Received: 21 February 2024 | Accepted: 26 July 2024

DOI: 10.1111/1462-2920.16690

BRIEF REPORT

ENVIRONMENTAL MICROBIOLOGY

Comparative genomics identifies key adaptive traits of sponge-associated microbial symbionts

Paul A. O'Brien[1,2,3,4] | **Steven J. Robbins**[4] | **Shangjin Tan**[5,6] | **Laura Rix**[4] | **David J. Miller**[7,8] | **Nicole S. Webster**[2,4,9] | **Guojie Zhang**[10,11] | **David G. Bourne**[1,2,3]

[1]College of Science and Engineering, James Cook University, Townsville, Queensland, Australia

[2]Australian Institute of Marine Science, Townsville, Queensland, Australia

[3]AIMS@JCU, Townsville, Queensland, Australia

[4]School of Chemistry and Molecular Biosciences, Australian Centre for Ecogenomics, The University of Queensland, St Lucia, Queensland, Australia

[5]BGI-Shenzhen, Shenzhen, China

[6]BGI Research, Wuhan, China

[7]ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Queensland, Australia

[8]Centre for Tropical Bioinformatics and Molecular Biology, James Cook University, Townsville, Queensland, Australia

[9]Institute for Marine and Antarctic studies, University of Tasmania, Hobart, Tasmania, Australia

[10]Centre for Evolutionary & Organismal Biology, and Women's Hospital, Zhejiang University School of Medicine, Hangzhou, China

[11]Liangzhu Laboratory, Zhejiang University Medical Center, Hangzhou, China

Correspondence
David G. Bourne, College of Science and Engineering, James Cook University, Townsville, QLD, Australia.
Email: david.bourne@jcu.edu.au

Guojie Zhang, Centre for Evolutionary & Organismal Biology, and Women's Hospital, Zhejiang University School of Medicine, Hangzhou 310058, China.
Email: guojiezhang@zju.edu.cn

Funding information
AIMS@JCU; Beijing Genomics Institute

Abstract

Sponge microbiomes are often highly diverse making it difficult to determine which lineages are important for maintaining host health and homeostasis. Characterising genomic traits associated with symbiosis can improve our knowledge of which lineages have adapted to their host and what functions they might provide. Here we examined five microbial families associated with sponges that have previously shown evidence of cophylogeny, including *Endozoicomonadaceae, Nitrosopumilaceae, Spirochaetaceae, Microtrichaceae* and *Thermoanaerobaculaceae*, to better understand the mechanisms behind their symbiosis. We compared sponge-associated genomes to genomes found in other environments and found that sponge-specific clades were enriched in genes encoding many known mechanisms for symbiont survival, such as avoiding phagocytosis and defence against foreign genetic elements. We expand on previous knowledge to show that glycosyl hydrolases with sulfatases and sulfotransferases likely form multi-enzyme degradation pathways to break and remodel sulfated polysaccharides and reveal an enrichment in superoxide dismutase that may prevent damage from free oxygen radicals produced by the host. Finally, we identified novel traits in sponge-associated symbionts, such as urea metabolism in *Spirochaetaceae* which was previously shown to be rare in the phylum Spirochaetota. These results identify putative mechanisms by which symbionts have adapted to living in association with sponges.

INTRODUCTION

Genomes of symbiotic microorganisms can 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 lifestyle (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 (Agashe & Shankar, 2014). Successful symbionts have also developed mechanisms to evade the host immune response, for example secretion of eukaryote-like proteins (ELPs) can interfere with cellular processes such as phagocytosis (Jernigan & Bordenstein, 2014; Reynolds & Thomas, 2016). Similarly, the host immune response can involve the generation of reactive oxygen species released from phagocytes, such as superoxide, and the Cu/Zn family of superoxide dismutase (SOD) may protect bacteria from phagocyte killing as well as oxy-radical damage (Broxton & Culotta, 2016).

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

Though many genomic features considered advantageous in symbiotic bacteria are also observed in non-symbiotic microbes, certain traits appear to be enriched in symbionts if encounters are more frequent. For example, symbionts of sponges are exposed to increased amounts of mobile genetic elements due to the high seawater filtration rate of their sponge host (Jahn et al., 2019), and the enrichment of genes that encode components of the restriction–modification (RM) and CRISPR-*cas* systems is proposed to be a consequence of this (Fan et al., 2012; Horn et al., 2016). 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 (Gao et al., 2014; Kamke et al., 2014; Thomas et al., 2010; Zhang et al., 2019). Thus, it is likely that new symbiotic traits can be uncovered using comparative genomics when comparing the genomes of specific lineages of sponge symbionts to closely related genomes from non-sponge environments (Cooke et al., 2019; Díez-Vives et al., 2018; Zhang et al., 2019).

A recent genomic analysis of the sponge microbiome compared a phylogenetically diverse range of sponge symbionts to a wide range of prokaryote taxa from seawater, which required symbiont traits to be enriched across disparate groups (Robbins et al., 2021). However, this broad analysis may obscure traits that are unique to specific microbial lineages. Therefore, we provide a comparative enrichment analysis targeting sponge symbionts which have previously shown evidence for cophylogeny in coral reef invertebrates (O'Brien et al., 2021), as these may have increased likelihood of showing genomic signals of host adaptations. Using publicly available genomes of each sponge symbiont and genomes from non-sponge environments, we reconstructed the phylogeny to identify how the evolutionary history of sponge-associated genomes compared to those of non-sponge genomes. Following this, we looked for gene enrichment patterns in sponge symbiont clades against non-sponge clades to identify genomic evidence of adaptation to the sponge environment.

EXPERIMENTAL PROCEDURES

Genome curation, phylogeny and taxonomic classification

Five prokaryote families were selected for analysis based on evidence of cophylogenetic signatures as detailed in O'Brien et al. (2021), including *Endozoicomonadaceae, Spirochaetaceae, Nitrosopumilaceae* (formerly in Thaumarchaeota), *Microtrichaceae* and *Thermoanaerobaculaceae*. Genomes for each

symbiont group were obtained by first mining a set of sponge-derived metagenome-assembled genomes (MAGs) from O'Brien et al. (2023) for those classified within the families of our selected list of symbionts. To ensure consistent taxonomic classification between the MAGs used in this study and the 16S rRNA amplicons identified as cophylogenetic taxa in O'Brien et al. (2021), we extracted the 16S rRNA gene sequences from our MAGs and reclassified them using GraftM (v 0.13.1) (Boyd et al., 2018) and the Silva v132 16S rRNA gene database (Tables S1–S5). Second, we included all sponge-derived MAGs consistent with the taxonomic classifications above in a comprehensive list of sponge symbionts compiled by Robbins et al. (2021) to give a final set of sponge-associated symbiont genomes (Tables S1–S5).

To compare our sponge symbiont genomes to closely related non-sponge genomes, we first identified suitable candidates by constructing a phylogenomic tree and including all entries of the same genome classifications as our symbionts in the Genome Taxonomy Database (GTDB, release 95) (Parks et al., 2022). The phylogenomic tree was computed using GTDB-tk (v1.4.0) (Chaumeil et al., 2020) with the de_novo_wf command, which uses Fast Tree (Price, Dehal, & Arkin, 2010) to estimate phylogeny from 122 and 120 bacterial and archaeal marker genes respectively. This method allowed us to phylogenetically visualize closely related genomes and sister clades of our selected symbiont groups. Closely related genomes were then retrieved and added to our set of symbiont genomes, which were dereplicated at 95% average nucleotide identity (ANI) to represent operational species boundaries (Jain et al., 2018), yielding 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. Finally, 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. Phylogenetic outgroups were chosen by including the next closest lineage to the genomes within our enrichment analysis within GTDB.

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 (except for two *Endozoicomonadaceae* genomes that were included at 81% and 84% completeness 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., 2018) with the Kyoto Encyclopedia of Genes and Genomes (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 groups. Statistical validation was performed using two metrics, (a) Fisher's exact test with a Benjamini-Hochberg correction for multiple comparisons to test for enrichment by the number of genomes containing the gene in each group, and (b) Mann–Whitney U test with a Benjamini-Hochberg correction for multiple comparisons to test for 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).

