Family matters: The genomes of conserved bacterial symbionts provide insight into specialized metabolic relationships with their sponge host

Samantha C. Waterworth[a,b], Jarmo-Charles J. Kalinski [b], Luthando S. Madonsela[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

Keywords: Latrunculiidae, Spirochete, Tethybacterales, pyrroloiminoquinone

Running title: Genomics of *Tsitsikamma favus* sponge microbial symbionts

Competing Interests

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

Abstract (200 words)

Sponges have evolved complex associations with conserved bacteria symbionts that contribute to the fitness of the host. Sponges of the family, Latrunculiidae, produce a diverse array of cytotoxic pyrroloiminoquinone alkaloids and their microbiomes are dominated by a conserved spirochete and betaproteobacterium. We investigated the microbiome of *Tsitsikamma favus*, a latrunculid species that produces predominantly tsitsikammamines. A total of 50 Metagenome-Assembled-Genomes (MAGs) were recovered from the four metagenomes, including representatives of the conserved spirochete and betaproteobacterium symbionts. We predict that the Sp02-3, one of two spirochetes unique to latrunculid species, is likely an anaerobic symbiont that produces vitamin B6. Dysbiosis of the spirochete population correlates with the chemistry associated with distinct sponge chemotypes, implying a role in pyrroloiminoquinone production. The betaproteobacteria symbiont belongs to the newly defined Tethybacterales order and likely aids in the removal of nitrate. There is also evidence to suggest that the two symbionts may work in tandem to remove sulfate from the sponge. Additionally, we analyzed seventeen Tethybacterales genomes and identified a third family within the order. The phylogeny of Tethybacterales is incongruent with that of the sponge hosts suggesting that the association of Tethybacterales and their sponge hosts likely occurred multiple times over their evolutionary history.

Introduction

Marine sponges are often host to a diverse array of microorganisms and have formed specialized associations with bacteria that aid in nutrient cycling or defense through the production of bioactive compounds[1]. One of the most dominant and widespread sponge symbionts are the bacteria within the phylum Poribacteria[2]. Initially identified in 2004[3], single-cell genomics and comparative genomics revealed that Poribacteria carry several genes associated with a

symbiont lifestyle[4, 5] and likely degrade carbohydrates and 1,2-propanediol[6, 7]. More recently, free-living relatives were identified, and comparative genomics showed distinct differences in the genomic repertoire between the free-living (*Pelagiporibacteria*) and the sponge associated (*Entoporibacteria*) Poribacteria linages[8].

Similarly, betaproteobacteria are often conserved in sponge microbiomes[9] and dominant betaproteobacteria have previously been observed in several marine sponge species[10–18], with some predicted to be endosymbiotic[10, 15, 18]. In *Amphimedeon queenslandica*, the betaproteobacterial symbiont is present in all stages of the sponge life cycle[19] and genomic insight suggested a role in carbohydrate metabolism[18]. Recent findings have indicated that these betaproteobacteria belong to a group of mostly sponge-specific bacteria within the proposed Tethybacterales order[20].

Sponges of the family, Latrunculiidae (Demospongiae, Poecilosclerida), are known for their cytotoxic pyrroloiminoquinone alkaloids, including makaluvamines, discorhabdins and tsitsikammamines[21, 22]. The microbiomes of latrunculid sponges are highly conserved and dominated by betaproteobacteria and spirochetes taxa[23]. In *Tsitsikamma favus*, sponge microbial communities are dominated by two sponge-specific bacterial species defined by their 16S rRNA gene sequence, namely clones Sp02-1 and Sp02-3, classified within the Betaproteobacteria and Spirochaete subphyla, respectively[23–25]. Additionally, the betaproteobacterial symbiont Sp02-1 was included within the newly proposed Tethybacterales order based on its 16S rRNA gene sequence[20]. The phylogenetic relationship between dominant betaproteobacterium Operational Taxonomic Units (OTUs) closely follows their latrunculid hosts[23]. This study aimed to characterize the genomes of these sponge-associated bacteria to shed light on factors that drive

their conservation. Metagenomic sequence analysis of the microbiomes of four *Tsitsikamma favus* sponge specimens enabled the characterization of the dominant bacterial species and we present data that implicates the involvement of spirochetes in pyrroloiminoquinone production.

METHODS AND MATERIALS

Sponge Collection. Sponge specimens were collected by SCUBA or Remotely Operated Vehicle (ROV) from multiple locations within Algoa Bay, the South African National Parks Garden Route National Park, Tsitsikamma Marine Protected Area, and off Bouvet Island in the South Atlantic Ocean. 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 listed in Table S1. Sponges were stored on ice during collection and subsamples were preserved in RNALater (Invitrogen) and stored at -20 °C.

DNA extraction. Small sections of each sponge (approx. 2cm^3) were pulverized in 2ml sterile artificial seawater (ASW) 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).

Sponge identification. Sponges were dissected, thin sections and spicules mounted on microscope slides and examined to allow species identification, see species description in Parker-Nance et al. 2019[26–28]. Molecular barcoding (28S rRNA gene) was also performed for several of the sponge specimens (Fig. S1) as described previously[23].

16S rRNA gene amplicon sequencing and analysis. 16S rRNA gene amplicons were prepared and analyzed as described in Kalinski et al., 2019[25]. ANOSIM analysis of the sponge-associated

microbial communities was performed using the commercial program Primer7. Indicator species analysis of microbial communities associated with sponge chemotypes was performed using the "indicspecies" package in R.

Metagenomic sequencing and analysis. Shotgun metagenomic sequencing was performed for four *T. favus* sponge specimens using Ion Torrent platforms. The TIC2016-050A sample was analyzed on an Ion S5 530 Chip and the other three samples (TIC2018-003B, TIC2016-050C and TIC2018-003D) were analyzed using an Ion P1.1.17 chip. Adapters were trimmed and trailing bases (15 nts) were iteratively removed if the average quality score for the last 30 nts was lower than 16 using Trimmomatic v. 0.39. Resultant metagenomic datasets were assembled into contiguous sequences (contigs) with SPAdes version 3.12.0[29] using the --iontorrent and --only-assembler options. Contigs were clustered into genomic bins using Autometa[30] and manually curated for optimal completion and purity. Validation of the bins was performed using CheckM v1.0.12.[31]. Of the 50 recovered genome bins, 5 were high quality, 13 were of medium quality and 32 were of low quality in accordance with MIMAG standards[32] (Table S2).

Acquisition and assembly of spirochete symbiont genomes from other host systems. Genomes of *Sphaerochaeta coccoides* DSM 17374 (GCA_000208385.1), *Salinispira pacifica* (GCA_000507245.1), *Spirochaeta africana* (CP003282.1), *Spirochaeta perfilievii* strain P (CP035807.1) and *Spirochaeta thermophila* DSM 6192 (CP001698.1) were downloaded from the NCBI database. Additional MAGs were retrieved from the JGI database (3300004122_5, 3300005978_6, 3300006913_10, 3300010314_9 and 3300027532_25).

Acquisition and assembly of betaproteobacterium symbiont genomes from other marine sponges. 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 (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[33] and assembled using SPAdes v3.14[29] in --meta mode and resultant contigs were binned using Autometa[30]. This resulted in a total of 393 additional genome bins (Table S2). The quality of all bins was assessed using CheckM[31] and bins were taxonomically classified with GTDB-Tk[34] (Table S2). A total of 27 bins were classified as "AqS2" and were considered likely members of the newly proposed Tethybacterales order[20]. However, ten low quality bins were not used in downstream analyses.

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[35]. Genomes were additionally uploaded individually to autoMLST[36] 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[34] with database release95. ANI between the four spirochete bins was calculated using FastANI (https://github.com/ParBLiSS/FastANI) with default parameters.

Genome annotation and metabolic potential analysis. All bins and downloaded genomes were annotated using Prokka 1.13[37] with NCBI compliance enabled. Protein-coding amino-acid sequences from genomic bins were annotated against the KEGG database using kofamscan[38] with output in mapper format. Potential Biosynthetic Gene Clusters (BGCs) were identified by

uploading Genome bins 003D_7 and 003B_4 to the antiSMASH web server[39] with all options enabled. Predicted amino acid sequences of genes within each identified gene cluster were aligned against the nr database using BLASTn[35] to identify the closest homologs. Protein sequences of genes within each identified gene cluster were aligned against the nr database using BLASTn[35] to identify the closest homolog.

Identification of host-associated and unique genes in putative symbiont genome bins. To identify "host-associated" genes, orthologous groups of putative Sp02-3 spirochete bins and reference genomes were found using OMA (v.2.4.1)[40]. Genes were considered "host-associated" if the gene was present in at least two of the three Sp02-3 genome bins and at least five of the seven host-associated genomes and absent in the free-living spirochetes.

A custom database of genes from all non-Sp02-3 *T. favus*-associated bacterial bins was created using the "makedb" option in Diamond[41] to identify genes that were unique to the Sp02-3 spirochete symbionts. In order to be exhaustive, genes from low-quality genomes (except low-quality spirochete bin 050A_2), small contigs (<3000 bp) and unclustered contigs were included in this database. Sp02-3 genes were aligned using diamond blast[41]. 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" Sp02-3 genes annotated as "hypothetical" (both Prokka and NCBI nr database annotations) were removed. Finally, we compared Prokka annotation strings between the three Sp02-3 genomes and all other *T. favus* associated genome bins and excluded any Sp02-3 genes that were found to have the same annotation as a gene in one of the other bins. Unique Sp02-1 genes were identified as for the Sp02-3 genes.

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[42]. 16S rRNA genes were analyzed using BLASTn[35]. The rate of synonymous substitutions was calculated as performed previously[43]. We illustrated the matrix of pairwise divergence values as a tree using MEGAX[44]. Concatenated amino acid and nucleotide sequences of the 18 orthologous genes were aligned using MUSCLE v.3.8.155[45] and the evolutionary history inferred using the UPMGA method[46] in MEGAX[44] with 10000 bootstrap replicates.

Chemical profiling of *T. favus* and *T. michaeli* sponges. Approximately 5 cm^3 pieces of *T. favus* sponge material of sponge specimens were removed from freshly collected sponge material and extracted in 10 mL methanol (MeOH) on the day of collection. Samples were filtered (0.2 μm) and analyzed at 2 μL injection volumes. The samples of the *T. michaeli* specimens collected in 2015 were prepared by extracting approx. 2 cm^3 wedges, cut from frozen sponge material, with 12 mL MeOH, followed by filtration (0.2 μm). These samples were analyzed with injection volumes of 5 μL. The *T. michaeli* specimens collected in 2017 were extracted with DCM-MeOH (2/1 v/v) and dried *in vacuo.* Samples were prepared at 10 mg/mL in MeOH, followed by filtration (0.2 μm) and analyzed using injection volumes of 1 μL. Correspondingly, other sponge specimens indicated in the Principal Component Analysis (PcoA) were processed. The High-Resolution Tandem Mass Spectrometry (LC-MS-MS) analyses were carried out on a Bruker Compact QToF mass spectrometer (Bruker, Rheinstetten, Germany) coupled to a Dionex UHPLC (Thermo Fisher Scientific, Sunnyvale, CA, USA) using an electrospray ionization probe in positive mode as described previously [7]. The chromatograph was equipped with a Kinetex polar C-18 column (3.0

x 100 mm, 2,6 µm; Phenomenex, Torrance, CA, USA) and operated at a flow rate of 0.300 mL/min using a mixture of water and acetonitrile (LC-MS) grade solvents (Merck Millipore, Johannesburg, South Africa), both adjusted with 0.1 % formic acid (Sigma-Aldrich, Johannesburg, South Africa)[25]. The Principal Component Analysis was performed as part of a molecular networking job on the GNPS platform[47]. Visualization of the base peak chromatograms was done using MZMine2 (v. 2.5.3)[48].

