbiomes were associated with a higher level of non-neutrality (M. Sieber et al., 2018). Potential reasons of non-neutrality include the presence of a more sophisticated immune system in cnidarians providing active selection on certain microbial taxa, or that the microbiome in these cases are more transient, or a combination of these. In summary, neutral population dynamics filtered through

phylogenetically related host-traits likely result in, or at least contribute to, the observed patterns of phyllosymbiosis. This does not necessarily mean that this pattern is unimportant or not contributing to coevolution at the hologenome level, and it may be that the community of microbes that follow these patterns are responsible for broad ecological functions (Rivett & Bell, 2018). On the other hand, microbes that deviate from these patterns may be responsible for more specific functions, and will be of high interest to those trying to identify symbionts and coevolution at the microbial species or strain level.

1.3.2. *Codivergence: microbial phylogeny and host phylogeny are congruent*

The second criterion in assessing host-microbial coevolution is whether individual microbial lineages have matching phylogenies with the host (Clark et al., 2000; Fenn & Blaxter, 2004; Nishiguchi et al., 1998). Codivergence implies a tightly coupled, long-term interaction between two species, and can potentially identify beneficial symbionts (or parasites) that have coevolved with the host (Moran, 2006). However it is also important to recognise that codivergence can arise due to processes other than coevolution, such as one species adaptively tracking another which would imply evolution is not reciprocal, or two species responding independently to the same speciation event or environmental stress (Moran & Sloan, 2015). In known cases of coevolution, phylogenies of hosts and their microbial symbionts are congruent (Bandi et al., 1998; Baumann et al., 1997; Nishiguchi et al., 1998). However, in complex and uncharacterised systems, this strategy can be reversed to identify potential symbionts. Therefore, the main value of investigating codivergence in complex associations is to identify those specific microbes on which to focus further attention.

Codivergence has been demonstrated in the case of *Hydra viridissima*, a freshwater relative of marine cnidarians, and its photosymbiont *Chlorella* (Deines & Bosch, 2016). In this system, photosynthetically fixed carbohydrates from *Chlorella* are transported to its host

(Mews, 1980), and phylogenetic analysis of 6 strains of *Hydra viridissima* and their vertically transmitted symbionts resulted in clear congruency of host and symbiont topologies (Kawaida et al., 2013). In more complex systems, patterns of codivergence have been illustrated in the gut microbiota of hominids (Moeller et al., 2016). Analysis of faecal samples from humans, wild chimpanzees, wild bonobos and wild gorillas, showed that four clades of bacteria from the dominant families *Bacteroidaceae* and *Bifidobacteriaceae* codiverged with host phylogeny. Importantly, this example illustrates one possible way of identifying codivergence in complex holobionts where the symbionts are unknown. Since bacteria from the families *Bacteroidaceae*, *Bifidobacteriaceae* and *Lachnospiraceae* are known to dominate the gut of hominids, multiple primer sets targeting each individual family were utilised, and phylogenetic analysis of each family was completed independently. Furthermore, instead of using the relatively slowly diverging 16S rRNA gene, the fast evolving and variable gene, DNA gyrase subunit B was used for bacteria phylogenetics. Similar methods may be applied to complex marine invertebrates such as coral and sponges, where 16S rRNA gene studies have identified prominent bacteria.

Within complex marine invertebrate holobionts, codivergence has been most clearly demonstrated in cold-water sponges in the family Latrunculiidae. The microbiomes of six species within this family were each dominated by a single betaproteobacterial OTU, and the phylogeny of this OTU was highly congruent with that of the host (Matcher et al., 2017). Furthermore, gene expression analysis suggests that the dominant betaproteobacteria are active members of the microbiome as opposed to dormant or non-viable, however whether or not this potential symbiont and its host show metabolic collaboration is unknown and highlights an example warranting further investigation. The microbiomes of many other marine invertebrates are dominated by members of the genus *Endozoicomonas* (Neave et al., 2016). A pan genomic analysis of the genomes of seven *Endozoicomonas* strains representing a broad range of hosts

(corals, sponges and sea slugs) provided some evidence for codivergence (Neave, Michell, et al., 2017). Strikingly, the two closely related corals, *Stylophora pistillata* and *Pocillopora verrucosa*, hosted *Endozoicomonas* with highly similar genomes. A second, large-scale study (Pollock et al., 2018), found that *Endozoicomonas* within the coral tissues showed strong signals of codivergence with its host, however were grouped into two major divisions; host specific and host generalist. The presence of a host generalist clade may partly explain why patterns of codivergence did not hold when samples of *Stylophora pistillata* and *Pocillopora verrucosa* were collected across 28 reefs world-wide (Neave, Rachmawati, et al., 2017). Furthermore, the genome of *Endozoicomonas* is large and appears to be adapted to a planktonic lifestyle (Neave et al., 2016). Having a free-living stage to the *Endozoicomonas* lifecycle suggests a facultative relationship with corals and will limit the extent of codivergence.

Codivergence may also occur between two symbionts within the microbial community associated with a single host. An interesting example occurs in lower termites, which live in a symbiotic relationship with flagellate protozoa that are essential for the breakdown of lignocellulose obtained from wood particles (Brune, 2014). Within the hindgut, these flagellate protozoa are associated with endosymbiotic prokaryotes, and while the functional basis of this relationship is unclear, matching phylogenies of flagellate host and prokaryote symbiont indicate codivergence (Ikeda-Ohtsubo & Brune, 2009). The microbiomes of many marine invertebrates also include both eukaryotes and prokaryotes that appear to closely interact with one another. For example, the symbiotic algae Symbiodiniaceae, which reside in the endoderm of the coral tissue, are producers of dimethylsulfoniopropionate (DMSP) that is likely metabolised by bacteria within the holobiont (Raina et al., 2009). Symbiodiniaceae and bacteria are also linked through the nitrogen cycle, where diazotrophs within the holobiont are postulated to fix nitrogen such that it can be used by the endo-symbiotic algae (Lema et al., 2012; Rädcker et al., 2015). Furthermore, the potential for a core microbiome associated with

Symbiodiniaceae appears likely, with bacteria affiliated to *Marinobacter*, *Labrenzia* and *Chromatiaceae* present across 18 cultures of Symbiodiniaceae spanning 5 genera (Lawson et al., 2018). A range of other marine invertebrates, including soft corals, sponges and molluscs, also host Symbiodiniaceae and it would be valuable to investigate whether Symbiodiniaceae show codivergence and coevolution with prokaryotes in these systems.

1.3.3. Metabolic Collaboration: Intimate association between host and microbe

A third key feature of coevolution is that host and microbe collaborate in a way that is mutually beneficial (Wilson & Duncan, 2015). This is often related to the metabolic function of the microbe, with the host facilitating or complementing that function. This could be a specialised cell or organ to host microbial symbionts (McFall-Ngai, 2008), a shared metabolic pathway to produce essential vitamins or amino acids (Russell et al., 2013), or microbial regulation of certain metabolites produced by the host (Kennedy et al., 2017). Metabolic collaboration should be validated where potential candidates for coevolution have been identified through population dynamics and codivergence, as reciprocal evolution necessitates an interaction between the two species. A key step demonstrating an interaction, and therefore identifying potential reciprocal evolution, is to look at the genome and transcriptomes of the host and symbionts for evidence of integrated metabolism, combined with targeted *in situ* visualisation of metabolite passage to support metabolic collaboration.

Sharpshooters, a group of xylem-feeding insects, provide an elegant example of metabolic collaboration between a host and bacterial symbionts. Sharpshooters host two microbial symbionts, *Baumannia cicadellinicola* and *Sulcia muelleri*, in their specialised bacteriocyte cells (Moran & Baumann, 2000), and both symbionts show patterns of codivergence with their host (Takiya et al., 2006). The genomes of *B. cicadellinicola* and *S. muelleri* predict the synthesis of vitamins and essential amino acids respectively, which are

deficient in the diet of sharpshooters (D. Wu et al., 2006). Furthermore, not only do the roles of these two symbionts appear to complement each other in terms of supplementing the host diet, but each symbiont also appears dependent on the other. Circumstantial evidence suggests that similar functional relationships may exist amongst marine invertebrates, and the characterisation of these should be a high priority.

Some examples of metabolic collaboration in complex marine invertebrate holobionts are provided by sponges. Genome and transcriptome data from *Cymbastela concentrica* and two of its bacterial symbionts (novel genomes of the *Phyllobacteriaceae* and *Nitrosopumilales*) suggest that creatine and creatinine produced by sponge metabolism are likely to be degraded to the amino acid glycine by its symbionts (Moitinho-Silva et al., 2017). Furthermore, gene expression data suggests that the urea produced by creatine degradation by the *Phyllobacteriaceae* symbiont may be transported and degraded by a third bacterial symbiont in the genus *Nitrospira* (Moitinho-Silva et al., 2017). The potential for metabolic collaboration also exists between the sponge *Theonella swinhoei* and its symbiont belonging to *Candidatus Entotheonella*. The genome of *Ca. Entotheonella* possesses the repertoire for production of almost all amino acids as well as rare coenzymes, however additional research is needed to understand if these products are used by the host (Lackner et al., 2017). While the following does not constitute metabolic collaboration, sponge symbionts also appear to interact with their host through eukaryotic-like proteins (ELPs). For example, microbial symbionts associated with different sponges often contain genes encoding for ELPs, some of which are phylogenetically similar to those found in sponges, and appear to inhibit phagocytosis (Nguyen et al., 2014; Reynolds & Thomas, 2016). Furthermore, additional functional domains associated with ELPs suggest these proteins are transported to the outer membrane where they are maintained and potentially used in bacteria-host interactions (Díez-Vives et al., 2017). A symbiosis maintained through host-bacterial interactions such as this emphasises the potential

for coevolution to take place, although itself does not demonstrate reciprocal evolution. Finally, characterisations based on metagenomic and metatranscriptomic data sets require functional validation using techniques such as stable isotope probing (SIP) (Berry & Loy, 2018). For example, using ^{14}C - and ^{13}C -labelled bicarbonate in combination with autoradiography and nanoscale secondary ion mass spectrometry (nanoSIMS), symbionts of the colonial ciliate *Zoothamnium niveum* were shown to fix inorganic carbon and translocate organic carbon to its host (Volland et al., 2018). In the advent of new technology associated with SIP, future research would benefit from validating putative microbial functions implied by genomic research.

1.3.4. Core microbiome and the potential of viruses

A core microbial community, i.e., one that has high intra species stability, is often the primary focus of microbial ecologists trying to separate functionally important taxa from commensals or short-term visitors (Hernandez-Agreda et al., 2017). While a few bacterial lineages have been shown to occur across a large number of corals and other invertebrate species (Ainsworth et al., 2015; Neave et al., 2016), a defined and stable core community remains elusive. It may be that a core community from a taxonomic perspective does not exist, but rather a core functional capacity exists across diverse lineages. In marine sponges for example, different host species associate with different symbionts that perform equivalent functions (Fan et al., 2012). Namely, host specific microbes among different sponge species appear to use different enzymes to perform the same functions in processes such as denitrification and ammonium oxidation. However, functional redundancy in microbial ecosystems may not be as common as previously thought, as rare microbial phylotypes have been implicated in specific microbial pathways while more abundant phylotypes are positively correlated with broader metabolic functions such as respiration (Rivett & Bell, 2018). This may have important implications when looking at neutral population dynamics, as those rare taxa

that are present more often than expected could be responsible for key microbial functions. A core community would have obvious implications for coevolution, as universally associated microbes are more likely to have coevolved with their host. If present, reconstructing phylogenetic relationships of core taxa can illustrate whether microbes also diverge in parallel with their host, leading to further investigations that utilise integrated genomic techniques to identify core functional genes and pathways.

While research on the microbiome of marine invertebrates has focussed mostly on prokaryotes and microbial eukaryotes (Box 2), there is increasing recognition of the importance of viruses as components of the holobiont, adding to the complexity of an already challenging system (Weynberg et al., 2014). Viruses are the most abundant biological entities in the oceans (Wommack & Colwell, 2000) and are likely to play important roles in host-microbial coevolution, as bacteria commonly acquire genes for symbiosis or pathogenicity through lateral gene transfer from viruses (Ochman & Moran, 2001). For example, the bacterium *Hamiltonella defensa* is a common symbiont of aphids providing defence against wasp parasitism. However, toxin-encoding genes required for aphid protection only occur after infection from a lysogenic lambdoid bacteriophage (Oliver et al., 2009). Thus, it is feasible that coevolution of host and symbiont can be made possible through the initial acquisition of symbiont genes from viruses. Furthermore, viruses structure bacterial communities through processes such as cell lysis, thereby adding another selective pressure to invertebrate holobionts (Bettarel et al., 2014). A recent study found that viral communities of corals and sponges are specific to their host species and are distinct from the viral communities inhabiting the surrounding seawater (Laffy et al., 2018). Viruses of the order *Caudovirales* (tailed bacteriophages) were found across all viromes in the study, often being the dominant member, thus a host specific virome combined with a host specific microbiome could be associated with viral selection and pressure. As a result, by influencing microbial community structure, viruses can have major effects on

coevolution within the holobiont. The extent to which viruses influence marine invertebrate holobionts is still unknown, however future research on reef holobionts would benefit from including analyses of both the viral and prokaryotic communities.

Box 2 – Symbiodiniaceae: an obligate symbiont and a coevolved partner?

Dinoflagellates from the family Symbiodiniaceae (LaJeunesse et al., 2018) are common symbionts of many different marine invertebrates including cnidarians, sponges, molluscs and protozoans (Stat et al., 2006). These photosynthetic dinoflagellates provide their host with fixed carbon and in return gain inorganic nutrients and a suitable living environment, creating a remarkable symbiosis that is responsible for the foundation of coral reef ecosystems (Rowan, 1998; Stat et al., 2006). The symbiotic lifestyle often leads to a reduction of genome size and, although the genomes of Symbiodiniaceae are large by comparison with those of many other eukaryotic microbes, they are among the smallest for dinoflagellates. The relatively small genomes typical of the Symbiodiniaceae suggest some degree of adaptation towards life inside the host (Aranda et al., 2016), despite the fact that many members of this family are known to have a free-living stage (LaJeunesse et al., 2018; Stat et al., 2006). An important exception to this life cycle is the dinoflagellate formerly known as clade C15, which is vertically transmitted in coral hosts, and culturing experiments suggest that it is unlikely that the strain can survive outside the host environment (Krueger & Gates, 2012). Moreover, this symbiont appears to have lost its genomic potential for motility, a likely adaptation to life inside a host (Krueger & Gates, 2012).

1.4. Challenges, further considerations and conclusions

Illustrating reciprocal adaptation of one lineage in response to another is extremely challenging in complex symbiotic systems. Whilst meeting the basic criteria set out in this thesis does not prove coevolution, it would provide support for coevolution in host-microbial systems where little is known about the evolutionary origins. In doing so, it is also likely that obligate microbes can be differentiated from transient members of the holobiont. Many factors need to be considered, including common ancestry, the origins of the host-microbial association and the estimated times of divergence. The butterfly-plant example (Box 1) highlights the necessity to distinguish the possibility of microbes colonising their host after host evolution has taken place. In the case of the Aphid-*Buchnera* symbiosis, the origin of infection has been dated at 150-250 MYA, when aphids first diverged from a common ancestor, and *Buchnera* form a monophyletic group that is exclusively associated with aphids (Baumann et al., 1997; Moran & Baumann, 2000). Within hominids, divergence times were calculated for gut bacteria that show codivergence with their host and were found to coincide with host evolution. Furthermore, the hominid-microbe association appears to have arisen from a common ancestor of all African great apes

Vertical versus horizontal microbial acquisition may also influence patterns of evolution and should be considered within any study on host-microbial coevolution. Generally, microbes that are acquired vertically, i.e., passed from parent to offspring, are more likely to have coevolved with their host. This is the case for many insect endosymbionts, and their loss of a free-living stage and subsequent adaptation to the host environment determines many of the coevolution signals previously detailed (Fisher et al., 2017; Moran & Baumann, 2000; D. Wu et al., 2006). For example, the endosymbionts *Buchnera* have been passed from parent to offspring for over 100 million years and as the endosymbiont evolved, it lost many genes required for life outside of the host (Baumann et al., 1997). Such patterns may be far more difficult to observe in microbes acquired from the environment (horizontal transmission).

Codiversification is more difficult to detect in horizontally acquired symbionts as selection pressures include environmental forces that act in concert with the host-imposed pressures. Invertebrates such as cnidarians and sponges can acquire microbial symbionts through both vertical and horizontal transmission (Ceh et al., 2013; Leite et al., 2017; Sharp et al., 2007, 2010, 2012), and focussing initially on vertically transmitted microbes would simplify the search for coevolutionary signals.

Consideration of genetic markers and key traits of symbiosis could also be useful for identifying potentially coevolved symbionts. For example, many vertically transmitted endosymbionts have reduced genome sizes compared to their free-living relatives, since many genes may become redundant during adaptation to the host environment (Fisher et al., 2017; Moran & Baumann, 2000). Some microbial symbionts are also housed in bacteriocytes or other specialised compartments, and microbial aggregates resembling such associations have been detected in both corals and sponges (Maldonado, 2007; Work & Aeby, 2014). Microbes housed in these specialised cells represent priority candidates in the search for coevolved relationships. Other trends such as lower G + C content, high isoelectric points and quickly evolving proteins relative to free living bacteria, are all features of insect endosymbionts (D. Wu et al., 2006). Exploring these traits in more complex systems may also have some utility in the search for coevolved symbionts. Furthermore, observing support for host-symbiont coevolution may require careful choice of appropriate genetic markers due to different divergence rates. In particular, it has been suggested immune genes should be targeted as they are rapidly evolving and likely to directly influence the microbial community (Brucker & Bordenstein, 2012). Additionally, unresolved genealogies of host and microbe may further confuse patterns of host-microbial coevolution, thus robust phylogenetic trees and markers are critical to illustrate codivergence.

To begin investigating host-microbial coevolution in complex holobionts it may be useful to unify around a number of model organisms. Marine sponges present an ideal starting point for investigating coevolution in complex systems for a variety of reasons. Firstly, they may represent the earliest animal lineage to have diverged and host highly stable microbial communities, increasing the likelihood of discovering coevolved symbionts. Secondly, metagenomic analyses are currently better developed in sponges than in other marine invertebrates with complex microbiomes, providing a solid platform for which to investigate coevolution. Lastly, some evidence of coevolution already exists, with sponges exhibiting codivergence and metabolic collaboration and some species hosting microbial cells within bacteriocytes. However, as yet, no research has tied all the aforementioned traits to a single holobiont species.

In this era of climate change and environmental degradation impacting heavily on marine ecosystems (Hughes, Barnes, et al., 2017; Hughes, Kerry, et al., 2017), there is an urgent need to better understand the microbial processes that underpin invertebrate health and evolution. Following the criteria set out in this chapter will not only explore evidence for coevolution, but also provide a better understanding of how microbial communities are structured and identify potentially beneficial symbionts which can be targeted using genomic techniques to elucidate their specific roles within the holobiont.

1.5. Thesis aims and chapter structure

The overall objective of this thesis was to explore host-microbe coevolution in coral reef invertebrates, thereby improving our understanding of the complex microbial structure and functions of these symbioses. To address this, chapter 2 investigated the patterns of phyllosymbiosis across thirty species of coral reef invertebrates encompassing corals, octocorals, sponges and ascidians, to understand if host evolutionary history helps shape the

microbiome. In chapter 3, the pattern of cophylogeny was used in an attempt to uncover symbionts of coral reef invertebrates that have a higher likelihood to affect host fitness. Microbial function is often predicted using metagenomics, and chapter 4 identified the expression of common metabolic pathways in the sponge microbiome using genome-centred metatranscriptomics to validate metagenomic hypotheses of sponge-microbe interactions. Chapter 5 used gene enrichment patterns to demonstrate how symbionts of sponges with strong cophylogenetic patterns differ from closely related microbes, and in doing so, highlighted signatures of microbial adaptation to the sponge and unique microbial functions. Finally, chapter 6 clarified where this thesis has enhanced our understanding of marine invertebrate microbiomes, identified limitations and areas for future research, and ends with concluding remarks on the coevolution of coral reef invertebrates with their microbiota and why this is important.

Chapter 2: Diverse coral reef invertebrates exhibit patterns of phylosymbiosis

This chapter is published as:

O'Brien, P. A., Tan, S., Yang, C., Frade, P. R., Andreakis, N., Smith, H. A., ... & Bourne, D. G. (2020). Diverse coral reef invertebrates exhibit patterns of phylosymbiosis. *The ISME journal*, 14(9), 2211-2222.

2.1. Summary

The previous chapter outlined the scope of the thesis and provided a framework for studying coevolution. In this chapter, I investigate patterns of phylosymbiosis in corals, octocorals, sponges and ascidians and thereby identify deterministic traits in the microbiota that could arise under coevolution. I show that these invertebrate groups demonstrated a signal of phylosymbiosis, and that this signal was strongest in sponges, which typically show a stable microbiome with low intraspecific variability compared to other invertebrate groups. This work improves our knowledge of how the microbiota of coral reef invertebrates is structured in the context of host evolutionary history.

2.2. Introduction

Phylosymbiosis occurs when microbial community relationships reflect the evolutionary history of the host (Brooks et al., 2016; Brucker & Bordenstein, 2013; Lim & Bordenstein, 2020). The term was first coined to describe the impact of a host phylogenetic signal on gut microbial community relationships in *Nasonia* parasitoid wasps (Brucker & Bordenstein, 2011, 2013), and the phenomenon has since been investigated in a diverse range of taxa and environments, e.g. the gut microbiomes of mammals and insects (Brooks et al., 2016; Kohl et al., 2018; Ochman et al., 2010b), the skin microbiome of ungulates (Ross et al., 2018), the endolithic microbiome of coral (Pollock et al., 2018) and the root microbiome of plants (Yeoh et al., 2017). These studies have confirmed that phylosymbiosis occurs in the simplest as well as the most diverse microbial communities and the discovery of virus/host phylosymbioses (Leigh et al., 2018) demonstrates that the phenomenon is not limited to prokaryotes. As phylosymbiosis has become more frequently observed, the mechanisms underpinning these patterns are of increasing interest.

Evolutionary processes such as codivergence and coevolution are distinct from phylosymbiosis, establishing the need of an alternative term (Brooks et al., 2016). Namely, phylosymbiosis is a pattern observed at one moment in time and space, which does not assume a stable evolutionary association between a host and its microbiota or congruent ancestral splits, nor does it assume vertical transmission of microbial symbionts (van Opstal & Bordenstein, 2019). While it is possible that different evolutionary processes contribute to the mechanisms behind phylosymbiosis (Franzenburg et al., 2013; Pollock et al., 2018), complex and dynamic systems that acquire high numbers of microbes from the environment are likely structured by other mechanisms. For example, horizontal transmission of microbes filtered through phylogenetically congruent host traits, biogeography of a host and the microbiota, and dispersal of microbes among conspecifics all potentially contribute to observed phylosymbiosis patterns (Douglas & Werren, 2016; Franzenburg et al., 2013; Groussin et al., 2017; Mazel et al., 2018). These explanations are not necessarily mutually exclusive. Within a complex microbiome where both vertical and horizontal transmission occurs among obligate and facultative microbial members, phylosymbiosis is expected to rely on multiple mechanisms (Kohl et al., 2018; Lim et al., 2019).

Despite the extensive literature supporting phylosymbiotic relationships, host phylogeny does not always correlate with microbial community (dis)similarity. For example, in contrast to other mammals, no significant congruence was observed between skin microbiome composition and host phylogeny in the case of carnivores (Ross et al., 2018). Similarly, no phylosymbiotic signal could be detected in the case of the intestinal microbiota of 59 neotropical birds (Hird et al., 2015) and the gut microbiomes of bats are more similar to birds than other mammals (Song et al., 2020). There are multiple reasons why phylosymbiosis may not occur. Firstly, factors such as environment and diet may obscure phylosymbiotic signals, which have been successfully controlled for in some studies (Brooks et al., 2016;

Brucker & Bordenstein, 2011). Secondly, in some cases, host genotype exert strong effects on microbiome composition that are independent of host phylogeny (Glasl et al., 2019; Pollock et al., 2018; Thomas et al., 2016). Finally, host physiology can structure the microbiome (Amato, G. Sanders, et al., 2019), however physiological traits may not always be consistent with host phylogeny (Amato, Mallott, et al., 2019). Therefore, patterns of phylosymbiosis may be dependent on a certain host taxonomic level (i.e., host family), where host genotype effects are reduced and host physiological traits and phylogeny are congruent.

Reef invertebrates provide interesting opportunities for testing hypotheses of phylosymbiosis, as they often host diverse microbial communities acquired by combinations of vertical and horizontal transmission (Apprill et al., 2009; O'Brien et al., 2019; Sharp et al., 2012; Webster et al., 2010) that can be dynamic among different environments (Morrow et al., 2014; Ziegler et al., 2017). Here, I first characterised the microbiomes of four groups of coral reef invertebrates; scleractinian corals, octocorals, sponges and ascidians. I then tested three recommended analyses to investigate phylosymbiosis; a) comparison of intraspecific and interspecific variation in microbiome composition, b) comparison of the topology of host phylogeny and hierarchical clustering of its associated microbial community, and c) correlation of host phylogenetic distance with microbial community dissimilarity (Lim & Bordenstein, 2020; Mazel et al., 2018). I hypothesised that a phylosymbiotic signal will be found across all four groups to show that host phylogeny is a dominant factor in microbiome structure of reef invertebrates. Through an improved understanding of microbial community dynamics using phylosymbiosis our knowledge of how a microbiome is structured and maintained in complex marine holobionts will be enhanced (O'Brien et al., 2019).

2.3. Materials and Methods

2.3.1. Sample collection

Tissue samples from 3-5 replicates of 30 species of coral reef invertebrates (12 corals, 10 octocorals, 5 sponges and 3 ascidians) were collected on SCUBA from 7 locations across the central and northern sectors of the Great Barrier Reef (GBR) (Appendix A; Table S2.1; Figure S2.1). On sampling trips to Broadhurst reef, Davies Reef and Orpheus Island, August 2017 (Table S2.1), adult colonies no larger than 30 x 30cm were collected using hammer and chisel and returned to the reef after sampling. Alternatively, sampling of invertebrates was performed *in situ*. On the surface, colonies/samples were isolated and placed in running seawater (0-2 hours) until processing. Each invertebrate was sampled for 3-5 fragments approximately 5 cm in length using either a hammer and chisel or dive knife (coral), or sterile razor blades (all other invertebrates). Additionally, seawater samples were collected from the central GBR sites in August 2017 as an environmental control (Table S2.1). All samples were collected under the marine parks permits G12/35236.1 and G15/37574.1

2.3.2. Sample processing and preservation

Fragments were rinsed in autoclaved calcium- and magnesium-free seawater (CMFSW; NaCl 26.2g, KCl 0.75g, Na₂SO₄ 1g, NaHCO₃ 0.042g, per 1L) to remove any loosely attached microbes. For scleractinian coral, tissue was removed from the skeleton by pressurised air into approximately 30 mL of CMFSW. Coral blastate was homogenised by vortex for 1min and 2 x 2 mL aliquots were kept for DNA extraction. Aliquots were centrifuged for 10 min at 10,000 x g, the supernatant removed, and tissue pellet was either snap frozen in liquid nitrogen or preserved in 1 mL dimethyl sulfoxide-EDTA salt saturated solution (DESS) and kept at -80°C (Table S2.1). For octocorals and sponges, fragments were cut into small pieces approximately 0.5 x 0.5cm³ using a sterile razor blade, snap frozen in a 2 mL cryovial and stored at -80°C until DNA extraction. Alternatively, a 15 mL falcon tube with ~7 mL DESS was filled with the dissected tissue until approximately a 1:1 ratio of tissue:DESS was reached. The ascidians

Lissoclinum patella and *Polycarpa aurata* were dissected longitudinally and the tunic layer removed and snap frozen as described above. Colonies of the remaining ascidian *Didemnum molle* were dissected into 3 equal parts as the tunic was too small to isolate and preserved in 1 mL DESS and kept at -80°C. Seawater was collected from each site (excluding the Ribbon reefs and Osprey reef) approximately 1m above the benthos at the area of sample collection using 4 x 5L retractable water bottles (washed and sterilized with 10% hydrochloric acid). Approximately 2-3L were then filtered through 0.22µm Sterivex filters and stored at -80°C (where -80°C was not available, samples were stored at -20°C for 1-5days before being transferred to -80°C upon returning to the laboratory).

2.3.3. DNA extraction and sequencing

Approximately 0.05g of tissue was used for DNA extraction using the DNeasy PowerBiofilm Kit (QIAGEN Pty Ltd, VIC Australia 3148). Extraction was performed following the manufacturers protocol with the BioSpec Mini-Beadbeater-96 used for mechanical lysis at 3-5 cycles of 30 – 60sec depending on the difficulty to break down the tissue. Genomic DNA was sent to the Ramaciotti Centre for Genomics (UNSW, Sydney Australia) for 16S rRNA amplicon sequencing on the Illumina MiSeq platform using the modified V4 region primer set, 515F (GTGYCAGCMGCCGCGGTAA) (Parada et al., 2016) and 806R (GGACTACNVGGGTWTCAAT) (Apprill et al., 2015). Samples were prepared for sequencing with the Earth Microbiome Project's 16S Illumina Amplicon protocol and sequencing was performed following the standard Illumina protocol for 16S rRNA gene amplicon library prep. Sequencing of the host phylogenetic markers COI, 18S rRNA and ITS1 was performed at the Beijing Genome Institute following the BGISEQ-500 library prep protocol on the BGISEQ-SE400 module. COI (~712 bp), 18S (~470 bp) and ITS1 (~288 bp) were amplified using the primer pairs, LCO1490 (GGTCAACAAATCATAAAGATATTGG)

and HCO2198 (TAAACTTCAGGGTGACCAAAAAATCA) for COI (Vrijenhoek, 1994) and V4_18S_Next.For (CCAGCASCYGC GGTAATTCC) and V4_18S_Next.Rev.B (ACTBTCGYTCTTGATYARNGA) modified from Pirredda et al. (Piredda et al., 2017) for 18S rRNA. For ITS1, the custom primers 18S-F1759 (GGTGAACCTGCGGAWGGATC) and 5.8S-R40 (CGCASYTDGCTGCGTTCTTC) were designed by retrieving all available sequences from the target species and aligning them using MAFFT (Katoh et al., 2019). Full length barcodes were assembled from single-end 400bp reads using the HIFI-SE pipeline (Yang et al., 2018).

2.3.4. 16S rRNA gene amplicon analysis

Sequences were analysed using QIIME2 (v 2018.4) (Bolyen et al., 2019) by first demultiplexing reads and denoising following the DADA2 pipeline (Callahan et al., 2016). Taxonomic assignment was performed using a Naïve Bayes classifier pre-trained on the Silva 132 99% OTU database modified to the V4 region primer set 515F/806R. The resulting amplicon sequence variant (ASV) table was filtered for chloroplast, mitochondrial and eukaryotic sequences. A phylogenetic tree was reconstructed using the qiime fragment-insertion sepp command (QIIME2 v 2019.1), which places the ASVs into a larger, well curated 16S rRNA reference phylogeny containing >200,000 representative tips (GreenGenes 13.8, 99% OTU) (Janssen et al., 2018). The resulting tree was then trimmed to the original reference sequences and used for subsequent UniFrac analyses. ASV and taxonomic tables were imported into R studio v.3.5.0 (Team, 2018) for further analysis with extensive use of the packages ‘phyloseq’ (McMurdie & Holmes, 2013), ‘vegan’ (Oksanen et al., 2019), ‘ggplot2’ (Wickham, 2016), ‘ggtree’ (Yu et al., 2017), ‘ape’ (Paradis & Schliep, 2018), ‘phangorn’ (Schliep, 2011) and ‘dplyr’ (Wickham et al., 2019).

2.3.5. Characterisation of microbial diversity and composition

The following analyses were conducted at the ASV level, excluding visual representations of relative abundance. Relative abundance for each microbial phylum was calculated and grouped by invertebrate taxonomy to give a broad overview of microbial profiles of each invertebrate group. Additionally, the top 25 most abundant microbial families across the entire data set were shown to give an overview of the lower taxonomic levels. As the taxonomic profile of the blanks was sufficiently different from the marine invertebrate profiles, with only 0.4% of sequences present in the top 25 family level ASVs, these samples were removed from further analysis. Rarefaction curves were calculated and plotted to illustrate the total diversity of ASVs captured against the sampling effort. Alpha diversity was calculated using both species richness (total number of ASVs retrieved per sample) and Shannon-Wiener diversity index on a dataset rarefied to 3500 sequences (equal to the sample with the lowest number of sequences). Beta diversity was calculated on non-rarefied data using the Bray-Curtis dissimilarity measure by first standardizing the data by the species maximum and then by the sample total (Wisconsin double-standardisation). This method of normalisation was chosen for beta diversity as transforming data to proportions returns the most accurate Bray-Curtis dissimilarities (McKnight et al., 2019). The resulting dissimilarity scores were visualised using non-metric multidimensional scaling (NMDS) to observe overall patterns in microbial community structure among the different invertebrates. Analysis of variance (ANOVA) and a post-hoc Tukey's test with unplanned comparisons and a Bonferroni correction was used for significance testing of alpha diversity while Permutational Multivariate Analysis of Variance (PERMANOVA) was used for beta diversity using the pairwiseAdonis function for post-hoc analysis.

2.3.6. Host phylogenetic reconstructions

Representative sequences for COI, 18S rRNA and ITS1 from each species in each taxonomic group were aligned separately using MUSCLE (Edgar, 2004) and then concatenated using DAMBE (Xia & Xie, 2001). Concatenated octocoral and sponge alignments were further curated using Gblocks (Castresana, 2000) to remove poorly aligned, high gap regions. Evolutionary model selection was performed using JModelTest2 (Darriba et al., 2012) (Appendix A; Table S2.2) and phylogenetic analysis was conducted in Mr Bayes v3.2.7 (Ronquist et al., 2012) using the outgroups *Carteriospongia foliascens* for corals, octocorals and ascidians and *Cladiella sp.* for sponges. Outgroups were selected based on their low phylogenetic relatedness to the ingroup and low variability in microbiome composition among sample replicates. Evolutionary history was inferred using Bayesian inference with the Markov Chain Monte Carlo (MCMC) method using two independent runs of 5,000,000 generations and all models converged at <0.01.

2.3.7. *Phylosymbiosis analysis*

The 16S rRNA gene dataset was subsampled to each taxonomic group and analysed independently. Gorgonians did not contain enough species within the dataset to compare host phylogeny with microbial composition and were added to the soft coral data set to create an octocoral group. Intraspecific against interspecific variability of microbiome composition was compared using pairwise comparisons of Bray-Curtis dissimilarity between each sample. Welches t-test was used for significance testing following an arcsine transformation to normalise the 0-1 distribution, while an ANOVA and post-hoc Tukey's test with unplanned comparisons and a Bonferroni correction were used to test for significant differences in intraspecific variation among invertebrate groups.

Microbial dendrograms were built in QIIME2 using the qiime diversity beta-rarefaction command. Within each invertebrate ASV table subset, all ASVs that appear 2 times or less and

those that are present in only one sample were removed to reduce noise from potentially spurious and transient ASVs. Each sample was then pooled by host species and rarefied over 1000 iterations to the host species with the lowest number of reads following the method of Brooks et al. (Brooks et al., 2016). Hierarchical clustering of host species from the resulting table was performed using the UPGMA clustering method based on Bray-Curtis dissimilarity and both weighted and unweighted UniFrac distances. Microbial dendrograms along with phylogenetic trees and pooled ASV tables were imported into Rstudio for analysis.

To assess topological congruency, host phylogenetic tree topology was compared to the microbial dendrograms using the normalised Robinson-Foulds (nRF) metric, where 0 is complete congruence and 1 is no congruence. Branch lengths were removed in host phylogenetic trees for visualisation and a significance value was calculated using the RFmeasures function (Mazel et al., 2018) with 9999 permutations. Correlation between host phylogenetic distance and microbial dissimilarity was analysed by first creating a distance matrix of pairwise phylogenetic distances between each host species and distance matrices of Bray-Curtis dissimilarity and weighted and unweighted UniFrac distances using the pooled ASV tables. A Mantel test was used to test for correlation between host and microbial distance matrices using Pearson correlation with 9999 permutations. A similarity percentages (SIMPER) analysis was used to identify which ASVs were contributing to dissimilarity between host species that showed incongruence.

2.3.8. Data and code availability

All microbial data has been made available at the NCBI Sequence Read Archive under the BioProject accession number PRJNA577361 and host sequence data is available at the CNGB Sequence Archive under the accession number N_000000252.1 - N_000000348.1. Code used for the analysis is available at <https://github.com/paobrien>.

2.4. Results

2.4.1. Sample collection and sequencing

Field collections resulted in a total of 161 samples across 30 species of reef invertebrates (Table S2.1). Additionally, eight seawater samples, two blank extractions and two sequencing positive controls were sequenced. For 16S rRNA amplicon sequencing, this yielded a total of 10,415,183 reads in 173 samples, which was reduced to 8,611,147 high quality reads following quality control and denoising. For host phylogeny, successful COI sequences were obtained for all 30 species, however 18S rRNA sequencing was unsuccessful for *Acropora formosa*, *Acropora hyacinthus*, *Diploastrea heliopora*, *Heteroxenia sp.* and *Isis hippuris* and ITS1 sequencing was unsuccessful for *Lissoclinum patella* and *Didemnum mole*. As a result, ITS1 was not used for ascidian phylogeny.

2.4.2. Characterisation of microbial diversity and composition

Rarefaction curves for each sample approached asymptotes, illustrating that total ASV richness for each sample was captured (Appendix A; Figure S2.2). However, rarefaction to the sample with the lowest number of reads (*Isis hippuris*: 3323 reads; excluding blanks) resulted in a loss in diversity in some samples. Nonetheless, overall trends showed that both ASV richness and ASV diversity (Shannon-Wiener Index) were significantly different across the broad taxonomic associations (richness; ANOVA; $F_{(5, 163)} = 7.01$, $p < 0.001$; Figure 2.1) (Shannon diversity; ANOVA; $F_{(5, 163)} = 4.64$, $p < 0.001$; Figure 2.1). Post-hoc comparisons revealed that seawater had a significantly higher ASV richness than the ascidians ($p = 0.024$), while coral had a significantly higher ASV richness than ascidians ($p = 0.006$), soft corals ($p = 0.006$) and sponges ($p = 0.003$). For ASV diversity, post-hoc comparisons revealed an increase in diversity in coral compared to the ascidians ($p = 0.009$) and soft corals ($p = 0.046$),

and an increase in seawater compared to the ascidians ($p = 0.014$). However, unlike richness, no difference was seen in ASV diversity between corals and sponges ($p = 1.0$).

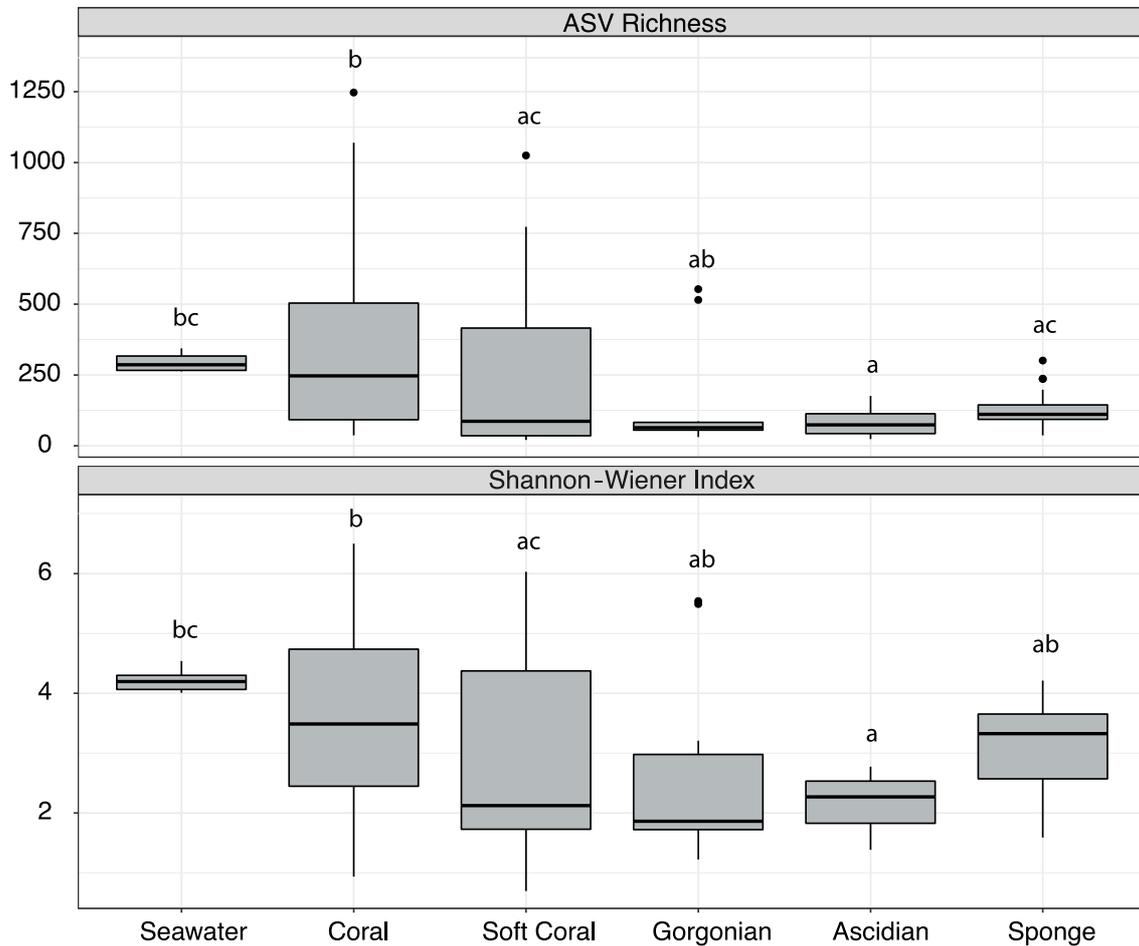


Figure 2.1. ASV richness (top panel) and Shannon-Wiener diversity index (bottom panel) for each invertebrate group and seawater. Letters indicate groups which are significantly different from each other.

A total of 62 microbial phyla were observed across the invertebrate groups and microbial profiles showed a high degree of uniformity at the phylum level. Microbial taxonomy mentioned herein are ASV sequences affiliated to that taxonomic classification, with *Proteobacteria*, *Cyanobacteria* and *Bacteroidetes* among the dominant phyla across all marine invertebrates (Appendix A; Figure S2.3). However, differences were evident even at the broad taxonomic level, with the octocorals (soft coral and gorgonians) hosting a higher relative

abundance of *Tenericutes* (mean = 4.71% ± 1.63 SE & 11.12% ± 6.28 SE respectively) compared to other invertebrates, while sponges were associated with more *Chloroflexi*, *Acidobacteria* and *Cyanobacteria* (mean = 19.09% ± 2.29 SE, 9.86% ± 1.64 SE & 28.31 % ± 3.67 SE respectively).

Relative abundance at the family level indicated far more variation in taxonomic profiles among the invertebrate groups (Figure 2.2). The three groups of anthozoans (coral, soft coral and gorgonian) were clearly different to the other marine invertebrate classifications and mostly dominated by the common *Endozoicomonadaceae* (mean = 33.52% ± 4.19 SE, 38.41% ± 4.58 SE & 42.88% ± 10.74 SE respectively). Sponges consisted of a high relative abundance of *Cyanobiaceae* (mean = 27.87% ± 3.74 SE), comprised of the commonly found cyanobacteria *Prochlorococcus* and *Synechococcus* (Silva database classification), as did seawater (mean = 32.26% ± 2.46 SE). Ascidians appeared more variable, with *Rhodobacteraceae*, *Porticoccaceae*, *Cyclobacteriaceae* and unclassified *Alphaproteobacteria*, all abundant within the top 25 bacteria at the family level.

Between sample variability (beta-diversity) showed there was an overall weak clustering of samples by their broad taxonomic classifications (Figure 2.3). Particularly the 3 anthozoans (coral, soft coral and gorgonian) and ascidians had low homogeneity in microbial composition. Comparatively, sponge and seawater samples formed clusters that indicated

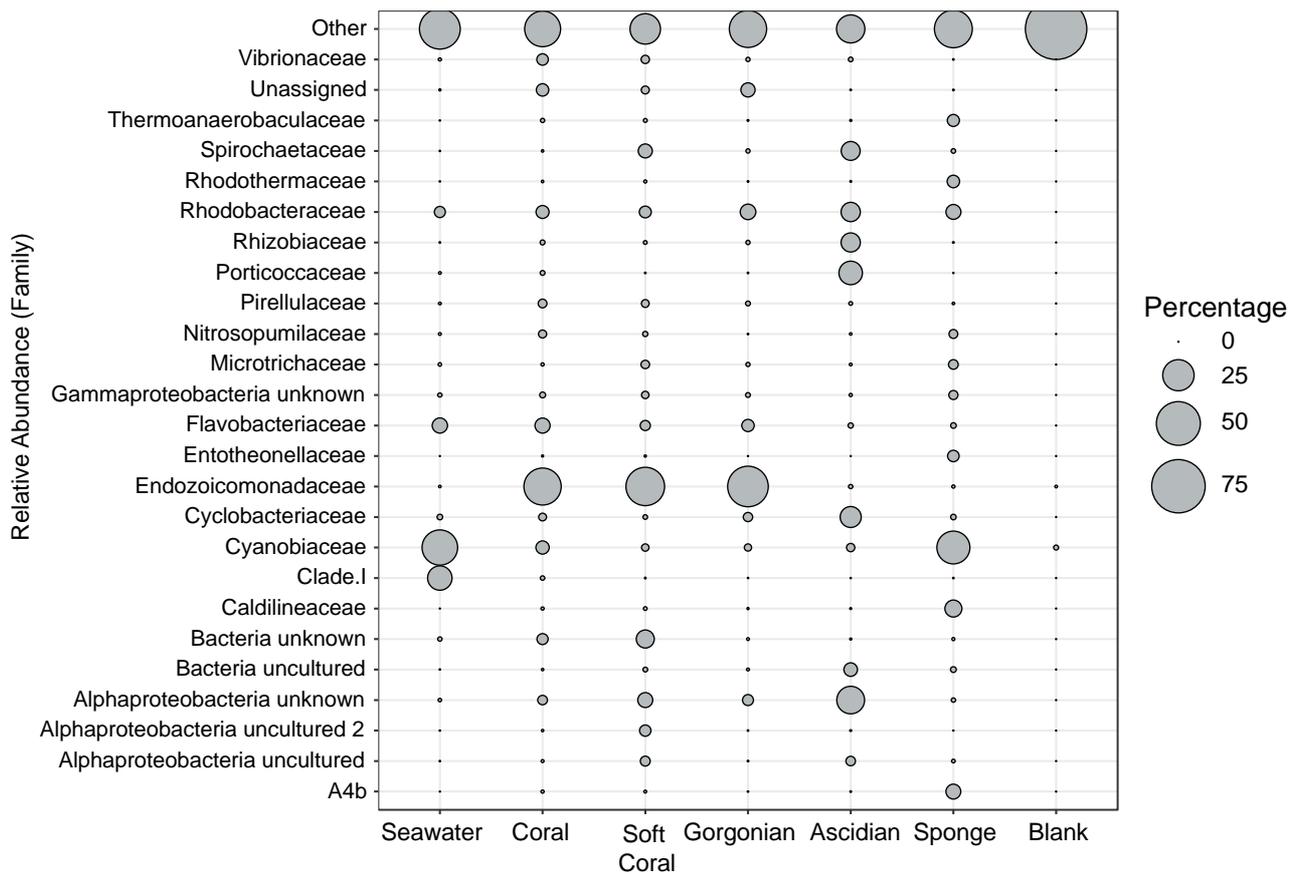


Figure 2.2. Relative abundance of the top 25 prokaryotic families found across each invertebrate group as well as seawater and blank extractions.

consistent microbial composition across samples. Microbial composition was confirmed statistically to be associated with host taxonomy (PERMANOVA; $F_{(5,163)} = 2.58$, $p < 0.001$), however only a small amount of variation in the data was explained by the broad taxonomic classification ($R^2 = 0.073$). When samples were instead grouped by host species, the amount of variation explained increased dramatically (PERMANOVA; $F_{(30,138)} = 2.01$, $R^2 = 0.30$, $p < 0.001$). Lastly, beta-diversity analysis showed there was a significant association to collection site (PERMANOVA; $F_{(6,162)} = 1.90$, $R^2 = 0.066$, $p < 0.001$), however only a small amount of variation could be explained by this variable, and since many species were collected from only one reef, it is likely the variation is due to species-specific microbiomes.



Figure 2.3. Bray-Curtis dissimilarity based on microbial composition visualised using NMDS. Each symbol represents a sample where colour is the associated host and shape is reef zone where sample was collected.

2.4.3. Assessment of phyllosymbiosis among coral reef invertebrates

All four marine invertebrate groups showed lower intraspecific Bray-Curtis dissimilarity in microbial composition compared to interspecific Bray-Curtis dissimilarity (Coral; $t_{(364)} = 13.53$, $p < 0.001$, Octocoral; $t_{(302)} = 18.84$, $p < 0.001$; Sponge; $t_{(200)} = 34.80$, $p < 0.001$, Ascidian; $t_{(69)} = 19.09$, $p < 0.001$), confirming lower microbiome variability among conspecifics (Figure 2.4). Further, intraspecific variation was significantly different among the invertebrate groups (ANOVA; $F_{(3,818)} = 231.15$, $p < 0.001$), with the exception of the ascidians

and octocorals ($t=1.85, p = 0.39$), highlighting sponges and coral with the highest and lowest microbiome homogeneity respectively.

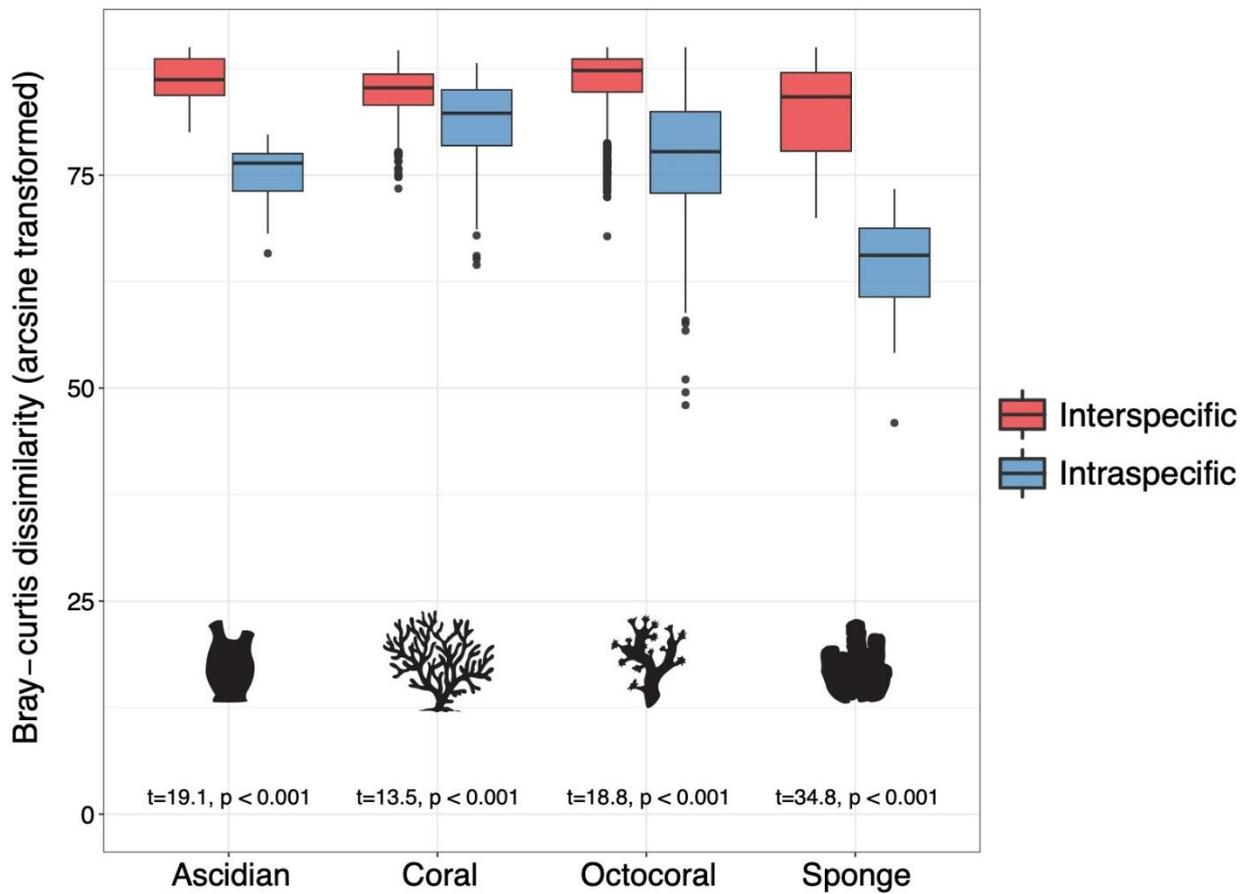


Figure 2.4. Intraspecific and interspecific Bray-Curtis dissimilarity scores for each invertebrate group.

Comparing the topology of host phylogenetic trees with the corresponding microbial dendrograms (nRF test) and measuring the correlation of host phylogenetic distance with microbial dissimilarity (Mantel test) further revealed significant levels of phylosymbiosis across all four groups of invertebrates (Table 2.1). Patterns of phylosymbiosis were significant in sponges using all tests and metrics (Figure 2.5a, Appendix A; Figure S2.4), while Bray-Curtis and weighted UniFrac metrics found significant patterns of phylosymbiosis using the nRF and Mantel tests in corals (Figure 2.5b, Appendix A; Figure 2.S5a) and octocorals (Figure 2.5c, Appendix A; Figure S2.6a). Using the unweighted UniFrac distance, phylosymbiosis

patterns were significant only using the Mantel test but not the nRF test for coral (Appendix A; Figure S2.5b) and octocoral (Appendix A; Figure S2.6b) and no patterns were detected in the ascidians (Appendix A; Figure S2.7b). Perfect congruency between host phylogeny and microbial dendrograms was observed in the ascidians using both the Bray-Curtis and weighted UniFrac metrics (Figure 2.5d, Appendix A; Figure S2.7a). Despite this, no significant phylosymbiosis was observed using the Mantel test. This opposing result is likely due to the low sample size combined with marked differences in microbial composition among the three ascidians (Appendix A; Figure S2.8).

Table 2.1. Normalised Robinson-Foulds (nRF) and mantel statistics across Bray-Curtis, weighted and unweighted UniFrac beta-diversity metrics. Non-significant values are highlighted in grey.

	Bray-Curtis	Weighted UF	Unweighted UF
Sponge - nRF	RF = 0.02, p < 0.001	RF = 0.4, p = 0.006	RF = 0.4, p = 0.01
Sponge - mantel	r = 0.71, p < 0.001	r = 0.78, p = 0.006	r = 0.75, p = 0.03
Coral - nRF	RF = 0.69, p < 0.001	RF = 0.69, p < 0.001	RF = 0.92, p = 0.15
Coral - mantel	r = 0.37, p = 0.02	r = 0.38, p = 0.01	r = 0.42, p = 0.03
Octocoral - nRF	RF = 0.64, p < 0.001	rRF = 0.82, p = 0.02	RF = 0.91, p = 0.24
Octocoral - mantel	r = 0.23, p < 0.001	r = 0.36, p < 0.001	r = 0.25, p < 0.001
Ascidian - nRF	RF = 0, p < 0.001	RF = 0, p < 0.001	RF = 0.5, p = 0.34
Ascidian - mantel	r = -0.03, p = 0.63	r = 0.46, p = 0.17	r = 0.18, p = 0.46

A select few species were collected from multiple locations and showed contrasting results in relation to phylosymbiosis. The sponge *Ircinia ramosa* and octocoral *Sarcophyton sp.* were collected from two locations and both correctly formed a clade with their conspecifics (Figures 2.5, S2.4 & S2.6), which was supported by uniform microbial profiles (Appendix A; Figures S2.9 & S2.10). Conversely, the octocoral *Sinularia sp.* and the coral species *Porites*

cylindrica and *Seriatopora hystrix* did not form clades with their conspecifics from different locations and there was a reduced overall phyllosymbiotic signal (Figures 2.5, S2.5 & S2.6). A SIMPER analysis revealed that shifts in the relative abundance of ASVs assigned to *Endozoicomonadaceae* were consistently the top contributors to the dissimilarity observed between species collected from two sites (Table S2.3; Figure S2.11). For example, *Porites cylindrica* collected from the Palm Islands (PI) had a dramatic reduction in *Endozoicomonadaceae* compared to those collected from the Ribbon Reefs (RR), where the mean relative abundance of *Endozoicomonadaceae* fell from 82.9% (\pm 4.32 SE) to 3.31% (\pm 1.69 SE). Similarly, the microbial profile of *Sinularia* collected from RR differed from the two *Sinularia* species collected from PI, with colonies from RR hosting a lower relative abundance of *Endozoicomonadaceae* and a higher relative abundance of unknown bacteria and *Fusobacteriaceae* (Figure S2.11).

Additional incongruences were observed among the groups where sample location was not a factor. The overwhelming majority of extant corals fall into one of two major clades, the Robusta and Complexa. This split was only partially reflected in the Bray-Curtis and weighted UniFrac microbial dendrograms, although in most cases, species within a genus or family clustered together (Figures 2.5b & S2.5). Similarly, host phylogeny was recapitulated in the microbiome of only certain clades of octocorals using Bray-Curtis and weighted UniFrac metrics, such as the microbiome of *Briareum* and species within the family Alcyoniidae (*Sarcophyton*, *Sinularia* and *Cladiella*), with the exception of *Sinularia* collected from the RR (Figure 2.5c & S2.6). However, no congruence was seen between gorgonian phylogeny and microbial composition, which can again be attributed to ASVs assigned to *Endozoicomonadaceae* (Table S2.3; Figure S2.10). Lastly, although the signal of phyllosymbiosis in sponges was strong and robust across all analyses, the main incongruence was due to an unclassified *Ircinia* sp, which did not form a clade with its sister species in the

host phylogeny (Figure 2.5a), and highlights the unresolved phylogenetic relationships among the *Ircinia* (Pöppe et al., 2010).

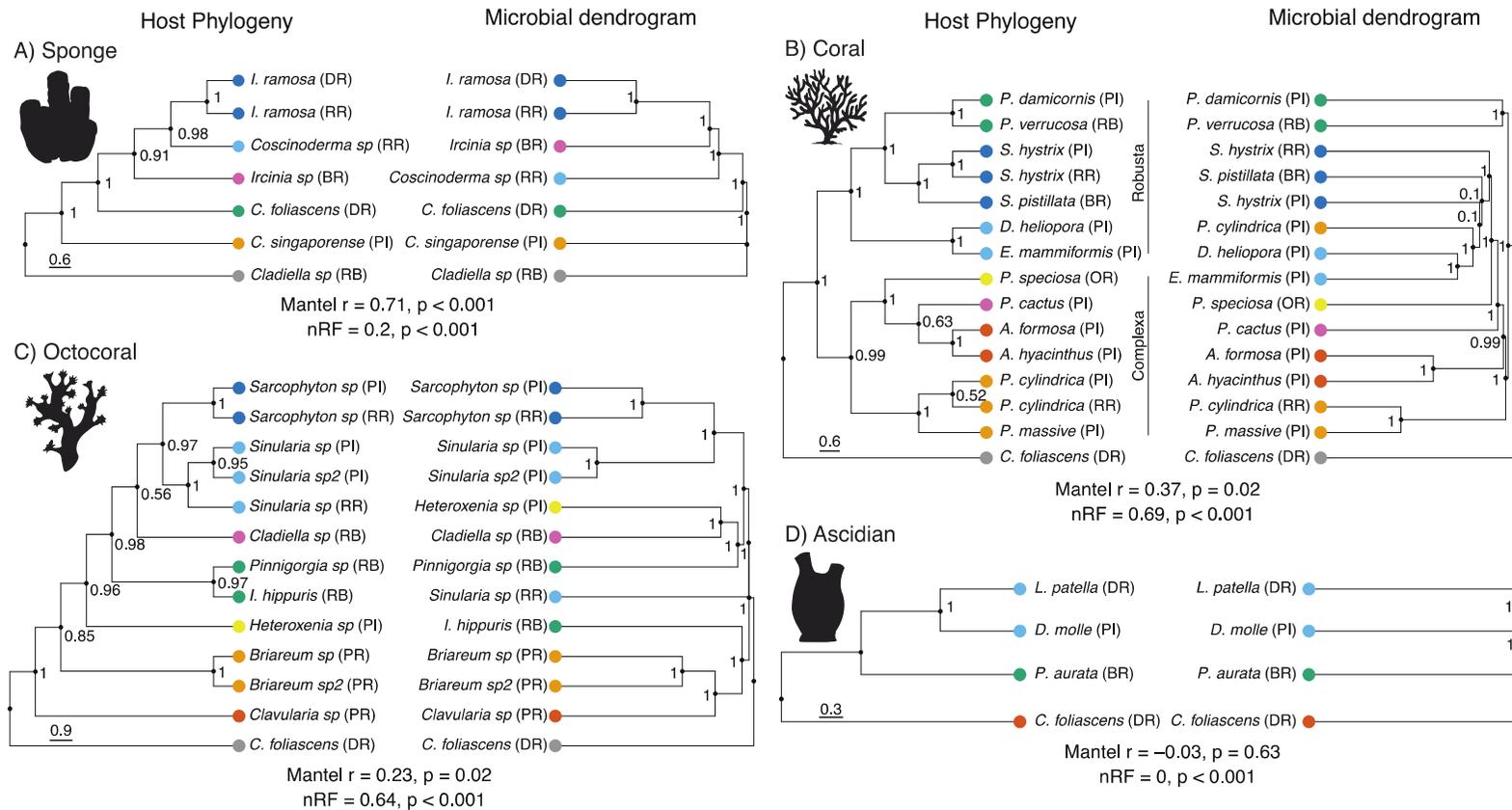


Figure 2.5. Host phylogeny inferred from COI, 18S rRNA and ITS1 sequences compared to a microbial dendrogram based on Bray-Curtis dissimilarity for microbial composition of each host species. *Cladiella* sp. was used as an outgroup for 5A) sponges, while *C. foliascens* was used as an outgroup for 5B) coral, 5C) octocoral and 5D) ascidians. Numbers at nodes reflect posterior probability for clade support in the host tree and jackknife support values in dendrograms. Branch tips are coloured to reflect clades in host phylogeny. Initials in brackets next to species names refer to collection site. BR = Broadhurst Reef, DR = Davies Reef, OR = Osprey Reef, PI = Palm Islands (Orpheus and Pelorus), PR = Pandora Reef, RB = Rib Reef, RR = Ribbon Reefs.

2.5. Discussion

This chapter evaluated the signal of phylosymbiosis in diverse coral reef invertebrates, finding evidence that host evolutionary history helps shape the microbiome in sponges, corals, octocorals and ascidians. By testing three commonly used methods for phylosymbiosis analysis, I showed that all groups have lower intraspecies microbiome variability compared to interspecies. This was combined with greater topological congruency between host phylogeny and the microbial dendrogram than would be expected by chance, except when using the unweighted UniFrac distance in corals, octocorals and ascidians. Interestingly, all invertebrate groups but the ascidians exhibited a significant correlation between host phylogenetic distance and microbial dissimilarity across all beta-diversity metrics.

2.5.1. *Sponges demonstrated a strong signal of phylosymbiosis*

These results demonstrated that sponges have a strong signature of phylosymbiosis, which likely reflects the uniform microbiome structure in sponges compared to other coral reef invertebrates (Thomas et al., 2016). This was observed through low intraspecific variation and high homogeneity in the microbiome when the same species was collected from different reefs. Sponges are also known to have a relatively stable microbiome in response to temporal variation and environmental perturbations (Erwin et al., 2015; Glasl et al., 2018; Luter et al., 2012). A stable microbiome may lead to a strong phylosymbiotic signal if there is less influence from the surrounding environment, leaving host factors to be the primary structuring element of the microbiome (Glasl et al., 2018). Importantly, while sister species were included in the analysis, overall the sponges sampled here span a larger phylogenetic diversity compared to the other groups, which may increase the chance to observe phylosymbiosis. These results agree with previous conclusions of a significant correlation between host phylogeny and

microbiome dissimilarity and validate a prominent role of host phylogeny in shaping the sponge microbiome (Easson & Thacker, 2014; Thomas et al., 2016).

2.5.2. *The signal of phylosymbiosis was weaker in coral*

A signal of phylosymbiosis was demonstrated in coral, which was characterised by a tendency of corals of the same genus or family to cluster together. However, incongruences were observed where the same species was collected from two different locations, primarily due to a shift in the relative abundance of *Endozoicomonadaceae*. Shifts in *Endozoicomonadaceae* have been documented previously, normally in response to host stress (Meyer et al., 2014; Morrow et al., 2014). As shifts in the microbial community can often precede visual signs of an unhealthy holobiont (Bourne et al., 2008; Pollock et al., 2019), it is plausible the decrease in *Endozoicomonadaceae* is linked to an unknown event. Secondly, coral tissue samples are often contaminated by the coral mucus, which is known to have a dynamic microbial community shifting in composition between new and aged mucus (Glasl et al., 2016). However, bacteria within the tissues of corals are housed within coral associated microbial aggregates (CAMAs) and these communities likely have a more stable association with the host (Wada et al., 2019; Work & Aeby, 2014). Therefore, developing approaches to target tissue specific microbes could be beneficial to understanding phylosymbiosis and other questions related to microbial symbiosis in corals.

Similar clustering of coral microbiomes has been observed in Caribbean corals. This partially reflected coral phylogeny, as congenics showed comparatively low microbial dissimilarity and the two major coral clades tended to cluster together, however inconsistencies were seen when looking at the species level (Sunagawa et al., 2010), and reflect the results seen here on the GBR. Further evidence of phylosymbiosis in coral was found in an analysis of 691 coral samples collected Australia wide (Pollock et al., 2018). The endolithic microbial

community showed the strongest signal and was the best predictor of the deep phylogeny between the Robusta and Complexa clades. Tissue microbiomes also illustrated evidence of phylosymbiosis, however the signal was absent in the coral's surface mucus layer. This emphasises an increasing strength of phylosymbiosis where direct environmental factors are reduced. Additionally, a small number of microbial lineages, including those within *Endozoicomonadaceae*, demonstrated co-phylogeny with their host, while other clades had a more generalist host distribution. It is possible that host specialist clades play a minor role in phylosymbiosis through codivergence and future work should aim to untangle the mechanisms behind phylosymbiosis (Mazel et al., 2018).

2.5.3. *The first direct evidence for phylosymbiosis in octocorals*

Research on the microbiome structure of octocorals is limited compared to corals, and I show for the first-time direct evidence for phylosymbiosis. The phylosymbiotic signal in octocorals was similar to corals and incongruences also occurred when there was a shift in the relative abundance of *Endozoicomonadaceae*. Octocorals are known to have a more stable and less diverse microbial community than hard corals (Van De Water et al., 2018), consistent with my finding that overall microbial diversity was lower and microbiome uniformity higher in octocorals compared to hard corals. While this likely influences the phylosymbiotic signal, a direct comparison between octocorals and corals (and other invertebrate groups) cannot be drawn due to the differences in phylogenetic relatedness between host species. Further, the phylogenetic markers used in this study were chosen to capture both mitochondrial and nuclear evolution across a broad range of diverse species. However, octocorals have poorly understood phylogenetic relationships, with little concordance between morphological, nuclear and mitochondrial data (McFadden et al., 2010). The incorporation of alternative phylogenetic markers optimised for each taxonomic group may further improve analyses of phylosymbiosis

and comparisons among groups. Finally, octocoral identification in the field is extremely challenging especially when trying to resolve to species level (Quattrini et al., 2019). Despite these limitations, I still observed a significant signal of phylosymbiosis, which is likely to strengthen with improved phylogenetic relationships and species identification.