Gene search using tblastx

To manually search for genes that may have been missed in our annotation pipeline above, reference sequences for each gene (*tauABC*) were retrieved from KEGG. A custom blast database was built using the genomes of each symbiont group using blast v2.11 (Altschul et al., 1990), and tblastx was used to perform a translated protein search of each gene to each symbiont database. Significant matches for gene homology were determined by a minimum percent identity of 30%, alignment length of 100, e-value of 1e-03 and a bit score of 50 (Pearson, 2013).

RESULTS

Phylogeny and genome characteristics of sponge-associated symbionts

A total of 13 *Endozoicomonadaceae* (five sponge), 57 *Microtrichaceae* (42 sponge), 62 *Nitrosopumilaceae* (16 sponge), 48 *Spirochaetaceae* (nine sponge), and 32 *Thermoanaerobaculaceae* (14 sponge) genomes were included in the analysis based on their 16S rRNA gene taxonomic classifications and phylogenomic placement (Tables S1–S5), representing a total of 16 sponge host species. A phylogenomic tree showed

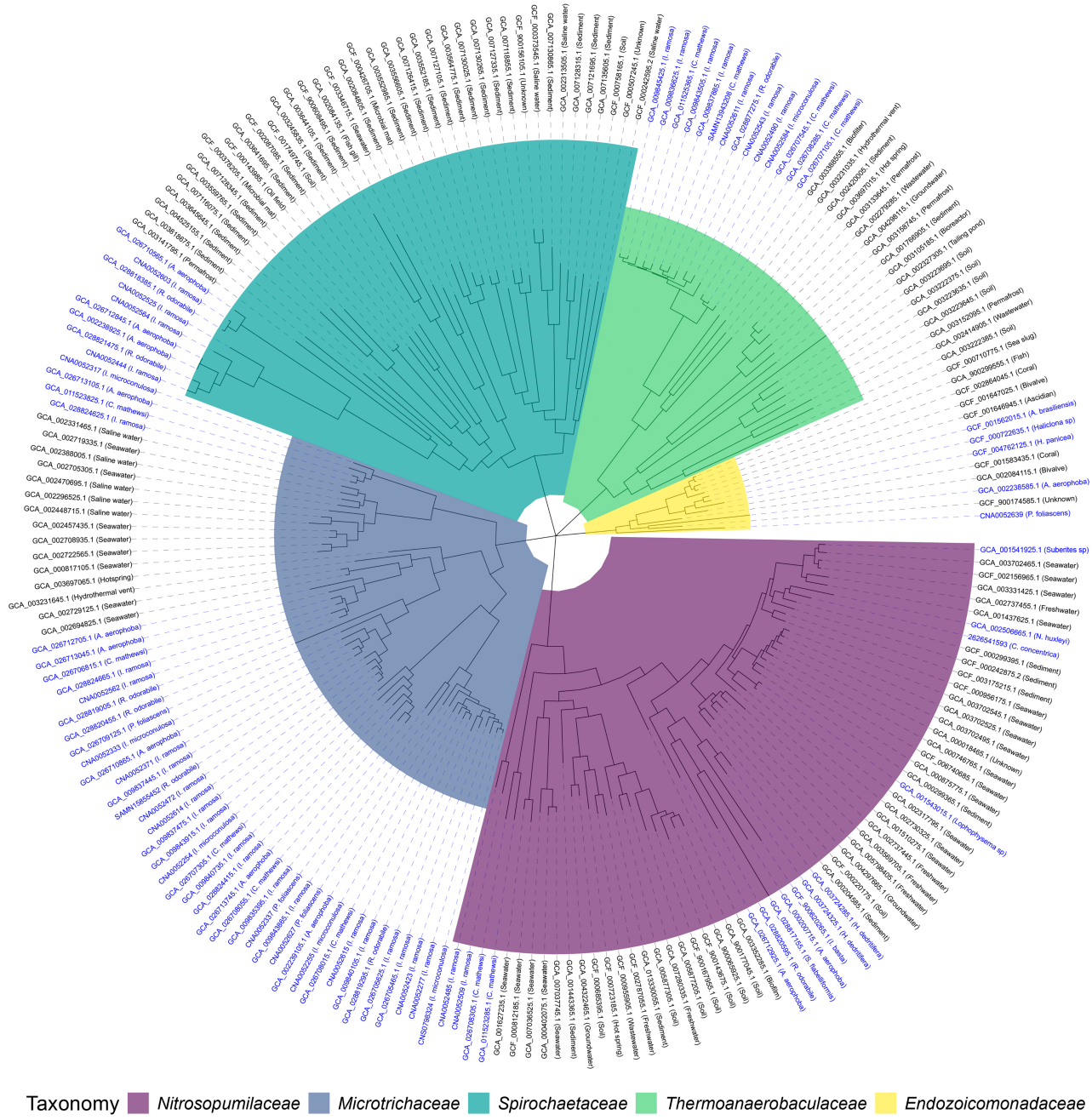

FIGURE 1 Phylogenomic tree of sponge and non-sponge associated microbial genomes. Branch tips in blue indicate sponge-associated genomes while branch tips in black indicate non-sponge-associated genomes. Tip labels refer to the genome assembly ID with isolation source or sponge host species in parentheses and clades are coloured by microbial family.

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 1). Moreover, we note that the 16S classification of one *Thermoanaerobaculaceae* clade was Holophagae Subgroup 7, however, this was included in the analysis based on its close phylogenomic relationship within the non-sponge *Thermoanaerobaculaceae* clade. Sponge-derived genomes of *Nitrosopumilaceae* were associated with multiple lineages of free-living microbes, however, one sponge cluster was found consisting of seven genomes (Figure 1). Similarly, sponge-derived genomes of *Endozoicomonadaceae* were closely related to symbionts retrieved from other marine invertebrates (all other genomes) and did not cluster in an exclusive sponge clade (Figure 1).

Genome size was greater and GC content higher in sponge-associated *Spirochaetaceae* and *Microtrichaceae*