RESULTS AND DISCUSSION

Sponge-associated bacterial communities and chemical profiles

We used 16S rRNA gene amplicon sequencing to profile 61 sponge specimens' bacterial communities, representing three genera of the family Latrunculidae (*Tsitsikamma, Cyclacanthia and Latruculia*) and their surrounding water columns. Reads were clustered into OTUs at a distance of 0.03: a proxy for bacterial species. In agreement with previous studies[23–25], we observed that different latrunculid species are associated with distinct bacterial communities significantly different from the surrounding water (Table S4). Each of the sponge species is associated with a distinct microbial community, where each community is dominated by a betaproteobacterium (Fig. S2). *T. favus, T. nguni and T. pedunculata* are all dominated by OTU1 and a closely related betaproteobacterium, OTU2, dominate *T. michaeli* sponges. The *L. algoaensis* sponge and *L. apicalis* sponges harbor their own dominant betaproteobacteria (OTU21 and OTU7 respectively).

Six OTUs were present in all sixty-one sponge specimens (Fig. 1A), three of which were closely related to the dominant betaproteobacterium Sp02-1 and two shared the greatest sequence identity to spirochete clones Sp02-3 and Sp02-15, all of which had been previously identified in *T.*

favus[24]. Remarkably, these bacteria were also present (although in low abundance) in the three *L. apicalis* sponges collected from Bouvet Island, approximately 2850 km southwest of Algoa Bay in the Southern Atlantic Ocean. The sixth conserved OTU, OTU17, is most closely related to uncultured clones isolated from seawater collected from the Santa Barbara channel (MG013048.1) and the North Yellow sea (FJ545595.1). This OTU was also present in our seawater samples (Fig. S2) and may represent a ubiquitous marine bacterium.

We used LC-MS to evaluate the pyrroloiminoquinone suites present in crude chemical extracts derived from each sponge specimen and obtained distinct chemical profiles for each species (Fig 1B). In agreement with previous studies[25], two distinct chemotypes were observed in *T. favus* sponges, and now also in *T. michaeli* sponges collected from Evans Peak and Riy Banks reefs in Algoa Bay, respectively (Fig. 1B, Fig. S3). Chemotype I of *T. favus* has been shown to contain mostly tsitsikammamines (*m/z* 304.10 and 318.12) and brominated discorhabdins (*m/z* 541.85), while chemotype II produced predominantly makaluvamines (*m/z* 188.08, 265.99, 202.09 and 280.00) [25]. The chemical profiles of *T. michaeli* chemotype I (Riy Banks) were found to be similar to those of *T. favus* chemotype I, albeit with a paucity of tsitsikammamines, whereas chemotype II of *T. michaeli* (Evans Peak) produced distinct signals matching unprecedented, hydroxylated discorhabdins (*m/z* 324.13 and 322.12). An indicator species analysis showed there was a significant shift in the abundance of several OTUs between chemotypes (Table S5). The most notable shift shared in both species was a significant increase in the abundance of OTU5 (clone Sp02-15) in *T. favus* chemotype II ($p = 0.005$, R = 0.51) and *T. michaeli* chemotype II ($p = 0.03$, R = 0.87) (Fig. 1C). Therefore, it is likely that the Sp02-15 spirochete plays a role in the chemistry of the latrunculid sponges.

Characterization of putative spirochete Sp02-3 genome bin

Spirochaetes have only been identified as dominant within *Clathrina clathrus* sponges[49] and are present in a lower abundance within several other sponges[50–52], yet their function remains unknown. Similarly, spirochaetes have been identified as dominant in bacterial communities associated with several corals[53–55], where they are predicted to be involved in the fixation of carbon or nitrogen, as performed by spirochaete symbionts in termite guts[53, 55, 56].

Following assembly and binning of four metagenomic datasets, 16S rRNA and 23S rRNA gene sequences were extracted from individual bins. Four genomes were identified as potentially representative of the conserved spirochete Sp02-3 (OTU3), sharing 99.52% sequence identity with the 16S rRNA gene reference sequence (HQ241788.1), and an average of 97% ANI (Table S6). *Salinispira pacifica*, previously identified as a close relative of Sp02-3 [6], was identified as the closest characterized relative of the putative Sp02-3 genomes. In both placement and *de novo* mode, autoMLST assigned *S. pacifica* L21-RPul-D2 as the closest relative to the putative Sp02-3 genome (67.1% ANI) (Fig. 2) and GTDB-Tk placed the bin within the *Salinispira* genus, confirming the 16S rRNA taxonomic classification.

Validation using CheckM showed that genome bin 003D_7 was of the highest quality (Table 1), which was used in the subsequent analysis as a representative genome of Sp02-3. The Sp02-3 genome is approximately 2.48 Mb in size, with an average GC content of 49.87%. The bin comprises 95 contigs, the longest of which was 98 374 bp, and an N50 of 37 020 bp. The Sp02-3 genome is substantially smaller than that of *S. pacifica* (2.48 Mbp vs. 3.78Mb, respectively), with a marginally lower GC content than *S. pacifica* (49.87% vs. 51.9%). The putative Sp02-3 bin's coding density is 77.05%, which is lower than the average bacterial coding density of 85 -

90%[57]. Finally, the identification of 81% of essential genes[43] suggests that this genome is likely complete and may be reduced. Given the divergence of the putative Sp02-3 genomes (Fig. 2) and low (<94%) shared identity[58] (Fig. S4) with other spirochetes, we propose a new taxonomic classification using Bin 003D_7 as the type material. Bin 003D_7 is of high quality as defined by MiMAG standards, with full-length 16S (1524 bp), full-length 23S (2979 bp), two copies of the 5S rRNA genes (111 bp each), as well as 40 tRNAs. We propose the name "*Candidatus* Latrunculis rolihlahla" which acknowledges both the sponge family with which it is associated (Latrunculiidae) and "rolihlahla", which means "troublemaker" in isiXhosa.

Annotation of the genes in the *Ca.* L. rolihlahla genome against the KEGG database indicated that glycolysis and aerobic pyruvate oxidation pathways were present. These genes are usually identified in aerobic bacteria but were also detected in its close relative *S. pacifica*, which is an aerotolerant anaerobe. Very few genes associated with oxidative phosphorylation were identified in Sp02-3 genomes. However, genes *atpEIK* and *ntpABDE* were detected in representative Sp02-3 genome bins 003D_7, 003B_7 and 050C_7 and may be used either to serve as a proton pump or for the anaerobic production of ATP, as proposed for both *S. pacifica* and *Thermus thermophilus[59]*. The *Ca.* L. rolihlahla genome includes all genes required for the assimilatory reduction of sulfate to sulfite and two copies of genes encoding sulfite exporters. No genes involved in dissimilatory sulfate reduction were detected and therefore it is unlikely that *Ca.* L. rolihlahla uses sulfur for anaerobic respiration. It is possible that the produced sulfite may be exported and potentially used by other symbiotic bacteria. *Ca.* L. rolihlahla genome carries a single putative terpene BGC that is also present in *S. pacifica* L21-RPul-D2 and predicted to produce a red-orange pigment hypothesized to protect the bacterium against oxidative stress[59]. The

terpene, if expressed, may perform a similar protective function for *Ca.* L. rolihlahla, which appears to be a facultative anaerobe.

Identification of unique genes in *Ca.* L. rolihlahla relative to *T. favus* metagenome

We identified a total of 23 genes that were functionally unique to the *Ca.* L. rolihlahla genomes, relative to the other bacteria of the *T. favus* metagenomes. Nine of these were common to all three *Ca.* L. rolihlahla genomes (Table S7). Eight of these unique genes encode proteins potentially involved in the transport and metabolism of sugars. Sugar transport/uptake plays an important role in symbiotic interactions[60–63]. Therefore, the presence of these genes would suggest that sugars are similarly important to the spirochete-sponge interaction. Additionally, two unique genes encoding proteins associated with nitrogen fixation were detected (*rnfD* and *nirD:* Electron transport to nitrogenase [64]), but no core genes involved in nitrogen metabolism were found in the spirochete genomes.

The putative *Ca.* L. rolihlahla symbionts also possess additional unique genes predicted to encode an unspecified methyltransferase, a tetratricopeptide repeat protein and *N*-acetylcysteine and a pyridoxal 5'-phosphate synthase (subunits PdxST). This would imply that the spirochete is exclusively capable of producing the active form of vitamin B6, as *pdxA*, *pdxB*, *serC*, *pdxJ*, or *pdxH* genes (DXP-dependent pathway)[65] could not be detected in the rest of the metagenome. Vitamin B6 acts as a cofactor to a wide variety of enzymes across all kingdoms of life [65, 66] and may potentially be used by spirochete symbionts as a competitive advantage in the sponge host[67].

Comparison of putative *Ca.* L. rolihlahla genome bin with other symbiotic spirochetes

Spirochete MAGs from two sponges (*Aplysina aerophoba* and *Iophon methanophila* M2), a *Neotermes castaneus* termite, and three species of gutless marine worms were acquired along with five free-living species and compared to Sp02-3 bin (Table S2). We identified a total of 4037 orthologous genes (OGs) between the 14 genomes/bins and the hierarchical clustering of gene presence/absence showed that the *Ca.* L. rolihlahla symbionts did not cluster with *S. pacifica* and other free-living spirochetes (Fig 3A). However, the *Ca.* L. rolihlahla bins shared the greatest abundance of orthologous genes with free-living spirochete *S. pacifica* (~21% average shared OGs) (Fig. 3B) and less than 14% with any of the host-associated spirochetes which would confirm *S. pacifica* as a close relative and suggests that the difference in gene repertoire (Fig. 3A) may be due to gene loss in the *Ca.* L. rolihlahla bins.

Only six genes were shared between the *Ca.* L. rolihlahla genomes and the host-associated spirochetes that were not present in the free-living species. There may be more functionally similar genes present between the genomes, but as identification of orthologous genes is based on shared sequence, sequence drift may confound this identification and result in an underestimation of functional conservation. The shared genes were predicted to encode L-fucose isomerase, L-fucose mutarotase, xylose import ATP-binding protein XylG, ribose import permease protein RbsC, an ABC transporter permease and autoinducer two import system permease protein LsrD. Fucose mutarotase and isomerase perform the first two steps in fucose utilization and are responsible for converting beta-fucose to fuculose[68]. The role of fucosylated glycans in mammalian guts as points of adherence for various bacteria has been well documented[69–72] and it is possible that *Ca.* L. rolihlahla spirochetes are similarly capable of binding fucose in their respective hosts. The remaining genes are likely involved in the uptake and utilization of sugars.

Characterization of putative betaproteobacteria Sp02-1 genome bin

Genome bin 003B_4, from sponge specimen TIC2018-003B, carried a 16S rRNA gene sequence that shared 99.86% identity with *T. favus*-associated betaproteobacterium Sp02-1 clone (HQ241787.1). The next closest relatives were uncultured betaproteobacterium clones from a *Xestospongia muta* and *Tethya aurantium* sponges (Fig. S5). Bins 050A_14, 050C_6 and 003D_6 were also identified as possible representatives of the betaproteobacterium Sp02-1 based on their predicted phylogenetic relatedness (Table S2); however, they were of low quality and not used in downstream analyses.