2.5.4. Inconsistent results for phylosymbiosis in ascidians

Ascidians showed complete congruence between the host phylogeny and microbial dendrogram for both Bray-Curtis and weighted UniFrac metrics, yet no correlation existed between host phylogenetic distance and microbial dissimilarity. These results therefore do not provide strong support for phylosymbiosis in the group, yet they highlight the need for multiple lines of evidence when evaluating phylosymbiosis (Lim & Bordenstein, 2020). For example, I find that when sample numbers are low, particularly when marked changes are observed among the microbiomes of host species, the dendrogram approach was more sensitive to patterns of phylosymbiosis compared to the Mantel test. Furthermore, unweighted UniFrac methods were unable to identify a phylosymbiotic signal in the ascidians and had the least power to identify a signal across all invertebrate groups, which agrees with previous conclusions on weighted and unweighted beta-diversity metrics (Mazel et al., 2018). As this method does not account for the abundance of ASVs, it is less likely to identify beta-diversity patterns in highly diverse microbiomes that are dominated by a relatively small number of bacteria.

2.5.5. Patterns of phylosymbiosis occurred despite may uncontrolled factors

This study overwhelmingly found that host phylogeny is reflected in the microbiome of marine invertebrates, particularly notable when considering several confounding factors. Sampling of the reef invertebrates occurred over four field trips that spanned a one-year timeframe, potentially obscuring phylosymbiosis patterns due to seasonal influences on the

microbiome (J. Li et al., 2014). Furthermore, these samples are from wild colonies collected from multiple locations on the GBR which introduces local environmental differences including water quality and the pelagic communities that serve as host diet. Preservation methods also varied across organisms, including snap freezing and the use of salt saturated dimethyl sulfoxide-EDTA. While these preservation approaches have been shown to have little effect on the microbial composition of coral, it could have influenced alpha diversity (Hernandez-agreda et al., 2018). Finally, sample representation differed among the four groups and likely has an important impact on the strength of the phylosymbiosis signal. For example, only three ascidian species (and one outgroup) were used whereas four related species are recommended (Lim & Bordenstein, 2020). Had more species been included in the analysis, with a larger number of taxonomic sister species, a more reliable representation of phylosymbiosis would likely have been achieved.

2.5.6. Conclusions

This is the first study to systematically assess phylosymbiosis among diverse groups of marine invertebrates. I identified a phylosymbiotic signal across all invertebrate groups with multiple methods, of which sponges consistently showed a significant signal using all beta-diversity metrics. Increased intraspecific variability of the microbiome in both scleractinian corals and octocorals was often associated with a change in the relative abundance of *Endozoïcomonadaceae*. This microbial family is characterized by host-specialist and host-generalist clades and is assumed to be a dynamic member of the coral holobiont (Pollock et al., 2018). Host-specialist clades may contribute to phylosymbiosis in corals and octocorals through codivergence, while host-generalist clades obscure the signal through host infidelity. Here, I provide a foundation to begin exploring the mechanisms behind phylosymbiosis and further our understanding on host-microbe symbiosis and co-evolution in marine invertebrates.

Chapter 3: Testing cophylogeny between coral reef invertebrates and their bacterial and archaeal symbionts

This chapter is published as:

O'Brien, P. A., Andreakis, N., Tan, S., Miller, D. J., Webster, N. S., Zhang, G., & Bourne, D. G. (2021). Testing cophylogeny between coral reef invertebrates and their bacterial and archaeal symbionts. *Molecular Ecology*. 30, 3768– 3782.

3.1. Summary

Chapter 2 looked at evidence for coevolution considering the microbiota as a unit of selection. In this chapter, I focus on individual microbial lineages to identify those taxa that show high congruency with host phylogeny. Microbes that show patterns of cophylogeny are more likely to be closely associated with the host and represent important symbionts within the microbiome. Further, these may have a larger contribution to the patterns of phylosymbiosis observed in chapter 2. I also subset the analysis to corals, octocorals and sponges, as there was low phylogenetic representation among the ascidians which resulted in low statistical power for the analyses conducted in chapter 2. This chapter identifies five groups of microbes that consistently show a significant cophylogenetic signal and highlights priority taxa for further exploration; *Endozoicomonadaceae*, *Microtrichaceae*, *Nitrosopumiliaceae*, *Spirochaetaceae* and *Thermoanaerobaculaceae*.

3.2. Introduction

Eukaryotic life often depends on microbial symbionts (herein generically referred to as bacteria and archaea) for survival. In the marine realm, bacteria can provide a range of functions for their host including stimulating the metamorphosis of larvae (Ericson et al., 2019), aiding camouflage through bioluminescence (McFall-Ngai, 2014) and providing essential nutrients through bacterial heterotrophy or autotrophy (Rix et al., 2020), all of which highlight the critical importance of host-microbe relationships. In complex host-associated microbial ecosystems, it can be difficult to understand which members of the community are tightly coupled with host health as they coexist amongst a diverse community of commensal and transient bacteria. Previous studies have defined the “core microbiome” using the assumption that a microbial operational taxonomic unit (OTU) found in a high percentage of species replicates is likely to be an important member of the microbiota that can persist in a

symbiotic relationship with the host through space and time (Astudillo-García et al., 2017; Hernandez-Agreda et al., 2017; Shade & Handelsman, 2012). However we may achieve greater insights into the (core) microbiome by incorporating host evolutionary relationships to address questions related to codivergence and using network analyses to understand microbial co-occurrence patterns (Astudillo-García et al., 2020; Shade & Handelsman, 2012).

Codivergence (also known as cospeciation) occurs when two organisms diverge or speciate in parallel and is a signature of an intimate and long-standing symbiosis (Clark et al., 2000; Moeller et al., 2016; O'Brien et al., 2019). The obligate relationship between the pea aphid and its bacterial endosymbiont *Buchnera* is a classic codivergence example in which host and symbiont have coevolved a mutualistic and functional dependence (Baumann et al., 1997; Monnin et al., 2020). Conversely, the parallel evolution between pocket gophers and chewing lice has become a textbook example of cospeciation in a parasitic symbiosis (Hafner et al., 2003). Such examples in nature are rare, as they are often complicated by host promiscuity, independent speciation of a host or symbiont and/or the extinction of host species or symbionts (Balbuena et al., 2013; Groussin et al., 2020). However, cophylogenetic analyses can be used to test whether host-symbiont associations are non-random by assessing the phylogenetic congruence of interacting species (Balbuena et al., 2013).

Traditionally, cophylogenetic studies have been used to confirm coevolved host-microbe relationships where an understanding of the symbiosis already exists. For example, the bobtail squid symbiosis with *Vibrio fischeri* was well known prior to the establishment of congruent host-microbe phylogenies, and this discovery strengthened the picture of a long-term, coevolved relationship (Nishiguchi et al., 1998). Similarly, in the complex microbial community of the primate gut (including humans), *Bifidobacter* species are represented as important symbionts, and were later confirmed to cospeciate with their host (Moeller et al.,

2016). These examples suggest that important symbioses can be identified through cophylogenetic studies by detailing a long-standing and tightly coupled relationship. Therefore, an alternative approach which would be particularly valuable in species with highly diverse microbiota is to identify cophylogeny first, and subsequently investigate the nature of the symbiosis. For example, a cophylogenetic analysis was used to identify a range of bacteria in the gut microbiota of mammals that show strong cophylogenetic patterns, many of which have no known function but are now priority candidates for further exploration (Youngblut et al., 2019).

Network analyses are another useful tool for deconstructing interactions among members of a complex ecosystem such as a microbiome (Berry & Widder, 2014; Gysi & Nowick, 2020). In particular, when no single organism exerts control over a community, networks of co-occurrence may better explain patterns of host-microbe assemblages. For example, the pathogenesis of irritable bowel disease is not attributed to a single microbe and co-occurrence modelling has illustrated that gut bacteria co-occur in different communities in diseased patients compared to healthy (Baldassano & Bassett, 2016). The identification of sub-networks and hub species within the host microbiota may further elucidate important groups of microbes, as they either co-occur or have a disproportionate influence over the microbial community. Further, co-occurring sub-groups may indicate functional guilds of microbes, which would otherwise be missed if concentrating on a single lineage or the microbial community as a whole (Youngblut et al., 2019). Combining this strategy with a cophylogenetic analysis may further tease apart microbial interactions and help to understand the assemblage of a complex microbial community.

Here I investigated the microbial profiles associated with corals, octocorals and sponges under a cophylogenetic framework to further our understanding of microbial community

assemblages and identify potential codiverging symbionts. I first assessed the fit of the core microbiota of each host group to a cophylogenetic model and then analysed each microbial genus and family independently to observe which groups show stronger phylogenetic signals. These relationships were scrutinised using the local Moran's I measure of spatial autocorrelation to identify specific amplicon sequence variants (ASVs) that are clustered according to host phylogeny. Finally, I applied a co-occurrence network analysis to the core microbiota dataset to gain a deeper understanding of the interactions among microbial members and observe whether co-occurring microbes are also distributed in accordance with host phylogeny. This chapter demonstrated that within the complexity of coral reef invertebrate microbiomes, exists a highly structured (core) community with implications for environmental adaptation and evolution.

3.3. Materials and Methods

3.3.1. Data collection and sequencing

This chapter used the sequence data generated in chapter 2 with minor modifications. Namely, host phylogenetic data now includes every host species replicate, rather than one representative from each species. Further, ascidian data was not included in this chapter due to the low phylogenetic representation. Overall, the final dataset consisted of twelve coral, ten octocoral and five sponge species (n=3-5 biological replicates per species) and eight seawater samples with a subset of host species collected from two locations (Table S2.2). Each sample consisted of a bacterial 16S rRNA gene profile paired to the multi-locus COI, 18S rRNA and ITS1 host phylogeny. This design allowed for direct comparison of host phylogenetic relationships with microbial phylogeny while incorporating details on host-microbe interactions.

3.3.2. Host sequence alignments, model selection and phylogenetic reconstructions

Alignments of host sequences were computed for each marker separately in MUSCLE (<http://www.ebi.ac.uk/Tools/muscle/index.html>) (Edgar, 2004) and alignments were refined by eye. Following exploratory, independent phylogenies for each DNA region and species group, the three alignments were concatenated and the best-fitting substitution model for maximum likelihood (ML) analyses given the alignment was identified in jModelTest 2 (Darriba et al., 2012). Unpartitioned model-based heuristic searches for ML were inferred for the full dataset in raxmlGUI 2.0 (Edler et al., 2019) and nodal support was tested against 1000 bootstrap pseudo-replicates.

3.3.3. 16S rRNA data pre-processing and filtering

Raw 16S rRNA gene sequences were clustered into ASVs and phylogeny reconstructed using QIIME2 (v 2018.4; (Bolyen et al., 2019) during chapter 2. To further explore the cophylogenetic patterns, a second ASV phylogeny was reconstructed using a de novo phylogenetic method implemented in QIIME2 (v 2018.4), to ensure that any novel clades not represented in the reference phylogeny are not impacting cophylogenetic patterns. ASVs were de novo aligned using MAFFT (Katoh & Standley, 2013), the alignment masked to remove poorly aligned columns and the final tree was inferred using FastTree 2 (M. N. Price et al., 2010). Both trees were subsequently used for all cophylogenetic analyses while all figures involving ASV phylogeny were represented using the fragment-insertion tree from chapter 2. The resulting ASV table, taxonomic table and phylogenetic trees were imported into RStudio for analysis (Team, 2018) and pre-processing and filtering was conducted using ‘Phyloseq’ (McMurdie & Holmes, 2013). Rarefaction was first applied using the value of the sample with the lowest number of reads (3300). To obtain a core ASV dataset, I filtered out low abundance ASVs (≤ 10 / sample) then split the data by host species and removed ASVs that were present

in less than 50% of host species replicates (eg. 2 of 5 replicates). The resulting tables were then joined to create a core set of ASVs, defined here as ASVs present in greater than 50% of samples within each host species. A presence value of 50% was chosen to retain sufficient ASVs for robust analysis while removing ASVs that were only sporadically associated with a given host species and therefore less likely to be an important member of the microbiota. To obtain independent datasets of each microbial genus and family, I opted against the core filtering method to increase the number of genera/families being tested and to increase the number of ASVs within each genus/family and gain more statistical power. I first removed low abundance ASVs that had a total read count of ten or less from the rarefied dataset. A subset of the resulting ASV data was then made for each genus and family, and any genus or family with less than ten ASVs were removed due to low phylogenetic representation and sample size.

3.3.4. Cophylogenetic analysis using PACo and Moran's I

To assess cophylogeny I followed a pipeline developed by Youngblut et al., 2019 with minor modifications. The Procrustean Approach to Cophylogeny (PACo; (Balbuena et al., 2013) model was implemented in R (Hutchinson et al., 2017) and was used to test the global fit of cophylogeny for all samples against all core ASVs and again between each host group and their core ASVs. To account for unequal sample sizes and clade representation as well as phylogenetic uncertainty, I used a sensitivity approach where each host tree was subsampled 100 times, randomly selecting one sample per host species. I then used the PACo model on each host tree subset along with the ASV phylogeny and a presence-absence interaction matrix. Each model ran 999 permutations using the 'quasiswap' method, which does not assume that the microbe phylogeny tracks the host or vice-versa (a more conservative model). Results were then summarised to report the mean p -value following a Benjamini-Hocking (B-H) adjustment and residual sums of squares for the 100 data subsets. To identify the relative contributions of

each host-microbe link to the cophylogenetic model, the `paco_links` command was used on each `paco` model output (i.e., each host tree subset). Each contribution is weighed by jackknife estimation of their residual sums of squares, where a lower residual value indicates a better fit to cophylogeny. Analysis of variance (ANOVA) was used to test for differences in residual values among the host species following fourth root transformation.

To further breakdown patterns of cophylogeny, ASVs within each microbial genus and family were tested separately to identify which groups of microbes had the strongest cophylogenetic fit and were therefore more likely to have an intimate symbiosis. These taxonomic levels were chosen to maintain high taxonomic resolution while still retaining sufficient ASVs to compare host phylogeny against microbe phylogeny. The PACo model was used in the same sub-sampling process as described above and each microbial genus/family was compared to all host samples. Each genus and family were ranked by their mean B-H adjusted *p*-value and mean sums of squares to find the bacterial groups which showed the strongest fit to cophylogeny. To determine which ASVs had the strongest phylogenetic signal I applied the local Moran's *I* spatial autocorrelation coefficient (Moran's *I*; (Anselin, 1995; Gittleman & Kot, 1990) implemented in the R package 'phylosignal' (Keck et al., 2016). Using the core set of ASVs as 'traits' associated with each host, the presence and abundance of core ASVs was tested to see whether it was affected by the phylogenetic relationship of the host while also including the hotspots (host species) where these ASVs were clustered. The sensitivity approach was again employed as described above and applied Moran's *I* to each host subset. To assess which ASVs were most affected by host phylogeny, only the ASVs that had a significant correlation in $\geq 50\%$ of host sub-sampled data was retained, following a B-H adjustment for multiple comparisons. All phylogenetic figures were plotted using 'ggtree' (Yu et al., 2017) and 'dendextend' (Galili, 2015) with seawater controls used in relative abundance plots to identify if cophylogenetic ASVs were present in the surrounding seawater.

3.3.5. Co-occurrence network analysis

To identify co-occurring subgroups of microbes, the pipeline developed by Youngblut et al., 2019 was followed with minor modifications. Using the core subset of ASVs across all host groups, a co-occurrence network analysis was conducted using the ‘cooccur’ package in R (Griffith et al., 2016). Firstly, the core ASV data was transformed to a binary presence-absence matrix and then the observed and expected frequencies of co-occurrence between each ASV pair using a probabilistic model of species co-occurrence were calculated (Veech, 2013). Only the significantly co-occurring ASV pairs at a conservative $p < 0.01$ following B-H p -adjustment were retained to reduce the likelihood of false-positives. The resulting output was used to identify co-occurring sub-networks of ASVs by calculating the centrality of the major hub ASVs and the betweenness centrality (number of shortest paths through a node) using the R packages ‘igraph’ and ‘tidygraph’ (Csardi & Nepusz, 2006; Pedersen, 2018) and plotted using ‘ggraph’(Pedersen, 2020). To determine whether sub-networks were implicated in cophylogeny, the proportion of ASVs within each sub-network that also had a significant host phylogenetic signal were identified.

3.4. Results

3.4.1. Host sequence alignments, model selection and phylogenetic reconstructions

The final host species alignment consisted of 141 sequences which spanned up to 1,347 nucleotides across the 18S rRNA, COI and ITS1 concatenated gene regions. The HKY+G ($g=0.53$) was the best fitting model of evolution identified by jModelTest 2. Substantial phylogenetic signal and phylogenetic congruence among single marker datasets were revealed and the phylogenetic reconstructions were always in agreement with the taxonomic

expectations at the genus and family levels for each host species group (Appendix B; Figure S3.1).

3.4.2. Coral reef invertebrates and their symbionts display a pattern of cophylogeny

Following data pre-processing and filtering, a core set of ASVs (defined here as ASVs that are present in 50% of samples within a host species) were identified. The strict filtering criteria and core threshold reduced the total number of ASVs from 15 575 to 108 in coral, 6766 to 114 in octocoral, 2583 to 150 in sponges and from 18 556 to 360 across all samples combined (see Table S3.1 for number of core ASVs per host species). Despite heavily reducing the number of ASVs, core ASVs still represented 51-79% of the total reads within the dataset (Table 3.1). Interestingly, sponges retained the largest proportion of ASVs while corals retained the least using this core threshold, which may reflect the increased intraspecies variation of the coral microbiota compared to sponges (Table 3.1).

Table 3.1. Filtering and PACo model results for core ASVs associated with each host group and all samples combined. Results for PACo model are obtained from the filtered core data and mean p-value is calculated following a Benjamini-Hochberg adjustment for multiple comparisons. FI refers to results using the fragment-insertion phylogeny while DN refers to the de novo phylogeny.

Host Group	Core filtering				PACo results			
	No. of ASVs	No. of core ASVs	Proportion of ASVs retained (%)	Relative abundance of core ASVs (%)	mean p-value (FI)	residual ss (FI)	mean p-value (DN)	residual ss (DN)
Coral	15 575	108	0.69	51.48	0.001	0.955	0.001	0.953
Octocoral	6 766	114	1.68	64.77	0	0.92	0	0.917
Sponge	2 583	150	5.81	79.4	0.004	0.977	0.022	0.982
All samples	18 556	360	1.94	63.44	0	0.95	0	0.949

Applying the PACo model to all samples using their filtered core ASVs and then again to each host group and their core ASVs separately, identified that the core set of ASVs fit a model of cophylogeny using both phylogenetic methods ($p < 0.05$; Table 3.1; Figure 3.1).

Residuals for each host-microbe interaction were significantly affected by host species (ANOVA; $F_{(31, 1862)} = 384.37, p < 0.001$; all host species) and the residual contribution from an ASV to the overall fit could be relatively high or low depending on which host it was associated with (Figure 3.2; Appendix B; Figure S3.2). This indicates that at the core ASV level, cophylogenetic interactions may be driven more by the host than the microbe. Interestingly, the pattern of relative contributions to the overall residuals sums of squares differed when host groups were analysed separately or together (Figure 3.2; Appendix B; Figure S3.3). For example, the sponge *Coelocarteria singaporensis* had a poorer fit to cophylogeny compared with other sponge species when analysed separately (Figure S3.3c), however when all samples were included, this species showed a better fit to cophylogeny compared to other sponge species (Figure 3.2). This inconsistency is likely due to the distant relatedness of *C. singaporensis* to the other sponges, causing it to act as an outgroup when the analysis was confined to sponges. Within the corals, there was a stronger divide between species when all samples were analysed together, with the acroporiids and pocilloporiids showing a relatively poorer fit to cophylogeny compared to coral species within the families Agariciidae, Merulinidae and Poritidae, although the patterns were consistent between both analyses (Figure 3.2; Figure S 3.3a).

3.4.3. Microbial groups with strong cophylogenetic signals

Further investigations of the microbial cophylogenetic patterns were restricted to thirty-one genera and fifty-four families after filtering low abundance reads and removing microbial genera/ families with a low number of ASV representatives. Applying the same cophylogenetic

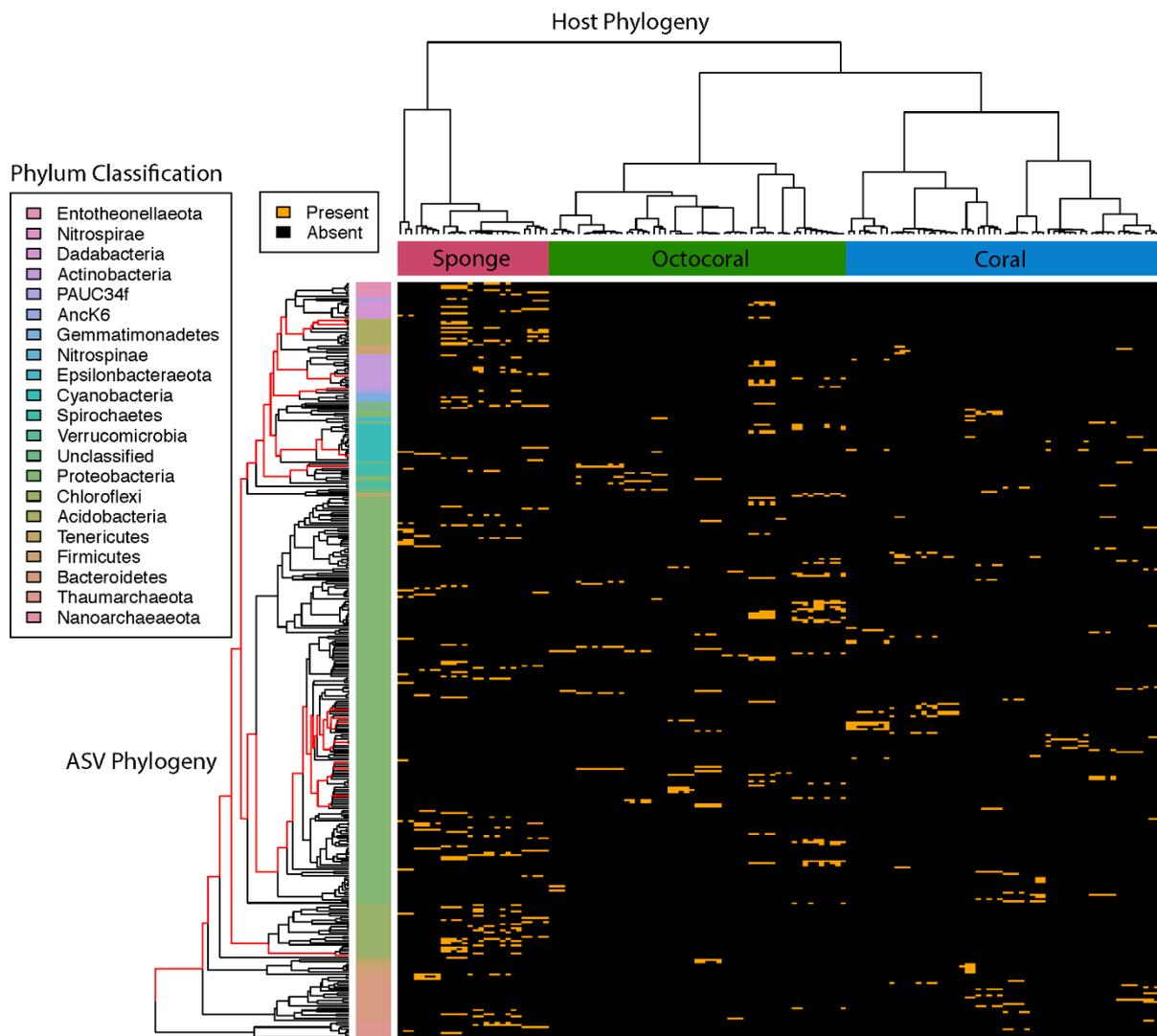


Figure 3.1. A presence-absence heatmap of each core ASV found across all host species samples. Rows are ordered by ASV phylogeny (y-axis dendrogram) and columns are order by host phylogeny (x-axis dendrogram). ASV phylogeny is coloured by phylum level classification while host phylogeny is coloured by phylum for sponges and subclass for corals and octocorals. Branches highlighted in red indicate ASVs that were significantly distributed by host phylogeny using local Moran's I spatial autocorrelation.

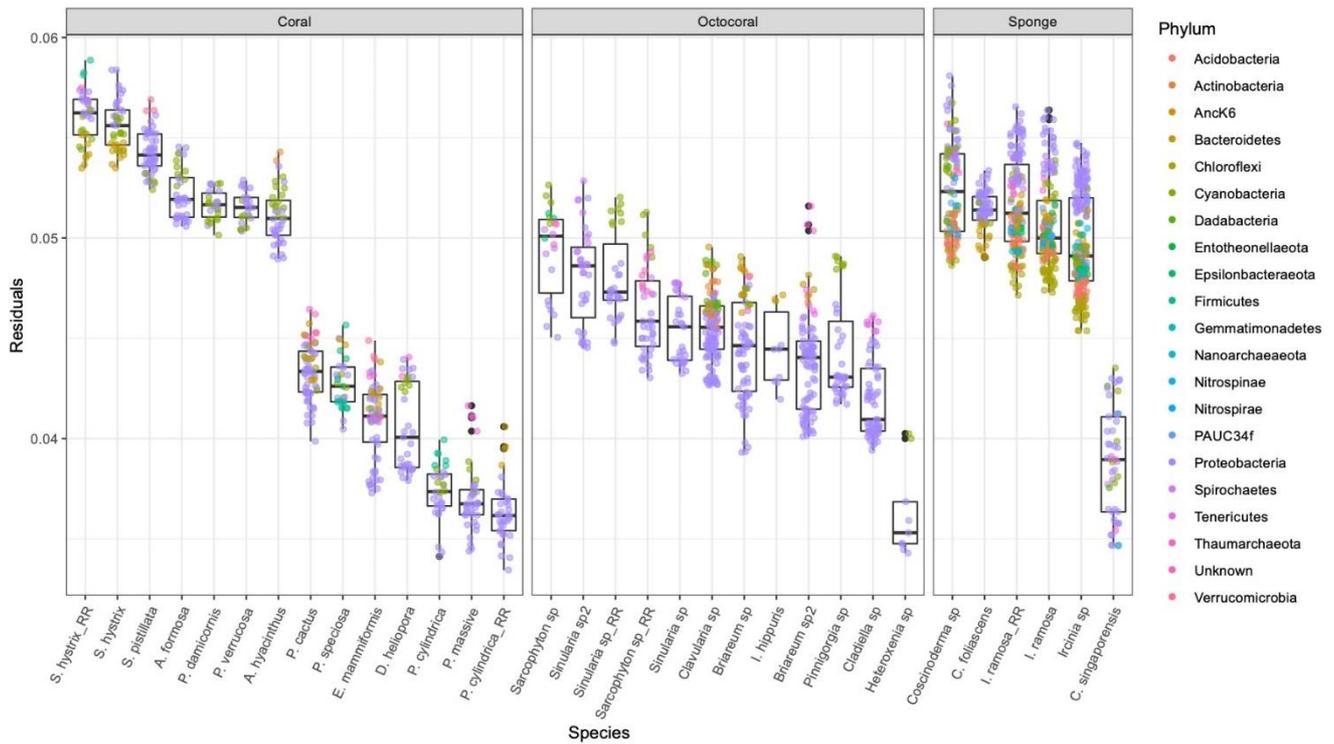


Figure 3.2. Residual contributions of each host-microbe link to the overall fit of cophylogeny. Boxplots are grouped by host species and residuals are overlaid and coloured by microbe phylum. Where a species was collected from two locations, ‘RR’ indicates the replicate from the ribbon reefs.

model above across all host species, five genera displayed a significant fit to cophylogeny ($p < 0.05$; Table 3.2, Table S3.2) while ten microbial families showed a significant fit to cophylogeny using both phylogenetic methods ($p < 0.05$; Table 3.2, Appendix B; Table S3.2). This increase is likely due to the higher number of ASVs being tested within each family compared to genus. Overall, five groups of microbes (Family level classification; *Endozoicomonadaceae*, *Spirochaetaceae*, *Nitrosopumilaceae*, *Thermoanaerobaculaceae* and *Microtrichaceae*) fitted a model of cophylogeny at both the genus and family taxonomic levels using both the fragment insertion and de novo phylogenies.

Table 3.2 Microbial genera and family that had a significant fit to the cophylogenetic model using both phylogenetic methods. Microbial taxa are ordered by mean sums of squares, with a lower value indicating a stronger fit. Mean p-value represents mean adjusted p-value of the 100 data subsets. Taxonomic classification is based on the Silva database.

Genus	Mean p-value	Std Dev	Mean SS	Std Dev	Mean p-value	Std Dev	Mean SS	Std Dev
	<i>Fragment insertion</i>				<i>De novo</i>			
<i>Spirochaeta 2</i>	0.015	0.043	0.512	0.134	0.013	0.029	0.550	0.113
<i>Subgroup 10; Thermoanaerobaculaceae</i>	0.006	0.010	0.647	0.100	0.024	0.038	0.715	0.069
<i>Candidatus Nitrosopumilus</i>	0.003	0.016	0.855	0.033	0.000	0.002	0.837	0.040
<i>Endozoicomonas</i>	0.000	0.000	0.864	0.009	0.000	0.000	0.898	0.005
<i>Sva0996 marine group; Microtrichaceae</i>	0.014	0.018	0.891	0.016	0.013	0.020	0.873	0.023
Family								
<i>Spirochaetaceae</i>	0.025	0.098	0.666	0.087	0.002	0.004	0.636	0.053
<i>Thermoanaerobaculaceae</i>	0.012	0.025	0.667	0.095	0.018	0.035	0.712	0.064
<i>Woeseiaceae</i>	0.000	0.000	0.687	0.054	0.000	0.000	0.667	0.062
<i>Microtrichaceae</i>	0.000	0.000	0.839	0.022	0.000	0.000	0.820	0.022
<i>Endozoicomonadaceae</i>	0.000	0.000	0.855	0.009	0.000	0.000	0.880	0.008
<i>Cyclobacteriaceae</i>	0.015	0.026	0.899	0.020	0.007	0.031	0.885	0.027
<i>Nitrosopumilaceae</i>	0.008	0.042	0.899	0.022	0.003	0.011	0.900	0.022
<i>Cyanobiaceae</i>	0.009	0.009	0.948	0.007	0.033	0.037	0.953	0.008
<i>Rhodobacteraceae</i>	0.000	0.000	0.956	0.007	0.001	0.003	0.963	0.006
<i>Flavobacteriaceae</i>	0.002	0.006	0.974	0.004	0.003	0.013	0.973	0.004

The local Moran's *I* measure of spatial autocorrelation (Anselin, 1995; Gittleman & Kot, 1990) identified if any particular ASVs were clustered according to host phylogeny. Across all host species, only fourteen ASVs were found to be significantly distributed by host phylogeny with local host phylogenetic clusters (Figures 3.1, 3.3 & 3.4). In support of the analysis above, ten of these ASVs belong to the microbial families / genera (all except *Nitrosopumilaceae*) that show strong levels of cophylogeny and further supports their putative close host association. The topology of the five groups of microbes with strong cophylogenetic patterns showed that *Endozoicomonadaceae*, *Spirochaetaceae* and *Microtrichaceae* exhibited clustering by host phylogeny when looking at the relative abundance of each ASV across all hosts (Figure 3.4). Analysis of the seawater controls showed that these ASVs were mostly absent from the surrounding environment, with only twelve of the *Endozoicomonadaceae* ASVs, three of the *Microtrichaceae* and none of *Spirochaetaceae* retrieved from seawater. The remaining two families, *Nitrosopumilaceae* and *Thermoanaerobaculaceae*, appeared to

include more clades with a generalist distribution despite having a significant fit to cophylogeny (Figure 3.4).

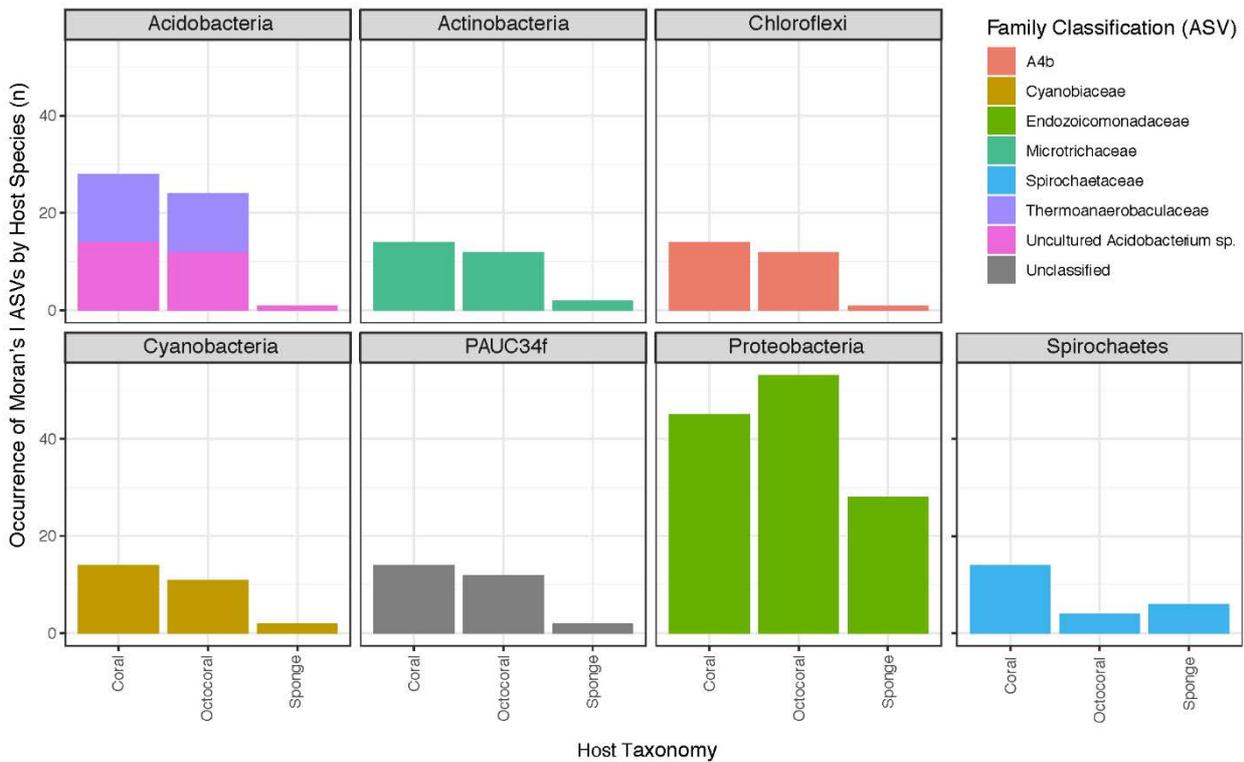


Figure 3.3. Taxonomic classification of ASVs that were significantly distributed by host phylogeny using local Moran's *I* measure of spatial autocorrelation. Occurrence refers to the number of host-ASV associations. *Endozoicomonadaceae* contains seven unique ASVs while the remaining families all contained one each.

3.4.4. Distinct microbial co-occurrence sub-networks within each host group

Network analysis identified co-occurring microbes across all samples in the core ASV dataset. Four communities of ASVs formed positively co-occurring sub-networks, with each host group dominated by a different sub-network (Figure 3.5; Appendix B; Figure S3.4). Sub-network 1 was the most diverse, comprising 24 microbial families and was almost entirely

located within the sponges. This sub-network also contained a dominant hub ASV affiliated to PAUC34f with relatively high centrality betweenness (Figure 3.5a). In contrast, sub-network 3 did not contain any ASV hubs and was mostly made up of ASVs affiliated to *Endozoicomonadaceae*, with a small proportion of ASVs affiliated to *Spirochaetaceae*, and was found exclusively within the octocorals (Figure 3.5b). Corals consisted mostly of sub-network 2, however this community was also present in the octocorals, and to a lesser extent the sponges (Figure 3.5b). Finally, sub-networks 1-3 consisted of a small number of ASVs that were significantly associated with host phylogeny using the Moran's *I* statistic (Table 3.3). Conversely, sub-network 4 was found in a comparatively small number of species (3 octocorals and 1 coral; Figure 3.5b) and contained only 3 ASVs that were connected by one edge each. Hence, this sub-network is not very robust and may in fact be part of sub-network 2 (Figure 3.5a).

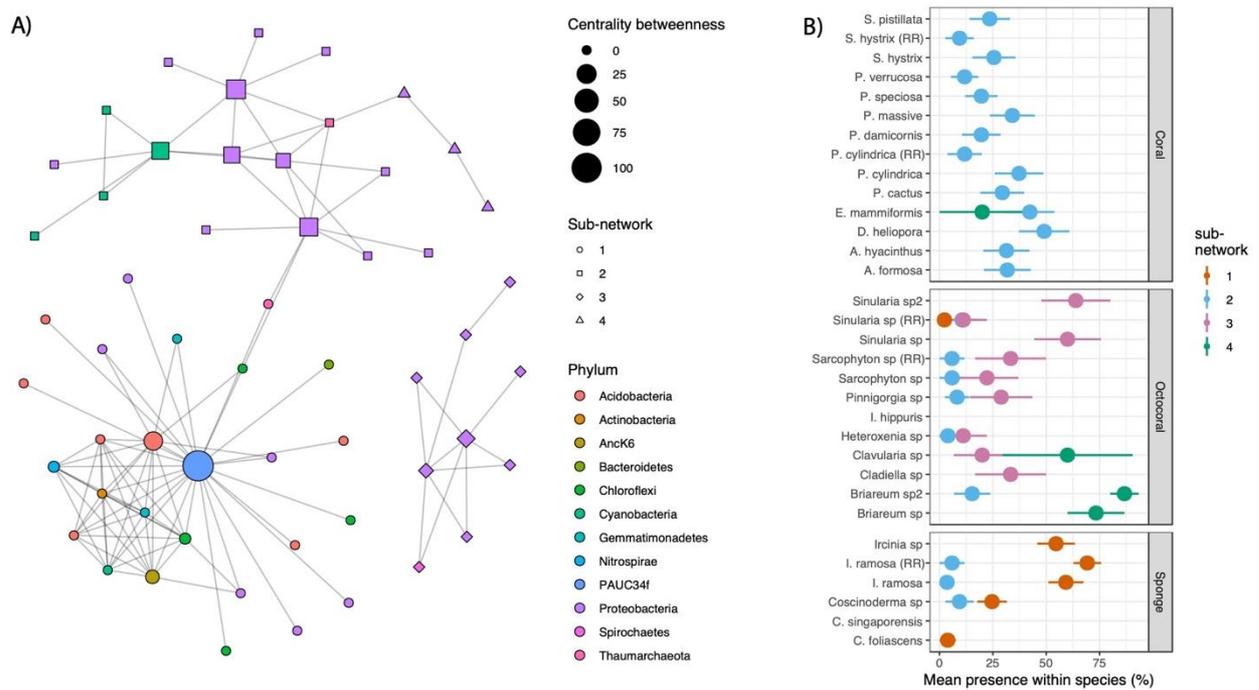


Figure 3.5. Four sub-networks of positively co-occurring ASVs within the core microbiome across all species. A) Presence-absence co-occurrence network with node shape reflective of each sub-network, node colour reflecting the phylum of the ASV node, and node size indicating centrality betweenness, i.e., how often the shortest path between two nodes traverses the focal node. B) Presence of each sub-

network within each host species. Circles represent the mean fraction of host species replicates the sub-network ASVs are found, lines represent the standard deviation. Where the same species has been collected from two locations, those collected from the Ribbon Reefs are indicated with 'RR'.

Table 3.3. Summary of the number of ASVs in each sub-network community and the percentage of those ASVs that were found to be significantly related to host phylogeny.

Sub-network	Number of ASVs	Number of Moran's / ASVs	Proportion Moran's / ASVs
1	26	6	23.08
2	17	3	17.65
3	9	4	44.44
4	3	0	0.00

3.5. Discussion

These results demonstrate that the community of microbes commonly associated with coral reef invertebrates fit a model of cophylogeny and a small number of microbial groups have strong cophylogenetic patterns when analysed independently. Further, each marine invertebrate group appears to be associated with sub-networks of co-occurring microbial communities, and these groups include microbial taxa that are phylogenetically clustered among hosts. Identification of important symbionts is a major challenge in any host species with high microbial complexity. Here I show that implementing a cophylogenetic approach is a powerful tool to explore the high microbial diversity inhabiting marine invertebrates, facilitating identification of groups of microbes that are more likely to have an intimate relationship with their host.

3.5.1. Cophylogenetic patterns suggest host-microbe codivergence and may contribute to patterns of phyllosymbiosis

Previously I observed that corals, octocorals and sponges show patterns of phylosymbiosis (O'Brien et al., 2020). However, it is unclear whether the mechanisms behind phylosymbiosis are due to codivergence or host-filtering processes (Mazel et al., 2018; Moran & Sloan, 2015). My cophylogenetic analysis supports that host-microbe codivergence is plausible in marine invertebrates and is confined to a small subset of microbial groups that likely have strong links to host fitness. Similar conclusions were drawn from thirty-two scleractinian taxa where a small number of bacterial genera, including *Endozoicomonas*, showed cophylogenetic patterns and were credited with contributing to the phylosymbiotic signal (Pollock et al., 2018). However, similar to patterns of phylosymbiosis, cophylogenetic patterns could still be driven by horizontally acquired microbes that are filtered through host phenotypic traits, particularly those traits that are congruent with host phylogeny (Lim & Bordenstein, 2020; Mazel et al., 2018). Therefore, although my observations are in line with what is expected during host-microbe codivergence, I cannot rule out that the phylosymbiotic and cophylogenetic patterns observed in these marine taxa are caused by host filtering processes.

3.5.2. *Cophylogenetic patterns reflect host generalist and host specialist microbial species*

The corals, octocorals and sponges tested here fit a model of cophylogeny, yet the strength of the signal was governed primarily by host species rather than by the microbe. Hence, ASVs affiliated to a particular taxonomy could have a relatively high or low contribution to the cophylogenetic pattern depending on which host the ASV was retrieved from. This may reflect differences in partner specialisation within a bacterial genus or family, encompassing both host generalist and host specialist species (Chomicki et al., 2020). This was observed when each microbial group was investigated individually, with different clades of cophylogenetic microbes showing either host specialist or generalist distributions. A similar observation was recorded in corals, where *Endozoicomonas* showed a strong cophylogenetic

signal despite a division into host generalist and specialist clades (Pollock et al., 2018). Further, *Thaumarchaeota* symbionts found in sponges consist of generalist and specialist taxa with varying degrees of metabolic specialisation to the host environment based on their host distribution (Zhang et al., 2019). However, caution is needed when interpreting residual patterns. Here, I define a core microbiota as ASVs that are present in 50% of host species replicates, with the aim of investigating patterns from microbes that are commonly associated with a host. Changing core thresholds has the potential to impact microbial patterns (Astudillo-García et al., 2020), and my 50% threshold was selected due to the low number of species replicates (3-5). Ideally, studies would use a higher number of replicates allowing for different core thresholds to be tested.

3.5.3. The strength of cophylogeny correlates with a host species tolerance to environmental stress

Host species demonstrated variable levels of cophylogeny (relatively stronger or weaker), and this showed an interesting correlation with the ability of a species to tolerate environmental disturbances (e.g. increases in water temperature/ $p\text{CO}_2$). This divide was most evident in corals, where species within the genus *Porites* had the strongest fit while those within the families Pocilloporidae and Acroporidae had the poorest fit to cophylogeny. *Porites* is often reported as a stress tolerant genus with a stable microbial community while pocilloporid and acroporid corals are more susceptible to environmental disturbances (Fabricius et al., 2011; Gardner et al., 2019; Hughes et al., 2018; O'Brien et al., 2018). Similarly, the sponge *Coelocarteria singaporensis* had the strongest cophylogenetic fit of all sponge species (when samples from all hosts were analysed) and is known to tolerate extreme environments with low pH (Botté et al., 2019; Morrow et al., 2014). However, it should also be acknowledged this sponge species was distantly related to the other sponges, which likely contributed to the

observed difference in residuals. Although microbial composition in corals and sponges has been correlated to environmental tolerance previously (Webster & Reusch, 2017; Ziegler et al., 2017, 2019), causal links to tolerance traits are still required and presents an important avenue for future research.

3.5.4. Five groups of microbes show a significant fit to cophylogeny when analysed independently

Independent assessment of the cophylogenetic fit for each microbial genus and family identified five groups of microbes with the potential for important host associations. *Endozoicomonas* has been reported as an important marine symbiont and my results support previous conclusions of strong cophylogenetic patterns in coral (Neave et al., 2016; Pollock et al., 2018; Van De Water et al., 2018). Genomic evidence indicates potential roles for this genus within the coral holobiont such as carbohydrate metabolism, amino acid synthesis and sulfur cycling through breakdown of both dimethylsulfoniopropionate (DMSP) and dimethyl sulfoxide (DMSO) (Neave, Michell, et al., 2017; Robbins et al., 2019; Tandon et al., 2020). However, the characterisation of the functional importance of *Endozoicomonas spp.* is hindered by the phylogenetic and metabolic diversity within the genus. Host relationships with *Endozoicomonas* are extremely diverse (Neave et al., 2016), and therefore how they interact likely differs between associations, with no general rule or model. Further, their relatively large genomes (2.3 – 6.3 Mb; (Neave, Michell, et al., 2017)a) and presence in the surrounding seawater imply that at least some *Endozoicomonas* species may have facultative rather than obligate relationships with their hosts (Neave, Rachmawati, et al., 2017; Weber et al., 2019).

Spirochaetaceae and *Nitrosopumiliaceae* are increasingly reported as important members of invertebrate microbiomes (Matcher et al., 2017; Robbins et al., 2019; Van De Water et al., 2018; Zhang et al., 2019). In non-marine hosts such as termites, *Spirochaetes* of

the genus *Treponema* (family: *Spirochaetaceae*) have been implicated in nitrogen fixation (Lilburn et al., 2001), while in the marine environment, *Spirochaeta* appear to metabolise carbon sources in the gill of lucinid clams (Lim et al., 2019). Further, *Spirochaeta* were identified as a key member of the holobiont in *Tsitsikamma* and *Cyclacanthia* sponges where cophylogenetic patterns were observed (Matcher et al., 2017) and metagenomic analysis of the sponge-associated Spirochaete suggested roles in vitamin B6 and pyrroloiminoquinone production (Waterworth et al., 2020). The Thaumarcheota genus *Nitrosopumilus* (family: *Nitrosopumiliaceae*) is characterised by its ability to convert ammonia to nitrite, providing an essential step in nitrogen cycling (Könneke et al., 2005). This genus has been reported as a key member of the sponge microbiota, where it was the only taxon identified capable of performing the ammonia oxidation step in nitrification (Engelberts et al., 2020; Robbins et al., 2021). Similarly, in the Antarctic sponge *Leucetta antarctica*, *Nitrosopumilaceae* genomes were found to encode ammonia monooxygenase, the key enzyme in the ammonia oxidation pathway (Moreno-Pino et al., 2020). Additional roles inferred from *Nitrosopumilaceae* genomes retrieved from corals and sponges include carbon fixation through the 3-hydroxypropionate/ 4-hydroxybutyrate (HP-HB) cycle (Engelberts et al., 2020; Moreno-Pino et al., 2020; Robbins et al., 2019), indicating *Nitrosopumilaceae* fulfil both unique and redundant functions within the sponge microbiome. The strong cophylogenetic patterns recovered in *Spirochaetaceae* and *Nitrosopumiliaceae* clearly suggest the relevance of these families in the invertebrate microbiome. Whether this association contributes positively to the fitness and survival of the host remains to be confirmed.

Microtrichaceae and *Thermoanaerobaculaceae* are less known as microbial symbionts but also showed strong cophylogenetic patterns. Although these bacterial families have been found in marine invertebrates previously (Friel et al., 2020; Sacristán-Soriano et al., 2020), no studies have yet assessed their functions within a host environment. Recent refinement of the

taxonomic structure of *Acidobacteria* places *Thermoanaerobaculaceae* as a novel family in subdivision 23, members of which are characterised as thermophilic and neutrophilic anaerobes that may use sugars, organic acids and proteinaceous compounds as growth substrates (Dedysh & Yilmaz, 2018). The *Microtrichaceae* are placed within the *Actinobacteria* and although this phylum contains many known symbionts associated with hosts as diverse as insects to mammals (Lewin et al., 2016), there is currently insufficient information to infer any role for this family within these reef invertebrates. Importantly, the cophylogenetic patterns associated with the microbes highlighted here are found across a diverse group of host species. Future research would benefit from uncovering these patterns within a larger number of closely related host species, to identify if the same symbionts show cophylogenetic patterns at different levels of host taxonomy.

3.5.5. *Limitations of cophylogenetic studies*

This cophylogenetic analysis has uncovered intriguing patterns related to host-microbe symbiosis, however it is important to acknowledge two potential caveats. Firstly, cophylogenetic studies have previously used the V4 region of the 16S rRNA gene to profile the microbiome of a large number of host species (Pollock et al., 2018; Youngblut et al., 2019). While this is currently the most practical approach, amplicons are short marker sequences with limited phylogenetic information. I attempted to strengthen my results by using two phylogenetic methods, however interpretation must still be cautious with an understanding that this estimation of microbial phylogeny is based on changes in the V4 region of the 16S rRNA gene. Secondly, the null hypothesis of a cophylogenetic model is that there is no relationship between the host and microbe distance matrices. Therefore, neither the nature of the symbiosis, eg, mutualism, parasitism or commensalism, nor the mechanism responsible for that pattern is revealed. For example, cospeciation occurring in the gut microbiota of mammals can be

explained by allopatric speciation of the host, associated with limited dispersal of symbionts and subsequent diversification of bacteria (Groussin et al., 2020). While these host species are not geographically isolated from one another, the host environment likely acts as a barrier to the microbial taxa that can successfully establish, potentially promoting the radiation of certain taxa within a host group. Moreover, although the above patterns suggest these symbionts are linked to host health, the effect may not be positive. For example, *Spirochaetes* are well suited to a parasitic lifestyle (Johnson, 2013) and their cophylogenetic distribution may indicate a harmful symbiosis rather than a beneficial one. Parasitic and mutualistic symbioses are equally important when trying to understand the impact of a symbiosis on host fitness and I suggest that future studies seek to understand the nature of the symbioses identified here.

3.5.6. Host groups are associated with unique sub-networks of co-occurring microbes

Along with cophylogenetic patterns, I found that microbes commonly associated with these reef invertebrates positively co-occurred in sub-networks that are distributed by host taxonomy. Of particular interest is sub-network 3, which consisted entirely of ASVs affiliated to *Endozoicomonadaceae* and *Spirochaetaceae*, had the highest proportion of ASVs with strong phylogenetic distributions and was found exclusively in the octocorals. These two bacterial groups are often found within octocorals, particularly those within the family *Alcyoniidae*. For example, *Lobophytum pauciflorum* is a close relative of *Sinularia* and *Sarcophytum* represented in this study, and the microbiota associated with this species is dominated by *Endozoicomonadaceae* and *Spirochaetaceae* (Wessels et al., 2017). Sub-network 1 was dominant in the sponges and possessed a hub ASV affiliated to PAUC34f, which may indicate a key member within the sponge microbiota as it is closely connected to other members. Previous studies have implicated co-occurring microbes as potential functional guilds (Youngblut et al., 2019), however further research is needed to conclude whether

Endozoicomonadaceae and *Spirochaetaceae*, along with other cooccurring taxa, exploit the same host resources or form syntrophic communities. It must also be acknowledged that co-occurrence patterns may arise due to shared habitat preferences (Berry & Widder, 2014), and thus the sub-networks observed here could be a product of the host environment in which they are found. Further, bacteria within coral and sponge tissues are known to form aggregates and the co-occurrence of ASVs here may represent these structures (Maldonado, 2007; Wada et al., 2019). Overall, network analysis suggests each of these host groups are associated with a unique sub-network of co-occurring microbes and include microbial taxa with strong host phylogenetic distributions.

3.5.7. Conclusions

The cophylogenetic patterns evident in reef invertebrates indicates congruent divergence between host species and a small number of symbionts and points to well-established relationships that have persisted through time. Further, reef species that are typically known to tolerate increases in water temperature and $p\text{CO}_2$ showed relatively stronger cophylogenetic patterns. While not the focus of this study, future research may benefit from exploring how the evolution of the microbiome is related to strategies of host survivorship. Finally, I reveal that sub-networks of co-occurring microbes are confined to particular host groups and include microbial taxa that are strongly distributed by host phylogeny. Taken together, I identify a structured microbial symbiosis within the complexity of coral reef invertebrate microbiomes that likely represent key members and underpin host fitness.

**Chapter 4: Validation of key sponge symbiont pathways using
genome-centric metatranscriptomics**

4.1. Summary

Amplicon sequencing was used in the previous chapters to identify patterns of phyllosymbiosis and codivergence across diverse marine invertebrates. This approach was well suited to analyse the large sample sizes (e.g. >100) but did not allow for inference of microbial function. The aims of this chapter were to a) identify the similarity and differences in metabolic pathways within the microbiomes of different sponges and b) identify whether functional predictions from metagenomic data are accurate and validated by gene expression profiles. Genome-centric metatranscriptomics was used to characterise microbial metabolism within the sponges *Ircinia microconulosa* and *Phyllospongia foliascens*, as well as a metagenomic characterisation of *Ircinia ramosa*. Expression of entire pathways for carbon fixation and multiple sulfur compound transformations were observed, along with gene expression of complete gene pathways for nitrification, denitrification and nitrate reduction in *I. microconulosa*. Expression of the pathway for vitamin B biosynthesis was common and spread across many microbial phyla, however in some cases only the partial pathway was observed, and multiple microbial taxa were needed for complete biosynthesis. This work emphasises the dynamic nature of metabolic events across the sponge microbiota and provide further evidence that the microbiota contribute to the health of their sponge hosts via nutrient exchange.

4.2. Introduction

Marine sponges are key members of coral reef ecosystems. By filtering large volumes of seawater, sponges make dissolved organic matter (DOM) available to higher trophic levels, thereby providing a nutrient source in an otherwise nutrient poor environment (De Goeij et al., 2013). Part of that trophic link is performed by the sponge-associated microbes. For instance, in high-microbial abundance (HMA) sponges, up to 87% of DOM can be assimilated by its

microbial community, demonstrating the importance of the microbiota in underpinning the ecological function of sponges (Rix et al., 2020).

Sponges are one of the earliest diverging animal lineages and the sponge-microbe symbiosis likely represents one of the most basal host-microbe symbioses. This long-standing host-microbe relationship has likely contributed to the species specific and stable microbial associations observed today (Thomas et al., 2016). For example, the sponge microbiota shows stability under environmental disturbances such as changes in water quality, temperature and salinity (Glasl et al., 2018; Luter et al., 2014; Strand et al., 2017). Furthermore, there is a strong signal of phylosymbiosis in sponges, where the microbiota demonstrates relatively low community variation within a host species and is more dissimilar as host species become more divergent (O'Brien et al., 2020; Thomas et al., 2016). Within the sponge microbiota, a number of microbial lineages have shown strong cophylogenetic signals, indicating a tightly coupled host-microbe relationship that underpins the health of the host (Matcher et al., 2017; O'Brien et al., 2021). Additionally, in some cases sponges vertically transmit a proportion of their microbes, ensuring the symbiosis persists through generations (Björk et al., 2019; Schmitt et al., 2008; Usher et al., 2001). This well-established, deeply divergent and tightly coupled host-microbe relationship, along with the importance of their microbiota, make sponges an ideal choice for host-microbe symbiosis research.

A wealth of knowledge of putative sponge-associated microbial activities has been generated through metagenomic studies. For example, pathways involved in carbon fixation, carbohydrate metabolism, sulfur and nitrogen cycling as well as biosynthesis of many B vitamins are encoded in sponge associated microbial genomes and putatively supplement host metabolism (Engelberts et al., 2020; Kamke et al., 2013; Moreno-Pino et al., 2020; Robbins et al., 2021; Thomas et al., 2010). In addition, the genomes of many sponge symbionts are

enriched in genes that potentially facilitate life in a host environment, such as eukaryote-like proteins (ELPs) and restriction enzymes, which help symbionts evade phagocytosis and defend against harmful mobile genetic elements that might otherwise inhibit survival within a host (Reynolds & Thomas, 2016; Robbins et al., 2021). Further, some sponge symbionts contain the genomic repertoire for secretion systems (Engelberts et al., 2020; Robbins et al., 2021), which can be used by both beneficial and pathogenic symbionts to interact with host cells (Costa et al., 2015).

Metagenomic studies are highly valuable for inferring functional potential from DNA coding sequences in the microbiome, however application of metatranscriptomics, which employs the sequencing and analysis of microbial mRNA to establish putative microbial activity, can be used to validate inferences based on metagenomic data. Metatranscriptomics has been applied to show expression of microbially mediated B vitamin and amino-acid synthesis as well as nitrification and carbon fixation in sponges (Fiore et al., 2015; Radax et al., 2012). However, these prior studies used a gene-centric approach, precluding direct linkage of gene activity and microbial taxonomy. A genome-centric approach identifies which microbial genome is responsible for the observed transcriptomic activity and can elucidate potential syntrophy. For example, genome-centric metatranscriptomics was used to link the oxidation of ammonia to nitrite by sponge-associated Thaumarchaeota symbionts and the subsequent oxidation of nitrite to nitrate by members of the bacterial genus *Nitrospira* (Moitinho-Silva et al., 2017).

In this chapter, I used metagenomic sequencing of three common Great Barrier Reef sponges, *Ircinia ramosa*, *Ircinia microconulosa* and *Phyllospongia foliascens* (formerly *Carteriospongia*; (Abdul Wahab et al., 2021) to identify metabolic pathways potentially utilised by sponge symbionts. Additionally, I used parallel metatranscriptomic sequencing of

I. microconulosa and *P. foliascens* followed by genome-centric analyses to validate microbial functions. These sponge species were selected for their high microbial abundance (HMA), which indicates a role for the microbiome in supporting the host ecological function (Rix et al., 2020). Further, *I. microconulosa* is characterised by a distinct copper blue throughout its mesohyl, due to an uncharacterised, vertically transmitted microbial symbiont (Wahab et al., 2016). *Phyllospongia* are phototrophic sponges where vertically transmitted cyanobacterial symbionts potentially contribute up to 50% of the sponge's energy requirements (Abdul Wahab et al., 2021). Notably, *Ircinia microconulosa* and *Phyllospongia foliascens* have not previously been characterised using genome-centric metatranscriptomics.

4.3. Materials and methods

4.3.1. Sample collection and preservation

Five replicates of three sponge species (*Phyllospongia foliascens*, *Ircinia ramosa* and *Ircinia microconulosa*) were collected at a depth between 5 to 10 m from Davies Reef (-18.82°, 147.65°) and Broadhurst Reef (-18.97°, 147.72°) in the central Great Barrier Reef (GBR) Australia in August 2017. An additional five replicates of *I. ramosa* were collected from the Ribbon Reefs (-14.79°, 145.69°) in the northern sector of the GBR, Australia in October 2017. Sponges were placed in holding tanks with running seawater on the research vessel until sampling and returned to the reef once completed. A small section of each specimen was dissected and rinsed in autoclaved calcium- and magnesium-free seawater (CMFSW; NaCl 26.2g, KCl 0.75g, Na₂SO₄ 1g, NaHCO₃ 0.042g), then cut into small pieces approximately 0.5 x 0.5 cm². Dissected pieces were then added to a 15 mL falcon tube half filled with dimethylsulfoxide-EDTA salt saturated solution (DESS) until the tube was full, reaching a 1:1 DESS:tissue ratio and stored at -20°C for metagenomic analysis. Three replicates from *P.*

foliascencs and *I. microconulosa* were additionally preserved for metatranscriptomic analysis using the same approach.

4.3.2. Microbial cell enrichment by cell separation for metagenomics

Separation of Prokaryotic and Eukaryotic cells was performed following the method of (Thomas et al., 2010) with minor modifications. Briefly, approximately 2-4 g of sponge sample was transferred to a 50 mL falcon tube containing 15 mL of autoclaved CMFSW and agitated for 5-10 mins at 99 rpm at room temperature using the intelli-mixer RM-2 (Bartelt Instruments Pty Ltd) to remove loosely attached microbes. The washed sample was transferred to 2 x 15 mL Falcon tubes half filled with fresh autoclaved CMFSW and homogenised using the Bio-Gen PRO200 Rotary Homogenizer (Pro Scientific Inc) for 10 mins. The homogenate was pooled and filter-sterilised collagenase added to a final concentration of 0.5 mg/mL. Samples were agitated using the intelli-mixer RM-2 at 99 rpm for 30 min to further break apart the sponge cells. The sponge homogenate was then filtered through a 100 µm sieve and collected into a new Falcon tube, centrifuged at 100 x g for 15 min at 4°C and then centrifuged twice at 300 x g for 15 min at 4°C. The supernatant was filtered twice through an 8 µm and twice through a 5 µm filter and then centrifuged for 20 min at 8,800 x g at 4°C to pellet the microbial cells. Two wash steps were performed by adding 10 mL of 10 mM Tris-NaCl, pH 8 to the pellet and then resuspended by vortex for 5 sec, followed by centrifugation for 20 min at 8,000 x g at 4°C to re-pellet the cells. Finally, the cell pellet was recovered in 1 mL of 10 mM Tris-NaCl, pH 8 and divided into two 500 µl aliquots and stored at -20°C.

4.3.3. Total DNA extraction for metagenomics

One 500 µl aliquot of cell separated sponge tissue was used for DNA extraction. Contaminating host DNA was first reduced by the addition of 1 µl of DNAase 1 (NE bio labs),

50 μ l of DNase buffer and incubated for 30 min at 37°C. DNAase 1 was heat inactivated by adding 5 μ l of 0.5 M EDTA and incubated at 75°C for 10 min and then proceeded directly to DNA extraction.

Microbial cells were pelleted at 20,000 x g for 5 min at 4°C, supernatant discarded, and cells resuspended in 500 μ l of Lysis buffer (50 mM Tris-HCl, 40 mM EDTA, 0.75 M Sucrose, MilliQ water). 75 μ l of lysozyme (100 mg/mL) was added and samples incubated for 1 hour at 37°C. Microbial cells then underwent 3x freeze-thaw cycles using liquid nitrogen and a heat-block set to 65°C. 100 μ l of sodium dodecyl sulfate (SDS; 10% solution) was added and incubated at 65°C for 3 mins. After cooling to room temperature, 7 μ l of RNase A (10 mg/mL) was added and samples incubated for 30 min at 37°C. Proteinase K (20 μ l, 20 mg/mL) was added and samples further incubated for 1 hour at 37°C. Equal volumes of Phenol:Chloroform:IAA (25:24:1) was added, mixed by inversion and centrifuged at 16,000 x g for 10 min. The aqueous phase was recovered and a second Phenol:Chloroform:IAA phase separation step was performed as above. The above phase separation steps were repeated using equal volumes of Chloroform:IAA (24:1) and aqueous phase recovered. 50 μ l of Sodium Acetate (3 M) was added and DNA precipitated by adding equal volumes of cold Isopropanol and incubated at 4°C overnight. Precipitated DNA was pelleted at 20,000 x g for 30 min at 4°C, supernatant removed and 1 mL of 75% ethanol added to wash the pellet. DNA was spun again at 18,000 x g for 10 min at 4°C, the supernatant discarded, and a second wash step was performed. Following removal of ethanol, samples were air dried for 10 min and DNA resuspended in 10 mM Tris-HCl and cleaned using the Zymo gDNA Clean and Concentrator kit following the manufacturers protocol. DNA was quantified using the Quantus Fluorometer (Promega) and quality checked using the NanoDrop 2000 (Thermofisher) and running a small aliquot on a 1.5% agarose gel.

4.3.4. Total RNA extraction and microbial mRNA enrichment for metatranscriptomics

Approximately 0.25 g of sponge tissue was crushed using liquid nitrogen and a mortar and pestle and total RNA isolated from crushed tissue using the RNeasy PowerMicrobiome Kit (Qiagen) following the manufacturer's protocol. Total RNA was cleaned using the RNeasy MinElute Cleanup kit (Qiagen) and quality and quantity of total RNA were checked using the Agilent 2200 TapeStation on the High Sensitivity RNA ScreenTape system. DNA was isolated from approximately 0.05 g of sponge tissue using the DNeasy PowerBiofilm Kit (Qiagen) as per the manufacturer's protocol for the purpose of synthesising rRNA probes.

Microbial mRNA enrichment was performed following an adapted protocol from (Stewart et al., 2010). In summary, biotinylated host species-specific rRNA probes were synthesised by PCR amplification of the LSU and SSU rRNA coding regions of bacterial, archaeal and eukaryotic gDNA from corresponding samples (Appendix C; Table S4.1). Each sample was amplified 5 times then pooled and cleaned using the NucleoSpin Gel and PCR clean-up kit (Macherey-Nagal). Equal volumes of each host-sample replicate were then combined to create host-species specific anti-sense RNA probes. Biotinylated aRNA probes were then hybridised with the complementary rRNA in the total RNA sample, and the hybridised rRNA subtracted using streptavidin-coated magnetic beads. Microbial mRNA was then further enriched by subtraction of poly A-tailed (eukaryotic) mRNA using oligo(dT)-coated magnetic beads. Enriched microbial mRNA was cleaned using the RNeasy MinElute Cleanup kit (Qiagen) and verified using the Agilent 2200 TapeStation on the High Sensitivity RNA ScreenTape system to ensure signature rRNA peaks had been successfully reduced. The remaining mRNA was amplified using the MessageAmp II Bacteria kit (Ambion) by polyadenylation of the mRNA followed by reverse transcription using the T7-BpmI-d(T)16-VN primer. The resulting antisense RNA was then transcribed to double stranded cDNA using

the SuperScript III First Strand Synthesis System (Invitrogen) for first strand synthesis and the SuperScript Double Stranded cDNA kit (Invitrogen) for second-strand cDNA synthesis. Finally, poly-A tails were digested from cDNA using a BpmI restriction enzyme and subsequently cleaned using the NucleoSpin Gel and PCR clean-up kit. For full details of the protocol including reagents and kits purchased, refer to the supplementary material of (Sato et al., 2017).

4.3.5. Library preparation and sequencing

Metagenomic and metatranscriptomic libraries were prepared using the MGIEasy Universal DNA Library Prep Set (BGI, Shenzhen, China) and sequenced on the MGISEQ-2000 platform under the 100 bp paired-end mode at BGI Australia (Queensland, Australia). An additional technical replicate of sample CS70 (*Ircinia ramosa*, Ribbon Reefs) was prepared for metagenomic sequencing, resulting in a total of 21 metagenomic libraries and 6 metatranscriptomic libraries being sequenced. Both replicates of sample CS70 were subsequently used for assembly and binning.

4.3.6. Bioinformatics pipeline for metagenomic analysis

Raw reads were pre-processed using SOAPnuke v2.1.0 (Y. Chen et al., 2018) to remove adaptor sequences and low-quality reads and each sample was assembled individually using the metaSPAdes option in the SPAdes genome assembler v3.14.0 (Nurk et al., 2017). Cleaned reads from each sample were mapped back to each assembly of the same host species using CoverM v. 0.5.0, utilising the short-read option within minimap2 alignment program (H. Li, 2018). The resulting BAM files were used for differential coverage estimation using the `jgi_summarize_bam_contig_depths` script implemented through MetaBAT v2.15 (Kang et al., 2019). Metagenomic binning was performed for each assembly using both MetaBAT v1 (Kang

et al., 2015) (with three different parameters: sensitive, specific and superspecific) and MetaBAT v2 executed through MetaBAT v2.15, as well as MaxBin v2.2.7 (Y.-W. Wu et al., 2016) and CONCOCT v1.1.0 (Alneberg et al., 2014), giving a total of six sets of bins per sample. A non-redundant final set of bins was then selected using DAS tool v1.1.2 (C. M. K. Sieber et al., 2018), by ranking each genome based on a single-copy gene scoring function. Quality of each of the refined set of bins was further scrutinised using CheckM v1.1.2 (Parks et al., 2015) and finally all bins were consolidated and dereplicated at 95 average nucleotide identity (ANI) using CoverM v 0.5.0, retaining only those bins with >50% completion and <10% contamination.