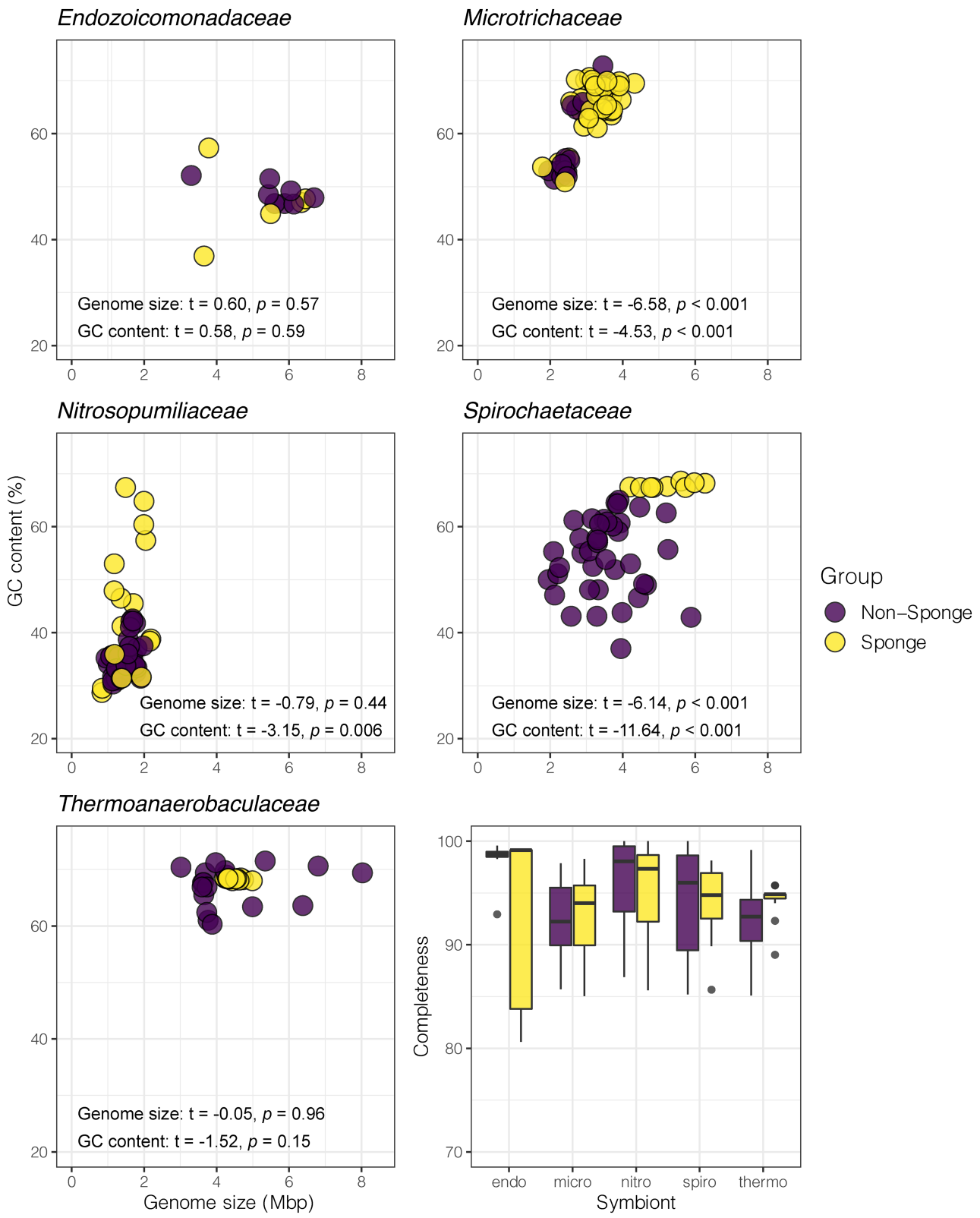

FIGURE 2 Genome size and GC content for all sponge-associated and non-sponge-associated genomes for each microbial family. Genome completeness for sponge-associated and non-sponge-associated genomes for each microbial group (last panel).

compared to non-sponge relatives (Figure 2; Table 1), suggesting that these sponge symbionts have not undergone genome streamlining. Although there was no statistical difference between genome sizes of sponge and non-sponge-derived *Nitrosopumilaceae*, two sponge-associated strains had particularly small genomes of 0.84 Mbp (97.6% and 98.1% completeness), highlighting the potential for genome reduction (Figure 2; Table 1).

TABLE 1 Statistical summary for genome size and GC content for sponge-associated and non-sponge-associated genomes in all five microbial groups.

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

TABLE 2 Overview of the total number of enriched genes for sponge-associated and non-sponge-associated genomes for all five microbial groups.

Family	Group	CAZY	KO	PFAM
Spirochaetaceae	Sponge	8	331	317
	Non-sponge	8	425	511
Microtrichaceae	Sponge	0	312	367
	Non-sponge	2	134	140
Thermoanaerobaculaceae	Sponge	4	538	545
	Non-sponge	7	435	510
Endozoicomonadaceae	Sponge	0	0	0
	Non-sponge	0	0	0
Nitrosopumilaceae	Sponge	0	7	16
	Non-sponge	0	15	27

Note: Enriched genes are included if the number of genomes that contain the gene within each group are statistically different.

Similarly, these two genomes had the lowest GC content of all *Nitrosopumilaceae* genomes, with 28.7% and 29.5%, indicating these two outlier genomes may be obligate symbionts (Giovannoni et al., 2014). 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 2; Table 1). Finally, genome size and GC content within the *Endozoicomonadaceae* were similar between sponge and non-sponge genomes (Figure 2; Table 1), likely because non-sponge genomes were still derived from symbionts of other marine invertebrates, including corals, bivalves, an ascidian, and echinoderm (Table S1).

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

The *Spirochaetaceae*, *Microtrichaceae* and *Thermoanaerobaculaceae* all had greater than 300 enriched genes in sponge-associated genomes compared to non-sponge genomes when analysing the KO and Pfam annotations (Table 2). Comparatively, only 7 and 17 genes were enriched in sponge-associated *Nitrosopumilaceae* genomes for KO and Pfam annotations respectively, while no enrichment patterns were observed for *Endozoicomonadaceae* in any annotation database. Sponge-associated *Spirochaetaceae* and *Thermoanaerobaculaceae* additionally showed eight and four enriched CAZymes respectively, however, no CAZymes were enriched in the remaining microbial groups in sponge-associated genomes (Table 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 our phylogenomic trees for these three families. Given the lack of enriched genes in *Endozoicomonadaceae* and *Nitrosopumilaceae* genomes, we focussed the comparative genomic analysis on the *Spirochaetaceae*, *Microtrichaceae* and *Thermoanaerobaculaceae* genomes.

Avoiding the host immune response and infection from foreign DNA

Our symbionts were enriched in a range of genes that potentially facilitate 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). 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 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 encoded the five ARPs, however only one