Genome bin 003B_4 served as a representative of the Sp02-1 bacterium. Bin 003B_4 is approximately 2.95 Mbp in size and medium quality per MIMAG standards [32] (Table S2). There was 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 [57]. An abundance of pseudogenes and low coding density, is usually an indication that the genome in question may be undergoing genome reduction [73] similar to other betaproteobacteria in the proposed order of Tethybacterales [20].

The Sp02-1 genome had all genes necessary for glycolysis, PRPP biosynthesis and most genes required for the citrate cycle and oxidative phosphorylation were identified. The genome has the genes necessary to biosynthesize only valine, leucine, isoleucine, tryptophan, phenylalanine, tyrosine and ornithine amino acids and 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[74–76].

Sp02-1 was the only genome to carry all genes necessary for the periplasmic conversion of nitrate to nitrous oxide (*napAB*, *nirK*, *norB*) in the metagenomes. No genes encoding nitrous oxide reductase (*nosZ*) were detected and therefore, full denitrification may not be possible. However, this loss of a functional *nosZ* gene has been observed in other denitrifying bacteria resulting in the release of nitrous oxide, which is reduced to nitrogen by other bacteria in the environment[77] or as in the case of *Ircinia ramosa* sponge-associated bacteria, the reduction stops following production of nitrous oxide and aids in the removal of nitrogen from the sponge system[78]. The putative Sp02-1 was also the only genome in all four sponge metagenomes that appeared to include both sulfite reductase genes *cysI* and *cysJ* as observed in many other sponge-associated betaproteobacteria[20]. It carried no other genes encoding sulfur-reducing enzymes. However, the potential expression of *cysIJ* would enable the complete reduction of the sulfite potentially produced by the putative spirochete symbiont Sp02-3 to sulfide, which can potentially aid in the removal of sulfur from the system.

Genes unique to the Sp02-1 genome relative to the *T. favus* sponge metagenome

A total of 13 genes unique to the Sp02-1 genome were identified using the same approach as the spirochete symbionts. Two of these genes were predicted to encode an ABC transporter permease subunit that was likely involved in glycine betaine and proline betaine uptake and a 5-oxoprolinase subunit PxpA, respectively (Table S8). The presence of the genes suggests that the Sp02-1 genome can acquire proline and convert it to glutamate[79] in addition to glutamate already produced via glutamate synthase. Other unique genes encoded a restriction endonuclease subunit and site-specific DNA-methyltransferase which would presumably aid in defense against foreign DNA. In contrast to this supposed function, at least seven of the unique genes are predicted to be associated with phages, including an anti-restriction protein ArdA, which may inhibit the action of restriction-

modification systems[80]. 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 involved in cell-adhesion and protein-protein interactions[81, 82] and may help facilitate the symbiosis between the Sp02-1 betaproteobacterium and *T. favus*.

Comparison of putative Sp02-1 and other Tethybacterales

Several betaproteobacteria sponge symbionts have been described to date and these bacteria are thought to have functionally diversified following the initiation of their ancient partnership[20]. Twelve genomes/MAGs of Tethybacterales were downloaded from the JGI database. Additionally, we assembled and binned metagenomic data from thirty-six sponge SRA datasets, covering fourteen sponge species (Table S3) and recovered an additional fourteen AqS2-like genomes (Table S2). Ten of these bins were low quality, so Bin 003B_4 (Sp02-1) and sixteen medium quality Tethybacterales bins/genomes were used for further analysis (Table 2).

Phylogeny of the Tethybacterales symbionts inferred using single-copy markers revealed a deep branching clade of these sponge-associated symbionts (Fig. 4) and that bin 003B_4 clustered within the Persebacteraceae family (Fig. 4), members of which dominate the microbial community of their respective sponge hosts[10, 23, 24, 83]. We additionally identified what appears to be a third family, consisting of symbionts associated with *C. singaporensis* and *Cinachyrella* sponge species. Assessment of shared AAI indicates that these genomes represent a new family, sharing an average of 80% AAI (Table S9)[58]. Additionally, these three families share less than 89% sequence similarity with respect to their 16S rRNA sequences and intra-clade differences of less than 92%. Therefore, they may represent novel classes within the Tethybacterales order (Table

S9)[58]. In keeping with naming the families after Oceanids of Greek mythology[20], we propose Kalypsobacteraceae, which means “hidden”.

We identified 4306 orthologous genes between all seventeen Tethybacterales genomes, with only eighteen genes common to all the genomes. 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. 5A). Thirteen of the shared genes between all Tethybacterales genomes encoded ribosomal proteins or those involved in energy production. Genes encoding chorismate synthase were conserved across all seventeen genomes and suggest that tryptophan production may be conserved among these bacteria. According to a recent study, *Dysidea etheria* and *A. queenslandica* sponges cannot produce tryptophan, which is an essential amino acid, which may indicate a central role for the Tethybacterales symbionts[84]. This would be a beneficial advantage to the Latrunculiidae sponges, as tryptophan is the predicted starter molecule for pyrroloiminoquinone compounds[21, 25].

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[85–89]. Internal changes in oxygen levels[90], anthropogenic activity[91] and temperature changes[60, 92, 93] 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 adapted to their fluctuating environment.

Alignment of the Tethybacterales genes against the KEGG database (Table S10) revealed some trends that differentiated the three families (Fig. 5B): (1) the genomes of the proposed Kalypsobacteraceae 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 differed to some extent in what can be transported in and out of the symbiont cell (Fig. 5C). Proposed members of the Kalypsobacteraceae are exclusively capable of transporting hydroxyproline which may suggest a role in collagen degradation[94]. The Tethybacteraceae and Persebacteraceae appear able to transport spermidine, putrescine, taurine, and glycine which, in combination with their potential ability to reduce nitrates, may suggest a role in C-N cycling[95]. All three families transport various amino acids as well as phospholipids and heme. The transport of amino acids between symbiont and sponge host has previously been observed[96] and may provide the Tethybacterales with a competitive advantage over other sympatric microorganisms[97] and possibly allow the sponge hosts to regulate the symbioses via regulation of the quantity of amino acids available for symbiont uptake[98]. Similarly, the transfer of heme in the iron-starved ocean environment between sponge host and symbiont could provide a selective advantage[99]. The Tethybacteraceae were distinct in their ability to transport sugars. As mentioned earlier, the transport of sugars plays an important role in symbiotic interactions[60–63] and it is possible that this family of symbionts require sugars from their sponge hosts.

Finally, we wanted to examine the approximate divergence pattern between these symbionts and whether they diverged in a pattern similar to that of their host species. The eighteen shared genes were used to estimate the rate of synonymous substitution, which provides an approximation for

the pattern of divergence between the species[100]. We found that the estimated divergence pattern (Fig. 6) and the phylogeny of the host sponges was incongruent. Phylogenetic trees inferred using single-copy marker genes (Fig. 4) and the comprehensive 16S rRNA tree published by Taylor and colleagues[20] confirm this lack of congruency between symbiont and host phylogeny. Other factors such as collection site or depth could not explain the observed trend. This would suggest that these sponges likely acquired a free-living Tethybacterales common ancestor from the environment at different time points throughout their evolution. There is evidence that suggests co-evolution of betaproteobacteria symbionts with their sponge host in *Tsitsikamma[23]*, *Tethya[12]*, *Tedania[101]* and *A. queenslandica[19]* sponge species and so we conclude that the symbiont coevolved with its respective host subsequent to acquisition.

CONCLUSION

The sponges within the Latrunculiidae family host closely related conserved spirochetes and betaproteobacteria. The reason for this consistent conservation over both time and geographic space had not yet been elucidated. This study has shown that the conserved spirochete may not only play a role as a nutritional symbiont through its production of vitamin B6, but that it and the closely related spirochete Sp02-15 likely play a role in the production or modification of cytotoxic pyrroloiminoquinone alkaloids that are a signature of these sponges. We are currently working toward extracting the genome of spirochete Sp02-15. Similarly, the conservation of the betaproteobacterium appears to be driven by its ability to reduce nitrates and potentially produce tryptophan for pyrroloiminoquinone production. This work has expanded our understanding of the Tethybacterales and the functional specialization of the various families within this new order. Finally, this study found that the phylogeny of the Tethybacterales is incongruent with that of the sponge host phylogeny, which suggests that the association of Tethybacterales and their sponge

hosts did not occur once but rather likely occurred multiple times over evolutionary history. It is likely that this symbiosis is more ubiquitous than previously thought and addition of new genomes will no doubt benefit our understanding of this new family and their evolutionary relationship with marine sponges.

ACKNOWLEDGEMENTS

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.J.K.) and by an NRF Innovation and Rhodes University Henderson PhD Scholarships. J-C.J.K. and L.M. were supported by a SARChI PhD Fellowships (UID: 87583) and 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. 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 thank Ceridwen Fraser (University of Otago) and Rachel Downey (Australian National University) for subsamples of *L. apicalis* samples

collected during the Antarctic Circumnavigation Expedition. 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.

Competing Interests

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

REFERENCES

1. Kiran GS, Sekar S, Ramasamy P, Thinesh T, Hassan S, Lipton AN, et al. Marine sponge microbial association: Towards disclosing unique symbiotic interactions. *Mar Environ Res* 2018; **140**: 169–179.
2. Lafi FF, Fuerst JA, Fieseler L, Engels C, Goh WWL, Hentschel U. Widespread distribution of poribacteria in demospongiae. *Appl Environ Microbiol* 2009; **75**: 5695–5699.
3. Fieseler L, Horn M, Wagner M, Hentschel U. Discovery of the novel candidate phylum 'Poribacteria' in marine sponges. *Appl Environ Microbiol* 2004; **70**: 3724–3732.
4. Siegl A, Kamke J, Hochmuth T, Piel J, Richter M, Liang C, et al. Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. *ISME J* 2011; **5**: 61–70.

5. Kamke J, Rinke C, Schwientek P, Mavromatis K, Ivanova N, Sczyrba A, et al. The candidate phylum Poribacteria by single-cell genomics: New insights into phylogeny, cell-compartmentation, eukaryote-like repeat proteins, and other genomic features. *PLoS One* 2014; **9**: e87353.
6. Kamke J, Sczyrba A, Ivanova N, Schwientek P, Rinke C, Mavromatis K, et al. Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *ISME J* 2013; **7**: 2287–2300.
7. Jahn MT, Markert SM, Ryu T, Ravasi T, Stigloher C, Hentschel U, et al. Shedding light on cell compartmentation in the candidate phylum Poribacteria by high resolution visualisation and transcriptional profiling. *Sci Rep* 2016; **6**: 35860.
8. Podell S, Blanton JM, Neu A, Agarwal V, Biggs JS, Moore BS, et al. Pangenomic comparison of globally distributed Poribacteria associated with sponge hosts and marine particles. *ISME J* 2019; **13**: 468–481.
9. Cleary DFR, Swierts T, Coelho FJRC, Polónia ARM, Huang YM, Ferreira MRS, et al. The sponge microbiome within the greater coral reef microbial metacommunity. *Nat Commun* 2019; **10**: 1644.
10. Croué J, West NJ, Escande M-L, Intertaglia L, Lebaron P, Suzuki MT. A single betaproteobacterium dominates the microbial community of the crambescidine-containing sponge *Crambe crambe*. *Sci Rep* 2013; **3**: 2583.
11. Thiel V, Neulinger SC, Staufenberger T, Schmaljohann R, Imhoff JF. Spatial distribution of sponge-associated bacteria in the Mediterranean sponge *Tethya aurantium*. *FEMS Microbiol Ecol* 2007; **59**: 47–63.

