# Comparative genomics provides insight into the function of broad-host range sponge symbionts

**Samantha C. Waterworth[a,b], Shirley Parker-Nance[b,c], Jason C. Kwan[a], Rosemary A. Dorrington* [b.d]**

[a]Division of Pharmaceutical Sciences, University of Wisconsin, 777 Highland Ave., Madison, Wisconsin 53705, USA

[b]Department of Biochemistry and Microbiology, Rhodes University, Makhanda, South Africa

[c]South African Environmental Observation Network, Elwandle Coastal Node, Port Elizabeth, South Africa

[d]South African Institute for Aquatic Biodiversity, Makhanda, South Africa

***Correspondence:**

Rosemary A. Dorrington

r.dorrington@ru.ac.za

**Running title:** Genomics of sponge-associated bacterial symbionts

**Competing Interests**

The authors declare no competing interests, financial or otherwise, in relation to the work described here.

**ABSTRACT**

As the oldest extant metazoans, sponges (Phylum *Porifera*) have been forming symbiotic relationships with microbes that may date back as far as 700 million years. Most symbionts are conserved within a narrow host range and perform specialized functions. However, there are widely distributed bacterial taxa such as *Poribacteria*, *SAUL* and *Tethybacterales* that are found in a broad range of invertebrate hosts. Here, we added eleven new genomes to the *Tethybacterales* order, identified a novel family, and show that functional potential differs between the three *Tethybacterales* families. We compare the *Tethybacterales* with the well-characterized *Entoporibacteria* and show that these broad-host range, sponge-associated bacteria likely perform distinct functions within their hosts and that their respective phylogenies are incongruent with their host phylogenies. These results suggests that ancestors of these bacteria may have undergone multiple association events, rather than a single association event followed by co-evolution.

**IMPORTANCE**

Marine sponges often form symbiotic relationships with bacteria that fulfil a specific need within the sponge holobiont, and these symbionts are often conserved within a narrow range of related taxa. To date, there exist only three know bacterial taxa (*Entoporibacteria*, *SAUL* and *Tethybacterales*) that are globally distributed and found in a broad range of sponge hosts, and little is known about the latter two. Understanding what distinguishes these broad-host range symbionts from specialized symbionts will provide insight into the mechanisms by which sponges form these symbioses. We show that the functional potential of broad-host range symbionts is conserved at a family level and that these symbionts have been acquired several times over evolutionary history.

This contrasts with specialized symbionts, where function is often a strain-specific trait and have co-evolved with their host following a single association event.

**Keywords:** *Latrunculiidae, Tethybacterales, Poribacteria*, Symbiosis, *Porifera*, Comparative Genomics.

## INTRODUCTION

Sponges (Phylum *Porifera*) are the oldest living metazoans and they play an important role in maintaining the health of marine ecosystems(1). Sponges are remarkably efficient filter feeders, acquiring nutrients via phagocytosis of particulate matter, compromising mainly microbes, from the surrounding water(2). Since their emergence almost 700 million years ago, sponges have evolved close associations with microbial symbionts that provide services essential for the fitness and survival of the host in diverse ecological niches(3). These symbionts are involved in a diverse array of beneficial processes, including the cycling of nutrients(4, 5) such as nitrogen(5–10), sulfur (11, 12), phosphate(13, 14), the acquisition of carbon(15, 16), and a supply of vitamins(17–20) and amino acids(17). They can play a role in the host sponge life cycle, such as promoting larval settlement(21). In addition, some symbionts provide chemical defenses against predators and biofouling through the production of bioactive compounds(22–25). In turn, the sponge host can provide symbionts with nutrients and minerals, such as creatinine and ammonia as observed in *Cymbastela concentrica* sponges(8). In most cases, these specialized functions are performed by bacterial populations that are conserved within a given host taxon.

As filter-feeders, sponges encounter large quantities of bacteria and other microbes. How sponges are able to distinguish between prey bacteria and those of potential benefit to the sponge, and the establishment of symbiotic relationships, is still not well understood, but the structure and composition of bacterial lipopolysaccharide, peptidoglycan, or flagellin may aid the host sponge in distinguishing symbionts from prey(26). Sponge hosts encode an abundance of Nucleotide-binding domain and Leucine-rich Repeat (NLR) receptors, that recognize different microbial ligands and potentially allow for distinction between symbionts, pathogens, and prey(27). Additionally, it has recently been shown that phages produce ankyrins which modulate the sponge immune response and allow for colonization by bacteria(28).

Symbionts are often specific to their sponge host with enriched populations relative to the surrounding seawater(29). However, there are a small number of "cosmopolitan" symbionts that are ubiquitously distributed across phylogenetically distant sponge hosts. The *Poribacteria* and "*sponge-associated unclassified lineage*" (*SAUL*)(30) are examples of such cosmopolitan bacteria species. The phylum *Poribacteria* were thought to be exclusively found in sponges(31). However, the identification of thirteen putative *Poribacteria*-related metagenome-assembled genomes (MAGs) from ocean water samples(32), led to the reclassification and distinction of sponge-associated *Entoporibacteria* and free-living *Pelagiporibacteria* within the phylum(33). *Entoporibacteria* are associated with phylogenetically divergent sponge hosts in distant geographic locations, with no apparent correlations between their phylogeny and that of their sponge host or location(33, 34). Different *Poribacteria* phylotypes have been detected within the same sponge species(35). The *Entoporibacteria* carry several genes(36, 37) that encode enzymes responsible for the degradation of carbohydrates, and metabolism of sulfates and uronic acid(38–

40), prompting the hypothesis that these bacteria may be involved in the breakdown of the proteoglycan host matrix(40). However, subsequent analyses of *Poribacteria* transcriptomes from the mesohyl of *Aplysina aerophoba* sponges showed that genes involved in carbohydrate metabolism were not highly expressed(39). Instead, there was a higher expression of genes involved in 1,2-propanediol degradation and import of vitamin B12, which together suggest that the bacterium may import vitamin B12 as a necessary cofactor for anaerobic 1,2-propanediol degradation and energy generation(39).

The *SAUL* bacteria belong to the larger taxon of candidate phylum *PAUC34f* and have been detected, although at low abundance, in several sponge species(41). Host-associated *SAUL* bacteria were likely acquired by eukaryotic hosts (sponges, corals, tunicates) at different evolutionary time points and are phylogenetically distinct from their planktonic relatives(41). Previous investigations into the only two SAUL bacteria genomes provided evidence to suggest that these symbionts may play a role in the degradation of host and algal carbohydrates, as well as the storage of phosphate for the host during periods of phosphate limitation(30).

Recently, a third group of ubiquitous sponge-associated betaproteobacterial symbionts were described(42). The proposed new order, the *Tethybacterales*, comprises two families, the *Tethybacteraceae* and the *Persebacteraceae*(42). Based on assessment of MAGs representative of different species within these two families, it was shown that the bacteria within these families had functionally radiated, as they coevolve with their specific sponge host(42). The *Tethybacterales* are distributed both globally and with phylogenetically diverse sponges that represent both high microbial abundance (HMA) and low microbial abundance (LMA) sponges(42). These bacteria

have also been detected in other marine invertebrates, ocean water samples, and marine sediment(42) suggesting that these symbionts may have, at one point, been acquired from the surrounding environment.

*Tethybacterales* are conserved in several sponge microbiomes(43) and can be the numerically dominant bacterial population(12, 44–51), with some predicted to be endosymbiotic(12, 44, 49). In *Amphimedeon queenslandica*, the AqS2 symbiont, *Amphirhobacter heronislandensis* (Family Tethybacteraceae), is co-dominant with sulfur-oxidizing *Gammaproteobacteria* AqS1. *A. heronislandensis* AqS2, is present in all stages of the sponge life cycle(52) and appears to have a reduced genome(12). Interestingly, the AqS2 MAG shares some functional similarity with the codominant AqS1, including the potential to generate energy via carbon monoxide oxidation, assimilate sulfur and produce most essential amino acids(12). However, these sympatric symbionts differ significantly in what metabolites they could possibly transport(12).

There are thus two types of sponge-associated symbionts: those that are conserved within a narrow host range, and those that have a broader host range, such as the *SAUL*, *Entoporibacteria* and *Tethybacterales*. What is currently unknown, is whether these symbionts fulfil the same roles regardless of their host or respond to the needs of the host? In this study we sought to provide answers to these questions, using the dominant, conserved Tethybacterales (strain Sp02-1) symbiont of *Tsitsikamma favus* sponge species (Family *Latrunculiidae*)(53–56) as a springboard into a deeper investigation. Here, we report a comparative study using new and existing *Tethybacterales* genomes and show that functional potential follows that of their taxonomic ranking rather than host-specific adaptation. We also show that the *Tethybacterales* and

*Poribacteria* have distinct functional repertoires, that these bacterial families can co-exist in a single host and that the *Tethybacterales* may represent a more ancient lineage of ubiquitous sponge-associated symbionts.

## MATERIALS AND METHODS

**Sponge Collection and taxonomic identification.** Sponge specimens were collected by SCUBA or Remotely Operated Vehicle (ROV) from multiple locations within Algoa Bay (Port Elizabeth), the Garden Route National Park, and the Tsitsikamma Marine Protected Area. Collection permits were acquired prior to collections from the Department Environmental Affairs (DEA) and the Department of Environment, Forestry and Fisheries (DEFF) under permit numbers: 2015: RES2015/16 and RES2015/21; 2016:RES2016/11; 2017:RES2017/43; 2018: RES2018/44. Collection metadata are provided in Table S1. Sponge specimens were stored on ice during collection, and thereafter at -20 °C. Subsamples collected for DNA extraction were preserved in RNALater (Invitrogen) and stored at -20 °C. Sponge specimens were dissected, thin sections generated and spicules mounted on microscope slides and examined to allow species identification, as done previously (57–59). Molecular barcoding (28S rRNA gene) was also performed for several of the sponge specimens (Fig. S1) as described previously(53).

**Metagenomic sequencing and analysis.** Small sections of each preserved sponge (approx. 2cm$^3$) were pulverized in 2ml sterile artificial seawater (24.6 g NaCl, 0.67 g KCl, 1.36 g $CaCl_2.2H_2O$, 6.29 g $MgSO_4.7H_2O$, 4.66 g $MgCl_2.6H_2O$ and 0.18 g $NaHCO_3$, distilled $H_2O$ to 1 L) with a sterile mortar and pestle. The resultant homogenate was centrifuged at 16000 rpm for 1 min to pellet cellular material. Genomic DNA (gDNA) was extracted using the ZR Fungal/Bacterial DNA

MiniPrep kit (D6005, Zymo Research). Shotgun metagenomic sequencing was performed for four *T. favus* sponge specimens using Ion Torrent platforms. Shotgun metagenomic libraries, of reads 200bp in length, were prepared for each of the four sponge samples respectively (TIC2016-050A, TIC2018-003B, TIC2016-050C and TIC2018-003D) using an Ion P1.1.17 chip. Additional sequence data of 400bp was generated for TIC2016-050A using an Ion S5 530 Chip. TIC2016-050A served as a pilot experiment and we wanted to identify which read length was best for our investigations. However, we did not want to waste additional sequence data and included it when assembling the TIC2016-050A metagenomic contigs so the 400bp reads were included in the assembly of these metagenomes. Metagenomic datasets were assembled into contiguous sequences (contigs) with SPAdes version 3.12.0(60) using the --iontorrent and --only-assembler options. Contigs were clustered into genomic bins using Autometa(61) and manually curated for optimal completion and purity. Validation of the bins was performed using CheckM v1.0.12.(62). Of the 50 recovered genome bins, 5 were of high quality, 13 were of medium quality and 32 were of low quality in accordance with MIMAG standards(63) (Table S2).