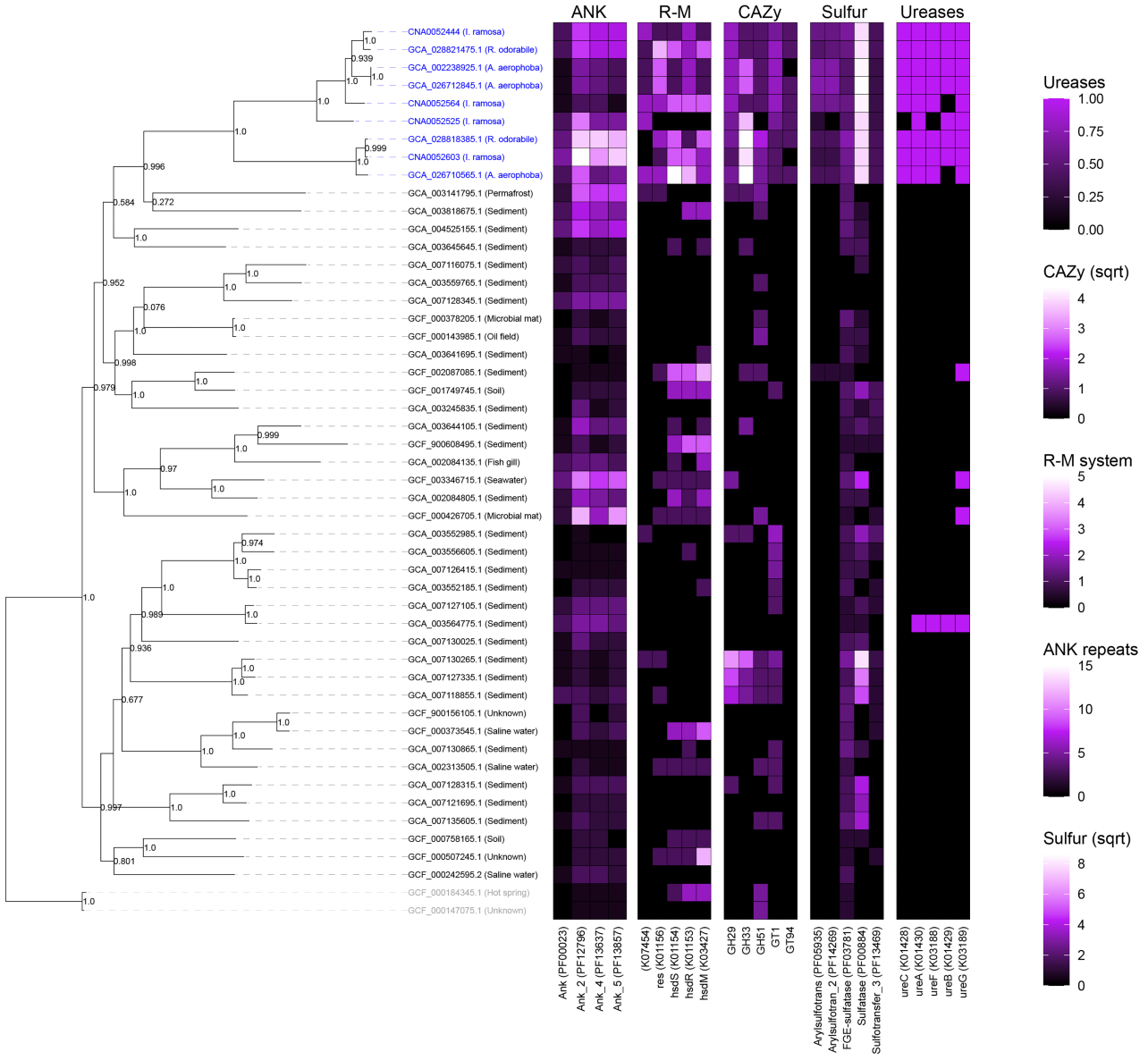

FIGURE 3 Phylogenomic relationships and gene enrichment patterns in sponge-associated and non-sponge-associated *Spirochaetaceae* genomes. Branches with blue labels indicate sponge-associated genomes, branches with black labels indicate non-sponge-associated genomes while grey labels indicate outgroups and weren't included in the enrichment analysis. Tip labels refer to the genome assembly ID with isolation source or sponge host species in parentheses. 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 and are labelled by their gene codes (where available). Data that has been square root transformed for illustration is indicated with (sqrt). ANK, ankyrin repeat proteins; CAZy, carbohydrate-active enzymes; R–M, restriction–modification system enzymes; Sulfur, sulfatases and sulfatransferases.

was enriched compared to the non-sponge genomes (Figure 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 encoding ELPs (Figure 5). Finally, all three microbial families were enriched in the Cu-Zn family of SOD (extracellular *SodC*), and while the Fe-Mn family (intracellular *SodAB*) 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 3 and 5). While *Thermoanaerobaculaceae* also contained restriction enzymes, only two of these were enriched in sponge-associated genomes (Figure 4). Similarly, genes classified as *cas* enzymes associated with the CRISPR-Cas system were abundant in our dataset. Of these, two *cas* enzymes were enriched in our 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* encoded *cas* enzymes and none showed any enrichment.

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 encoded in all sponge-associated *Microtrichaceae* and *Thermoanaerobaculaceae* genomes and were enriched compared to non-sponge genomes. Cadherins are another adhesion molecule and both *Microtrichaceae* and *Thermoanaerobaculaceae* genomes from sponges were enriched in a cadherin domain, while *Microtrichaceae* were additionally enriched in a cadherin-like beta-sandwich domain, suggesting high potential for this form of adhesion (Figure 5). Finally, we note that in some cases, genes encoding fibronectin and cadherin domains were found on the same coding sequence, suggesting potential fusion of these domains.

Secretion systems (SS) can be used to interact with either adjacent microbes or host cells, however, our data showed that SS are mostly absent from our sponge symbionts, except for type II SS which are mostly involved in nutrient acquisition (Nivaskumar & Francetic, 2014). All *Thermoanaerobaculaceae* genomes encoded proteins associated with type II SS, however the number of copies of type II SS genes were enriched in non-sponge associated genomes (Figure 4). While fewer type II SS proteins were encoded in *Microtrichaceae* genomes, these were enriched in sponge-associated genomes and all symbionts possessed the type II SS gene, which was almost absent in non-sponge genomes. Finally, proteins with type II SS annotations were mostly absent from *Spirochaetaceae* and hence no enrichment was observed for those associated with sponges.

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 encoded an abundance of glycosyl hydrolases (GH), in particular those that act on sialic acids (GH33; sialidase) and fucose (GH29; fucosidase) (Figure 3). Genes encoding carbohydrate esterases (CE) were also abundant in *Spirochaetaceae* genomes, however these were not enriched compared to non-sponge genomes. *Thermoanaerobaculaceae* genomes from sponges were also enriched for GH33 and additionally showed enrichment for the CE family 10, which was encoded in all sponge-derived genomes and only half of the non-sponge genomes (Figure 4). *Microtrichaceae* showed no enrichment for carbohydrate-active enzymes and GHs were encoded in less than half of sponge-derived genomes. However, CE family 10 was encoded in a large proportion of sponge-associated genomes (81%), indicating capacity for carbohydrate metabolism.

We observed enrichment in genes annotated as sulfatases and sulfotransferases in sponge symbiont genomes, suggesting these may act with the CAZymes outlined above to break down and remodel sulfated polysaccharides. Specifically, we found that *Spirochaetaceae* genomes from sponges are highly enriched in sulfatases with an average of 64 copies per genome and additionally show an enrichment in sulfotransferases and arylsulfotransferases (Figure 3). Similarly, sulfatases were encoded in all *Thermoanaerobaculaceae* genomes, however, copy numbers were far higher in sponge-associated genomes, with an average of 57 copies compared to 13 in non-sponge genomes (Figure 4). The sulfatase modifying enzyme was also encoded in all *Thermoanaerobaculaceae* genomes but again demonstrated gene copy number enrichment in