12. Waterworth SC, Jiwaji M, Kalinski J-CJ, Parker-Nance S, Dorrington RA. 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* 2017; **15**.
13. Cárdenas CA, Bell JJ, Davy SK, Hoggard M, Taylor MW. Influence of environmental variation on symbiotic bacterial communities of two temperate sponges. *FEMS Microbiol Ecol* 2014; **88**: 516–527.
14. Steinert G, Taylor MW, Deines P, Simister RL, de Voogd NJ, Hoggard M, et al. In four shallow and mesophotic tropical reef sponges from Guam the microbial community largely depends on host identity. *PeerJ* 2016; **4**: e1936.
15. Webster NS, Wilson KJ, Blackall LL, Hill RT. Phylogenetic diversity of bacteria associated with the marine sponge *Rhopaloeides odorabile*. *Appl Environ Microbiol* 2001; **67**: 434–444.
16. Cleary DFR, Becking LE, de Voogd NJ, Pires ACC, Polónia ARM, Egas C, et al. Habitat- and host-related variation in sponge bacterial symbiont communities in Indonesian waters. *FEMS Microbiol Ecol* 2013; **85**: 465–482.
17. Trindade-Silva AE, Rua C, Silva GGZ, Dutilh BE, Moreira APB, Edwards RA, et al. Taxonomic and functional microbial signatures of the endemic marine sponge *Arenosclera brasiliensis*. *PLoS One* 2012; **7**: e39905.
18. Gauthier M-EA, Watson JR, Degnan SM. Draft genomes shed light on the dual bacterial symbiosis that dominates the microbiome of the coral reef sponge *Amphimedon queenslandica*. *Frontiers in Marine Science* 2016; **3**: 196.
19. Fieth RA, Gauthier M-EA, Bayes J, Green KM, Degnan SM. Ontogenetic changes in the bacterial symbiont community of the tropical demosponge *Amphimedon queenslandica*: Metamorphosis is a new beginning. *Frontiers in Marine Science* 2016; **3**: 228.

20. Taylor JA, Palladino G, Wemheuer B, Steinert G, Sipkema D, Williams TJ, et al. Phylogeny resolved, metabolism revealed: Functional radiation within a widespread and divergent clade of sponge symbionts. *ISME J* 2020.

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

22. Hu J-F, Fan H, Xiong J, Wu S-B. Discorhabdins and pyrroloiminoquinone-related alkaloids. *Chem Rev* 2011; **111**: 5465–5491.

23. Matcher GF, Waterworth SC, Walmsley TA, Matsatsa T, Parker-Nance S, Davies-Coleman MT, et al. Keeping it in the family: Coevolution of latrunculid sponges and their dominant bacterial symbionts. *Microbiologyopen* 2017; **6**.

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

25. Kalinski J-CJ, Waterworth SC, Noundou XS, Jiwaji M, Parker-Nance S, Krause RWM, et al. Molecular networking reveals two distinct chemotypes in pyrroloiminoquinone-producing *Tsitsikamma favus* sponges. *Mar Drugs* 2019; **17**.

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

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

28. Hamlyn-Harris R, Queensland Museum, Hamlyn-Harris R. Memoirs of the Queensland Museum. 1996; **40 (1996)**.

29. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012; **19**: 455–477.

30. Miller IJ, Rees ER, Ross J, Miller I, Baxa J, Lopera J, et al. Autometa: Automated extraction of microbial genomes from individual shotgun metagenomes. *Nucleic Acids Res* 2019; **47**: e57.

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

32. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 2017; **35**: 725–731.

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

34. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 2019.

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

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

37. Seemann T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 2014; **30**: 2068–2069.

38. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *bioRxiv* 2019.

39. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, et al. antiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* 2019; **47**: W81–W87.

40. Altenhoff AM, Levy J, Zarowiecki M, Tomiczek B, Warwick Vesztrocy A, Dalquen DA, et al. OMA standalone: Orthology inference among public and custom genomes and transcriptomes. *Genome Res* 2019; **29**: 1152–1163.

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

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

43. Waterworth SC, Flórez LV, Rees ER, Hertweck C, Kaltenpoth M, Kwan JC. Horizontal Gene Transfer to a Defensive Symbiont with a Reduced Genome in a Multipartite Beetle Microbiome. *MBio* 2020; **11**.

44. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionarygenetics analysis across computing platforms. *Mol Biol Evol* 2018; **35**: 1547–1549.

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

46. Sneath PHA, Sokal RR. Numerical taxonomy. The principles and practice of numerical classification. 1973. W.H. Freeman and Co. San Francisco.

47. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* 2016; **34**: 828–837.

48. Pluskal T, Castillo S, Villar-Briones A, Oresic M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 2010; **11**: 395.

49. Neulinger SC, Stöhr R, Thiel V, Schmaljohann R, Imhoff JF. New phylogenetic lineages of the Spirochaetes phylum associated with *Clathrina* species (Porifera). *J Microbiol* 2010; **48**: 411–418.

50. Villegas-Plazas M, Wos-Oxley ML, Sanchez JA, Pieper DH, Thomas OP, Junca H. Variations in Microbial Diversity and Metabolite Profiles of the Tropical Marine Sponge *Xestospongia muta* with Season and Depth. *Microb Ecol* 2019; **78**: 243–256.

51. Isaacs LT, Kan J, Nguyen L, Videau P, Anderson MA, Wright TL, et al. Comparison of the bacterial communities of wild and captive sponge *Clathria prolifera* from the Chesapeake Bay. *Mar Biotechnol* 2009; **11**: 758–770.

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

53. van de Water JAJM, Melkonian R, Junca H, Voolstra CR, Reynaud S, Allemand D, et al. Spirochaetes dominate the microbial community associated with the red coral *Corallium rubrum* on a broad geographic scale. *Sci Rep* 2016; **6**: 27277.

54. Wessels W, Sprungala S, Watson S-A, Miller DJ, Bourne DG. The microbiome of the octocoral *Lobophytum pauciflorum*: Minor differences between sexes and resilience to short-term stress. *FEMS Microbiol Ecol* 2017; **93**.

55. Lawler SN, Kellogg CA, France SC, Clostio RW, Brooke SD, Ross SW. Coral-associated bacterial diversity is conserved across two deep-sea *Anthothela* species. *Front Microbiol* 2016; **7**: 458.

56. Lilburn TG, Kim KS, Ostrom NE, Byzek KR, Leadbetter JR, Breznak JA. Nitrogen fixation by symbiotic and free-living spirochetes. *Science* 2001; **292**: 2495–2498.

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

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

59. Ben Hania W, Joseph M, Schumann P, Bunk B, Fiebig A, Spröer C, et al. Complete genome sequence and description of *Salinispira pacifica* gen. nov., sp. nov., a novel spirochaete isolated form a hypersaline microbial mat. *Stand Genomic Sci* 2015; **10**: 7.

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

61. Ekman M, Picossi S, Campbell EL, Meeks JC, Flores E. A *Nostoc punctiforme* sugar transporter necessary to establish a Cyanobacterium-plant symbiosis. *Plant Physiol* 2013; **161**: 1984–1992.

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

63. Rix L, Ribes M, Coma R, Jahn MT, de Goeij JM, van Oevelen D, et al. Heterotrophy in the earliest gut: A single-cell view of heterotrophic carbon and nitrogen assimilation in sponge-microbe symbioses. *ISME J* 2020.

64. Jeong HS, Jouanneau Y. Enhanced nitrogenase activity in strains of *Rhodobacter capsulatus* that overexpress the rnf genes. *J Bacteriol* 2000; **182**: 1208–1214.

65. Richts B, Rosenberg J, Commichau FM. A survey of pyridoxal 5'-phosphate-dependent proteins in the Gram-positive model bacterium *Bacillus subtilis*. *Front Mol Biosci* 2019; **6**: 32.

66. Fleischman NM, Das D, Kumar A, Xu Q, Chiu H-J, Jaroszewski L, et al. Molecular characterization of novel pyridoxal-5'-phosphate-dependent enzymes from the human microbiome. *Protein Sci* 2014; **23**: 1060–1076.

67. Pita L, Rix L, Slaby BM, Franke A, Hentschel U. The sponge holobiont in a changing ocean: From microbes to ecosystems. *Microbiome* 2018; **6**: 46.

68. Higgins MA, Boraston AB. Structure of the fucose mutarotase from *Streptococcus pneumoniae* in complex with L-fucose. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2011; **67**: 1524–1530.

69. Pickard JM, Chervonsky AV. Intestinal fucose as a mediator of host-microbe symbiosis. *J Immunol* 2015; **194**: 5588–5593.

70. Becerra JE, Yebra MJ, Monedero V. An L-fucose operon in the probiotic *Lactobacillus rhamnosus* GG is involved in adaptation to gastrointestinal conditions. *Appl Environ Microbiol* 2015; **81**: 3880–3888.

71. Robbe C, Capon C, Coddeville B, Michalski J-C. Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *Biochem J* 2004; **384**: 307–316.

72. Coyne MJ, Reinap B, Lee MM, Comstock LE. Human symbionts use a host-like pathway for surface fucosylation. *Science* 2005; **307**: 1778–1781.

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

74. Feng H, Edwards N, Anderson CMH, Althaus M, Duncan RP, Hsu Y-C, et al. Trading amino acids at the aphid-*Buchnera* symbiotic interface. *Proc Natl Acad Sci U S A* 2019; **116**: 16003–16011.

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

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

77. Harter J, El-Hadidi M, Huson DH, Kappler A, Behrens S. Soil biochar amendment affects the diversity of *nosZ* transcripts: Implications for N_2O formation. *Sci Rep* 2017; **7**: 3338.

78. Engelberts JP, Robbins SJ, de Goeij JM, Aranda M, Bell SC, Webster NS. Characterization of a sponge microbiome using an integrative genome-centric approach. *ISME J* 2020; **14**: 1100–1110.

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

80. Chen K, Reuter M, Sanghvi B, Roberts GA, Cooper LP, Tilling M, et al. ArdA proteins from different mobile genetic elements can bind to the *Eco*KI Type I DNA methyltransferase of *E. coli* K12. *Biochim Biophys Acta* 2014; **1844**: 505–511.

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

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

83. Said Hassane C, Fouillaud M, Le Goff G, Sklirou AD, Boyer JB, Trougakos IP, et al. Microorganisms associated with the marine sponge *Scopalina hapalia*: A reservoir of bioactive molecules to slow down the aging process. *Microorganisms* 2020; **8**.

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

85. Voth W, Jakob U. Stress-Activated Chaperones: A First Line of Defense. *Trends Biochem Sci* 2017; **42**: 899–913.

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

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

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

89. Gennaris A, Ezraty B, Henry C, Agrebi R, Vergnes A, Oheix E, et al. Repairing oxidized proteins in the bacterial envelope using respiratory chain electrons. *Nature* 2015; **528**: 409–412.

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

91. Efremova SM, Itskovich VB, Parfenova V, Drucker VV, Müller WEG, Schröder HC. Lake Baikal: A unique place to study evolution of sponges and their stress response in an environment nearly unimpaired by anthropogenic perturbation. *Cell Mol Biol* 2002; **48**: 359–371.

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

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

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

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

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

97. Hosie AHF, Allaway D, Galloway CS, Dunsby HA, Poole PS. *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* 2002; **184**: 4071–4080.

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

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

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

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

102. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 2007; **24**: 1586–1591.