To assign each metagenome assembled genome (MAG) taxonomy and obtain a phylogenomic tree of MAGs, the Genome Taxonomy Database software toolkit (GTDB-tk) was used by running the ‘Classify’ workflow. Briefly, this pipeline identifies 120 and 122 bacterial and archaeal marker genes respectively for multiple sequence alignment and phylogenetic inference. Genome classification is achieved by placement of each genome into the GTDB-tk reference tree using pplacer (Matsen et al., 2010). Genome coverage (relative abundance) was calculated in CoverM v0.5.0 using the ‘coverm genome’ command with default parameters, which mapped the cleaned reads from each sample back to the final set of 95% ANI dereplicated MAGs.

4.3.7. Bioinformatics pipeline for metatranscriptomic analysis

Raw reads were quality checked using SOAPnuke v2.1.0 to remove adaptor sequences and low-quality reads and further pre-processed to remove rRNA reads using SortMeRNA v2.1 (Kopylova et al., 2012). A custom reference database for each host species was configured by taking the full set of MAGs from each host species and dereplicating each separately at 95ANI (as described above), and finally concatenating into a single reference file for each host species.

Cleaned metatranscriptomic reads from each host species were aligned to the reference file of the same host species to produce BAM files using CoverM v0.5.0 with the minimap-2 alignment program (as above). BAM files were sorted by name using samtools v1.11 (H. Li et al., 2009) and reference files annotated using EnrichM v0.6.3 (Boyd et al., 2019) to obtain a gff file. Both the sorted bam files and gff annotation files were then imported into HTSeq v0.9.1 (Anders et al., 2015) to quantify the number of reads aligned to each gene in the annotation file.

Following rRNA removal, reads were assembled using Trinity v2.9.1 (Grabherr et al., 2011) to provide an understanding of taxonomic classification of all mRNA reads and therefore assess eukaryotic contamination. Predicted genes with a length ≥ 100 bp from assemblies were mapped against the National Centre for Biotechnology Information nonredundant (NCBI-nr) database using DIAMOND (Buchfink et al., 2015) and taxonomic classification identified using BLAST (Altschul et al., 1990) after filtering genes with an E-value $> 1e-5$, bit score < 60 , percent identity $< 50\%$ and mapped length $< 50\%$ for queries.

4.3.8. Metabolic pathways reconstruction

A list of common metabolic pathways found in sponge associated microbes was compiled based on previous work by (Engelberts et al., 2020) and refined by manually curating each pathway based on their Kyoto Encyclopedia of Genes and Genomes (KEGG) maps. These included six prokaryote autotrophic carbon fixation pathways (Wood-Ljungdahl pathway, Dicarboxylate/4-hydroxybutyrate cycle, 3-hydroxypropionate bicycle, 3-hydroxypropionate/4-hydroxybutyrate cycle, reductive citric acid cycle and the Calvin-Benson-Bassham cycle), nitrogen metabolism (nitrogen fixation, nitrification, denitrification and dissimilatory nitrate reduction), sulfur metabolism (taurine transport and oxidation, dissimilatory sulfate reduction and thiosulfate oxidation) and B-vitamin biosynthesis.

Annotation files from both metagenomic and metatranscriptomic datasets were subset to the genes of interest and then combined using a custom R script. The final dataset consisted of each MAG and their gene annotation, and indicated whether the gene was absent, found but not expressed or expressed. Lastly, while the metatranscriptomic analysis is focussed on *I. microconulosa* and *P. foliascens*, I additionally include a description of the metagenome of *I. ramosa* as a number of genomes are used for comparative analysis in the subsequent chapter 5.

4.4. Results

4.4.1. Sequencing, metagenomic binning and taxonomy

Sequencing effort yielded a total of 243 – 310 million reads from each metagenomic library ($n = 21$) while sequencing of each metatranscriptomic library yielded 541 – 663 million reads ($n = 6$). A total of 1290 metagenome assembled genomes (MAGs), including 1195 med-high quality MAGs (>50% completion <10% contamination), were retrieved across the three species of sponges. Following dereplication, the total number was reduced to 781 med-high quality MAGs at 99% average nucleotide identity (ANI) and 415 med-high quality MAGs at 95% ANI (Appendix C; Table S4.2). Mapping raw reads from each sample to the 415 MAGs (95 ANI) showed that between 75.6% and 91.9% of reads successfully aligned to the final set of MAGs, indicating a good representation of the total microbial community in each sample (Table S4.2). This was highest for *I. microconulosa* with $89.0 \pm 0.9\%$ (mean \pm standard error) of reads mapping to the metagenome and lowest for *I. ramosa* with $82.4 \pm 2.2\%$, while $86.0 \pm 0.4\%$ of reads from *C. foliascens* mapped to the metagenome.

The 415 MAGs across all three sponge species represented 21 bacterial phyla and one archaeal phylum classified as Thermoproteota (formerly Thaumarchaeota) (Figure 4.1).

Proteobacteria was the most diverse phylum, consisting of 154 MAGs (37.1%), while the archaeal phylum included only two MAGs and was found exclusively in the *I. ramosa* microbiota at a low relative abundance of $0.3 \pm 0.1\%$. Cyanobacteria was the most abundant phylum in *I. ramosa* representing $28.7 \pm 10.7\%$ of the microbiota, with one MAG from the genus *Synechococcus* making up $18.70 \pm 2.91\%$ of the relative abundance. Chloroflexota were

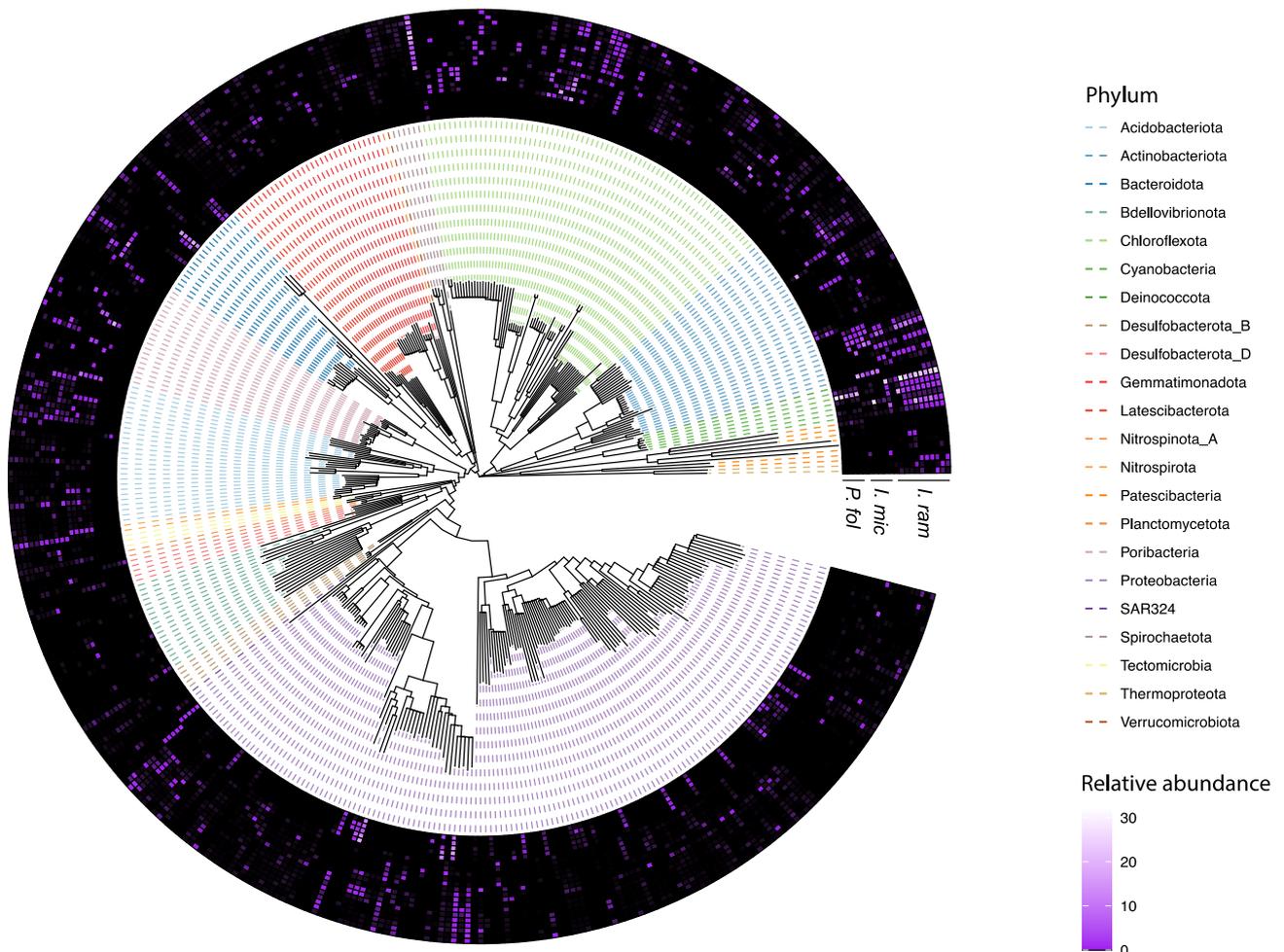


Figure 4.1. Phylogeny of 415 MAGs isolated from *I. ramosa*, *I. microconulosa* and *P. foliascens* illustrating the large diversity of microbial taxa found across the three sponge species. Tree clades are coloured by microbial phylum while outer heatmap indicates the relative abundance of each MAG in each sponge sample.

also highly abundant in the *I. ramosa* microbiota with a relative abundance of $19.7 \pm 2.4\%$. Within the Chloroflexota, one MAG classified as family A4b was particularly dominant and made up $10.0 \pm 2.1\%$ of the microbiota. Finally, Actinobacteria and Proteobacteria each represented significant contributions to the overall *I. ramosa* microbiota, with a relative abundance of $11.2 \pm 2.9\%$ and 9.9 ± 0.9 respectively. The microbiota of *I. microconulosa*, like *I. ramosa*, was dominated by Chloroflexota and Actinobacteriota with a relative abundance of $33.5 \pm 1.2\%$ and $23.9 \pm 4.9\%$ respectively. Cyanobacteria were however less dominant, making up only $5.7 \pm 5.5\%$ of the microbiota, while Poribacteria showed an increase in relative abundance compared to *I. ramosa*, with $9.6 \pm 3.6\%$. *P. foliascens* was largely dominated by Proteobacteria, making up $42.3 \pm 5.1\%$ of the microbiota. Cyanobacteria also contributed heavily with a relative abundance of $20.3 \pm 3.0\%$ with one dominant *Synechococcus* MAG contributing $19.8 \pm 2.6\%$ of the sequenced community. *P. foliascens* further differed from the other two sponges with Bacteroidota having a high relative abundance of $17.2 \pm 3.2\%$, which included the second most abundant MAG overall, at $13.5 \pm 1.6\%$.

4.4.2. Metatranscriptomic mapping results

Following rRNA removal, metatranscriptomic reads were reduced to 201-296 million reads per sample. Mapping the remaining mRNA reads to the corresponding metagenomic reference files revealed $81.0 \pm 1.6\%$ of reads did not align to the *P. foliascens* metagenome while $85 \pm 3.3\%$ of reads did not align to the *I. microconulosa* metagenome (Figure 4.2), likely due to eukaryotic mRNA contamination, based on taxonomic classification of assembled metatranscriptomic contigs (Appendix C; Table S3). Additionally, the strict mapping criteria for mRNA reads and potential for errors in identifying coding sequences in the metagenome may further reduce mapping success (L. Hao et al., 2020). Of the mRNA reads that successfully

aligned, $3.1 \pm 0.2\%$ and $4.3 \pm 1.7\%$ of reads in *P. foliascens* and *I. microconulosa* respectively were ambiguously aligned (i.e., mapped to multiple genes) and were therefore excluded.

Analysis of the successfully aligned reads showed that taxon-specific gene expression was disproportionate to the relative abundance of microbial genomes (Figure 4.2), indicating that the most abundant taxa do not necessarily contribute most to microbial activity. Gene expression was most active in Cyanobacteria in the *P. foliascens* microbiome, making up $76.0 \pm 3.9\%$ of all reads that aligned to the *P. foliascens* metagenome (Figure 4.2). Proteobacteria and Bacteroidota genes were also highly expressed making up $20 \pm 2.9\%$ and $4 \pm 0.73\%$ of total gene expression respectively. Gene expression in the *I. microconulosa* microbial community was more broadly distributed compared to *P. foliascens*, which is also reflective of its more diverse microbial community. Microbial activity was largely associated with Poribacteria and Proteobacteria, consisting of $22 \pm 10.4\%$ and $21 \pm 0.6\%$ of total gene expression respectively. While species replicates generally had a consistent level of gene expression based on microbial taxonomy, one *I. microconulosa* replicate showed a lower expression of Poribacteria and a higher expression of Cyanobacteria compared to the other two replicates (Figure 4.2).

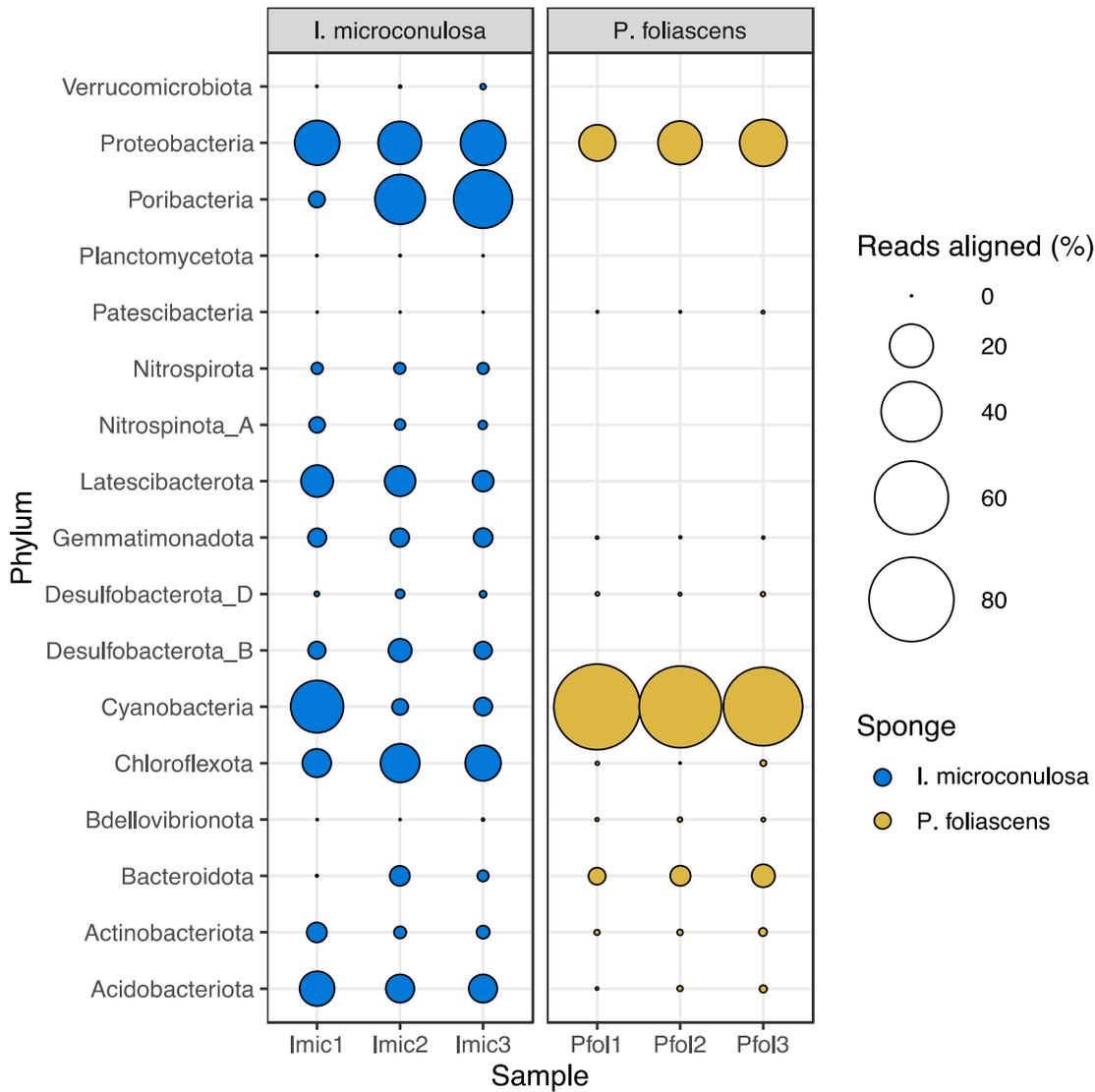


Figure 4.2. Relative abundance of metatranscriptomic reads that successfully aligned to the reference MAGs illustrating a larger diversity of microbes are active within the *I. microconulosa* microbiota compared to *P. foliascens*. Reads are grouped by microbial phylum and the size of the circle is proportional to the relative abundance. *I. microconulosa* is represented in blue while *P. foliascens* is represented in yellow.

4.4.3. Nutrient cycling in the sponge microbiome

All three sponge microbiomes had the metabolic potential for carbon fixation through the Calvin cycle and complete expression of this pathway was observed in *Synechococcus* (Phylum: Cyanobacteria) symbionts in both *I. microconulosa* and *P. foliascens* (Figure 4.3; Appendix C; Figure S4.1). Similarly, most of the genes necessary for carbon fixation through

the reductive citric acid (rTCA) cycle were present in the metagenomes of both *Ircinia* species. In the case of *I. microconulosa*, most of the rTCA transcripts mapped to Nitrospirales MAGs (Phylum: Nitrospirota), with the exception of 2-oxoglutarate synthase. Transcripts for Nitrospinota symbionts also expressed all steps with the exception of the genes for pyruvate and 2-Oxoglutarate production as well as the conversion of citrate to acetyl CoA. *P. foliascens* symbionts expressed the majority of steps in the rTCA cycle but no MAGs contained the genes necessary to form pyruvate or convert citrate to acetyl CoA. Near complete pathways for carbon fixation through the 3-hydroxypropionate (3-HP) bicycle were found in the metagenomes of all three sponges, however the reduction of malonyl-CoA to propionyl-CoA was absent from all symbionts, potentially representing a key step missing in this pathway. Nonetheless, the remainder of this pathway was expressed largely by Proteobacteria in both *I. microconulosa* and *P. foliascens* metatranscriptomes. Thermoproteota MAGs were only recovered from *I. ramosa* and key enzymes necessary for the HP-HB pathway were not detected, including those responsible for malonyl-CoA reduction, though previous studies have demonstrated the presence of the HP-HB cycle in Thaumacheota symbionts of sponges (Engelberts et al., 2020; Moitinho-Silva et al., 2017).

Genes for nitrogen fixation were absent from the microbiota of all three sponges, however both *Ircinia* species possessed microbes with the genomic potential for nitrate reduction, denitrification and nitrification. The first step in denitrification is the reduction of nitrate to nitrite by nitrate reductase, which was expressed in both Alphaproteobacteria and Gammaproteobacteria symbionts within the *I. microconulosa* microbiota (Figure 4.3; Appendix C; Figure S4.2). Alphaproteobacteria and Gammaproteobacteria MAGs additionally contained genes for the remaining steps of denitrification although gene expression was incomplete, while both Desulfobacterota and Gammaproteobacteria expressed genes to reduce nitrite to ammonia. MAGs from multiple Alphaproteobacteria families encoded the gene for

nitrous oxide reductase (*nosZ*), responsible for reducing nitrous oxide to nitrogen, and expression was observed in the *UBA828* family. For nitrification, ammonia oxidation was possible through Desulfobacterota, where MAGs contained all three genes of ammonia monooxygenase (*amoABC*) and both *amoA* and *amoB* were expressed (Figure S4.2). Genes for the remaining steps in nitrification (oxidation of hydroxylamine and nitrite) were expressed in Alphaproteobacteria and Gammaproteobacteria, while Nitrospirales also expressed genes for nitrite oxidation. Conversely, nitrogen metabolism was largely absent in the *P. foliascens* microbiota, where pathways for nitrate reduction, denitrification and nitrification were incomplete (Figure S4.2). Nonetheless, genomic potential was observed for denitrification where genes for nitrate reduction were retrieved in Rhodospirillales (Alphaproteobacteria) and expressed in Gammaproteobacteria (family *UBA10353*). Additionally, expression of nitrous oxide reductase was also observed in Alphaproteobacteria (family *Bin65*).

Genes for sulfur metabolism were present in all three sponge metagenomes, which included sulfate reduction, thiosulfate oxidation and taurine transport and oxidation. In both *P. foliascens* and *I. microconulosa*, gene expression was observed for taurine transport and its metabolism via the ABC transport system (*tauABC*) and taurine dioxygenase (*tauD*) in both Gammaproteobacteria and Alphaproteobacteria MAGs, while Nitrospinota also expressed these genes in *I. microconulosa* (Figure 4.3; Appendix C; Figure S4.3). Similarly, Gammaproteobacteria MAGs in both sponges showed expression for sulfate reduction and potential for thiosulfate oxidation, however one gene was missing from the SOX complex for thiosulfate oxidation in *P. foliascens* and was found but not expressed in *I. microconulosa*. In *P. foliascens*, *Rhodobacteraceae* expressed the full pathway for thiosulfate oxidation, however these genes were found but not expressed in *Rhodobacteraceae* MAGs from *I. microconulosa*. Taken together, the metatranscriptomic profiles of *I. microconulosa* and *P. foliascens* were highly consistent in their ability to metabolise sulfur.

4.4.4. Biosynthesis of B vitamins

The metagenomes of the three sponge species contained the necessary genes for all B vitamin synthesis including the bioactive forms for thiamine (B₁), riboflavin (B₂), pantothenate (B₅), pyridoxine (B₆), biotin (B₇) and cobalamin (B₁₂) as well as thiamine and biotin transporters. The *P. foliascens* metatranscriptomic profile illustrated that MAGs classified as Gammaproteobacteria (including *Endozoicomonadaceae*), Acidobacteriota and Desulfobacterota expressed the necessary genes to contribute to thiamine synthesis via metabolism of the aminoimidazole ribotide precursor, as well as the production of thiamine phosphate and its conversion to thiamine diphosphate (Figure 4.3, Appendix C; Figure S4.4). However, thiamine synthesis also requires 2-carboxy-4-methyl-thiazole-5-yl and only Desulfobacterota expressed the necessary genes to produce this compound. Interestingly, it was the Alphaproteobacteria, Chloroflexota and Actinobacteria that showed expression/partial expression of thiamine transport, suggesting the majority of those involved in synthesising thiamine may not necessarily transport it. Within the *I. microconulosa* metatranscriptome, expression of aminoimidazole ribotide metabolism and the metabolism of thiamine phosphate to thiamine diphosphate was also widespread encompassing six phyla (Figure 4.3, Figure S4.4). However, it was only Desulfobacterota and Nitrospirota that expressed genes for the production of 2-carboxy-4-methyl-thiazole-5-yl. Additionally, full expression of thiamine transport genes was observed in Alphaproteobacteria (*Rhodobacteriaceae*), Gammaproteobacteria (*HK1*) Chloroflexota and Actinobacteriota.

The partial biosynthetic pathways for riboflavin were expressed in almost all phyla in both *P. foliascens* and *I. microconulosa* metatranscriptomes, however the ability to incorporate ribulose 5-phosphate for full pathway expression was far more limited (Appendix C; Figure S5.5). In *P. foliascens*, only *Shewanellaceae* MAGs contained genes for the complete pathway

and expression of these genes (Figure 4.3). Similarly, expression for partial biosynthetic pathways for pantothenate was widespread encompassing the majority of microbial phyla in both sponges (Appendix C; Figure S4.6). However, it was only the *Endozoicomonadaceae* MAGs that showed expression of the full pathway, with the incorporation of alanine in the *P. foliascens* metatranscriptome, while no MAGs showed expression of the full pathway in *I. microconulosa* (Figure 4.3). Psuedomonadales (Gammaproteobacteria) within the *P. foliascens* microbiome expressed the necessary genes for pyridoxine synthesis using D-Erythrose-4-phosphate (Appendix C; Figure S4.7), and though the same pathway was present in the Psuedomonadales genomes recovered from *I. microconulosa*, expression of this pathway was not detected (Figure 4.3). *Endozoicomonadaceae* recovered from *P. foliascens* also possessed the pyridoxine synthesis pathway, though no expression was detected for the enzyme pyridoxine 5-phosphate synthase. Importantly however, pyridoxine can be synthesised using alternative compounds from glycolysis and the pentose phosphate pathway and this was expressed in Acidobacteriota, Chloroflexota and Actinobacteria genomes from both sponges.

Acidobacteriota, Desulfobacterota, Proteobacteria and Cyanobacteria were all expressing genes for biotin synthesis in the *P. foliascens* microbiome, as well as Latescibacterota and Poribacteria in *I. microconulosa* (Figure 4.3, Appendix C; Figure S4.8). Of these, only *Synechococcus* from *P. foliascens* additionally had the necessary genes for biotin transport, which were partially expressed. Similarly, only *Synechococcus* and *HK1* (Gammaproteobacteria) showed expression of biotin transporters within the *I. microconulosa* metatranscriptome, however biotin synthesis was only partially expressed in *HK1*. Finally, cobalamin metabolism from cobyrinate a,c-diamide was detected in the metatranscriptomes of both sponge species (Figure 4.3, Appendix C; Figure S4.9), performed by multiple Alphaproteobacteria and Gammaproteobacteria, as well Acidobacteriota in *I. microconulosa*. However, cobyrinate a,c-diamide production via both the precorrin (aerobic) pathway and co-

precorrin (anaerobic) pathways were only expressed in *Rhodobacteraceae* (Alphaproteobacteria), and expression of precorrin 3B synthase (aerobic pathway) was not detected. Despite containing the necessary genes, Alphaproteobacteria MAGs derived from *I. microconulosa* did not show complete expression of either pathways for cobyrinate a,c-diamide production.

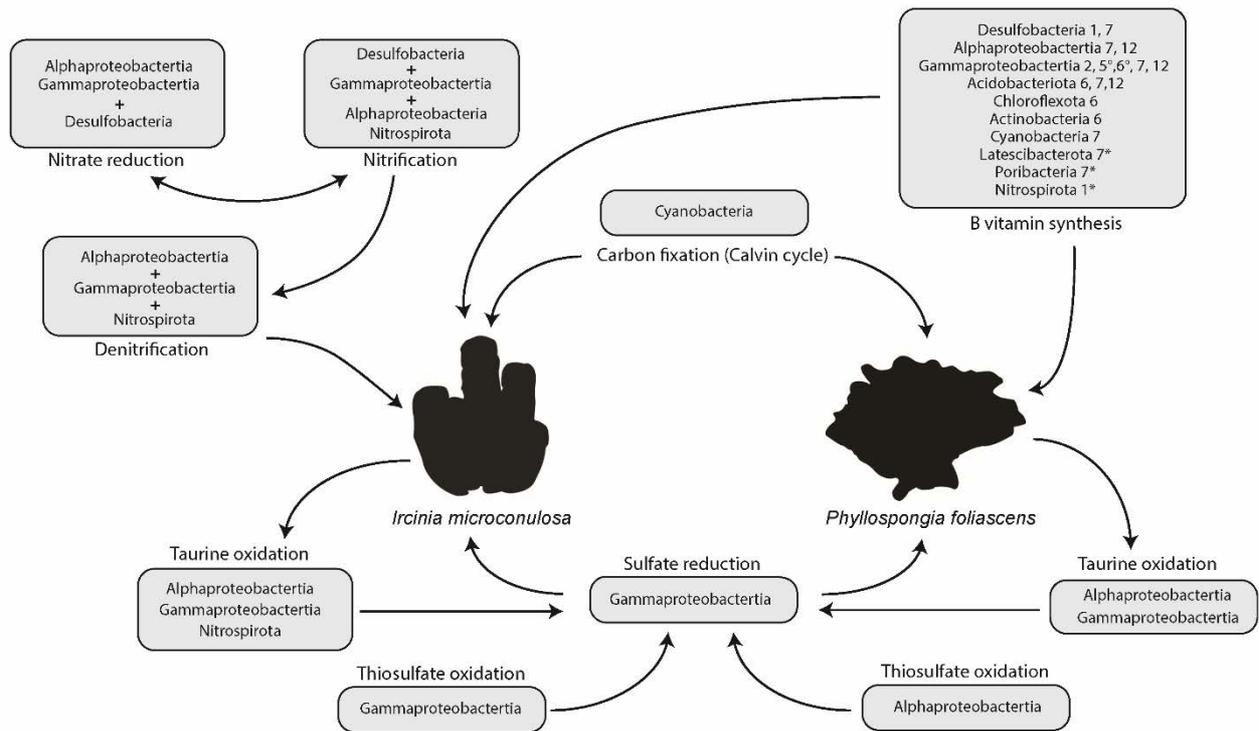


Figure 4.3. Overview of complete and near complete (>80%) gene expression for metabolic pathways found within the microbiomes of *Ircinia microconulosa* and *Phyllospongia foliascens*. The microbiota of both sponges showed expression for multiple transformations of sulfur along with carbon fixation and B-vitamin synthesis. However, only the *I. microconulosa* microbiota showed expression of nitrogen cycling. Arrows indicate proposed transfer of nutrients within the holobiont. A '+' symbol is used to indicate where microbes from multiple taxonomic classifications are needed to complete a pathway. B-vitamin synthesis pathways are numerically labelled adjacent to the corresponding microbe and those labelled with an asterisk indicate pathways only found in *I. microconulosa* while open circles indicate pathways only found in *P. foliascens*.

4.5. Discussion

Coral reefs generally exist in oligotrophic environments and therefore reef fauna must obtain and recycle nutrients efficiently to contribute to the high productivity of these ecosystems. Sponges play a critical role in supporting this productivity through the assimilation of DOM that couples benthic and pelagic trophic interactions (De Goeij et al., 2013; Rix et al., 2020). My genome-centred metatranscriptomic analysis revealed pathways for nutrient metabolism and potential provision to the host are expressed within the microbiome of coral reef sponges. Further, metagenomic data was a good predictor of microbial activity, highlighting the importance of obtaining metagenomes from reef species to understand reef dynamics.

4.5.1. Autotrophic carbon fixation in sponge symbionts

Expression of complete autotrophic carbon fixation was observed in *Synechococcus* MAGs with carbon fixation implicated as an important carbon source for the host and other members of the microbiome (Burgsdorf et al., 2021; Freeman et al., 2013). The Calvin cycle is responsible for the carbon fixation process during photosynthesis and my analysis confirmed that the phototrophic cyanobacterium, *Synechococcus*, was largely responsible for carbon fixation in these coral reef sponges. Interestingly, each sponge was dominated by a distinct *Synechococcus* population with genome reconstruction indicating that MAGS shared less than 95% ANI. Further, the abundance of *Synechococcus* was far higher in *I. ramosa* and *P. foliascens* microbiomes compared to *I. microconulosa*. The presence of *Synechococcus* in sponges has been documented extensively, including high abundance within the microbial community of *Stylissa carteri* collected from the Red Sea, which showed high expression of Cyanobacteria photosystems and carbon fixation (Moitinho-Silva et al., 2014). Similarly, genomic analysis of the abundant symbiont *Ca. Synechococcus spongarium* within the

microbiota of *P. foliascens* collected from the Red Sea suggests that its genome had undergone streamlining and contained the necessary genes for carbon fixation through the Calvin cycle (Gao et al., 2014). This *Synechococcus* genome was closely related to the genome of *Ca. Synechococcus spongarium*, indicating this *Synechococcus-Phyllospongia* symbiosis has been maintained across the vast geographic range the sponge is found.

Additional pathways for autotrophic carbon fixation were near complete in these sponge microbiomes, highlighting the potential for alternative sources of fixed carbon. Previously, Nitrospirota from *Ircinia ramosa* had been documented as possessing the necessary genes for the rTCA cycle (Engelberts et al., 2020), while expression of this pathway was observed in Nitrospirota from *Cymbastela concentrica* (Moitinho-Silva et al., 2017). Here, the microbiomes of both *Ircinia* species contained most of the genes required for the rTCA cycle and Nitrospirota symbionts of *I. microconulosa* had near complete expression of this pathway. These coral reef sponges also showed potential for carbon fixation through the 3-hydroxypropionate (3-HP) bicycle pathway, mediated by their Proteobacteria symbionts. While this pathway hasn't been identified in sponges previously, the similar 3-hydroxypropionate/4-hydroxybutyrate (HP-HB) cycle has been shown in Thermoproteota symbionts from *I. ramosa* and *C. concentrica* (Engelberts et al., 2020; Moitinho-Silva et al., 2017). Thermoproteota MAGs retrieved here did not contain key enzymes for the HP-HB pathway, such as those responsible for malonyl-CoA reduction, which also represented genes lacking from the 3-HP bicycle. Given that these MAGs are not 100% complete, it is unknown whether these missing enzymes are due to an absent pathway or genes that were not assembled.

4.5.2. Nitrogen metabolism is more common in the *Ircinia* sp. microbiome than *Phyllospongia foliascens*

Within an oligotrophic environment, nitrogen is of particular interest as it is a limiting element for primary productivity in the ocean (Falkowski, 1997; Tyrrell, 1999). Given no sponge-associated microbes had the capability of nitrogen fixation, assimilation of inorganic nitrogen may provide an important source of nitrogen for the sponge holobiont. The microbiota of both *Ircinia* species revealed evidence for multiple transformations of nitrogen including denitrification, nitrate reduction and nitrification. Previously, *Nitrosopumiliaceae* has been suggested as a key sponge symbiont of *Ircinia ramosa* given it was the only microbe containing ammonia oxidising genes (Engelberts et al., 2020). My analysis confirms that *Nitrosopumiliaceae* MAGs from *I. ramosa* contain ammonia oxidising genes, however within the closely related *Ircinia microconulosa*, ammonia oxidation was expressed by Desulfobacterota and no *Nitrosopumiliaceae* MAGs were retrieved. This suggests ammonia oxidation may be achieved through alternative ammonia oxidising prokaryotes when *Nitrosopumiliaceae* are not present. While Desulfobacterota likely contribute to nitrification, these MAGs did not contain the genes necessary to oxidise hydroxylamine to nitrite. It is feasible the remaining steps in nitrification are carried out by the Gammaproteobacteria family *HK1*, which expressed hydroxylamine oxidation as well as nitrite oxidation to nitrate. Additional bacteria, such as *Rhodobacteraceae* and Nitrospirales, also expressed genes for nitrite oxidation suggesting a large availability of nitrate. The resulting nitrate may then be used for multiple processes such as denitrification or nitrate reduction to either cycle the nitrogen back to ammonia (nitrate reduction) or release it as nitrogen (denitrification).

Evidence of denitrification is rarely observed in sponge metagenomic data as nitrous oxide reductase genes (*nosZ*) necessary for the final step in denitrification step to convert nitrous oxide to nitrogen, are often missing (Engelberts et al., 2020). Of the few studies that have identified the *nosZ* gene in sponge symbionts, *Pseudovibrio* belonging to the *Alphaproteobacteria* class is one symbiont with the complete set of genes for denitrification

(Bondarev et al., 2013). Here I observed the expression of nitrous oxide reduction in the Alphaproteobacteria family *UBA828* associated with *I. microconulosa* and the Alphaproteobacteria family *Bin65* in *P. foliascens*, while additional Alphaproteobacteria MAGs also contained this gene. Alphaproteobacteria also showed expression of the first steps of denitrification, both nitrate and nitrite reduction, however only Pseudomonadales and Entotheonellales MAGs contained the necessary genes to convert nitric oxide to nitrous oxide. Reducing nitrate to nitrite is also necessary for dissimilatory nitrate reduction resulting in ammonia. My analysis showed that both Gammaproteobacteria and Desulfobacterota expressed the pathway for nitrite conversion to ammonia, highlighting further capabilities of nitrogen cycling in *I. microconulosa*. Within the *P. foliascens* microbiota, genes for nitrogen metabolism were far less common and only the expression of nitrate and nitrous oxide reduction was observed. Hence in contrast to the *Ircinia* species, it is unlikely the *P. foliascens* holobiont obtains its nitrogen through microbial metabolism and may instead rely on exogenous sources of nitrogen. Alternatively, recent work on viral ecogenomics showed that HMA sponges, including *P. foliascens*, contained viromes that were enriched for genes related to nitrogen metabolism (Pascelli et al., 2020). However, further research is needed to understand how this might affect the nitrogen requirements of the host.

4.5.3. Evidence of widespread sulphur metabolism in the sponge microbiome

Sulphur is an essential nutrient required for microbial synthesis of certain amino acids (cysteine, cystine and methionine), vitamins (thiamine and biotin) and enzymes (Soda, 1987). One mechanism that sponge-associated microbes obtain sulfur is through sponge derived taurine, where sponge symbionts are often enriched in taurine dioxygenases and many also possess taurine transporters (Robbins et al., 2021). Here, all three sponge microbiomes contained the necessary genes to transport taurine across the cell membrane as well as taurine

dioxygenase to reduce taurine to sulfite. The expression profiles of both *P. foliascens* and *I. microconulosa* showed these genes were mostly active in Alphaproteobacteria MAGs, as well as Nitrospinota and the Gammaproteobacteria family *HK1* in *I. microconulosa*. Hence, the oxidation of host derived taurine by diverse bacteria potentially provides the sponge holobiont with a reliable source of sulfite which can be converted to either sulfide or sulfate within the dissimilatory sulfate reduction pathway. Genes for the dissimilatory sulfate reduction pathway were also present in all three sponge microbiomes, which were fully expressed by the *HK1* family in both *P. foliascens* and *I. microconulosa*. Additionally, multiple Alphaproteobacteria and Gammaproteobacteria MAGs showed partial expression of this pathway, where sulfite can be converted to sulfate. Finally, thiosulfate oxidation to sulfate via the SOX complex has been identified in sponge metagenomes and is of interest as thiosulfate may be produced by incomplete oxidation of sulfides (Engelberts et al., 2020). The metatranscriptomes showed full expression of the SOX complex in *Rhodobacteraceae* from *P. foliascens* and near complete expression in *HK1* from *I. microconulosa*. These results show active sulfur cycling in the sponge holobiont and verify previous metagenomic conclusions of widespread potential for sulfur metabolism.

4.5.4. Potential for supplementation of B-vitamins by sponge symbionts

Since animals cannot synthesize their own B-vitamins they must acquire these essential cofactors from their diet or from microbial symbionts. Although sponges may obtain B-vitamins from filter feeding, previous metagenomic studies have identified widespread potential for microbial biosynthesis and provisioning suggesting supplementation of B-vitamins (Engelberts et al., 2020; Fiore et al., 2015; Thomas et al., 2010). My analysis shows all three sponge microbiomes contain the necessary genes for biosynthesis of six essential B-vitamins; thiamine (B₁), riboflavin (B₂), pantothenate (B₅), pyridoxine (B₆), biotin (B₇) and

cobalamin (B₁₂). However, while vitamin biosynthesis appears widespread, only a small number of MAGs expressed the complete pathway or key steps in the biosynthesis pathways. For example, riboflavin and pantothenate synthesis were only expressed by *Shewanellaceae* and *Endozoicomonadaceae* MAGs respectively, despite widespread expression for the partial biosynthetic pathways of these vitamins. Similarly, thiamine biosynthesis requires multiple reaction pathways, and while partial thiamine synthesis is widespread, only Desulfobacterota and Nitrospirota expressed key steps for its production. It is possible that biosynthesis may be shared among multiple microbial phyla along different pathways, however this still likely requires key lineages to perform specific steps. Of particular interest was the family *HK1*, which expressed both thiamine and biotin transporter genes, as well as *Synechococcus* and *Rhodobacteraceae* with expressed biotin transporter genes, suggesting these bacteria might provision thiamine and biotin to the host respectively. There are other avenues for vitamin transport that may not be identified here, as recent reports have shown cobalamin can be transported into the mitochondria via alternative ABC transporters (McDonald et al., 2017). However future studies would benefit from tracking biosynthesis of B-vitamins by sponge symbionts to provide further evidence on whether they are translocated to the host. For example, *in vitro* experiments can be used to illustrate that vitamin production from one bacterium can stimulate the growth of another auxotrophic bacteria (Soto-Martin et al., 2020).

4.5.5. Conclusions

Genome-centred metatranscriptomic results confirm microbial functions and pathways are expressed within the sponge holobiont, and although the metatranscriptomic mapping rate was low, this was consistent with previous studies using environmental metagenomes (Hao et al., 2020). Importantly, incomplete expression of pathways or multi-enzyme complexes was observed despite the full suite of genes being assembled, likely a result of low microbial mRNA

recovery. Furthermore, methods using poly(A) subtraction of eukaryotic mRNA reads may inadvertently remove prokaryote mRNA reads as they are tagged with a poly(A) sequence before decay (Dendooven et al., 2020). Hence further work to optimise the subtraction of rRNA and eukaryotic mRNA within metatranscriptomic laboratory protocols would be highly beneficial to understanding the meta-metabolism within a host microbiome such as a sponge. For example, using cell separation protocols to enrich prokaryote cells prior to RNA extraction thereby removing the need for poly(A) subtraction may produce more accurate results. In other cases, missing genes within metagenomes resulted in incomplete pathways. It is possible these genes weren't assembled, since the minimum requirement for MAGs were greater than 50% complete with less than 10% contamination. However, it could also represent metabolic collaboration with the host and therefore combining host genomes and transcriptomes with microbial data may provide a unique perspective on how microbes interact with their host.

The microbiome is integral to the provision and efficient recycling of nutrients to the host, facilitating the success of sponges within reef ecosystems. Metagenome hypothesised functions were given for three sponge species (*I. ramosa*, *I. microconulosa* and *P. foliascens*) and metatranscriptomic validation was subsequently provided for *I. microconulosa* and *P. foliascens*, including pathways involved in carbon fixation, nitrogen and sulfur metabolism and biosynthesis of B-vitamins. I show that *Synechococcus* appears to be an important symbiont in all three sponges, where full expression of carbon fixation through the Calvin cycle was detected as well as biosynthesis and transport of B-vitamins. Similarly, the Gammaproteobacteria family *HK1* was also responsible for multiple functions, where it was involved in nitrification, sulfate reduction, thiosulfate oxidation and expressed both thiamine biosynthesis and transport. I also show that within the *I. microconulosa* microbiome, Desulfobacterota expressed genes for ammonia oxidation, which was previously confined to the Thermoproteota in sponges. Similarly, the expression of nitrous oxide reduction was

observed in Alphaproteobacteria and these genes have been rarely detected in sponge metagenomes. This work demonstrates the activity of microbial communities within coral reef sponges and highlights microbes with key functions of limited redundancy, while further clarifying the role of the microbiota in providing nutrients in a nutrient poor ocean.

**Chapter 5: Comparative genomics identifies key adaptive traits of
sponge associated microbial symbionts**

5.1. Summary

An overview of microbial metabolism within the sponge microbiome was provided in chapter 4, validating the broader functions of the microbial community and illustrating how this community might interact with its host. To understand how microbes may adapt to a host associated lifestyle, this current chapter focusses on five families of prokaryote symbionts; *Endozoicomonadaceae*, *Nitrosopumiliaceae*, *Spirochaetaeaceae*, *Microtrichaeaceae* and *Thermoanaerobaculaceae*, each identified in chapter 3 as having tight associations with their hosts through strong cophylogenetic patterns. Sponge specific clades were found to be enriched in many of the known mechanisms for symbiont survival, such as avoiding phagocytosis and defence against foreign genetic elements. Importantly, additional sponge symbiont characteristics, such as an enrichment in superoxide dismutase that prevent damage from free oxygen radicals was observed. In addition, a number of unique traits in sponge associated symbionts were identified, such as urea metabolism in *Spirochaetaeaceae* which was previously shown to be rare or absent in the Spirochaete phylum. These results highlight the mechanisms by which symbionts have adapted to living in association with sponges and show that these microbes have their own unique set of symbiont characteristics

5.2. Introduction

The genomes of symbiotic microorganisms differ significantly from those of closely related free-living relatives (Moran & Baumann, 2000). Identifying these characteristics can help distinguish if a host associated microbe has adapted to a symbiotic lifestyle or is a transient member of the community. Marine sponges often host complex microbial communities with the capacity to perform a range of metabolic functions that may reflect their adaptation to the host environment (Robbins et al., 2021; Webster & Thomas, 2016). Thus, sponge symbionts

represent interesting candidates to investigate the genomic signatures that facilitate adaptation to a host-associated lifestyle and identify how they underpin host health.

Obligate symbiotic bacteria, such as those associated with insects, have developed unique characteristics that define these symbionts and allow them to persist and colonise a host. For example, genome reduction combined with low GC content is a common trait once a symbiont evolves towards an intracellular life (McCutcheon & Moran, 2007). Genes that are no longer necessary are lost along with non-coding sequences, resulting in smaller genomes with high coding density that are skewed towards AT nucleotides, which may be due to a GC to AT bias in mutation rates (Agashe & Shankar, 2014). Similarly, successful symbionts often secrete eukaryote like proteins (ELPs) that facilitate protein-protein interactions that interfere with cellular processes such as phagocytosis (Jernigan & Bordenstein, 2014; Reynolds & Thomas, 2016). Furthermore, the host immune response can involve the generation of reactive oxygen species released from phagocytes, which obligate symbionts may break down with enzymes such as superoxide dismutase (SOD) (Broxton & Culotta, 2016; Peskin et al., 1998). Symbionts also exist in an environment rich in foreign genetic elements and bacteriophages that can cause harmful infections or cell mortality. Therefore, prokaryotic defences such as clustered regularly interspaced short palindromic repeats (CRISPR) with associated *cas* proteins, as well as restriction modification (RM) systems can represent additional genomic characteristics of symbionts (Horvath & Barrangou, 2010; Oliveira et al., 2014).

Along with microbial defences, symbionts have also devised strategies to attach and interact with host cells and tissues. Fibronectin binding proteins anchored to the cell wall of some symbionts can be used to attach to the fibronectin present within the extracellular matrix of a host (Hymes & Klaenhammer, 2016). Similarly, secretion systems are common characteristics of both beneficial and parasitic symbionts, allowing them to directly interact

with their host through injection of proteins across cell membranes (Costa et al., 2015). Finally, symbionts of a particular host can typically make use of the host resources. For example, carbohydrates derived from dissolved organic matter (DOM) may be present in the extracellular matrix of the sponge and symbionts have demonstrated large potential for carbohydrate degradation (Kamke et al., 2013; Robbins et al., 2021). Taken together, the above traits help a symbiont to survive within a host environment such as a sponge.

Importantly, many genomic features considered advantageous in symbiotic bacteria may also be seen in non-symbiotic microbes but are enriched in symbionts. For example, symbionts of sponges are exposed to a large number of mobile genetic elements due to the high seawater filtration rate of their sponge host (Jahn et al., 2019), and consequently an enrichment in genes that encode RM systems has been observed (Robbins et al., 2021). Similarly, amoebocyte cells selectively feed on microbes attempting to infect a sponge (Maldonado et al., 2010), and research has shown an enrichment in genes encoding ELPs in symbiont genomes that can interfere with amoebocyte phagocytosis (Robbins et al., 2021; Thomas et al., 2010). Thus, it is likely that additional symbiotic traits can be uncovered using comparative genomics when comparing the genomes of sponge symbionts to closely related genomes from non-sponge environments.

A recent comprehensive analysis of the sponge microbiome highlighted a range of symbiont characteristics when comparing the genomes of sponge-associated microbes to seawater microbes (Robbins et al., 2021). However, such a meta-analysis may miss finer details that can be revealed when comparing specific groups of interest. For example, the cyanobacterial symbiont *Synechococcus spongiarum* is confined to sponges and comparing the genome to free-living relatives revealed that this strain has one of the highest GC contents, a reduced genome size and an enrichment in ELPs (Gao et al., 2014). Similarly, *Rickettsialles* is

commonly found in marine invertebrates and an enrichment analysis showed these bacterial symbionts have many of the hallmarks associated with pathogenicity (Klinges et al., 2019). Thus, not only can these analyses indicate how a symbiont might adapt to life within a host, they may also provide an understanding towards the nature of the symbiosis.

Here, I provide a comparative enrichment analysis using metagenomic datasets from coral reef sponges. Specific symbionts were targeted which have previously shown evidence for cophylogeny in coral reef invertebrates (O'Brien et al., 2021). Using additional publicly available genomes of each symbiont from non-sponge environments, I reconstructed the phylogeny to identify how the evolutionary history of sponge-associated genomes compared to those of non-sponge genomes. Following this, I looked for gene enrichment in sponge symbiont clades against non-sponge clades to identify genomic signatures of adaptation to the sponge environment.

5.3. Materials & Methods

5.3.1. Laboratory methods and bioinformatics pre-processing

Methods for sample collection, DNA isolation protocols along with bioinformatic pipelines for metagenomic assembly, binning and classification of genomes have all been described in chapter 4.

5.3.2. Genome curation, phylogeny and taxonomic classification

Five prokaryote families were selected for analysis based on evidence of cophylogenetic signatures as detailed in chapter 3; *Endozoicomonadaceae*, *Spirochaetaceae*, *Nitrosopumiliaceae*, *Microtrichaceae* and *Thermoanaerobaculaceae*. Genomes for each symbiont group were obtained by first mining the set of sponge derived metagenome assembled genomes (MAGs) from chapter 4 (95ANI dereplicated) for those classified within

the selected list of sponge symbionts. Given the differences in taxonomic classifications between the Genome Taxonomy Database (GTDB; (Chaumeil et al., 2020) and the Silva 16S rRNA gene database (used in Chapter 3), I extracted 16S rRNA gene sequences from the MAGs and classified them using GraftM (v 0.13.1; (Boyd et al., 2018) and the Silva v132 16S rRNA gene database to ensure the correct symbionts identified in the 16S rRNA data were analysed. Additionally, I included all sponge derived MAGs identified as the same GTDB taxonomic classification in a comprehensive list of sponge symbionts compiled by (Robbins et al., 2021) to give a final set of sponge associated prokaryote genomes.

To compare the sponge symbiont genomes to closely related non-sponge genomes, I first identified suitable candidates by constructing a phylogenomic tree and including all GTDB entries of the same genome classification. The phylogenomic tree was computed using GTDB-tk with the `de_novo_wf` command, which uses Fast Tree (M. N. Price et al., 2010) to estimate microbe phylogeny from 122 and 120 bacterial and archaeal marker genes respectively. This method allowed for the visualisation of closely related genomes and sister clades of the selected symbiont groups. Closely related genomes were retrieved and added to the set of symbiont genomes, which was further dereplicated at 95ANI to remove duplicates, giving a final set of genomes for each taxonomic group. The isolation source for each GTDB genome was obtained from the NCBI database and any GTDB genome that was identified as sponge derived was added to the sponge group while all other genomes were grouped as non-sponge. Lastly, a phylogenomic tree was constructed (as detailed above) a second time using all dereplicated genomes to observe the phylogenetic relationships between sponge and non-sponge derived microbes. Outgroups were chosen by including the next closest lineage to the genomes within the enrichment analysis within the GTDB database.

5.3.3. Enrichment analysis between sponge associated and non-sponge associated genomes

All genomes were first quality checked using CheckM (v 1.1.3; (Parks et al., 2015) and any genome that was <85% complete or had >10% contamination was removed from the dataset to ensure a more robust comparison (with the exception of one *Endozoicomonadaceae* genome that was included at 83% complete due to the low number of MAGs). Genome characteristics commonly associated with symbiotic microbes, such as GC content and genome size, were compared between sponge and non-sponge groups using the results from CheckM. All genomes were then annotated using EnrichM (v 0.5.0; Boyd et al., 2019) with the KEGG orthologue database (KO), the protein family's database (Pfam) and the carbohydrate active enzymes database (CAZy). Finally, EnrichM's 'enrichment' function was used to compare the KO, Pfam and CAZy annotations between sponge associated and non-sponge associated genomes. This allowed identification of the genes present in each genome, along with the number of copies of each gene, and calculation of whether a particular gene was enriched in either the sponge associated or non-sponge associated groups. Statistical validation was performed using a Welch's t-test following p-value correction for multiple comparisons using two metrics; a) enrichment by comparison of the number of genomes containing the gene in each group, and b) enrichment by the number of copies of each gene per genome in each group. Enrichment figures showing genome trees and heatmaps were plotted in Rstudio (v 3.5.0; (Team, 2018) using the package 'ggtree' (Yu et al., 2017).

5.4. Results

5.4.1. Phylogeny and genome characteristics of sponge associated symbionts

A total of thirteen Endozoicomonadaceae (five sponge), fifty-seven Microtrichaceae (forty-two sponge), sixty-two Nitrosopumilaceae (sixteen sponge), forty-eight Spirochaetaceae (nine sponge) and thirty-two Thermoanaerobaculaceae (fourteen sponge) genomes were included in the analysis based on their 16S rRNA gene taxonomic classifications (Table S5.1-

5.5). A phylogenomic tree showed that sponge associated Spirochaetaceae and Thermoanaerobaculaceae formed single monophyletic clades, whereas sponge associated Microtrichaceae formed multiple clades that were restricted to genomes assembled from sponges (Figure 5.1). Sponge derived genomes of Nitrosopumiliaceae were associated with multiple lineages of free-living microbes, however one sponge cluster was found consisting of seven genomes (Figure 5.1). This may reflect the generalist nature previously observed in some Nitrosopumiliaceae sponge symbionts (Zhang et al., 2019). Similarly, sponge derived genomes of Endozoicomonadaceae were closely related to symbionts from other marine invertebrates and did not cluster in an exclusive sponge clade (Figure 5.1).

In general, genome size was greater and GC content higher in sponge associated Spirochaetaceae and Microtrichaceae compared to non-sponge relatives (Figure 5.2; Table 5.1), suggesting that these sponge symbionts have not undergone genome streamlining. Although on average there was no statistical difference between genome sizes of sponge and non-sponge derived Nitrosopumiliaceae, two sponge associated strains isolated from *Coscinoderma sp.* had particularly small genomes of 0.84 Mbp, highlighting the potential for genome reduction (Figure 5.2; Table 5.1). Similarly, on average sponge associated Nitrosopumiliaceae genomes were higher in GC content, however the two genomes isolated from the *Coscinoderma sp.* were the lowest of all Nitrosopumiliaceae genomes, with a GC content of 28.7% and 29.5%, indicating these two outlier genomes may be obligate symbionts.

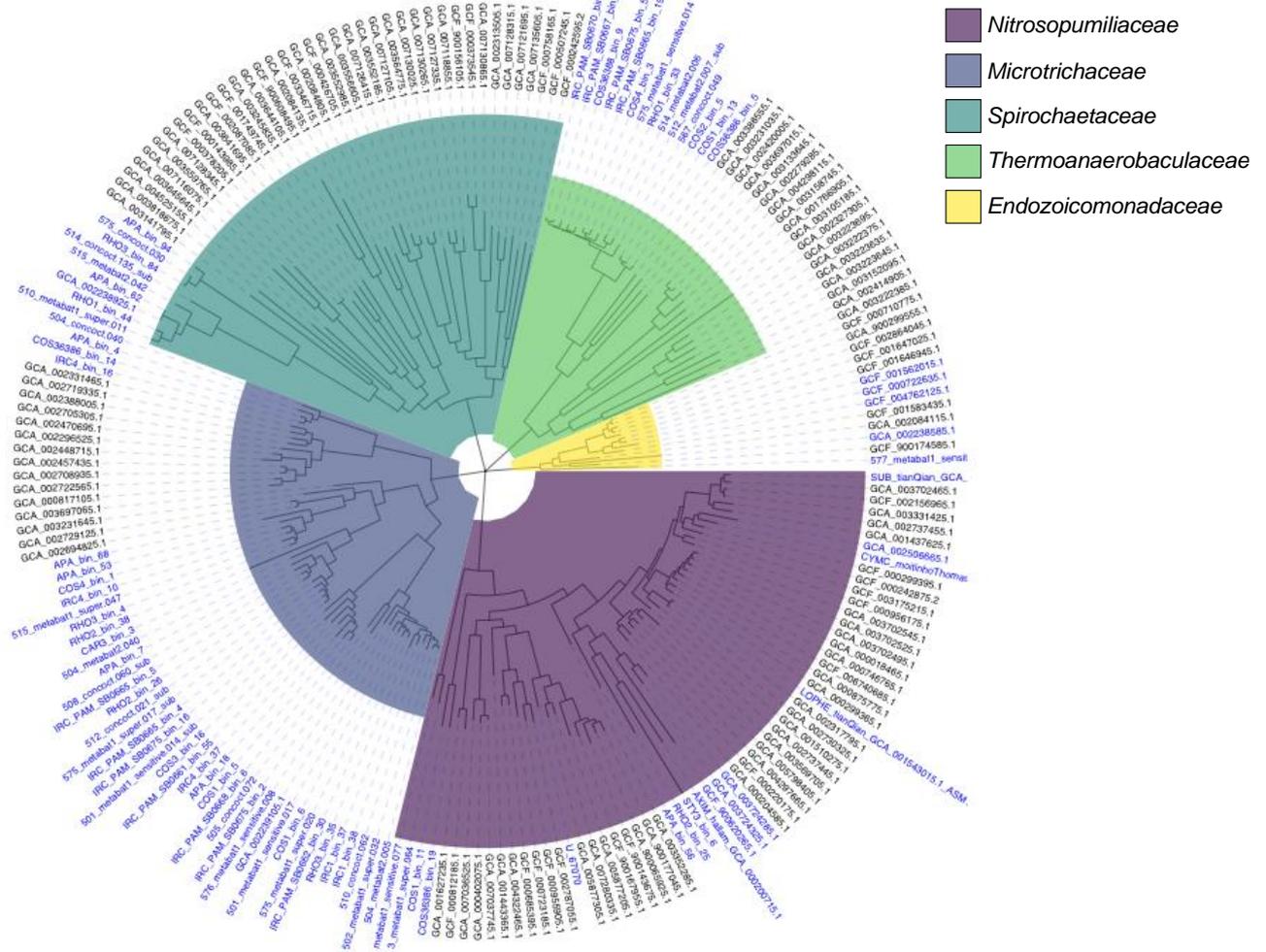


Figure 5.1. Phylogenomic tree of all sponge and non-sponge associated microbial genomes. Sponge-associated clades can be seen for *Microtrichaceae*, *Spirochaetaceae* and *Thermoanaerobaculaceae* while *Endozoicomonadaceae* and *Nitrosopumiliaceae* show multiple sponge-associated lineages. Branch tips in blue indicate sponge associated genomes while branch tips in black indicate non-sponge associated genomes. Clades are coloured by microbial family.

Genomes of *Thermoanaerobaculaceae* showed no difference in average size or GC content, however, non-sponge genomes were far more variable compared to sponge associated genomes (Figure 5.2; Table 5.1). Finally, genome size and GC content within the *Endozoicomonadaceae* were similar between sponge and non-sponge genomes (Figure 5.2;

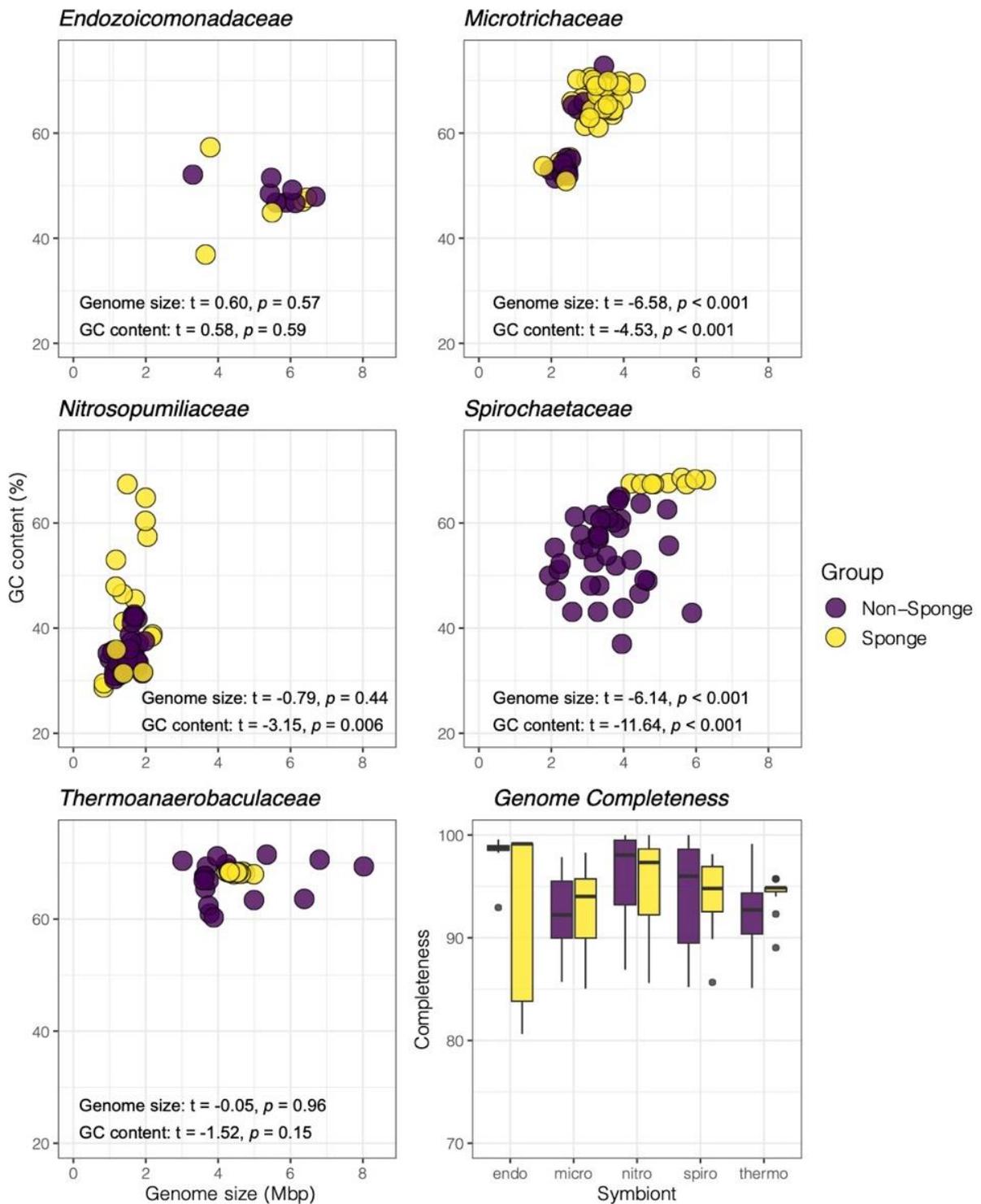


Figure 5.2. Genome size and GC content for all sponge associated and non-sponge associated genomes for each microbial group. In general, sponge associated symbionts do not have smaller genomes than free-living relatives, indicating a facultative symbiont lifestyle. Last panel shows genome completeness for sponge associated and non-sponge associated genomes for each microbial group.

Table 5.1), likely because non-sponge genomes were still derived from symbionts of marine invertebrates, including corals, bivalves an ascidian and echinoderm (Table S5.1).

Table 5.1. Statistical summary for genome size and GC content for sponge associated and non-sponge associated genomes in all five microbial groups. Std er. = standard error of the mean, min = minimum value, max = maximum value.

Group	Family	Mean size	Std er. size	Min size	Max size	Mean GC	Std er. GC	Min GC	Max GC
Sponge	Endo	5.14	0.61	3.65	6.45	46.76	3.26	36.9	57.3
Non-Sponge	Endo	5.57	0.36	3.3	6.69	48.69	0.75	46.7	52.1
Sponge	Micro	3.28	0.08	1.79	4.33	65.45	0.73	50.9	70.6
Non-Sponge	Micro	2.49	0.09	1.96	3.45	56.97	1.73	51.4	72.8
Sponge	Nitro	1.55	0.11	0.84	2.19	44.9	3.17	28.7	67.4
Non-Sponge	Nitro	1.46	0.04	0.95	1.97	34.81	0.47	30.3	42.6
Sponge	Spiro	5.23	0.24	4.19	6.27	67.76	0.16	67.4	68.6
Non-Sponge	Spiro	3.54	0.14	1.95	5.88	54.35	1.14	37	65
Sponge	Thermo	4.48	0.06	4.22	4.99	68.31	0.05	68	68.5
Non-Sponge	Thermo	4.46	0.32	3.02	8.02	67.03	0.84	60.3	71.5

5.4.2. Overview of gene enrichment patterns in sponge associated and non-sponge associated microbial genomes

The Spirochaetaceae, Microtrichaceae and Thermoanaerobaculaceae all had a far greater number of enriched genes compared to the Endozoicomnadaceae and Nitrosopumiliaceae when analysing the Pfam and KO annotations (Table 5.2). Spirochaetaceae and Thermoanaerobaculaceae additionally showed a small number of enriched CAZy genes, however no CAZy genes were enriched in the remaining microbial groups in sponge associated genomes (Table 5.2). This greater potential to identify genomic adaptations to the host in Spirochaetaceae, Microtrichaceae and Thermoanaerobaculaceae may, in part, be due to the sponge specific clades observed in the phylogenomic trees for these three families. Given these results, and since Endozoicomnadaceae and Nitrosopumiliaceae genomes have been characterised extensively in the past (Haber et al., 2021; Neave, Michell, et al., 2017; Tandon et al., 2020; Zhang et al., 2019), I focus my efforts on the analysis of Spirochaetaceae, Microtrichaceae and Thermoanaerobaculaceae microbial groups.

Table 5.2. Overview of the total number of enriched genes for sponge associated and non-sponge associated genomes for all five microbial groups. Enriched genes are included if the number of genomes that contain the gene within each group are statistically different. CAZY = Carbohydrate active enzymes, KO = genes annotated using the Kegg Orthology database, PFAM = genes annotated using the Protein Families database.