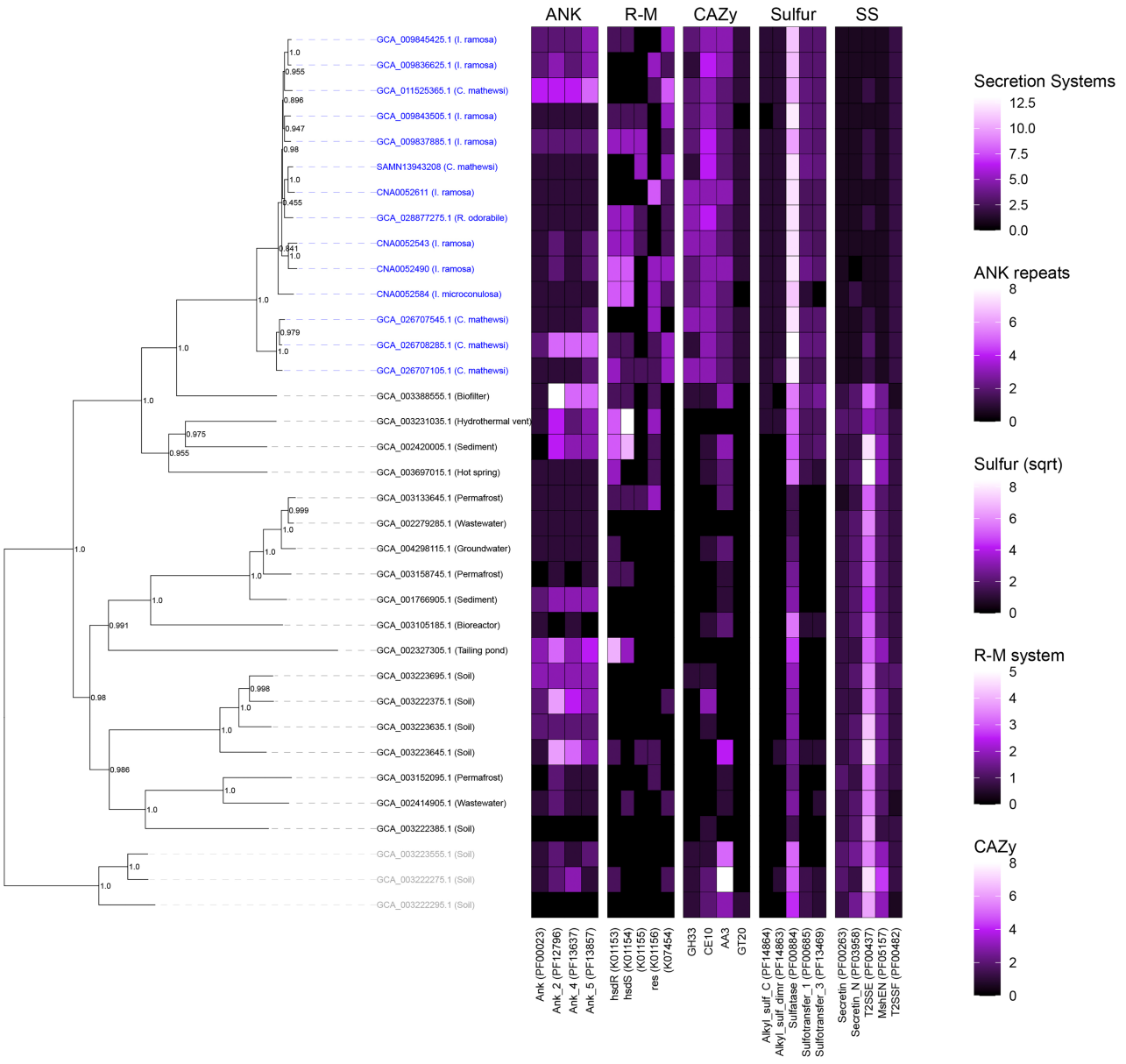

FIGURE 4 Phylogenomic relationships and gene enrichment patterns in sponge-associated and non-sponge-associated *Thermoanaerobaculaceae* genomes. Branches with blue labels indicate sponge-associated genomes, branches with black labels indicate non-sponge-associated genomes while grey labels indicate outgroups and weren't included in the enrichment analysis. Tip labels refer to the genome assembly ID with isolation source or sponge host species in parentheses. Numbers at nodes indicate support values. Heatmap indicates copy numbers for a selection of genes in key symbiotic signatures and are labelled by their gene codes (where available). ANK, ankyrin repeat proteins; CAZy, carbohydrate-active enzymes; R–M, restriction–modification system enzymes; SS, secretion systems; Sulfur, sulfatases and sulfatransferases.

sponge-associated genomes. Sulfatases were encoded 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.

Additional enrichment patterns in genes of interest

Taurine dioxygenase was encoded in all sponge-associated genomes of *Spirochaetaceae* and *Thermoanaerobaculaceae* and was absent in all non-sponge genomes. In the case of *Microtrichaceae*, most genomes encoded taurine dioxygenase (69%), however, this was not enriched compared to non-sponge genomes. Although our annotation pipeline did not

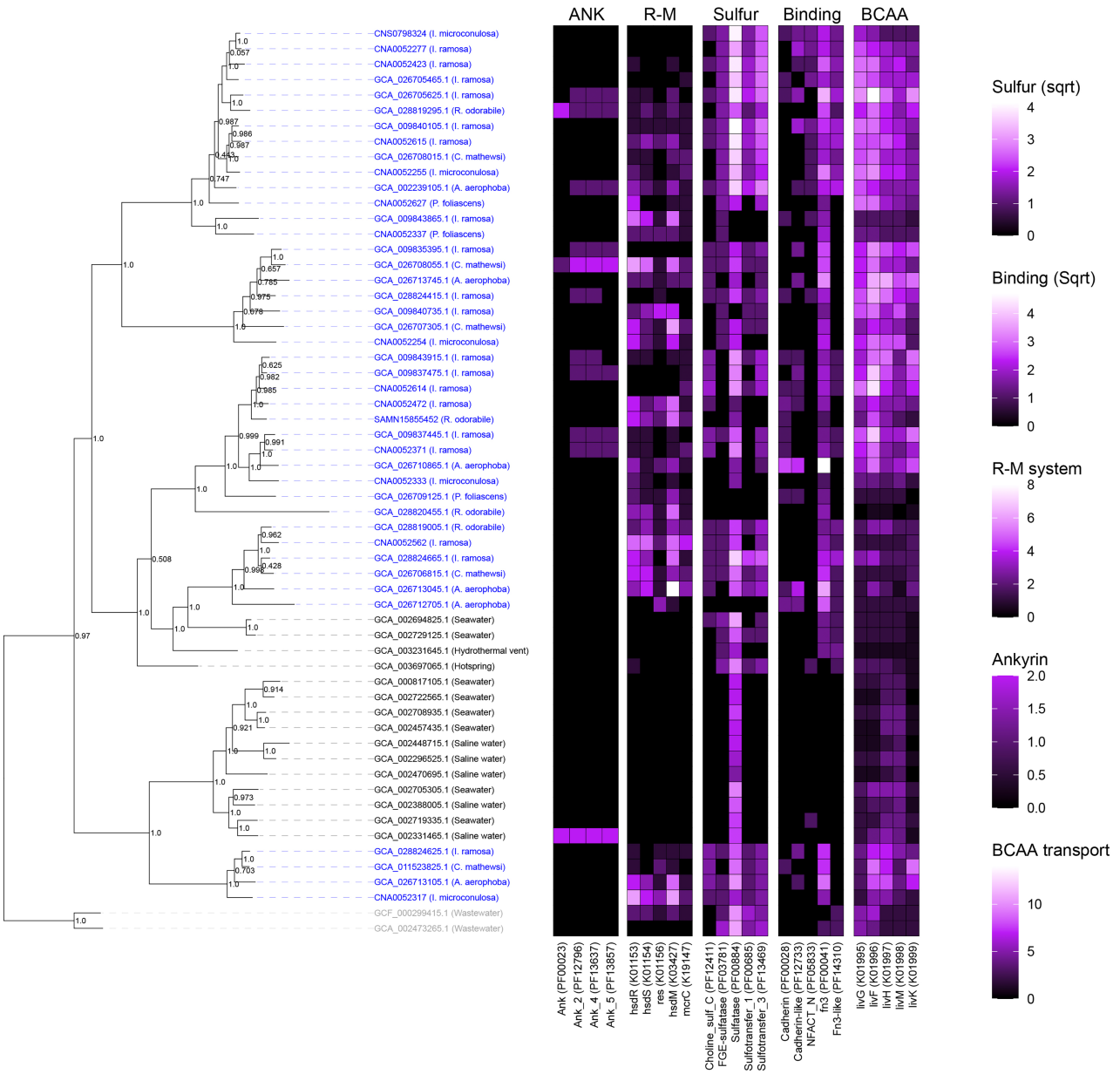

FIGURE 5 Phylogenomic relationships and gene enrichment patterns in sponge-associated and non-sponge-associated *Microtrichaceae* genomes. Branches with blue labels indicate sponge-associated genomes, branches with black labels indicate non-sponge-associated genomes while grey labels indicate outgroups and weren't included in the enrichment analysis. Tip labels refer to the genome assembly ID with isolation source or sponge host species in parentheses. Numbers at nodes indicate support values. Heatmap indicates copy numbers for a selection of genes in key symbiotic signatures and are labelled by their gene codes (where available). ANK, ankyrin repeat proteins; BCAA, branch chain amino acid; R–M, restriction–modification system enzymes; Sulfur, sulfatases and sulfatransferases.

identify genes classified as the taurine transport system (*tauABC*), all symbiont genomes (except for one spirochaete) were found to encode at least one of *tauABC* homologues using tblastx (Tables S6–S8), suggesting these symbionts can import and metabolize host-derived taurine. Interestingly, the reduction of taurine results in sulfite and our 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 and we found that *Spirochaetaceae* genomes from sponges were enriched for all three urease subunits (*ureABC*) (Figure 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,

genes encoding ureases along with accessory proteins and the transport system were rarely found in *Microtrichaceae* and *Thermoanaerobaculaceae* genomes. Although genes involved in urea metabolism may also be present in other microbial phyla from sponges, our results suggest that urea metabolism by sponge-associated *Spirochaetaceae* is a key function provided by these symbionts.

One of the most heavily enriched genes in sponge symbionts was phytanoyl-CoA dioxygenase (*phyH*) in sponge-associated *Spirochaetaceae*, 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 8 and 14 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).