TABLE AND FIGURE LEGENDS

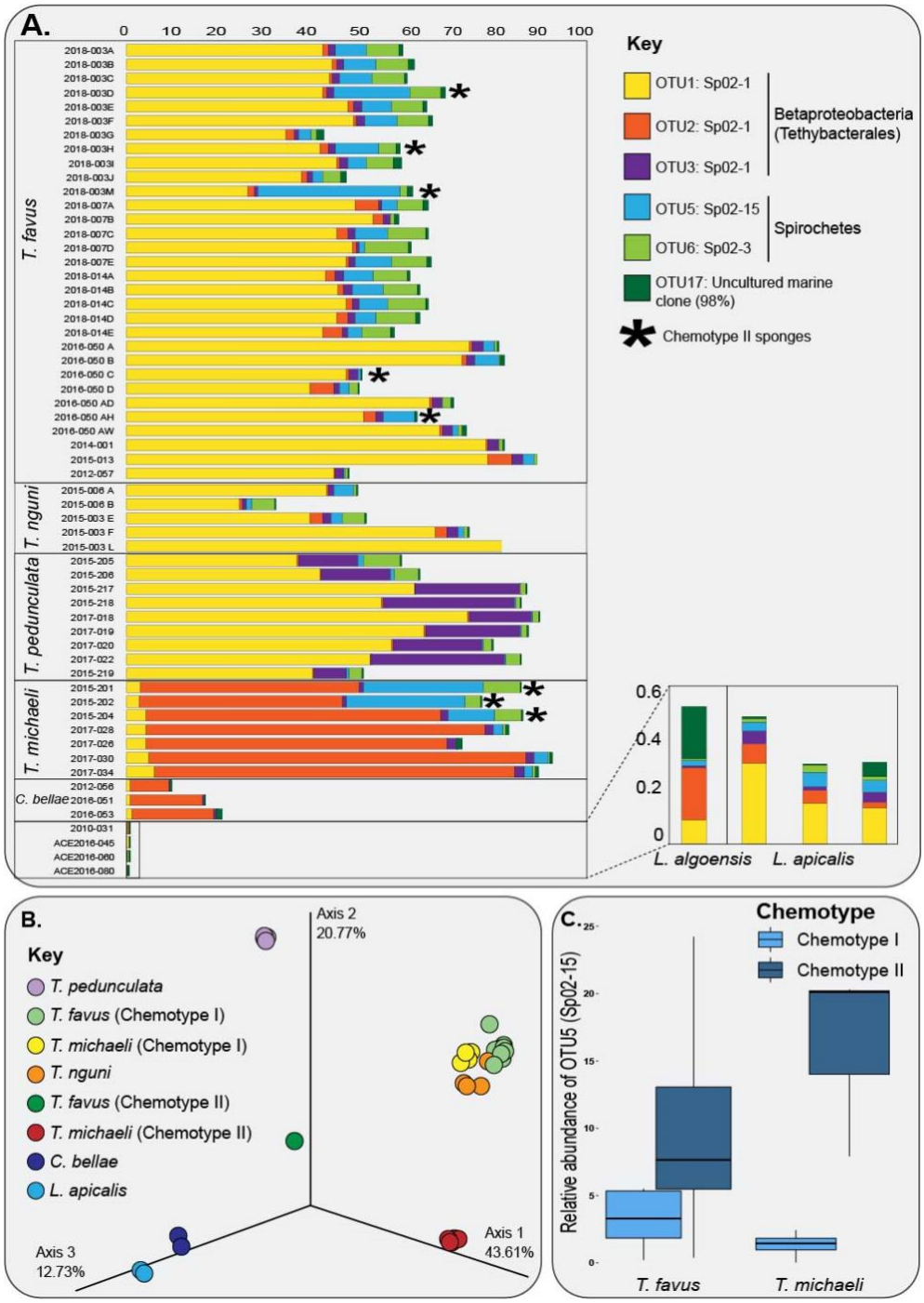

Figure 1. The relative abundance of conserved bacterial symbionts in all latrunculid sponges and chemical profiles. A total of six OTUs were present in all 61 latrunculid sponges investigated here, 58 of which were collected off the eastern coast of South Africa and three were collected from Bouvet Island in the Southern Ocean. **A)** Analysis of the 16S rRNA gene sequence amplicon data revealed the presence of six bacterial species conserved across all sponge samples. The abundance of conserved OTUs (clustered at a distance of 0.03) relative to the total reads per sample is shown **B)** Principal Component Analysis (using Bray-Curtis dissimilarity metrics) of pyrroloiminoquinone

compounds in latrunculid sponge specimens illustrates the distinct chemistry associated with the different sponge species, as well as the different chemotypes observed in *T. favus* and *T. michaeli* sponges. **C)** An indicator species analysis revealed that the abnormal Chemotype II sponges (denoted with *) were associated with a significant increase in the abundance of OTU5 (spirochete Sp02-15).

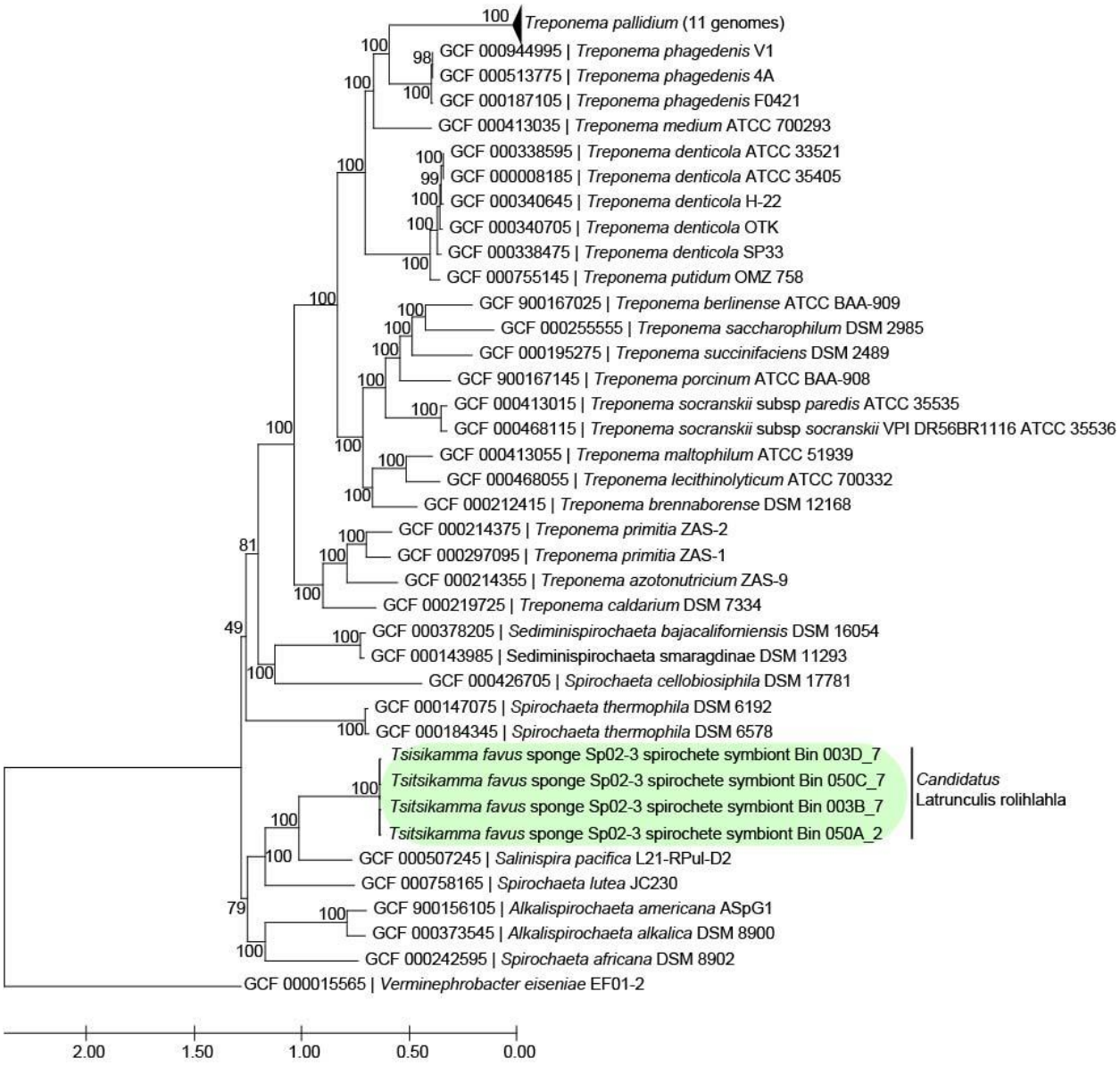

Figure 2. Putative spirochete Sp02-3 genomes form a distinct clade. A multi-locus *de novo*, concatenated species tree of putative *T. favus* associated Sp02-3 spirochete symbiont genomes (highlighted in green). The scale bar indicates the number of substitutions per site of the concatenated genes.

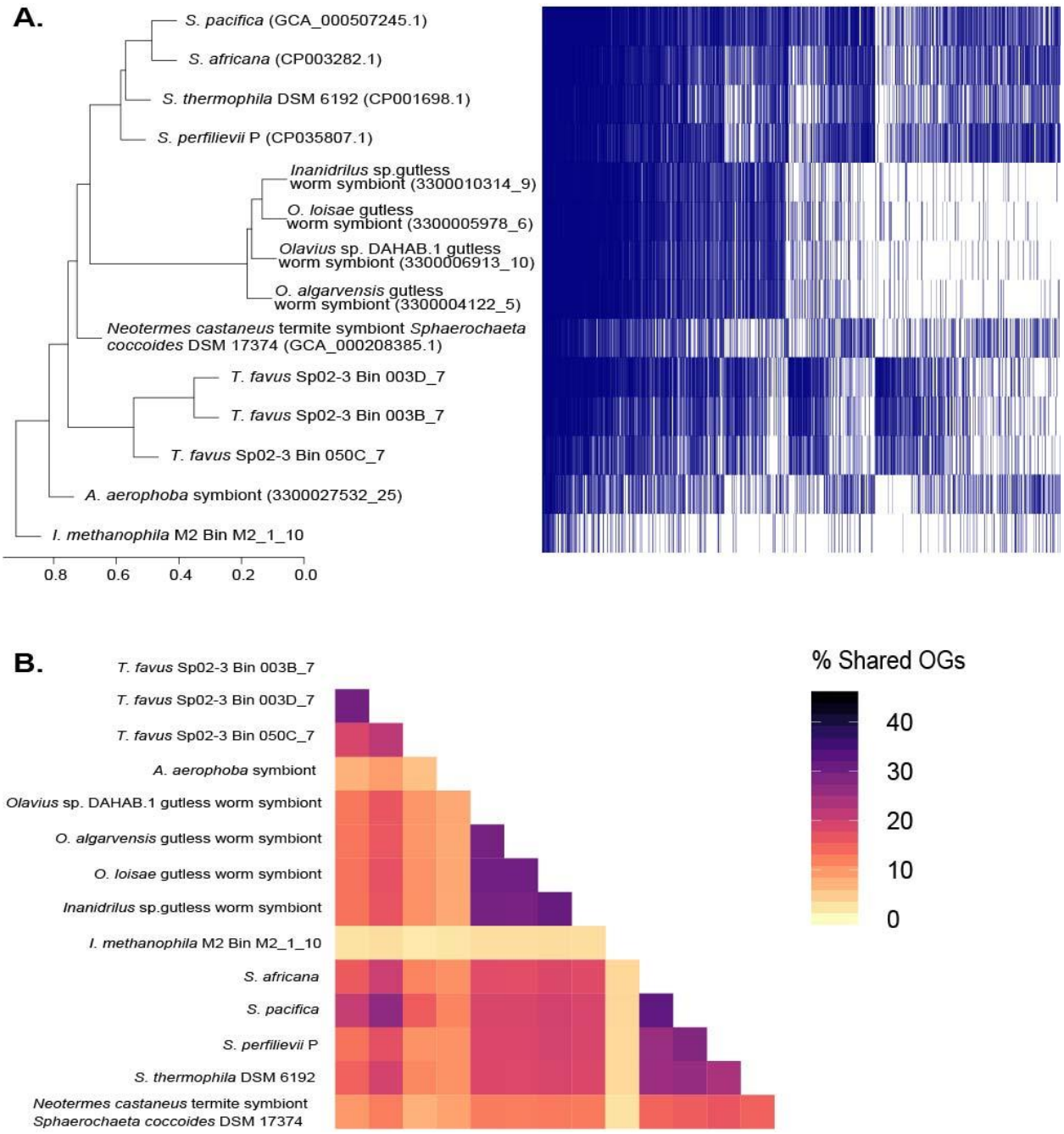

Figure 3. Gene repertoires of free-living and symbiotic spirochetes. A) Hierarchical clustering of shared orthologous genes between genomes showed that *T. favus* spirochete symbionts cluster with other host-associated spirochetes, but **B)** share the greatest number of orthologous genes with *S. pacifica*, their closest phylogenetic relative.