		CAZY	KO	PFAM
Spiro	Sponge	8	331	317
	Non-sponge	8	425	511
Micro	Sponge	0	312	367
	Non-sponge	2	134	140
Thermo	Sponge	4	538	545
	Non-sponge	7	435	510
Endo	Sponge	0	0	0
	Non-sponge	0	0	0
Nitro	Sponge	0	7	16
	Non-sponge	0	15	27

5.4.3. Avoiding the host immune response and infection from foreign DNA

Symbionts were enriched in a range of genes potentially facilitating the avoidance of the host immune response. Specifically, this included multiple eukaryote-like proteins (ELPs) in the form of ankyrin repeat proteins (ARPs) and a WD40-like beta propeller repeat, along with superoxide dismutase (SOD) which can provide additional protection from phagocytes. Five genes classified as ARPs were found across all Spirochaetaceae genomes, with four of them found in all nine sponge associated genomes, and two of these genes had significantly higher number of copies per genome than the non-sponge Spirochaetaceae (Figure 5.3). The WD40 repeat not only had a higher number of copies per genome, but all sponge associated Spirochaetaceae contained the gene as opposed to only 17.8% of non-sponge Spirochaetaceae. Sponge associated Thermoanaerobaculaceae all contained the five ARPs, however only one

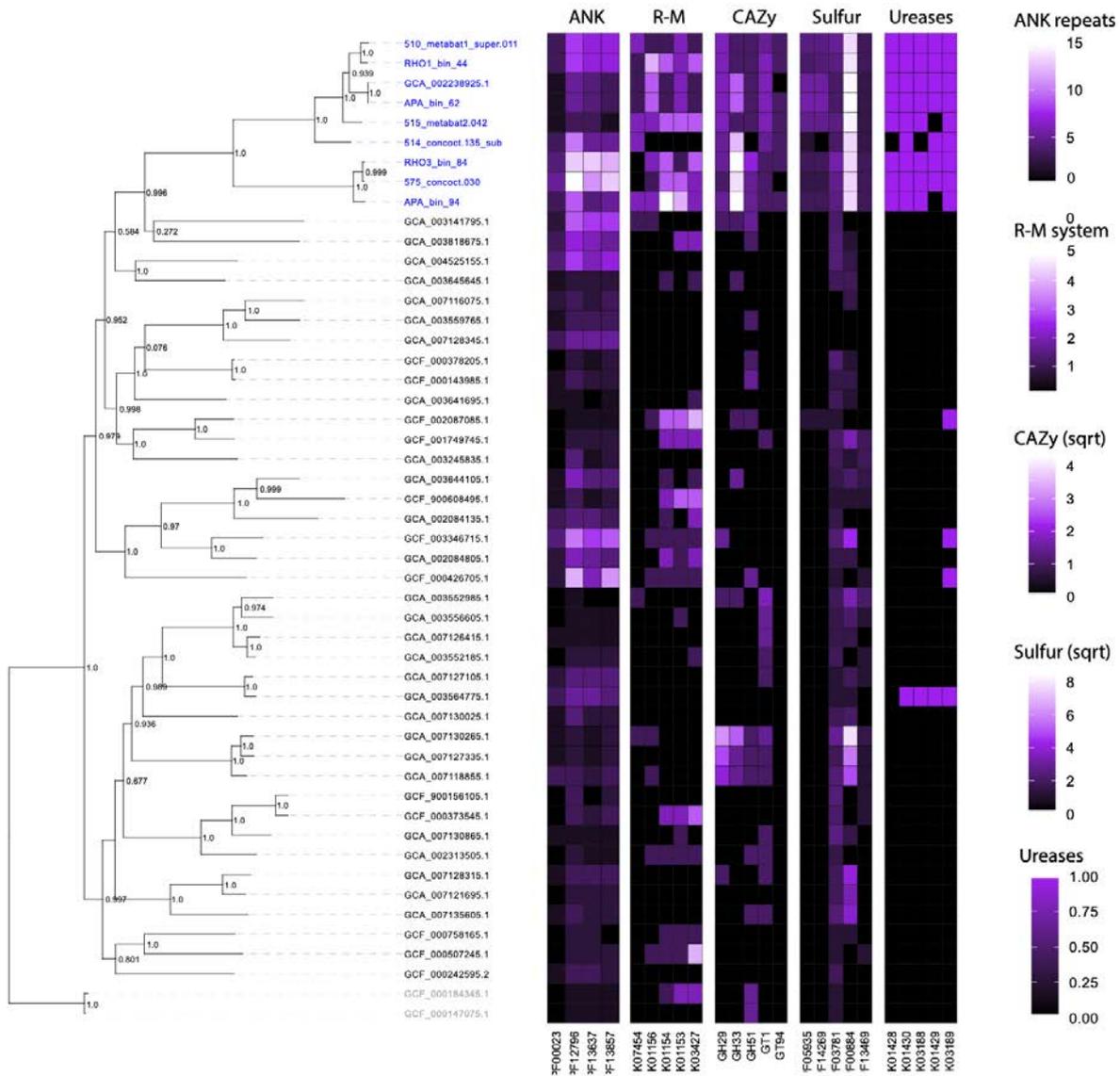


Figure 5.3. Phylogenomic relationships and gene enrichment patterns in sponge associated and non-sponge associated Spirochaetaceae genomes. Branches with blue labels indicate sponge associated genomes, branches with black labels indicate non-sponge associated genomes while grey labels indicate outgroups and weren't included in the enrichment analysis. Numbers at nodes indicate branch support values using the Shimodaira-Hasegawa test. Heatmap indicates copy numbers for a selection of genes in key symbiotic signatures. Data that has been square root transformed for illustration is indicated with (sqrt). Abbreviations: ANK = Ankyrin repeat proteins, R-M = restriction modification system enzymes, CAZy = carbohydrate active enzymes, Sulfur = sulfatases and sulfatransferases.

was enriched compared to the non-sponge genomes (Figure 5.4). Likewise, all Thermoanaerobaculaceae genomes from sponges contained the WD40-like beta propeller repeat, however this gene was not enriched compared to non-sponge genomes. While the above ELPs were also present in Microtrichaceae, there was no enrichment in those genomes that were derived from sponges with less than half of the genomes containing ELPs (Figure 5.5). Finally, all three microbial families were enriched in the Cu-Zn family of SOD, and while the Fe-Mn family was also abundant, this was only enriched in the sponge associated Microtrichaceae.

Genes related to defence against foreign DNA were enriched in Spirochaetaceae and Microtrichaceae from sponges, including multiple genes that were classified as restriction enzymes (Figures 5.3 & 5.5). While Thermoanaerobaculaceae also contained restriction enzymes, only two of these were enriched in sponge-associated genomes (Figure 5.4). Similarly, genes classified as cas enzymes associated with the CRISPR-cas system were abundant in the dataset. Of these, two cas enzymes were enriched in the sponge associated Spirochaetaceae, while five cas enzymes were enriched in sponge associated Microtrichaceae, which were absent from non-sponge genomes. Conversely, less than half of sponge associated Thermoanaerobaculaceae contained cas enzymes and none showed any enrichment.

5.4.4. Mechanisms of symbiont attachment and interaction with the host

Fibronectin binding proteins are potentially used to bind to host tissues and were common within Spirochaetaceae from both groups; however, it was the fibronectin type III domains that were significantly enriched in sponge-associated genomes. Similarly, fibronectin type III domains were found in all sponge associated Microtrichaceae and Thermoanaerobaculaceae and were enriched compared to non-sponge genomes. Cadherins are another adhesion molecule and may fuse to other domains such as fibronectin type III

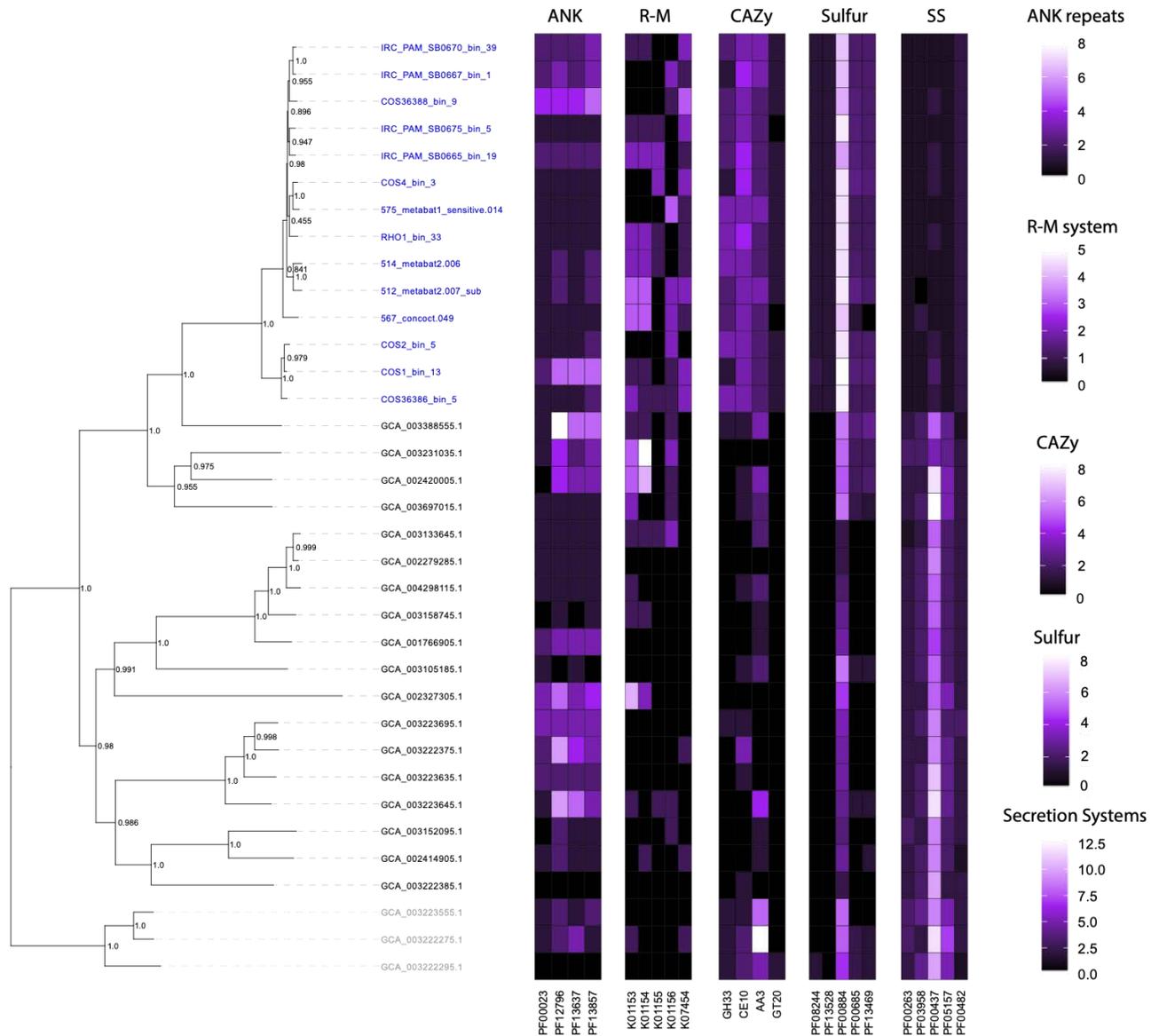


Figure 5.4. Phylogenomic relationships and gene enrichment patterns in sponge associated and non-sponge associated Thermoanaerobaculaceae genomes. Branches with blue labels indicate sponge associated genomes, branches with black labels indicate non-sponge associated genomes while grey labels indicate outgroups and weren't included in the enrichment analysis. Numbers at nodes indicate support values. Heatmap indicates copy numbers for a selection of genes in key symbiotic signatures. Abbreviations: ANK = Ankyrin repeat proteins, R-M = restriction modification system enzymes, CAZy = carbohydrate active enzymes, Sulfur = sulfatasases and sulfatransferases

(Anantharaman & Aravind, 2010). Both Microtrichaceae and Thermoanaerobaculaceae from sponges were enriched in a cadherin domain while Microtrichaceae were also enriched in a

adherin-like beta sandwich domain, suggesting high potential for this form of adhesion (Figure 5.5).

Secretion systems can be used to interact with either adjacent microbes or host cells and the data showed that all Thermoanaerobaculaceae genomes contained proteins associated with the type II secretion system, however the number of copies of type II SS genes were enriched in non-sponge associated genomes (Figure 5.4). While fewer type II secretion system proteins were found in Microtrichaceae genomes, these were enriched in sponge-associated genomes where all symbionts possessed the type II SS gene, which were almost absent in non-sponge genomes. Interestingly, proteins with secretion system annotations were mostly absent from Spirochaetaceae and hence no enrichment was observed for those associated with sponges.

5.4.5. Enrichment of genes related to the metabolism of carbohydrates

Gene enrichment for the breakdown of carbohydrates was most notable within the sponge associated Spirochaetaceae, which possessed an abundance of glycosyl hydrolases (GH), in particular those that act on sialic acids (GH33; sialidase) and fucose (GH29; fucosidase) (Figure 5.3). Carbohydrate esterases (CE) were also abundant in Spirochaetaceae genomes, however these were not enriched compared to non-sponge genomes. Thermoanaerobaculaceae from sponges were also enriched for GH33 and additionally showed an enrichment for the CE family 10, which was found in all sponge derived genomes and only half of non-sponge genomes (Figure 5.4). Microtrichaceae showed no enrichment for carbohydrate active enzymes and GHs were found in less than half of sponge derived genomes. However, CE family 10 and auxiliary activity (AA) family 3 were found in a large proportion

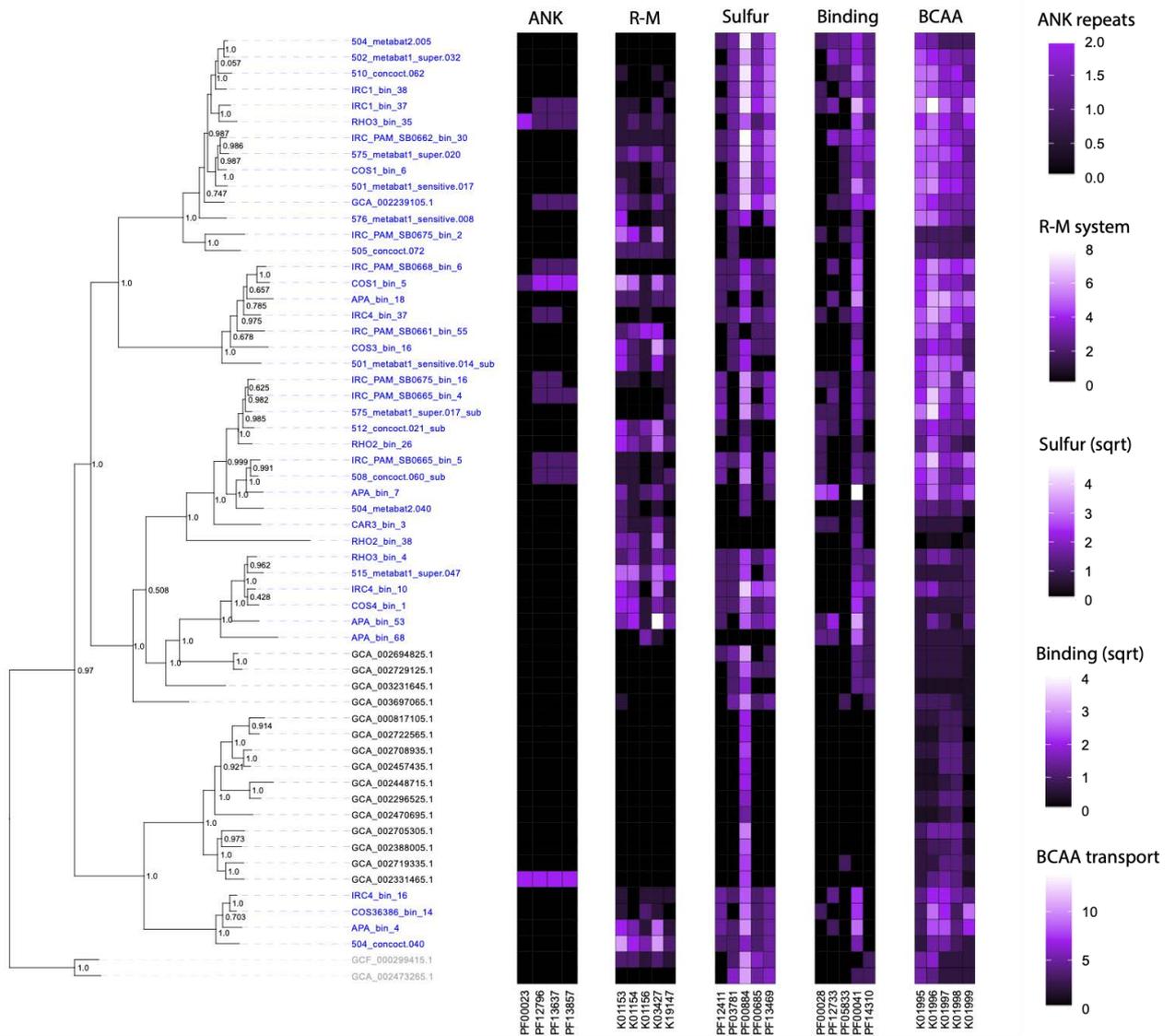


Figure 5.5. Phylogenomic relationships and gene enrichment patterns in sponge associated and non-sponge associated Microtrichaceae genomes. Branches with blue labels indicate sponge associated genomes, branches with black labels indicate non-sponge associated genomes while grey labels indicate outgroups and weren't included in the enrichment analysis. Numbers at nodes indicate support values. Heatmap indicates copy numbers for a selection of genes in key symbiotic signatures. Abbreviations: ANK = Ankyrin repeat proteins, R-M = restriction modification system enzymes, Sulfur = sulfatases and sulfatransferases, BCAA = branch chain amino acid.

of sponge-associated genomes (81% and 88% respectively), indicating capacity for carbohydrate metabolism.

I observed an enrichment in both sulfatases and sulfotransferases in sponge symbiont genomes, suggesting these may act with the CAZy genes to breakdown and remodel sulfated polysaccharides. Specifically, I found that Spirochaetaceae genomes from sponges are highly enriched in sulfatases with an average of sixty-four copies per genome and additionally show an enrichment in sulfotransferases and arylsulfotransferases (Figure 3). Similarly, sulfatases were found in all Thermoanaerobaculaceae genomes, however, copy numbers were far higher in sponge-associated genomes, with an average of fifty-seven copies compared to thirteen in non-sponge genomes (Figure 4). The sulfatase modifying enzyme was also present in all Thermoanaerobaculaceae genomes though again demonstrated gene copy number enrichment in sponge-associated genomes. Sulfatases were present in the majority of Microtrichaceae genomes (86% of sponge-associated), however copy numbers were far less than Spirochaetaceae and Thermoanaerobaculaceae genomes, with an average of seven copies per genome and were not enriched compared to non-sponge genomes (Figure 5). Finally, sulfotransferases were enriched in both Microtrichaceae and Thermoanaerobaculaceae sponge-associated genomes.

5.4.6. Additional enrichment patterns in genes of interest

Taurine metabolism has been of particular interest in sponge microbiology, with symbionts potentially using this as a source of sulfur. I found that taurine dioxygenase was present in all sponge-associated genomes of Spirochaetaceae and Thermoanaerobaculaceae and absent in all non-sponge genomes. However, neither of these two microbial families additionally possessed the taurine transport system. While most sponge associated Microtrichaceae contained taurine dioxygenase (69%), this was not enriched compared to non-sponge genomes and most genomes lacked the taurine transport system. Interestingly, reduction of taurine results in sulfite and the data shows that for all three families of microbes,

sponge-associated genomes were enriched in a nitrite/sulfite reductase ferredoxin-like half domain. Two copies of this repeat are found in nitrite and sulfite reductases and are key to the biosynthetic assimilation of sulfur and nitrogen.

Sponges and some associated microbes can produce urea as nitrogenous waste, which needs to be degraded or expelled. I found that Spirochaetaceae genomes from sponges were enriched for all three of urease subunits (ureABC) (Figure 5.3), along with two accessory proteins (ureFG) and all components of the urea transport system (urtABCDE). By contrast, very few non-sponge Spirochaetaceae contained any urease genes. Further, ureases along with accessory proteins and the transport system were rarely observed in Microtrichaceae and Thermoanaerobaculaceae genomes. Although genes involved in urea metabolism may also be present in other microbial phyla from sponges, these results suggest that urea metabolism in sponge associated Spirochaetaceae is a unique function within the Spirochaete phylum.

Previous studies have shown that Poribacteria from sponges contain large numbers of genes encoding phytanoyl-CoA dioxygenase (phyH) (Kamke et al., 2014). The results show that Spirochaetaceae genomes from sponges are also heavily enriched in phyH, containing on average 129 copies compared to less than one for non-sponge genomes. Sponge associated Thermoanaerobaculaceae were also enriched in phyH, with all genomes carrying the gene, however in this case far fewer copies (eight per genome) were present than in the Spirochaetaceae. Interestingly, this trend was not replicated for Microtrichaceae, where the gene was found in nearly all sponge and non-sponge genomes, with an average of eight and fourteen copies respectively. Finally, while most Microtrichaceae genomes encoded a branched-chain amino acid transport system, copy numbers for these genes were enriched in Microtrichaceae genomes isolated from sponges (Figure 5).

5.5. Discussion

Microbial symbionts from sponges show characteristic genomic signatures indicating adaptation to a host associated environment. Here, I analysed the phylogenomic and genomic content for five families of microbes that previously showed evidence of cophylogeny with coral reef invertebrates. Genomes classified as Endozoicomonadaceae and Nitrosopumiliaceae did not cluster into sponge associated clades and subsequently very little genetic enrichment was observed in sponge-associated microbial genomes. The remaining microbial families, Spirochaetaceae, Microtrichaceae and Thermoanaerobaculaceae all showed monophyletic sponge clades and an enrichment for genes across a range of symbiont characteristics. Hence, the genomic evidence for their adaptation to the sponge environment was described, which demonstrated their metabolic potential and highlighted functions which may be interpreted as beneficial to the host.

5.5.1. Sponge symbionts may escape phagocytosis using both eukaryote like proteins and superoxide dismutase

Eukaryote like proteins (ELPs) such as ankyrin repeat proteins (ARPs) have received considerable attention in sponge symbiosis as experimental evidence showed that *E. coli* containing ELPs were able to avoid phagocytosis by amoeba cells (Nguyen et al., 2014; Reynolds & Thomas, 2016). More recently, ARPs from bacteriophages were also shown to modulate eukaryote immune response leading to reduced phagocytosis of bacteria in a potential tripartite symbiosis (Jahn et al., 2019). Analysis showed that Spirochaetaceae and Thermoanaerobaculaceae were both enriched for ARPs, and while these were not enriched in Microtrichaceae, they were still present in most genomes. These results are consistent with previous analyses showing an enrichment of ARPs in sponge symbionts (Kamke et al., 2014; Robbins et al., 2021; Thomas et al., 2010), and symbionts of many other invertebrates (Jernigan & Bordenstein, 2014). For example, metagenomic data from the coral *Porites lutea* revealed an

enrichment of ARPs within the genomes of associated bacteria, and the genome of the *Drosophila melanogaster* symbiont, *Wolbachia pipientis*, has among the highest number of copies of ARPs of any prokaryote (D. Wu et al., 2006). In addition, ARPs secreted by bacterial pathogens have been shown to facilitate host infection (Habyarimana et al., 2008; C. T. D. Price et al., 2010). Taken together, an enrichment in ARPs appears to be a common signature of symbiotic microbes from sponges and other invertebrate hosts (Jernigan & Bordenstein, 2014).

Microbial genomes may encode additional ELPs that govern protein-protein interactions thereby regulating cellular processes such as phagocytosis (Reynolds & Thomas, 2016). Of these, the WD40 beta propeller repeat is among the most abundant protein domains in eukaryote genomes and mediates molecular recognition events (Xu & Min, 2011). The results showed that both Spirochaetaceae and Thermoanaerobaculaceae genomes contained a WD40-like beta propeller repeat, however this was only enriched in the Spirochaetaceae genomes. This has been observed in symbionts previously, where *Poribacteria* from both sponges and corals, as well as *Endozoicomonas* from coral, demonstrated a high number of WD40 domains within their genomes (Kamke et al., 2014; Robbins et al., 2021). These data are consistent with previous genomic analyses, suggesting that ELPs are a common mechanism for sponge symbionts to avoid being consumed by the host (Reynolds & Thomas, 2016).

One method to avoid phagocyte killing that has received little attention is through the use of superoxide dismutase (SOD). Phagocytes may generate large amounts of reactive oxygen species that control the growth of infecting microbes, and SOD can protect against oxygen radical damage and help cells survive phagocytosis (Battistoni, 2003; Broxton & Culotta, 2016). For example, the pathogen *Salmonella typhimurium* showed decreased survival against macrophages in response to a knockout mutation of the *sodC* gene (De Groote et al., 1997).

Similarly, the pathogen *Mycobacterium tuberculosis* showed an up-regulation of *sodC* in response to macrophages (D’Orazio et al., 2001). Data here shows that all three microbial families were enriched in the Cu-Zn family of SOD indicating a protection against superoxide radicals that are known to be present within sponges (Peskin et al., 1998). While an enrichment in the Mn-Fe family of SOD was also observed, these are generally intracellular, as opposed to the extracellular/periplasmic Cu-Zn family and are therefore less likely to be involved in protection against extracellular oxy-radical damage (Broxton & Culotta, 2016). Superoxide may also be produced outside of phagocytosis in an effort to control microbial infection and the Cu-Zn family can show flexibility in its requirements to assist bacterial colonisation (Battistoni, 2003).

5.5.2. Sponge symbionts enriched in enzymes associated with the restriction modification system and CRISPR-cas system

Microbial symbionts must deal with infection from mobile genetic elements (MGEs). Sponge symbionts are particularly at risk of MGEs as large volumes of seawater are filtered by the sponge host which exposes the symbionts to phage transposable elements and plasmids (Horn et al., 2016). MGEs infecting prokaryotes can be costly if they are incorporated into the chromosome and disrupt cellular function, or in the case of phage infection, result in cell death (Rankin et al., 2011). Two potential defence strategies against infection are restriction-modification (RM) systems and the CRISPR-cas system, both of which act to cleave foreign nucleic acids at specific recognition sites using nucleases (Horvath & Barrangou, 2010; Oliveira et al., 2014). Here, all three families showed an enrichment in endonucleases of the RM system (restriction enzymes) and both Spirochaetaceae and Microtrichaceae genomes were enriched in *cas* genes, which encode functional domains of the CRISPR-cas system such as nucleases, helicases and polymerases (Horvath & Barrangou, 2010). While these are a common

defence method for prokaryotes against MGE infection, an enrichment in the associated enzymes appears to be a signature trait among sponge symbionts (Horn et al., 2016; Robbins et al., 2021).

5.5.3. *Symbiont attachment through fibronectins and cadherins*

A host associated lifestyle frequently involves some form of adhesion to the host, through either direct microbe-host cell attachment or biofilm formation. Previous research has suggested that sponge symbionts might form this attachment through cell adhesion molecules encoded in fibronectin type III and cadherin domains (Kamke et al., 2014; Robbins et al., 2021). Fibronectin is a glycoprotein widely distributed among animals where it can play a major role in cell adhesion and bind to extracellular proteins such as collagen (Pankov & Yamada, 2002). All three microbial families were enriched in the fibronectin type III domain, suggesting that secretion of fibronectin may be a common form of adhesion for symbionts to attach to sponge collagen. Although not enriched, all except one Spirochaetaceae genome derived from sponges also encoded a fibronectin binding protein. An enrichment in the fibronectin type III domain may provide more binding sites for binding proteins in biofilm formation as well as attachment to the host. Similar methods have been described in pathogenic bacteria, including pathogenic Spirochaetes, which have developed fibronectin binding proteins anchored to the cell wall, allowing them to attach to host fibronectin (Cullen et al., 2004; Hymes & Klaenhammer, 2016; Schwarz-Linek et al., 2004).

Cadherins (calcium-dependent adhesion proteins) are transmembrane glycoproteins with adhesive properties that can be exploited by symbionts (Dash et al., 2021). Of particular interest is the discovery that bacterial cadherins are capable of both homophilic and heterophilic interactions (Fraiberg et al., 2010), suggesting the possibility that symbiont produced cadherins can be used for adhesion to either host derived or symbiont cadherins. An

enrichment in cadherin domains in both Microtrichaceae and Thermoanaerobaculaceae from sponges was observed, suggesting another mechanism for host attachment and biofilm formation. Moreover, the Microtrichaceae genomes were enriched in a cadherin-like beta sandwich domain. This domain is widespread in prokaryotes and often fused to other domains such as fibronectin type III (Anantharaman & Aravind, 2010), suggesting a combination of cadherins and fibronectins are used in host attachment.

5.5.4. Type II secretion systems may help degrade biopolymers in the extracellular matrix

Host-microbe interactions can be mediated through bacterial secretion systems (SS); nanomachines that secrete a range of substrates, including proteins and DNA, as part of an environmental response by bacteria (Costa et al., 2015). Of the seven types of SS, enrichment was seen only for type II SS proteins in Microtrichaceae and, although all Thermoanaerobaculaceae genomes also contained these genes, copy numbers were enriched in non-sponge genomes. The main function of type II SS is to secrete hydrolysing enzymes into the extracellular environment, which can be important for bacterial growth and survival in a host as they degrade biopolymers such as carbohydrates, lipids and proteins (Nivaskumar & Francetic, 2014). Therefore, it is possible both Microtrichaceae and Thermoanaerobaculaceae use the type II SS to scavenge such compounds in the sponge extracellular matrix. Interestingly, SS were mostly absent from Spirochaetaceae and are not common in sponge symbionts in general (Robbins et al., 2021). Thus, the use of type II SS in Microtrichaceae and Thermoanaerobaculaceae appears to be a rare trait for host interaction in these sponge associated symbionts.

5.5.5. Multienzyme degradation of carbohydrates including glycosyl hydrolases and sulfatases

Assimilation of dissolved organic matter (DOM) is an important ecological role of sponges on coral reefs. This function can be partially achieved through the microbiota (Campana et al., 2021; Rix et al., 2020), and previous studies have shown some sponge symbionts have an increased capacity to degrade carbohydrates (Kamke et al., 2013; Robbins et al., 2021). In other systems, Spirochaetes are well known for their ability to metabolise carbohydrates (Warnecke et al., 2007) and here an enrichment in glycosyl hydrolases (GH) in Spirochaetaceae genomes was found, which also contained abundant carbohydrate esterases (CE) although these weren't enriched. Similarly, Thermoanaerobaculaceae showed an enrichment in both GH and CE, however enzymes for carbohydrate metabolism were mostly undetected in Microtrichaceae. Of particular interest is the enrichment in sialidase (GH33), which appears to be common among sponge symbionts, potentially allowing them to use the sialic acid found in the sponge mesohyl (Robbins et al., 2021). Similarly, an enrichment in fucosidase (GH29) may be an adaptation to the fucose found in coral mucus, which makes up part of the DOM assimilated by coral reef sponges (Hadaidi et al., 2019; Rix et al., 2016). Thus, it is possible that Spirochaetaceae and Thermoanaerobaculaceae contribute to the ecological role of sponges given their enrichment in carbohydrate active enzymes.

The Spirochaetaceae and Thermoanaerobaculaceae genomes also showed an enrichment in genes that allow them to metabolise sulfated polysaccharides present in the sponges mesohyl. In sponges, sulfated polysaccharides play important roles in cell aggregation and maintain the structural integrity of the sponge (Vilanova et al., 2009; Zierer & Mourão, 2000), and these sulfated polysaccharides could be synthesised, degraded or remodelled by sulfatases in combination with sulfotransferases. For example, the synthesis of sulfated polysaccharides requires a sulfotransferase to graft a sulfate ester group onto an existing carbohydrate, while the degradation involves a sulfatase to cleave the sulfate ester group as well as a GH to cleave the glycosidic linkages of the carbohydrate (Helbert, 2017). Similar

multi-enzyme degradation pathways of carbohydrates can be found in the human gut, where the symbiont *Bacteroidetes thetaiotaomicron* uses both sulfatases and GHs to utilise mucin glycoproteins as a nutrient source during colonisation (Luis et al., 2021). Given the enrichment in both sulfatases and sulfotransferases, along with their co-occurrence with enriched GHs described above, it is possible these symbionts are involved in both degrading and synthesising sulfated polysaccharides. Further, an enrichment in sulfatases and sulfotransferases has been described in other sponge symbionts such as the Poribacteria (Kamke et al., 2013; Slaby et al., 2017), suggesting that sponge symbionts could play important roles in the cellular structure and organisation of the sponge.

5.5.6. Sponge associated Spirochaetaceae unique in their potential for urea degradation

Marine invertebrates generally excrete nitrogenous waste as ammonia (ammonotelic), however sponges may also excrete this as urea (Morley et al., 2016). Similarly, microbial metabolism, such as the degradation of creatine, may also produce urea as a by-product within the sponge holobiont (Moitinho-Silva et al., 2017). Thus, urea degradation by microbial symbionts may play a role in removal of sponge waste products. Spirochaetes are not known for their ability to breakdown urea (Solomon et al., 2010), yet the sponge associated Spirochaetaceae genomes showed an enrichment of ureases and urea transport genes. This may represent a unique Spirochaete function that has been acquired within the sponge microbiome, and likely assists the holobiont in maintaining homeostasis. Urea degradation has been suggested as a common trait of sponge symbionts and an alternative explanation is that urea metabolism represents a method to release nitrogen to the microbiome (Moitinho-Silva et al., 2017). For example, cleavage of urea results in ammonia which could then be used as a nitrogen source for ammonia oxidising microbes that occupy the sponge microbiome. Therefore, urea

degradation may not only remove metabolic waste but additionally recycle the nitrogen within the microbiome.

5.5.7. High copy numbers of phytanoyl-CoA dioxygenase in sponge symbionts

One particularly striking pattern I observed was an enrichment in phytanoyl-CoA dioxygenase (phyH). This was most obvious in the Spirochaetaceae genomes, where those assembled from sponges had 129 copies on average per genome. Although not as common as some of the other symbiont traits, similarly high copy numbers have been observed in Poribacteria genomes from sponges (Kamke et al., 2014). However, the low amino acid identity (AAI) similarity among proteins within this domain (Kamke et al., 2014), and the high versatility of oxidative reactions (Schofield & McDonough, 2007), has meant the relevance of such high copy numbers is not clearly understood. Functional characterisation of the bacterial phyH genes have suggested roles as diverse as involvement in quorum sensing and metabolism of dissolved organic phosphorus (Y. Hao et al., 2010; Martinez et al., 2010), and further work would benefit from resolving the function of phyH in sponge associated microbes.

5.5.8. Sponge symbionts investigated here likely represent a facultative symbiosis

Reduced genome size and low GC content are often described as characteristics of obligate intracellular symbionts, particularly those symbionts that have undergone host-microbe coevolution (Moran & Baumann, 2000). Results presented here are consistent with previous estimates of sponge symbiont genome sizes and showed that genome size and GC content were generally higher in sponge symbionts compared to their free-living counterparts (Horn et al., 2016). This suggests that the sponge associated symbionts investigated here are unlikely to be obligate intracellular associates and may have acquired new genes to facilitate host interactions rather than lost redundant genes. Additionally, this may reflect that these

symbionts have a free-living stage and therefore retained the necessary genes for both free-living and host-associated lifestyles. Similarly, it suggests that symbionts are horizontally transmitted, since genome size negatively correlates with host dependence for vertically transmitted but not horizontally transmitted symbionts (Fisher et al., 2017). An exception was two genomes from Nitrosopumiliaceae isolated from the sponge *Coscinoderma sp.*, which had genome sizes far smaller (<0.85Mbp) and GC content far lower (<30%) than any other symbiont genome in this study. These two genomes did not fall within the general sponge clade and did not share gene enrichment patterns with other sponge symbionts. This may reflect the specialist/generalist dichotomy previously described within the Nitrosopumiliaceae (Zhang et al., 2019), and future studies may benefit from research aimed specifically towards these potentially intracellular, obligate symbionts.

5.5.9. Conclusions

From this study and others, it appears that carbohydrate degradation using glycosyl hydrolases (in particular sialidases), avoiding phagocytosis with ELPs, defence against MGEs using restriction enzymes and attachment to the host with fibronectins are all common symbiont traits, which have potentially been acquired through lateral gene transfer within the sponge microbiome (Robbins et al., 2021). These enrichment patterns suggest a multitude of ways symbionts have adapted to thrive in a hostile sponge environment. The sponge symbiont molecular repertoire was expanded in this study to show an enrichment of SOD, which is potentially involved in protecting against phagocytosis and oxidative damage. It is also demonstrated that sulfatases and sulfotransferases potentially work with glycosyl hydrolases to break down and remodel carbohydrates. Finally, urea metabolism within the Spirochaetaceae appears to be unique to sponge-associated Spirochaetes.

Many of the mechanisms used to survive the host environment have been characterised in pathogenic bacteria, but this alone does not suggest these symbionts are pathogenic/parasitic, since mutualistic symbionts often employ the same tools for host colonisation (Ochman & Moran, 2001). Likewise, the metabolic capability of these symbionts could also be interpreted as beneficial or parasitic. For example, abundant sulfatases, sulfotransferases and GHs could be seen as beneficial if symbionts are degrading, remodelling and synthesising sulfated polysaccharides that could then be used by the sponge for cell aggregation. Alternatively, this could be seen as parasitic if symbionts are simply taking advantage of the sponge resources which would otherwise be used by the sponge. In any case, the metabolism of sulfated polysaccharides along with other carbohydrates highlights that sponge symbionts are important for the ecological role in metabolising DOM (De Goeij et al., 2013). Although a well-established suite of symbiotic traits exists, such as genome streamlining and reduced GC content, results from this chapter show that sponge-specific symbionts carry their own unique characteristics that reflect their evolution towards a sponge associated lifestyle.

**Chapter 6: General discussion. Does the microbiome of coral reef
invertebrates coevolve with their host?**

6.1. Summary

Research conducted in this thesis has significantly improved our understanding of host-microbe interactions within coral reef invertebrates. Firstly, the thesis expands our understanding of how the microbiota associated with reef invertebrates is assembled by illustrating the importance of host evolutionary history. Secondly, the use of a unique tool for uncovering potentially important symbionts through the application of cophylogeny is demonstrated. Thirdly, metatranscriptomics was used to validate metabolic pathways that are critical to facilitating interactions between the sponge host and microbiota. Finally, comparative genomics is used to illustrate how key symbionts differ from free living relatives indicating potential adaptation of the sponge microbiome to a host associated lifestyle. In this final discussion, I recap the questions this thesis aimed to address, synthesise results that contributed to resolving these questions and suggest ways future research could further expand our understanding of host-microbe coevolution. Finally, I deconstruct the evidence that suggests coral reef invertebrate microbiomes coevolve with the host.

6.2. Coral reef invertebrate microbiomes are a complex web of interactions

The microbiomes of coral reef invertebrates are complex, with hundreds to thousands of microbial taxa challenging the identification of important lineages and their metabolic functions that underpin host fitness (see chapter 1). Many studies have focused on understanding how the host microbiota is structured and influenced by the surrounding environment (Lima et al., 2020), though the evolutionary history of the host has been largely overlooked. This is despite research on terrestrial invertebrates and mammals illustrating the importance of host phylogeny on the microbiota for over a decade (Brucker & Bordenstein, 2011; Ley et al., 2008; Ochman et al., 2010b). In this thesis, I showed that even when environmental variables are not controlled, a signal of the host phylogeny is reflected in the

assemblage of the microbiota in coral reef invertebrates (chapter 2). Although there are many potential drivers of phyllosymbiosis (Mallott & Amato, 2021), there are two main hypotheses on why this pattern arises. The first argues that host traits are often phylogenetically related, and these traits can act as a filter to microbes attempting to colonise a particular niche of the host (Mazel et al., 2018). The second argues that phyllosymbiosis could arise when dominant symbionts codiverge with their host and are faithfully transmitted across generations (Brucker & Bordenstein, 2012). In either case, the pattern of phyllosymbiosis suggests that the same microbes persistently colonise a host and thus likely exert a selection pressure. Importantly these alternative explanations need not be mutually exclusive.

While phyllosymbiosis is a useful analysis to understand microbial patterns at the community level, it does not allow us to identify individual lineages which may be of a comparatively high importance to the host. Previous research had often achieved this by looking to the core microbiome (Astudillo-García et al., 2017; Hernandez-Agrede et al., 2017). The assumption being that microbial taxa which are found to be consistently associated with a host are more likely to underpin host health, or even exert a selective pressure. In this thesis, I went beyond the core microbiome and uncovered microbes that had a congruent evolutionary history with the host along with those that cooccur in sub-networks. Despite limited host specialists among the microbial taxa of coral reef invertebrates, a cophylogenetic analysis makes it possible to identify taxa that interact more frequently with a particular host species, and subsequently test if the phylogenetic relationships are congruent (Blasco-Costa et al., 2021). This allowed me to implement strict criteria in the search for important symbionts using a tool that had been previously underutilised for studies of complex symbiosis. From the results of chapter 3, along with the investigation on phyllosymbiosis (chapter 2), I was able to demonstrate that within the complex microbial consortia there is a deterministic structure with a subset of microbes that likely have a larger influence on host health.

While chapters 2 & 3 provided important insights into the structure of the invertebrate host microbiome and identified potentially important lineages, these methods are not able to discern microbial function or reveal the molecular mechanisms behind host-microbe interaction. To date, much of this knowledge has been produced using metagenomic approaches (Engelberts et al., 2020; Glasl et al., 2020; Robbins et al., 2019). However, metagenomics can only infer functional potential based on encoded enzymes, protein families and metabolic pathways. Therefore, in this thesis I sought to validate some of the commonly held hypotheses of microbial functions, particularly in sponges, by mapping metatranscriptomic data to symbiont genomes to reveal which genes are actively expressed. Applying a genome-centric metatranscriptomic approach, I was able to show which microbial taxa are involved in a particular metabolic pathway. This reinforced previous results, such as cyanobacteria being responsible for fixing carbon within the sponge holobiont (Gao et al., 2014; Moitinho-Silva et al., 2014), and demonstrated new results, such as desulfobacteria oxidising ammonia. Further, these results were consistent with metagenomic predictions and this protocol will serve as a useful method for understanding microbial function in future studies.

In the final chapter, I aimed to test a hypothesis that was developed in the thesis - do symbionts with cophylogenetic patterns have an intimate association with the host? One way to test this was to compare the genomes of sponge symbionts to the genomes of closely related microbes isolated from other environments (in most cases free-living). This analysis showed that sponge symbionts were enriched in many of the classic symbiotic signatures, such as ELPs, R-M enzymes and fibronectin domains (Robbins et al., 2021). However, it also uncovered patterns that had not previously been presented, such as an enrichment in superoxide dismutase and ureases along with urea transporters. Together, these enrichment patterns suggested ways in which a symbiont has adapted to living within the sponge microbiome and highlighted

potential functions that could be beneficial to the host. Further, many symbionts of sponges share similar characteristics despite being divergent lineages. This suggests some degree of HGT occurring within the sponge microbiome, which is likely exacerbated under a phyllosymbiotic microbial assemblage, as this suggests continuous colonisation of the same microbial lineages.

6.3. Future research priorities

Coevolution is the reciprocal evolution of interacting species. Given this thesis focussed on the microbiota of coral reef invertebrates to better understand community assemblage and function, it can only reflect one side of the coevolution story (Groussin et al., 2020). To complete the picture, the next phase would be to sequence the host genome and transcriptome to identify any evidence that the host is a) adapting in response to the microbiome, or b) working together with the microbiome through processes such as metabolic collaboration. Such an approach is necessary for the full characterisation of host-microbe symbioses where coevolution has been identified. For example, through sequencing the genomes and transcriptomes of both the pea aphid host and its microbial symbionts, it was possible to confirm shared metabolic pathways for amino acid synthesis, fulfilling a nutrient requirement which allows the host to occupy a specialised niche (Consortium, 2010; Hansen & Moran, 2011). The metatranscriptomic analysis in chapter 4 identified many pathways that were partially expressed and combining this data with a host genome and transcriptome may uncover similar findings. Further, the microbiota and their host may collaborate in other ways, i.e., use each-others compounds even when metabolic pathways are not shared. However, whether or not this arises from reciprocal evolution, or fosters reciprocal evolution, would be an interesting focal point for future research.

Future studies would also benefit from using microscopy methods to visualise the symbiosis (Engelberts et al., 2021). For example, a host might house its symbionts in specialised cells and in return a microbe may provide certain nutrients or metabolites. Using targeted *in situ* visualisation (e.g. secondary-ion mass spectrometry [SIMS]) of metabolite passage would support theories of carbon translocation from symbiont to sponge, or biosynthesis of B vitamins translocated from symbiont to sponge (Hudspith et al., 2021). Combining these methods with fluorescence *in situ* hybridisation (FISH) would reveal the localisation of symbionts and identify if they form aggregates similar to symbionts seen within bacteriocytes (Engelberts et al., 2021). There is some evidence to suggest this is the case in corals and sponges (Maldonado, 2007; Work & Aeby, 2014), and future studies would benefit from characterising which symbionts are distributed this way and how widespread this is among hosts. Finally, microscopy methods such as FISH can be useful to understand if microbes are vertically transmitted by observing the transmitted microbes in gametes or larvae in the case of brooding colonies (Damjanovic et al., 2020). This has implications for symbiosis and coevolution studies as microbes that are faithfully passed from generation to generation have a greater likelihood of being critical for host health.

The genomics analysis presented in this thesis (chapters 4 & 5) focussed on sponges and thus these conclusions cannot necessarily be extrapolated to other invertebrates such as coral. One of the major bottlenecks for studying the microbiome of corals is that successfully obtaining a comprehensive metagenome is difficult. In sponges, for example, microbial cells can easily be enriched before sequencing, removing much of the host contamination prior to bioinformatic analysis (Thomas et al., 2010). However, no such method currently exists for coral despite years of research and further development of alternative methods of microbial enrichment, such as digestion enzymes to remove host DNA, would be beneficial to coral microbiologists. Similarly, methods for ribosomal RNA subtraction could be further optimised

to improve metatranscriptomic protocols. Synthesising probes to target specific host nucleic acids is time consuming and laborious with variable outcomes (chapter 4). Ribosomal depletion kits are becoming more effective and could be considered for future studies. Finally, while metatranscriptomics is an ideal tool to investigate microbial gene expression, gene expression doesn't necessarily lead to translated proteins and future studies may use additional omics methods, such as proteomics or metabolomics, to have a more comprehensive understanding of microbial metabolism (Engelberts et al., 2021).

6.4. Concluding remarks: Why it's important to understand whether coral reef invertebrates coevolve with their microbiome

Data generated from this comprehensive study contributes to answering the question: do coral reef invertebrates coevolve with their microbiome? I have shown that a) evolutionary history is reflected in the microbiome, b) a subset of microbes display strong patterns of cophylogeny, and c) the genomes of those microbes show evidence of adaptation to the sponge host (Figure 6.1). While further evidence is required to unambiguously answer this question, the information presented in this thesis makes a stronger case for host-microbe coevolution.

A major counter argument to attributing the patterns of phylosymbiosis, cophylogeny and gene enrichment to coevolution, is that the microbe(s) could be adaptively tracking the host, with no influence on host evolution. The results in this thesis do not disprove this hypothesis, however some perspectives can be gained when considering other host-microbe systems. For example, the human gut draws similar comparisons in terms of the metabolic complexity provided by the microbiota, which is believed to have been a driving evolutionary force behind the host-microbe coevolution in humans (Van den Abbeele et al., 2011). Here, the patterns of phylosymbiosis, cophylogeny and gene enrichment suggest persistent colonisation of a host by the same microbial community over many generations. This microbial community

is also the source of a comprehensive metabolic capacity (chapter 4), which interacts with the host and likely applies a selective pressure on the host. Thus, it is feasible the microbiota could affect the host phenotype. However, less obvious is whether the host phenotype affected by the microbiota is heritable (Mallott & Amato, 2021).

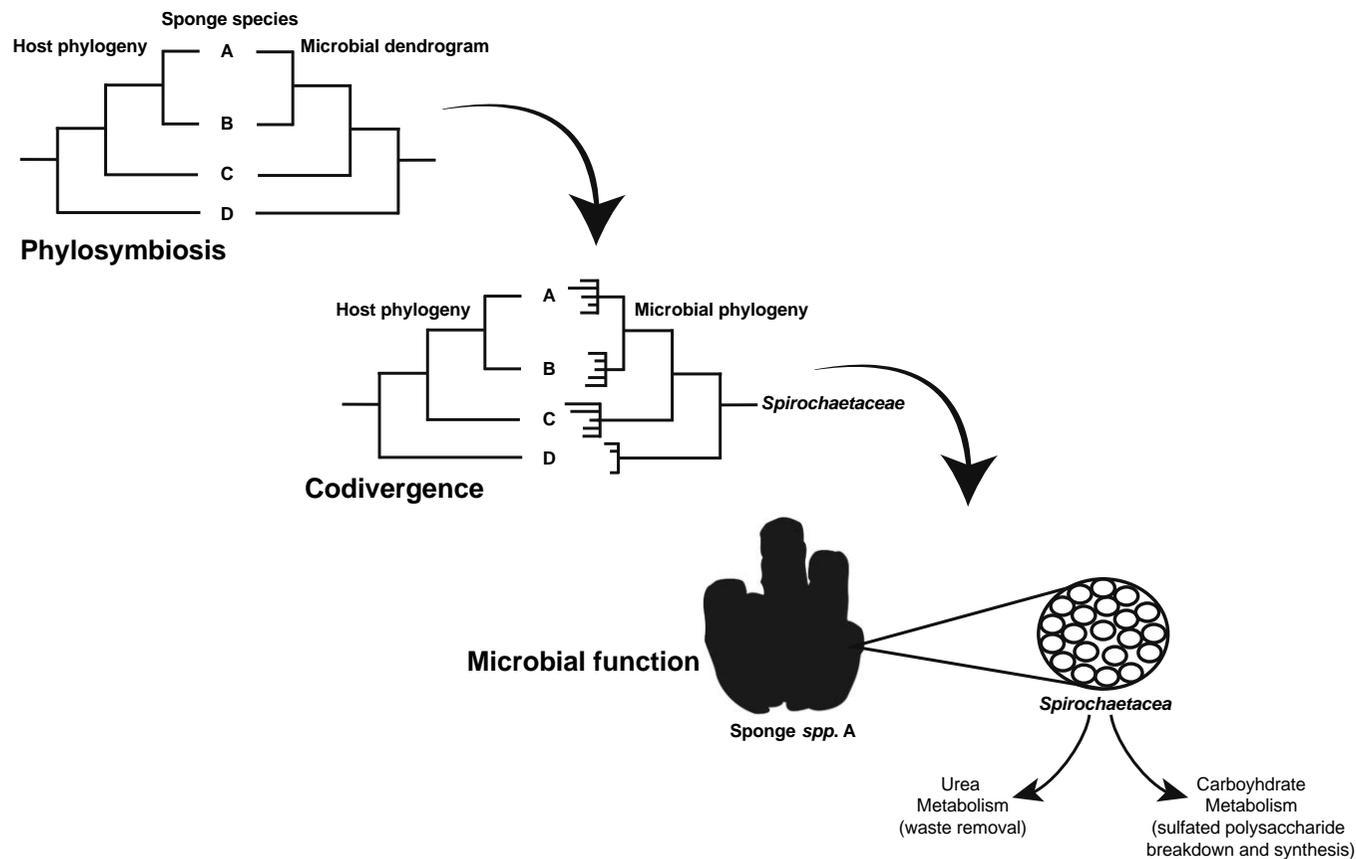


Figure 6.1. Overview of the three major lines of evidence that support host-microbe coevolution in sponges (adapted from Figure 1.2). The microbiome of the host is reflected in the host phylogeny (phylosymbiosis), a subset of those microbes show evidence of codivergence (for example, the *Spirochaetaceae*), the genomes of these microbes suggest an adaptation to the host with unique functions (for example, urea metabolism).

Understanding the complexity of microbial symbioses in coral reef invertebrates is a critical step to ensuring coral reefs exist in future generations. This thesis took a unique perspective to address these challenges, synthesising evidence of coevolution to greatly improve our knowledge behind the formation and function of the microbiome. Linking the patterns described here to mechanisms of host survival and evolution are a major goal for coevolution studies (Blasco-Costa et al., 2021), and this thesis has laid the foundations for future research to pursue these questions in coral reef invertebrates. As new and innovative methods of reef restoration and adaptation are being explored (Bay et al., 2019), particularly those including microbial interventions, it has never been more timely for a comprehensive understanding of the coral reef invertebrate holobiont.

References

- Abdul Wahab, M. A., Wilson, N. G., Prada, D., Gomez, O., & Fromont, J. (2021). Molecular and morphological assessment of tropical sponges in the subfamily Phyllospongiinae, with the descriptions of two new species. *Zoological Journal of the Linnean Society*, *193*(1), 319–335.
- Agashe, D., & Shankar, N. (2014). The evolution of bacterial DNA base composition. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *322*(7), 517–528.
- Ainsworth, T. D., Krause, L., Bridge, T., Torda, G., Raina, J. B., Zakrzewski, M., Gates, R. D., Padilla-Gamiño, J. L., Spalding, H. L., Smith, C., Woolsey, E. S., Bourne, D. G., Bongaerts, P., Hoegh-Guldberg, O., & Leggat, W. (2015). The coral core microbiome identifies rare bacterial taxa as ubiquitous endosymbionts. *ISME Journal*, *9*(10), 2261–2274. <https://doi.org/10.1038/ismej.2015.39>
- Alneberg, J., Bjarnason, B. S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J., Andersson, A. F., & Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods*, *11*(11), 1144–1146.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.
- Amato, K. R., G. Sanders, J., Song, S. J., Nute, M., Metcalf, J. L., Thompson, L. R., Morton, J. T., Amir, A., J. McKenzie, V., Humphrey, G., Gogul, G., Gaffney, J., L. Baden, A., A.O. Britton, G., P. Cuzzo, F., Di Fiore, A., J. Dominy, N., L. Goldberg, T., Gomez, A., ... R. Leigh, S. (2019). Evolutionary trends in host physiology outweigh dietary niche in

structuring primate gut microbiomes. *ISME Journal*, 13(3), 576–587.
<https://doi.org/10.1038/s41396-018-0175-0>

Amato, K. R., Mallott, E. K., McDonald, D., Dominy, N. J., Goldberg, T., Lambert, J. E., Swedell, L., Metcalf, J. L., Gomez, A., Britton, G. A. O., Stumpf, R. M., Leigh, S. R., & Knight, R. (2019). Convergence of human and Old World monkey gut microbiomes demonstrates the importance of human ecology over phylogeny. *Genome Biology*, 20(1), 1–12. <https://doi.org/10.1186/s13059-019-1807-z>

Anantharaman, V., & Aravind, L. (2010). Novel eukaryotic enzymes modifying cell-surface biopolymers. *Biology Direct*, 5(1), 1–8.

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2), 166–169.

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, 27(2), 93–115.

Apprill, A., Marlow, H. Q., Martindale, M. Q., & Rappé, M. S. (2009). The onset of microbial associations in the coral *Pocillopora meandrina*. *ISME Journal*, 3(6), 685–699.
<https://doi.org/10.1038/ismej.2009.3>

Apprill, A., McNally, S., Parsons, R., & Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology*, 75(2), 129–137. <https://doi.org/10.3354/ame01753>

Aranda, M., Li, Y., Liew, Y. J., Baumgarten, S., Simakov, O., Wilson, M. C., Piel, J., Ashoor, H., Bougouffa, S., Bajic, V. B., Ryu, T., Ravasi, T., Bayer, T., Micklem, G., Kim, H., Bhak, J., LaJeunesse, T. C., & Voolstra, C. R. (2016). Genomes of coral dinoflagellate

- symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Scientific Reports*, 6(August), 1–15. <https://doi.org/10.1038/srep39734>
- Archibald, J. M. (2015). Endosymbiosis and eukaryotic cell evolution. *Current Biology*, 25(19), R911–R921. <https://doi.org/10.1016/j.cub.2015.07.055>
- Astudillo-García, C., Bell, J. J., Montoya, J. M., Moitinho-Silva, L., Thomas, T., Webster, N. S., & Taylor, M. W. (2020). Assessing the strength and sensitivity of the core microbiota approach on a highly diverse sponge reef. *Environmental Microbiology*, 22, 3985–3999. <https://doi.org/10.1111/1462-2920.15185>
- Astudillo-García, 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(4), 1450–1462. <https://doi.org/10.1111/1462-2920.13647>
- Balbuena, J. A., Míguez-Lozano, R., & Blasco-Costa, I. (2013). PACo: A Novel Procrustes Application to Cophylogenetic Analysis. *PLoS ONE*, 8(4). <https://doi.org/10.1371/journal.pone.0061048>
- Baldassano, S. N., & Bassett, D. S. (2016). Topological distortion and reorganized modular structure of gut microbial co-occurrence networks in inflammatory bowel disease. *Scientific Reports*, 6(1), 1–14. <https://doi.org/10.1038/srep26087>
- Bandi, C., Anderson, T. J. C., Genchi, C., & Blaxter, M. L. (1998). Phylogeny of Wolbachia in filarial nematodes. *Proceedings of the Royal Society B: Biological Sciences*, 265(1413), 2407–2413. <https://doi.org/10.1098/rspb.1998.0591>

- Battistoni, A. (2003). Role of prokaryotic Cu,Zn superoxide dismutase in pathogenesis. *Biochemical Society Transactions*, 31(6), 1326–1329. <https://doi.org/10.1042/bst0311326>
- Baumann, P., Moran, N. A., & Baumann, L. (1997). The evolution and genetics of aphid endosymbionts. *BioScience*, 47(1), 12–20. <https://doi.org/10.2307/1313002>
- Bay, L. K., Rocker, M., Boström-Einarsson, L., Babcock, R., Buerger, P., Cleves, P., Harrison, D., Negri, A., Quigley, K., & Randall, C. J. (2019). *Reef restoration and adaptation program: intervention technical summary. a report provided to the Australian government by the reef restoration and adaptation program.*
- Berry, D., & Loy, A. (2018). Stable isotope probing of human and animal microbiome function. *Trends in Microbiology*, 13(1), 8–10. <https://doi.org/10.1016/j.tim.2018.06.004>
- Berry, D., & Widder, S. (2014). Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Frontiers in Microbiology*, 5, 1–14. <https://doi.org/10.3389/fmicb.2014.00219>
- Bettarel, Y., Bouvier, T., Nguyen, H. K., & Thu, P. T. (2014). The versatile nature of coral-associated viruses. *Environmental Microbiology*, 17(10), 3433–3439. <https://doi.org/10.1111/1462-2920.12579>
- Björk, J. R., Díez-Vives, C., Astudillo-García, C., Archie, E. A., & Montoya, J. M. (2019). Vertical transmission of sponge microbiota is inconsistent and unfaithful. *Nature Ecology and Evolution*, 3(8), 1172–1183. <https://doi.org/10.1038/s41559-019-0935-x>
- Blasco-Costa, I., Hayward, A., Poulin, R., & Balbuena, J. A. (2021). Next-generation cophylogeny: unravelling eco-evolutionary processes. *Trends in Ecology and Evolution*, 36(10), 907–918. <https://doi.org/10.1016/j.tree.2021.06.006>

- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., & Asnicar, F. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857.
- Bondarev, V., Richter, M., Romano, S., Piel, J., Schwedt, A., & Schulz-Vogt, H. N. (2013). The genus *Pseudovibrio* contains metabolically versatile bacteria adapted for symbiosis. *Environmental Microbiology*, *15*(7), 2095–2113.
- Bordenstein, S. R., & Theis, K. R. (2015). Host biology in light of the microbiome: Ten principles of holobionts and hologenomes. *PLoS Biology*, *13*(8), 1–23. <https://doi.org/10.1371/journal.pbio.1002226>
- Botté, E. S., Nielsen, S., Abdul Wahab, M. A., Webster, J., Robbins, S., Thomas, T., & Webster, N. S. (2019). Changes in the metabolic potential of the sponge microbiome under ocean acidification. *Nature Communications*, *10*(1), 1–10. <https://doi.org/10.1038/s41467-019-12156-y>
- Bourne, D., Iida, Y., Uthicke, S., & Smith-keune, C. (2008). Changes in coral-associated microbial communities during a bleaching event. *ISME Journal*, *2*, 350–363. <https://doi.org/10.1038/ismej.2007.112>
- Boyd, J. A., Woodcroft, B. J., & Tyson, G. W. (2018). GraftM: a tool for scalable, phylogenetically informed classification of genes within metagenomes. *Nucleic Acids Research*, *46*(10), e59–e59.
- Brooks, A. W., Kohl, K. D., Brucker, R. M., van Opstal, E. J., & Bordenstein, S. R. (2016). *Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across*

- Host Evolutionary History. *PLoS Biology*, 14(11), 1–29.
<https://doi.org/10.1371/journal.pbio.2000225>
- Broxton, C. N., & Culotta, V. C. (2016). SOD Enzymes and Microbial Pathogens: Surviving the Oxidative Storm of Infection. *PLoS Pathogens*, 12(1), 8–13.
<https://doi.org/10.1371/journal.ppat.1005295>
- Brucker, R. M., & Bordenstein, S. R. (2011). The roles of host evolutionary relationships (genus: *Nasonia*) and development in structuring microbial communities. *Evolution*, 66(2), 349–362.
- Brucker, R. M., & Bordenstein, S. R. (2012). Speciation by symbiosis. *Trends in Ecology and Evolution*, 27(8), 443–451. <https://doi.org/10.1016/j.tree.2012.03.011>
- Brucker, R. M., & Bordenstein, S. R. (2013). The Hologenomic Basis of Speciation: Gut Bacteria Cause Hybrid Lethality in the Genus *Nasonia*. *Science*, 341, 667–669.
<https://doi.org/10.1126/science.1240659>
- Brune, A. (2014). Symbiotic digestion of lignocellulose in termite guts. *Nature Reviews Microbiology*, 12(3), 168–180. <https://doi.org/10.1038/nrmicro3182>
- Brune, A., & Dietrich, C. (2015). The Gut Microbiota of Termites: Digesting the Diversity in the Light of Ecology and Evolution. *Annual Review of Microbiology*, 69(1), 145–166.
<https://doi.org/10.1146/annurev-micro-092412-155715>
- Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12(1), 59–60.
- Burgsdorf, I., Sizikov, S., Squatrito, V., Britstein, M., Slaby, B. M., Cerrano, C., Handley, K. M., & Steindler, L. (2021). Lineage-specific energy and carbon metabolism of sponge

symbionts and contributions to the host carbon pool. *The ISME Journal*, May.
<https://doi.org/10.1038/s41396-021-01165-9>

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. <https://doi.org/10.1038/nmeth.3869>

Campana, S., Busch, K., Hentschel, U., Muyzer, G., & de Goeij, J. M. (2021). DNA-stable isotope probing (DNA-SIP) identifies marine sponge-associated bacteria actively utilizing dissolved organic matter (DOM). *Environmental Microbiology*, *23*(8), 4489–4504. <https://doi.org/10.1111/1462-2920.15642>

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, *17*(4), 540–552.

Ceh, J., van Keulen, M., & Bourne, D. G. (2013). Intergenerational Transfer of Specific Bacteria in Corals and Possible Implications for Offspring Fitness. *Microbial Ecology*, *65*(1), 227–231. <https://doi.org/10.1007/s00248-012-0105-z>

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020). *GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database*. Oxford University Press.

Chen, C., Tseng, C., Chen, C. A., & Tang, S. (2011). The dynamics of microbial partnerships in the coral *Isopora palifera*. *The ISME Journal*, *5*(4), 728–740. <https://doi.org/10.1038/ismej.2010.151>

Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., Li, Y., Ye, J., Yu, C., & Li, Z. (2018). SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience*, *7*(1), gix120.

- Chomicki, G., Kiers, E. T., & Renner, S. S. (2020). The Evolution of Mutualistic Dependence. *Annual Review of Ecology, Evolution, and Systematics*, 51(1), 409–432. <https://doi.org/10.1146/annurev-ecolsys-110218-024629>
- Clark, M. A., Moran, N. A., Baumann, P., & Wernegreen, J. J. (2000). Cospeciation between Bacterial Endosymbionts (Buchnera) and a Recent Radiation of Aphids (Uroleucon) and Pitfalls of Testing for Phylogenetic Congruence. *Evolution*, 54(2), 517–525.
- Collins, S. M., Surette, M., & Bercik, P. (2012). The interplay between the intestinal microbiota and the brain. *Nature Reviews Microbiology*, 10(11), 735–742. <https://doi.org/10.1038/nrmicro2876>
- Consortium, I. A. G. (2010). Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biology*, 8(2), e1000313.
- Costa, T. R. D., Felisberto-Rodrigues, C., Meir, A., Prevost, M. S., Redzej, A., Trokter, M., & Waksman, G. (2015). Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nature Reviews Microbiology*, 13(6), 343–359.
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695(5), 1–9.
- Cullen, P. A., Haake, D. A., & Adler, B. (2004). Outer membrane proteins of pathogenic spirochetes. *FEMS Microbiology Reviews*, 28(3), 291–318. <https://doi.org/10.1016/j.femsre.2003.10.004>
- Damjanovic, K., Menéndez, P., Blackall, L. L., & van Oppen, M. J. H. (2020). Mixed-mode bacterial transmission in the common brooding coral *Pocillopora acuta*. *Environmental Microbiology*, 22(1), 397–412.

- Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772.
- Dash, S., Duraivelan, K., & Samanta, D. (2021). Cadherin-mediated host–pathogen interactions. *Cellular Microbiology*, *23*(5), e13316.
- De Goeij, J. M., Van Oevelen, D., Vermeij, M. J. A., Osinga, R., Middelburg, J. J., De Goeij, A. F. P. M., & Admiraal, W. (2013). Surviving in a marine desert: the sponge loop retains resources within coral reefs. *Science*, *342*(6154), 108–110.
- De Groote, M. A., Ochsner, U. A., Shiloh, M. U., Nathan, C., McCord, J. M., Dinauer, M. C., Libby, S. J., Vazquez-Torres, A., Xu, Y., & Fang, F. C. (1997). Periplasmic superoxide dismutase protects *Salmonella* from products of phagocyte NADPH-oxidase and nitric oxide synthase. *Proceedings of the National Academy of Sciences*, *94*(25), 13997–14001.
- Dedysh, S. N., & Yilmaz, P. (2018). Refining the taxonomic structure of the phylum Acidobacteria. *International Journal of Systematic and Evolutionary Microbiology*, *68*(12), 3796–3806. <https://doi.org/10.1099/ijsem.0.003062>
- Deines, P., & Bosch, T. C. G. (2016). Transitioning from microbiome composition to microbial community interactions: The potential of the metaorganism hydra as an experimental model. *Frontiers in Microbiology*, *7*, 1–7. <https://doi.org/10.3389/fmicb.2016.01610>
- Dendooven, T., Luisi, B. F., & Bandyra, K. J. (2020). RNA lifetime control, from stereochemistry to gene expression. *Current Opinion in Structural Biology*, *61*, 59–70.
- Díez-Vives, C., Moitinho-Silva, L., Nielsen, S., Reynolds, D., & Thomas, T. (2017). Expression of eukaryotic-like protein in the microbiome of sponges. *Molecular Ecology*, *26*(5), 1432–1451. <https://doi.org/10.1111/mec.14003>

- D’Orazio, M., Folcarelli, S., Mariani, F., Colizzi, V., Rotilio, G., & Battistoni, A. (2001). Lipid modification of the Cu, Zn superoxide dismutase from *Mycobacterium tuberculosis*. *Biochemical Journal*, *359*(1), 17–22.
- Douglas, A. E., & Werren, J. H. (2016). Holes in the Hologenome: Why Host-Microbe Symbioses Are Not Holobionts. *MBio*, *7*(2), e02099-15. <https://doi.org/10.1128/mBio.02099-15>.Invited
- Easson, C. G., & Thacker, R. W. (2014). Phylogenetic signal in the community structure of host-specific microbiomes of tropical marine sponges. *Frontiers in Microbiology*, *5*(OCT), 1–11. <https://doi.org/10.3389/fmicb.2014.00532>
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. <https://doi.org/https://doi.org/10.1093/nar/gkh340>
- Edler, D., Klein, J., Antonelli, A., & Silvestro, D. (2019). raxmlGUI 2.0 beta: a graphical interface and toolkit for phylogenetic analyses using RAxML. *BioRxiv*, 800912.
- Ehrlich, P. R., & Raven, P. H. (1964). *Butterflies and Plants: A Study in Coevolution*. *18*(4), 586–608. <https://doi.org/10.1079/IVP2001184>
- Engelberts, J. P., Robbins, S. J., Damjanovic, K., & Webster, N. S. (2021). Integrating novel tools to elucidate the metabolic basis of microbial symbiosis in reef holobionts. *Marine Biology*, *168*(12), 1–19. <https://doi.org/10.1007/s00227-021-03952-6>
- Engelberts, J. P., Robbins, S. J., de Goeij, J. M., Aranda, M., Bell, S. C., & Webster, N. S. (2020). Characterization of a sponge microbiome using an integrative genome-centric approach. *ISME Journal*, *14*(5), 1100–1110. <https://doi.org/10.1038/s41396-020-0591-9>

- Ericson, C. F., Eisenstein, F., Medeiros, J. M., Malter, K. E., Cavalcanti, G. S., Zeller, R. W., Newman, D. K., Pilhofer, M., & Shikuma, N. J. (2019). A contractile injection system stimulates tubeworm metamorphosis by translocating a proteinaceous effector. *ELife*, *8*, 1–19. <https://doi.org/10.7554/eLife.46845>
- Erwin, P. M., Coma, R., Paula, L., Serrano, E., Ribes, M., & Mar, P. (2015). Stable symbionts across the HMA-LMA dichotomy: low seasonal and interannual variation in sponge-associated bacteria from taxonomically. *FEMS Microbiology Ecology*, *91*, 1–11. <https://doi.org/10.1093/femsec/fiv115>
- Erwin, P. M., Pineda, M. C., Webster, N., Turon, X., & López-Legentil, S. (2014). Down under the tunic: Bacterial biodiversity hotspots and widespread ammonia-oxidizing archaea in coral reef ascidians. *ISME Journal*, *8*(3), 575–588. <https://doi.org/10.1038/ismej.2013.188>
- Fabricius, K. E., Langdon, C., Uthicke, S., Humphrey, C., Noonan, S., De'ath, G., Okazaki, R., Muehllehner, N., Glas, M. S., & Lough, J. M. (2011). Losers and winners in coral reefs acclimatized to elevated carbon dioxide concentrations. *Nature Climate Change*, *1*(3), 165–169. <https://doi.org/10.1038/nclimate1122>
- Falkowski, P. G. (1997). Evolution of the nitrogen cycle and its influence on the biological sequestration of CO₂ in the ocean. *Nature*, *387*(6630), 272–275.
- 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), E1878–E1887. <https://doi.org/10.1073/pnas.1203287109>

- Fenn, K., & Blaxter, M. (2004). Are filarial nematode Wolbachia obligate mutualist symbionts? *Trends in Ecology and Evolution*, 19(4), 163–166. <https://doi.org/10.1016/j.tree.2004.01.002>
- Fiore, C. L., Labrie, M., Jarett, J. K., & Lesser, M. P. (2015). Transcriptional activity of the giant barrel sponge, *Xestospongia muta* Holobiont: Molecular evidence for metabolic interchange. *Frontiers in Microbiology*, 6(APR). <https://doi.org/10.3389/fmicb.2015.00364>
- Fisher, R. M., Henry, L. M., Cornwallis, C. K., Kiers, E. T., & West, S. A. (2017). The evolution of host-symbiont dependence. *Nature Communications*, 8(May), 15973. <https://doi.org/10.1038/ncomms15973>
- Fraiberg, M., Borovok, I., Weiner, R. M., & Lamed, R. (2010). Discovery and characterization of cadherin domains in *Saccharophagus degradans* 2-40. *Journal of Bacteriology*, 192(4), 1066–1074.
- Franzenburg, S., Walter, J., Künzel, S., Wang, J., Baines, J. F., & Bosch, T. C. G. (2013). Distinct antimicrobial peptide expression determines host species-specific bacterial associations. *Proceedings of the National Academy of Sciences*, 110(39), E3730-E3738. <https://doi.org/10.1073/pnas.1304960110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1304960110>
- 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? *ISME Journal*, 7(6), 1116–1125. <https://doi.org/10.1038/ismej.2013.7>
- Friedman, W. E. (2009). The meaning of Darwin's "abominable mystery." *American Journal of Botany*, 96(1), 5–21. <https://doi.org/10.3732/ajb.0800150>

- Friel, A. D., Neiswenter, S. A., Seymour, C. O., Bali, L. R., McNamara, G., Leija, F., Jewell, J., & Hedlund, B. P. (2020). Microbiome shifts associated with the introduction of wild Atlantic horseshoe crabs (*Limulus polyphemus*) into a touch-tank exhibit. *BioRxiv*.
- Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, *31*(22), 3718–3720.
- Gao, Z. M., Wang, Y., Tian, R. M., Wong, Y. H., Batang, Z. B., Al-Suwailem, A. M., Bajic, V. B., & Qian, P. Y. (2014). Symbiotic Adaptation Drives Genome Streamlining of the Cyanobacterial Sponge Symbiont “Candidatus *Synechococcus spongiarum*”. *MBio*, *5*(2), 1–11. <https://doi.org/10.1128/mBio.00079-14>
- Gardner, S. G., Camp, E. F., Smith, D. J., Kahlke, T., Osman, E. O., Gendron, G., Hume, B. C. C., Pogoreutz, C., Voolstra, C. R., & Suggett, D. J. (2019). Coral microbiome diversity reflects mass coral bleaching susceptibility during the 2016 El Niño heat wave. *Ecology and Evolution*, *9*(3), 938–956.
- Gil-Agudelo, D. L., Myers, C., Smith, G. W., & Kim, K. (2006). Changes in the microbial communities associated with *Gorgonia ventalina* during aspergillosis infection. *Diseases Of Aquatic Organisms*, *69*, 89–94.
- Gittleman, J. L., & Kot, M. (1990). Adaptation: statistics and a null model for estimating phylogenetic effects. *Systematic Zoology*, *39*(3), 227–241.
- Glasl, B., Herndl, G. J., & Frade, P. R. (2016). The microbiome of coral surface mucus has a key role in mediating holobiont health and survival upon disturbance. *The ISME Journal*, *10*(9), 2280–2292. <https://doi.org/10.1038/ismej.2016.9>

- Glasl, B., Robbins, S., Frade, P. R., Marangon, E., Laffy, P. W., Bourne, D. G., & Webster, N. S. (2020). Comparative genome-centric analysis reveals seasonal variation in the function of coral reef microbiomes. *ISME Journal*, *14*(6), 1435–1450. <https://doi.org/10.1038/s41396-020-0622-6>
- Glasl, B., Smith, C. E., Bourne, D. G., & Webster, N. S. (2018). Exploring the diversity-stability paradigm using sponge microbial communities. *Scientific Reports*, *8*, 1–9. <https://doi.org/10.1038/s41598-018-26641-9>
- Glasl, B., Smith, C. E., Bourne, D. G., & Webster, N. S. (2019). Disentangling the effect of host-genotype and environment on the microbiome of the coral *Acropora tenuis*. *PeerJ*, *2019*(2), 1–18. <https://doi.org/10.7717/peerj.6377>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., & Zeng, Q. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, *29*(7), 644.
- Griffith, D. M., Veech, J. A., & Marsh, C. J. (2016). Cooccur: Probabilistic species co-occurrence analysis in R. *Journal of Statistical Software*, *69*(1), 1–17. <https://doi.org/10.18637/jss.v069.c02>
- Groussin, M., Mazel, F., & Alm, E. J. (2020). Co-evolution and Co-speciation of Host-Gut Bacteria Systems. *Cell Host and Microbe*, *28*(1), 12–22. <https://doi.org/10.1016/j.chom.2020.06.013>
- Groussin, M., Mazel, F., Sanders, J. G., Smillie, C. S., Lavergne, S., Thuiller, W., & Alm, E. J. (2017). Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nature Communications*, *8*(14319), 1–12. <https://doi.org/10.1038/ncomms14319>

- Gysi, D. M., & Nowick, K. (2020). Construction, comparison and evolution of networks in life sciences and other disciplines. *Journal of the Royal Society Interface*, *17*(166). <https://doi.org/10.1098/rsif.2019.0610>
- Haber, M., Burgsdorf, I., Handley, K. M., Rubin-Blum, M., & Steindler, L. (2021). Genomic Insights Into the Lifestyles of Thaumarchaeota Inside Sponges. *Frontiers in Microbiology*, *11*(January), 1–18. <https://doi.org/10.3389/fmicb.2020.622824>
- Habyarimana, F., Al-khodor, S., Kalia, A., Graham, J. E., Price, C. T., Garcia, M. T., & Kwaik, Y. A. (2008). Role for the Ankyrin eukaryotic-like genes of *Legionella pneumophila* in parasitism of protozoan hosts and human macrophages. *Environmental Microbiology*, *10*(6), 1460–1474.
- Hadaidi, G., Gegner, H. M., Ziegler, M., & Voolstra, C. R. (2019). Carbohydrate composition of mucus from scleractinian corals from the central Red Sea. *Coral Reefs*, *38*(1), 21–27. <https://doi.org/10.1007/s00338-018-01758-5>
- Hafner, M. S., Demastes, J. W., Spradling, T. A., & Reed, D. L. (2003). Cophylogeny between pocket gophers and chewing lice. *Tangled Trees: Phylogeny, Cospeciation, and Coevolution*. University of Chicago Press, Chicago, 195–218.
- Hansen, A. K., & Moran, N. A. (2011). Aphid genome expression reveals host–symbiont cooperation in the production of amino acids. *Proceedings of the National Academy of Sciences*, *108*(7), 2849–2854.
- Hao, L., Michaelsen, T. Y., Singleton, C. M., Dottorini, G., Kirkegaard, R. H., Albertsen, M., Nielsen, P. H., & Dueholm, M. S. (2020). Novel syntrophic bacteria in full-scale anaerobic digesters revealed by genome-centric metatranscriptomics. *ISME Journal*, *14*(4), 906–918. <https://doi.org/10.1038/s41396-019-0571-0>

- Hao, Y., Winans, S. C., Glick, B. R., & Charles, T. C. (2010). Identification and characterization of new LuxR/LuxI-type quorum sensing systems from metagenomic libraries. *Environmental Microbiology*, *12*(1), 105–117.
- Helbert, W. (2017). Marine polysaccharide sulfatases. *Frontiers in Marine Science*, *4*, 6.
- Hernandez-Agreda, A., Gates, R. D., & Ainsworth, T. D. (2017). Defining the Core Microbiome in Corals' Microbial Soup. *Trends in Microbiology*, *25*(2), 125–140. <https://doi.org/10.1016/j.tim.2016.11.003>
- Hernandez-agreda, A., Leggat, W., Ainsworth, T. D., & Hernandez-agreda, A. (2018). *A Comparative Analysis of Microbial DNA Preparation Methods for Use With Massive and Branching Coral Growth Forms*. *9*(September), 1–14. <https://doi.org/10.3389/fmicb.2018.02146>
- Herre, E. A., Knowlton, N., Mueller, U. G., & Rehner, S. A. (1999). The evolution of mutualisms: exploring the paths between conflict and cooperation. *Trends in Ecology and Evolution*, *14*(2), 49–53.
- Hird, S. M., Sánchez, C., Carstens, B. C., & Brumfield, R. T. (2015). Comparative gut microbiota of 59 neotropical bird species. *Frontiers in Microbiology*, *6*(1403), 1–16. <https://doi.org/10.3389/fmicb.2015.01403>
- Horn, H., Slaby, B. M., Jahn, M. T., Bayer, K., Moitinho-Silva, L., Förster, 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*(NOV), 1–15. <https://doi.org/10.3389/fmicb.2016.01751>

- Horvath, P., & Barrangou, R. (2010). CRISPR/Cas, the immune system of Bacteria and Archaea. *Science*, 327(5962), 167–170. <https://doi.org/10.1126/science.1179555>
- Hughes, T. P., Barnes, M. L., Bellwood, D. R., Cinner, J. E., Cumming, G. S., Jackson, J. B. C., Kleypas, J., Van De Leemput, I. A., Lough, J. M., Morrison, T. H., Palumbi, S. R., Van Nes, E. H., & Scheffer, M. (2017). Coral reefs in the Anthropocene. *Nature*, 546(7656), 82–90. <https://doi.org/10.1038/nature22901>
- Hughes, T. P., Kerry, J. T., Álvarez-Noriega, M., Álvarez-Romero, J. G., Anderson, K. D., Baird, A. H., Babcock, R. C., Beger, M., Bellwood, D. R., Berkelmans, R., Bridge, T. C., Butler, I. R., Byrne, M., Cantin, N. E., Comeau, S., Connolly, S. R., Cumming, G. S., Dalton, S. J., Diaz-Pulido, G., ... Wilson, S. K. (2017). Global warming and recurrent mass bleaching of corals. *Nature*, 543(7645), 373–377. <https://doi.org/10.1038/nature21707>
- Hughes, T. P., Kerry, J. T., Baird, A. H., Connolly, S. R., Dietzel, A., Eakin, C. M., Heron, S. F., Hoey, A. S., Hoogenboom, M. O., Liu, G., McWilliam, M. J., Pears, R. J., Pratchett, M. S., Skirving, W. J., Stella, J. S., & Torda, G. (2018). Global warming transforms coral reef assemblages. *Nature*, 556(7702), 492–496.
- Hutchinson, M. C., Cagua, E. F., Balbuena, J. A., Stouffer, D. B., & Poisot, T. (2017). paco: implementing Procrustean Approach to Cophylogeny in R. *Methods in Ecology and Evolution*, 8(8), 932–940. <https://doi.org/10.1111/2041-210X.12736>
- Hymes, J. P., & Klaenhammer, T. R. (2016). Stuck in the middle: Fibronectin-binding proteins in gram-positive bacteria. *Frontiers in Microbiology*, 7(SEP), 1–9. <https://doi.org/10.3389/fmicb.2016.01504>

- Ikeda-Ohtsubo, W., & Brune, A. (2009). Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and “*Candidatus Endomicrobium trichonymphae*.” *Molecular Ecology*, *18*(2), 332–342. <https://doi.org/10.1111/j.1365-294X.2008.04029.x>
- Jahn, M. T., Arkhipova, K., Markert, S. M., Stigloher, C., Lachnit, T., Pita, L., Kupczok, A., Ribes, M., Stengel, S. T., Rosenstiel, P., Dutilh, B. E., & Hentschel, U. (2019). A Phage Protein Aids Bacterial Symbionts in Eukaryote Immune Evasion. *Cell Host and Microbe*, *26*(4), 542-550.e5. <https://doi.org/10.1016/j.chom.2019.08.019>
- Janssen, S., McDonald, D., Gonzalez, A., Navas-Molina, J. A., Jiang, L., Xu, Z. Z., Winker, K., Kado, D. M., Orwoll, E., Manary, M., Mirarab, S., & Knight, R. (2018). Phylogenetic Placement of Exact Amplicon Sequences Improves Associations with Clinical Information. *MSystems*, *3*(3), 1–14.
- Janz, N., & Nylin, S. (1998). Butterflies and Plants: A Phylogenetic Study. *Evolution*, *52*(2), 486–502.
- Jernigan, K. K., & Bordenstein, S. R. (2014). Ankyrin domains across the Tree of Life. *PeerJ*, *2*(i), e264. <https://doi.org/10.7717/peerj.264>
- Johnson, R. C. (2013). *Biology of parasitic spirochaetes*. Academic Press.
- 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). <https://doi.org/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. *ISME Journal*, 7(12), 2287–2300. <https://doi.org/10.1038/ismej.2013.111>
- Kang, D. D., Froula, J., Egan, R., & Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*, 3, e1165.
- Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., & Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, 7, e7359.
- Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). *MAFFT online service : multiple sequence alignment , interactive sequence choice and visualization*. 20(July 2017), 1160–1166. <https://doi.org/10.1093/bib/bbx108>
- Katoh, K., & Standley, D. M. (2013). Mafft multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/doi:10.1093/molbev/mst010>.
- Kawaida, H., Ohba, K., Koutake, Y., Shimizu, H., Tachida, H., & Kobayakawa, Y. (2013). Symbiosis between hydra and chlorella: Molecular phylogenetic analysis and experimental study provide insight into its origin and evolution. *Molecular Phylogenetics and Evolution*, 66(3), 906–914. <https://doi.org/10.1016/j.ympev.2012.11.018>
- Keck, F., Rimet, F., Bouchez, A., & Franc, A. (2016). PhyloSignal: An R package to measure, test, and explore the phylogenetic signal. *Ecology and Evolution*, 6(9), 2774–2780. <https://doi.org/10.1002/ece3.2051>

- Kennedy, P. J., Cryan, J. F., Dinan, T. G., & Clarke, G. (2017). Kynurenine pathway metabolism and the microbiota-gut-brain axis. *Neuropharmacology*, *112*, 399–412. <https://doi.org/10.1016/j.neuropharm.2016.07.002>
- Klinges, J. G., Rosales, S. M., McMinds, R., Shaver, E. C., Shantz, A. A., Peters, E. C., Eitel, M., Wörheide, G., Sharp, K. H., Burkepile, D. E., Silliman, B. R., & Vega Thurber, R. L. (2019). Phylogenetic, genomic, and biogeographic characterization of a novel and ubiquitous marine invertebrate-associated Rickettsiales parasite, *Candidatus Aquarickettsia rohweri*, gen. nov., sp. nov. *ISME Journal*, *13*(12), 2938–2953. <https://doi.org/10.1038/s41396-019-0482-0>
- Kohl, K. D., Dearing, M. D., & Bordenstein, S. R. (2018). Microbial communities exhibit host species distinguishability and phyllosymbiosis along the length of the gastrointestinal tract. *Molecular Ecology*, *27*(8), 1874–1883. <https://doi.org/10.1111/mec.14460>
- Könneke, M., Bernhard, A. E., De La Torre, J. R., Walker, C. B., Waterbury, J. B., & Stahl, D. A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature*, *437*(7058), 543–546. <https://doi.org/10.1038/nature03911>
- Kopylova, E., Noé, L., & Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, *28*(24), 3211–3217.
- Koren, O., & Rosenberg, E. (2006). Bacteria associated with mucus and tissues of the coral *Oculina patagonica* in summer and winter. *Applied and Environmental Microbiology*, *72*(8), 5254–5259. <https://doi.org/10.1128/AEM.00554-06>
- Krueger, T., & Gates, R. D. (2012). Cultivating endosymbionts - Host environmental mimics support the survival of *Symbiodinium C15* ex hospite. *Journal of Experimental Marine Biology and Ecology*, *413*, 169–176. <https://doi.org/10.1016/j.jembe.2011.12.002>

- Lackner, G., Peters, E. E., Helfrich, E. J. N., & Piel, J. (2017). Insights into the lifestyle of uncultured bacterial natural product factories associated with marine sponges. *Proceedings of the National Academy of Sciences*, *114*(3), E347–E356. <https://doi.org/10.1073/pnas.1616234114>
- Laffy, P. W., Wood-charlson, E. M., Turaev, D., Jutz, S., Pascelli, C., Bell, S. C., Peirce, T. E., Weynberg, K. D., Oppen, M. J. H. Van, Rattei, T., & Webster, N. S. (2018). Reef invertebrate viromics : diversity , host specificity and functional capacity. *Environmental Microbiology*, *00*. <https://doi.org/10.1111/1462-2920.14110>
- LaJeunesse, T. C., Parkinson, J. E., Gabrielson, P. W., Jeong, H. J., Reimer, J. D., Voolstra, C. R., & Santos, S. R. (2018). Systematic Revision of Symbiodiniaceae Highlights the Antiquity and Diversity of Coral Endosymbionts. *Current Biology*, *28*(16), 2570-2580.e6. <https://doi.org/10.1016/j.cub.2018.07.008>
- Lawson, C. A., Raina, J. B., Kahlke, T., Seymour, J. R., & Suggett, D. J. (2018). Defining the core microbiome of the symbiotic dinoflagellate, Symbiodinium. *Environmental Microbiology Reports*, *10*(1), 7–11. <https://doi.org/10.1111/1758-2229.12599>
- Leigh, B. A., Bordenstein, S. R., Brooks, A. W., Mikaelyan, A., & Bordenstein, S. R. (2018). Finer-Scale Phyllosymbiosis: Insights from Insect Viromes. *MSystems*, *3*(6), e00131-18.
- Leite, D. C. A., Leão, P., Garrido, A. G., Lins, U., Santos, H. F., Pires, D. O., Castro, C. B., van Elsas, J. D., Zilberberg, C., Rosado, A. S., & Peixoto, R. S. (2017). Broadcast spawning coral *Mussismilia Hispida* can vertically transfer its associated bacterial core. *Frontiers in Microbiology*, *8*(FEB), 1–12. <https://doi.org/10.3389/fmicb.2017.00176>

- Lema, K. A., Willis, B. L., & Bourne, D. G. (2012). Corals form characteristic associations with symbiotic nitrogen-fixing bacteria. *Applied and Environmental Microbiology*, 78(9), 3136–3144. <https://doi.org/10.1128/AEM.07800-11>
- Lewin, G. R., Carlos, C., Chevrette, M. G., Horn, H. A., McDonald, B. R., Stankey, R. J., Fox, B. G., & Currie, C. R. (2016). Evolution and Ecology of Actinobacteria and Their Bioenergy Applications. *Annual Review of Microbiology*, 70, 235–254. <https://doi.org/10.1146/annurev-micro-102215-095748>
- Ley, R. E., Hamady, M., Lozupone, C., Turnbaugh, P. J., Ramey, R. R., Bircher, J. S., Schlegel, M. L., Tucker, T. A., Schrenzel, M. D., Knight, R., & Gordon, J. I. (2008). Evolution of mammals and their gut microbes. *Science*, 320(5883), 1647–1651. <https://doi.org/10.1126/science.1155725>
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094–3100.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079.
- Li, J., Chen, Q., Long, L. J., Dong, J. De, Yang, J., & Zhang, S. (2014). Bacterial dynamics within the mucus, tissue and skeleton of the coral *Porites lutea* during different seasons. *Scientific Reports*, 4, 1–8. <https://doi.org/10.1038/srep07320>
- Lilburn, T. G., Kim, K. S., Ostrom, N. E., Byzek, K. R., Leadbetter, J. R., & Breznak, J. A. (2001). Nitrogen fixation by symbiotic and free-living spirochetes. *Science*, 292(5526), 2495–2498. <https://doi.org/10.1126/science.1060281>

- Lim, S. J., & Bordenstein, S. R. (2020). An introduction to phylosymbiosis. *Proceedings of the Royal Society B*, 287(1922), 20192900.
- Lim, S. J., Davis, B. G., Gill, D. E., Walton, J., Nachman, E., Engel, A. S., Anderson, L. C., & Campbell, B. J. (2019). Taxonomic and functional heterogeneity of the gill microbiome in a symbiotic coastal mangrove lucinid species. *ISME Journal*, 13(4), 902–920. <https://doi.org/10.1038/s41396-018-0318-3>
- Lima, L. F. O., Weissman, M., Reed, M., Papudeshi, B., Alker, A. T., Morris, M. M., Edwards, R. A., de Putron, S. J., Vaidya, N. K., & Dinsdale, E. A. (2020). Modeling of the coral microbiome: the influence of temperature and microbial network. *MBio*, 11(2), e02691-19.
- Littman, R. A., Willis, B. L., Pfeffer, C., & Bourne, D. G. (2009). Diversities of coral-associated bacteria differ with location, but not species, for three acroporid corals on the Great Barrier Reef. *FEMS Microbiology Ecology*, 68(2), 152–163. <https://doi.org/10.1111/j.1574-6941.2009.00666.x>
- Luis, A., Baslé, A., Byrne, D., Wright, G., London, J., Chunsheng, J., Karlsson, N., Hansson, G., Eysers, P., & Czjzek, M. (2021). *Sulfated host glycan recognition by carbohydrate sulfatases of the human gut microbiota*.
- Luter, H. M., Gibb, K., & Webster, N. S. (2014). Eutrophication has no short-term effect on the *Cymbastela stipitata* holobiont. *Frontiers in Microbiology*, 5, 216.
- Luter, H. M., Whalan, S., & Webster, N. S. (2012). Thermal and sedimentation stress are unlikely causes of brown spot syndrome in the coral reef sponge, *Ianthella basta*. In *PLoS ONE* (Vol. 7, p. e39779).

- Maldonado, M. (2007). Intergenerational transmission of symbiotic bacteria in oviparous and viviparous demosponges, with emphasis on intracytoplasmically-compartmented bacterial types. *Journal of the Marine Biological Association of the United Kingdom*, 87(6), 1701–1713. <https://doi.org/10.1017/S0025315407058080>
- Maldonado, M., Zhang, X., Cao, X., Xue, L., Cao, H., & Zhang, W. (2010). Selective feeding by sponges on pathogenic microbes: a reassessment of potential for abatement of microbial pollution. *Marine Ecology Progress Series*, 403, 75–89.
- Mallott, E. K., & Amato, K. R. (2021). Host specificity of the gut microbiome. *Nature Reviews Microbiology*, 19(10), 639–653. <https://doi.org/10.1038/s41579-021-00562-3>
- Martinez, A., Tyson, G. W., & DeLong, E. F. (2010). Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional screening and metagenomic analyses. *Environmental Microbiology*, 12(1), 222–238.
- 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. *MicrobiologyOpen*, 6(2), 1–13. <https://doi.org/10.1002/mbo3.417>
- Matsen, F. A., Kodner, R. B., & Armbrust, E. V. (2010). pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, 11(1), 1–16.
- Mazel, F., Davis, K. M., Loudon, A., Kwong, W. K., Groussin, M., & Parfrey, L. W. (2018). Is Host Filtering the Main Driver of Phylosymbiosis across the Tree of Life? *MSystems*, 3(5), e00097-18. <https://doi.org/10.1128/msystems.00097-18>

- McCutcheon, J. P., & Moran, N. A. (2007). Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(49), 19392–19397. <https://doi.org/10.1073/pnas.0708855104>
- McDonald, M. K., Fritz, J. A., Jia, D., Scheuchner, D., Snyder, F. F., Stanislaus, A., Curle, J., Li, L., Stabler, S. P., Allen, R. H., Mains, P. E., & Gravel, R. A. (2017). Identification of ABC transporters acting in vitamin B12 metabolism in *Caenorhabditis elegans*. *Molecular Genetics and Metabolism*, *122*(4), 160–171. <https://doi.org/10.1016/j.ymgme.2017.11.002>
- McFadden, C. S., Sánchez, J. A., & France, S. C. (2010). Molecular phylogenetic insights into the evolution of octocorallia: A review. *Integrative and Comparative Biology*, *50*(3), 389–410. <https://doi.org/10.1093/icb/icq056>
- McFall-Ngai, M. (2008). Hawaiian bobtail squid. *Current Biology*, *18*(22), 1043–1044. <https://doi.org/10.1016/j.cub.2008.08.059>
- McFall-Ngai, M. (2014). Divining the Essence of Symbiosis: Insights from the Squid-Vibrio Model. *PLoS Biology*, *12*(2), 1–6. <https://doi.org/10.1371/journal.pbio.1001783>
- McFall-Ngai, M., Hadfield, M. G., Bosch, T. C. G., Carey, H. V., Domazet-Lošo, T., Douglas, A. E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S. F., Hentschel, U., King, N., Kjelleberg, S., Knoll, A. H., Kremer, N., Mazmanian, S. K., Metcalf, J. L., Neilson, K., Pierce, N. E., ... Wernegreen, J. J. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences*, *110*(9), 3229–3236. <https://doi.org/10.1073/pnas.1218525110>

- McKnight, D. T., Huerlimann, R., Bower, D. S., Schwarzkopf, L., Alford, R. A., & Zenger, K. R. (2019). Methods for normalizing microbiome data: An ecological perspective. *Methods in Ecology and Evolution*, *10*(3), 389–400. <https://doi.org/10.1111/2041-210X.13115>
- McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, *8*(4). <https://doi.org/10.1371/journal.pone.0061217>
- Mews, L. K. (1980). The Green Hydra Symbiosis . III . The Biotrophic transport of Carbohydrate from Alga to Animal. *Proceedings of the Royal Society of London*, *209*(1176), 377–401.
- Meyer, J. L., Paul, V. J., & Teplitski, M. (2014). Community Shifts in the Surface Microbiomes of the Coral *Porites astreoides* with Unusual Lesions. *PLoS ONE*, *9*(6), e100316. <https://doi.org/10.1371/journal.pone.0100316>
- Moeller, A. H., Caro-Quintero, A., Mjungu, D., Georgiev, A. V, Lonsdorf, E. V, Muller, M. N., Pusey, A. E., Peeters, M., Hahn, B. H., & Ochman, H. (2016). Cospeciation of gut microbiota with hominids. *Science (New York, N.Y.)*, *353*(6297), 380–382. <https://doi.org/10.1126/science.aaf3951>
- Moitinho-Silva, L., Díez-Vives, C., Batani, G., Esteves, A. I. S., Jahn, M. T., & Thomas, T. (2017). Integrated metabolism in sponge-microbe symbiosis revealed by genome-centered metatranscriptomics. *ISME Journal*, *11*(7), 1651–1666. <https://doi.org/10.1038/ismej.2017.25>
- Moitinho-Silva, L., Seridi, L., Ryu, T., Voolstra, C. R., Ravasi, T., & Hentschel, U. (2014). Revealing microbial functional activities in the Red Sea sponge *Stylissa carteri* by

- metatranscriptomics. *Environmental Microbiology*, 16(12), 3683–3698.
<https://doi.org/10.1111/1462-2920.12533>
- Monnin, D., Jackson, R., Kiers, E. T., Bunker, M., Ellers, J., & Henry, L. M. (2020). Parallel Evolution in the Integration of a Co-obligate Aphid Symbiosis. *Current Biology*, 30(10), 1949–1957.e6. <https://doi.org/10.1016/j.cub.2020.03.011>
- Moran, N. A. (2006). Symbiosis. *Current Biology*, 16(20), 866–871.
- Moran, N. A., & Baumann, P. (2000). Bacterial endosymbionts in animals. *Current Opinion in Microbiology*, 3(3), 270–275. [https://doi.org/10.1016/S1369-5274\(00\)00088-6](https://doi.org/10.1016/S1369-5274(00)00088-6)
- Moran, N. A., & Sloan, D. B. (2015). The Hologenome Concept: Helpful or Hollow? *PLoS Biology*, 13(12), 1–10. <https://doi.org/10.1371/journal.pbio.1002311>
- Moreno-Pino, M., Cristi, A., Gillooly, J. F., & Trefault, N. (2020). Characterizing the microbiomes of Antarctic sponges: a functional metagenomic approach. *Scientific Reports*, 10(1), 1–12. <https://doi.org/10.1038/s41598-020-57464-2>
- Morley, S. A., Berman, J., Barnes, D. K. A., Carbonell, C. de J., Downey, R. V., & Peck, L. S. (2016). Extreme phenotypic plasticity in metabolic physiology of Antarctic demosponges. *Frontiers in Ecology and Evolution*, 3(JAN), 1–10. <https://doi.org/10.3389/fevo.2015.00157>
- Morrow, K. M., Bourne, D. G., Humphrey, C., Botté, E. S., Laffy, P., Zaneveld, J., Uthicke, S., Fabricius, K. E., & Webster, N. S. (2014). Natural volcanic CO₂seeps reveal future trajectories for host-microbial associations in corals and sponges. *ISME Journal*, 9, 1–15. <https://doi.org/10.1038/ismej.2014.188>

- Neave, M. J., Apprill, A., Ferrier-Pagès, C., & Voolstra, C. R. (2016). Diversity and function of prevalent symbiotic marine bacteria in the genus *Endozoicomonas*. *Applied Microbiology and Biotechnology*, *100*(19), 8315–8324. <https://doi.org/10.1007/s00253-016-7777-0>
- Neave, M. J., Michell, C. T., Apprill, A., & Voolstra, C. R. (2017). *Endozoicomonas* genomes reveal functional adaptation and plasticity in bacterial strains symbiotically associated with diverse marine hosts. *Scientific Reports*, *7*, 40579. <https://doi.org/10.1038/srep40579>
- Neave, M. J., Rachmawati, R., Xun, L., Michell, C. T., Bourne, D. G., Apprill, A., & Voolstra, C. R. (2017). Differential specificity between closely related corals and abundant *Endozoicomonas* endosymbionts across global scales. *The ISME Journal*, *11*(1), 186–200. <https://doi.org/10.1038/ismej.2016.95>
- Nguyen, M. T. H. D., Liu, M., & Thomas, T. (2014). Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. *Molecular Ecology*, *23*(6), 1635–1645. <https://doi.org/10.1111/mec.12384>
- Nishiguchi, M. K., Ruby, E. G., & McFall-Ngai, M. J. (1998). Competitive dominance among strains of luminous bacteria provides an unusual form of evidence for parallel evolution in sepiolid squid-vibrio symbioses. *Applied and Environmental Microbiology*, *64*(9), 3209–3213.
- Nivaskumar, M., & Francetic, O. (2014). Type II secretion system: a magic beanstalk or a protein escalator. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, *1843*(8), 1568–1577.
- Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Research*, *27*(5), 824–834.

- O'Brien, P. A., Andreakis, N., Tan, S., Miller, D. J., Webster, N. S., Zhang, G., & Bourne, D. G. (2021). Testing cophylogeny between coral reef invertebrates and their bacterial and archaeal symbionts. *Molecular Ecology*, *30*(15), 3768–3782. <https://doi.org/10.1111/mec.16006>
- O'Brien, P. A., Smith, H. A., Fallon, S., Fabricius, K., Willis, B. L., Morrow, K. M., & Bourne, D. G. (2018). Elevated CO₂ Has Little Influence on the Bacterial Communities Associated With the pH-Tolerant Coral, Massive *Porites* spp. *Frontiers in Microbiology*, *9*, 1–12. <https://doi.org/10.3389/fmicb.2018.02621>
- O'Brien, P. A., Tan, S., Yang, C., Frade, P. R., Andreakis, N., Smith, H. A., Miller, D. J., Webster, N. S., Zhang, G., & Bourne, D. G. (2020). Diverse coral reef invertebrates exhibit patterns of phyllosymbiosis. *ISME Journal*, *14*(9), 2211–2222. <https://doi.org/10.1038/s41396-020-0671-x>
- O'Brien, P. A., Webster, N. S., Miller, D. J., & Bourne, D. G. (2019). Host-Microbe Coevolution: Applying Evidence from Model Systems to Complex Marine Invertebrate Holobionts. *MBio*, *10*(1), 1–14. <https://doi.org/10.1128/mBio.02241-18>
- Ochman, H., & Moran, N. A. (2001). Genes Lost and Genes Found: Evolution of Bacterial Pathogenesis and Symbiosis. *Science*, *292*(5519), 1096–1099. <https://doi.org/10.1126/science.1058543>
- Ochman, H., Worobey, M., Kuo, C. H., Ndjango, J. B. N., Peeters, M., Hahn, B. H., & Hugenholtz, P. (2010a). Evolutionary relationships of wild hominids recapitulated by gut microbial communities. *PLoS Biology*, *8*(11), 3–10. <https://doi.org/10.1371/journal.pbio.1000546>

- Ochman, H., Worobey, M., Kuo, C. H., Ndjango, J. B. N., Peeters, M., Hahn, B. H., & Hugenholtz, P. (2010b). Evolutionary relationships of wild hominids recapitulated by gut microbial communities. *PLoS Biology*, 8(11), 3–10. <https://doi.org/10.1371/journal.pbio.1000546>
- Oksanen, J., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2019). *vegan: Community Ecology Package. R package version 2.5-4*.
- Oliveira, P. H., Touchon, M., & Rocha, E. P. C. (2014). The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Research*, 42(16), 10618–10631. <https://doi.org/10.1093/nar/gku734>
- Oliver, K. M., Degnan, P. H., Hunter, M. S., & Moran, N. A. (2009). Bacteriophages encode factors required for protection in a symbiotic mutualism. *Science*, 325(5943), 992–994. <https://doi.org/10.1126/science.1174463>
- Pankov, R., & Yamada, K. M. (2002). Fibronectin at a glance. *Journal of Cell Science*, 115(20), 3861–3863.
- Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, 18(5), 1403–1414. <https://doi.org/10.1111/1462-2920.13023>
- Paradis, E., & Schliep, K. (2018). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528.

- 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.
- Pascelli, C., Laffy, P. W., Botté, E., Kupresanin, M., Rattei, T., Lurgi, M., Ravasi, T., & Webster, N. S. (2020). Viral ecogenomics across the Porifera. *Microbiome*, 8(1), 1–22.
- Paterson, S., Vogwill, T., Buckling, A., Benmayor, R., Spiers, A. J., Thomson, N. R., Quail, M., Smith, F., Walker, D., Libberton, B., Fenton, A., Hall, N., & Brockhurst, M. A. (2010). Antagonistic coevolution accelerates molecular evolution. *Nature*, 464(7286), 275–278. <https://doi.org/10.1038/nature08798>
- Pedersen, T. L. (2018). Tidygraph: a tidy API for graph manipulation. *R Package Version*, 1(0).
- Pedersen, T. L. (2020). ggraph: An implementation of grammar of graphics for graphs and networks. *R Package Version 2.0.3*, 1.
- Peskin, A. V, Labas, Y. A., & Tikhonov, A. N. (1998). Superoxide radical production by sponges *Sycon* sp. *FEBS Letters*, 434(1–2), 201–204.
- Piredda, R., Tomasino, M. P., D’Erchia, A. M., Manzari, C., Pesole, G., Montresor, M., Kooistra, W. H. C. F., Sarno, D., & Zingone, A. (2017). Diversity and temporal patterns of planktonic protist assemblages at a Mediterranean Long Term Ecological Research site. *FEMS Microbiology Ecology*, 93(1), fiw200. <https://doi.org/https://doi.org/10.1093/femsec/fiw200>
- Pollock, F. J., Lamb, J. B., Van De Water, J. A. J. M., Smith, H. A., Schaffelke, B., Willis, B. L., & Bourne, D. G. (2019). Reduced diversity and stability of coral-associated bacterial

- communities and suppressed immune function precedes disease onset in corals. *Royal Society Open Science*, 6(6), 190355. <https://doi.org/10.1098/rsos.190355>
- Pollock, F. J., McMinds, R., Smith, S., Bourne, D. G., Willis, B. L., Medina, M., Thurber, R. V., & Zaneveld, J. R. (2018). Coral-associated bacteria demonstrate phylosymbiosis and cophylogeny. *Nature Communications*, 9(4921), 1–13. <https://doi.org/10.1038/s41467-018-07275-x>
- Pöppe, J., Sutcliffe, P., Hooper, J. N. A., Wörheide, G., & Erpenbeck, D. (2010). CO I barcoding reveals new clades and radiation patterns of indo-pacific sponges of the family irciniidae (Demospongiae: Dictyoceratida). *PLoS ONE*, 5(4), 1–6. <https://doi.org/10.1371/journal.pone.0009950>
- Price, C. T. D., Al-Khodor, S., Al-Quadani, T., & Abu Kwaik, Y. (2010). Indispensable role for the eukaryotic-like ankyrin domains of the ankyrin B effector of *Legionella pneumophila* within macrophages and amoebae. *Infection and Immunity*, 78(5), 2079–2088.
- Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2—approximately maximum-likelihood trees for large alignments. *PloS One*, 5(3), e9490.
- Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., ... Zoetendal, E. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285), 59–65. <https://doi.org/10.1038/nature08821>
- Quattrini, A. M., Wu, T., Soong, K., Jeng, M., Benayahu, Y., & Mcfadden, C. S. (2019). A next generation approach to species delimitation reveals the role of hybridization in a cryptic species complex of corals. *BMC Evolutionary Biology*, 19(116), 1–19.

- Radax, R., Rattei, T., Lanzen, A., Bayer, C., Rapp, H. T., Urich, T., & Schleper, C. (2012). Metatranscriptomics of the marine sponge *Geodia barretti*: Tackling phylogeny and function of its microbial community. *Environmental Microbiology*, *14*(5), 1308–1324. <https://doi.org/10.1111/j.1462-2920.2012.02714.x>
- Rädecker, N., Pogoreutz, C., Voolstra, C. R., Wiedenmann, J., & Wild, C. (2015). Nitrogen cycling in corals: The key to understanding holobiont functioning? *Trends in Microbiology*, *23*(8), 490–497. <https://doi.org/10.1016/j.tim.2015.03.008>
- Raina, J. B., Tapiolas, D., Willis, B. L., & Bourne, D. G. (2009). Coral-associated bacteria and their role in the biogeochemical cycling of sulfur. *Applied and Environmental Microbiology*, *75*(11), 3492–3501. <https://doi.org/10.1128/AEM.02567-08>
- Rankin, D. J., Rocha, E. P. C., & Brown, S. P. (2011). What traits are carried on mobile genetic elements, and why. *Heredity*, *106*(1), 1–10. <https://doi.org/10.1038/hdy.2010.24>
- Reynolds, D., & Thomas, T. (2016). Evolution and function of eukaryotic-like proteins from sponge symbionts. *Molecular Ecology*, *25*(20), 5242–5253. <https://doi.org/10.1111/mec.13812>
- Rivett, D. W., & Bell, T. (2018). Abundance determines the functional role of bacterial phylotypes in complex communities. *Nature Microbiology*. <https://doi.org/10.1038/s41564-018-0180-0>
- Rix, L., De Goeij, J. M., Mueller, C. E., Struck, U., Middelburg, J. J., Van Duyl, F. C., Al-Horani, F. A., Wild, C., Naumann, M. S., & Van Oevelen, D. (2016). Coral mucus fuels the sponge loop in warm-and cold-water coral reef ecosystems. *Scientific Reports*, *6*(1), 1–11.

- Rix, L., Ribes, M., Coma, R., Jahn, M. T., de Goeij, J. M., van Oevelen, D., Escrig, S., Meibom, A., & Hentschel, U. (2020). Heterotrophy in the earliest gut: a single-cell view of heterotrophic carbon and nitrogen assimilation in sponge-microbe symbioses. *ISME Journal*. <https://doi.org/10.1038/s41396-020-0706-3>
- Robbins, S. J., Singleton, C. M., Chan, C. X., Messer, L. F., Geers, A. U., Ying, H., Baker, A., Bell, S. C., Morrow, K. M., Ragan, M. A., Miller, D. J., Forêt, S., Ball, E., Beeden, R., Berumen, M., Aranda, M., Ravasi, T., Bongaerts, P., Hoegh-Guldberg, O., ... Tyson, G. W. (2019). A genomic view of the reef-building coral *Porites lutea* and its microbial symbionts. *Nature Microbiology*, *4*(12), 2090–2100. <https://doi.org/10.1038/s41564-019-0532-4>
- Robbins, S. J., Song, W., Engelberts, J. P., Glasl, B., Slaby, B. M., Boyd, J., Marangon, E., Botté, E. S., Laffy, P., Thomas, T., & Webster, N. S. (2021). A genomic view of the microbiome of coral reef demosponges. *ISME Journal*. <https://doi.org/10.1038/s41396-020-00876-9>
- Ronquist, F., Teslenko, M., Mark, P. Van Der, Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*(3), 539–542.
- Ross, A. A., Müller, K. M., Weese, J. S., & Neufeld, J. D. (2018). Comprehensive skin microbiome analysis reveals the uniqueness of human skin and evidence for phyllosymbiosis within the class Mammalia. *Proceedings of the National Academy of Sciences*, 1–10. <https://doi.org/10.1073/pnas.1801302115>

- Röthig, T., Costa, R. M., Simona, F., Baumgarten, S., Torres, A. F., Radhakrishnan, A., Aranda, M., & Voolstra, C. R. (2016). Distinct Bacterial Communities Associated with the Coral Model *Aiptasia* in Aposymbiotic and Symbiotic States with Symbiodinium. *Frontiers in Marine Science*, 3(November). <https://doi.org/10.3389/fmars.2016.00234>
- Rowan, R. (1998). Diversity and Ecology of Zooxanthellae on Coral Reefs. *Journal of Phycology*, 34(3), 407–417. <https://doi.org/10.1046/j.1529-8817.1998.340407.x>
- Russell, C. W., Bouvaine, S., Newell, P. D., & Douglas, A. E. (2013). Shared metabolic pathways in a coevolved insect-bacterial symbiosis. *Applied and Environmental Microbiology*, 79(19), 6117–6123. <https://doi.org/10.1128/AEM.01543-13>
- Ryan, M. F., & Byrne, O. (1988). Plant-insect coevolution and inhibition of acetylcholinesterase. *Journal of Chemical Ecology*, 14(10), 1965–1975. <https://doi.org/10.1007/BF01013489>
- Sacristán-Soriano, O., Turon, X., & Hill, M. (2020). *Microbiome structure of ecologically important bioeroding sponges (family Clionaidae): the role of host phylogeny and environmental plasticity*.
- Sato, Y., Ling, E. Y. S., Turaev, D., Laffy, P., Weynberg, K. D., Rattei, T., Willis, B. L., & Bourne, D. G. (2017). Unraveling the microbial processes of black band disease in corals through integrated genomics. *Scientific Reports*, 7(September 2016), 40455. <https://doi.org/10.1038/srep40455>
- Schliep, K. P. (2011). phangorn: phylogenetic analysis in R. *Bioinformatics*, 27(4), 592–593.
- Schmitt, S., Angermeier, H., Schiller, R., Lindquist, N., & Hentschel, U. (2008). Molecular microbial diversity survey of sponge reproductive stages and mechanistic insights into

- vertical transmission of microbial symbionts. *Applied and Environmental Microbiology*, 74(24), 7694–7708.
- Schofield, C. J., & McDonough, M. A. (2007). Structural and mechanistic studies on the peroxisomal oxygenase phytanoyl-CoA 2-hydroxylase (PhyH). *Biochemical Society Transactions*, 35(5), 870–875.
- Schöttner, S., Hoffmann, F., Cárdenas, P., Rapp, H. T., Boetius, A., & Ramette, A. (2013). Relationships between Host Phylogeny, Host Type and Bacterial Community Diversity in Cold-Water Coral Reef Sponges. *PLoS ONE*, 8(2), 1–11. <https://doi.org/10.1371/journal.pone.0055505>
- Schwarz-Linek, U., Höök, M., & Potts, J. R. (2004). The molecular basis of fibronectin-mediated bacterial adherence to host cells. *Molecular Microbiology*, 52(3), 631–641.
- Shade, A., & Handelsman, J. (2012). Beyond the Venn diagram: The hunt for a core microbiome. *Environmental Microbiology*, 14(1), 4–12. <https://doi.org/10.1111/j.1462-2920.2011.02585.x>
- Sharp, K. H., Distel, D., & Paul, V. J. (2012). Diversity and dynamics of bacterial communities in early life stages of the Caribbean coral *Porites astreoides*. *The ISME Journal*, 6(4), 790–801. <https://doi.org/10.1038/ismej.2011.144>
- Sharp, K. H., Eam, B., John Faulkner, D., & Haygood, M. G. (2007). Vertical transmission of diverse microbes in the tropical sponge *Corticium* sp. *Applied and Environmental Microbiology*, 73(2), 622–629. <https://doi.org/10.1128/AEM.01493-06>

- Sharp, K. H., Ritchie, K. B., Schupp, P. J., Ritson-Williams, R., & Paul, V. J. (2010). Bacterial acquisition in juveniles of several broadcast spawning coral species. *PLoS ONE*, 5(5), 1–6. <https://doi.org/10.1371/journal.pone.0010898>
- Sieber, C. M. K., Probst, A. J., Sharrar, A., Thomas, B. C., Hess, M., Tringe, S. G., & Banfield, J. F. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology*, 3(7), 836–843.
- Sieber, M., Pita, L., Weiland-Bräuer, N., Dirksen, P., Wang, J., Mortzfeld, B., Franzenburg, S., Schmitz, R. A., Baines, J. F., Fraune, S., Hentschel, U., Schulenburg, H., Bosch, T. C. G., & Traulsen, A. (2018). The Neutral Metaorganism. *BioRxiv*, 367243. <https://doi.org/10.1101/367243>
- 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. *ISME Journal*, 11(11), 2465–2478. <https://doi.org/10.1038/ismej.2017.101>
- Sloan, W. T., Lunn, M., Woodcock, S., Head, I. M., Nee, S., & Curtis, T. P. (2006). Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environmental Microbiology*, 8(4), 732–740. <https://doi.org/10.1111/j.1462-2920.2005.00956.x>
- Soda, K. (1987). Microbial sulfur amino acids: an overview. *Methods in Enzymology*, 143, 453–459.
- Solomon, C. M., Collier, J. L., Berg, G. M., & Glibert, P. M. (2010). Role of urea in microbial metabolism in aquatic systems: A biochemical and molecular review. *Aquatic Microbial Ecology*, 59(1), 67–88. <https://doi.org/10.3354/ame01390>

- Song, S. J., Sanders, J. G., Delsuc, F., Metcalf, J., Amato, K., Taylor, M. W., Mazel, F., Lutz, H. L., Winker, K., Graves, G. R., Humphrey, G., Gilbert, J. A., Hackett, S. J., White, K. P., Skeen, H. R., Kurtis, S. M., Withrow, J., Braile, T., Miller, M., ... Blanton, J. M. (2020). Comparative Analyses of Vertebrate Gut Microbiomes Reveal Convergence between Birds and Bats. *Mbio*, *11*(1), 1–14.
- Soto-Martin, E. C., Warnke, I., Farquharson, F. M., Christodoulou, M., Horgan, G., Derrien, M., Faurie, J.-M., Flint, H. J., Duncan, S. H., & Louis, P. (2020). Vitamin biosynthesis by human gut butyrate-producing bacteria and cross-feeding in synthetic microbial communities. *MBio*, *11*(4), e00886-20.
- Stat, M., Carter, D., & Hoegh-Guldberg, O. (2006). The evolutionary history of Symbiodinium and scleractinian hosts-Symbiosis, diversity, and the effect of climate change. *Perspectives in Plant Ecology, Evolution and Systematics*, *8*(1), 23–43. <https://doi.org/10.1016/j.ppees.2006.04.001>
- Stewart, F. J., Ottesen, E. A., & DeLong, E. F. (2010). Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *The ISME Journal*, *4*(7), 896–907. <https://doi.org/10.1038/ismej.2010.18>
- Strand, R., Whalan, S., Webster, N. S., Kutti, T., Fang, J. K.-H., Luter, H. M., & Bannister, R. J. (2017). The response of a boreal deep-sea sponge holobiont to acute thermal stress. *Scientific Reports*, *7*(1), 1–12.
- Sunagawa, S., Woodley, C. M., & Medina, M. (2010). Threatened corals provide underexplored microbial habitats. *PLoS ONE*, *5*(3), 1–7. <https://doi.org/10.1371/journal.pone.0009554>

- Takiya, D. M., Tran, P. L., Dietrich, C. H., & Moran, N. A. (2006). Co-cladogenesis spanning three phyla: Leafhoppers (Insecta: Hemiptera: Cicadellidae) and their dual bacterial symbionts. *Molecular Ecology*, *15*(13), 4175–4191. <https://doi.org/10.1111/j.1365-294X.2006.03071.x>
- Tandon, K., Lu, C. Y., Chiang, P. W., Wada, N., Yang, S. H., Chan, Y. F., Chen, P. Y., Chang, H. Y., Chiou, Y. J., Chou, M. S., Chen, W. M., & Tang, S. L. (2020). Comparative genomics: Dominant coral-bacterium *Endozoicomonas acroporae* metabolizes dimethylsulfoniopropionate (DMSP). *ISME Journal*, *14*(5), 1290–1303. <https://doi.org/10.1038/s41396-020-0610-x>
- Team, R. C. (2018). *R: A language and environment for statistical computing* (3.5.0). R Foundation for Statistical Computing.
- Theis, K. R., Dheilly, N. M., Klassen, J. L., Brucker, R. M., Baines, J. F., Bosch, T. C. G., Cryan, J. F., Gilbert, S. F., Goodnight, C. J., Lloyd, E. A., Sapp, J., Vandenkoornhuyse, P., Zilber-Rosenberg, I., Rosenberg, E., & Bordenstein, S. R. (2016). Getting the Hologenome Concept Right: an Eco-Evolutionary Framework for Hosts and Their Microbiomes. *MSystems*, *1*(2), e00028-16. <https://doi.org/10.1128/mSystems.00028-16>
- 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., Victor Lopez, J., ... Webster, N. S. (2016). Diversity, structure and convergent evolution of the global sponge microbiome. *Nature Communications*, *7*(11870), 1–12. <https://doi.org/10.1038/ncomms11870>

- Thomas, T., Rusch, D., DeMaere, M. Z., Yung, P. Y., Lewis, M., Halpern, A., Heidelberg, K. B., Egan, S., Steinberg, P. D., & Kjelleberg, S. (2010). Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *ISME Journal*, 4(12), 1557–1567. <https://doi.org/10.1038/ismej.2010.74>
- Tyrrell, T. (1999). The relative influences of nitrogen and phosphorus on oceanic primary production. *Nature*, 400(6744), 525–531.
- 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(1), 9–13.
- Van De Water, J. A. J. M., Allemand, D., & Ferrier-Pagès, C. (2018). Host-microbe interactions in octocoral holobionts - recent advances and perspectives. *Microbiome*, 6(64), 1–28. <https://doi.org/10.1186/s40168-018-0431-6>
- Van den Abbeele, P., Van de Wiele, T., Verstraete, W., & Possemiers, S. (2011). The host selects mucosal and luminal associations of coevolved gut microorganisms: A novel concept. *FEMS Microbiology Reviews*, 35(4), 681–704. <https://doi.org/https://doi.org/10.1111/j.1574-6976.2011.00270.x>
- van Opstal, E. J., & Bordenstein, S. R. (2019). Phylosymbiosis Impacts Adaptive Traits in *Nasonia* Wasps. *MBio*, 10(4), 1–11. <https://doi.org/10.1128/mbio.00887-19>
- Van Valen, L. (1974). Molecular evolution as predicted by natural selection. *Journal of Molecular Evolution*, 3(2), 89–101. <https://doi.org/10.1007/BF01796554>
- Veech, J. A. (2013). A probabilistic model for analysing species co-occurrence. *Global Ecology and Biogeography*, 22(2), 252–260.

- Vilanova, E., Coutinho, C. C., & Mourao, P. A. S. (2009). Sulfated polysaccharides from marine sponges (Porifera): an ancestor cell–cell adhesion event based on the carbohydrate–carbohydrate interaction. *Glycobiology*, *19*(8), 860–867.
- Volland, J. M., Schintlmeister, A., Zambalos, H., Reipert, S., Mozetič, P., Espada-Hinojosa, S., Turk, V., Wagner, M., & Bright, M. (2018). NanoSIMS and tissue autoradiography reveal symbiont carbon fixation and organic carbon transfer to giant ciliate host. *ISME Journal*, *12*(3), 714–727. <https://doi.org/10.1038/s41396-018-0069-1>
- Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol*, *3*(5), 294–299.
- Wada, N., Ishimochi, M., Matsui, T., Pollock, F. J., Tang, S., Ainsworth, T. D., Willis, B. L., Mano, N., & Bourne, D. G. (2019). Characterization of coral- associated microbial aggregates (CAMAs) within tissues of the coral *Acropora hyacinthus*. *Scientific Reports*, *9*, 1–13. <https://doi.org/10.1038/s41598-019-49651-7>
- Wahab, M. A. A., De Nys, R., Holzman, R., Schneider, C. L., & Whalan, S. (2016). Patterns of reproduction in two co-occurring Great Barrier Reef sponges. *Marine and Freshwater Research*, *68*(7), 1233–1244.
- Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T. H., Stege, J. T., Cayouette, M., McHardy, A. C., Djordjevic, G., Aboushadi, N., Sorek, R., Tringe, S. G., Podar, M., Martin, H. G., Kunin, V., Dalevi, D., Madejska, J., Kirton, E., Platt, D., ... Leadbetter, J. R. (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature*, *450*(7169), 560–565. <https://doi.org/10.1038/nature06269>

- Waterworth, S. C., Kalinski, J. C. J., Madonsela, L. S., Parker-Nance, S., Kwan, J. C., & Dorrington, R. A. (2020). Family matters: The genomes of conserved bacterial symbionts provide insight into specialized metabolic relationships with their sponge host. In *bioRxiv*. <https://doi.org/10.1101/2020.12.09.417808>
- Weber, L., Gonzalez-díaz, P., Armenteros, M., & Apprill, A. (2019). The coral ecosphere : A unique coral reef habitat that fosters coral – microbial interactions. *Limnology and Oceanography*, 1–16. <https://doi.org/10.1002/lno.11190>
- Webster, N. S., & Reusch, T. B. H. (2017). Microbial contributions to the persistence of coral reefs. *ISME Journal*, 11(10), 2167–2174. <https://doi.org/10.1038/ismej.2017.66>
- Webster, N. S., Taylor, M. W., Behnam, F., Lückner, S., Rattei, T., Whalan, S., Horn, M., & Wagner, M. (2010). Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environmental Microbiology*, 12(8), 2070–2082. <https://doi.org/10.1111/j.1462-2920.2009.02065.x>
- Webster, N. S., & Thomas, T. (2016). The sponge hologenome. *MBio*, 7(2), 1–14. <https://doi.org/10.1128/mBio.00135-16>
- Wessels, W., Sprungala, S., Watson, S. A., Miller, D. J., & Bourne, D. G. (2017). The microbiome of the octocoral *Lobophytum pauciflorum*: Minor differences between sexes and resilience to short-term stress. *FEMS Microbiology Ecology*, 93(5), 1–13. <https://doi.org/10.1093/femsec/fix013>
- Weynberg, K. D., Wood-Charlson, E. M., Suttle, C. A., & van Oppen, M. J. H. (2014). Generating viral metagenomes from the coral holobiont. *Frontiers in Microbiology*, 5(MAY), 1–11. <https://doi.org/10.3389/fmicb.2014.00206>

- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wickham, H., François, R., Henry, L., & Müller, K. (2019). *dplyr: A Grammar of Data Manipulation* (R package version 0.8.0.1).
- Wilson, A. C. C., & Duncan, R. P. (2015). Signatures of host/symbiont genome coevolution in insect nutritional endosymbioses. *Proceedings of the National Academy of Sciences*, *112*(33), 10255–10261. <https://doi.org/10.1073/pnas.1423305112>
- Wommack, K. E., & Colwell, R. R. (2000). Virioplankton: Viruses in Aquatic Ecosystems. *Microbiology and Molecular Biology Reviews*, *64*(1), 69–114. <https://doi.org/10.1128/MMBR.64.1.69-114.2000>
- Work, T. M., & Aeby, G. S. (2014). Microbial aggregates within tissues infect a diversity of corals throughout the Indo-Pacific. *Marine Ecology Progress Series*, *500*, 1–9. <https://doi.org/10.3354/meps10698>
- Wu, D., Daugherty, S. C., Van Aken, S. E., Pai, G. H., Watkins, K. L., Khouri, H., Tallon, L. J., Zaborsky, J. M., Dunbar, H. E., Tran, P. L., Moran, N. A., & Eisen, J. A. (2006). Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biology*, *4*(6), 1079–1092. <https://doi.org/10.1002/adsc.200505252>
- Wu, Y.-W., Simmons, B. A., & Singer, S. W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*, *32*(4), 605–607.

- Xia, X., & Xie, Z. (2001). DAMBE: software package for data analysis in molecular biology and evolution. *Journal of Heredity*, 92(4), 371–373. <https://doi.org/https://doi.org/10.1093/jhered/92.4.371>
- Xu, C., & Min, J. (2011). Structure and function of WD40 domain proteins. *Protein and Cell*, 2(3), 202–214. <https://doi.org/10.1007/s13238-011-1018-1>
- Yang, C., Tan, S., Meng, G., Bourne, D. G., O'Brien, P. A., Xu, J., Liao, S., Chen, A., Chen, X., & Liu, S. (2018). Access COI barcode efficiently using high throughput Single-End 400 bp sequencing. *BioRxiv*, 498618. <https://doi.org/10.1101/498618>
- Yeoh, Y. K., Dennis, P. G., Paungfoo-Lonhienne, C., Weber, L., Brackin, R., Ragan, M. A., Schmidt, S., & Hugenholtz, P. (2017). Evolutionary conservation of a core root microbiome across plant phyla along a tropical soil chronosequence. *Nature Communications*, 8(215), 1–9. <https://doi.org/10.1038/s41467-017-00262-8>
- Youngblut, N. D., Reischer, G. H., Walters, W., Schuster, N., Walzer, C., Stalder, G., Ley, R. E., & Farnleitner, A. H. (2019). Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. *Nature Communications*, 10(1), 1–15. <https://doi.org/10.1038/s41467-019-10191-3>
- Yu, G., Smith, D., Zhu, H., Guan, Y., & Tsan-Yuk Lam, T. (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1), 28–36.
- Zaneveld, J., Turnbaugh, P. J., Lozupone, C., Ley, R. E., Hamady, M., Gordon, J. I., & Knight, R. (2008). Host-bacterial coevolution and the search for new drug targets. *Current Opinion in Chemical Biology*, 12(1), 109–114. <https://doi.org/10.1016/j.cbpa.2008.01.015>

- Zhang, S., Song, W., Wemheuer, B., Reveillaud, J., Webster, N., & Thomas, T. (2019). Comparative Genomics Reveals Ecological and Evolutionary Insights into Sponge-Associated Thaumarchaeota. *MSystems*, 4(4), 1–16. <https://doi.org/10.1128/msystems.00288-19>
- Ziegler, M., Grupstra, C. G. B., Barreto, M. M., Eaton, M., BaOmar, J., Zubier, K., Al-Sofyani, A., Turki, A. J., Ormond, R., & Voolstra, C. R. (2019). Coral bacterial community structure responds to environmental change in a host-specific manner. *Nature Communications*, 10(1). <https://doi.org/10.1038/s41467-019-10969-5>
- Ziegler, M., Seneca, F. O., Yum, L. K., Palumbi, S. R., & Voolstra, C. R. (2017). Bacterial community dynamics are linked to patterns of coral heat tolerance. *Nature Communications*, 8, 1–8. <https://doi.org/10.1038/ncomms14213>
- Zierer, M. S., & Mourão, P. A. S. (2000). A wide diversity of sulfated polysaccharides are synthesized by different species of marine sponges. *Carbohydrate Research*, 328(2), 209–216.
- Zilber-Rosenberg, I., & Rosenberg, E. (2008). Role of microorganisms in the evolution of animals and plants: The hologenome theory of evolution. *FEMS Microbiology Reviews*, 32(5), 723–735. <https://doi.org/10.1111/j.1574-6976.2008.00123.x>

Appendix A. Supporting tables and figures for chapter 2

Table S2.1. Metadata for each sample collection. Preservation method refers to snap frozen in liquid nitrogen or preserved in salt-saturated DMSO (see methods). Collection sites Orpheus Island and Pelorus Island have been grouped as one collection site (Palm Islands) due to their close proximity and environmental conditions.

Taxa	Species	Replicates	Preservation	Collection Site	Reef Zone	Date Collected
Coral	<i>Acropora hyacinthus</i>	3	DMSO	Orpheus Island	Inshore	May 2017
Coral	<i>Pocillopora damicornis</i>	3	DMSO	Orpheus Island	Inshore	May 2017
Coral	<i>Seriatopora hystrix</i>	3	DMSO	Orpheus Island	Inshore	May 2017
Coral	<i>Porites cylindrica</i>	3	DMSO	Orpheus Island	Inshore	May 2017
Coral	<i>Diploastrea heliophora</i>	3	DMSO	Orpheus Island	Inshore	May 2017
Coral	massive <i>Porites sp.</i>	5	Frozen	Orpheus Island	Inshore	August 2017
Coral	<i>Pocillopora verrucosa</i>	5	Frozen	Rib Reef	Mid shelf	August 2017
Coral	<i>Stylophora pistillata</i>	5	Frozen	Broadhurst Reef	Mid shelf	August 2017
Coral	<i>Acropora formosa</i>	5	Frozen	Orpheus Island	Inshore	August 2017
Coral	<i>Pavona cactus</i>	5	Frozen	Orpheus Island	Inshore	August 2017
Coral	<i>Echinopora mammiformis</i>	5	Frozen	Orpheus Island	Inshore	August 2017
Coral	<i>Seriatopora hystrix</i>	5	DMSO	Ribbon Reef 10	Offshore	Oct 2017
Coral	<i>Porites cylindrica</i>	5	DMSO	Ribbon Reef 10	Offshore	Oct 2017
Coral	<i>Pachyseris speciosa</i>	3	DMSO	Osprey Reef	Offshore	Oct 2017
Coral	<i>Seriatopora hystrix</i>	3	DMSO	Pelorus Island	Inshore	May 2018
Coral	massive <i>Porites sp.</i>	3	DMSO	Pelorus Island	Inshore	May 2018
Sponge	<i>Coelocarteria singaporensis</i>	3	Frozen	Orpheus Island	Inshore	May 2017
Sponge	<i>Ircinia ramosa</i>	5	Frozen	Davies Reef	Mid shelf	August 2017
Sponge	<i>Carteriospongia foliascens</i>	5	Frozen	Davies Reef	Mid shelf	August 2017
Sponge	<i>Ircinia sp.</i>	5	Frozen	Broadhurst Reef	Mid shelf	August 2017
Sponge	<i>Ircinia ramosa</i>	5	DMSO	Ribbon Reef 10	Offshore	Oct 2017
Sponge	<i>Coscinoderma sp.</i>	5	DMSO	Ribbon Reef 10	Offshore	Oct 2017
Soft coral	<i>Sarcophyton sp.</i>	3	Frozen	Orpheus Island	Inshore	May 2017
Soft coral	<i>Heteroxenia sp.</i>	3	Frozen	Pelorus Island	Inshore	May 2017
Soft coral	<i>Briareum sp.</i>	5	Frozen	Pandora Reef	Inshore	August 2017
Soft coral	<i>Briareum sp. 2</i>	5	Frozen	Pandora Reef	Inshore	August 2017
Soft coral	<i>Clavularia sp.</i>	5	Frozen	Pandora Reef	Inshore	August 2017
Soft coral	<i>Sinularia sp.</i>	5	Frozen	Orpheus Island	Inshore	August 2017
Soft coral	<i>Sinularia sp. 2</i>	5	Frozen	Orpheus Island	Inshore	August 2017
Soft coral	<i>Cladiella sp.</i>	5	Frozen	Rib Reef	Mid shelf	August 2017
Soft coral	<i>Sarcophyton sp.</i>	5	DMSO	Ribbon Reef 10	Offshore	Oct 2017
Soft coral	<i>Sinularia sp.</i>	5	DMSO	Ribbon Reef 10	Offshore	Oct 2017
Gorgonian	<i>Pinnigorgia sp.</i>	5	Frozen	Rib Reef	Mid shelf	August 2017
Gorgonian	<i>Isis hippuris</i>	5	Frozen	Rib Reef	Mid shelf	August 2017
Ascidian	<i>Didemnum molle</i>	3	DMSO	Pelorus Island	Inshore	May 2017
Ascidian	<i>Lissoclinum patella</i>	5	Frozen	Davies Reef	Mid shelf	August 2017
Ascidian	<i>Polycarpa aurata</i>	5	Frozen	Broadhurst Reef	Mid shelf	August 2017
Control	Seawater	1	Frozen	Pandora Reef	Inshore	August 2017
Control	Seawater	1	Frozen	Orpheus Island	Inshore	August 2017
Control	Seawater	1	Frozen	Rib Reef	Mid shelf	August 2017
Control	Seawater	1	Frozen	Davies Reef	Mid shelf	August 2017
Control	Seawater	1	Frozen	Broadhurst Reef	Mid shelf	August 2017
Control	Seawater	1	Frozen	Orpheus Island	Inshore	August 2017
Control	Seawater	1	Frozen	Orpheus Island	Inshore	August 2017
Control	Seawater	1	Frozen	Orpheus Island	Inshore	August 2017
Control	Positive control (MiSeq)	2	NA	NA	NA	NA
Control	Blank	2	NA	NA	NA	NA
Total		173				

Table S2.2. AIC and BIC selection results for evolutionary models. Model used was selected as the closest model available in the software Mr Bayes. Substitution rates for octocoral were set to ‘even’ to resolve polytomies. All models used were within the 100% confidence interval for the best performing models.

		AIC	BIC	Model Used
<i>Ascidan</i>	18S	TrN+G	HKY+G	HKY+G
	COI	TIM1+I	TIM1+I	GTR+I
<i>Coral</i>	18S	HKY+G	K80+G	HKY+G
	COI	HKY+G	HKY+G	HKY+G
	ITS	HKY+I+G	K80+G	HKY+G
<i>Octocoral</i>	18S	TrN+I+G	K80+G	HKY
	COI	TPM1uf+I+G	HKY+I+G	HKY
	ITS	SYM+G	K80+G	HKY
<i>Sponge</i>	18S	TrN+G	K80+G	HKY+G
	COI	TPM3uf+G	HKY+G	HKY+G
	ITS	TPM3uf+G	K80	HKY+G

Table S2.3. Similarity percentages (SIMPER) analysis between host species that showed incongruences between host phylogeny and microbial dissimilarity. Average refers to the amount of microbial dissimilarity attributed to that ASV while cumulative is the cumulative sum of the top 5 contributing ASVs. Microbial family depicted as ‘f_’ and genus as ‘g_’.

Comparison: Porities cylindrica (PI) & Porities cylindrica (RR)

Taxonomy	average	sd	cumulative
f_Endozoicomonadaceae;g_Endozoicomonas	0.21923	0.06339	0.22481425
f_Endozoicomonadaceae;g_Endozoicomonas	0.04636	0.01547	0.27235939
f_Rhodobacteraceae;g_HIMB11	0.02192	0.03246	0.29483691
f_Chlorobiaceae;g_Prosthecochloris	0.01784	0.03671	0.3131313
f_Chlorobiaceae;g_Prosthecochloris	0.01205	0.01718	0.3254911

Comparison: Seriatopora hystrix (PI) & Seriatopora hystrix (RR)

Taxonomy	average	sd	cumulative
f_Anaplasmataceae;g_Neorickettsia	0.074	0.111	0.080
f_Endozoicomonadaceae;g_Endozoicomonas	0.041	0.050	0.125
f_Cyanobiaceae;g_Synechococcus CC9902	0.038	0.032	0.166
f_Francisellaceae;g_Francisella	0.030	0.060	0.199
Unassigned	0.023	0.049	0.224

Comparison: Sinularia sp. PI & Sinularia sp. RR

Taxonomy	average	sd	cumulative
f_Endozoicomonadaceae;g_Endozoicomonas	0.351	0.120	0.353
Unknown Bacteria	0.096	0.122	0.449
f_Endozoicomonadaceae;g_Endozoicomonas	0.074	0.122	0.523
f_Fusobacteriaceae;g_Cetobacterium	0.053	0.081	0.576
f_Endozoicomonadaceae;g_Endozoicomonas	0.050	0.063	0.626

Comparison: Isis hippuis & Pinnigorgia sp.