DISCUSSION

Sponge symbionts investigated are likely to represent a facultative symbiosis

The phylogenomic and genomic content for three of five families of microbes that previously showed evidence of cophylogeny with coral reef invertebrates demonstrated characteristic genomic signatures indicating adaptation to a host-associated environment. Reduced genome size and low GC content are often described as characteristics of obligate intracellular symbionts. Our results are consistent with previous estimates of sponge symbiont genome sizes, with genome size and GC content generally higher in sponge symbionts compared to their free-living counterparts (Horn et al., 2016). (Horn et al., 2016). This suggests 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 retain the necessary genes for both free-living and host-associated lifestyles. Hence, although symbiont genomes were recovered from sponges, it is possible they also exist in the surrounding environment. One exception was two *Nitrosopumilaceae* genomes that had genome sizes far smaller (<0.85 Mbp) 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. Hence, future studies may benefit from research aimed specifically towards these potentially intracellular, obligate symbionts.

Despite previous reports of cophylogenetic patterns in *Endozoicomonadaceae* and *Nitrosopumiliaceae* (O'Brien et al., 2021; Pollock et al., 2018), we saw that the genomes analysed did not cluster into sponge-associated clades and subsequently very little genetic enrichment was observed in sponge-associated microbial genomes compared to genomes derived from other environments. In the case of *Endozoicomonadaceae*, this likely reflects that available genomes are from host-associated environments and therefore no gene enrichment is observed between those adapted to sponges compared to other hosts. Similarly, both *Endozoicomonadaceae* and *Nitrosopumilaceae* are thought to contain species with either specialist or generalist symbiosis associations (Pollock et al., 2018; Thomas et al., 2016; Zhang et al., 2019), and therefore genomes of generalist symbionts may not cluster with sponge-associated clades. In contrast, the remaining microbial families, *Spirochaetaceae*, *Microtrichaceae* and *Thermoanaerobaculaceae* all showed monophyletic sponge clades and enrichment for genes in a range of symbiont characteristics. We therefore focussed our efforts on describing the genomic evidence for their adaptation to the sponge environment and demonstrated their metabolic potential, highlighting functions which may be interpreted as beneficial to the host.

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). Our analysis showed that *Spirochaetaceae* and *Thermoanaerobaculaceae* were both enriched for ARPs and while these were not enriched in *Microtrichaceae*, they were still encoded in most genomes. Our 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 (Robbins et al., 2019), and the genome of the *Drosophila melanogaster* symbiont, *Wolbachia pipientis*, has among the highest number of copies of ARPs of any prokaryote (Wu et al., 2006). In addition, ARPs secreted by bacterial pathogens have been shown to facilitate host infection (Habyarimana et al., 2008; Price, Al-Khodor, et al., 2010). Taken together, an enrichment in ARPs appears to be a common signature of symbiotic microbes from sponges, consistent with 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). Our results showed that both *Spirochaetaceae* and *Thermoanaerobaculaceae* genomes encoded 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, *Ruegeria* from sponges and *Endozoicomonas* from coral, all encode a high number of WD40 domains within their genomes (Díez-Vives et al., 2018; Kamke et al., 2014; Robbins et al., 2019). Overall, our data suggests that these symbiont lineages have adapted to take advantage of ARPs and WD40 domains over alternative ELPs, such as tetratricopeptide repeats (TPRs), leucine-rich repeats (LRRs), and HEAT repeats, which can also be enriched in sponge symbionts (Kamke et al., 2014; Robbins et al., 2021; Zhang et al., 2019).

One method to avoid phagocyte killing that has received little attention in sponges is 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 oxy-radical damage and help cells survive phagocytosis (Battistoni, 2003; Broxton & Culotta, 2016; Guo et al., 2023). For example, the pathogen *Salmonella typhimurium* showed decreased survival against macrophages following a knockout mutation of the *SodC* gene (De Groote et al., 1997), while the pathogen *Mycobacterium tuberculosis* showed an up-regulation of *SodC* in response to macrophages (D'Orazio et al., 2001). Our data show that all three microbial families were enriched in the extracellular Cu-Zn family of SOD (*SodC*) indicating protection against superoxide radicals that can be generated by sponge archeocytes/amoebocytes (Mukherjee et al., 2015; Peskin et al., 1998). While we also observed enrichment in the Mn-Fe family of SOD, 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 to control microbial infection, or as a metabolic by-product unrelated to immune function, and the Cu-Zn family can show flexibility in its requirements to assist bacterial colonization (Battistoni, 2003).

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 (Stones & Krachler, 2016). Previous research has suggested that sponge symbionts might form this attachment through cell adhesion molecules such as fibronectin and cadherin (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). Our data show that all three microbial families have enrichment in the fibronectin type III domain, suggesting that 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. This adhesin may allow bacterial cells to attach to their own secreted fibronectin, where an enrichment in the fibronectin type III domain may provide more binding sites for biofilm formation as well as an attachment to the host (Robbins et al., 2021). 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. Our data show enrichment in cadherin domains in both *Microtrichaceae* and *Thermoanaerobaculaceae* from sponges, suggesting another mechanism for host attachment and biofilm formation. Moreover, we found that *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, suggesting a carbohydrate-binding function is used in host attachment (Anantharaman & Aravind, 2010).

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). We found that all three families showed 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, enrichment in the associated enzymes appears to be a signature trait among sponge symbionts (Horn et al., 2016; Robbins et al., 2021; Zhang et al., 2019).

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 metabolize carbohydrates (Warnecke et al., 2007), and here we found an enrichment in genes annotated as glycosyl hydrolases (GH) in *Spirochaetaceae* genomes, which also encoded abundant carbohydrate esterases (CE) although these weren't enriched. Similarly, *Thermoanaerobaculaceae* showed 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) hypothesised to break the sialic acid found in the sponge mesohyl, which appears to be common among sponge symbionts and shows a relatively higher gene expression than other GHs (O'Brien et al., 2023; Robbins et al., 2021). Similarly, an enrichment in fucosidase (GH29) may be an adaptation to the fucose found in coral mucus or released by macroalgae, which makes up part of the DOM assimilated by coral reef sponges (Hadaidi et al., 2019; Rix et al., 2016). Thus, Spirochaetaceae and Thermoanaerobaculaceae may contribute to the ecological role of sponges given their enrichment in carbohydrate-active enzymes.

The sponge-associated *Spirochaetaceae* and *Thermoanaerobaculaceae* genomes analysed here also had an enrichment in genes allowing them to metabolize 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 synthesized, 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 utilize mucin glycoproteins as a nutrient source during colonization (Luis et al., 2021). Given the enrichment in both sulfatases and sulfotransferases, along with their co-occurrence with enriched GHs described above, it seems likely these symbionts are involved in both degrading and synthesizing 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 organization of the sponge.

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 the removal of sponge waste products. Spirochaetes are not known for their ability to break down urea (Solomon et al., 2010), yet our

sponge-associated *Spirochaetaceae* genomes showed an enrichment of ureases and urea transport genes. This may represent an unusual 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 urea metabolism has also been proposed as a method to release nitrogen to the microbiome (Moitinho-Silva et al., 2017; Siegl et al., 2011; Su et al., 2013). For example, cleavage of urea results in ammonia which could then be used as a nitrogen source for ammonia oxidizing microbes that occupy the sponge microbiome. Therefore, urea degradation may not only remove metabolic waste but also recycle the nitrogen within the microbiome.

Potential biosynthetic assimilation of host-derived taurine

Taurine metabolism has been of particular interest in sponge microbiology, as evidence suggests some symbionts use host-derived taurine as a source of sulfur (Moeller et al., 2023; O'Brien et al., 2023). Both *Spirochaetaceae* and *Thermoanaerobaculaceae* were heavily enriched in taurine dioxygenase, highlighting the potential for taurine degradation to sulfite. Additionally, *Spirochaetaceae, Microtrichaceae* and *Thermoanaerobaculaceae* all had an enrichment in a sulfite reductase domain, suggesting that sulfite from taurine degradation could be used for the assimilation of sulfur into cellular components such as amino acids and coenzymes (Crane & Getzoff, 1996). Similar results have been observed in the microbiome of the sponge *Ianthella basta*, where some symbiont genomes encoded pathways for the transport and degradation of taurine and the reduction of sulfite, suggesting a gain of cellular sulfur via the degradation of host-derived taurine (Engelberts et al., 2023). Taken together, there is growing evidence that sponge symbionts have adapted to exploit the abundance of taurine in sponges.