**Acquisition and assembly of reference genomes.** The genome of *A. queenslandica* symbiont Aqs2 (GCA_001750625.1) was retrieved from the NCBI database. Similarly, other sponge associated *Tethybacterales* MAGs from the JGI database were downloaded and used as references (3300007741_3, 3300007056_3, 3300007046_3, 3300007053_5, 3300021544_3, 3300021545_3, 3300021549_5, 2784132075, 2784132054, 2814122900, 2784132034 and 2784132053).

Thirty-six raw read SRA datasets from sponge metagenomes were downloaded from the SRA database (Table S3). Illumina reads from these datasets were trimmed using Trimmomatic v0.39(64) and assembled using SPAdes v3.14(60) in --meta mode and resultant contigs were

binned using Autometa(61). This resulted in a total of 393 additional genome bins, the quality of which was assessed using CheckM(62) and taxonomically classified with GTDB-Tk(65) with database release95 (Table S2). A total of 27 bins were classified as AqS2 and were considered likely members of the newly proposed *Tethybacterales* order(42). However, 10 of the 27 bins were low quality and were not used in downstream analyses. In addition, 59 *Poribacteria* genome bins were downloaded from the NCBI database for functional comparison (Table S4) and three were used from the 393 genome bins generated in this study (*Geodia parva* sponge hosts).

**Taxonomic identification.** Partial and full-length 16S rRNA gene sequences were extracted from bins using barrnap 0.9 (https://github.com/tseemann/barrnap). Extracted sequences were aligned against the nr database using BLASTn(66). Genomes were additionally uploaded individually to autoMLST(67) and analyzed in both placement mode and *de novo* mode (IQ tree and ModelFinder options enabled, and concatenated gene tree selected). All bins and downloaded genomes were taxonomically identified using GTDB-Tk(65).

**Genome annotation and metabolic potential analysis.** All bins and downloaded genomes were annotated using Prokka 1.13(68) with NCBI compliance enabled. Protein-coding amino-acid sequences from genomic bins were annotated against the KEGG database using kofamscan(69) with output in mapper format. Custom Python scripts were used to summarize annotation counts (Find scripts here: https://github.com/samche42/Family_matters). Potential Biosynthetic Gene Clusters (BGCs) were identified by uploading genome bins to the antiSMASH web server(70) with all options enabled. Predicted amino acid sequences of genes within each identified gene

cluster were aligned against the nr database using BLASTn(66) to identify the closest homologs. Protein sequences of genes within each identified gene cluster were aligned against the nr database using BLASTn(66) to identify the closest homolog.

**Phylogeny and function of *Tethybacterales* species.** A subset of orthologous genes common to all medium quality *Tethybacterales* genomes/bins was created. Shared amino acid identity (AAI) was calculated with the aai.rb script from the enveomics package(71). 16S rRNA genes were analyzed using BLASTn(66). Functional genes were annotated against the KEGG database using kofamscan(69). Annotations were collected into functional categories and visualized in R (See https://github.com/samche42/Family_matters for all scripts). A Non-metric Multidimensional Scaling (NMDS) plot of the presence/absence metabolic counts was constructed using Bray-Curtis distance using the vegan package(72) in R. Analysis of Similarity (ANOSIM) analyses were also conducted using the vegan package in R using Bray-Curtis distance and 9999 permutations.

**Genome divergence estimates.** Divergence estimates were performed as described previously(73). Briefly, homologous genes in *Tethybacterales* genomes were identified using OMA v. 2.4.2.(74). A subset of homologous genes present in all genomes was created. Homologous genes were aligned using MUSCLE v.3.8.155(75) and clustered into fasta files representing each genome using merge_fastas_for_dNdS.py (See https://github.com/samche42/Family_matters for all scripts). Corresponding nucleotide sequences extracted from PROKKA annotations using multifasta_seqretriever.py. All stop codons were removed using remove_stop_codons.py. All nucleotide sequences, per genome, were concatenated to produce a single nucleotide sequence per genome using the union function from EMBOSS(76).

All amino acid sequences were similarly concatenated. This resulted in a single concatenated nucleotide sequence and a single concatenated amino acid sequence per genome. Concatenated nucleotide sequences were clustered into two fasta files (one nucleotide, one protein sequence) and then aligned using PAL2NAL(77). The resultant alignment was then run in codeml to produce pairwise synonymous substitution rates (dS). Divergence estimates can be determined by dividing pairwise dS values by a given substitution rate, and further divided by 1 million. Pairwise synonymous substitution rates can be found in Table S5. Pairwise divergence values were illustrated as a tree using MEGAX(78). Concatenated amino acid and nucleotide sequences of the 18 orthologous genes were aligned using MUSCLE v.3.8.155(75) and the evolutionary history inferred using the UPMGA method(79) in MEGAX(78) with 10000 bootstrap replicates.

**Identification of unique and host-associated genes in putative symbiont genome bins.** A custom database of genes from all bacterial bins (with the exception of the *Tethybacterales* Sp02-1 symbionts) was created using the "makedb" option in Diamond(80) to identify genes that were unique to the Sp02-1 symbionts. To be exhaustive and screen against the entire metagenome, genes from low-quality genomes (except low-quality Sp02-1 genomes), small contigs (<3000 bp) that were not included in binning and unclustered contigs (i.e., included in binning but were not placed within a bin) were included in this database. Sp02-1 genes were aligned using diamond blast(80). A gene was considered "unique" if the aligned hit shared less than 40% amino acid identity with any other genes from the *T. favus* metagenomes and had no significant hits against the nr database or were identified as pseudogenes. All "unique" Sp02-1 genes annotated as "hypothetical" (both Prokka and NCBI nr database annotations) were removed. Finally, we compared Prokka annotation strings between the three Sp02-1 genes and all other *T. favus* associated genome bins

and excluded any Sp02-1 genes that were found to have the same annotation as a gene in one of the other bins.

## DATA AVAILABILITY

All data can be accessed from the NCBI website under BioProject PRJNA508092.

## RESULTS AND DISCUSSION

The microbiomes of sponges of the *Latrunculiidae* family are highly conserved and are dominated by populations of related betaproteobacterial symbionts. Based on their 16S rRNA gene sequence, the *Betaproteobacteria* Sp02-1 symbionts from different latrunculid sponges are closely related and are likely members of the newly described *Tethybacterales*.

### Characterization of putative betaproteobacteria Sp02-1 genome bin

Genome bin 003B_4, from sponge specimen TIC2018-003B, included a 16S rRNA gene sequence that shared 99.86% identity with the *T. favus*-associated *Tethybacterales* Sp02-1 full-length 16S rRNA gene clone (HQ241787.1). The next closest relatives were uncultured betaproteobacterium 16S clones from a *Xestospongia muta* and *Tethya aurantium* sponges (Fig. S2). Bins 050A_14, 050C_6 and 003D_6 were also identified as possible representatives of the *Tethybacterales* Sp02-1 based on their predicted phylogenetic relatedness. However, they were of low quality and not used in downstream analyses.

Genome bin 003B_4 was used as a representative of the *Tethybacterales* Sp02-1 symbiont. Bin 003B_4 is approximately 2.95 Mbp in size and of medium quality per MIMAG standards(63)

(Table S2), and it has a notable abundance of pseudogenes (~25% of all genes), which resulted in a coding density of 65.27%, far lower than the average for bacteria(81). An abundance of pseudogenes and low coding density, is usually an indication that the genome in question may be undergoing genome reduction (82) similar to other betaproteobacteria in the proposed order of *Tethybacterales*(42).

The *Tethybacterales* Sp02-1 genome encodes all genes necessary for glycolysis, PRPP biosynthesis and most genes required for the citrate cycle and oxidative phosphorylation were detected in the gene annotations. Also present are the genes necessary to biosynthesize valine, leucine, isoleucine, tryptophan, phenylalanine, tyrosine and ornithine amino acids as well as genes required for transport of L-amino acids, proline and branched amino acids. This would suggest that this bacterium may exchange amino acids with the host, as observed previously in both insect and sponge-associated symbioses(8, 83, 84).

A total of 13 genes unique to the *Tethybacterales* Sp02-1 symbiont were identified (i.e., not identified elsewhere in the *T. favus* metagenomes). One gene was predicted to encode an ABC transporter permease subunit that was likely involved in glycine betaine and proline betaine uptake. A second gene encoded 5-oxoprolinase subunit PxpA (Table S6). The presence of these two genes suggests that the *Tethybacterales* Sp02-1 genome can acquire proline and convert it to glutamate(85) in addition to glutamate already produced via glutamate synthase. Other unique genes encode a restriction endonuclease subunit and site-specific DNA-methyltransferase, which would presumably aid in defense against foreign DNA. At least seven of the unique gene products are predicted to be associated with phages, including the anti-restriction protein ArdA. ArdA is a

protein that has previously been shown to mimic the structures of DNA normally bound by Type I restriction modification enzymes, which prevent DNA cleavage, and effectively results in anti-restriction activity(86). If functionally active in the *Tethybacterales* Sp02-1 symbiont, we speculate that this protein may similarly prevent DNA cleavage through its mimicry of the targeted DNA structures and protect the genome against Type I restriction modification enzymes. Finally, two of the unique genes were predicted to encode an ankyrin repeat domain-containing protein and a VWA domain-containing protein. These two proteins are known to be involved in cell-adhesion and protein-protein interactions(87, 88) and if active within the symbiont, they may help facilitate the symbiosis between the *Tethybacterales* Sp02-1 symbiont and the sponge host.

**Comparison of putative Sp02-1 with other Tethybacterales**

Several betaproteobacteria (*Tethybacterales*) sponge symbionts have been described to date and these bacteria are thought to have functionally diversified following the initiation of their ancient partnership(42). To test this hypothesis, we downloaded twelve genomes/MAGs of *Tethybacterales* (classified as AqS2 in GTDB) from the JGI database. Additionally, we assembled and binned metagenomic data from thirty-six sponge SRA datasets, covering fourteen sponge species and recovered an additional fourteen AqS2-like genomes. Ten of the total twenty-seven bins were of low quality, so Bin 003B_4 (Sp02-1) and sixteen medium quality *Tethybacterales* bins/genomes were used for further analysis (Table 1).