Taxonomy	average	sd	cumulative
f_Endozoicomonadaceae;g_Endozoicomonas	0.254	0.132	0.256
f_Endozoicomonadaceae;g_Endozoicomonas	0.070	0.084	0.326
f_Endozoicomonadaceae;g_Endozoicomonas	0.056	0.060	0.382
f_Entomoplasmatales Incertae Sedis;g_Candidatus Hepatoplasma	0.055	0.113	0.437
f_Entomoplasmatales Incertae Sedis;g_Candidatus Hepatoplasma	0.049	0.073	0.486

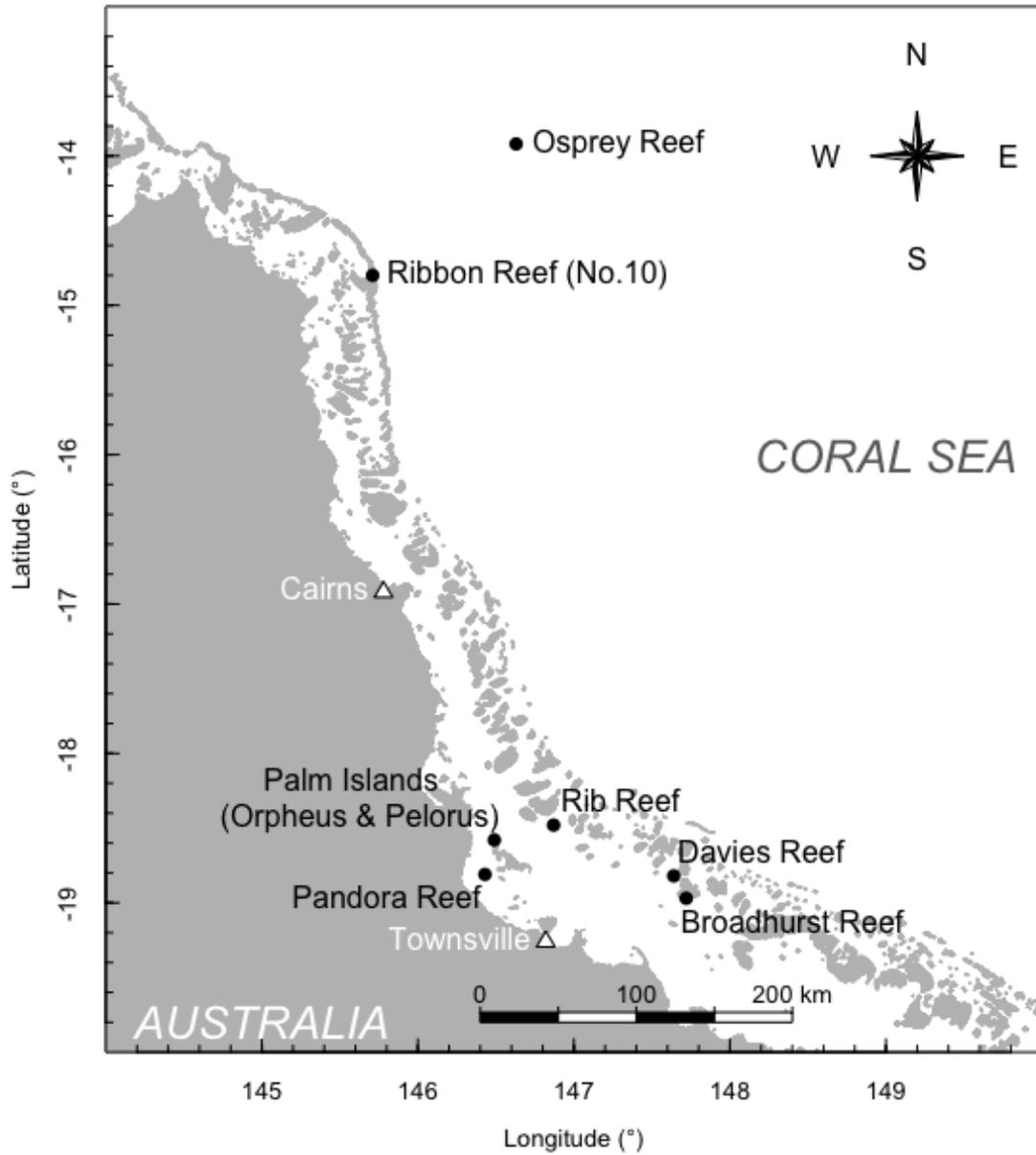


Figure S2.1. Sample collection sites across the central and northern sectors of the Great Barrier Reef. Collection sites marked with a black circle while nearby landmarks are indicated with a white triangle. Orpheus and Pelorus Islands are grouped as one site in the Palm Islands

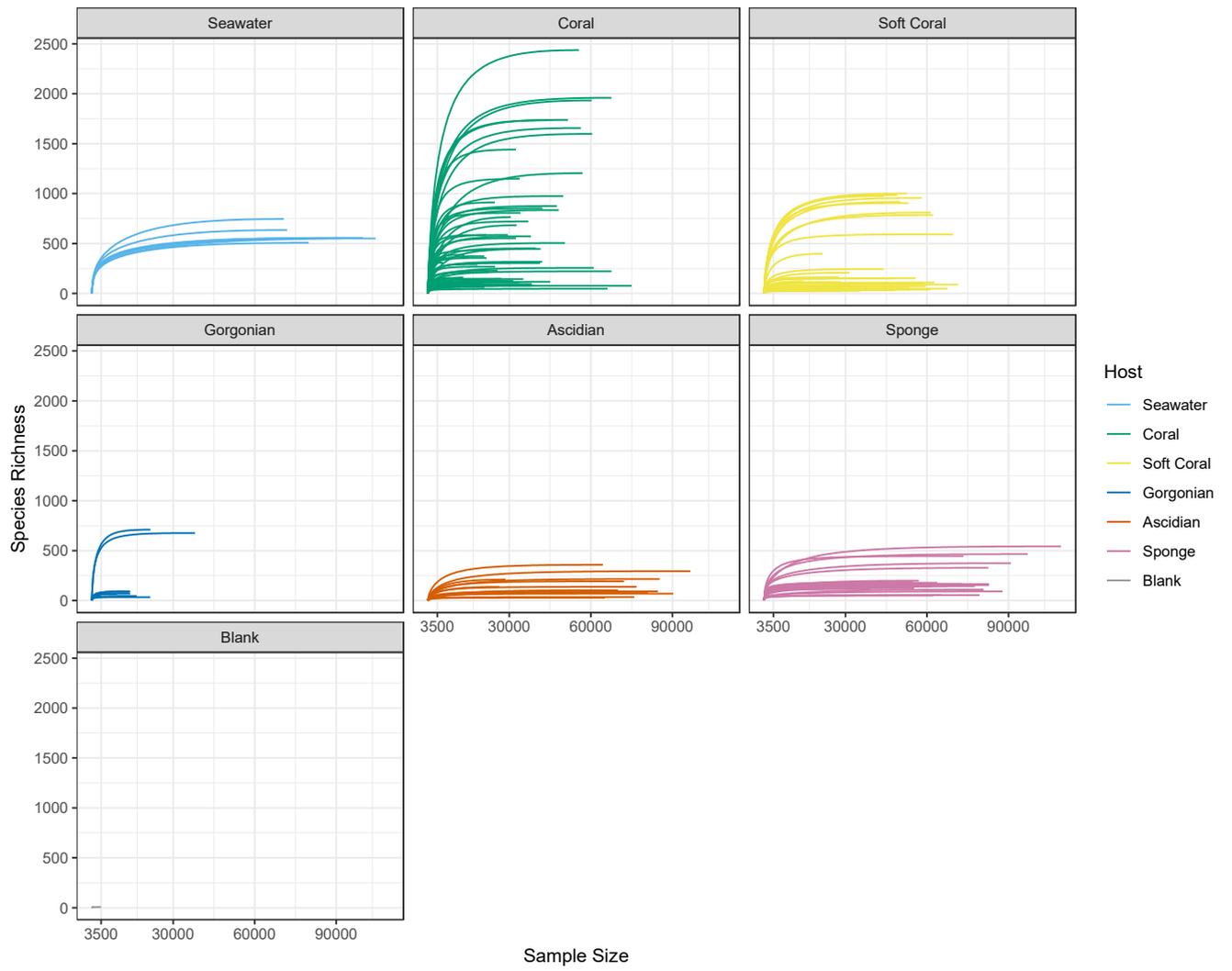


Figure S2.2. Rarefaction analysis of each sample within each invertebrate group as well as seawater and blank samples.

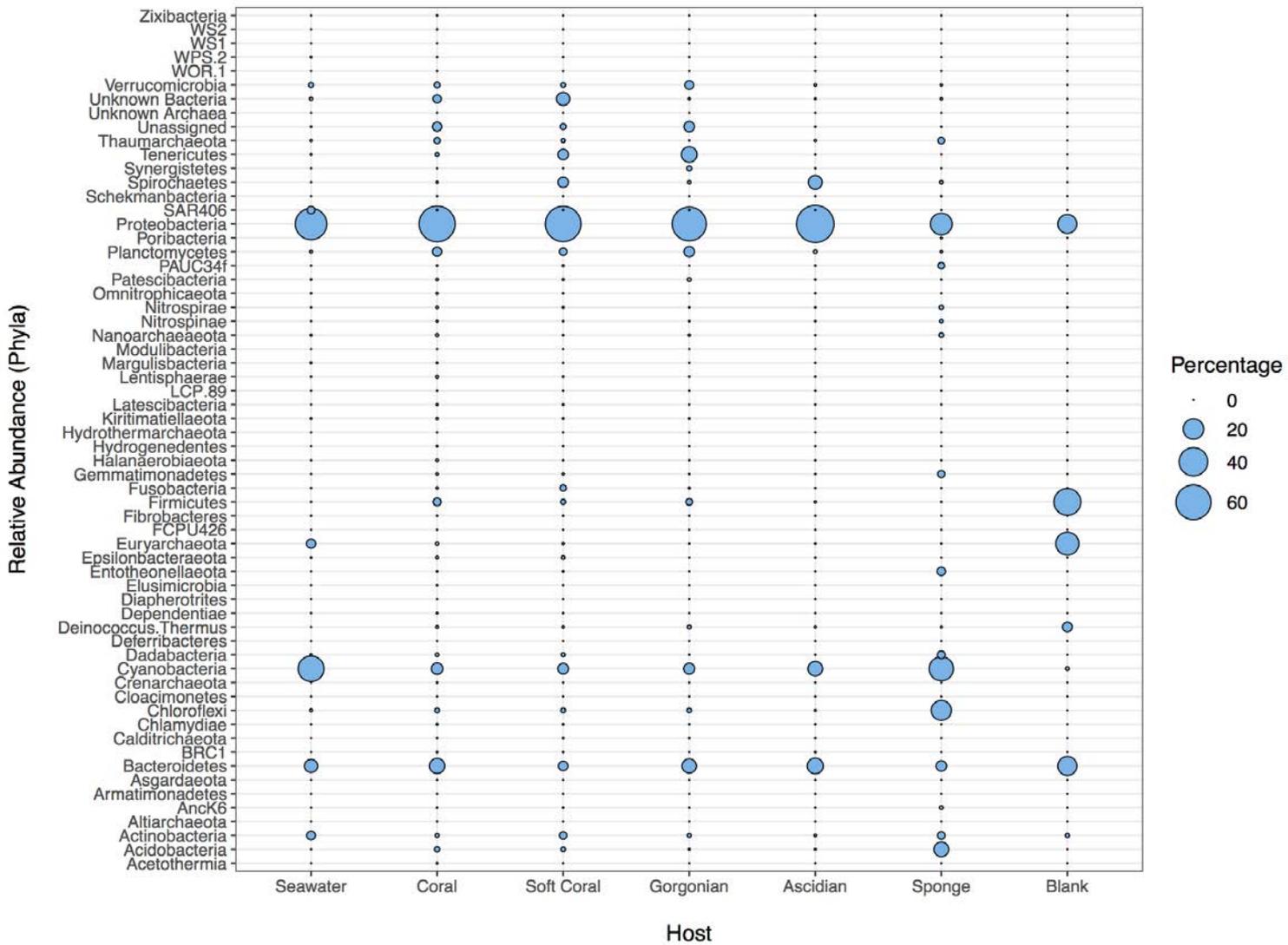


Figure S2.3. Relative abundance of the total prokaryotic phyla associated with each taxonomic group as well as seawater and blank samples (blank extractions and sequencing controls).

Figure S2.4a – Sponge weighted UniFrac

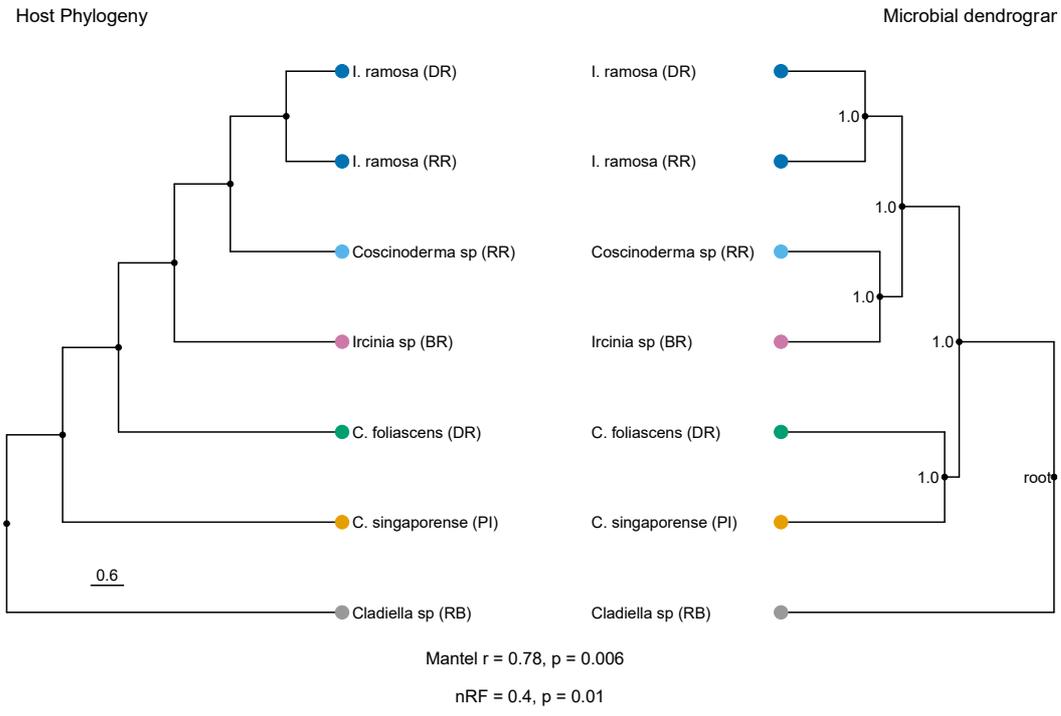


Figure S2.4b – Sponge unweighted UniFrac

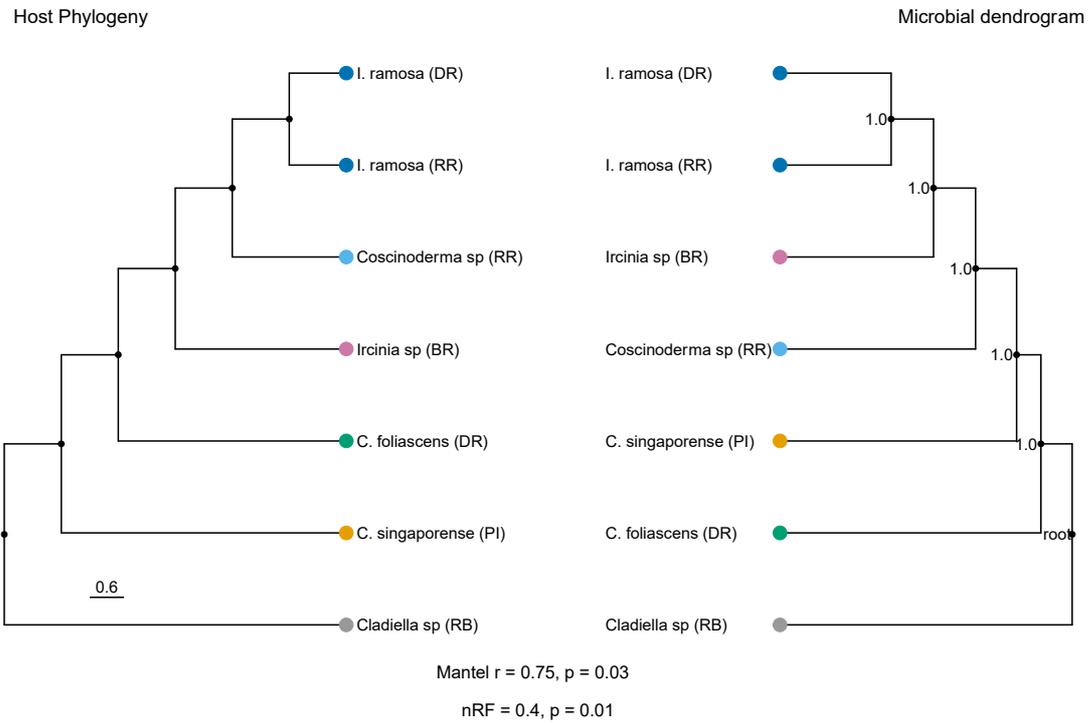


Figure S2.5a – Coral weighted UniFrac

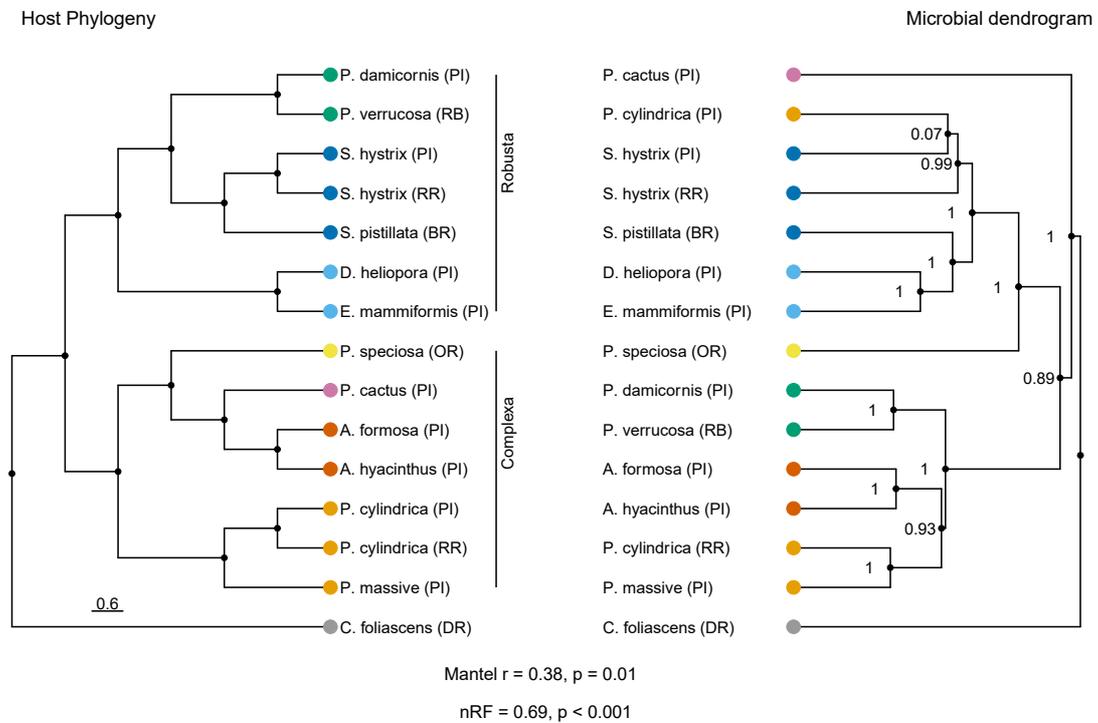


Figure S2.5b – Coral Unweighted UniFrac

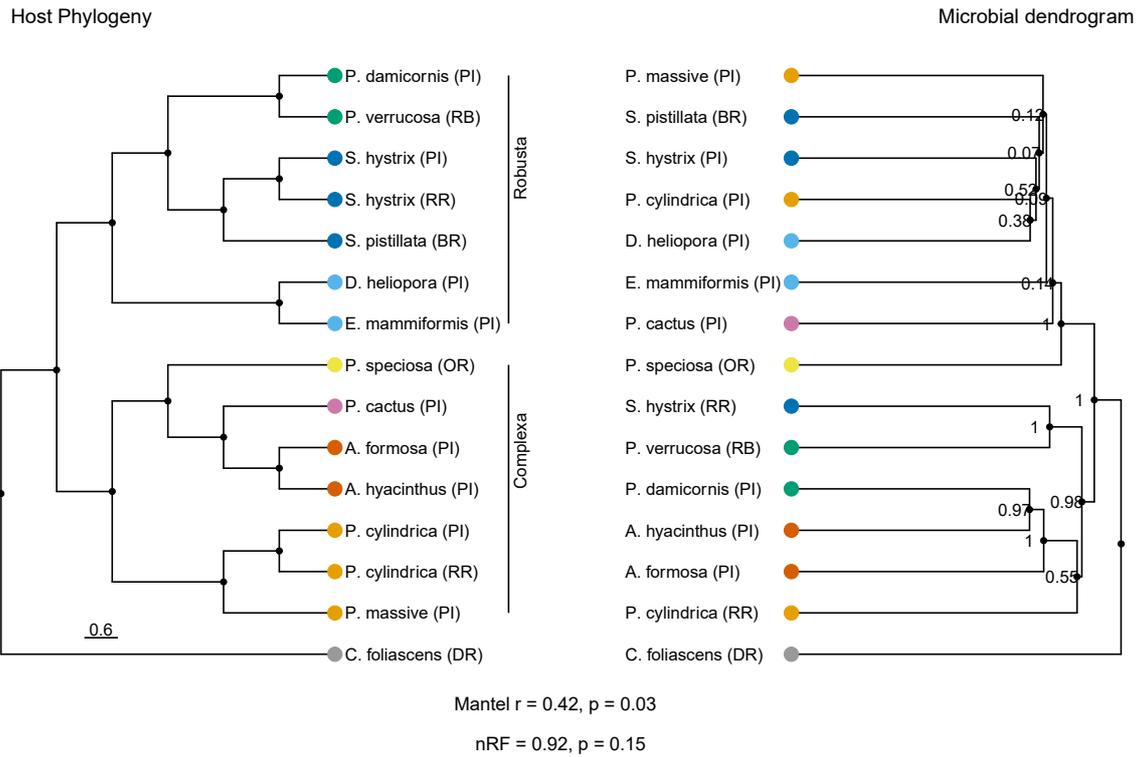


Figure S2.6a – Octocoral weighted UniFrac

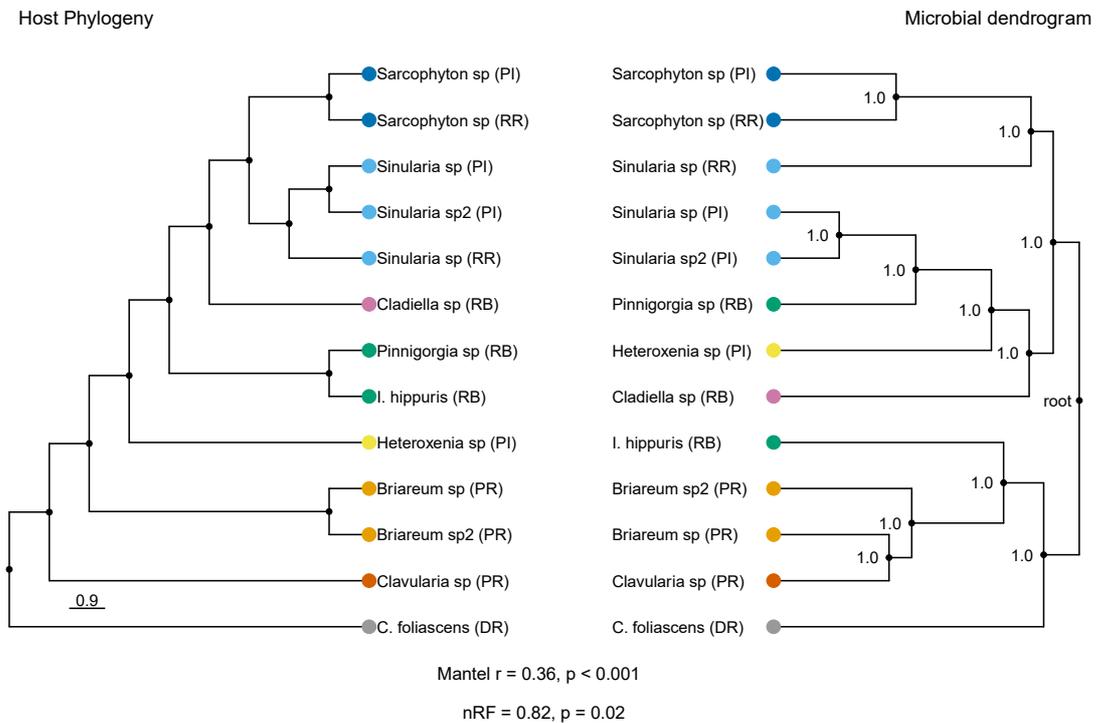


Figure S2.6b - Octocoral unweighted UniFrac

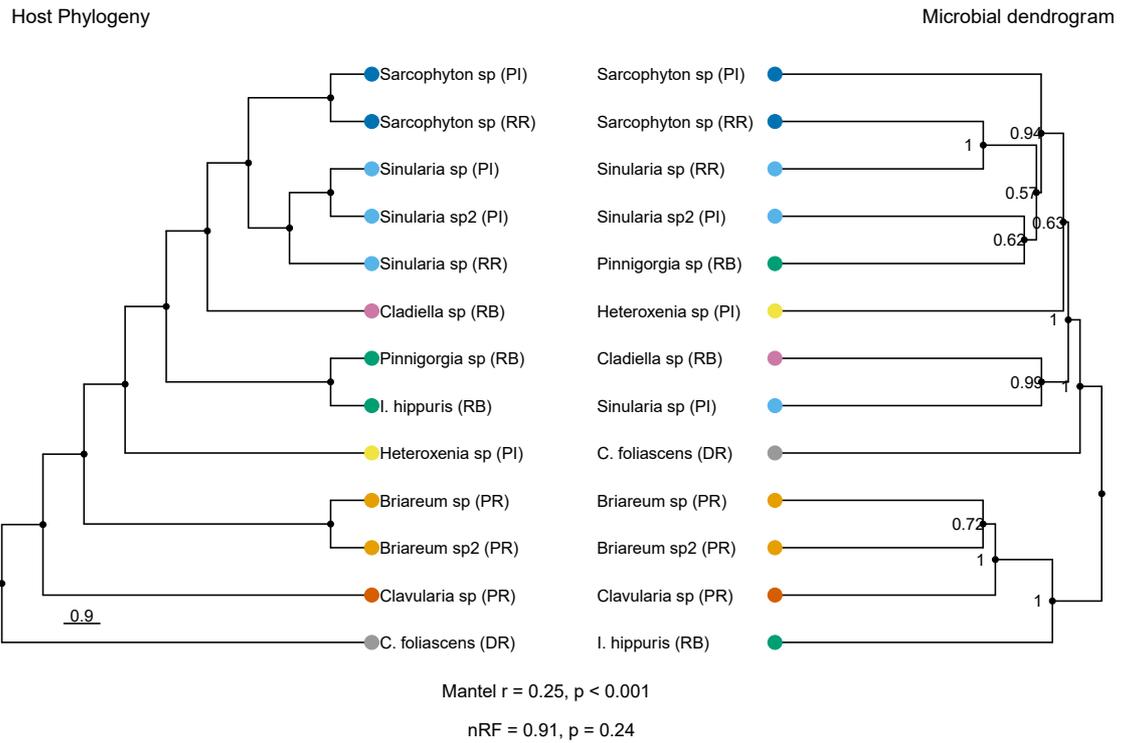


Figure S2.7a – Ascidian weighted UniFrac

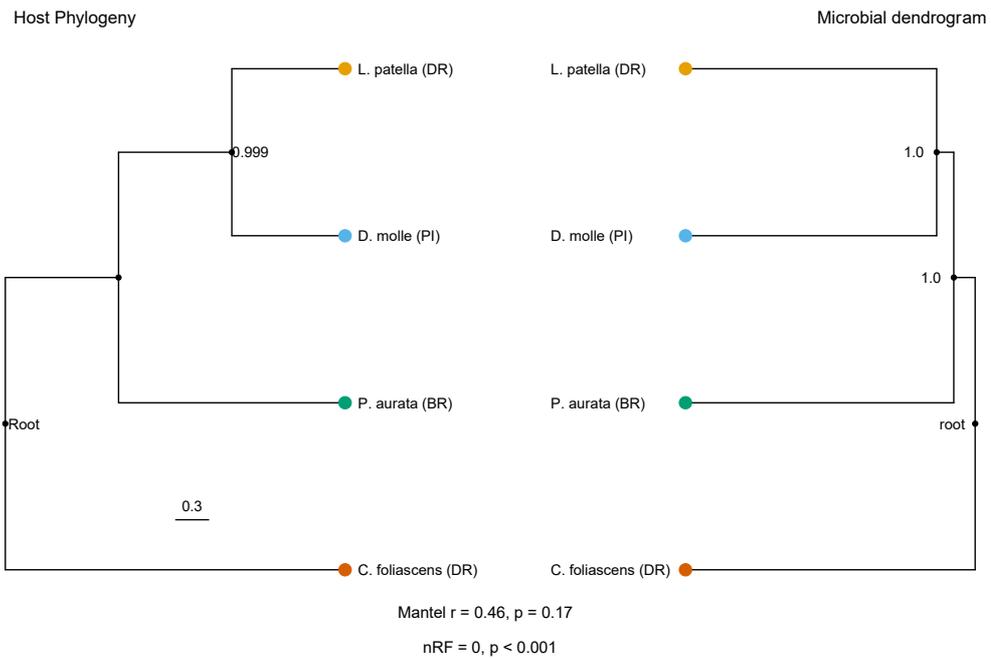
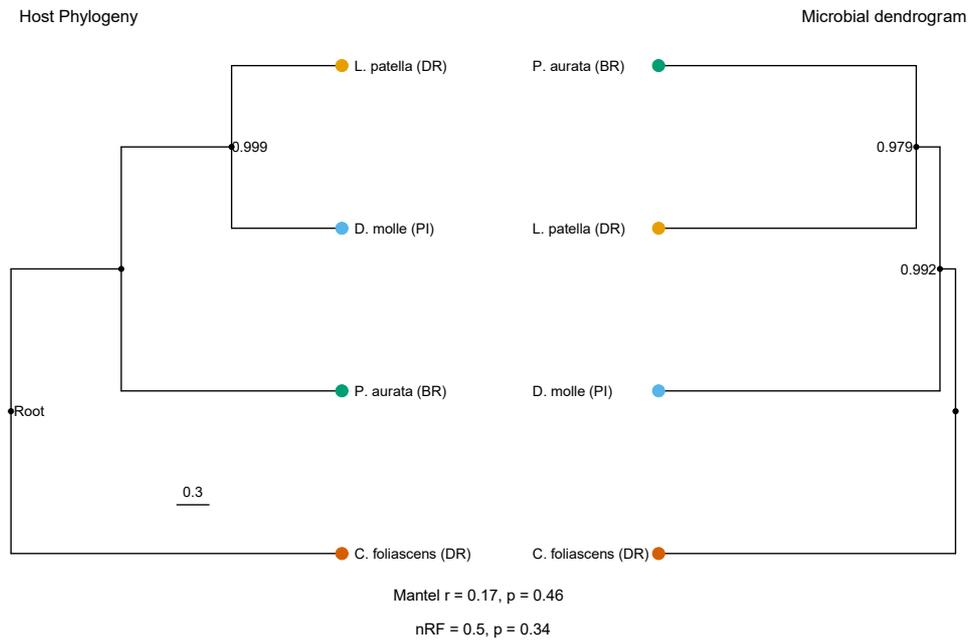


Figure S2.7b – Ascidian unweighted UniFrac



Figures S2.4-S2.7. Host phylogeny compared to a microbial dendrogram based on weighted (a) and unweighted (b) UniFrac distances for microbial composition of each host species. Numbers at nodes reflect posterior probability for clade support in the host tree and jackknife support values in dendrograms. Branch tips are coloured to reflect clades in host phylogeny. Initials in brackets next to species name refer to collection site. *BR* = Broadhurst Reef, *DR* = Davies Reef, *OR* = Osprey Reef, *PI* = Palm Islands (*Orpheus* and *Pelorus*), *PR* = Pandora Reef, *RB* = Rib Reef, *RR* = Ribbon Reefs.

Figure S2.8 – Ascidian

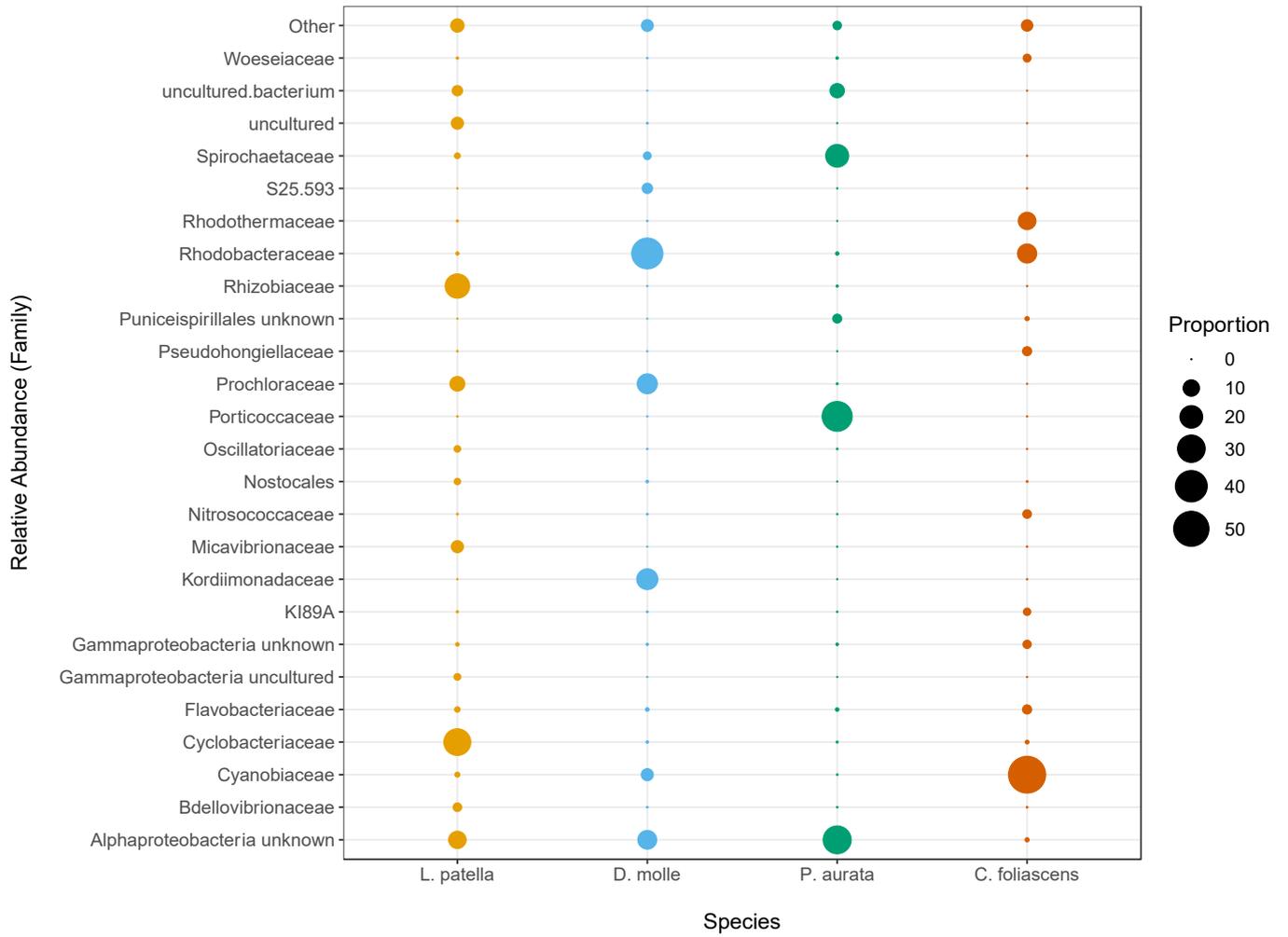


Figure S2.9 – Sponge

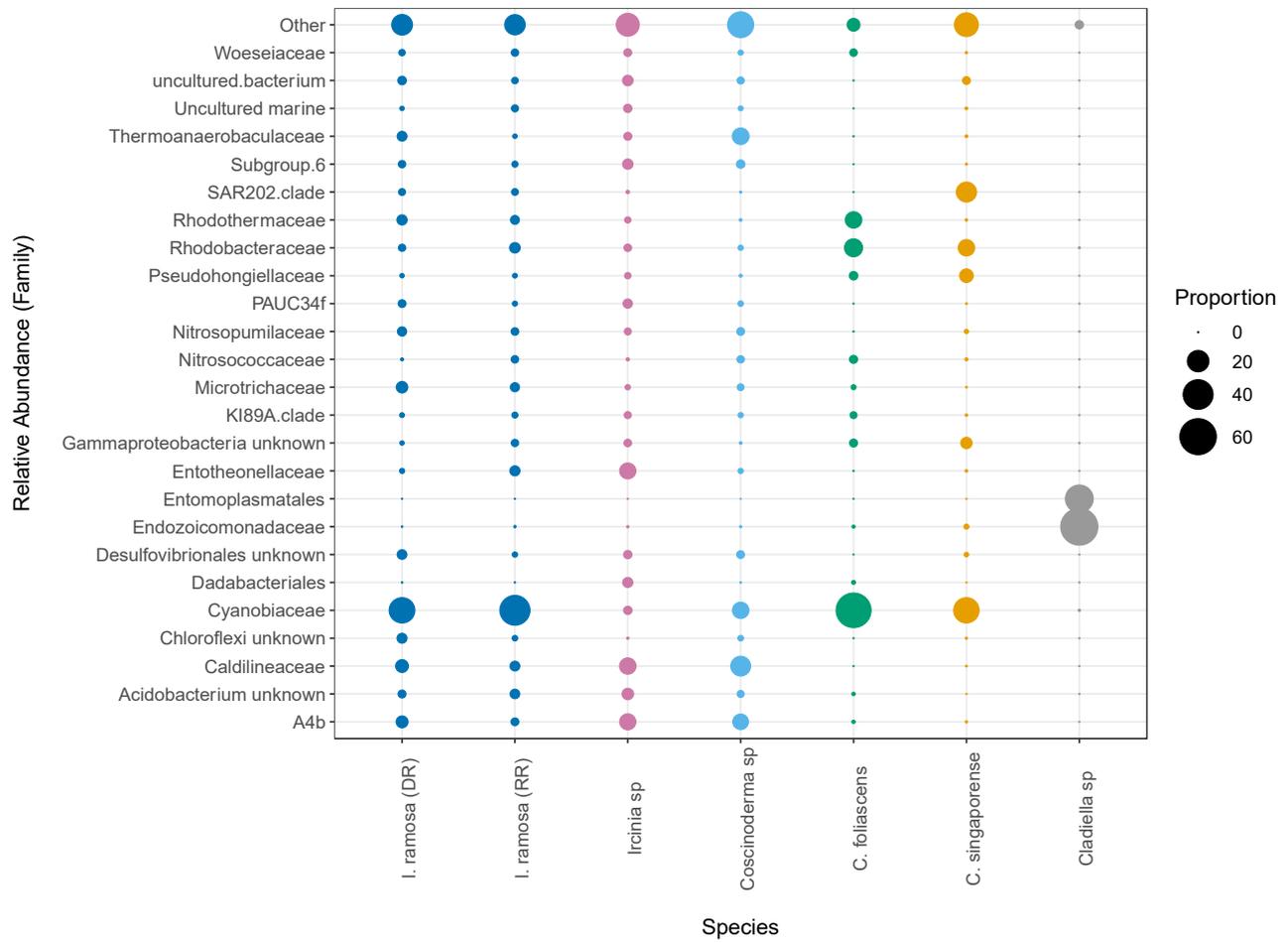


Figure S2.10 – Octocoral

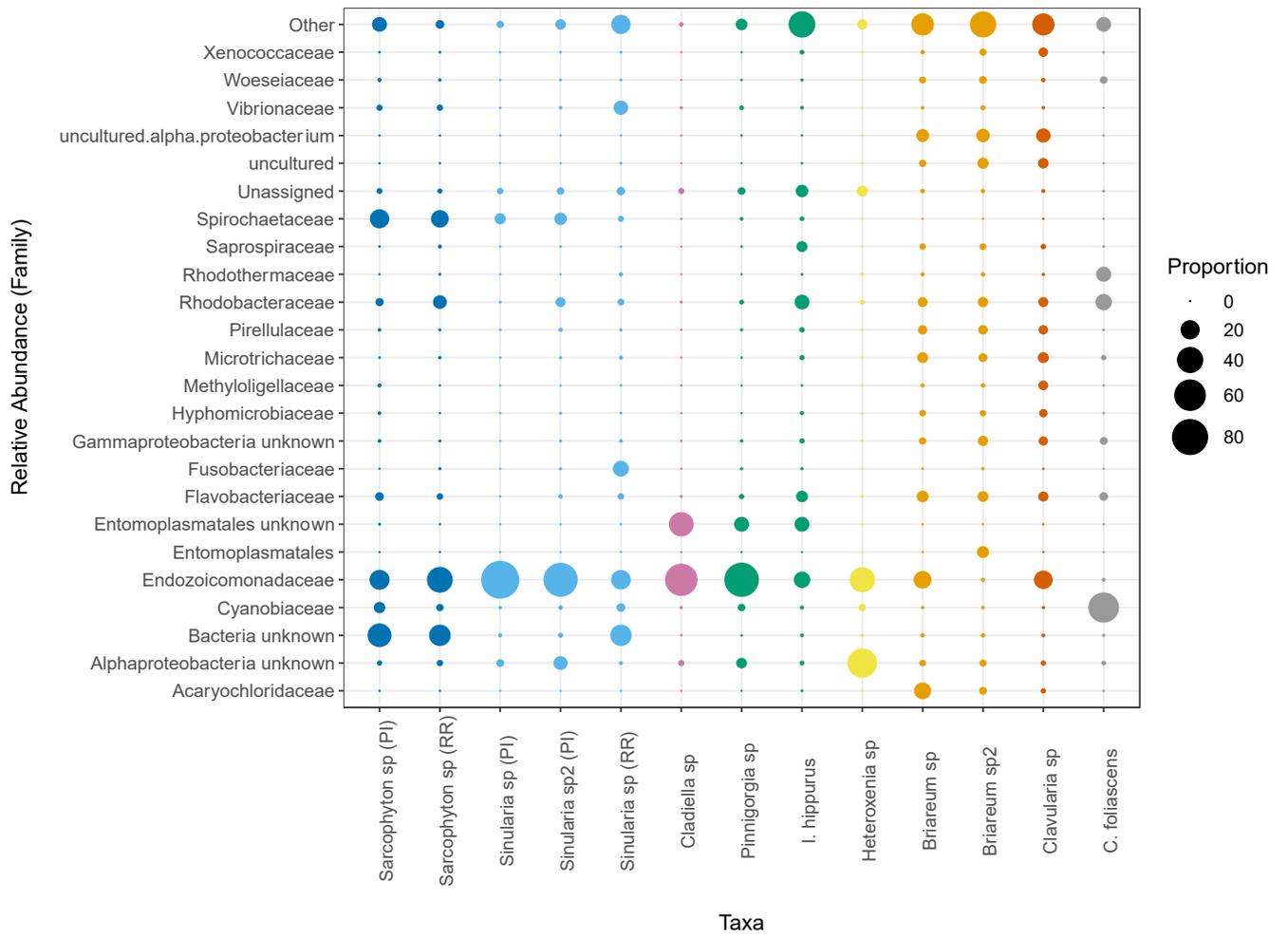
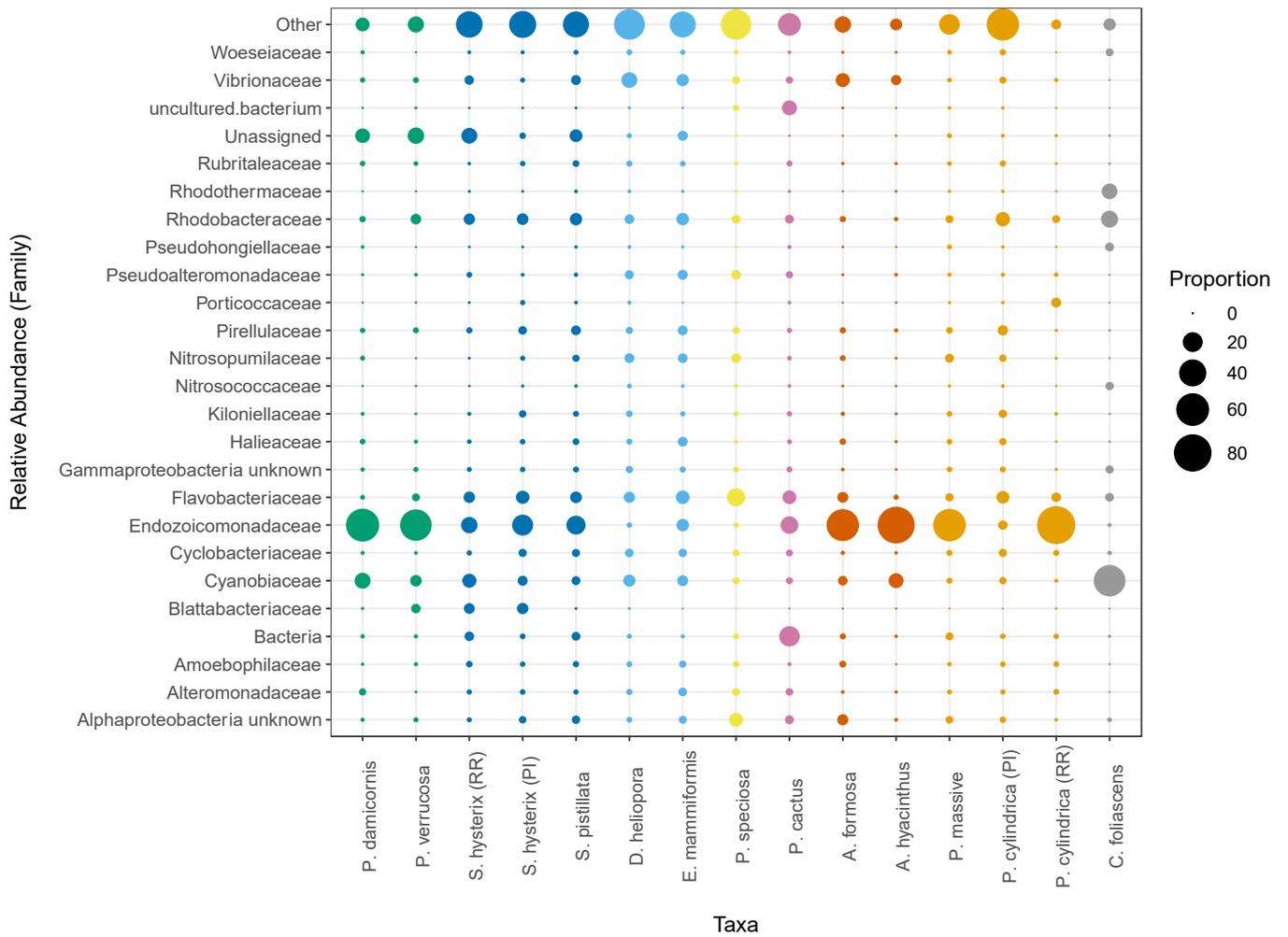


Figure S2.11 – Coral



Supplementary Figure S2.8-S2.11. Relative abundance of the top 25 prokaryotic families associated with each species of ascidian (S8), sponge (S9), octocoral (S10) and coral (S11). Colours refer to clades in host phylogeny

Appendix B. Supporting tables and figures for chapter 3

Table S3.1. Species list and associated metadata (adapted from Table S2.1) including number of amplicon sequence variants (ASVs) before and after core filtering.

Taxa	Species	Replicates	Collection Site	No. ASVs	No core ASVs	Proportion
Coral	<i>Acropora formosa</i>	5	Orpheus Island	537	9	1.68
Coral	<i>Acropora hyacinthus</i>	3	Orpheus Island	272	15	5.51
Coral	<i>Diploastrea heliopora</i>	3	Orpheus Island	2562	14	0.55
Coral	<i>Echinopora mammiformis</i>	5	Orpheus Island	1907	17	0.89
Coral	massive <i>Porites</i> sp.	5	Orpheus Island	2120	9	0.42
Coral	<i>Pachyseris speciosa</i>	3	Osprey Reef	1150	15	1.30
Coral	<i>Pavona cactus</i>	5	Orpheus Island	1423	17	1.19
Coral	<i>Pocillopora damicornis</i>	3	Orpheus Island	250	10	4.00
Coral	<i>Pocillopora verrucosa</i>	5	Rib Reef	580	6	1.03
Coral	<i>Porites cylindrica</i>	3	Orpheus Island	2618	13	0.50
Coral	<i>Porites cylindrica</i>	5	Ribbon Reef 10	298	9	3.02
Coral	<i>Seriatopora hystrix</i>	5	Ribbon Reef 10	774	10	1.29
Coral	<i>Seriatopora hystrix</i>	3	Orpheus Island	1019	17	1.67
Coral	<i>Stylophora pistillata</i>	5	Broadhurst Reef	2328	14	0.60
Octocoral	<i>Sarcophyton</i> sp.	3	Orpheus Island	224	8	3.57
Octocoral	<i>Heteroxenia</i> sp.	3	Pelorus Island	98	4	4.08
Octocoral	<i>Briareum</i> sp.	5	Pandora Reef	1766	23	1.30
Octocoral	<i>Briareum</i> sp. 2	5	Pandora Reef	2594	24	0.93
Octocoral	<i>Clavularia</i> sp.	5	Pandora Reef	1596	34	2.13
Octocoral	<i>Sinularia</i> sp.	5	Orpheus Island	105	7	6.67
Octocoral	<i>Sinularia</i> sp. 2	4	Orpheus Island	170	10	5.88
Octocoral	<i>Cladiella</i> sp.	5	Rib Reef	70	15	21.43
Octocoral	<i>Sarcophyton</i> sp.	5	Ribbon Reef 10	286	10	3.50
Octocoral	<i>Sinularia</i> sp.	5	Ribbon Reef 10	432	8	1.85
Octocoral	<i>Pinnigorgia</i> sp.	5	Rib Reef	203	10	4.93
Octocoral	<i>Isis hippuris</i>	5	Rib Reef	1068	3	0.28
Sponge	<i>Coelocarteria singaporensis</i>	3	Orpheus Island	154	20	12.99
Sponge	<i>Ircinia ramosa</i>	5	Davies Reef	237	41	17.30
Sponge	<i>Carteriospongia foliascens</i>	5	Davies Reef	87	17	19.54
Sponge	<i>Ircinia</i> sp.	5	Broadhurst Reef	208	52	25.00
Sponge	<i>Ircinia ramosa</i>	5	Ribbon Reef 10	289	37	12.80
Sponge	<i>Coscinoderma</i> sp.	5	Ribbon Reef 10	688	30	4.36
Control	Seawater	1	Pandora Reef	343	NA	NA
Control	Seawater	1	Rib Reef	327	NA	NA
Control	Seawater	1	Davies Reef	260	NA	NA
Control	Seawater	1	Broadhurst Reef	300	NA	NA
Control	Seawater	1	Orpheus Island	260	NA	NA
Control	Seawater	1	Orpheus Island	276	NA	NA
Control	Seawater	1	Orpheus Island	286	NA	NA
Control	Seawater	1	Orpheus Island	302	NA	NA
Total samples		149				

Table S3.2a-b. All microbial genera tested independently for their cophylogenetic fit ordered by mean adjusted p-value. Genera are Silva database classification

a) Fragment insertion phylogeny

Genus	Mean p-value	Std Dev	Mean SS	Std Dev
Endozoicomonas	0.000	0.000	0.864	0.009
Candidatus Nitrosopumilus	0.003	0.016	0.855	0.033
Subgroup 10	0.006	0.010	0.647	0.100
Sva0996 marine group	0.014	0.018	0.891	0.016
Spirochaeta 2	0.015	0.043	0.512	0.134
uncultured gamma proteobacterium	0.110	0.084	0.833	0.036
Maritimimonas	0.129	0.157	0.481	0.180
Woeseia	0.235	0.197	0.861	0.039
Ekhidna	0.240	0.158	0.729	0.087
AqS1	0.315	0.265	0.853	0.056
Bythopirellula	0.333	0.224	0.810	0.062
Pir4 lineage	0.400	0.250	0.806	0.061
Epulopiscium	0.492	0.302	0.844	0.065
uncultured organism	0.511	0.270	0.770	0.087
Vibrio	0.565	0.199	0.927	0.015
Filomicrobium	0.571	0.257	0.907	0.036
Muricauda	0.651	0.233	0.915	0.023
Thalassotalea	0.663	0.249	0.919	0.030
uncultured alpha proteobacterium	0.712	0.224	0.944	0.019
Pelagibius	0.737	0.287	0.918	0.042
uncultured Chloroflexus sp.	0.793	0.267	0.931	0.046
Rubritalea	0.824	0.260	0.896	0.045
Zeaxanthinibacter	0.831	0.306	0.866	0.044
Winogradskyella	0.848	0.247	0.903	0.049
Rhodopirellula	0.849	0.281	0.775	0.115
Synechococcus CC9902	0.870	0.253	0.967	0.008
Tenacibaculum	0.957	0.270	0.832	0.072
Roseibacillus	0.989	0.262	0.845	0.068
Aquibacter	0.991	0.270	0.885	0.032
Blastopirellula	0.991	0.271	0.920	0.022
uncultured Chloroflexi bacterium	0.993	0.252	0.881	0.050
Phycisphaera	1.000	0.263	0.884	0.052

b) De novo phylogeny

Genus	Mean p-value	Std Dev	Mean SS	Std Dev
Endozoicomonas	0	0	0.89754841	0.00541842
Candidatus Nitrosopumilus	0.00046346	0.00159942	0.83722935	0.03962837
Sva0996 marine group	0.01294826	0.01984344	0.87314344	0.02271334
Spirochaeta 2	0.01314764	0.02893225	0.54973503	0.11348887
uncultured gamma proteobacterium	0.01652746	0.02821652	0.828955	0.02623199
Subgroup 10	0.02354235	0.03783913	0.7147568	0.06941066
Maritimimonas	0.18109183	0.141814	0.61743654	0.13806493
AqS1	0.21715568	0.12985604	0.87105047	0.05039631
Pir4 lineage	0.30156128	0.20574908	0.80265877	0.05701447
Ekhidna	0.3647472	0.17568172	0.73311272	0.13539457
Pelagibius	0.40603807	0.30327612	0.87922409	0.05513641
Woeseia	0.42837622	0.24957171	0.87646385	0.04071044
Zeaxanthinibacter	0.44137178	0.30229575	0.8321911	0.07847418
uncultured organism	0.46863218	0.2770733	0.80629393	0.06719614
Phycisphaera	0.49883617	0.22655688	0.84898557	0.06361932
Vibrio	0.60037136	0.23358876	0.9232986	0.0186195
Epulopiscium	0.61941422	0.32285334	0.88390882	0.05664716
Aquibacter	0.66876886	0.25058346	0.86372535	0.03456827
Winogradskyella	0.69697363	0.28401244	0.91754091	0.05621405
Bythopirellula	0.74076572	0.30874365	0.89998718	0.07041909
Blastopirellula	0.76243963	0.25386699	0.91749053	0.02145461
uncultured Chloroflexus sp.	0.81345328	0.25570366	0.93146522	0.03908261
Muricauda	0.81431128	0.23077753	0.93230662	0.02312719
Rubritalea	0.86900596	0.2348349	0.92824689	0.03524112
Filomicrobium	0.92618655	0.21551517	0.96799248	0.01109778
uncultured Chloroflexi bacterium	0.95172172	0.20684291	0.94607369	0.02726422
Synechococcus CC9902	0.96649399	0.23211609	0.97212461	0.00894272
Thalassotalea	0.98905906	0.2737873	0.93609775	0.02607674
Seonamhaeicola	0.99099099	0.22240171	0.87147313	0.04967798
Tenacibaculum	0.99104104	0.28238155	0.84094452	0.06202091
Roseibacillus	0.995996	0.26697528	0.88564638	0.06856597
Rhodopirellula	0.99695696	0.2402421	0.82320426	0.09600309

Table S3.3a-b. All microbial families tested independently for their cophylogenetic fit ordered by mean adjusted p-value. Family are Silva database classification

a) Fragment insertion phylogeny

Family	Mean p-value	Std Dev	Mean SS	Std dev
Microtrichaceae	0	0	0.8204841	0.02231827
Endozoicomonadaceae	0	0	0.88037192	0.00835274
Woeseiaceae	3.003E-05	0.00017162	0.6668438	0.06195385
Rhodobacteraceae	0.00124314	0.00310754	0.96263694	0.00620144
Spirochaetaceae	0.00193575	0.0035884	0.63604243	0.05285384
Nitrosopumilaceae	0.00273331	0.01078946	0.90019767	0.02177247
Flavobacteriaceae	0.00285592	0.01308215	0.97303695	0.0041059
Cyclobacteriaceae	0.00737336	0.03084259	0.88535558	0.02705977
Thermoanaerobaculaceae	0.01823267	0.03489803	0.71240898	0.06374157
uncultured gamma proteobacterium	0.02205954	0.0405029	0.82845097	0.03030317
Caldilineaceae	0.03035333	0.06205071	0.78155439	0.06140666
Rhodothermaceae	0.03300059	0.04108034	0.75407903	0.06367784
Cyanobiaceae	0.03330858	0.03710601	0.95303516	0.00781094
Methyloligellaceae	0.03656509	0.05697419	0.75140682	0.10394676
Sandaracinaceae	0.06849168	0.07380639	0.65588458	0.13541007
Bdellovibrionaceae	0.08040591	0.08216856	0.38916301	0.12212088
Kiloniellaceae	0.11597451	0.11538883	0.93183838	0.01613573
Arenicellaceae	0.13703715	0.24401229	0.60797889	0.13659819
Xenococcaceae	0.16676387	0.08803083	0.86122673	0.03964996
Pirellulaceae	0.19809006	0.15559817	0.95881465	0.00754572
Rhizobiaceae	0.24412961	0.26236681	0.90923034	0.0319037
Saprosiraceae	0.2601042	0.18991291	0.90120283	0.0373684
Phycisphaeraceae	0.34324899	0.1833965	0.82025718	0.05117992
Cellvibrionaceae	0.35251675	0.1404089	0.7640246	0.0743806
Alteromonadaceae	0.35584665	0.21822828	0.82978698	0.05427192
Nitrosococcaceae	0.36185291	0.18032499	0.90336435	0.03189026
Hyphomonadaceae	0.38087908	0.19156781	0.848814	0.05836675
Entotheonellaceae	0.42480542	0.20845676	0.82643117	0.0839564
Shewanellaceae	0.449278	0.24531861	0.81000378	0.06913641
Terasakiellaceae	0.46453104	0.20978354	0.71454347	0.08120071
Nitricolaceae	0.47050585	0.23457238	0.855336	0.0427018
Amoebophilaceae	0.49850262	0.26153929	0.7927841	0.07711805
Gimesiaceae	0.50361704	0.27286875	0.87304097	0.0560165
Lachnospiraceae	0.5265412	0.27884414	0.8790167	0.04044319
uncultured alpha proteobacterium	0.5323778	0.2349155	0.95115591	0.02011795
Phormidiesmiaceae	0.64881384	0.27294617	0.8178217	0.10835855
Fusobacteriaceae	0.72745957	0.26063792	0.85765457	0.07439291
Peptostreptococcaceae	0.75318812	0.28264998	0.90045699	0.03737634
Rubinisphaeraceae	0.81400797	0.27156258	0.93804015	0.01959692
Halieaceae	0.82781128	0.25188798	0.9517991	0.00919382
DEV007	0.86560765	0.30179493	0.85231722	0.07187811
A4b	0.91129872	0.2585702	0.91023073	0.04305836
Hyphomicrobiaceae	0.92577957	0.25130611	0.96080222	0.01346965
Vibrionaceae	0.93406902	0.25468066	0.96150702	0.00982993
Sphingomonadaceae	0.93727273	0.25943505	0.85571322	0.06043133
Spongibacteraceae	0.94339339	0.24563301	0.87195355	0.05416095
Stappiaceae	0.94410047	0.30323563	0.83892146	0.0730031
Kordiimonadaceae	0.95375375	0.22306717	0.87789335	0.04671973
uncultured Chloroflexi bacterium	0.97543832	0.22947736	0.9414791	0.047404
Colwelliaceae	0.98905906	0.2737873	0.93609775	0.02607674
Clostridiaceae 1	0.99304365	0.27044931	0.89992061	0.0428815
Pseudoalteromonadaceae	0.99397397	0.247718	0.97438624	0.01412577
Unknown Family	0.99496496	0.25576754	0.83581078	0.0784975
Rubritaleaceae	1	0.25021731	0.93596831	0.02018752

b) De novo phylogeny

Family	Mean p-value	Std Dev	Mean SS	Std dev
Woeseiaceae	0.000	0.000	0.687	0.054
Microtrichaceae	0.000	0.000	0.839	0.022
Endozoicomonadaceae	0.000	0.000	0.855	0.009
Rhodobacteraceae	0.000	0.000	0.956	0.007
Flavobacteriaceae	0.002	0.006	0.974	0.004
Nitrosopumilaceae	0.008	0.042	0.899	0.022
Pirellulaceae	0.008	0.017	0.936	0.016
Cyanobiaceae	0.009	0.009	0.948	0.007
Thermoanaerobaculaceae	0.012	0.025	0.667	0.095
Cyclobacteriaceae	0.015	0.026	0.899	0.020
Spirochaetaceae	0.025	0.098	0.666	0.087
Colwelliaceae	0.026	0.018	0.838	0.032
Arenicellaceae	0.036	0.115	0.583	0.113
Kiloniellaceae	0.041	0.079	0.923	0.016
Xenococcaceae	0.075	0.058	0.831	0.052
Caldilineaceae	0.082	0.066	0.810	0.050
Terasakiellaceae	0.085	0.068	0.564	0.119
Sandaracinaceae	0.086	0.134	0.674	0.142
Rhodothermaceae	0.090	0.075	0.822	0.062
uncultured gamma proteobacterium	0.110	0.084	0.833	0.036
Methyloligellaceae	0.211	0.135	0.839	0.049
Lachnospiraceae	0.285	0.185	0.861	0.047
Cellvibrionaceae	0.290	0.179	0.718	0.068
Shewanellaceae	0.295	0.200	0.814	0.070
Nitrosococcaceae	0.298	0.237	0.876	0.040
Hyphomonadaceae	0.327	0.154	0.836	0.051
Haliaceae	0.335	0.180	0.934	0.014
Entotheonellaceae	0.447	0.270	0.811	0.087
Bdellovibrionaceae	0.548	0.354	0.546	0.098
DEV007	0.553	0.314	0.842	0.076
Gimesiaceae	0.562	0.288	0.851	0.057
Saprospiraceae	0.579	0.258	0.916	0.031
A4b	0.594	0.270	0.877	0.083
Alteromonadaceae	0.596	0.183	0.863	0.034
Cryomorphaceae	0.598	0.315	0.636	0.153
Phormidesmiaceae	0.641	0.296	0.808	0.110
Peptostreptococcaceae	0.686	0.293	0.896	0.032
Rhizobiaceae	0.725	0.291	0.936	0.022
uncultured alpha proteobacterium	0.745	0.291	0.964	0.022
Hyphomicrobiaceae	0.750	0.272	0.923	0.026
Stappiaceae	0.794	0.281	0.866	0.083
Amoebophilaceae	0.795	0.315	0.835	0.083
Vibrionaceae	0.889	0.278	0.953	0.011
Nitrincolaceae	0.896	0.259	0.870	0.038
Clostridiaceae 1	0.916	0.295	0.865	0.064
uncultured Chloroflexi bacterium	0.949	0.268	0.867	0.061
Unknown Family	0.969	0.251	0.827	0.112
Rubinisphaeraceae	0.985	0.295	0.935	0.020
Sphingomonadaceae	0.992	0.314	0.833	0.067
Rubritaleaceae	0.996	0.244	0.914	0.021
Pseudoalteromonadaceae	0.999	0.273	0.959	0.019
Phycisphaeraceae	1.000	0.300	0.854	0.051

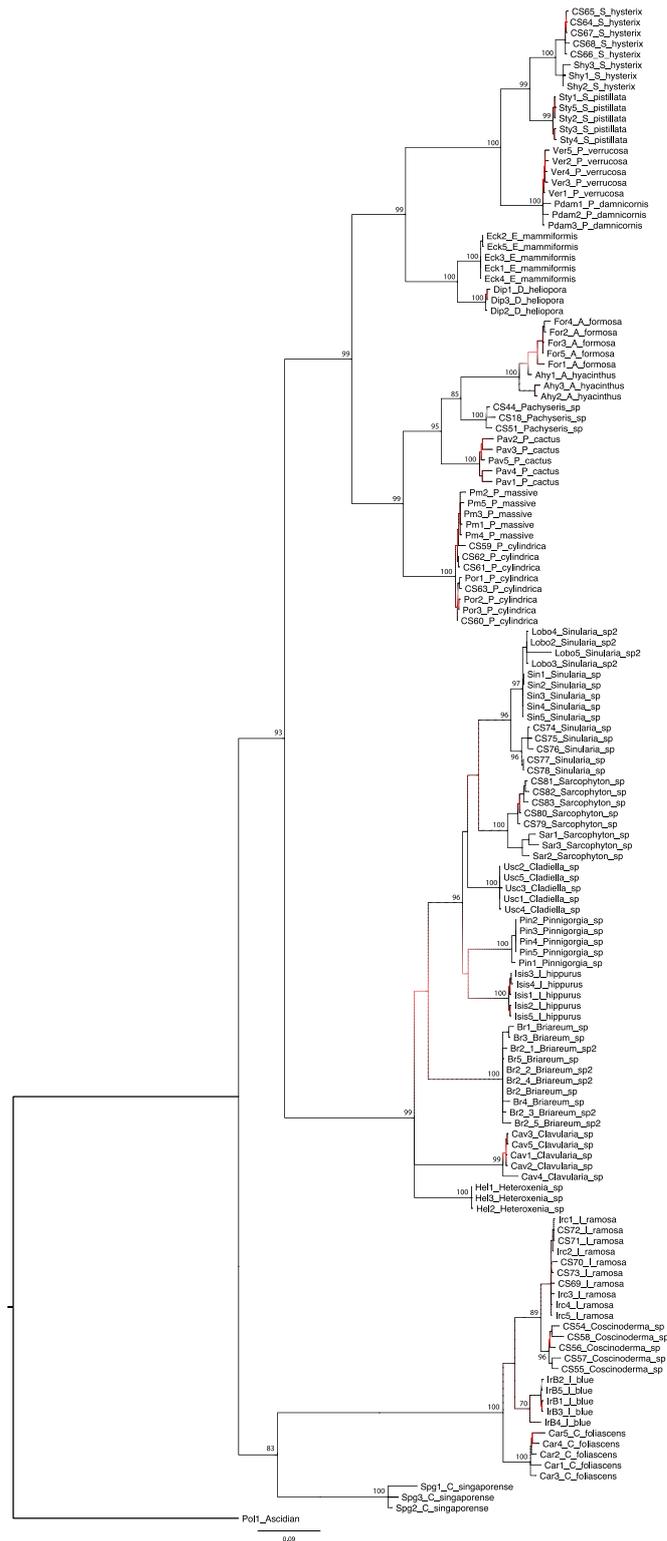


Figure S3.1. Host species phylogenetic tree used for the cophylogenetic analysis. Host tree includes each host species replicate and the ascidian *Polycarpa aurata* is used as an outgroup. Numbers on nodes denote bootstrap support (>70) calculated for individual clades on 1,000 replicates under the same Maximum Likelihood model and parameters as RAxML GUI (Edler et al. 2019). See materials and methods for further details.