High copy numbers of phytanoyl-CoA dioxygenase in sponge symbionts

One particularly striking pattern we observed in our data was an enrichment in phytanoyl-CoA dioxygenase (*phyH*). This was most obvious in our *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 characterization of the bacterial *phyH* genes has suggested roles as diverse as involvement in quorum sensing and metabolism of dissolved organic phosphorus (Hao et al., 2010; Martinez et al., 2010), and further work would benefit from resolving the function of *phyH* in sponge-associated microbes.

Conclusions

Mounting evidence suggests that carbohydrate degradation using glycosyl hydrolases (in particular sialidases), avoiding phagocytosis with ELPs, defence against MGEs using restriction enzymes, attachment to the host with fibronectins and taurine degradation and assimilation, are all common traits of sponge symbionts (Robbins et al., 2021). We further expand the sponge symbiont molecular repertoire to show an enrichment of SOD, which may protect against phagocytosis and oxidative damage. We also show that sulfatases and sulfotransferases potentially work with glycosyl hydrolases to break down and remodel carbohydrates. Finally, we show that genes involved in urea metabolism in *Spirochaetaceae* genomes appear to be a novel function within the Spirochaetes, limited to those associated with sponges. These enrichment patterns suggest a multitude of ways symbionts have adapted to thrive in a sponge environment. While evidence suggests some of these traits may be acquired through lateral gene transfer within the sponge microbiome (Robbins et al., 2021), we cannot rule out that some traits may also be the result of shared ancestry of sponge symbionts. Although a generalist suite of symbiotic traits exists, our results show that sponge-specific symbionts carry their own unique characteristics that reflect their evolution towards a sponge-associated lifestyle.

AUTHOR CONTRIBUTIONS

Paul A. O'Brien: Conceptualization; data curation; formal analysis; writing – original draft; methodology. **Steven J. Robbins:** Conceptualization; methodology; supervision; writing – review and editing. **Shangjin Tan:** Methodology; writing – review and editing. **Laura Rix:** Supervision; writing – review and editing. **David J. Miller:** Supervision; writing – review and editing. **Nicole S. Webster:** Conceptualization; supervision; writing – review and editing. **Guojie Zhang:** Conceptualization; writing – review and editing. **David G. Bourne:** Conceptualization; supervision; funding acquisition; writing – review and editing.

ACKNOWLEDGEMENTS

This work was funded by the Beijing Genome Institute. P.A.O. was supported by an AIMS@JCU postgraduate

research scholarship. Open access publishing facilitated by James Cook University, as part of the Wiley - James Cook University agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All genomes used in this study are publicly available at either the National Center for Biotechnology Information or the China National GeneBank DataBase (details in Tables S1–S5). All code used for the analysis along with outputs from the enrichment analysis are available at https://github.com/paobrien/SpongeSymbionts. The DOI for this repository: http://doi.org/10.5281/zenodo.13266492.

REFERENCES

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.

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.

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

Battistoni, A. (2003) Role of prokaryotic Cu, Zn superoxide dismutase in pathogenesis. *Biochemical Society Transactions*, 31(6), 1326–1329. Available from: https://doi.org/10.1042/bst0311326

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.

Broxton, C.N. & Culotta, V.C. (2016) SOD enzymes and microbial pathogens: surviving the oxidative storm of infection. *PLoS Pathogens*, 12(1), 8–13. Available from: https://doi.org/10.1371/journal.ppat.1005295

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. Available from: https://doi.org/10.1111/1462-2920.15642

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.

Cooke, I., Mead, O., Whalen, C., Boote, C., Moya, A., Ying, H. et al. (2019) Molecular techniques and their limitations shape our view of the holobiont. *Zoology*, 137, 125695.

Costa, T.R.D., Felisberto-Rodrigues, C., Meir, A., Prevost, M.S., Redzej, A., Trokter, M. et al. (2015) Secretion systems in gram-negative bacteria: structural and mechanistic insights. *Nature Reviews Microbiology*, 13(6), 343–359.

Crane, B.R. & Getzoff, E.D. (1996) The relationship between structure and function for the sulfite reductases. *Current Opinion in Structural Biology*, 6(6), 744–756.

Cullen, P.A., Haake, D.A. & Adler, B. (2004) Outer membrane proteins of pathogenic spirochetes. *FEMS Microbiology Reviews*, 28(3), 291–318. Available from: https://doi.org/10.1016/j.femsre.2003.10.004

Dash, S., Duraivelan, K. & Samanta, D. (2021) Cadherin-mediated host–pathogen interactions. *Cellular Microbiology*, 23(5), e13316.

De Groote, M.A., Ochsner, U.A., Shiloh, M.U., Nathan, C., McCord, J.M., Dinauer, M.C. et al. (1997) Periplasmic superoxide dismutase protects salmonella from products of phagocyte NADPH-oxidase and nitric oxide synthase. *Proceedings of the National Academy of Sciences of the United States of America*, 94(25), 13997–14001.

Díez-Vives, C., Esteves, A.I.S., Costa, R., Nielsen, S. & Thomas, T. (2018) Detecting signatures of a sponge-associated lifestyle in bacterial genomes. *Environmental Microbiology Reports*, 10(4), 433–443. Available from: https://doi.org/10.1111/1758-2229.12655

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.

Engelberts, J.P., Robbins, S.J., Herbold, C.W., Moeller, F.U., Jehmlich, N., Laffy, P.W. et al. (2023) Metabolic reconstruction of the near complete microbiome of the model sponge *Ianthella basta*. *Environmental Microbiology*, 25(3), 646–660.

Fan, L., Reynolds, D., Liu, M., Stark, M., Kjelleberg, S., Webster, N.S. et al. (2012) Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proceedings of the National Academy of Sciences of the United States of America*, 109(27), E1878–E1887. Available from: https://doi.org/10.1073/pnas.1203287109

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.

Gao, Z.M., Wang, Y., Tian, R.M., Wong, Y.H., Batang, Z.B., Al-Suwailem, A.M. et al. (2014) Symbiotic adaptation drives genome streamlining of the cyanobacterial sponge symbiont "*Candidatus Synechococcus spongiarum*". *MBio*, 5(2), 1–11. Available from: https://doi.org/10.1128/mBio.00079-14

Giovannoni, S.J., Cameron Thrash, J. & Temperton, B. (2014) Implications of streamlining theory for microbial ecology. *ISME Journal*, 8(8), 1553–1565.

Guo, L., Tang, J., Tang, M., Luo, S. & Zhou, X. (2023) Reactive oxygen species are regulated by immune deficiency and toll pathways in determining the host specificity of honeybee gut bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 120(33), e2219634120.

Habyarimana, F., Al-khodor, S., Kalia, A., Graham, J.E., Price, C.T., Garcia, M.T. et al. (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. Available from: https://doi.org/10.1007/s00338-018-01758-5

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.

Horn, H., Slaby, B.M., Jahn, M.T., Bayer, K., Moitinho-Silva, L., Förster, F. et al. (2016) An enrichment of CRISPR and other defense-related features in marine sponge-associated microbial metagenomes. *Frontiers in Microbiology*, 7(NOV), 1–15. Available from: 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. Available from: https://doi.org/10.1126/science.1179555

Hymes, J.P. & Klaenhammer, T.R. (2016) Stuck in the middle: fibronectin-binding proteins in gram-positive bacteria. *Frontiers in Microbiology*, 7, 1–9. Available from: https://doi.org/10.3389/fmicb.2016.01504

Jahn, M.T., Arkhipova, K., Markert, S.M., Stigloher, C., Lachnit, T., Pita, L. et al. (2019) A phage protein aids bacterial symbionts in

eukaryote immune evasion. *Cell Host and Microbe*, 26(4), 542–550.e5. Available from: https://doi.org/10.1016/j.chom.2019.08.019

Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T. & Aluru, S. (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, 9(1), 1–8.