**Table 1. General characteristics of putative Tethybacterales genomes/MAGs**

| Genome | Sponge host | Size (Mbp) | Complete (%) | Contam (%) | Quality | "Core" genes | Study/ Accession |
|---|---|---|---|---|---|---|---|
| "*Candidatus* Ukwabelana africanus" (003B_4) | *Tsitsikamma favus* | 2.96 | 72.92 | 3.56 | Medium | 84.52 | This study |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *"Candidatus* Regalo mexicanus" (ImetM1_9) | *Iophon methanophila* | 1.56 | 85.58 | 0.61 | Medium | 83.33 | This study |
| *"Candidatus* Regalo mexicanus" (ImetM2_1_1) | *Iophon methanophila* | 1.6 | 84.36 | 0.61 | Medium | 86.9 | This study |
| *Persebacter sydneyensis* (C29) | *Crella incrustans* | 1.52 | 82.87 | 0.61 | Medium | 78.57 | Taylor et al., 2021 JGI: 2784132075 |
| *Tethybacter castellensis* (ccb2r) | *Cymbastela concentrica* | 1.69 | 81 | 0.3 | Medium | 85.71 | Taylor et al., 2021 JGI: 2784132054 |
| *Beroebacter blanensis* (Crambe1) | *Crambe crambe* | 2.25 | 80.38 | 1.81 | Medium | 73.8 | Taylor et al., 2021 JGI: 2814122900 |
| *Amphirhobacter heronislandensis* (AqS2) | *Amphimedon queenslandica* | 1.61 | 71.13 | 1.52 | Medium | 80.95 | Gauthier et al., 2016 NCBI: GCA_001750625.1 |
| *Calypsobacter congwongensis* (B3) | *Scopalina sp.* | 1.05 | 64.33 | 0.04 | Medium | 53.57 | Taylor et al., 2021 JGI: 2784132034 |
| *Telestobacter tawharauni* (TSB1) | *Tethya stolonifera* | 1.24 | 56.3 | 0 | Medium | 57.14 | Taylor et al., 2021 JGI: 2784132053 |
| *"Candidatus* Hadiah malacca" (Csing_1) | *Coelocarteria singaporensis* | 1.58 | 65.47 | 0.61 | Medium | 73.81 | JGI: 3300007741_3 |
| *"Candidatus* Hadiah malacca" (Csing_2) | *Coelocarteria singaporensis* | 1.36 | 58.12 | 0 | Medium | 72.62 | JGI: 3300007056_3 |
| *"Candidatus* Hadiah malacca" (Csing_3) | *Coelocarteria singaporensis* | 1.44 | 57.04 | 0.61 | Medium | 76.19 | JGI: 3300007046_3 |
| *"Candidatus* Hadiah malacca" (Csing_5) | *Coelocarteria singaporensis* | 1.15 | 55.82 | 0.61 | Medium | 75.00 | JGI: 3300021544_3 |
| *"Candidatus* Hadiah malacca" (Csing_6) | *Coelocarteria singaporensis* | 1.36 | 58.12 | 0 | Medium | 72.62 | JGI: 3300021545_3 |
| *"Candidatus* Hadiah malacca" (Csing_7) | *Coelocarteria singaporensis* | 0.98 | 54.76 | 1.93 | Medium | 69.05 | JGI: 3300021549_5 |
| *"Candidatus* Dora taiwanensis" (CCyA_2_3) | *Cinachyrella sp.* | 3.08 | 84.92 | 1.83 | Medium | 78.57 | This study |
| *"Candidatus* Dora taiwanensis" (CCyB_3_2) | *Cinachyrella sp.* | 3.45 | 79.58 | 5.93 | Medium | 75.00 | This study |
| CCyA_16_0 | *Cinachyrella sp.* | 0.86 | 16.95 | 3.81 | Low | 35.71 | This study |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| CCyA_3_0 | *Cinachyrella sp.* | 0.62 | 29.31 | 0 | Low | 15.48 | This study |
| CCyA_3_39 | *Cinachyrella sp.* | 0.41 | 22.01 | 4.88 | Low | 36.90 | This study |
| CCyB_501_11 | *Cinachyrella sp.* | 0.04 | 7.51 | 0 | Low | 17.86 | This study |
| CCyB_502_83 | *Cinachyrella sp.* | 0.43 | 17.01 | 0 | Low | 21.43 | This study |
| CCyC_2_7 | *Cinachyrella sp.* | 1.81 | 47.5 | 0 | Low | 57.14 | This study |
| 050C_6 | *Tsitsikamma favus* | 2.21 | 24.03 | 4.51 | Low | 32.14 | This study |
| Csing_4 | *Coelocarteria singaporensis* | 1.07 | 39.35 | 0.61 | Low | 54.76 | JGI: 3300007053_5 |
| 050A_14 | *Tsitsikamma favus* | 6.08 | 61 | 19.4 | Low | 73.81 | This study |
| 003D_6 | *Tsitsikamma favus* | 0.32 | 0 | 0 | Low | 0 | This study |

First, the phylogeny of the *Tethybacterales* symbionts was determined using single-copy marker genes in autoMLST, revealing a deep branching clade of these sponge-associated symbionts and that bin 003B_4 clustered within the proposed *Persebacteraceae* family (Fig. 1). All members of the *Persebacteraceae* family dominate the microbial community of their respective sponge hosts(44, 53, 54, 89). We additionally identified what appears to be a third family, consisting of symbionts associated with *C. singaporensis* and *Cinachyrella* sponge species (Fig. 1). Assessment of shared AAI (Average Amino acid Identity) indicates that these genomes represent a new family, sharing an average of 80% AAI within the family (Table S7)(90). These three families share less than 89% sequence similarity with respect to their 16S rRNA sequences, with intra-clade differences of less than 92% (Table S8). Therefore, they may represent novel classes within the *Tethybacterales* order (90). In keeping with naming the families after Oceanids of Greek mythology(42), we propose the family name *Polydorabacteraceae*, which means “many gifts”. Additionally, we propose species names for the newly identified genera as follows: Bin 003B_4 is

a single representative of "*Candidatus* Ukwabelana africanus", Bin Imet_M1_9 and Bin ImetM2_1_1 are both representatives of "*Candidatus* Regalo mexicanus", Bin CCyA_2_3 and CCyB_3_2 are both representatives of "*Candidatus* Dora taiwanensis", and all six bins from *C. singaporensis* are representative of "*Candidatus* Hadiah malacca". In each case, the genus name means "gift from" in the local language (where possible) from where the host sponge was collected, and the species name reflects the region/country from which the sponge host was collected.

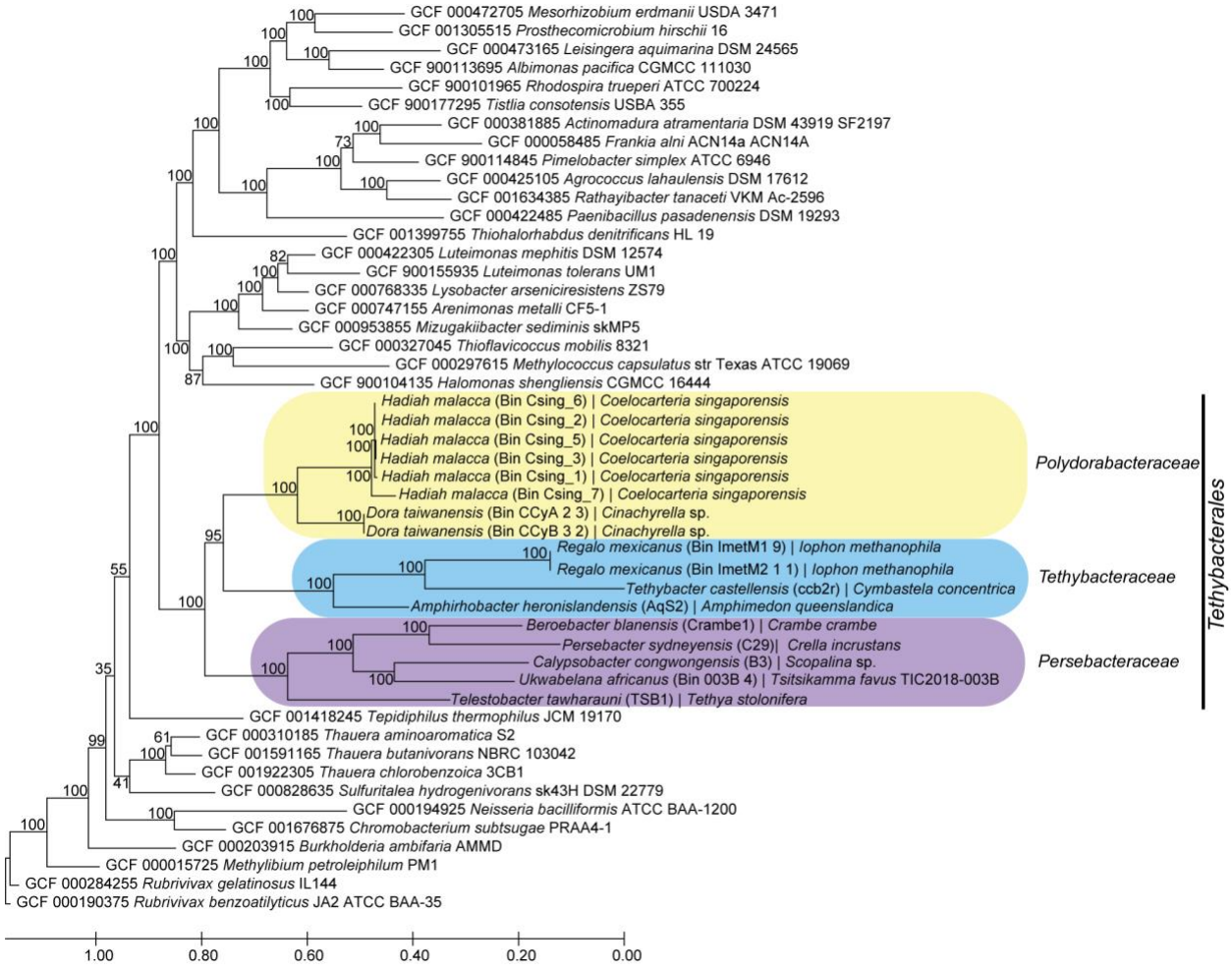


**Figure 1. Phylogeny of the *Tethybacterales* sponge symbionts.** Using autoMLST, single-copy markers were selected and used to delineate the phylogeny of these sponge-associated betaproteobacteria revealing a new family of symbionts in the *Tethybacterales* order. Additionally,

it was shown that the *T. favus* associated Sp02-1 symbiont belongs to the *Persebacteraceae* family. The phylogenetic tree was inferred using the de novo method in AutoMLST using a concatenated alignment with IQ Tree and ModelFinder enabled. Branch lengths are proportional to the number of substitutions per site.

We identified 4306 groups of orthologous genes between all seventeen *Tethybacterales* genomes, with only eighteen genes common to all the genomes. More shared genes were expected, but as several of the genomes investigated are incomplete, it is possible that additional common genes would be found if the genomes were complete. Hierarchical clustering of gene presence/absence data revealed that the gene pattern of Bin 003B_4 most closely resembled that of *Tethybacterales* genomes from *C. crambe*, *C. incrustans* and the *Scopalina* sp. sponges (family *Persebacteraceae*) (Fig. 2A). Thirteen of the shared genes between all *Tethybacterales* genomes encoded ribosomal proteins or those involved in energy production. Genes encoding chorismate synthase were found across all seventeen genomes and suggest that tryptophan production may be shared among these bacteria. According to a recent study, *Dysidea etheria* and *A. queenslandica* sponges cannot produce tryptophan (a possible essential amino acid), which may indicate a common role for the *Tethybacterales* symbionts as tryptophan producers(91). Several other shared genes were predicted to encode proteins involved in stress responses, including protein-methionine-sulfoxide reductase, ATP-dependent Clp protease and chaperonin enzyme proteins, which aid in protein folding or degradation under various stressors(92–96). Internal changes in oxygen levels(97), and temperature changes(98–100) are examples of stressors experienced by the sponge holobiont. It is unsurprising that this clade of largely sponge-specific *Tethybacterales* share the ability to deal with these many stressors as they adapt to their fluctuating environment.