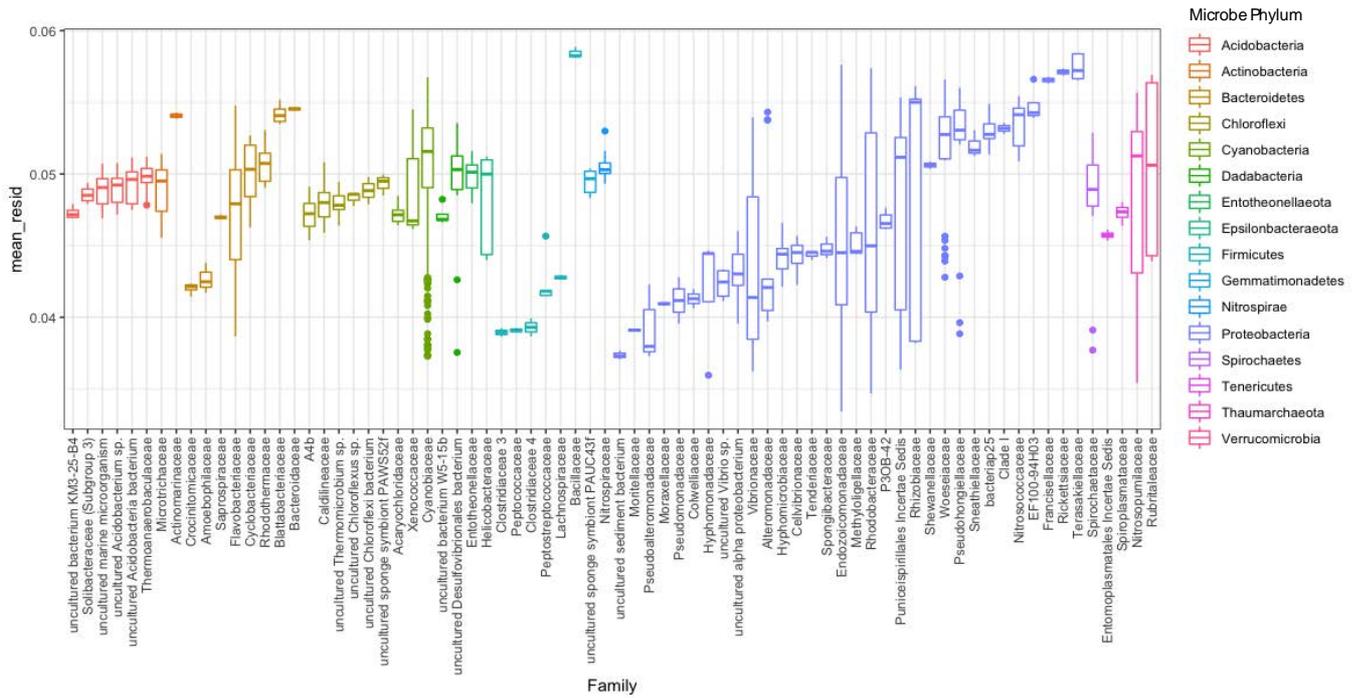


Figure S3.2. Mean residuals for each microbial family across all host species coloured by microbe phylum. For many families, residuals could be relatively high or low depending on which host that particular ASV was found in.

Figure 3.3a. Coral

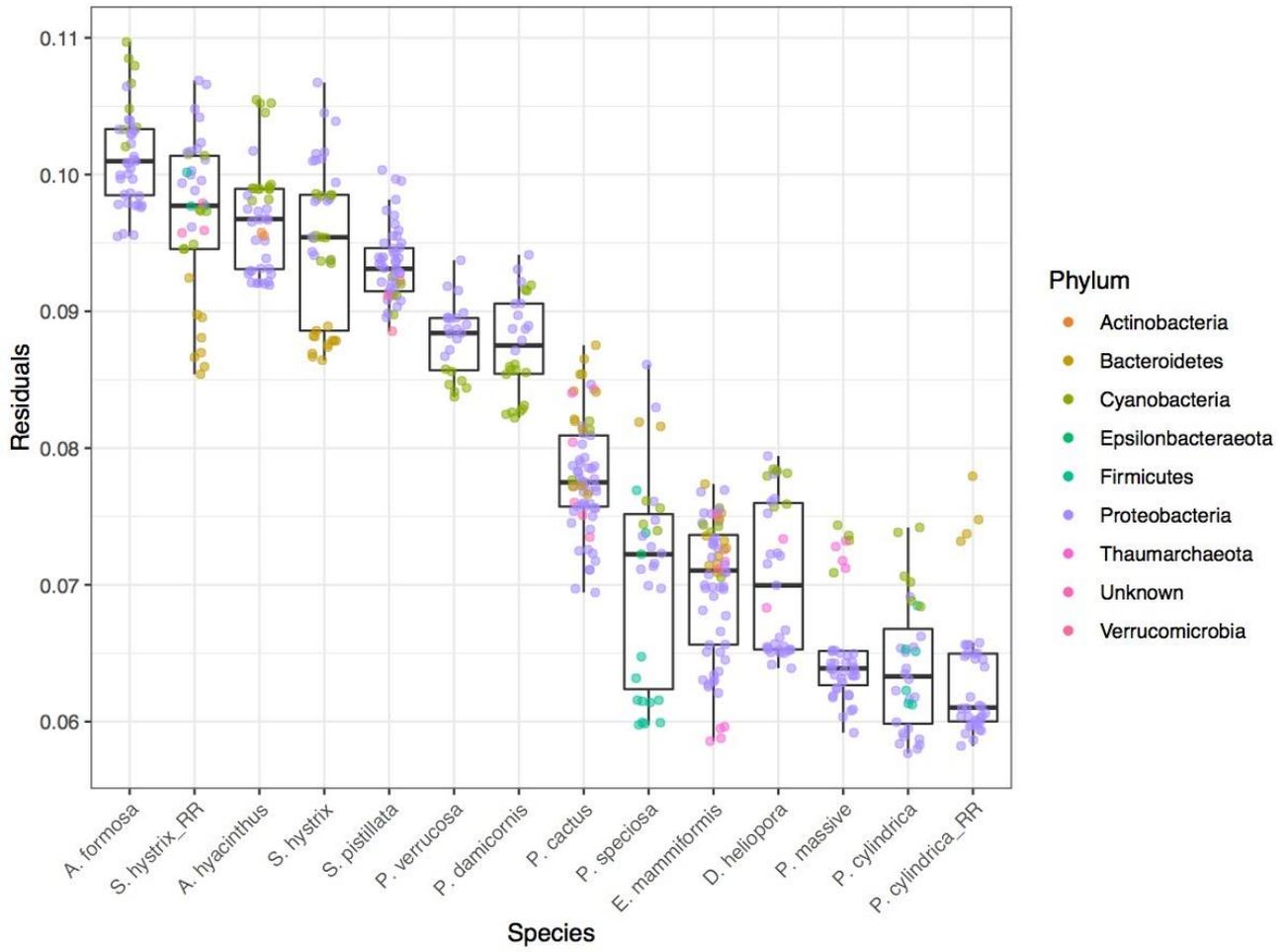


Figure 3.3b. Coral

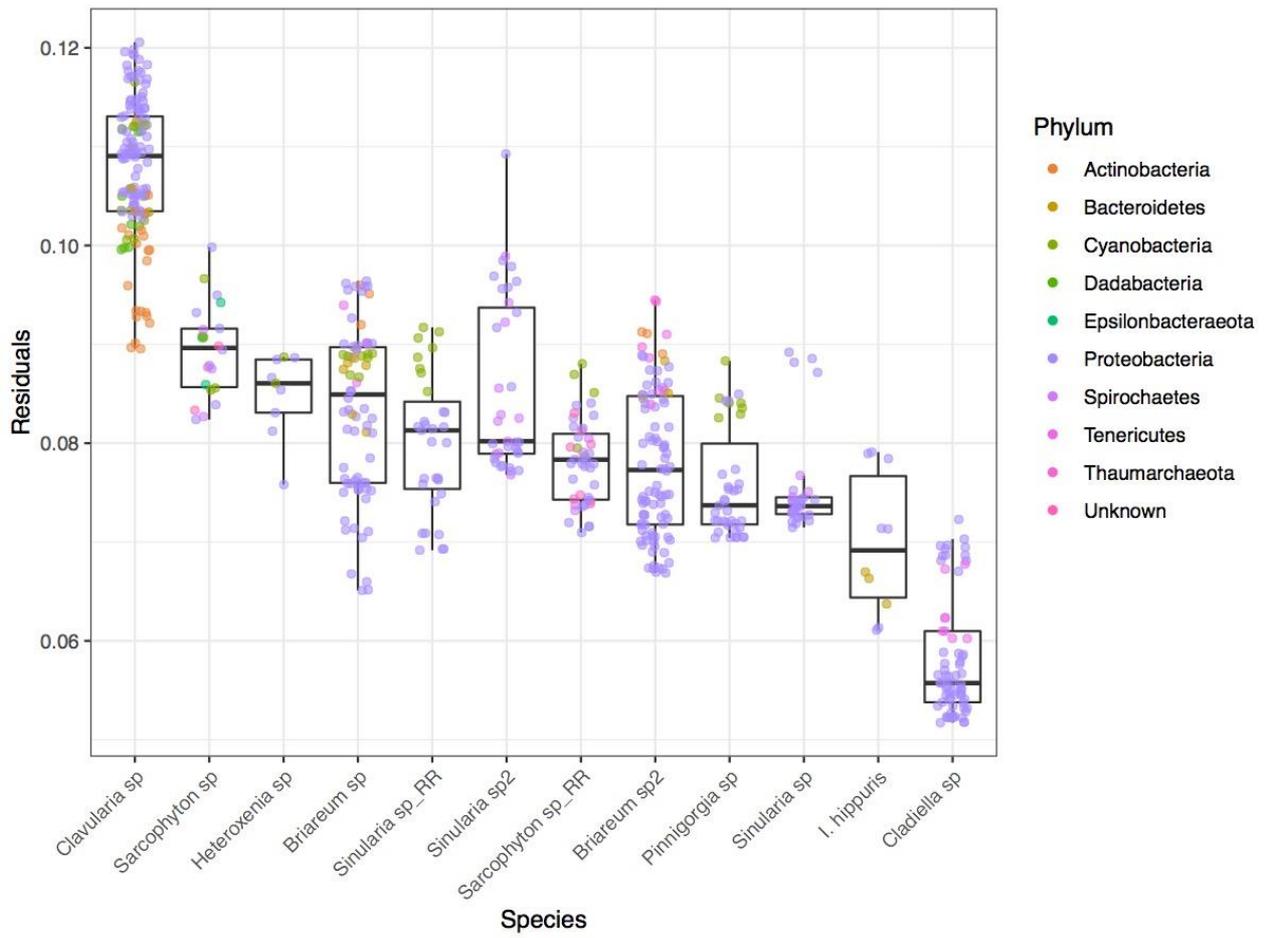
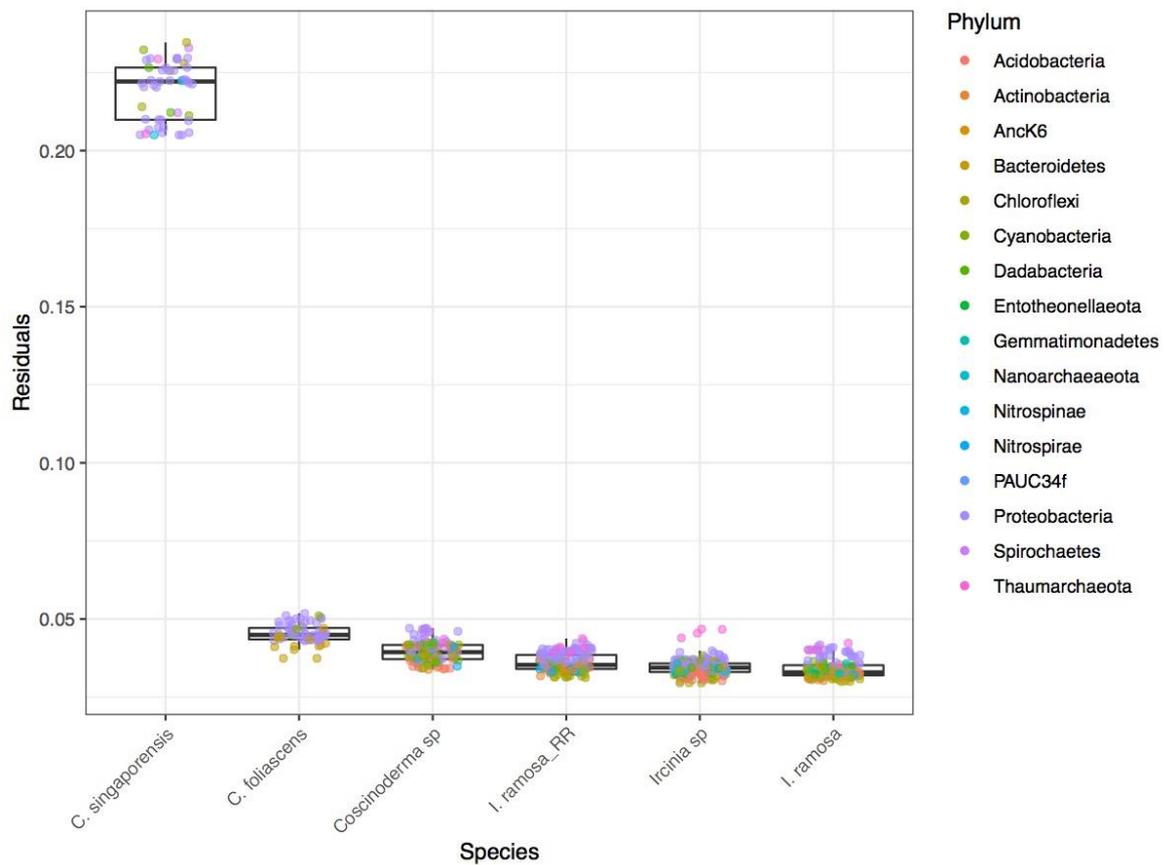


Figure 3.3c. Sponge



Figures S3.3a-c. Residual contributions of each host-microbe link to the overall fit of cophylogeny when each host group is analysed independently. Residuals are grouped by host species and each boxplot is overlaid with the residuals coloured by microbial phylum. Where a species was collected from two locations, 'RR' indicates the replicate from the ribbon reefs.

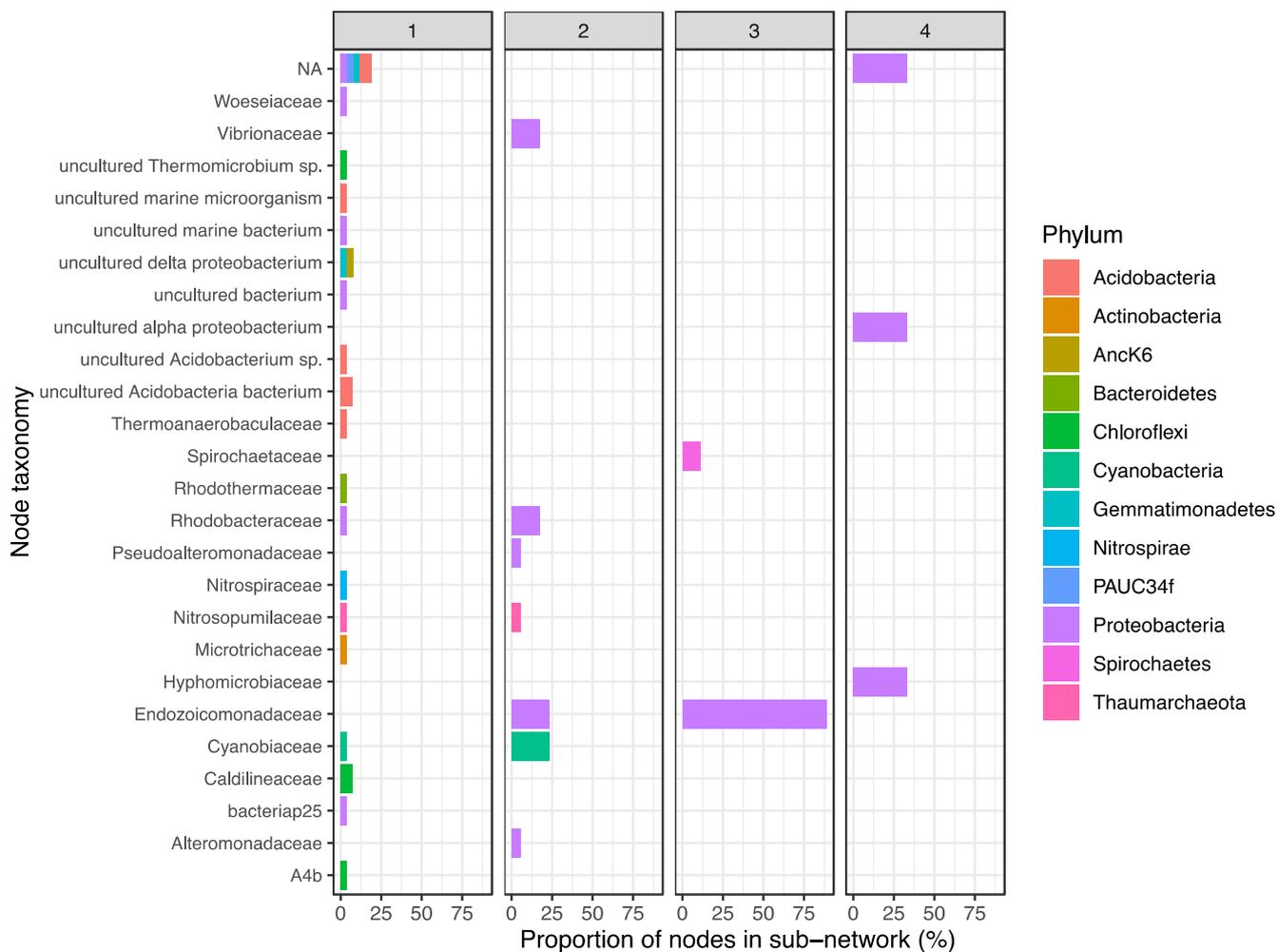


Figure S3.4. The taxonomic classification of each sub-network and the relative number of nodes belonging to each microbial family within a community

Appendix C: Supporting tables and figures for chapter 4

Table S4.1. Primer information for those used in the rRNA hybridisation and subtraction step for microbial mRNA enrichment during the metatranscriptomic laboratory protocol (Stewart et al., 2010)

Domain	Region	ID	Sequence
Bacteria	16S	Eub16S_27F	AGAGTTTGATCCTGGCTCAG
Bacteria	16S	Eub16S_1492R_T7	GCCAGTGAATTGIAATACGACTCACTATAAGGGACGGCTACCTGTTACGACTT
Bacteria	23S	Eub23S_189F	GAATGAAACATCTHAGTA
Bacteria	23S	Eub23S_2490R_T7	GCCAGTGAATTGIAATACGACTCACTATAAGGGCGACATCGAGGTGCCAAAC
Archaea	16S	Arch16S_21F	TCCGGTTGATCCYGCCGG
Archaea	16S	Arch16S_1492_T7	GCCAGTGAATTGIAATACGACTCACTATAAGGGGGYYACCTTGTACGACTT
Archaea	23S	Arch23S_F	ASAGGGTGAHARYCCCGTA
Archaea	23S	Arch23S_R	GCCAGTGAATTGIAATACGACTCACTATAAGGGCTGTCTCRGACGGTCTRAACCCA
Eukaryote	18S	Euk18S_1F	ACCTGGTTGATCCTGCCAG
Eukaryote	18S	Euk18S_1520R_T7	AATTATAATACGACTCACTATAAGATTTCYGCAGGTTACCTAC
Eukaryote	28S	Euk28S_26F	ACCCGCGAAYTTAAGCATA
Eukaryote	28S	Euk28S_3126R_T7	AATTATAATACGACTCACTATAAGATTCTGRYTTAGAGGCGTTCAG

Table S4.2. Taxonomic classification and genome statistics for 415 MAGs dereplicated at 95ANI

Link to table can be found at:

[TableS4.2.xlsx](#)

Table S4.3. Number and proportion of sequences classified within each taxonomic domain. Note that unclassified sequences are not shown and hence not included in the total

Tax	I. micronulosa_1		I. micronulosa_2		I. micronulosa_3	
	Sequence number	Percentage	Sequence number	Percentage	Sequence number	Percentage
Bacteria	23822976	32.01	16283802	19.34	19898971	30.03
Archaea	100813	0.14	12992	0.02	49558	0.07
Viruses	1800	0	1898	0	1801	0
Fungi	11371303	15.28	11606988	13.79	6588896	9.94
Other eukar	3587712	4.82	5842076	6.94	1634946	2.47
Total	74425612	52.25	84193089	40.09	66257110	42.51

Tax	P. foliascens_1		P. foliascens_2		P. foliascens_3	
	Sequence number	Percentage	Sequence number	Percentage	Sequence number	Percentage
Bacteria	10408151	11.67	7107274	10.68	8701807	11.63
Archaea	6951	0.01	4455	0.01	4230	0.01
Viruses	1601	0	870	0	1432	0
Fungi	31062938	34.83	25392540	38.15	27190686	36.34
Other eukar	10684213	11.98	9245744	13.89	10462839	13.98
Total	89172067	58.49	66566979	62.73	74825676	61.96

Figure S4.1a. *Phyllospongia foliascens*

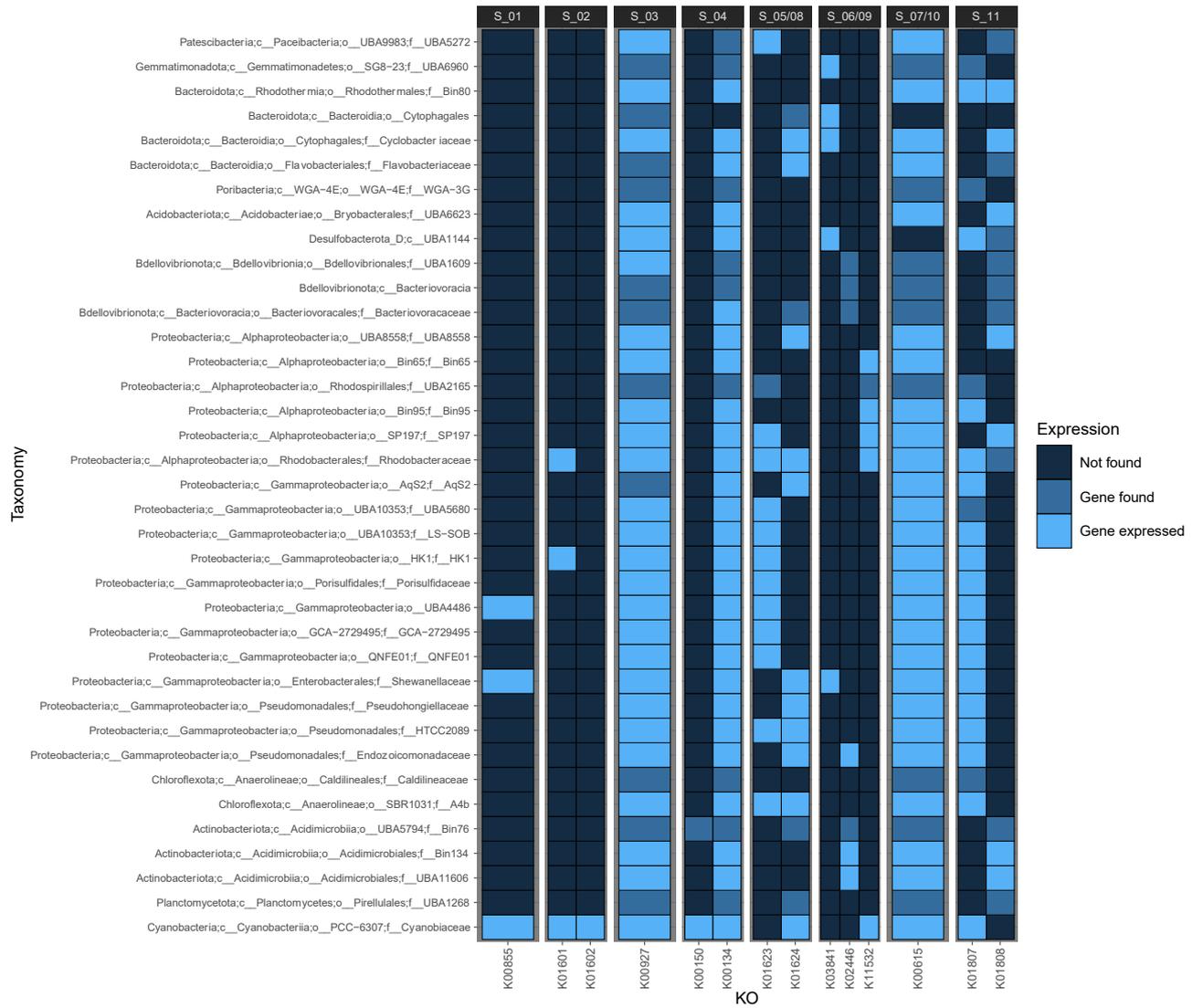


Figure S4.1b. *Ircinia microconulosa*

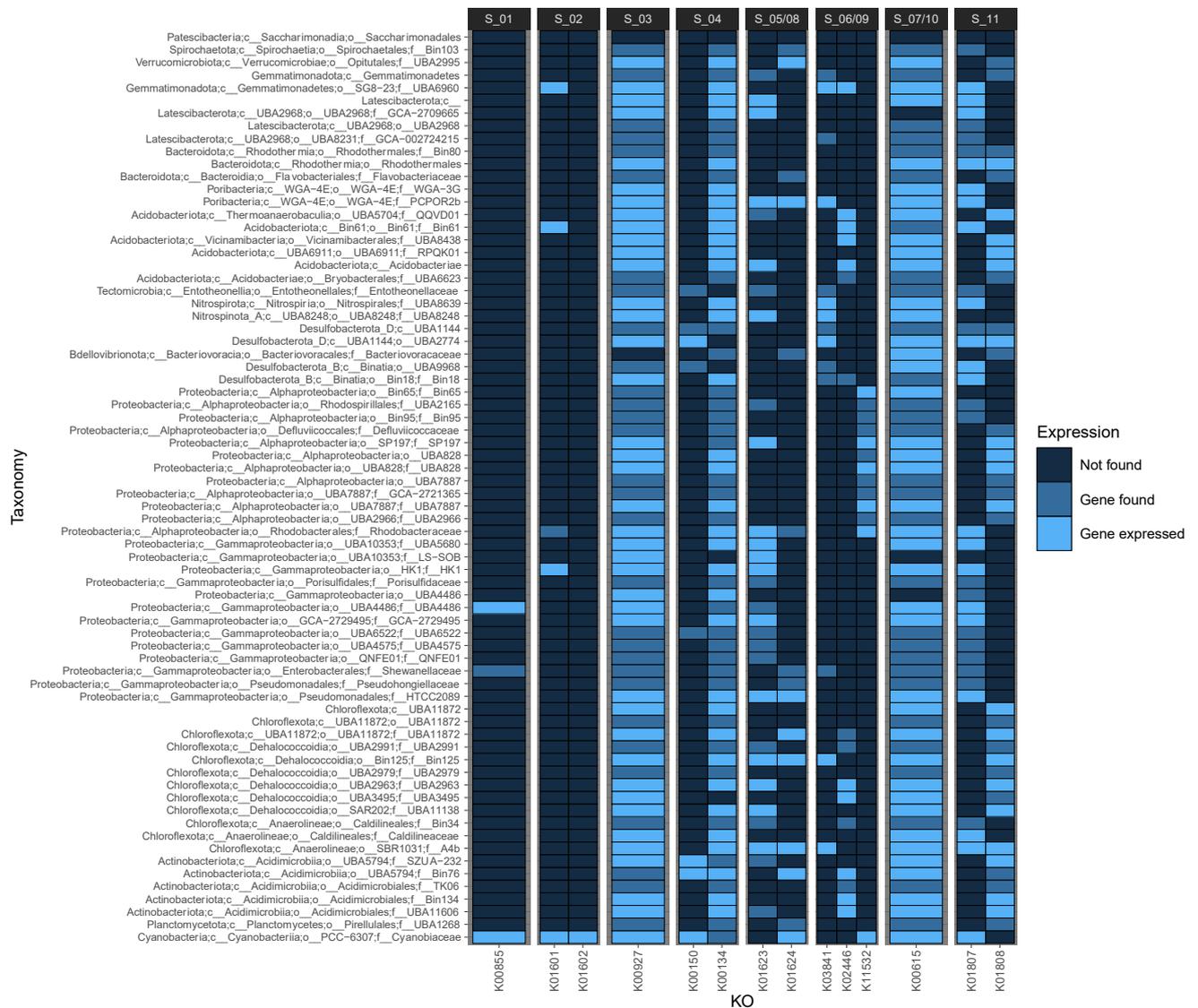


Figure S4.1. Microbial gene expression for the Calvin-Benson-Bassham cycle in a) *P. foliascens*, and b) *I. microconulosa*. Each step in the pathway is displayed in a separate column with the KEGG orthologue (KO) on the x-axis label. KOs involved in multiple steps are grouped together. Microbial taxonomy is grouped by family classification to improve readability and displayed on the y-axis. Heatmap is shown as presence-absence using three categories; Gene not found, Gene found (but not expressed), and Gene expressed.

Figure S4.1a. *Phyllospongia foliascens*

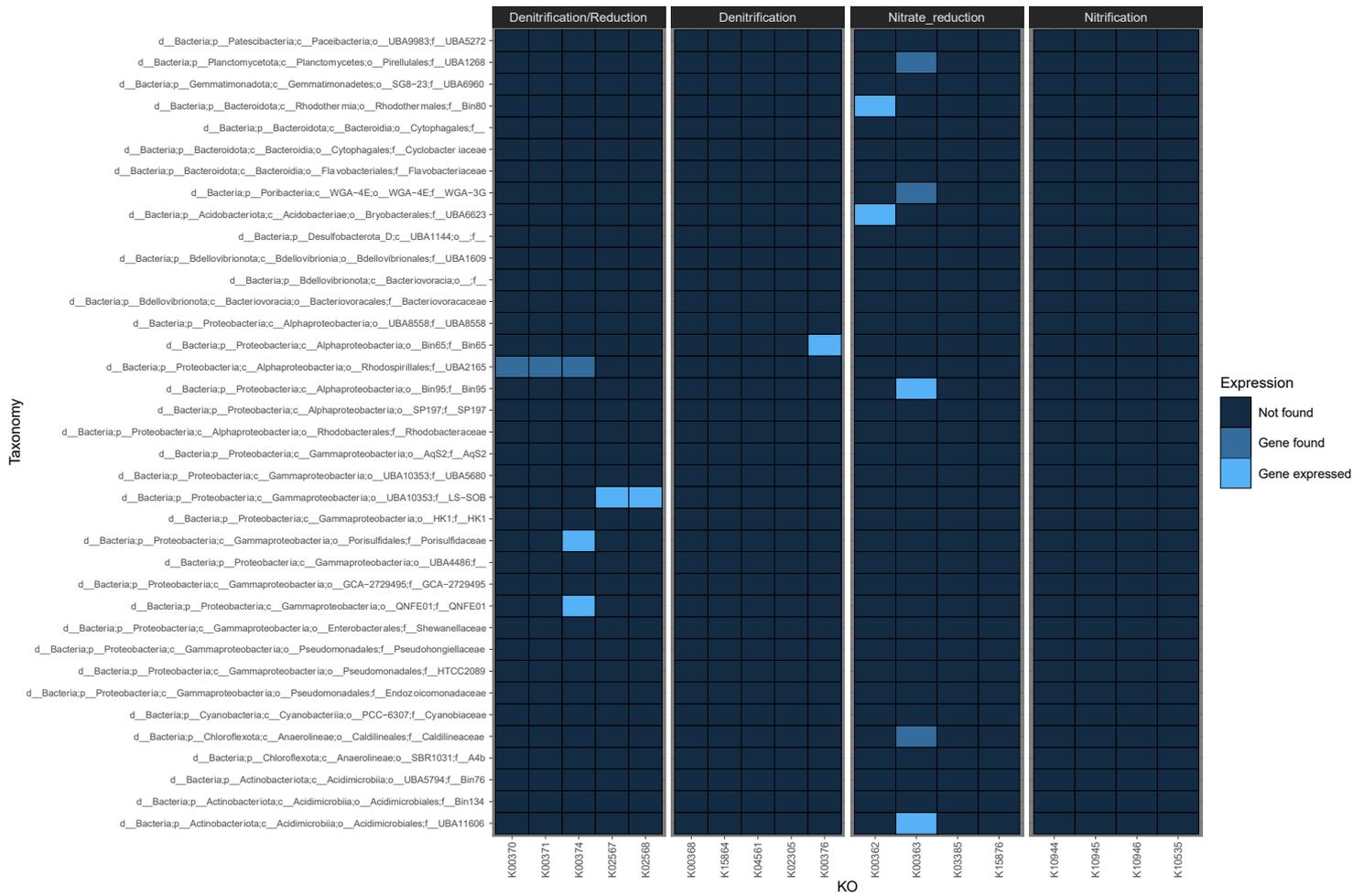


Figure S4.1b. *Ircinia microconulosa*

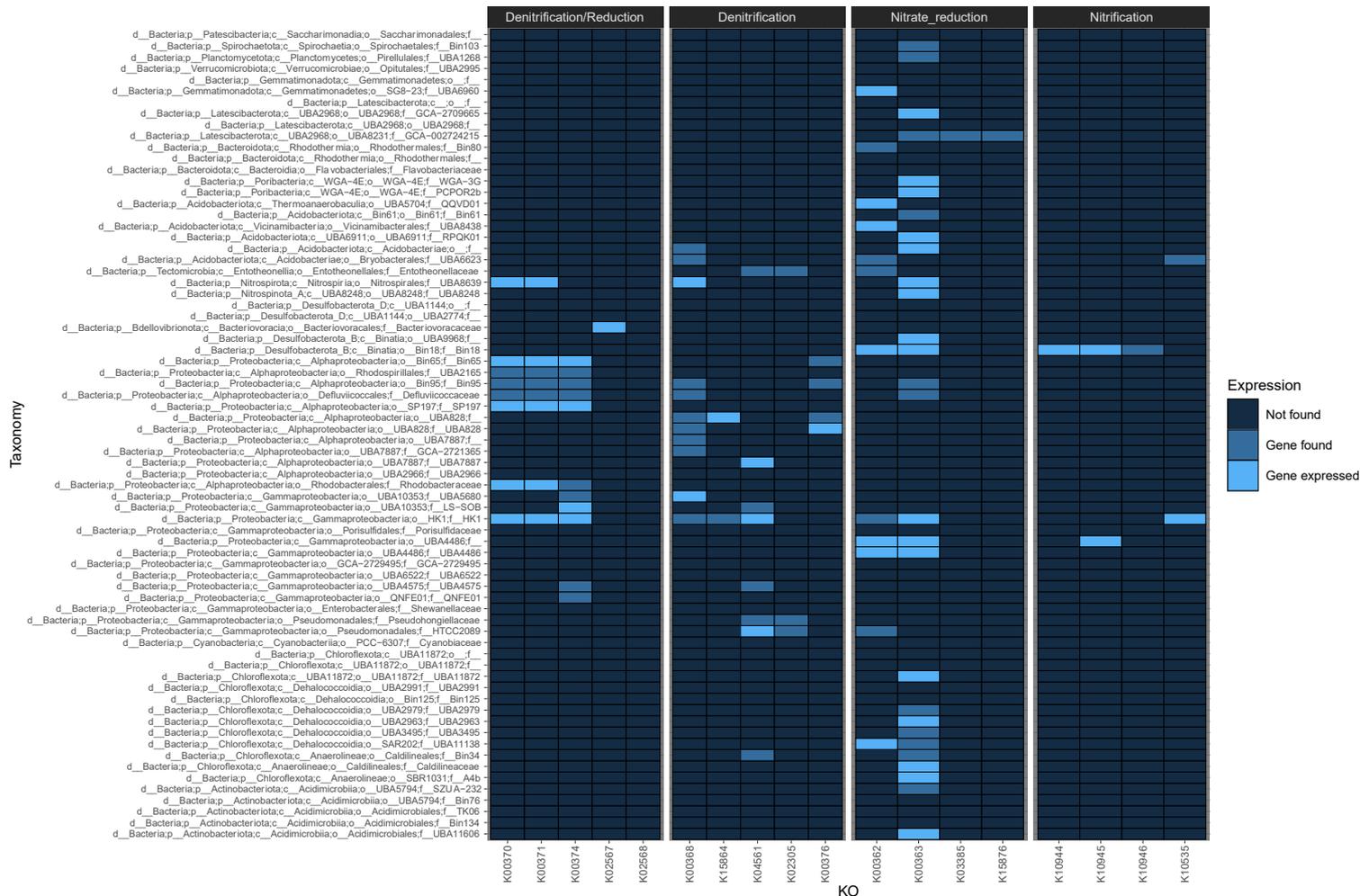


Figure S4.2. Microbial gene expression for Nitrogen metabolism in A) *P. foliascens*, and B) *I. microconulosa*. Each process of nitrogen metabolism is faceted in a separate column, with the KEGG orthologue (KO) on the x-axis label. KOs involved in multiple metabolic processes are grouped together. Microbial taxonomy is grouped by family classification to improve readability and displayed on the y-axis. Heatmap is shown as presence-absence using three categories; Gene not found, Gene found (but not expressed), and Gene expressed.

Figure S4.3a. *Phyllospongia foliascens*

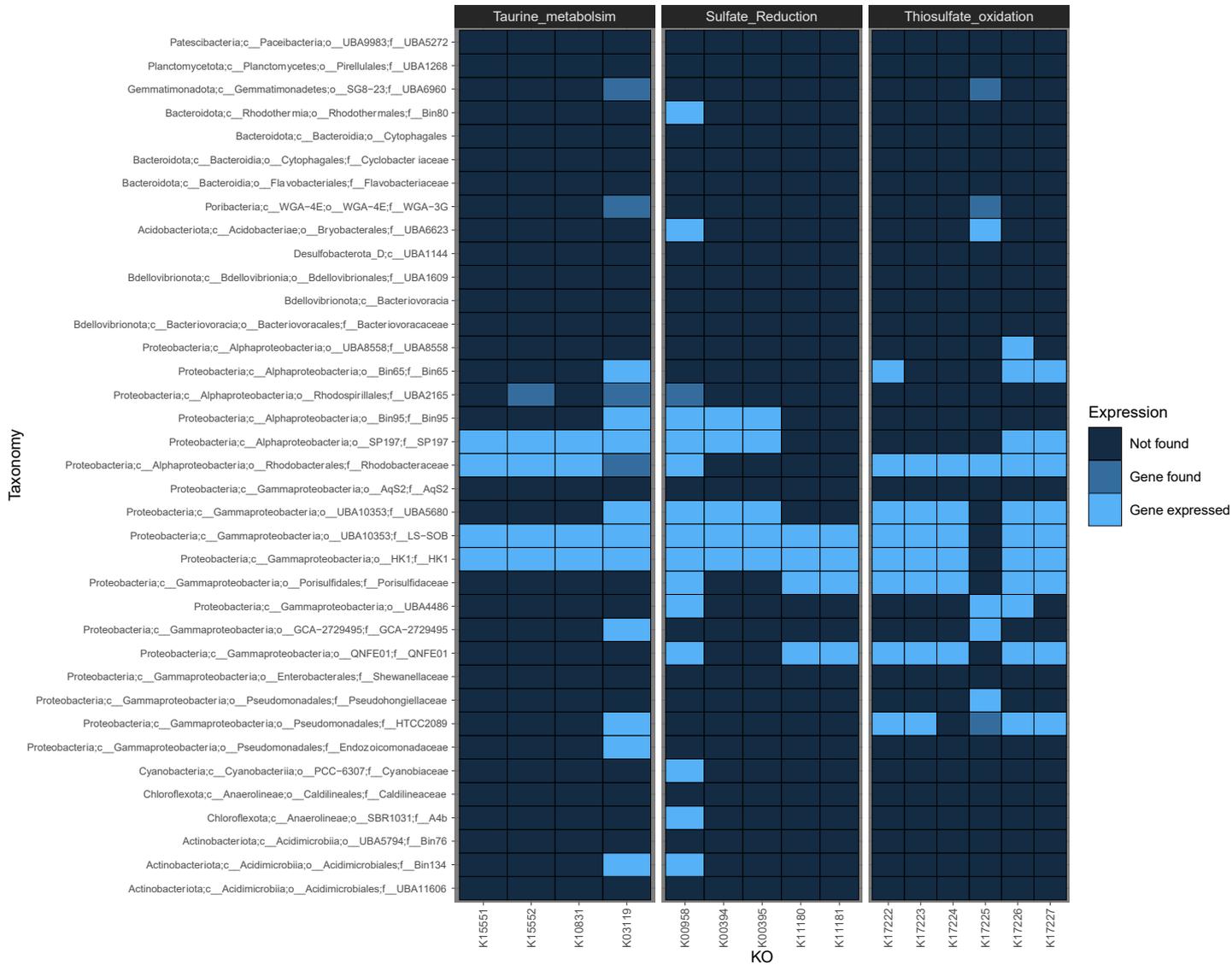


Figure S4.3b. *Ircinia microconulosa*

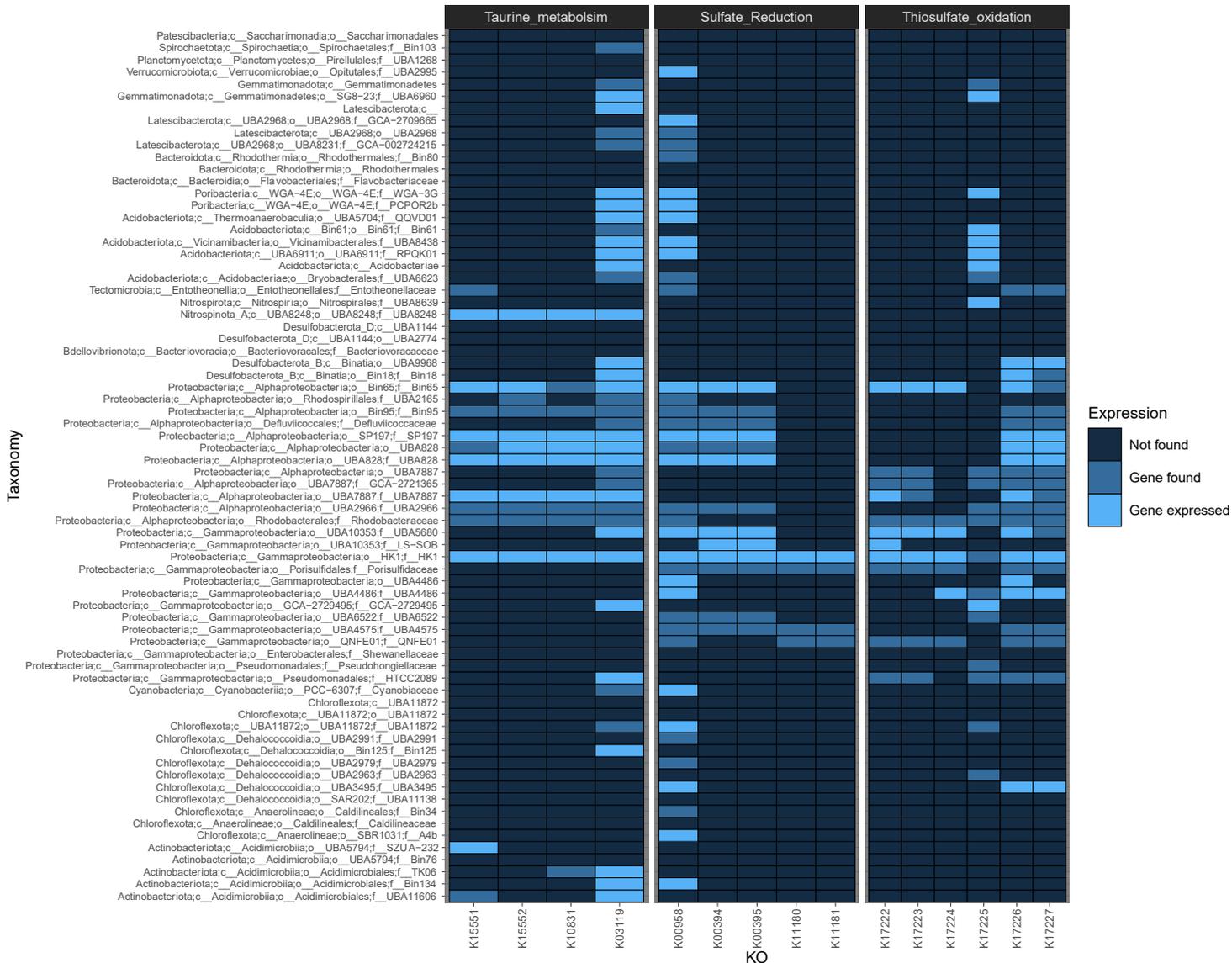


Figure S4.3. Microbial gene expression for Sulfur metabolism in A) *P. foliascens*, and B) *I. microconulosa*. Each process of Sulfur metabolism is faceted in a separate column, with the KEGG orthologue (KO) on the x-axis label. Microbial taxonomy is grouped by family classification to improve readability and displayed on the y-axis. Heatmap is shown as presence-absence using three categories; Gene not found, Gene found (but not expressed), and Gene expressed.

Figure S4.4a. *Phyllospongia foliascens*

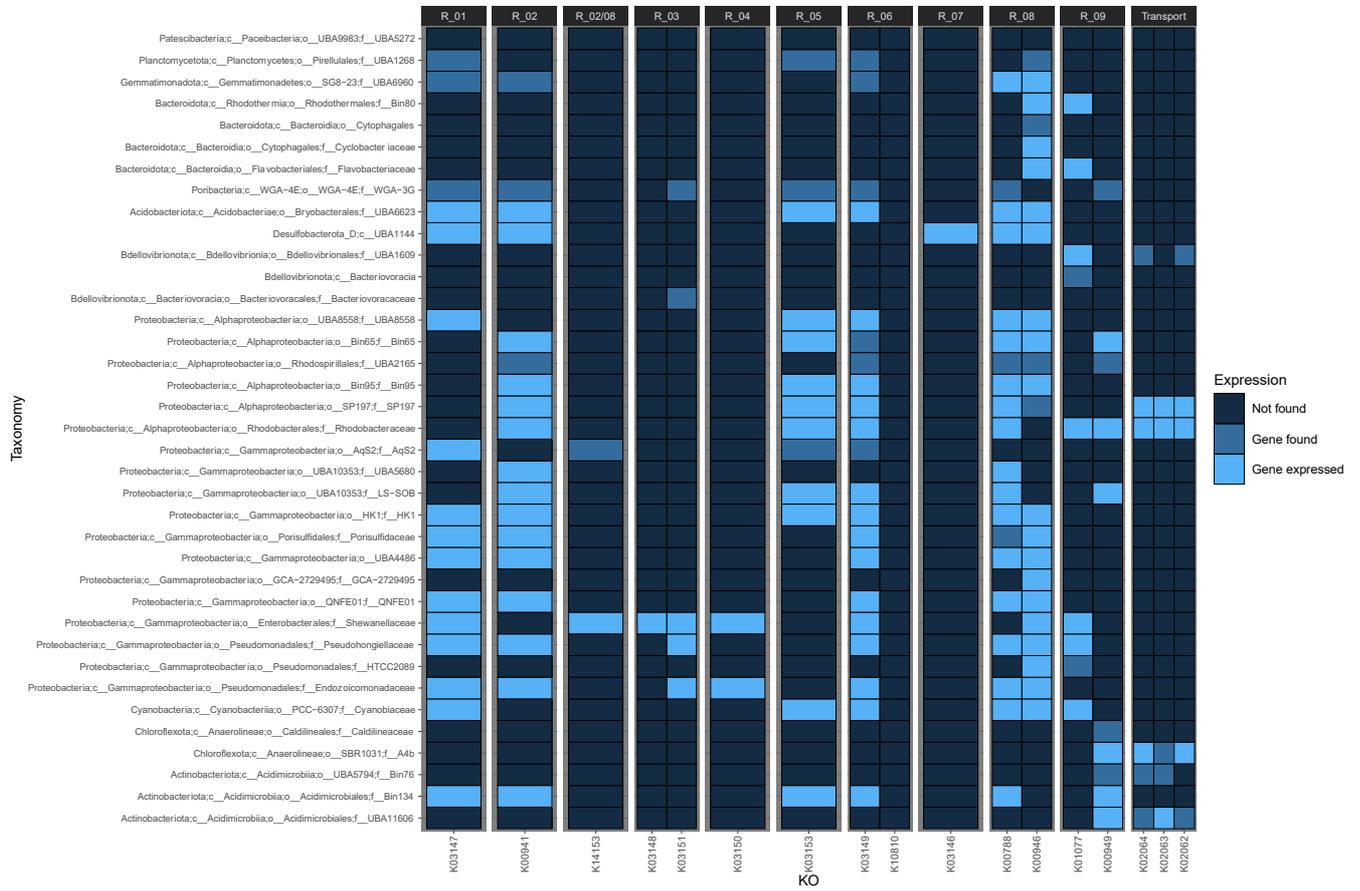


Figure S4.4b. *Ircinia microconulosa*

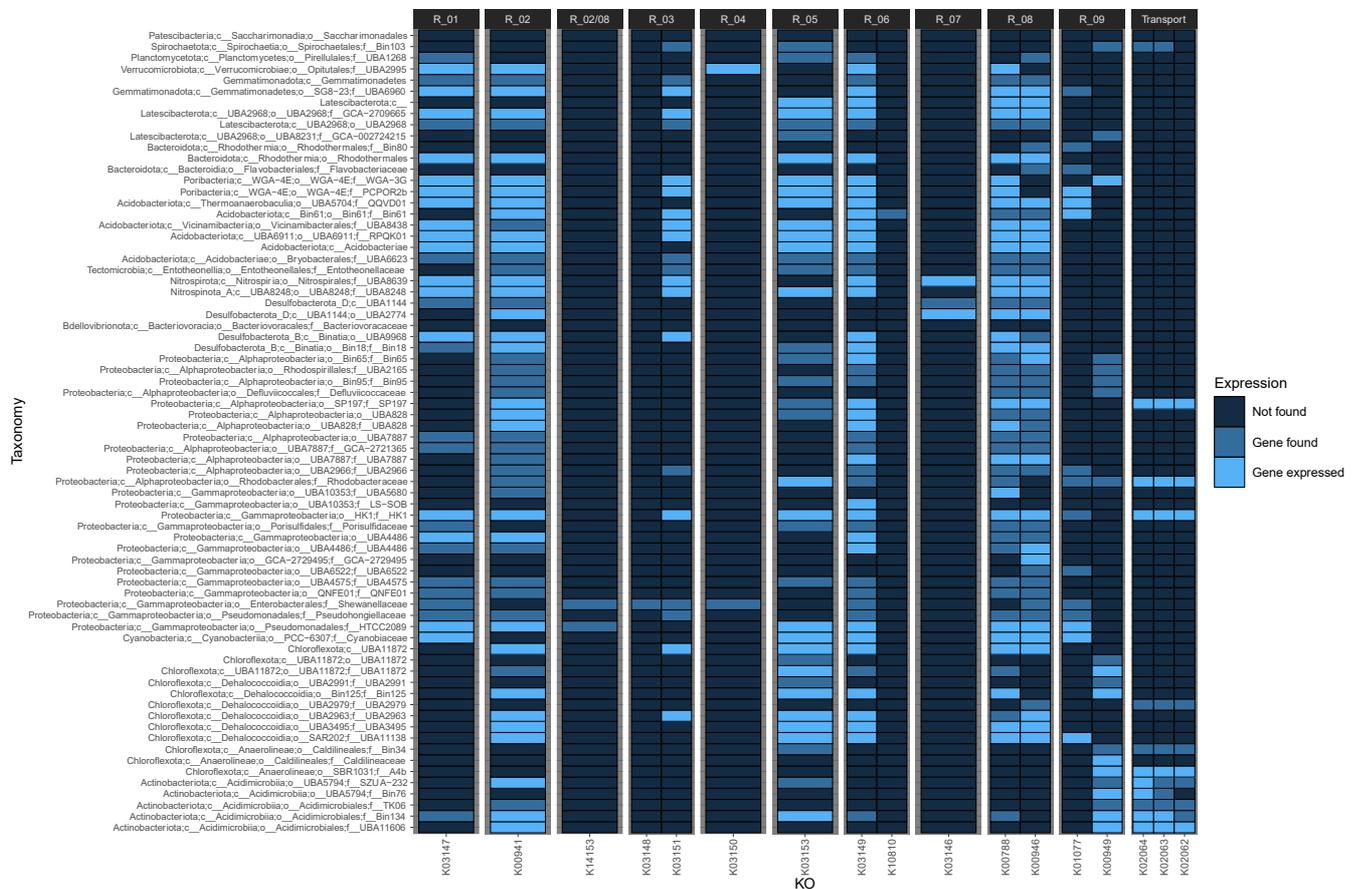


Figure S4.4. Microbial gene expression for thiamine (B₁) biosynthesis and transport in A) *P. foliascens*, and B) *I. microconulosa*. Thiamine biosynthesis involves the conversion of many reaction pathways and each reaction is facteted in a different column. KEGG orthologues (KO) are on the x-axis label and those involved in multiple reactions are grouped together. Microbial taxonomy is grouped by family classification to improve readability and displayed on the y-axis. Heatmap is shown as presence-absence using three categories; Gene not found, Gene found (but not expressed), and Gene expressed.

Figure S4.5a. *Phyllospongia foliascens*

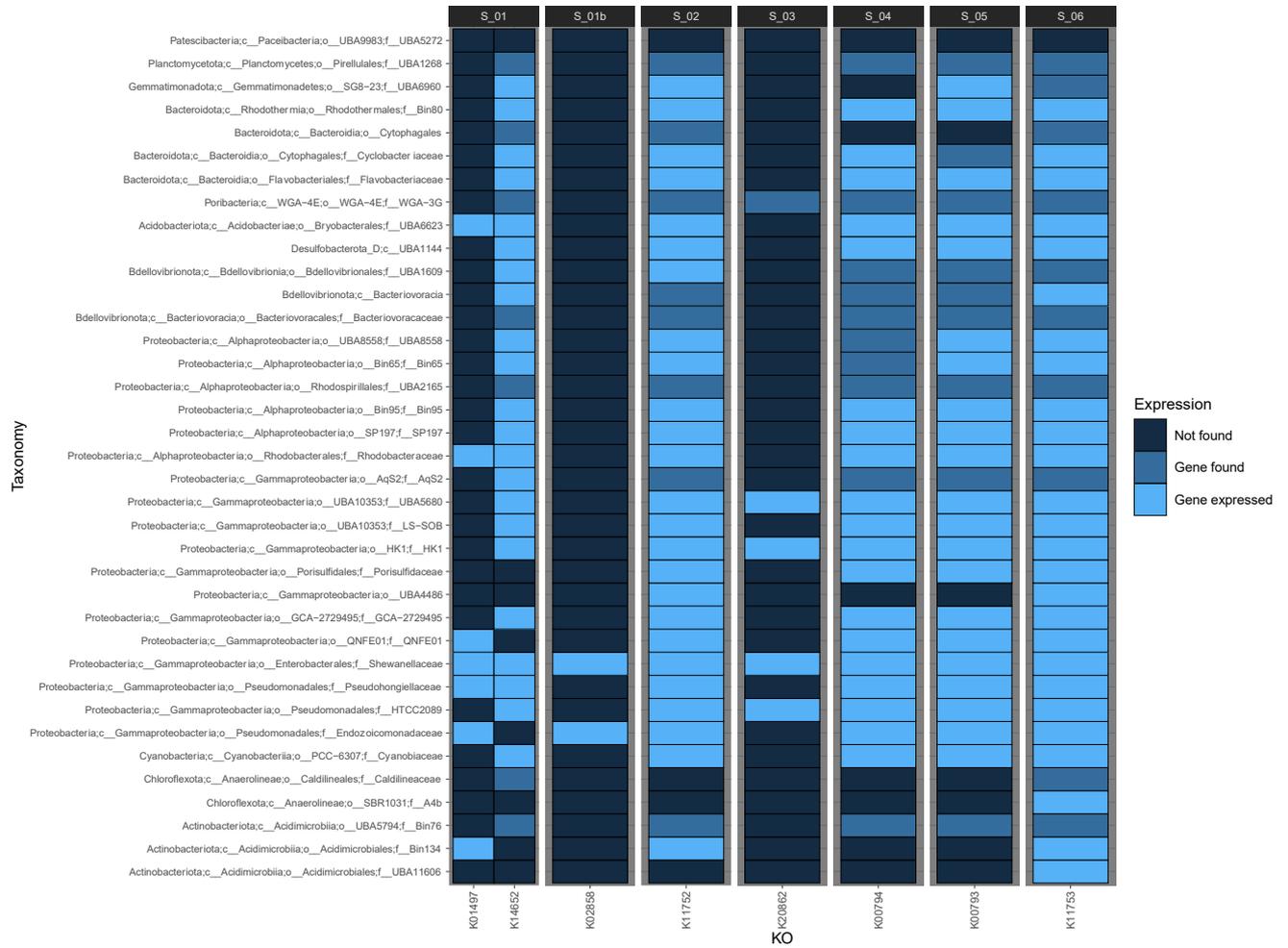


Figure S4.5b. *Ircinia microconulosa*

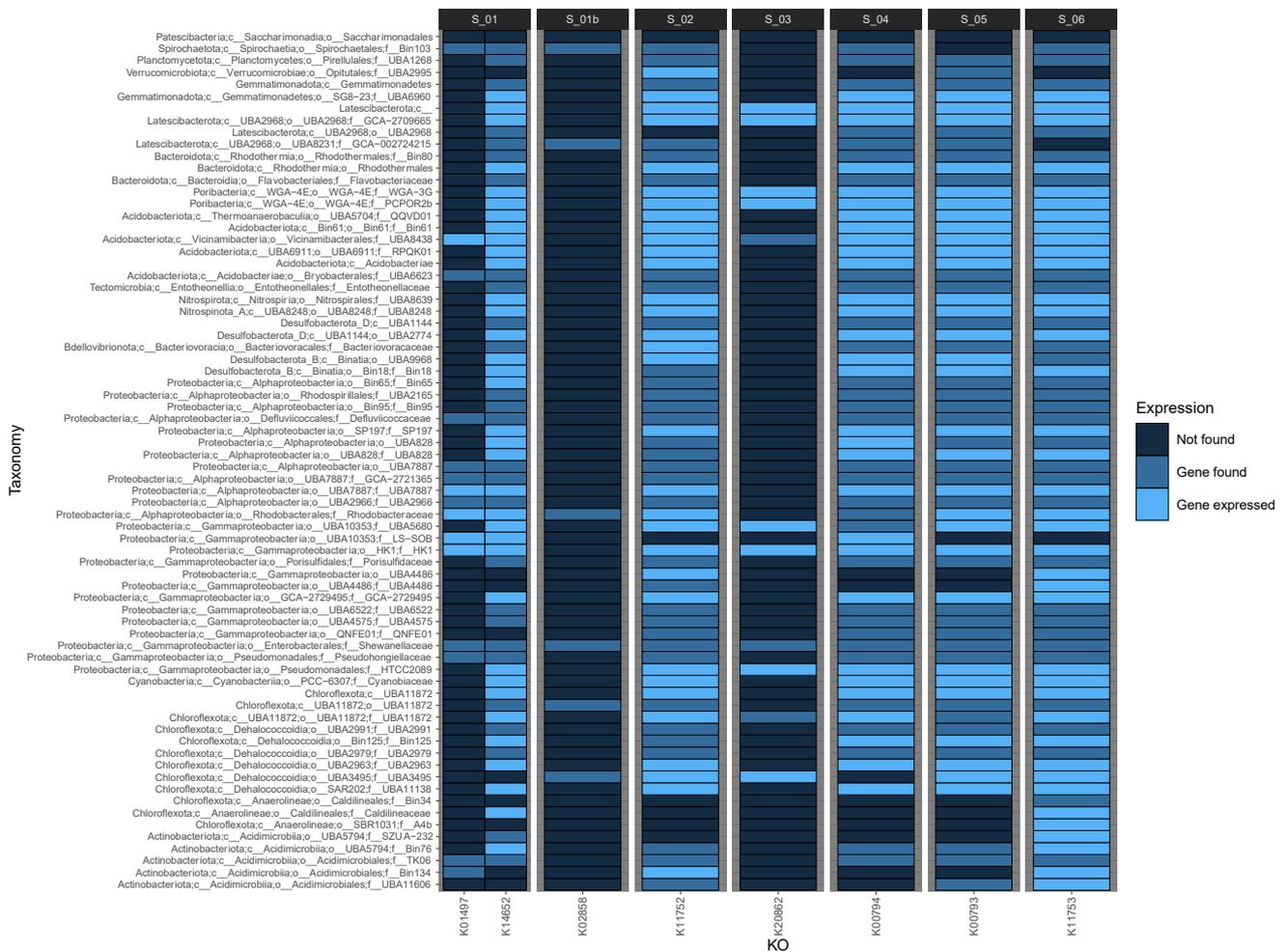


Figure S4.5. Microbial gene expression for riboflavin (B₂) biosynthesis in A) *P. foliascens*, and B) *I. microconulosa*. Each step in the pathway is displayed in a separate column with the KEGG orthologue (KO) on the x-axis label. Microbial taxonomy is grouped by family classification to improve readability and displayed on the y-axis. Heatmap is shown as presence-absence using three categories; Gene not found, Gene found (but not expressed), and Gene expressed.

Figure S4.6a. *Phyllospongia foliascens*

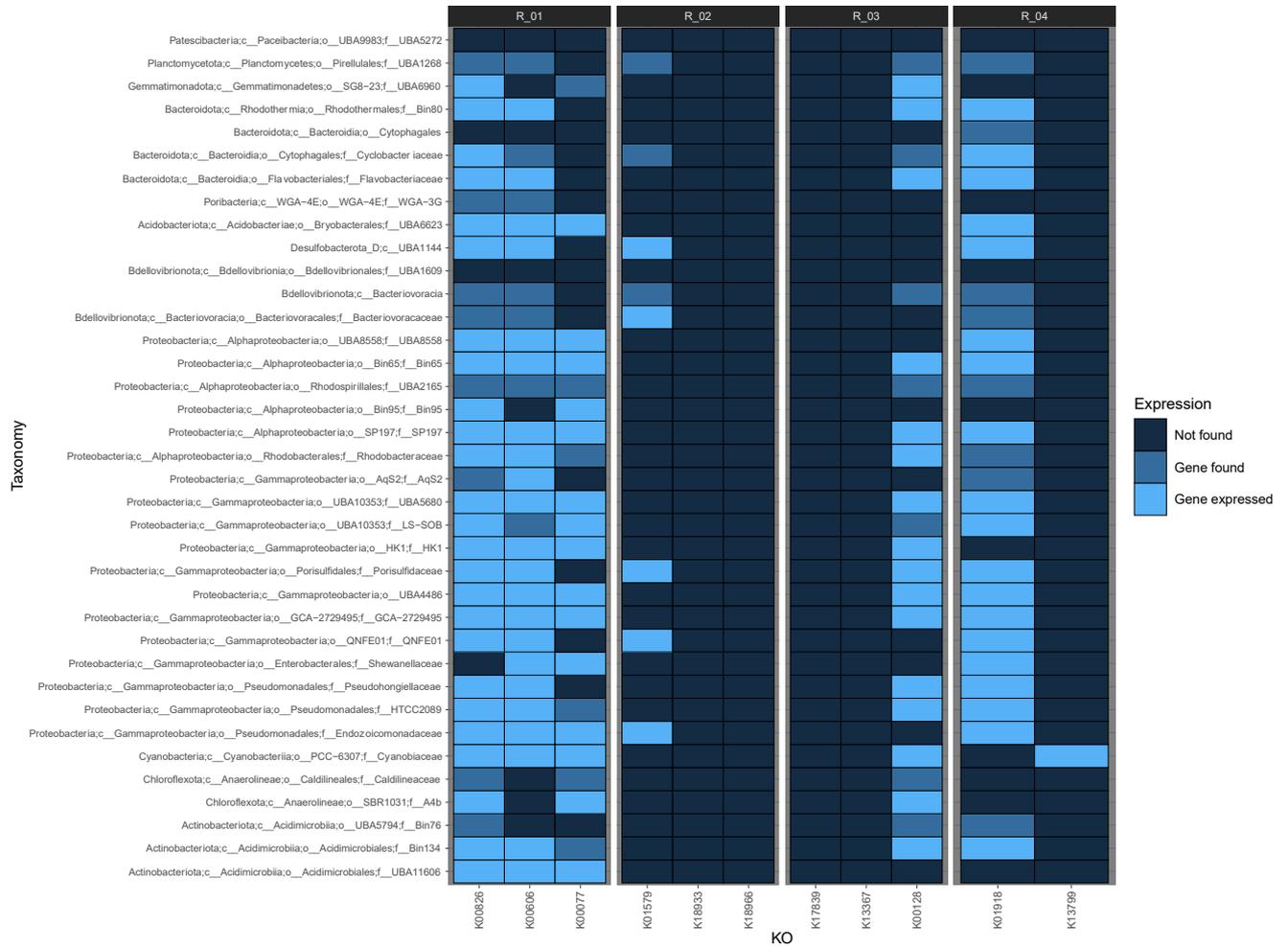


Figure S4.6b. *Ircinia microconulosa*

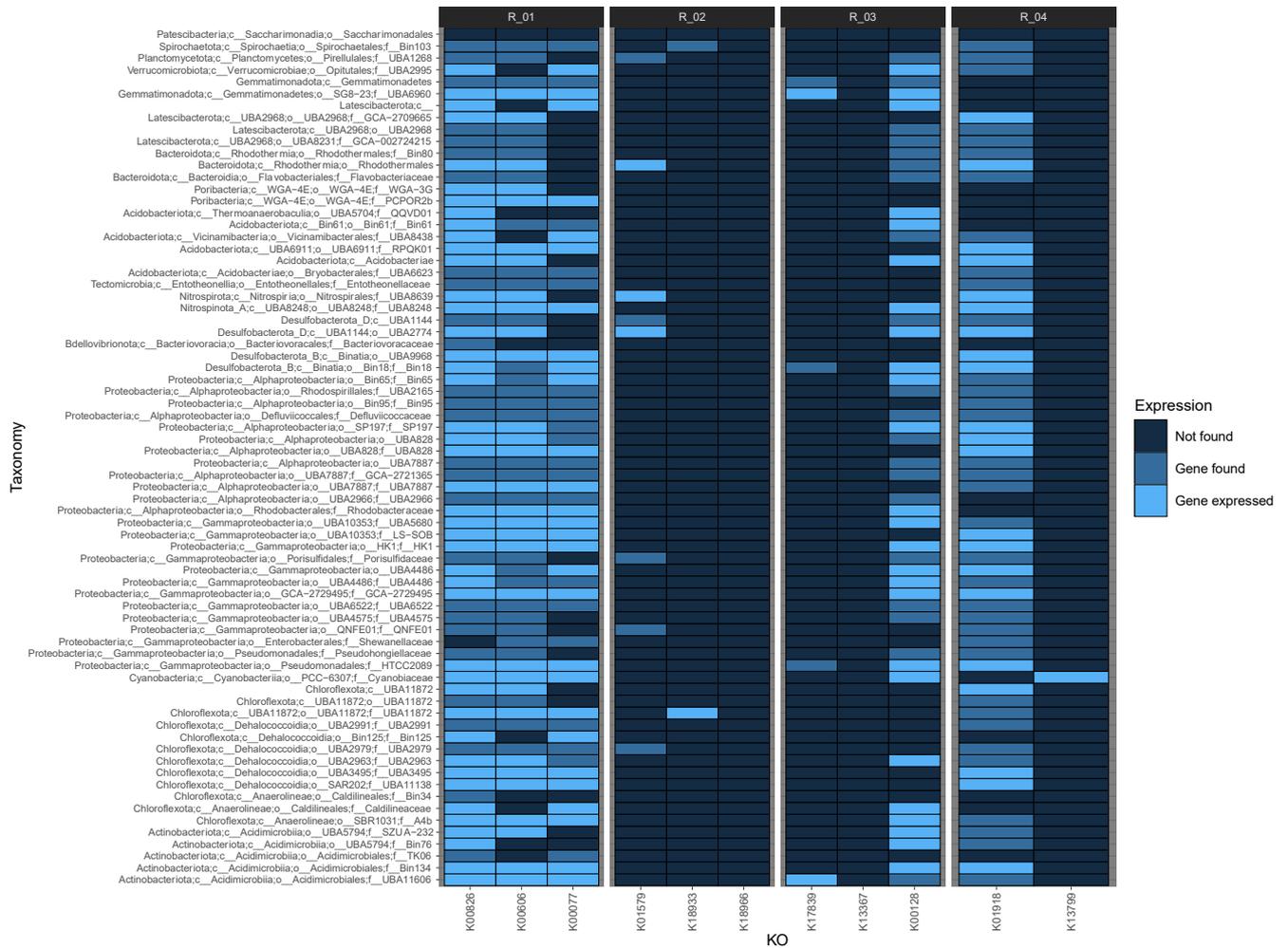


Figure S4.6. Microbial gene expression for pantothenate (B_5) biosynthesis in A) *P. foliascens*, and B) *I. microconulosa*. Pantothenate biosynthesis involves the conversion of multiple reaction pathways and each reaction is faceted in a different column. KEGG orthologues (KO) are on the x-axis label and those involved in multiple reactions are grouped together. Microbial taxonomy is grouped by family classification to improve readability and displayed on the y-axis. Heatmap is shown as presence-absence using three categories; Gene not found, Gene found (but not expressed), and Gene expressed.