Jernigan, K.K. & Bordenstein, S.R. (2014) Ankyrin domains across the tree of life. *PeerJ*, 2, e264. Available from: https://doi.org/10.7717/peerj.264

Kamke, J., Rinke, C., Schwientek, P., Mavromatis, K., Ivanova, N., Sczyrba, A. et al. (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), e87353. Available from: https://doi.org/10.1371/journal.pone.0087353

Kamke, J., Sczyrba, A., Ivanova, N., Schwientek, P., Rinke, C., Mavromatis, K. et al. (2013) Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *ISME Journal*, 7(12), 2287–2300. Available from: https://doi.org/10.1038/ismej.2013.111

Luis, A., Baslé, A., Byrne, D., Wright, G., London, J., Chunsheng, J. et al. (2021) Sulfated host glycan recognition by carbohydrate sulfatases of the human gut microbiota. *Nature Chemical Biology*, 18, 841–849.

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.

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.

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. Available from: https://doi.org/10.1073/pnas.0708855104

Moeller, F.U., Herbold, C.W., Schintlmeister, A., Mooshammer, M., Motti, C., Glasl, B. et al. (2023) Taurine as a key intermediate for host-symbiont interaction in the tropical sponge *Ianthella basta*. *ISME Journal*, 17(8), 1–16.

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. Available from: https://doi.org/10.1038/ismej.2017.25

Moran, N.A. & Baumann, P. (2000) Bacterial endosymbionts in animals. *Current Opinion in Microbiology*, 3(3), 270–275. Available from: https://doi.org/10.1016/S1369-5274(00)00088-6

Morley, S.A., Berman, J., Barnes, D.K.A., Carbonell, C.d.J., Downey, R.V. & Peck, L.S. (2016) Extreme phenotypic plasticity in metabolic physiology of Antarctic demosponges. *Frontiers in Ecology and Evolution*, 3(1), 1–10. Available from: https://doi.org/10.3389/fevo.2015.00157

Mukherjee, S., Ray, M. & Ray, S. (2015) Phagocytic efficiency and cytotoxic responses of Indian freshwater sponge (*Eunapius carteri*) cells isolated by density gradient centrifugation and flow cytometry: a morphofunctional analysis. *Zoology*, 118(1), 8–18.

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. Available from: https://doi.org/10.1111/mec.12384

Nivaskumar, M. & Francetic, O. (2014) Type II secretion system: a magic beanstalk or a protein escalator. *Biochimica et Biophysica Acta Molecular Cell Research*, 1843(8), 1568–1577.

O'Brien, P.A., Andreakis, N., Tan, S., Miller, D.J., Webster, N.S., Zhang, G. et al. (2021) Testing cophylogeny between coral reef invertebrates and their bacterial and archaeal symbionts. *Molecular Ecology*, 30(15), 3768–3782. Available from: https://doi.org/10.1111/mec.16006

O'Brien, P.A., Tan, S., Frade, P.R., Robbins, S.J., Engelberts, J.P., Bell, S.C. et al. (2023) Validation of key sponge symbiont pathways using genome-centric metatranscriptomics. *Environmental Microbiology*, 25, 3207–3224. Available from: https://doi.org/10.1111/1462-2920.16509

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. Available from: https://doi.org/10.1093/nar/gku734

Pankov, R. & Yamada, K.M. (2002) Fibronectin at a glance. *Journal of Cell Science*, 115(20), 3861–3863.

Parks, D.H., Chuvochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.-A. & Hugenholtz, P. (2022) GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Research*, 50(D1), D785–D794.

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.

Pearson, W.R. (2013) An introduction to sequence similarity ("homology") searching. *Current Protocols in Bioinformatics*, 42, 1–8. Available from: https://doi.org/10.1002/0471250953.bi0301s42

Peskin, A.V., Labas, Y.A. & Tikhonov, A.N. (1998) Superoxide radical production by sponges Sycon sp. *FEBS Letters*, 434(1–2), 201–204.

Pollock, F.J., McMinds, R., Smith, S., Bourne, D.G., Willis, B.L., Medina, M. et al. (2018) Coral-associated bacteria demonstrate phylosymbiosis and cophylogeny. *Nature Communications*, 9(4921), 1–13. Available from: https://doi.org/10.1038/s41467-018-07275-x

Price, C., Al-Khodor, S., Al-Quadan, 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.

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. Available from: 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. Available from: https://doi.org/10.1111/mec.13812

Rix, L., De Goeij, J.M., Mueller, C.E., Struck, U., Middelburg, J.J., Van Duyl, F.C. et al. (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. et al. (2020) Heterotrophy in the earliest gut: a single-cell view of heterotrophic carbon and nitrogen assimilation in sponge-microbe symbioses. *ISME Journal*, 14, 2554–2567. Available from: 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. et al. (2019) A genomic view of the reef-building coral *Porites lutea* and its microbial symbionts. *Nature Microbiology*, 4(12), 2090–2100. Available from: https://doi.org/10.1038/s41564-019-0532-4

Robbins, S.J., Song, W., Engelberts, J.P., Glasl, B., Slaby, B.M., Boyd, J. et al. (2021) A genomic view of the microbiome of coral reef demosponges. *ISME Journal*, 15, 1641–1654. Available from: https://doi.org/10.1038/s41396-020-00876-9

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.
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.
Siegl, A., Kamke, J., Hochmuth, T., Piel, J., Richter, M., Liang, C. et al. (2011) Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. *ISME Journal*, 5(1), 61–70.
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. Available from: https://doi.org/10.1038/ismej.2017.101
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. Available from: https://doi.org/10.3354/ame01390
Stones, D.H. & Krachler, A.M. (2016) Against the tide: the role of bacterial adhesion in host colonization. *Biochemical Society Transactions*, 44(6), 1571–1580.
Su, J., Jin, L., Jiang, Q., Sun, W., Zhang, F. & Li, Z. (2013) Phylogenetically diverse ure C genes and their expression suggest the urea utilization by bacterial symbionts in marine sponge *Xestospongia testudinaria*. *PLoS One*, 8(5), e64848.
Team, R.C. (2018) *R: a language and environment for statistical computing (3.5.0)*. Vienna: R Foundation for Statistical Computing. https://www.r-project.org/
Thomas, T., Moitinho-Silva, L., Lurgi, M., Björk, J.R., Easson, C., Astudillo-García, C. et al. (2016) Diversity, structure and convergent evolution of the global sponge microbiome. *Nature Communications*, 7(11870), 1–12. Available from: https://doi.org/10.1038/ncomms11870
Thomas, T., Rusch, D., DeMaere, M.Z., Yung, P.Y., Lewis, M., Halpern, A. et al. (2010) Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *ISME Journal*, 4(12), 1557–1567. Available from: https://doi.org/10.1038/ismej.2010.74
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.
Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T. et al. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature*, 450(7169), 560–565. Available from: https://doi.org/10.1038/nature06269
Webster, N.S. & Thomas, T. (2016) The sponge Hologenome. *MBio*, 7(2), 1–14. Available from: https://doi.org/10.1128/mBio.00135-16
Wu, D., Daugherty, S.C., Van Aken, S.E., Pai, G.H., Watkins, K.L., Khouri, H. et al. (2006) Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biology*, 4(6), 1079–1092. Available from: https://doi.org/10.1002/adsc.200505252
Xu, C. & Min, J. (2011) Structure and function of WD40 domain proteins. *Protein and Cell*, 2(3), 202–214. Available from: https://doi.org/10.1007/s13238-011-1018-1
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.
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. Available from: https://doi.org/10.1128/msystems.00288-19
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.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: O'Brien, P.A., Robbins, S.J., Tan, S., Rix, L., Miller, D.J., Webster, N.S. et al. (2024) Comparative genomics identifies key adaptive traits of sponge-associated microbial symbionts. *Environmental Microbiology*, 26(9), e16690. Available from: https://doi.org/10.1111/1462-2920.16690