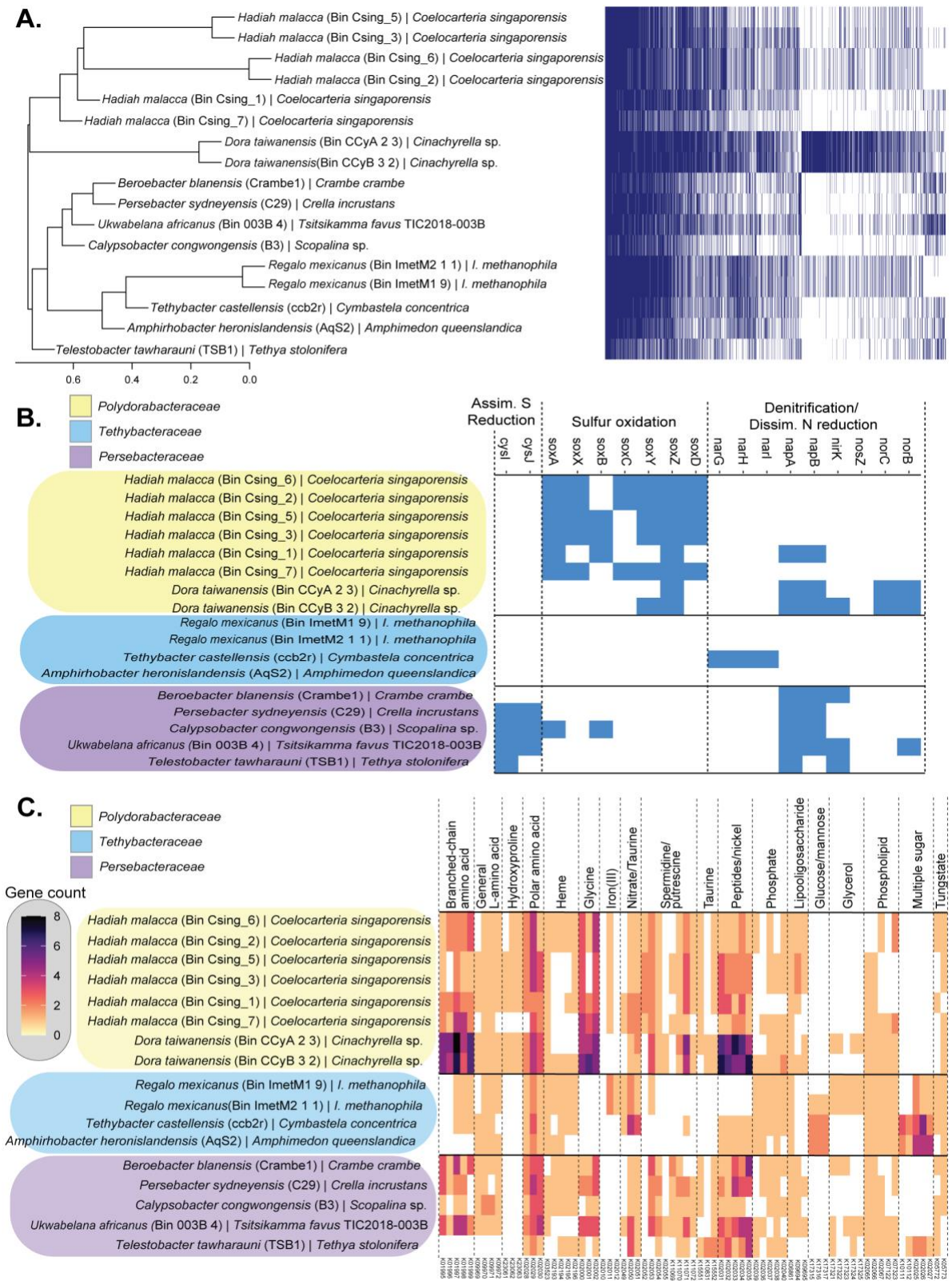


**Figure 2. Functional specialization of *Tethybacterales* families.** The newly proposed

*Tethybacterales* order appears to consist of three bacterial families. These families appear to have similar gene distribution (A) where the potential function of these genes indicates specialization in nutrient cycling (B) and solute transport (C).

Alignment against the KEGG database revealed some noteworthy trends that differentiated the three *Tethybacterales* families (Fig. 2B, Table S9): (1) the genomes of the proposed *Polydorabacteraceae* family include several genes associated with sulfur oxidation; (2) the *Persebacteraceae* are unique in their potential for reduction of sulfite (*cysIJ*) and (3) the *Tethybacteraceae* have the potential for cytoplasmic nitrate reduction (*narGHI*), while the other two families may perform denitrification. Similarly, the families differ to some extent in what can be transported in and out of the symbiont cell (Fig. 2C). Proposed members of the *Polydorabacteraceae* appear exclusively capable of transporting hydroxyproline which may imply a role in collagen degradation(101). The *Tethybacteraceae* and *Persebacteraceae* appear able to transport spermidine, putrescine, taurine, and glycine which, in combination with their potential to reduce nitrates, may suggest a role in C-N cycling(102). All three families transport various amino acids as well as phospholipids and heme. The exchange of amino acids between symbiont and sponge host has previously been observed(103) and may provide the *Tethybacterales* with a competitive advantage over other sympatric microorganisms(104) and possibly allow the sponge hosts to regulate the symbioses via regulation of the quantity of amino acids available for symbiont uptake(105). Similarly, the transfer of heme in the iron-starved ocean environment between sponge host and symbiont could provide a selective advantage as heme may act as a supply of iron(106). The *Tethybacteraceae* were distinct from the other two families in their potential to transport sugars. As mentioned earlier, the transport of sugars plays an important role in symbiotic

interactions(98, 107–109) and it is possible that this family of symbionts require sugars from their sponge hosts.

**Comparative analyses of functional potential between *Tethybacterales* and *Poribacteria***

We wanted to determine whether broad-host range sponge-associated symbionts have converged to perform similar roles in their sponge hosts. Accordingly, we annotated 62 *Poribacteria* genomes which consisted of 24 *Pelagiporibacteria* (free-living) and 38 *Entoporibacteria* (sponge-associated) genomes, and the 17 *Tethybacterales* genomes against the KEGG database. We catalogued the presence/absence of 896 unique genes spanning carbohydrate metabolism, methane metabolism, nitrogen metabolism, sulfur metabolism, phosphate metabolism and several transporter systems (Fig. 3, Table S9). Inspection of the functional potential in the *Tethybacterales* and *Poribacteria* revealed several insights (Fig. 3). The gene repertoires of the *Poribacteria* and the *Tethybacterales* are distinct from one another (Fig. S3, Table 2), with notable differences including the genes associated with denitrification, dissimilatory nitrate reduction, thiosulfate oxidation, hydroxyproline transport, glycine betaine/proline transport, glycerol transport, taurine transport, tungstate transport and glucose/mannose transport, all of which are present in the *Tethybacterales* and absent in the *Poribacteria* (Fig. 3). Conversely, several gene clusters were detected in the *Poribacteria* and absent in the *Tethybacterales*, including trehalose biosynthesis, the Entner-Doudoroff pathway, galactose degradation, phosphate metabolism, phosphonate transport, assimilatory sulfate reduction, molybdate transport, osmoprotectant transport, hydroxymethylpyrimidine transport (Fig. 3). It has been reported that both *Entoporibacteria* and *Pelagiporibacteria* include genes associated with denitrification(33), however we could not detect many genes associated with nitrogen metabolism in our analyses (Fig. 3).

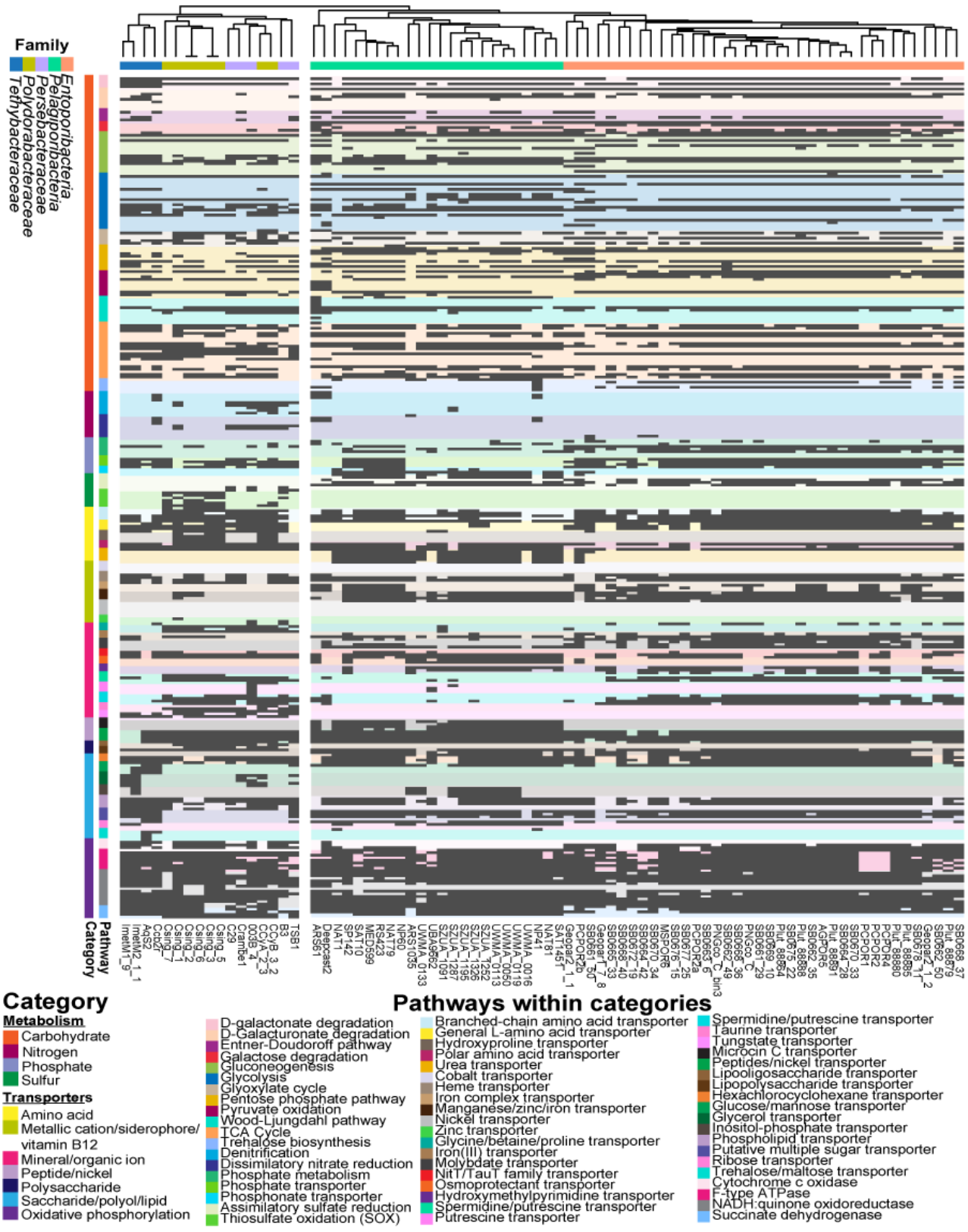


**Figure 3. Functional differences between *Tethybacterales* and *Poribacteria*.** Sponge-associated

*Tethybacterales* genomes include significantly different functional gene repertoires to those found in *Poribacteria*. Detailed presence/absence of metabolic genes (KEGG annotations) detected in *Tethybacterales* and *Poribacteria* genome bins. Genomes are listed at the bottom of the figure and clustered according to the presence/absence of functional genes (top). Taxonomic classification of each genome is indicated using a colored bar (top). Functional genes are collected into their respective pathways, which are further organized into larger functional categories. A colored key is provided (right) and is in the same order as that of the colored blocks in the graph (left).

We cross-checked gene annotations generated using Prokka (HAMAP database) and Blast (nr database). Genes associated with assimilatory nitrate reduction (*narB*, and *nirA*) were identified in *Poribacteria* using these alternate annotations, but we could not detect genes associated with denitrification in the *Poribacteria*. Conversely, genes associated with denitrification (*napAB* and *nirK*) were detected in the *Persebacteraceae* of the Tethybacterales in Prokka, Blast and KEGG annotations (Fig. 3), indicating that their absence in *Poribacteria* genomes was not an artefact of our analyses.