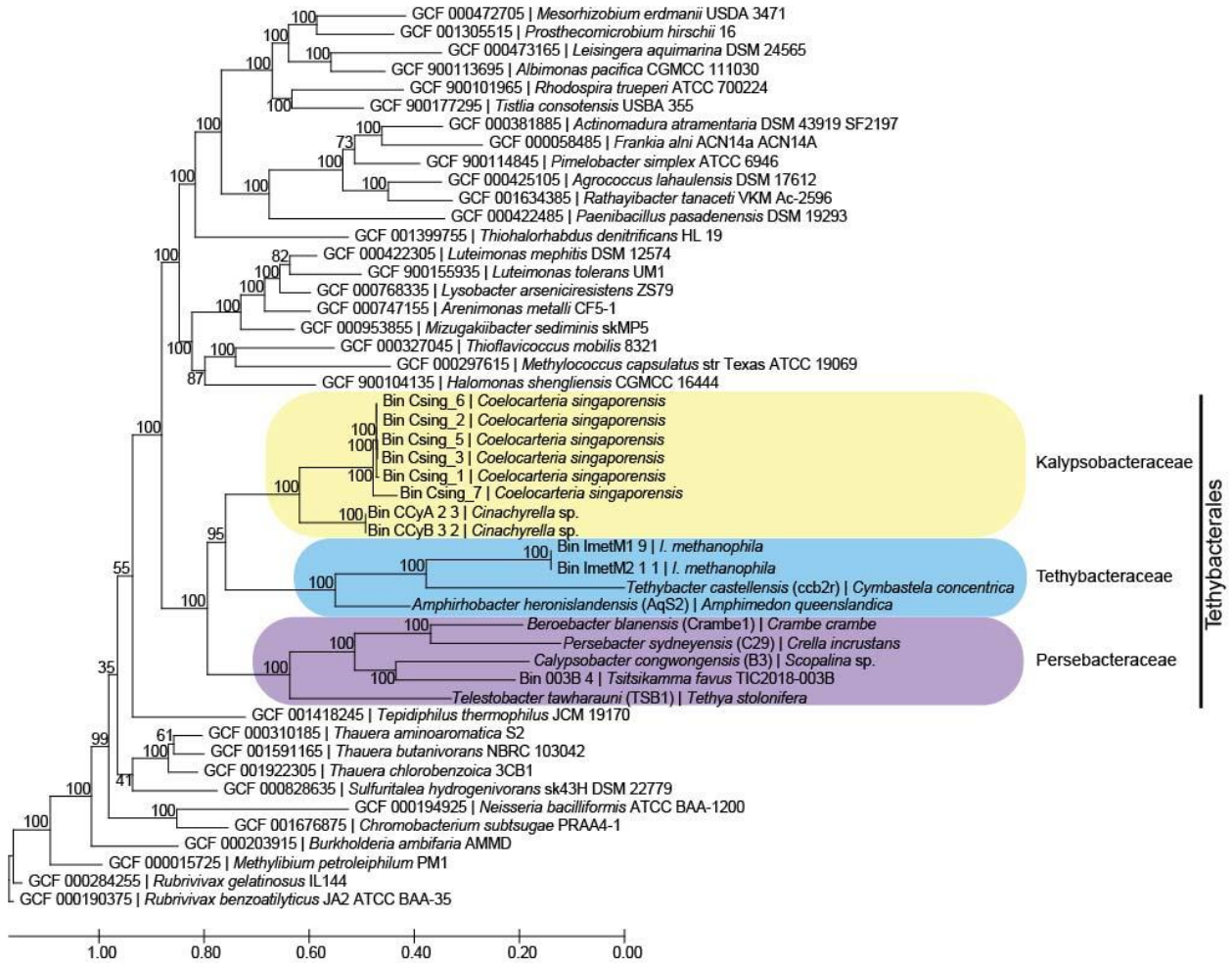

Figure 4. Phylogeny of the Tethybacterales sponge symbionts. Single-copy markers were used to delineate the phylogeny of these sponge-associated betaproteobacteria revealing a new family of symbionts in the Tethybacterales order. Additionally, it was shown that the *T. favus* associated Sp02-1 symbiont 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.

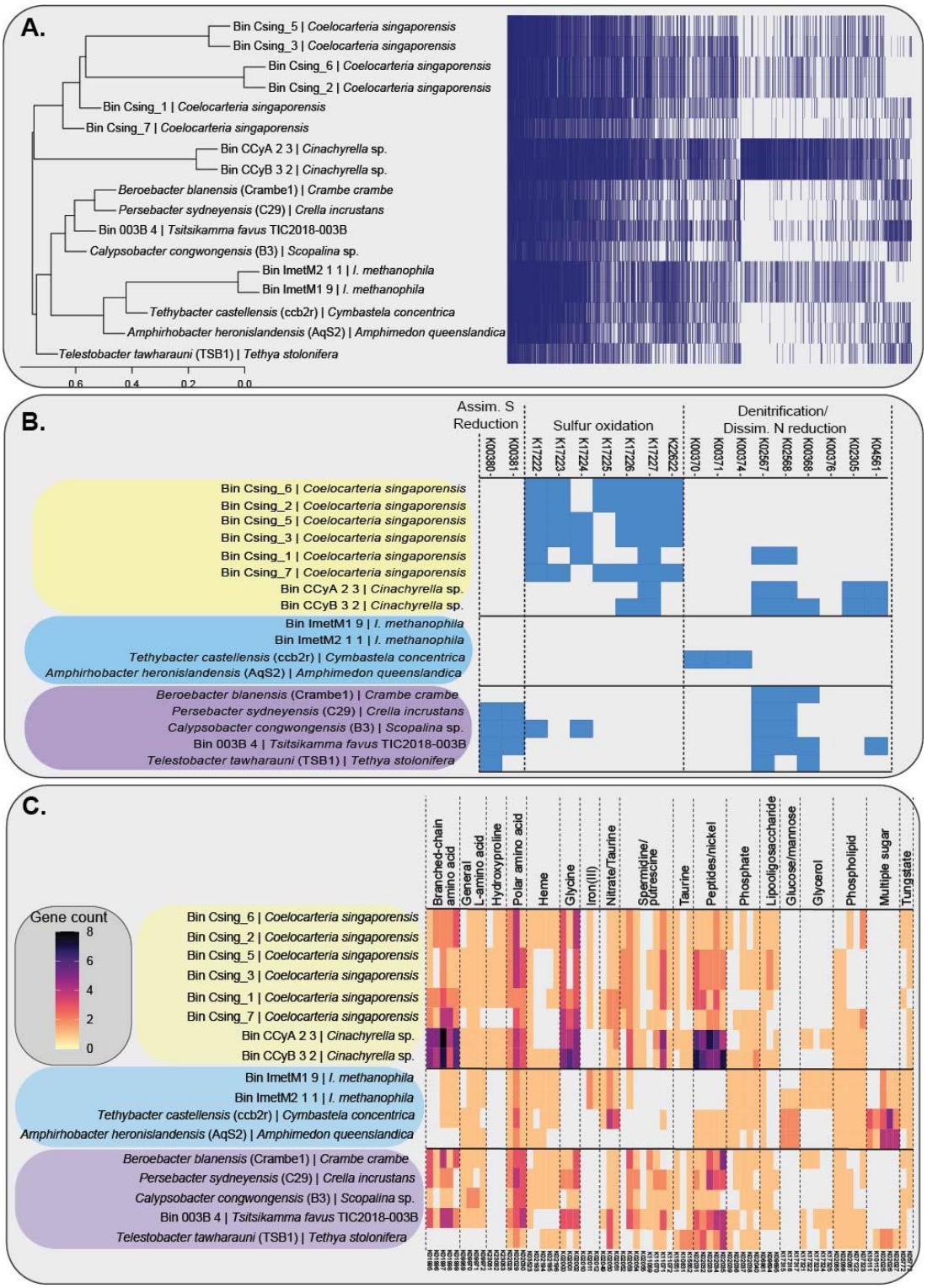

Figure 5. Functional specialization of Tethybacterales families. The newly proposed Tethybacterales order appears to consist of three bacterial families. These families appear to have similar gene distribution (**A**) where the potential function of these genes indicates specialization in nutrient cycling (**B**) and solute transport (**C**).

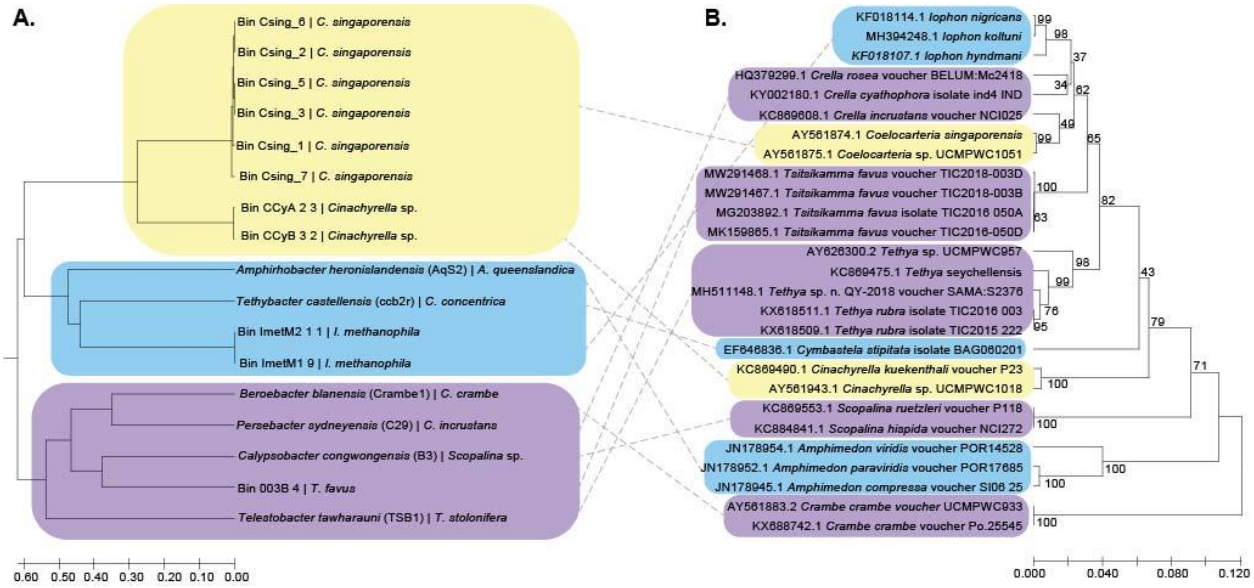

Figure 6. The divergence pattern of Tethybacterales and their host sponges. The divergence of the Tethybacterales is incongruent with the phylogeny of the host sponges. **(A)** Branch length of symbiont divergence estimates is proportional to the pairwise rate of synonymous substitution calculated (ML estimation) using a concatenation of 18 genes common to all 17 genomes. Rate of synonymous substitution was calculated using PAL2NAL and CodeML from the PAML package[102] and visualized in MEGAX[44]. **(B)** Phylogeny of host sponges (or close relatives thereof) was inferred 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.