Figure S4.7a. *Phyllospongia foliascens*

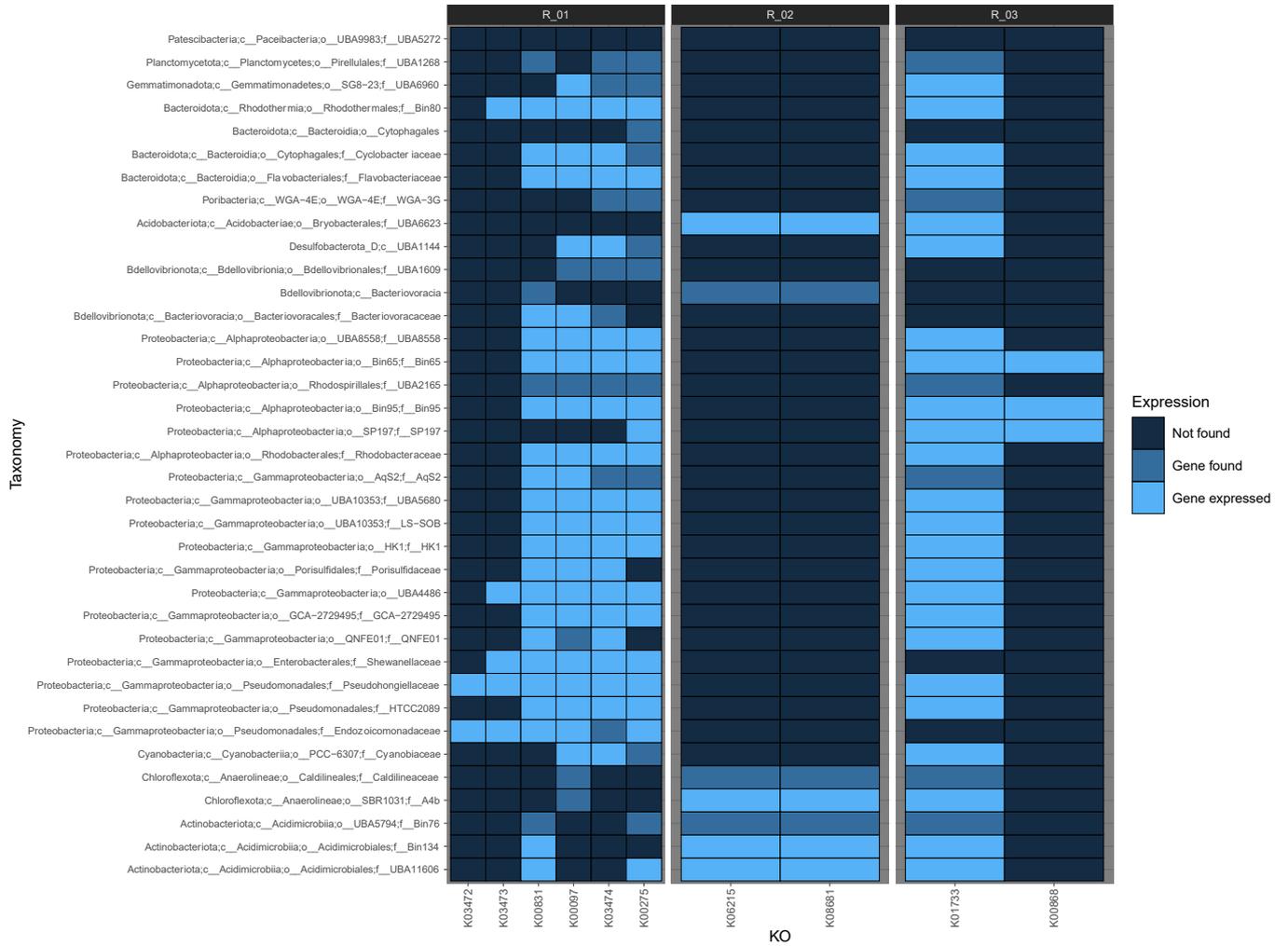


Figure S4.7b. *Ircinia microconulosa*

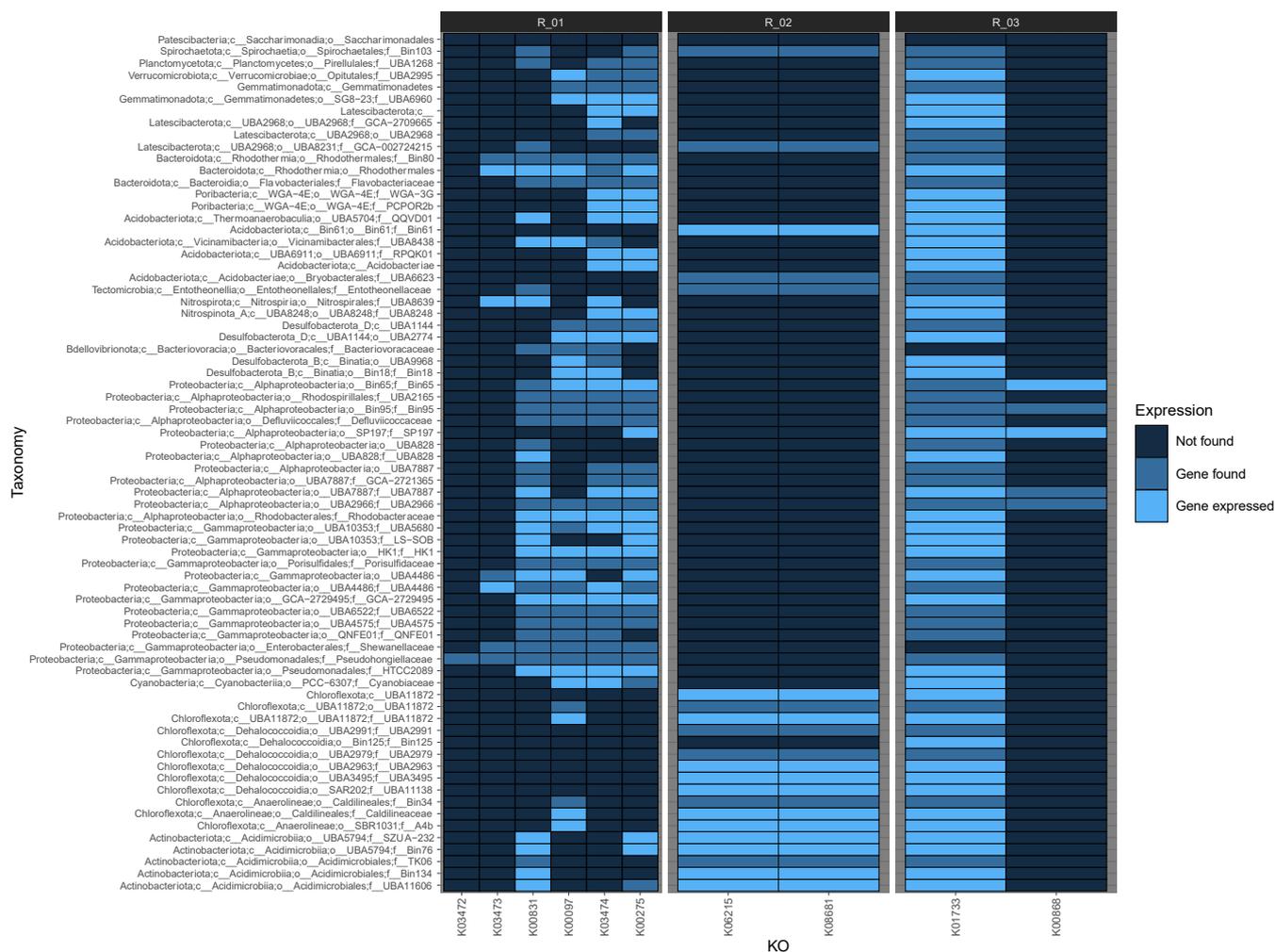


Figure S4.7. Microbial gene expression for pyridoxine (B₆) biosynthesis in A) *P. foliascens*, and B) *I. microconulosa*. Pyridoxine biosynthesis can occur through multiple reaction pathways and each reaction is faceted in a different column. KEGG orthologues (KO) are on the x-axis label and those involved in multiple reactions are grouped together. Microbial taxonomy is grouped by family classification to improve readability and displayed on the y-axis. Heatmap is shown as presence-absence using three categories; Gene not found, Gene found (but not expressed), and Gene expressed.

Figure S4.8a. *Phyllospongia foliascens*

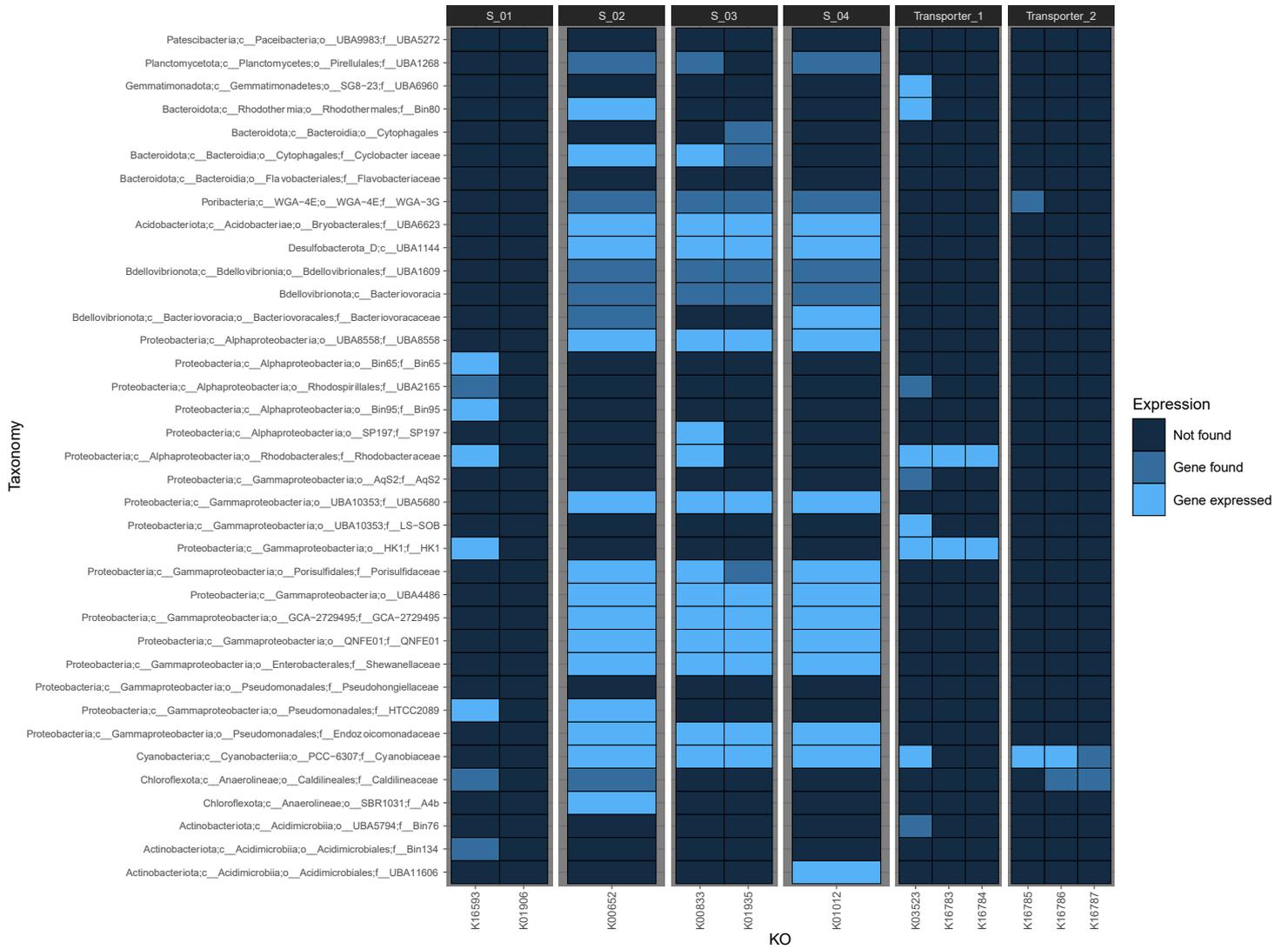


Figure S4.8b. *Ircinia microconulosa*

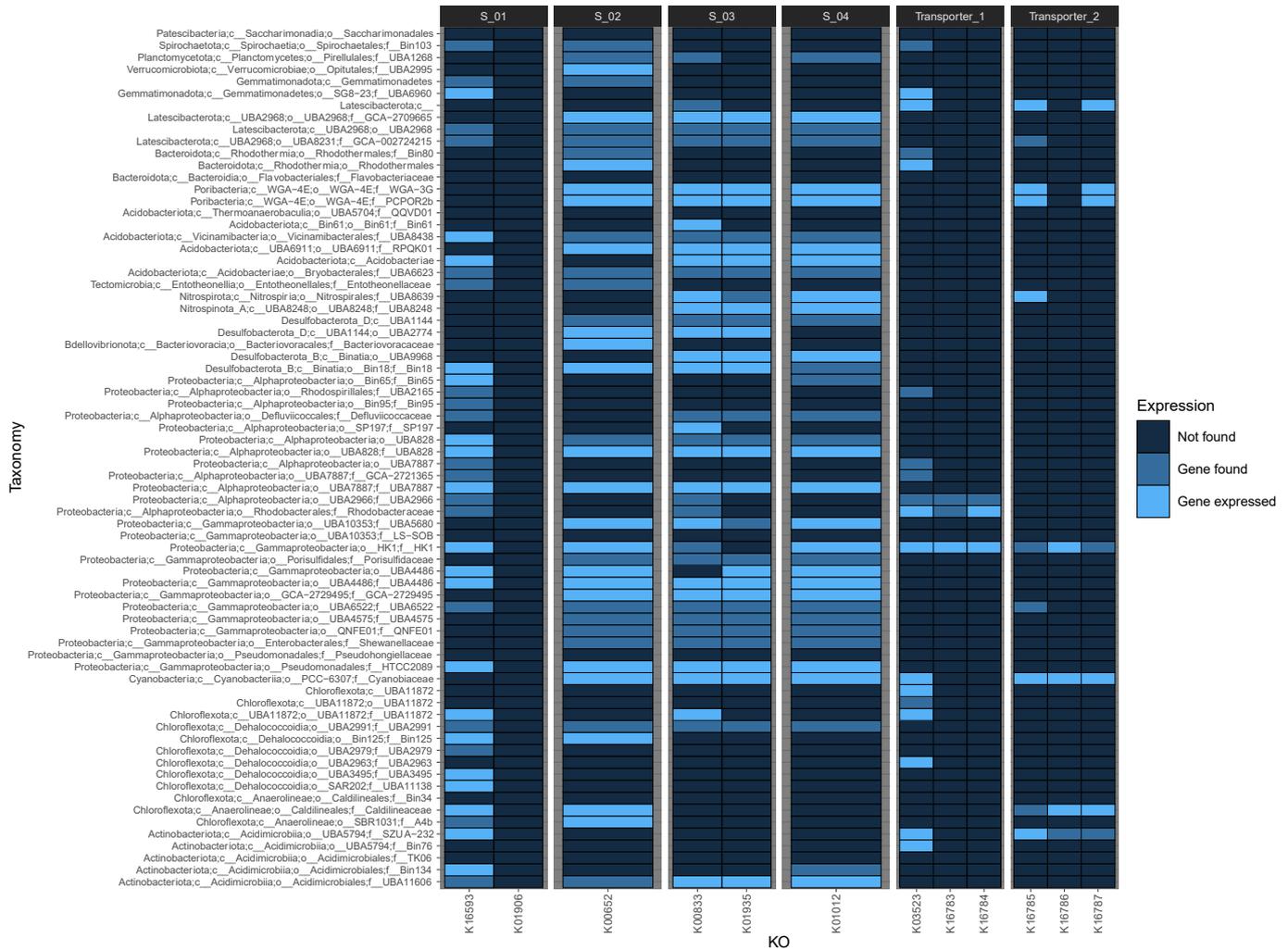


Figure S4.8. Microbial gene expression for biotin (B₇) biosynthesis and transport in A) *P. foliascens*, and B) *I. microconulosa*. Each step in the pathway is displayed in a separate column with the KEGG orthologue (KO) on the x-axis label. Microbial taxonomy is grouped by family classification to improve readability and displayed on the y-axis. Heatmap is shown as presence-absence using three categories; Gene not found, Gene found (but not expressed), and Gene expressed.

Figure S4.9a. *Phyllospongia foliascens*

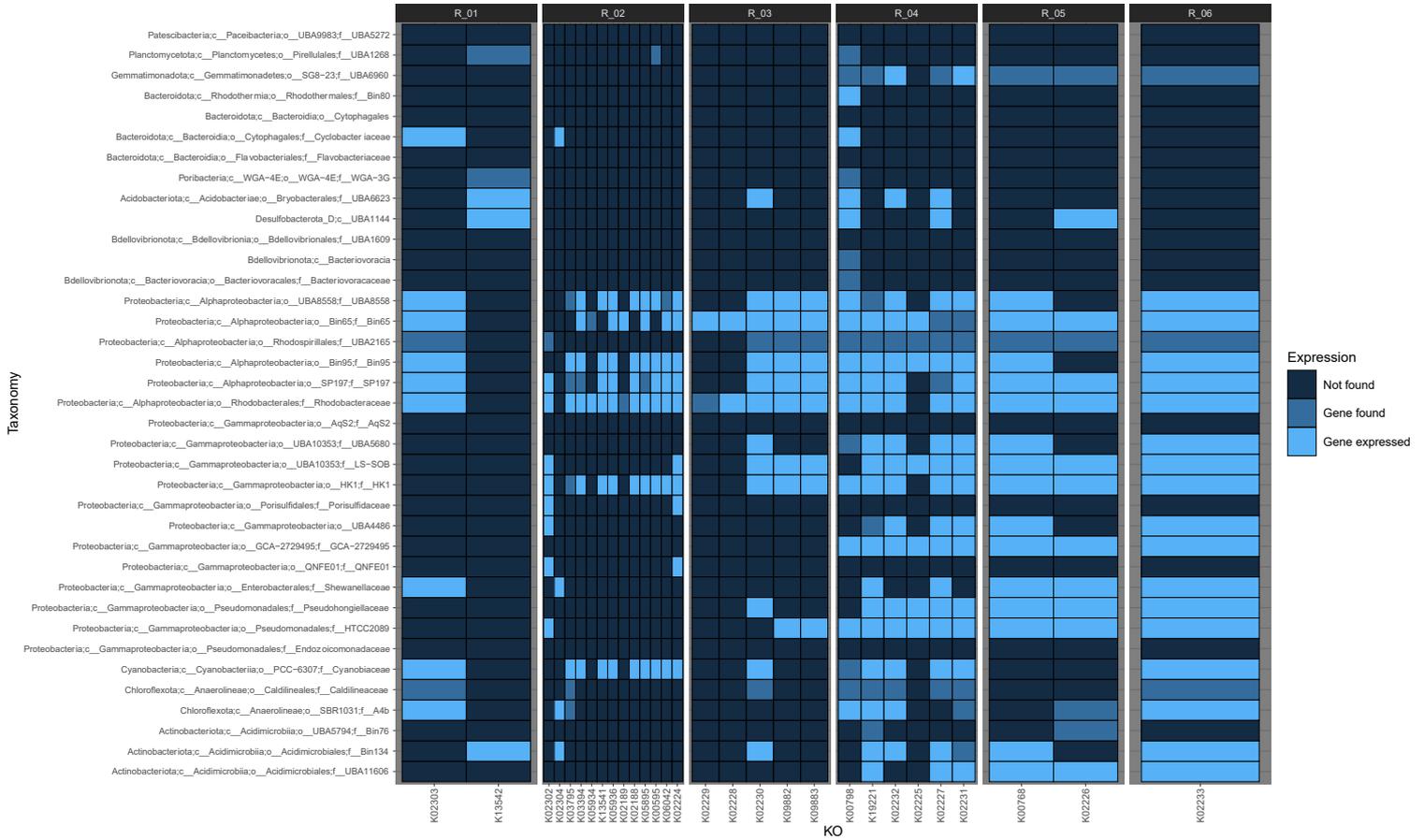


Figure S4.9b. *Ircinia microconulosa*

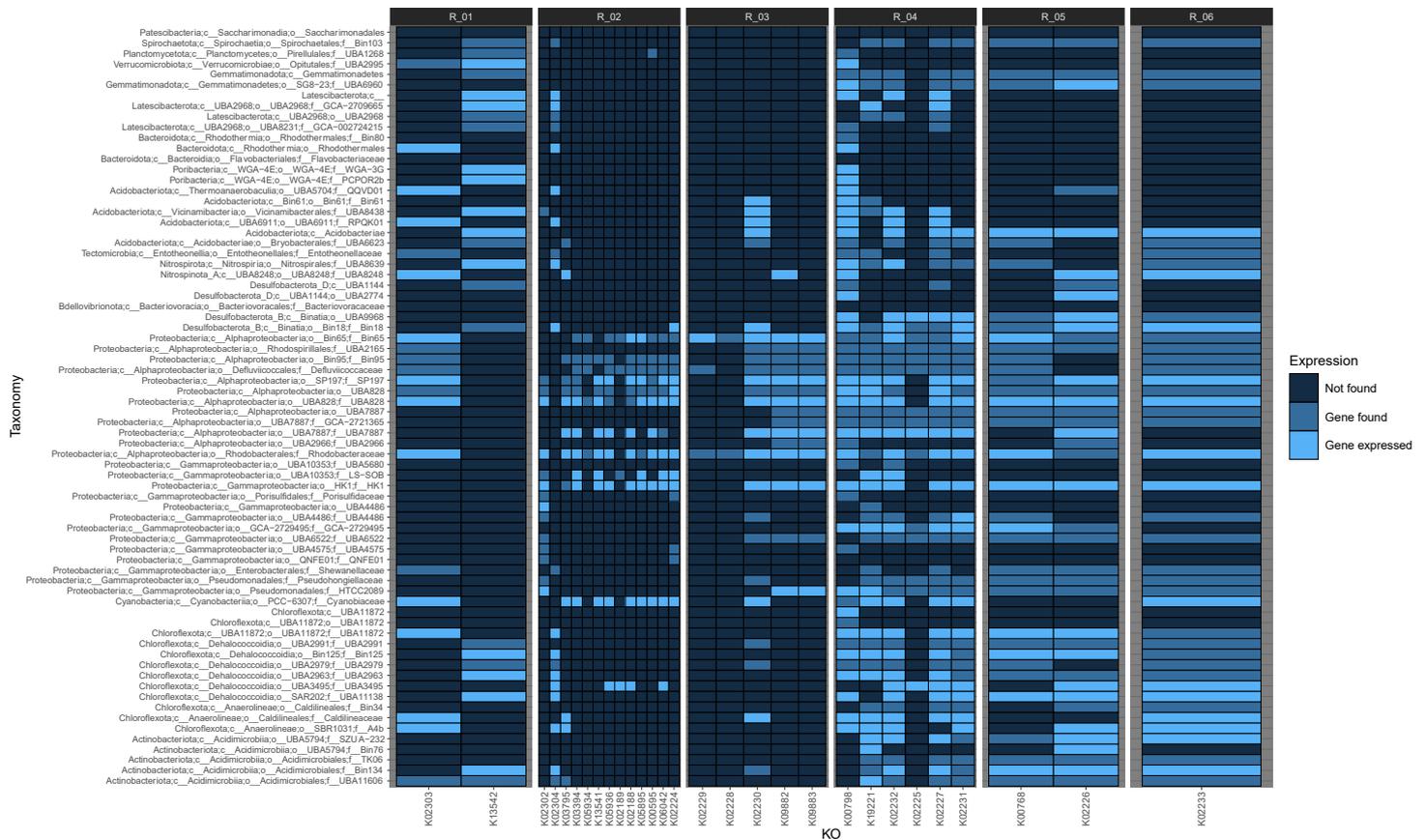


Figure S4.9. Microbial gene expression for cobalamin (B₁₂) biosynthesis in A) *P. foliascens*, and B) *I. microconulosa*. Cobalamin biosynthesis involves the conversion of multiple reaction pathways and each reaction is faceted in a different column. KEGG orthologues (KO) are on the x-axis label and those involved in multiple reactions are grouped together. Microbial taxonomy are grouped together. Heatmap is shown as presence-absence using three categories; Gene not found, Gene found (but not expressed), and Gene expressed.

Appendix D. Supporting tables and figures for chapter 5

Table S5.1. Genome characteristics and taxonomy for Endozoicomonadaceae

MAG_id	Group	Isolation_source	Completeness	Contamination	Genome_size_Mbp	GC	Classification_GTDB	Classification_16S
577_metabat1_sensitive.053	Sponge	Sponge	83.81	1.44	3.65	36.9	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Endozoicomonadaceae;g__s__	Bacteria;_Proteobacteria;_Gammaproteobacteria;_Oceanospirillales;_Endozoicomonadaceae;_Kistimonas
GCA_002084115.1	Non-Sponge	Bivalve	92.94	1.51	3.30	52.1	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Endozoicomonadaceae;g__LUC14-002-19-P2;s__LUC14-002-19-P2sp002084115	Bacteria;_Proteobacteria;_Gammaproteobacteria;_Oceanospirillales;_Endozoicomonadaceae;_Kistimonas
GCA_002238585.1	Sponge	Sponge	80.63	0	3.78	57.3	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Endozoicomonadaceae;g__Kistimonas;s__Kistimonas sp002238585	Bacteria;_Proteobacteria;_Gammaproteobacteria;_Oceanospirillales;_Endozoicomonadaceae;_Endozoicomonas
GCA_900299555.1	Non-Sponge	Fish_bream	98.61	5.11	5.88	46.8	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Endozoicomonadaceae;g__Endozoicomonas;s__Endozoicomonas sp900299555	Bacteria;_Proteobacteria;_Gammaproteobacteria;_Oceanospirillales;_Endozoicomonadaceae;_Endozoicomonas
GCF_000710775.1	Non-Sponge	Sea_slug	98.98	3.43	5.61	46.8	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Endozoicomonadaceae;g__Endozoicomonas;s__Endozoicomonas elysicola	Bacteria;_Proteobacteria;_Gammaproteobacteria;_Oceanospirillales;_Endozoicomonadaceae;_Endozoicomonas
GCF_000722635.1	Sponge	Sponge	99.14	1.66	6.34	47	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Endozoicomonadaceae;g__Endozoicomonas;s__Endozoicomonas numazuensis	Bacteria;_Proteobacteria;_Gammaproteobacteria;_Oceanospirillales;_Endozoicomonadaceae;_Endozoicomonas
GCF_001562015.1	Sponge	Sponge	99.14	2.05	6.45	47.7	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Endozoicomonadaceae;g__Endozoicomonas;s__Endozoicomonas arenosclerae	Bacteria;_Proteobacteria;_Gammaproteobacteria;_Oceanospirillales;_Endozoicomonadaceae;_Endozoicomonas;_uncultured_Spongiobacter_sp.
GCF_001583435.1	Non-Sponge	Coral	98.99	1.72	5.43	48.5	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Endozoicomonadaceae;g__Endozoicomonas;s__Endozoicomonas montiporae	Bacteria;_Proteobacteria;_Gammaproteobacteria;_Oceanospirillales;_Endozoicomonadaceae;_Endozoicomonas
GCF_001646945.1	Non-Sponge	Ascidian	98.28	1.81	6.13	46.7	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Endozoicomonadaceae;g__Endozoicomonas;s__Endozoicomonas ascidiicola	Bacteria;_Proteobacteria;_Gammaproteobacteria;_Oceanospirillales;_Endozoicomonadaceae;_Endozoicomonas;_uncultured_bacterium_graftm_1117
GCF_001647025.1	Non-Sponge	Bivalve	98.92	3.43	6.69	47.9	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Endozoicomonadaceae;g__Endozoicomonas;s__Endozoicomonas	Bacteria;_Proteobacteria;_Gammaproteobacteria;_Oceanospirillales;_Endozoicomonadaceae;_Endozoicomonas;_uncultured_bacterium_graftm_1117
GCF_002864045.1	Non-Sponge	Coral	98.56	1.39	6.05	49.2	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Endozoicomonadaceae;g__Endozoicomonas;s__Endozoicomonas acroporae	Bacteria;_Proteobacteria;_Gammaproteobacteria;_Oceanospirillales;_Endozoicomonadaceae;_Endozoicomonas;_uncultured_bacterium_graftm_1117
GCF_004340525.1	Outgroup	Unknown	99.52	1.32	3.91	62.2	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Halieaceae;g__Chromatococcus;s__Chromatococcus halotolerans	Bacteria;_Proteobacteria;_Gammaproteobacteria;_Cyllivibrionales;_Halieaceae;_Chromatococcus
GCF_004762125.1	Sponge	Sponge	99.14	1.7	5.49	44.9	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Endozoicomonadaceae;g__Endozoicomonas;s__Endozoicomonas sp004762125	Bacteria;_Proteobacteria;_Gammaproteobacteria;_Oceanospirillales;_Endozoicomonadaceae;_Endozoicomonas;_uncultured_bacterium_graftm_1117
GCF_900174585.1	Non-Sponge	no_source	99.57	2.64	5.47	51.5	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Endozoicomonadaceae;g__Parendozoicomonas;s__Parendozoicomonas haliclonae	Bacteria;_Proteobacteria;_Gammaproteobacteria;_Oceanospirillales;_Endozoicomonadaceae;_Endozoicomonas

Table S5.2. Genome characteristics and taxonomy for Microtrichaceae

MAG_id	Group	Isolation_source	Completeness	Contamination	Genome_size_Mbp	GC	Classification_GTDB	Classification_16S
501_metabat1_sensitive.014_sub	Sponge	Sponge	89.32	1.28	3.545251	67.9	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_bacterium_graftm_464
501_metabat1_sensitive.017	Sponge	Sponge	94.79	1.28	3.387132	65.3	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__Bin134;s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group
502_metabat1_super.032	Sponge	Sponge	89.46	1.28	3.024501	64.3	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__Bin134;s__	No_16S
504_concoct.040	Sponge	Sponge	96.58	2.14	3.69662	67	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__TK06;g__s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_actinobacteriu
504_metabat2.005	Sponge	Sponge	90.6	1.28	2.856713	64.2	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__Bin134;s__	No_16S
504_metabat2.040	Sponge	Sponge	89.74	1.28	2.974489	70.1	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_actinobacteriu m_graftm_64
505_concoct.072	Sponge	Sponge	94.02	1.28	2.233094	54.5	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__Bin134;s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group
508_concoct.060_sub	Sponge	Sponge	89.24	2.14	3.721398	69.4	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_actinobacteriu m_graftm_64
510_concoct.062	Sponge	Sponge	95.73	0.85	3.404532	64.6	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__Bin134;s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group
512_concoct.021_sub	Sponge	Sponge	88.75	2.14	3.435052	69.1	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__s__	No_16S
515_metabat1_super.047	Sponge	Sponge	96.58	1.28	2.926443	61.4	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__UBA11606;s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_bacterium_graftm_464
575_metabat1_super.017_sub	Sponge	Sponge	87.18	4.84	3.677933	68.8	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__s__	No_16S
575_metabat1_super.020	Sponge	Sponge	94.02	1.28	3.492288	64.5	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__Bin134;s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group
576_metabat1_sensitive.008	Sponge	Sponge	93.16	1.28	2.843642	66.6	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__Bin134;s__	No_16S
APA_bin_18	Sponge	Sponge	92.31	1.36	3.791969	69.3	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group
APA_bin_4	Sponge	Sponge	98.29	2.99	3.958302	66.4	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__TK06;g__s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_bacterium_graftm_464
APA_bin_53	Sponge	Sponge	96.58	1.28	3.693405	63.5	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__UBA11606;s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_bacterium_graftm_464
APA_bin_68	Sponge	Sponge	92.74	2.42	2.513536	55.4	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__UBA11606;s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_bacterium_graftm_464
APA_bin_7	Sponge	Sponge	93.16	2.99	4.327641	69.5	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__s__	No_16S
CAR3_bin_3	Sponge	Sponge	92.31	1.28	2.56484	66	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_actinobacteriu m_graftm_64
COS1_bin_5	Sponge	Sponge	95.73	1.28	3.907954	69.8	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_bacterium_graftm_464
COS1_bin_6	Sponge	Sponge	95.73	1.28	3.506906	64.4	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__Bin134;s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group

Table S5.2. Continued

COS36386_bin_14	Sponge	Sponge	97.44	2.14	3.319196	67.4	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__TK06;g__s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_actinobacteriu
COS3_bin_16	Sponge	Sponge	86.61	2.52	3.082623	70.6	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group
COS4_bin_1	Sponge	Sponge	95.73	1.28	3.043884	63	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__UBA11606;s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_bacterium_graftm_464
GCA_000817105.1	Non-Sponge	Seawater	90.17	0.85	2.109986	51.4	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__TK06;g__MedAcidi-G3;s__MedAcidi-G3 sp000817105	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_bacterium_graftm_464
GCA_002239105.1	Sponge	Sponge	96.58	0.43	3.684518	64.3	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__Bin134;s__Bin134 sp002239105	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group
GCA_002296525.1	Non-Sponge	Saline_water	96.15	2.14	2.45817	52.8	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__TK06;g__UBA6944;s__UBA6944 sp002296525	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_bacterium_graftm_464
GCA_002331465.1	Non-Sponge	Saline_water	86.56	2.99	2.415704	54.4	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__TK06;g__UBA2110;s__UBA2110 sp002331465	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_bacterium_graftm_464
GCA_002388005.1	Non-Sponge	Saline_water	92.23	3.42	2.422836	55.3	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__TK06;g__UBA2110;s__UBA2110 sp002388005	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_bacterium_graftm_464
GCA_002448715.1	Non-Sponge	Saline_water	95.01	3.85	2.224695	53	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__TK06;g__UBA6944;s__UBA6944 sp002448715	No_165
GCA_002457435.1	Non-Sponge	Seawater	87.65	0.85	1.958279	53	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__TK06;g__MedAcidi-G3;s__MedAcidi-G3 sp002457435	No_165
GCA_002470695.1	Non-Sponge	Saline_water	91.69	1.8	2.313555	51.9	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__TK06;g__UBA7388;s__UBA7388 sp002470695	No_165
GCA_002473265.1	Outgroup	Wastewater	67.59	2.85	3.370914	68.3	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Microtrichaceae;g__Microtrichix;s__Microtrichix sp002473265	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Candidatus_Microtrichix
GCA_002694825.1	Non-Sponge	Seawater	89.74	2.99	2.727227	64.6	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__UBA11606;s__UBA11606 sp002694825	No_165
GCA_002705305.1	Non-Sponge	Seawater	95.3	3.85	2.540084	55	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__TK06;g__UBA2110;s__UBA2110 sp002705305	Bacteria; Marinimicrobia_SAR406_clade; uncultured_bacterium_graftm_775
GCA_002708935.1	Non-Sponge	Seawater	95.73	4.7	2.408629	53	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__TK06;g__MedAcidi-G3;s__MedAcidi-G3 sp002708935	No_165
GCA_002719335.1	Non-Sponge	Seawater	85.7	2.14	2.319253	54.2	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__TK06;g__UBA2110;s__UBA2110 sp002719335	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_bacterium_graftm_464
GCA_002722565.1	Non-Sponge	Seawater	97.86	2.99	2.46767	51.9	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__TK06;g__MedAcidi-G3;s__MedAcidi-G3 sp002722565	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_bacterium_graftm_464

Table S5.2. Continued

GCA_002729125.1	Non-Sponge	Seawater	90.98	4.7	2.589833	65.3	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__UBA11606;s__UBA11606 sp002729125	No_16S
GCA_003231645.1	Non-Sponge	Hydrothermal_vent	94.02	1.28	2.896267	65.9	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__SZUA-87;s__SZUA-87 sp003231645	No_16S
GCA_003697065.1	Non-Sponge	Hot springs	96.58	3.13	3.453632	72.8	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__J010;s__J010 sp003697065	No_16S
IRC1_bin_37	Sponge	Sponge	93.16	1.28	3.726414	64.5	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__Bin134;s__	No_16S
IRC1_bin_38	Sponge	Sponge	94.87	1.28	3.114682	64.4	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__Bin134;s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group
IRC4_bin_10	Sponge	Sponge	95.73	1.28	3.29988	61.1	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__UBA11606;s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_bacterium_graftm_464
IRC4_bin_16	Sponge	Sponge	97.44	2.14	3.268807	67.3	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__TK06;g__s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_bacterium_graftm_464
IRC4_bin_37	Sponge	Sponge	93.59	2.14	3.192769	69.5	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__s__	No_16S
IRC_PAM_SB0661_bin_55	Sponge	Sponge	85.47	1.28	2.719624	70.2	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_actinobacterium_graftm_64
IRC_PAM_SB0662_bin_30	Sponge	Sponge	94.87	1.28	3.454628	64.7	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__Bin134;s__	No_16S
IRC_PAM_SB0665_bin_4	Sponge	Sponge	89.74	1.28	3.494407	68.8	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__s__	No_16S
IRC_PAM_SB0665_bin_5	Sponge	Sponge	91.45	1.28	3.905472	69	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__s__	No_16S
IRC_PAM_SB0668_bin_6	Sponge	Sponge	95.73	1.28	3.156972	70.1	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group; uncultured_bacterium_graftm_464
IRC_PAM_SB0675_bin_16	Sponge	Sponge	85.04	1.28	3.240544	69	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__s__	Bacteria; Cyanobacteria; Oxyphotobacteria; Synechococcales; Cyanobiaceae; [Synechococcus] spongiarum_group
IRC_PAM_SB0675_bin_2	Sponge	Sponge	94.87	1.28	1.785529	53.7	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__Bin134;s__	No_16S
RHO2_bin_26	Sponge	Sponge	88.89	1.45	3.567883	69.9	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group
RHO2_bin_38	Sponge	Sponge	95.73	1.28	2.406668	50.9	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__J010;s__	Bacteria
RHO3_bin_35	Sponge	Sponge	94.79	2.14	3.560049	65.4	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Bin134;g__Bin134;s__	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Sva0996_marine_group
RHO3_bin_4	Sponge	Sponge	96.58	1.28	3.061012	62.9	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__UBA11606;g__UBA11606;s__	No_16S
GCF_000299415.1	Outgroup	Wastewater	NA	NA	NA	NA	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Microtrichaceae;g__Microtrichix;s__Microtrichix parvicella	Bacteria; Actinobacteria; Acidimicrobia; Microtrichales; Microtrichaceae; Candidatus_Microtrichix

Table S5.3. Genome characteristics and taxonomy for Nitrosopumiliaceae

MAG_id	Group	Isolation_source	Completeness	Contamination	Genome_size_Mbp	GC	Classification_GTDB	Classification_16S
512_metabat1_sensitive.077	Sponge	Sponge	98.06	0	1.701044	45.5	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrosopumiliaceae;g__Cenarchaeum;s__	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_Candidatus_Nitrosopumilus_uncultured_crenarchaeote_graftm_5
513_metabat1_super.064	Sponge	Sponge	96.6	0	1.359681	46.5	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrosopumiliaceae;g__Cenarchaeum;s__	No_16S
APA_bin_56	Sponge	Sponge	91.59	0	2.186878	38.8	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrosopumiliaceae;g__Nitrosopumilus;s__	No_16S
AXIM_hallam_GCA_000200715.1	Sponge	Sponge	99.03	0	2.045086	57.4	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrosopumiliaceae;g__Cenarchaeum;s__Cenarchaeum symbiosum	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_Cenarchaeum
COS1_bin_11	Sponge	Sponge	97.57	0	0.836782	28.7	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrosopumiliaceae;g__Nitrosopumilus;s__	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_uncultured_marine_archaeon_graftm_1
COS36386_bin_19	Sponge	Sponge	98.06	0.49	0.84393	29.5	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrosopumiliaceae;g__Nitrosopumilus;s__	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_uncultured_marine_archaeon_graftm_1
CYMC_moitinhoThomas_67496.assembled	Sponge	Sponge	97.09	4.05	2.161854	38.4	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrosopumiliaceae;g__Nitrosopumilus;s__	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_Candidatus_Nitrosopumilus_uncultured_archaeon_graftm_56
GCA_000018465.1	Non-Sponge	Salt-water aquarium	100	0.97	1.645259	34.2	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrosopumiliaceae;g__Nitrosopumilus;s__Nitrosopumilus maritimus	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_Candidatus_Nitrosopumilus maritimus
GCA_000204585.1	Non-Sponge	Sediments SanFran Bay	98.06	0	1.772718	32.5	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrosopumiliaceae;g__Nitrosarchaeum;s__Nitrosarchaeum limnae	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_Candidatus_Nitrosarchaeum limnae
GCA_000299365.1	Non-Sponge	Marine sediment	94.66	0	1.639964	34.2	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrosopumiliaceae;g__Nitrosopumilus;s__Nitrosopumilus koreensis	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_Candidatus_Nitrosopumilus koreensis
GCA_000402075.1	Non-Sponge	Seawater	97.33	2.09	1.104438	35.7	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrosopumiliaceae;g__Nitrosopumilus;s__Nitrosopumilus sp000402075	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_uncultured_archaeon_graftm_8
GCA_000746765.1	Non-Sponge	Red Sea	86.89	0.97	1.360636	34.3	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrosopumiliaceae;g__Nitrosopumilus;s__Nitrosopumilus sp000746765	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_Candidatus_Nitrosopumilus sp000746765
GCA_000875775.1	Non-Sponge	Seawater Slovenia	100	0.97	1.713078	33.8	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrosopumiliaceae;g__Nitrosopumilus;s__Nitrosopumilus piranensis	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_Candidatus_Nitrosopumilus piranensis
GCA_001437625.1	Non-Sponge	Seawater Baltic Sea	99.51	0.97	1.270387	31	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrosopumiliaceae;g__Nitrosopumilus;s__Nitrosopumilus sp001437625	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_Candidatus_Nitrosopumilus sp001437625
GCA_001443365.1	Non-Sponge	Marine sediment	99.03	0	1.358434	34.4	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrosopumiliaceae;g__Nitrosopumilus;s__Nitrosopumilus sp001443365	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_uncultured_archaeon_graftm_8
GCA_001510275.1	Non-Sponge	Seawater Caspian sea	94.17	0	1.221161	32.8	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrosopumiliaceae;g__Nitrosopumilus;s__Nitrosopumilus sp001510275	Archaea
GCA_001627235.1	Non-Sponge	Seawater Red Sea	92.64	1.05	1.009131	34.1	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrosopumiliaceae;g__Nitrosopumilus;s__Nitrosopumilus sp001627235	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_Candidatus_Nitrosopumilus sp001627235
GCA_002317795.1	Non-Sponge	Seawater	93.85	0.97	1.906738	31.4	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrosopumiliaceae;g__Nitrosopumilus;s__Nitrosopumilus sp002317795	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_Candidatus_Nitrosopumilus_uncultured_archaeon_graftm_56
GCA_002506665.1	Sponge	deep-sea sponge Neamphius huxleyi	92.23	0	1.389729	41.2	d__Archaea;p__Thermoproteota;c__Nitrososphaeria;o__Nitrosopumiliaceae;g__Nitrosopumilus;s__Nitrosopumilus sp002506665	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_Candidatus_Nitrosopumilus_uncultured_archaeon_graftm_56

Table S5.3. Continued

GCA_002730325.1	Non-Sponge	Seawater	88.83	2.43	1.131154	31.7	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Nitrosopumilus;__Nitrosopumilus sp002730325	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_Candidatus_Nitrosopumilus
GCA_002737445.1	Non-Sponge	Lake Baikal	99.03	1.94	1.193764	31	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Nitrososphaerium;__Nitrososphaerium sp002737445	Archaea
GCA_002737455.1	Non-Sponge	Lake Baikal	99.03	0	1.139986	30.3	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Nitrosopumilus;__Nitrosopumilus sp002737455	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_Candidatus_Nitrosopumilus;_uncultured_thaumarchaeote
GCA_003331425.1	Non-Sponge	Seawater	100	0	1.235315	31.2	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Nitrosopumilus;__Nitrosopumilus sp002690535	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae
GCA_003352285.1	Non-Sponge	Seawater	99.03	0.97	1.689003	35.5	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Cenarchaeum;__Cenarchaeum sp003352285	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_uncultured_archaeon_graftm_8
GCA_003569705.1	Non-Sponge	biofilm from MA_RO membrane_Red_Sea	98.54	1.94	1.462514	33	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Nitrososphaerium;__Nitrososphaerium sp003569705	Archaea
GCA_003702465.1	Non-Sponge	Seawater Gulf of Mexico	93.2	0	1.127829	30.9	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Nitrosopumilus;__Nitrosopumilus sp003702465	No_165
GCA_003702495.1	Non-Sponge	Seawater Gulf of Mexico	95	1.94	1.266765	34.5	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Nitrosopumilus;__Nitrosopumilus sp003702495	No_165
GCA_003702525.1	Non-Sponge	Seawater Gulf of Mexico	92.72	2.91	1.279624	34.5	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Nitrosopumilus;__Nitrosopumilus sp003702525	No_165
GCA_003702545.1	Non-Sponge	Seawater Gulf of Mexico	91.26	0.97	1.224048	34.9	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Nitrosopumilus;__Nitrosopumilus sp003702545	No_165
GCA_003724285.1	Sponge	Sponge	89.81	0.07	1.176375	53	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Nitrosopumilus;__Nitrosopumilus sp003724285	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_Candidatus_Nitrosopumilus
GCA_003724325.1	Sponge	Sponge	92.23	0	1.17111	47.9	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Nitrososphaerium;__Nitrososphaerium sp003724325	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae;_Candidatus_Nitrosopumilus
GCA_004297665.1	Non-Sponge	Groundwater Tennessee	93.13	6.63	1.550815	33	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Nitrososphaerium;__Nitrososphaerium sp004297665	No_165
GCA_004322465.1	Non-Sponge	Groundwater Tennessee	91.76	2.91	1.681024	42.6	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Nitrosotenuis;__Nitrosotenuis sp004322465	No_165
GCA_005798405.1	Non-Sponge	Freshwater Fuxian Lake China	93.2	0	1.247066	33.4	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Nitrososphaerium;__Nitrososphaerium sp005798405	Archaea
GCA_005877205.1	Non-Sponge	Soil Angelo Coast USA	97.5	0	1.562348	38.7	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Nitrosotalea;__Nitrosotalea sp005877205	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosotaleales;_Nitrosotaleaceae
GCA_005877305.1	Non-Sponge	Soil Angelo Coast USA	92.64	1.94	1.329581	35.7	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__TA-20;__TA-20 sp005877305	No_165
GCA_007036525.1	Non-Sponge	Seawater Izu-Bonin Trench	88.11	0	0.94932	35.2	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__Nitrosopelagicus;__Nitrosopelagicus sp007036525	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosopumilales;_Nitrosopumilaceae
GCA_007037745.1	Non-Sponge	Seawater Izu-Bonin Trench	90.37	0	1.075201	35.5	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__TA-20;__TA-20 sp007037745	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosotaleales;_Nitrosotaleaceae
GCA_007280335.1	Non-Sponge	Freshwater Powell Lake British Columbia	94.17	0	1.296374	35.5	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososphaerales;f__Nitrosopumilaceae;g__TA-20;__TA-20 sp007280335	Archaea;_Thaumarchaeota;_Nitrososphaeria;_Nitrosotaleales;_Nitrosotaleaceae

Table S5.3. Continued

GCA_900065925.1	Non-Sponge	Soil	98.54	0	1.805304	37.1	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Nitrosotalea;s__Nitrosotalea devanaterra	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosota leales;_Nitrosotaleaceae;_C andidatus_Nitrosotalea
GCA_900177045.1	Non-Sponge	Wetland Soil	99.51	0	1.972585	37.5	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Nitrosotalea;s__Nitrosotalea okcheonensis	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosota leales;_Nitrosotaleaceae;_C andidatus_Nitrosotalea
GCF_000220175.1	Non-Sponge	Soil	100	0	1.607695	32.7	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Nitrosarchaeum;s__Nitrosarch aeum_koreense	Archaea
GCF_000242875.2	Non-Sponge	Sediment San Fran Bay estuary	92.39	1.94	1.572957	33.8	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Nitrosopumilus;s__Nitrosopum ilus_salaria	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosop umilales;_Nitrosopumilacea e
GCF_000299395.1	Non-Sponge	Sediment Arctic circle	97.09	0	1.690905	33.6	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Nitrosopumilus;s__Nitrosopum ilus_sediminis	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosop umilales;_Nitrosopumilacea e;_Candidatus_Nitrosopumil us;_uncultured_archaeon_gr aftm_56
GCF_000685395.1	Non-Sponge	Soil South Korea	99.03	0.97	1.762971	41.8	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Nitrosotenuis;s__Nitrosotenuis chungbukensis	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosop umilales;_Nitrosopumilacea e;_Candidatus_Nitrosotenui s;_uncultured_archaeon_gra ftm_149
GCF_000723185.1	Non-Sponge	terrestrial geothermal hot spring, Uzon Caldera	100	0.97	1.636125	42.2	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Nitrosotenuis;s__Nitrosotenuis sp000723185	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosop umilales;_Nitrosopumilacea e;_Candidatus_Nitrosotenui s
GCF_000812185.1	Non-Sponge	Seawater	99.51	0	1.232128	33.2	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Nitrosopelagicus;s__Nitrosopel agicus_brevis	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosop umilales;_Nitrosopumilacea e;_Candidatus_Nitrosopelag icus
GCF_000955905.1	Non-Sponge	wastewater treatment plant Beijing	100	1.94	1.620156	41	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Nitrosotenuis;s__Nitrosotenuis cloacae	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosop umilales;_Nitrosopumilacea e;_Candidatus_Nitrosotenui s
GCF_000956175.1	Non-Sponge	Seawater Slovenia	100	0	1.80309	33.4	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Nitrosopumilus;s__Nitrosopum ilus_adriaticus	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosop umilales;_Nitrosopumilacea e;_Candidatus_Nitrosopumil us
GCF_002156965.1	Non-Sponge	Seawater California	100	0	1.360076	31.4	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Nitrosopumilus;s__Nitrosopum ilus_catalinensis	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosop umilales;_Nitrosopumilacea e
GCF_002787055.1	Non-Sponge	freshwater aquarium biofilter Canada	99.68	1.94	1.697207	42.2	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Nitrosotenuis;s__Nitrosotenuis aquarius	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosop umilales;_Nitrosopumilacea e;_Candidatus_Nitrosotenui s
GCF_003175215.1	Non-Sponge	eelgrass sediment Japan	100	0	1.757809	33.8	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Nitrosopumilus;s__Nitrosopum ilus_sp003175215	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosop umilales;_Nitrosopumilacea e;_Candidatus_Nitrosopumil us
GCF_006740685.1	Non-Sponge	Seawater South Korea	100	0.97	1.49962	33.8	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Nitrosopumilus;s__Nitrosopum ilus_sp006740685	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosop umilales;_Nitrosopumilacea e;_Candidatus_Nitrosopumil us
GCF_900143675.1	Non-Sponge	soil wetland fen	99.51	0.97	1.598048	37.4	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Nitrosotalea;s__Nitrosotalea sinensis	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosota leales;_Nitrosotaleaceae;_C andidatus_Nitrosotalea
GCF_900167955.1	Non-Sponge	soil wetland fen	97.57	1.94	1.553159	36	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Nitrosotalea;s__Nitrosotalea bavarica	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosota leales;_Nitrosotaleaceae;_C andidatus_Nitrosotalea
GCF_900620265.1	Sponge	Sponge	99.03	0.97	1.99668	64.8	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Nitrosopumilus;s__Nitrosopum ilus_sp900620265	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosop umilales;_Nitrosopumilacea e;_Candidatus_Nitrosopumil us
LOPHE_tianQian_GCA_001543015.1_ASM154301v1	Sponge	Sponge	100	0	1.925182	31.6	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Nitrosopumilus;s__Nitrosopum ilus_sp001543015	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosop umilales;_Nitrosopumilacea e;_Candidatus_Nitrosopumil us;_uncultured_archaeon_gr aftm_56
RHO2_bin_25	Sponge	Sponge	94.17	0	1.990914	60.4	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Nitrosopumilus;s__	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosop umilales;_Nitrosopumilacea e;_Candidatus_Nitrosopumil us;_uncultured_archaeon_gr aftm_56

Table S5.3. Continued

STY3_bin_6	Sponge	Sponge	98.54	0	1.488362	67.4	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Cenarchaeum;s__Cenarchaeu m sp003724275	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosop umilales;_Nitrosopumilacea e;_Cenarchaeum;_unculture d_archaeon_graftm_145
SUB_tianQian_GCA_001541925.1_ASM154192v1	Sponge	Sponge	100	0	1.383621	31.4	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __Nitrosopumilus;s__Nitrosopum ilus sp001541925	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosop umilales;_Nitrosopumilacea e
U_67070	Sponge	Sponge	85.6	0	1.184172	35.9	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosopumilaceae;g __UBA8516;s__UBA8516 sp8516u	No_165
GCA_002898655.1	Outgroup	Ammonia-oxidizing_enrichment_culture	76.5	0.32	1.208457	41.8	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosocaldaceae;g __Nitrosocaldus;s__Nitrosocaldus sp002898655	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosoc aldales;_Nitrosocaldaceae; Candidatus_Nitrosocaldus
GCA_900248165.1	Outgroup	Hotspring_Italy	99.03	0	1.577284	41.6	d__Archaea;p__Thermoproteota; c__Nitrososphaeria;o__Nitrososp haerales;f__Nitrosocaldaceae;g __Nitrosocaldus;s__Nitrosocaldus cavascurensis	Archaea;_Thaumarchaeota; _Nitrososphaeria;_Nitrosoc aldales;_Nitrosocaldaceae; Candidatus_Nitrosocaldus

Table S5.4. Genome characteristics and taxonomy for Spirochaetaceae

MAG_id	Group	Isolation_source	Completeness	Contamination	Genome_size_Mbp	GC	Classification_GTDB	Classification_16S
510_metabat1_super.011	Sponge	Sponge	94.53	1.33	5.229259	67.6	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__Bin103.g__Bin103;s__	Bacteria; Spirochaetes; Spirochaetia; Spirochaetales; Spirochaetaceae; Spirochaeta 2; uncultured_bacterium_graftm_80
514_concoct.135_sub	Sponge	Sponge	85.66	0.93	4.194815	67.5	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__Bin103.g__Bin103;s__	No_165
515_metabat2.042	Sponge	Sponge	89.87	3.93	4.486424	67.4	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__Bin103.g__Bin103;s__	Bacteria; Spirochaetes; Spirochaetia; Spirochaetales; Spirochaetaceae; Spirochaeta 2; uncultured_bacterium_graftm_80
575_concoct.030	Sponge	Sponge	95.33	2.53	5.604275	68.6	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__Bin103.g__Bin103;s__	No_165
APA_bin_62	Sponge	Sponge	98.13	1.33	4.837189	67.4	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__Bin103.g__Bin103;s__	Bacteria; Spirochaetes; Spirochaetia; Spirochaetales; Spirochaetaceae; Spirochaeta 2; uncultured_bacterium_graftm_80
APA_bin_94	Sponge	Sponge	96.93	3.73	6.2703	68.2	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__Bin103.g__Bin103;s__	Bacteria; Spirochaetes; Spirochaetia; Spirochaetales; Spirochaetaceae; Spirochaeta 2; uncultured_bacterium_graftm_80
GCA_002238925.1	Sponge	Sponge	94.8	1.33	4.776897	67.4	NA	NA
GCA_002084135.1	Non-Sponge	Lucinid_gill	86.25	0.8	1.945966	50	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__Spirochaetaceae;g__UBA2779;s__UBA2779;sp002084135	Bacteria; Spirochaetes; Spirochaetia; Spirochaetales; Spirochaetaceae; Spirochaeta 2
GCA_002084805.1	Non-Sponge	Deep_sea_hydrothermal_vent_sediment	99.2	1.2	3.287989	43.1	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__Alkalispicrochaetaceae;g__Oceanispicrochaeta;s__Oceanispicrochaeta;sp002084805	No_165
GCA_002313505.1	Non-Sponge	Saline_water	85.2	0.8	3.813472	64.5	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__Fen-1364;g__Fen-1364;s__Fen-1364;sp003141795	No_165
GCA_003141795.1	Non-Sponge	Permafrost_active_layer_soil	89.74	3.07	5.200069	62.6	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__UBA9216;g__UBA9216;s__UBA9216;sp003245835	No_165
GCA_003245835.1	Non-Sponge	marine_hydrothermal_sulfide_sediment	100	6.32	4.433794	46.6	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__Alkalispicrochaetaceae;g__B15ed10-166;s__B15ed10-166;sp003552185	No_165
GCA_003552185.1	Non-Sponge	hypersaline_soda_lake_sediment	94.77	3.47	3.485397	61.4	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__Alkalispicrochaetaceae;g__B15ed10-166;s__B15ed10-166;sp003552985	No_165
GCA_003552985.1	Non-Sponge	hypersaline_soda_lake_sediment	95.2	4.67	3.928934	60.7	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__Alkalispicrochaetaceae;g__B15ed10-166;s__B15ed10-166;sp003556605	No_165
GCA_003556605.1	Non-Sponge	hypersaline_soda_lake_sediment	100	4.4	3.876057	59.1	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__PWKH01;g__PWKH01;s__PWKH01;sp003559765	No_165
GCA_003559765.1	Non-Sponge	hypersaline_soda_lake_sediment	88.06	1.24	2.093155	55.3	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__SLAA01;g__SLAA01;s__SLAA01;sp003964775	No_165
GCA_003564775.1	Non-Sponge	hypersaline_soda_lake_sediment	87.72	3.87	2.873866	55	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__B11-G9;g__B11-G9;s__B11-G9;sp003641695	No_165
GCA_003641695.1	Non-Sponge	Deep_sea_hydrothermal_vent_sediment	92.51	1.34	2.578246	43.1	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__Spirochaetaceae;g__UBA2779;s__UBA2779;sp003644105	No_165
GCA_003644105.1	Non-Sponge	Deep_sea_hydrothermal_vent_sediment	88	1.6	3.324968	48.1	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__B32-G16;g__B32-G16;s__B32-G16;sp003645645	No_165
GCA_003645645.1	Non-Sponge	Deep_sea_hydrothermal_vent_sediment	96.55	0	3.982661	43.8	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__RPPD01;g__RPPD01;s__RPPD01;sp003818675	No_165
GCA_003818675.1	Non-Sponge	Prairie_Pothole_Region_wetland_sediments	86.57	5.91	3.085686	48.1	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__SPDA01;g__SPDA01;s__SPDA01;sp004525155	No_165
GCA_004525155.1	Non-Sponge	Bothnian_Sea_sediment	89.01	1.25	3.176844	52.5	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__PWKH01;g__PWKH01;s__PWKH01;sp007116075	No_165
GCA_007116075.1	Non-Sponge	hypersaline_soda_lake_sediment	90.91	1.52	2.205354	51.1		

Table S5.4. Continued

GCF_003346715.1	Non-Sponge	Black_sea_water_bulgaria	99.73	3.73	5.884319	42.9	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__Spirochaetaceae;B;g__Oceanispirochaeta;s__Oceanispirochaeta sp003346715	Bacteria; Spirochaetes; Spirochaetia; Spirochaetales; Spirochaetaceae; Spirochaeta_2
GCF_900156105.1	Non-Sponge	Missing_type_strain_of_Alkalispichoeta_americana	98.4	0	3.309052	57.5	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__Alkalispichoetaeaceae;g__Alkalispichoeta;s__Alkalispichoeta americana	Bacteria; Spirochaetes; Spirochaetia; Spirochaetales; Spirochaetaceae; Alkalispichoeta
GCF_900608495.1	Non-Sponge	shallow_marine_sediment_Mediterranean_Sea	96	3.2	2.12133	47.1	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__Spirochaetaceae;B;g__UBA2779;s__UBA2779 sp900608495	No_165
RHO1_bin_44	Sponge	Sponge	97.73	2.93	5.730673	67.4	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__Bin103;g__Bin103;s__	Bacteria; Spirochaetes; Spirochaetia; Spirochaetales; Spirochaetaceae; Spirochaeta_2; uncultured_bacterium_graftm_80
RHO3_bin_84	Sponge	Sponge	92.53	3.33	5.975327	68.3	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;f__Bin103;g__s__	No_165
GCF_000147075.1	Outgroup	Unknown_isolate	100	0	2.472645	61.9	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;A;f__Spirochaetaceae;A;g__Spirochaeta;A;s__Spirochaeta_A thermophila_A	Bacteria; Spirochaetes; Spirochaetia; Spirochaetales; Spirochaetaceae; Spirochaeta_2
GCF_000184345.1	Outgroup	Hotspring	100	0	2.560222	60.9	d__Bacteria;p__Spirochaetota;c__Spirochaetia;o__Spirochaetales;A;f__Spirochaetaceae;A;g__Spirochaeta;A;s__Spirochaeta_A thermophila	Bacteria; Spirochaetes; Spirochaetia; Spirochaetales; Spirochaetaceae; Spirochaeta_2

Table S5.5. Genome characteristics and taxonomy for Thermoanaerobaculaceae

MAG_id	Group	Isolation_source	Completeness	Contamination	Genome_size_Mbp	GC	Classification_GTDB	Classification_16S
512_metabat2.007_sub	Sponge	Sponge	95.73	0.85	4.743285	68.2	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__UB A5704;f__QQVD01;g__s__	Bacteria; Acidobacteria; Thermoanaerobaculia; Thermoanaerobaculales; Thermoanaerobaculaceae; Subgroup_10
514_metabat2.006	Sponge	Sponge	94.87	3.42	4.371599	68.5	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__UB A5704;f__QQVD01;g__s__	Bacteria; Acidobacteria; Thermoanaerobaculia; Thermoanaerobaculales; Thermoanaerobaculaceae; Subgroup_10; uncultured_bacterium_gra
567_concoct.049	Sponge	Sponge	95.73	1.71	4.668207	68.5	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__UB A5704;f__QQVD01;g__s__	Bacteria; Acidobacteria; Thermoanaerobaculia; Thermoanaerobaculales; Thermoanaerobaculaceae; Subgroup_10; uncultured_bacterium_gra
575_metabat1_sensitive.014	Sponge	Sponge	94.59	1.71	4.258151	68.4	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__UB A5704;f__QQVD01;g__s__	Bacteria; Acidobacteria; Thermoanaerobaculia; Thermoanaerobaculales; Thermoanaerobaculaceae; Subgroup_10; uncultured_bacterium_gra
COS1_bin_13	Sponge	Sponge	89.03	7.05	4.98737	68	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__UB A5704;f__QQVD01;g__s__	Bacteria; Acidobacteria; Thermoanaerobaculia; Thermoanaerobaculales; Thermoanaerobaculaceae; Subgroup_10; uncultured_bacterium_gra
COS2_bin_5	Sponge	Sponge	94.02	1.71	4.435821	68	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__UB A5704;f__QQVD01;g__s__	Bacteria; Acidobacteria; Thermoanaerobaculia; Thermoanaerobaculales; Thermoanaerobaculaceae; Subgroup_10; uncultured_Acidobacteria_
COS36386_bin_5	Sponge	Sponge	92.31	1.71	4.621193	68.1	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__UB A5704;f__QQVD01;g__s__	Bacteria; Acidobacteria; Thermoanaerobaculia; Thermoanaerobaculales; Thermoanaerobaculaceae; Subgroup_10; uncultured_Acidobacteria_
COS36388_bin_9	Sponge	Sponge	94.87	1.71	4.223613	68.5	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__UB A5704;f__QQVD01;g__s__	Bacteria; Acidobacteria; Thermoanaerobaculia; Thermoanaerobaculales; Thermoanaerobaculaceae; Subgroup_10
COS4_bin_3	Sponge	Sponge	94.59	1.71	4.375714	68.3	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__UB A5704;f__QQVD01;g__s__	Bacteria; Acidobacteria; Thermoanaerobaculia; Thermoanaerobaculales; Thermoanaerobaculaceae; Subgroup_10
GCA_001766905.1	Non-Sponge	Sediment	90.6	0.85	3.610782	67.6	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__The rmoanaerobaculales;f__Thermo	Bacteria; Acidobacteria; Thermoanaerobaculia; Thermoanaerobaculales; Thermoanaerobaculaceae; Thermoanaerobaculum
GCA_002279285.1	Non-Sponge	Mine_wastewater	89.32	4.7	3.016655	70.4	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__The rmoanaerobaculales;f__Thermo	Bacteria; Acidobacteria; Thermoanaerobaculia; Thermoanaerobaculales; Thermoanaerobaculaceae; Thermoanaerobaculum; uncultured_bact
GCA_002327305.1	Non-Sponge	Suncor_tailing_pond	91.99	1.71	3.646212	65.5	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__UB A2201;f__UBA2201;g__UBA2201	No_16S
GCA_002414905.1	Non-Sponge	Waste_water_treatment_plant	87.08	4.27	4.270506	69	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__UB A5066;f__UBA5066;g__UBA5066	No_16S
GCA_002420005.1	Non-Sponge	Sediment	91.31	2.56	8.024947	69.4	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__UB A5704;f__UBA5704;g__UBA5704	No_16S
GCA_003105185.1	Non-Sponge	methane_bioreactor	94.11	7.26	4.236048	69.8	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__The rmoanaerobaculales;f__FEB-	Bacteria; Acidobacteria; Thermoanaerobaculia; Thermoanaerobaculales; Thermoanaerobaculaceae; Subgroup_23; uncultured_bacterium_gra
GCA_003133645.1	Non-Sponge	permafrost	90.29	1.71	3.69037	69.4	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__The rmoanaerobaculales;f__Thermo	No_16S
GCA_003152095.1	Non-Sponge	permafrost	93.7	2.56	3.637326	67.7	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__UB A5066;f__UBA5066;g__Fen-	No_16S
GCA_003158745.1	Non-Sponge	permafrost	92.69	1.71	3.735526	66.9	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__The rmoanaerobaculales;f__Thermo	No_16S
GCA_003222275.1	Outgroup	soil	92.09	5.13	6.756227	69.6	d__Bacteria;p__Acidobacteriota; c__Viciniabacteria_A;o__Fen-	No_16S
GCA_003222295.1	Outgroup	soil	75.52	4.27	5.279708	70.3	d__Bacteria;p__Acidobacteriota; c__Viciniabacteria_A;o__Fen-	No_16S
GCA_003222375.1	Non-Sponge	soil	85.11	4.27	3.767804	61	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__Gp 7-AA8;f__Gp7-AA8;g__Gp7-	No_16S
GCA_003222385.1	Non-Sponge	soil	93.95	2.56	3.600504	66.9	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__UB A5066;f__Gp7-AA6;g__Gp7-	Bacteria; Acidobacteria; Holophagae; Subgroup_7; uncultured_bacterium_graftm_635
GCA_003223555.1	Outgroup	soil	91.36	4.27	5.388591	71	d__Bacteria;p__Acidobacteriota; c__Viciniabacteria_A;o__Fen-	No_16S
GCA_003223635.1	Non-Sponge	soil	92.74	5.13	3.729145	62.4	d__Bacteria;p__Acidobacteriota; c__Thermoanaerobaculia;o__Gp 7-AA8;f__Gp7-	No_16S