Pairwise ANOSIM analysis (using Bray-Curtis distance) confirmed that the functional genetic repertoire (KEGG annotations) of the *Tethybacterales* bacteria showed a strong, significant dissimilarity to that of the sponge-associated *Entoporibacteria* and the free-living *Pelagiporibacteria* (Table 2). In addition, the *Polydorabacteraceae* and the *Persebacteraceae* were significantly different from one another, but the lower R-statistic would suggest that the dissimilarity is not as strong as between other groups in this analysis, while the *Tethybacteraceae* appear to be more functionally distinct from the other two *Tethybacterales* families.

**Table 2: Pairwise ANOSIM of presence/absence of KEGG-annotated functional genes in Poribacteria and Tethybacterales**

| Taxon A | Taxon B | p-value | R statistic |
|---|---|---|---|
| *Entoporibacteria* | *Polydorabacteraceae* | 0.0001 | 0.9965 |
| *Entoporibacteria* | *Pelagiporibacteria* | 0.0001 | 0.8881 |
| *Entoporibacteria* | *Persebacteraceae* | 0.0001 | 0.9985 |
| *Entoporibacteria* | *Tethybacteraceae* | 0.0001 | 0.9938 |
| *Polydorabacteraceae* | *Pelagiporibacteria* | 0.0001 | 0.9951 |
| *Polydorabacteraceae* | *Persebacteraceae* | 0.0045 | 0.3974 |
| *Polydorabacteraceae* | *Tethybacteraceae* | 0.0025 | 0.9908 |
| *Pelagiporibacteria* | *Persebacteraceae* | 0.0001 | 0.9961 |
| *Pelagiporibacteria* | *Tethybacteraceae* | 0.001 | 0.9741 |
| *Persebacteraceae* | *Tethybacteraceae* | 0.0085 | 0.7500 |

Taken together, these data suggest that the three *Tethybacterales* families and the *Entoporibacteria* lineages may each fulfil distinct functional or ecological niches within a given sponge host. Interestingly, some sponges such as *A. aerophoba* and *C. singaporenesis*, can play host to both *Tethybacterales* and *Entoporibacteria* species (Table S2) which provides further evidence that these symbionts may serve different purposes within their sponge host.

We investigated the respective approximate divergence pattern of the *Tethybacterales* and the *Entoporibacteria* and whether their divergence followed that of their sponge hosts. The eighteen homologous genes shared between the *Tethybacterales* were used to estimate the rate of synonymous substitution, which provides an approximation for the pattern of divergence between the species(110). We found that the estimated divergence pattern of the Tethybacterales (Fig. 4A)

and the phylogeny of the host sponges (Fig. 4B) was incongruent. Phylogenetic trees inferred using single-copy marker genes (Fig. 1), and the comprehensive 16S rRNA tree published by Taylor and colleagues(42) confirm this lack of congruency between symbiont and host phylogeny. Other factors such as collection site or depth could not explain the observed trend. Similar incongruence of symbiont and host phylogeny was observed for the *Entoporibacteria* (34 homologous genes used to estimate synonymous substitution rates) (Fig. 4C-D), in agreement with previous phylogenetic studies (31, 33, 34). This would suggest that these sponges likely acquired a free-living *Tethybacterales* common ancestor at different time points throughout their evolution, and that the same is true for the *Entoporibacteria*. Evidence of coevolution of betaproteobacteria symbionts within sponges families (46, 52, 53, 111) implies that *Tethybacterales* symbionts were likely acquired horizontally at various time points and may have coevolved with their respective hosts subsequent to acquisition.

Finally, the estimated rates of synonymous substitution of homologous genes were used to estimate the relative times at which the *Tethybacterales* and *Entoporibacteria* taxa began diverging, respectively. Regardless of substitution rate used, it was found that the sponge-associated *Tethybacterales* genomes began diverging from one another before the *Entoporibacteria* began diverging from one another (Table S5). If one accepts that divergence between exclusively sponge-associated bacterial lineages began when the common ancestor first associated with a sponge host, then the earlier divergence of sponge-associated *Tethybacterales* (relative to the *Entoporibacteria*) suggests that the *Tethybacterales* may have associated with sponges before the Poribacteria common ancestor and represent a more ancient symbiont. However, this hypothesis may prove false if additional *Entoporibacteria* lineages are discovered and added to the analyses, or other

factors such as mutation rates, time between symbiont acquisition and transition to vertical inheritance of symbionts or fossil records disprove this hypothesis.

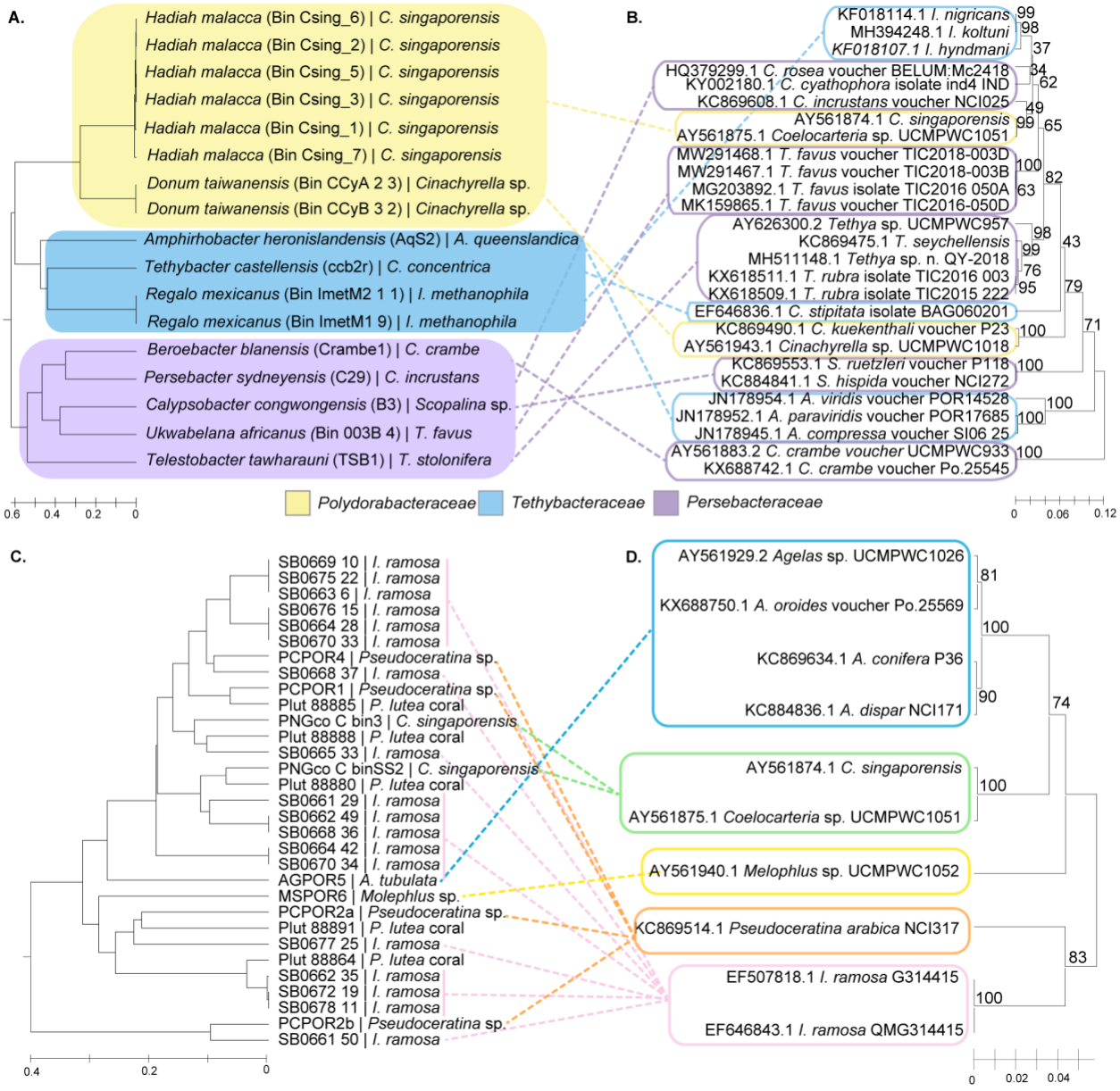


**Figure 4. The divergence pattern of sponge-associated *Tethybacterales*, *Entoporibacteria* and their respective host sponges.** The divergence of the *Tethybacterales* and *Entoporibacteria* is incongruent with the phylogeny of the host sponges. (A and C) Branch length of symbiont divergence estimates is proportional to the pairwise rate of synonymous substitution calculated

(ML estimation) using a concatenation of genes common to all genomes. Rate of synonymous substitution was calculated using PAL2NAL and CodeML from the PAML package and visualized in MEGAX. (B and D) Phylogeny of host sponges (or close relatives thereof) was inferred with 28S rRNA sequence data using the UPGMA method and Maximum Composite Likelihood model with 1000 bootstrap replicates. Branch lengths indicate the number of substitutions per site. All ambiguous positions were removed for each sequence pair (pairwise deletion option). Evolutionary analyses were conducted in MEGA X.

## CONCLUSION

Here, we have shown that the family to which a broad host-range symbiont belongs, such as the *Tethybacterales*, dictates the functional potential of the symbiont, whereas in narrow host-range symbionts, potential function is often a strain-specific trait. This work has expanded our understanding of the *Tethybacterales* and the possible functional specialization of the families within this new order. The *Tethybacterales* are functionally distinct from the *Poribacteria* which would suggest that although these bacteria are both ubiquitously associated with a wide range of sponge hosts, they likely have not converged to fulfil the same role. Instead, it would appear that these symbionts were selected by the various sponge hosts for existing functional capabilities. The incongruence of both *Tethybacterales* and *Entoporibacteria* suggests that their ancestors were horizontally acquired at different evolutionary timepoints, and co-evolution may have occurred following the establishment of the association. Estimates of when the *Tethybacterales* and *Entoporibacteria* began diverging from their respective common ancestors implied that *Tethybacterales* may have associated with a sponge host before the *Entoporibacteria* and therefore

the Tethybacterales may be an older sponge-associated symbiont. However, additional data is required to validate or disprove this hypothesis.

**FUNDING**

This research was funded by grants to R.A.D. from the South Africa Research Chair Initiative (SARChI) grant (UID: 87583), the NRF African Coelacanth Ecosystem Programme (ACEP) (UID: 97967) and the SARChI-led Communities of Practice Programme (GUN: 110612) from the South African National Research Foundation (NRF). S.C.W was supported by a Post-Doctoral fellowship from the Gordon and Betty Moore Foundation (Grant number 6920) (awarded to R.A.D and J.C.K.) and by an NRF Innovation and Rhodes University Henderson PhD Scholarships. S.P.-N. holds a NRF PDP (Grant number 101038). Opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to any of the above-mentioned donors.