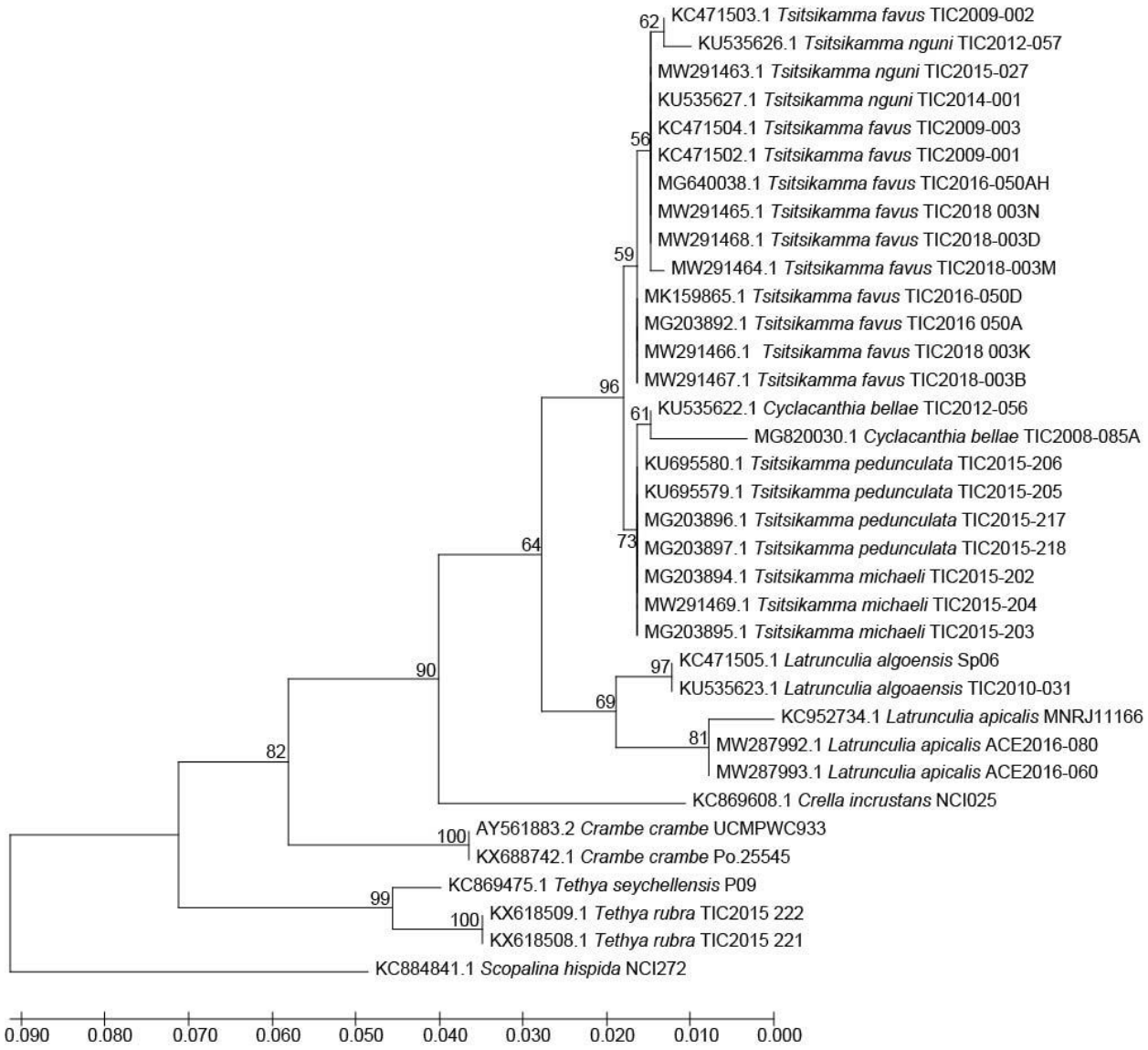

Figure S1. Phylogeny of sponges within the family Latrunculiidae. Sponge phylogeny was inferred using 28S rRNA sequence data using the Maximum-likelihood method and Tamura-Nei model with 1000 bootstrap replicates. Branch lengths indicate the number of substitutions per site. All ambiguous positions were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

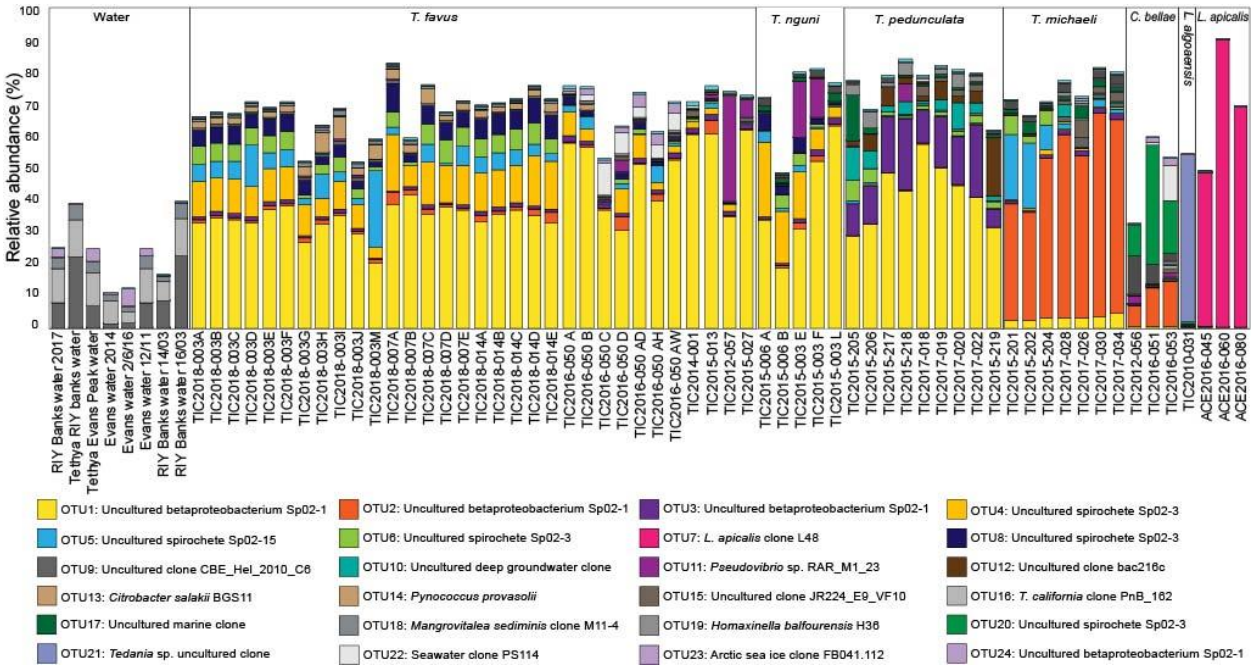

Figure S2. Latrunculid microbial community analysis. The top 24 OTUs (clustered at 0.03) associated with latrunculid sponges collected from Algoa Bay, South Africa and Bouvet Island in the Southern Ocean. Colored key shows the closest relative when aligned against the NCBI nr database

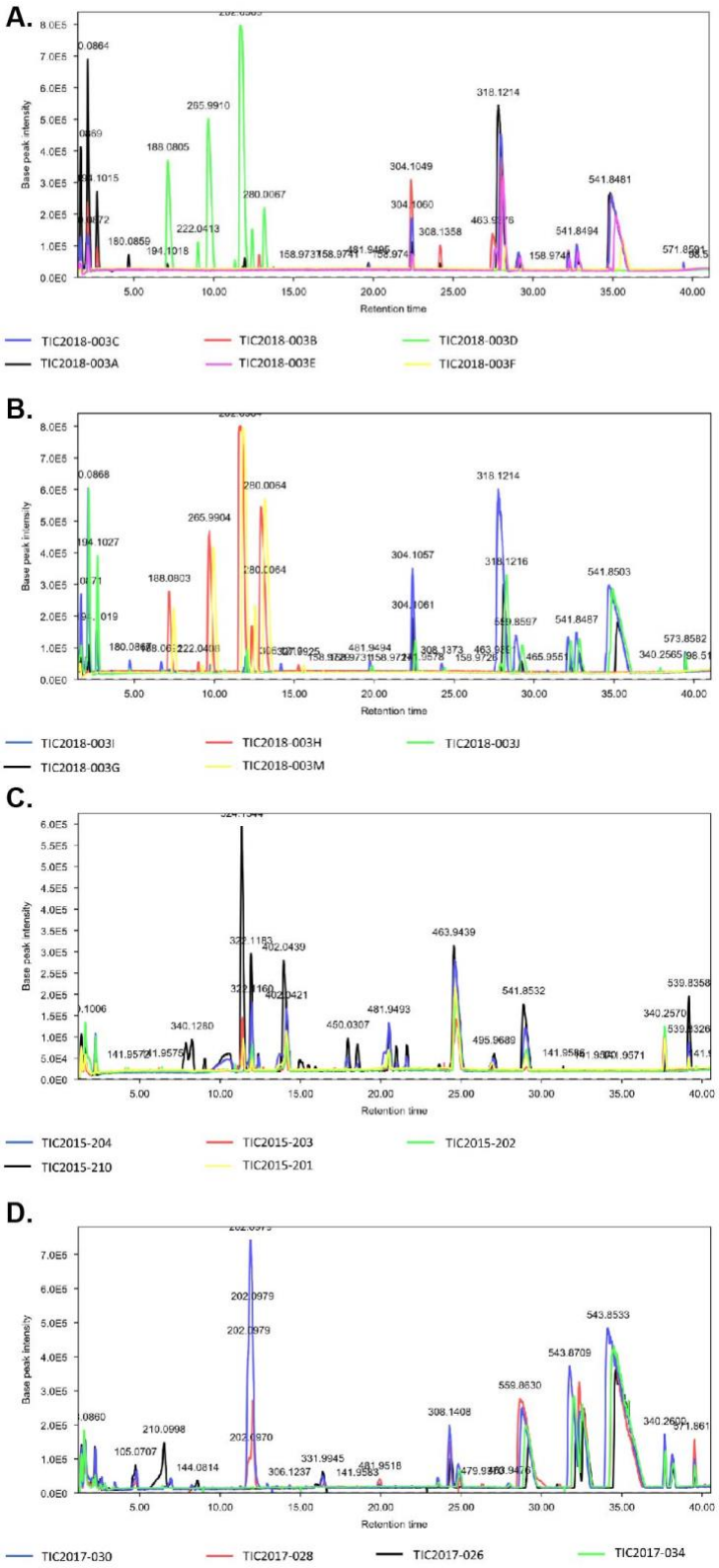

Figure S3. Base peak chromatograms of *Tsitsikamma* sponge chemotypes. Positive mode ESI-LC-MS base peak chromatograms of *T. favus* extracts exhibiting chemotype I or II (**A**-**B**) and *T. michaeli* extracts exhibiting chemotype I or II (**C**-**D**).