**ACKNOWLEDGEMENTS**

This research was performed in part using the computer resources and assistance of the UW-Madison Center for High Throughput Computing (CHTC) in the Department of Computer Sciences. The CHTC is supported by UW-Madison, the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, Wisconsin Institutes for Discovery and the National Science Foundation and is an active member of the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science. The authors also acknowledge the Center for High-Performance Computing (CHPC, South Africa) for providing computing facilities for bioinformatics data analysis. The authors acknowledge Gwynneth Matcher, Mr. Carel van Heerden and Ms. Alvera Vorster for their NGS technical

support. The authors thank Ryan Palmer, Skipper Koos Smith, Nicholas Riddin and Nicholas Schmidt (ACEP) for technical support and expertise during sponge collections. We thank the South African Environmental Observation Network (SAEON), Elwandle Coastal Node, and the Shallow Marine and Coastal Research Infrastructure (SMCRI) for the use of their research platforms and infrastructure and South African National Parks (SANParks) for their assistance and support.

## REFERENCES

1. Wilkins LGE, Leray M, O'Dea A, Yuen B, Peixoto RS, Pereira TJ, Bik HM, Coil DA, Duffy JE, Herre EA, Lessios HA, Lucey NM, Mejia LC, Rasher DB, Sharp KH, Sogin EM, Thacker RW, Vega Thurber R, Wcislo WT, Wilbanks EG, Eisen JA. 2019. Host-associated microbiomes drive structure and function of marine ecosystems. PLoS Biol 17:e3000533.

2. Taylor MW, Radax R, Steger D, Wagner M. 2007. Sponge-associated microorganisms: Evolution, ecology, and biotechnological potential. Microbiol Mol Biol Rev 71:295–347.

3. Webster NS, Taylor MW. 2012. Marine sponges and their microbial symbionts: Love and other relationships. Environ Microbiol 14:335–346.

4. Zhang F, Jonas L, Lin H, Hill RT. 2019. Microbially mediated nutrient cycles in marine sponges. FEMS Microbiol Ecol 95.

5. Karimi E, Slaby BM, Soares AR, Blom J, Hentschel U, Costa R. 2018. Metagenomic binning reveals versatile nutrient cycling and distinct adaptive features in alphaproteobacterial symbionts of marine sponges. FEMS Microbiol Ecol 94.

6. Hoffmann F, Radax R, Woebken D, Holtappels M, Lavik G, Rapp HT, Schläppy M-L, Schleper C, Kuypers MMM. 2009. Complex nitrogen cycling in the sponge *Geodia barretti*. Environ Microbiol 11:2228–2243.

7. Bayer K, Schmitt S, Hentschel U. 2006. Microbial nitrification in Mediterranean sponges: Possible involvement of ammonia-oxidizing Betaproteobacteria 2003.

8. Moitinho-Silva L, Díez-Vives C, Batani G, Esteves AI, Jahn MT, Thomas T. 2017. Integrated metabolism in sponge-microbe symbiosis revealed by genome-centered metatranscriptomics. ISME J 11:1651–1666.

9. Bayer K, Schmitt S, Hentschel U. 2008. Physiology, phylogeny and *in situ* evidence for bacterial and archaeal nitrifiers in the marine sponge *Aplysina aerophoba*. Environ Microbiol 10:2942–2955.

10. Chaib De Mares M, Jiménez DJ, Palladino G, Gutleben J, Lebrun LA, Muller EEL, Wilmes P, Sipkema D, van Elsas JD. 2018. Expressed protein profile of a *Tectomicrobium* and other microbial symbionts in the marine sponge *Aplysina aerophoba* as evidenced by metaproteomics. Sci Rep 8:11795.

11. Jensen S, Fortunato SAV, Hoffmann F, Hoem S, Rapp HT, Øvreås L, Torsvik VL. 2017. The relative abundance and transcriptional activity of marine sponge-associated microorganisms emphasizing groups involved in sulfur cycle. Microb Ecol 73:668–676.

12. Gauthier M-EA, Watson JR, Degnan SM. 2016. Draft genomes shed light on the dual bacterial symbiosis that dominates the microbiome of the coral reef sponge *Amphimedon queenslandica*. Frontiers in Marine Science 3:196.

13. Zhang F, Blasiak LC, Karolin JO, Powell RJ, Geddes CD, Hill RT. 2015. Phosphorus sequestration in the form of polyphosphate by microbial symbionts in marine sponges. Proc Natl Acad Sci U S A 112:4381–4386.

14. Colman AS. 2015. Sponge symbionts and the marine P cycle. Proc Natl Acad Sci U S A.

15. Wilkinson CR. 1983. Net primary productivity in coral reef sponges. Science 219:410–412.

16. Feng G, Li Z. 2019. Carbon and Nitrogen Metabolism of Sponge Microbiome, p. 145–169. *In* Li, Z (ed.), Symbiotic Microbiomes of Coral Reefs Sponges and Corals. Springer Netherlands, Dordrecht.

17. Fiore CL, Labrie M, Jarett JK, Lesser MP. 2015. Transcriptional activity of the giant barrel sponge, *Xestospongia muta* Holobiont: Molecular evidence for metabolic interchange. Front Microbiol 6:364.

18. Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS, Thomas T. 2012. Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. Proc Natl Acad Sci U S A 109:E1878-87.

19. Bayer K, Jahn MT, Slaby BM, Moitinho-Silva L, Hentschel U. 2018. Marine sponges as Chloroflexi hot spots: Genomic insights and high-resolution visualization of an abundant and diverse symbiotic clade. mSystems 3:6:e00150-18.

20. Karimi E, Keller-Costa T, Slaby BM, Cox CJ, da Rocha UN, Hentschel U, Costa R. 2019. Genomic blueprints of sponge-prokaryote symbiosis are shared by low abundant and cultivatable Alphaproteobacteria. Sci Rep 9:1999.

21. Song H, Hewitt OH, Degnan SM. 2020. Bacterial symbionts in animal development: Arginine biosynthesis complementation enables larval settlement in a marine sponge.

22. Helber SB, Hoeijmakers DJJ, Muhando CA, Rohde S, Schupp PJ. 2018. Sponge chemical defenses are a possible mechanism for increasing sponge abundance on reefs in Zanzibar. PLoS One 13:e0197617.

23. Lopanik NB. 2014. Chemical defensive symbioses in the marine environment. Funct Ecol 28:328–340.

24. Mori T, Cahn JKB, Wilson MC, Meoded RA, Wiebach V, Martinez AFC, Helfrich EJN, Albersmeier A, Wibberg D, Dätwyler S, Keren R, Lavy A, Rückert C, Ilan M, Kalinowski J, Matsunaga S, Takeyama H, Piel J. 2018. Single-bacterial genomics validates rich and varied specialized metabolism of uncultivated *Entotheonella* sponge symbionts. Proc Natl Acad Sci U S A 115:1718–1723.

25. Newman DJ, Cragg GM. 2020. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. J Nat Prod 83:770–803.

26. Pita L, Hoeppner MP, Ribes M, Hentschel U. 2018. Differential expression of immune receptors in two marine sponges upon exposure to microbial-associated molecular patterns. Sci Rep 8:16081.

27. Degnan SM. 2015. The surprisingly complex immune gene repertoire of a simple sponge, exemplified by the NLR genes: A capacity for specificity? Dev Comp Immunol 48:269–274.

28. Jahn MT, Arkhipova K, Markert SM, Stigloher C, Lachnit T, Pita L, Kupczok A, Ribes M, Stengel ST, Rosenstiel P, Dutilh BE, Hentschel U. 2019. A phage protein aids bacterial

symbionts in eukaryote immune evasion. Cell Host Microbe 26:542-550.e5.

29. Thomas T, Moitinho-Silva L, Lurgi M, Björk JR, Easson C, Astudillo-García C, Olson JB, Erwin PM, López-Legentil S, Luter H, Chaves-Fonnegra A, Costa R, Schupp PJ, Steindler L, Erpenbeck D, Gilbert J, Knight R, Ackermann G, Victor Lopez J, Taylor MW, Thacker RW, Montoya JM, Hentschel U, Webster NS. 2016. Diversity, structure and convergent evolution of the global sponge microbiome. Nat Commun 7:11870.

30. Astudillo-García C, Slaby BM, Waite DW, Bayer K, Hentschel U, Taylor MW. 2018. Phylogeny and genomics of SAUL, an enigmatic bacterial lineage frequently associated with marine sponges. Environ Microbiol 20:561–576.

31. Fieseler L, Horn M, Wagner M, Hentschel U. 2004. Discovery of the novel candidate phylum "*Poribacteria*" in marine sponges. Appl Environ Microbiol 70:3724–3732.

32. Tully BJ, Graham ED, Heidelberg JF. 2018. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Sci Data 5:170203.

33. Podell S, Blanton JM, Neu A, Agarwal V, Biggs JS, Moore BS, Allen EE. 2019. Pangenomic comparison of globally distributed *Poribacteria* associated with sponge hosts and marine particles. ISME J 13:468–481.

34. Lafi FF, Fuerst JA, Fieseler L, Engels C, Goh WWL, Hentschel U. 2009. Widespread distribution of poribacteria in demospongiae. Appl Environ Microbiol 75:5695–5699.

35. Steinert G, Gutleben J, Atikana A, Wijffels RH, Smidt H, Sipkema D. 2018. Coexistence of poribacterial phylotypes among geographically widespread and phylogenetically divergent sponge hosts. Environ Microbiol Rep 10:80–91.

36. Siegl A, Kamke J, Hochmuth T, Piel J, Richter M, Liang C, Dandekar T, Hentschel U. 2011. Single-cell genomics reveals the lifestyle of *Poribacteria*, a candidate phylum symbiotically associated with marine sponges. ISME J 5:61–70.

37. 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:e87353.

38. 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 J 7:2287–2300.

39. Jahn MT, Markert SM, Ryu T, Ravasi T, Stigloher C, Hentschel U, Moitinho-Silva L. 2016. Shedding light on cell compartmentation in the candidate phylum *Poribacteria* by high resolution visualisation and transcriptional profiling. Sci Rep 6:35860.

40. Hill MS, Sacristán-Soriano O. 2017. Molecular and functional ecology of sponges and their microbial symbionts, p. 105–142. *In* Carballo, JL, Bell, JJ (eds.), Climate Change, Ocean Acidification and Sponges: Impacts Across Multiple Levels of Organization. Springer International Publishing, Cham.

41. Chen ML, Becraft ED, Pachiadaki M, Brown JM, Jarett JK, Gasol JM, Ravin NV, Moser DP, Nunoura T, Herndl GJ, Woyke T, Stepanauskas R. 2020. Hiding in plain sight: The globally distributed bacterial candidate Phylum PAUC34f. Front Microbiol 11:376.

42. Taylor JA, Palladino G, Wemheuer B, Steinert G, Sipkema D, Williams TJ, Thomas T. 2020.

Phylogeny resolved, metabolism revealed: Functional radiation within a widespread and divergent clade of sponge symbionts. ISME J https://doi.org/10.1038/s41396-020-00791-z.

43. Cleary DFR, Swierts T, Coelho FJRC, Polónia ARM, Huang YM, Ferreira MRS, Putchakarn S, Carvalheiro L, van der Ent E, Ueng J-P, Gomes NCM, de Voogd NJ. 2019. The sponge microbiome within the greater coral reef microbial metacommunity. Nat Commun 10:1644.

44. Croué J, West NJ, Escande M-L, Intertaglia L, Lebaron P, Suzuki MT. 2013. A single betaproteobacterium dominates the microbial community of the crambescidine-containing sponge *Crambe crambe*. Sci Rep 3:2583.

45. Thiel V, Neulinger SC, Staufenberger T, Schmaljohann R, Imhoff JF. 2007. Spatial distribution of sponge-associated bacteria in the Mediterranean sponge *Tethya aurantium*. FEMS Microbiol Ecol 59:47–63.

46. Waterworth SC, Jiwaji M, Kalinski J-CJ, Parker-Nance S, Dorrington RA. 2017. A Place to Call Home: An Analysis of the Bacterial Communities in Two *Tethya rubra* Samaai and Gibbons 2005 Populations in Algoa Bay, South Africa. Mar Drugs 15.

47. Cárdenas CA, Bell JJ, Davy SK, Hoggard M, Taylor MW. 2014. Influence of environmental variation on symbiotic bacterial communities of two temperate sponges. FEMS Microbiol Ecol 88:516–527.

48. Steinert G, Taylor MW, Deines P, Simister RL, de Voogd NJ, Hoggard M, Schupp PJ. 2016. In four shallow and mesophotic tropical reef sponges from Guam the microbial community largely depends on host identity. PeerJ 4:e1936.

49. Webster NS, Wilson KJ, Blackall LL, Hill RT. 2001. Phylogenetic diversity of bacteria

associated with the marine sponge *Rhopaloeides odorabile*. Appl Environ Microbiol 67:434–444.

50. Cleary DFR, Becking LE, de Voogd NJ, Pires ACC, Polónia ARM, Egas C, Gomes NCM. 2013. Habitat- and host-related variation in sponge bacterial symbiont communities in Indonesian waters. FEMS Microbiol Ecol 85:465–482.

51. Trindade-Silva AE, Rua C, Silva GGZ, Dutilh BE, Moreira APB, Edwards RA, Hajdu E, Lobo-Hajdu G, Vasconcelos AT, Berlinck RGS, Thompson FL. 2012. Taxonomic and functional microbial signatures of the endemic marine sponge *Arenosclera brasiliensis*. PLoS One 7:e39905.

52. Fieth RA, Gauthier M-EA, Bayes J, Green KM, Degnan SM. 2016. Ontogenetic changes in the bacterial symbiont community of the tropical Demosponge *Amphimedon queenslandica*: Metamorphosis is a new beginning. Frontiers in Marine Science 3:228.

53. Matcher GF, Waterworth SC, Walmsley TA, Matsatsa T, Parker-Nance S, Davies-Coleman MT, Dorrington RA. 2017. Keeping it in the family: Coevolution of latrunculid sponges and their dominant bacterial symbionts. Microbiologyopen 6:00417

54. Walmsley TA, Matcher GF, Zhang F, Hill RT, Davies-Coleman MT, Dorrington RA. 2012. Diversity of bacterial communities associated with the Indian Ocean sponge *Tsitsikamma favus* that contains the bioactive pyrroloiminoquinones, tsitsikammamine A and B. Mar Biotechnol 14:681–691.

55. Antunes EM, Copp BR, Davies-Coleman MT, Samaai T. 2005. Pyrroloiminoquinone and related metabolites from marine sponges. Nat Prod Rep 22:62–72.

56. Kalinski J-CJ, Krause RWM, Parker-Nance S, Waterworth SC, Dorrington RA. 2021. Unlocking the Diversity of Pyrroloiminoquinones Produced by Latrunculid Sponge Species. Mar Drugs 19:68.

57. Samaai T, Kelly M. 2002. Family Latrunculiidae Topsent, 1922, p. 708–719. *In* Hooper, JNA, Van Soest, RWM, Willenz, P (eds.), Systema Porifera: A Guide to the Classification of Sponges. Springer US, Boston, MA.

58. Parker-Nance S, Hilliar S, Waterworth S, Walmsley T, Dorrington R. 2019. New species in the sponge genus *Tsitsikamma* (*Poecilosclerida*, *Latrunculiidae*) from South Africa. Zookeys 874:101–126.

59. Hamlyn-Harris R, Queensland Museum, Hamlyn-Harris R. 1996. Memoirs of the Queensland Museum 40.

60. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477.

61. Miller IJ, Rees ER, Ross J, Miller I, Baxa J, Lopera J, Kerby RL, Rey FE, Kwan JC. 2019. Autometa: Automated extraction of microbial genomes from individual shotgun metagenomes. Nucleic Acids Res 47:e57.

62. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055.

63. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Genome Standards Consortium, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Schriml L, Banfield JF, Hugenholtz P, Woyke T. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35:725–731.

64. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

65. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics https://doi.org/10.1093/bioinformatics/btz848.

66. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. 2008. NCBI BLAST: A better web interface. Nucleic Acids Res 36:W5-9.

67. Alanjary M, Steinke K, Ziemert N. 2019. AutoMLST: An automated web server for generating multi-locus species trees highlighting natural product potential. Nucleic Acids Res 47:W276–W282.

68. Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30:2068–

2069.

69. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. 2019. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. bioRxiv https://doi.org/10.1101/602110.

70. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87.

71. Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: A toolbox for specialized analyses of microbial genomes and metagenomes. e1900v1. PeerJ Preprints.

72. Dixon P. 2003. VEGAN, a package of R functions for community ecology. J Veg Sci 14:927–930.

73. Waterworth SC, Flórez LV, Rees ER, Hertweck C, Kaltenpoth M, Kwan JC. 2020. Horizontal gene transfer to a defensive symbiont with a reduced genome in a multipartite beetle microbiome. MBio 11.

74. Altenhoff AM, Levy J, Zarowiecki M, Tomiczek B, Warwick Vesztrocy A, Dalquen DA, Müller S, Telford MJ, Glover NM, Dylus D, Dessimoz C. 2019. OMA standalone: Orthology inference among public and custom genomes and transcriptomes. Genome Res 29:1152–1163.

75. Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797.

76. Rice P, Longden I, Bleasby A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet 16:276–277.

77. Suyama M, Torrents D, Bork P. 2006. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34:W609-12.

78. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35:1547–1549.

79. Sneath PHA, Sokal RR. 1973. Numerical taxonomy. The principles and practice of numerical classification.

80. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60.

81. McCutcheon JP, Moran NA. 2011. Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol 10:13–26.

82. Manzano-Marín A, Latorre A. 2016. Snapshots of a shrinking partner: Genome reduction in *Serratia symbiotica*. Sci Rep 6:32590.

83. Feng H, Edwards N, Anderson CMH, Althaus M, Duncan RP, Hsu Y-C, Luetje CW, Price DRG, Wilson ACC, Thwaites DT. 2019. Trading amino acids at the aphid-*Buchnera* symbiotic interface. Proc Natl Acad Sci U S A 116:16003–16011.

84. Knobloch S, Jóhannsson R, Marteinsson VÞ. 2020. Genome analysis of sponge symbiont "*Candidatus* Halichondribacter symbioticus" shows genomic adaptation to a host-dependent lifestyle. Environ Microbiol 22:483–498.

85. Niehaus TD, Elbadawi-Sidhu M, de Crécy-Lagard V, Fiehn O, Hanson AD. 2017. Discovery of a widespread prokaryotic 5-oxoprolinase that was hiding in plain sight. J Biol Chem 292:16360–16367.

86. Chen K, Reuter M, Sanghvi B, Roberts GA, Cooper LP, Tilling M, Blakely GW, Dryden DTF. 2014. ArdA proteins from different mobile genetic elements can bind to the EcoKI Type I DNA methyltransferase of E. coli K12. Biochim Biophys Acta 1844:505–511.

87. Boyd CD, Smith TJ, El-Kirat-Chatel S, Newell PD, Dufrêne YF, O'Toole GA. 2014. Structural features of the *Pseudomonas fluorescens* biofilm adhesin LapA required for LapG-dependent cleavage, biofilm formation, and cell surface localization. J Bacteriol 196:2775–2788.

88. Al-Khodor S, Price CT, Kalia A, Abu Kwaik Y. 2010. Functional diversity of ankyrin repeats in microbial proteins. Trends Microbiol 18:132–139.

89. Said Hassane C, Fouillaud M, Le Goff G, Sklirou AD, Boyer JB, Trougakos IP, Jerabek M, Bignon J, de Voogd NJ, Ouazzani J, Gauvin-Bialecki A, Dufossé L. 2020. Microorganisms associated with the marine sponge *Scopalina hapalia*: A reservoir of bioactive molecules to slow down the aging process. Microorganisms 8:1262

90. Konstantinidis KT, Rosselló-Móra R, Amann R. 2017. Uncultivated microbes in need of their own taxonomy. ISME J 11:2399–2406.

91. Munroe S, Sandoval K, Martens DE, Sipkema D, Pomponi SA. 2019. Genetic algorithm as an optimization tool for the development of sponge cell culture media. In Vitro Cell Dev Biol Anim 55:149–158.

92. Voth W, Jakob U. 2017. Stress-activated chaperones: A first line of defense. Trends Biochem Sci 42:899–913.

93. Gottesman S, Wickner S, Maurizi MR. 1997. Protein quality control: Triage by chaperones and proteases. Genes Dev 11:815–823.

94. Cohn MT, Ingmer H, Mulholland F, Jørgensen K, Wells JM, Brøndsted L. 2007. Contribution of conserved ATP-dependent proteases of *Campylobacter jejuni* to stress tolerance and virulence. Appl Environ Microbiol 73:7803–7813.

95. Thomsen LE, Olsen JE, Foster JW, Ingmer H. 2002. ClpP is involved in the stress response and degradation of misfolded proteins in *Salmonella enterica* serovar *Typhimurium*. Microbiology 148:2727–2733.

96. Gennaris A, Ezraty B, Henry C, Agrebi R, Vergnes A, Oheix E, Bos J, Leverrier P, Espinosa L, Szewczyk J, Vertommen D, Iranzo O, Collet J-F, Barras F. 2015. Repairing oxidized proteins in the bacterial envelope using respiratory chain electrons. Nature 528:409–412.

97. Lavy A, Keren R, Yahel G, Ilan M. 2016. Intermittent hypoxia and prolonged suboxia measured *in situ* in a marine sponge. Frontiers in Marine Science 3:263.

98. Fan L, Liu M, Simister R, Webster NS, Thomas T. 2013. Marine microbial symbiosis heats up: The phylogenetic and functional response of a sponge holobiont to thermal stress. ISME J 7:991–1002.

99. Simister R, Taylor MW, Tsai P, Fan L, Bruxner TJ, Crowe ML, Webster N. 2012. Thermal stress responses in the bacterial biosphere of the Great Barrier Reef sponge, *Rhopaloeides odorabile*. Environ Microbiol 14:3232–3246.

100. Guzman C, Conaco C. 2016. Gene expression dynamics accompanying the sponge thermal stress response. PLoS One 11:e0165368.

101. Tziveleka L-A, Ioannou E, Tsiourvas D, Berillis P, Foufa E, Roussis V. 2017. Collagen from the marine sponges *Axinella cannabina* and *Suberites carnosus*: Isolation and morphological, biochemical, and biophysical characterization. Mar Drugs 15.

102. Hanson BT, Hewson I, Madsen EL. 2014. Metaproteomic survey of six aquatic habitats: Discovering the identities of microbial populations active in biogeochemical cycling. Microb Ecol 67:520–539.

103. Shih JL, Selph KE, Wall CB, Wallsgrove NJ, Lesser MP, Popp BN. 2020. Trophic ecology of the tropical Pacific sponge *Mycale grandis* inferred from amino acid compound-specific isotopic analyses. Microb Ecol 79:495–510.

104. Hosie AHF, Allaway D, Galloway CS, Dunsby HA, Poole PS. 2002. *Rhizobium leguminosarum* has a second general amino acid permease with unusually broad substrate specificity and high similarity to branched-chain amino acid transporters (Bra/LIV) of the ABC family. J Bacteriol 184:4071–4080.

105. Prell J, White JP, Bourdes A, Bunnewell S, Bongaerts RJ, Poole PS. 2009. Legumes regulate *Rhizobium* bacteroid development and persistence by the supply of branched-chain amino acids. Proc Natl Acad Sci U S A 106:12477–12482.