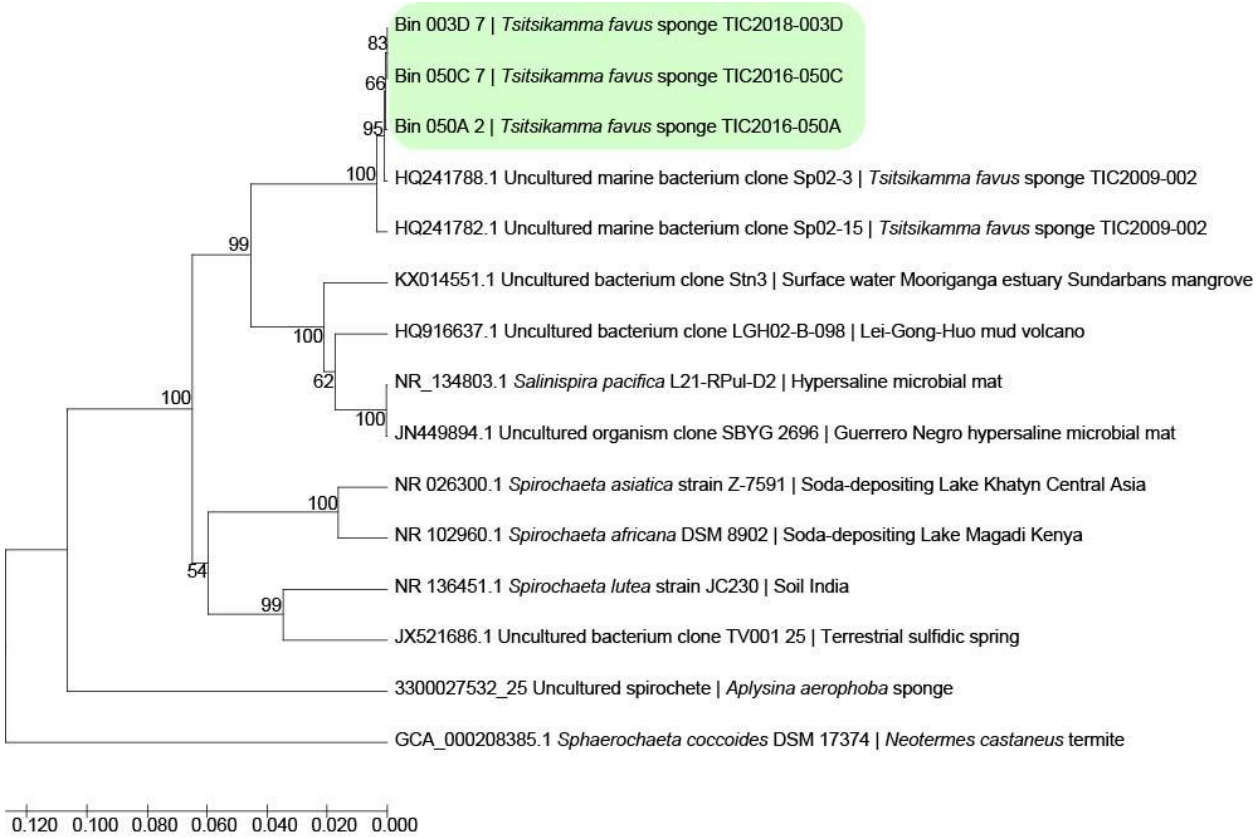

Figure S4. Phylogeny of *T. favus*-associated spirochete bins. The phylogenetic relationship between putative spirochete genome bins and closest relatives was based on 16S rRNA sequences and inferred using the UPGMA method. The percentage of replicate trees in which the associated taxa clustered together 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 12 16S rRNA gene sequences with a total of 1583 positions analyzed. All ambiguous positions were removed for each sequence pair (pairwise deletion option). Evolutionary analyses were conducted in MEGA X.

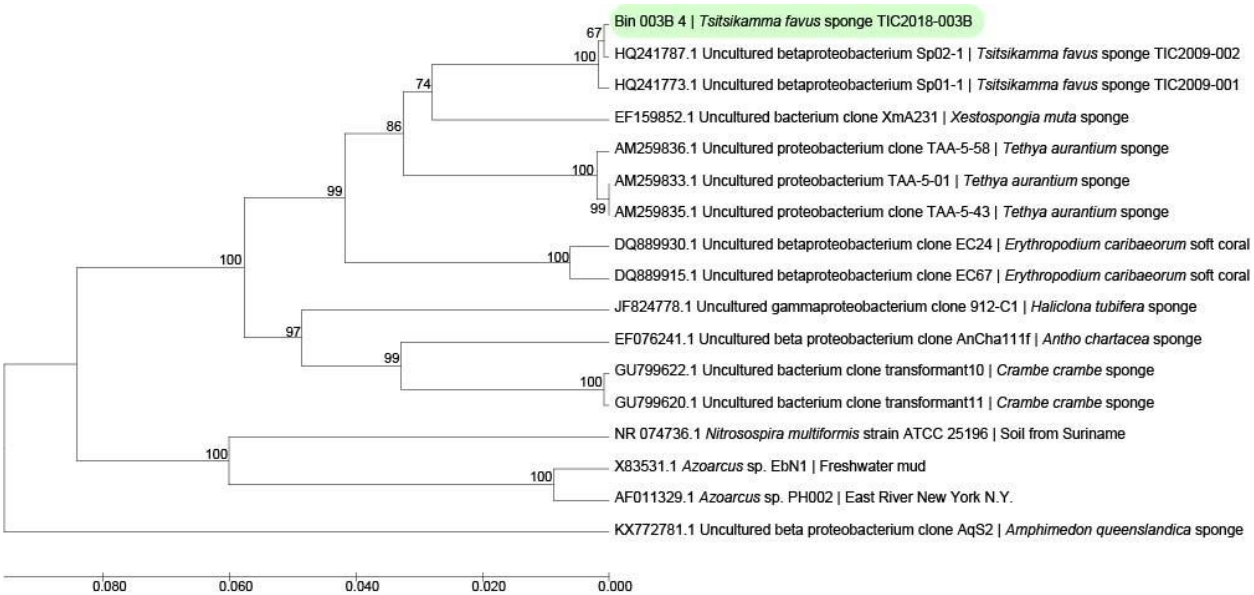

Figure S5. Phylogeny of *T. favus*-associated betaproteobacterium bin. The phylogenetic relationship between the putative betaproteobacteria Sp02-1 genome and closest relatives was based on 16S rRNA sequences and inferred using the UPGMA method. The percentage of replicate trees in which the associated 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.

Table 1. Characteristics of putative representative spirochete genomes of *T. favus* symbiont species

Bin	Size (Mbp)	Complete	Contam	Core genes	Quality	16S rRNA (% ID)	23S rRNA (% ID)	Sponge sample	Chemotype
003B_7	1.97	89.73%	2.67	78.57%	Medium	N/A	*Salinispira pacifica* L21-RPul-D2 (89.54%)	TIC2018-003B	Chemotype I
003D_7	2.48	96.93%	0.8	80.95%	High	Uncultured marine clone Sp02-3 (99.52%)	*Salinispira pacifica* L21-RPul-D2 (89.54%)	TIC2018-003D	Chemotype II
050A_2	2.73	89.69%	12.67	69.05%	Low	Uncultured marine clone Sp02-3 (99.52%)	*Salinispira pacifica* L21-RPul-D2 (89.54%)	TIC2016-050A	Chemotype I
050C_7	1.72	69.32%	1.65	67.86%	Medium	Uncultured marine clone Sp02-3 (99.52%)	N/A	TIC2018-050C	Chemotype II

Table 2. General characteristics of putative Tethybacterales genomes/MAGs

Genome	Sponge host	Size (Mbp)	Completeness	Contam	MIMAG Quality	"Core" genes
003B_4	*Tsitsikamma favus*	2.96	72.92	3.56	Medium	84.52
ImetM1_9	*Iophon methanophila*	1.56	85.58	0.61	Medium	83.33
ImetM2_1_1	*Iophon methanophila*	1.6	84.36	0.61	Medium	86.9
Persebacter sydneyensis (C29)	*Crella incrustans*	1.52	82.87	0.61	Medium	78.57
Tethybacter castellensis (ccb2r)	*Cymbastela concentrica*	1.69	81	0.3	Medium	85.71
Beroebacter blanensis (Crambe1)	*Crambe crambe*	2.25	80.38	1.81	Medium	73.8
Amphirhobacter heronislandensis (AqS2)	*Amphimedon queenslandica*	1.61	71.13	1.52	Medium	80.95
Calypsobacter congwongensis (B3)	*Scopalina sp.*	1.05	64.33	0.04	Medium	53.57
Telestobacter tawharauni (TSB1)	*Tethya stolonifera*	1.24	56.3	0	Medium	57.14
Csing_1	*Coelocarteria singaporensis*	1.58	65.47	0.61	Medium	73.81
Csing_2	*Coelocarteria singaporensis*	1.36	58.12	0	Medium	72.62
Csing_3	*Coelocarteria singaporensis*	1.445416	57.04	0.61	Medium	76.19
Csing_5	*Coelocarteria singaporensis*	1.15745	55.82	0.61	Medium	75.00
Csing_6	*Coelocarteria singaporensis*	1.36604	58.12	0	Medium	72.62
Csing_7	*Coelocarteria singaporensis*	0.981716	54.76	1.93	Medium	69.05
CCyA_2_3	*Cinachyrella sp.*	3.08	84.92	1.83	Medium	78.57
CCyB_3_2	*Cinachyrella sp.*	3.45	79.58	5.93	Medium	75.00
CCyA_16_0	*Cinachyrella sp.*	0.86	16.95	3.81	Low	35.71
CCyA_3_0	*Cinachyrella sp.*	0.62	29.31	0	Low	15.48
CCyA_3_39	*Cinachyrella sp.*	0.41	22.01	4.88	Low	36.90
CCyB_501_11	*Cinachyrella sp.*	0.04	7.51	0	Low	17.86
CCyB_502_83	*Cinachyrella sp.*	0.43	17.01	0	Low	21.43
CCyC_2_7	*Cinachyrella sp.*	1.81	47.5	0	Low	57.14
050C_6	*Tsitsikamma favus* sponge	2.211808	24.03	4.51	Low	32.14
Csing_4	*Coelocarteria singaporensis* sponge	1.079194	39.35	0.61	Low	54.76
050A_14	*Tsitsikamma favus* sponge	6.08233	61	19.4	Low	73.81
003D_6	*Tsitsikamma favus* sponge	0.326243	0	0	Low	0

Table S1. Collection metadata for all latrunculid sponges used in this study

Table S2. Metadata and taxonomic classification of all bins extracted from *T. favus* sponges

Table S3. SRR accessions for all Illumina datasets assembled and binned in this study.

Table S4. Pairwise ANOSIM analysis of OTUs (distance of 0.03) showed that bacterial communities associated with different sponge genera are significantly different ($p < 0.05$).

Table S5. Indicator Species Analysis found that there was a significant ($p < 0.05$) shift in the abundance of some OTU3 and OTU4 between Type I and Type II *T. favus* chemotypes.

Table S5. Shared average nucleotide identity (ANI) between putative Sp02-3 spirochete genome bins

Table S7. Unique genes common to all three putative Sp02-3 spirochete genome bins, relative to all other genes present in the four *T. favus* metagenomes.

Table S8. Genes unique to the putative betaproteobacteria symbiont Bin 003B4, relative to all other genes present in the four *T. favus* metagenomes.

Table S9. Two-way average amino acid identity and 16S rRNA gene sequence identity scores between Tethybacterales genomes

Table S10. Count data of KEGG annotations for all Tethybacterales genomes grouped by primary metabolic pathway.