106. Hogle SL, Barbeau KA, Gledhill M. 2014. Heme in the marine environment: From cells to the iron cycle. Metallomics 6:1107–1120.

107. Ekman M, Picossi S, Campbell EL, Meeks JC, Flores E. 2013. A *Nostoc punctiforme* sugar

transporter necessary to establish a Cyanobacterium-plant symbiosis. Plant Physiol 161:1984–1992.

108. Neave MJ, Michell CT, Apprill A, Voolstra CR. 2017. *Endozoicomonas* genomes reveal functional adaptation and plasticity in bacterial strains symbiotically associated with diverse marine hosts. Sci Rep 7:40579.

109. Rix L, Ribes M, Coma R, Jahn MT, de Goeij JM, 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 J https://doi.org/10.1038/s41396-020-0706-3.

110. Silva FJ, Santos-Garcia D. 2015. Slow and fast evolving endosymbiont lineages: Positive correlation between the rates of synonymous and non-synonymous substitution. Front Microbiol 6:1279.

111. Wu S, Ou H, Liu T, Wang D, Zhao J. 2018. Structure and dynamics of microbiomes associated with the marine sponge *Tedania* sp. during its life cycle. FEMS Microbiol Ecol 94.

## SUPPLEMENTARY FIGURES

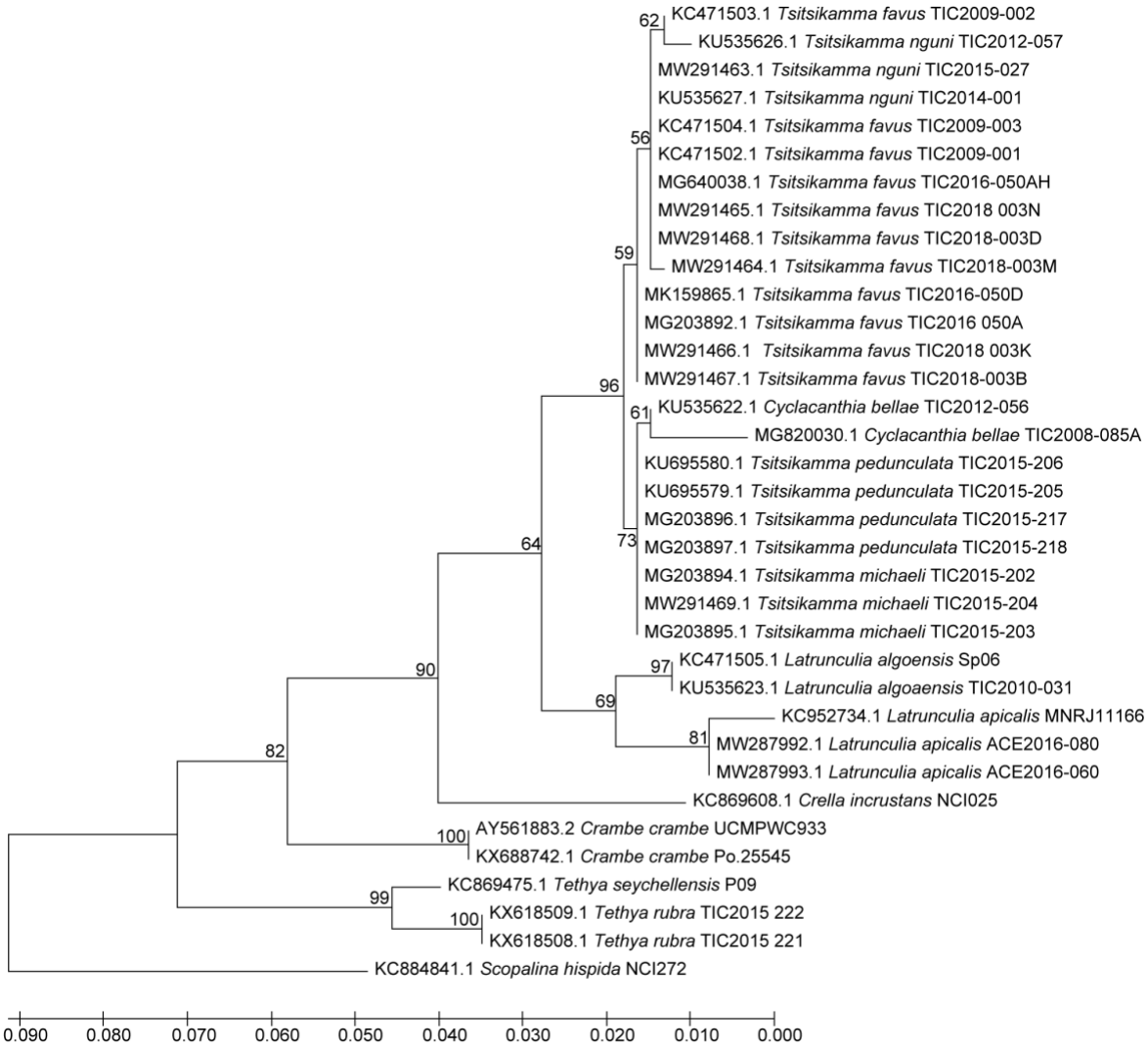


**Figure S1. Phylogeny of sponges within the family *Latrunculiidae*.** Sponge phylogeny was inferred using 28S rRNA sequence data using the Maximum-likelihood method and Tamura-Nei model with 1000 bootstrap replicates. Branch lengths indicate the number of substitutions per site. All ambiguous positions were removed for each sequence pair (pairwise deletion option). Evolutionary analyses were conducted in MEGA X.

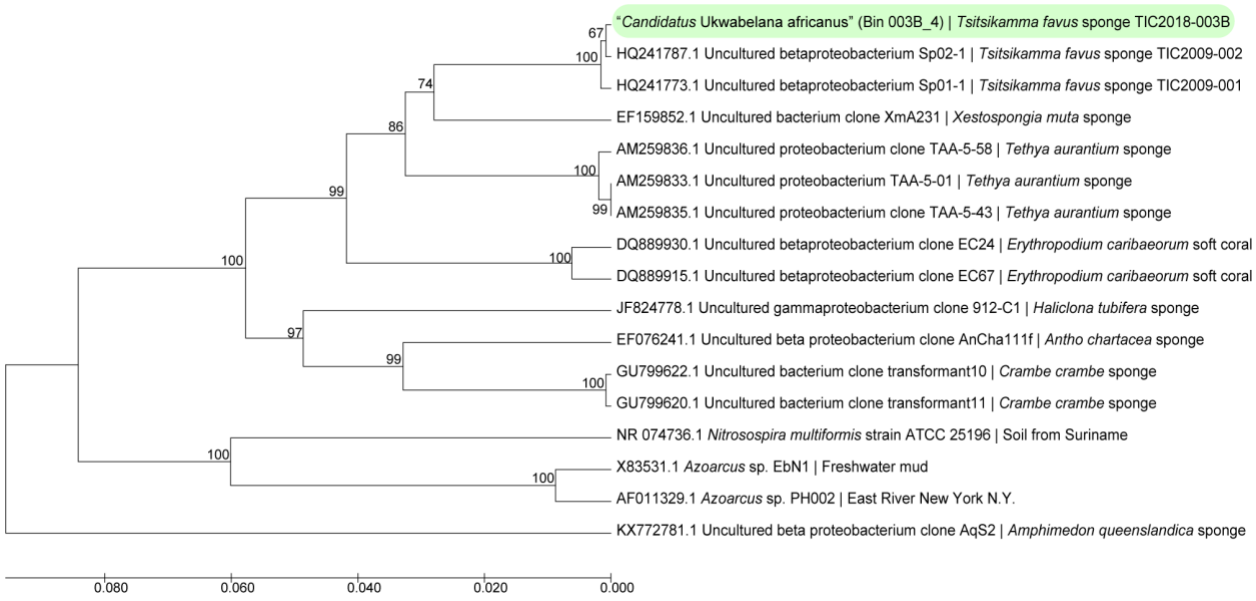


**Figure S2. Phylogeny of *T. favus*-associated betaproteobacterium bin.** The phylogenetic relationship between the putative betaproteobacteria Sp02-1 genome and closest relatives was based on 16S rRNA sequences and inferred using the UPGMA method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. This analysis involved 17 16S rRNA gene sequences with a total of 1291 positions analyzed. All ambiguous positions were removed for each sequence pair (pairwise deletion option). Evolutionary analyses were conducted in MEGA X.

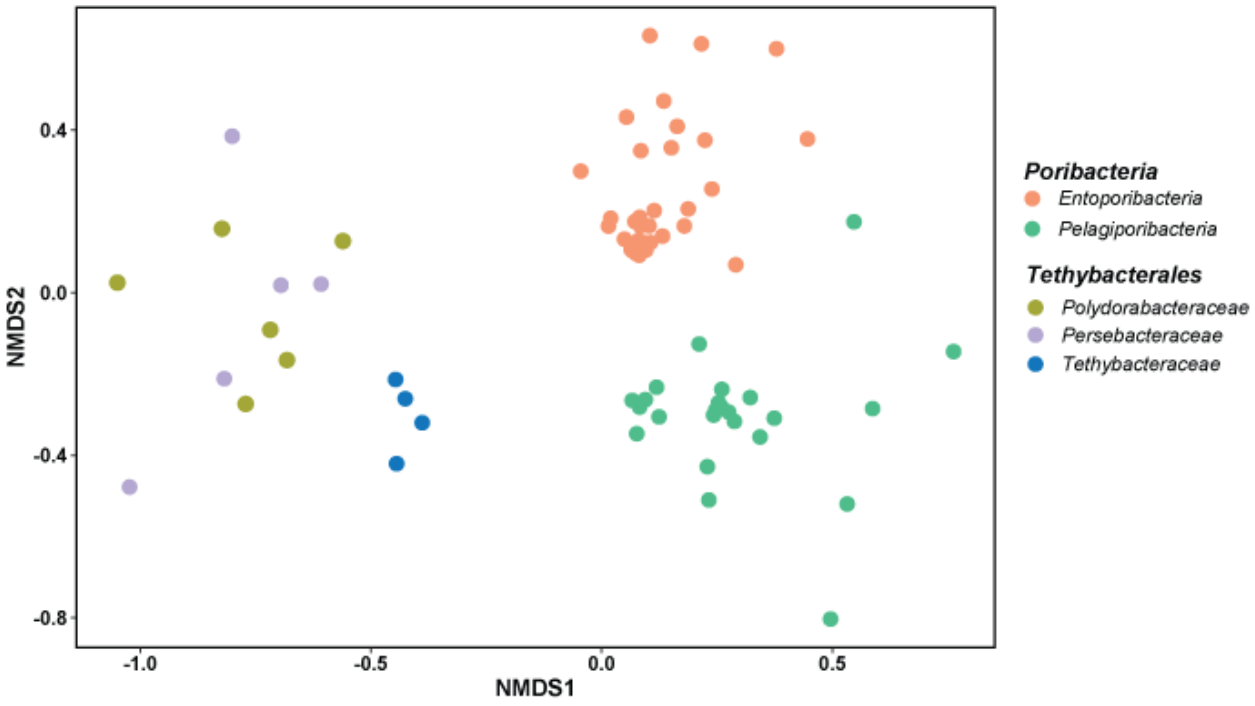


**Figure S3. The potential functional ability of the three Tethybacterales families and the two lineages within the Poribacteria appear to be distinct.** A Non-Metric Multi-Dimensional Scaling (NMDS) plot of the presence/absence metabolic counts from the Tethybacterales and Poribacteria calculated using Bray-Curtis distance.