Families matter: Comparative genomics provides insight into the function of phylogenetically diverse sponge associated bacterial symbionts

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

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

Background. Marine sponges play an important role in maintaining marine ecosystem health. As the oldest extant metazoans, sponges have been forming symbiotic relationships with microbes that may date back as far as 700 million years. These symbioses can be mutual or commensal in nature, and often allow the holobiont to survive in a given ecological niche. Most symbionts are conserved within a narrow host range and perform specialized functions. However, there are ubiquitously distributed bacterial taxa such as Poribacteria, SAUL and Tethybacterales that are found in a broad range of invertebrate hosts. What is not currently clear, is whether these ubiquitously distributed, sponge associated symbionts have evolutionarily converged to fulfil similar roles within their sponge host, or if they represent a more ancient lineage of specialist symbionts. Here, we aimed to determine the roles of codominant Tethybacterales and spirochete symbionts in Latrunculid sponges and use this system as a starting point to contrast the functional potential of ubiquitous and specialized symbionts, respectively.

Results. This study shows that the dominant betaproteobacterial symbiont Sp02-1 conserved in latrunculid sponges belongs to the newly proposed Tethybacterales order. Extraction of 11 additional Tethybacterales metagenome-assembled genomes (MAGs) revealed the presence of a third, previously unknown family within the Tethybacterales order. Comparative genomics revealed that the Tethybacterales taxonomic families dictate the functional potential, rather than adaptation to their host, and that ubiquitous sponge associated bacteria (Tethybacterales and *Entoporibacteria*) likely perform distinct functions within their host. The divergence of the Tethybacterales and *Entoporibacteria* is incongruent with their host phylogeny, which suggests that ancestors of these bacteria underwent multiple association events, rather than a single

association event followed by co-evolution. Finally, we investigated the potential role of the unique, dominant spirochete Sp02-3 symbiont in latrunculid sponges and found that it is likely a specialized nutritional symbiont that possibly provides active forms of vitamin B6 to the holobiont and may play a role in the host sponge chemistry.

Conclusions. This study shows that functional potential of Tethybacterales is dependent on their taxonomic families and that a similar trend may be true of the *Entoporibacteria*. We show that ubiquitous sponge-associated bacteria likely do not perform similar functions in their hosts, and that the Tethybacterales may be a more ancient lineage of sponge symbionts than the *Entoporibacteria*. In the case of the unusual dominant Spirochete symbiont, the abundance of which appears to be dependent on the sponge host family, evidence provided here suggests that it may play a role as a specialized nutritional symbiont.

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

Background

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], sulphur [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 classification 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 the broad-host range symbionts arose from a single association event, or multiple serendipitous events. Furthermore, if the latter were true, do 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 relationship between phylogenetically diverse co-dominant, conserved Tethybacterales and spirochetes symbionts of *Tsitsikamma favus* sponge species (Family Latrunculiidae)[53–56] as a springboard into a deeper investigation into the role of these two types of symbionts. 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.

METHODS AND MATERIALS

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. In addition, three *L. apicalis* specimens were collected by trawl net 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 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). Finally, genomes of symbiont and free-living species *Sphaerochaeta coccoides* DSM 17374 (GCA_000208385.1) (termite-associated), *Salinispira pacifica* (GCA_000507245.1),

Spirochaeta africana (CP003282.1), *Spirochaeta thermophila* DSM 6192 (CP001698.1) and *Spirochaeta perfilievii* strain P (CP035807.1) (outgroup) were downloaded from the NCBI database. Additional MAGs, representing spirochetes associated with different invertebrate hosts were retrieved from the JGI database (3300004122_5, 3300005978_6, 3300006913_10, 3300010314_9 and 3300027532_25). These reference genomes were chosen because they were either a identified relative of the *T. favus* associated spirochete (e.g. *S. pacifica*) or they are a known, characterized spirochete symbiont (e.g. *S. coccoides).* All genomes were assessed using CheckM[62] and bins were taxonomically classified with GTDB-Tk[65] (Table S2).

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]. 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[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 003D_7 and 003B_4 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 host-associated and unique genes in putative symbiont genome bins. "Host associated" genes were defined as genes that were detected exclusively in symbiont genomes and absent in free-living relatives of the symbionts. The reasoning being that genes that are exclusively present in symbionts may play a role in the symbiosis, as these gene functions would not be required in a free-living relative. To identify "host-associated" genes, orthologous groups of putative Sp02-3 spirochete bins and reference genomes were found using OMA (v.2.4.1)[74]. 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-Sp02-1 *T. favus*-associated bacterial bins was created using the "makedb" option in Diamond[80] to identify genes that were unique to the Sp02-1 spirochete 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 genomes and all other *T. favus* associated genome bins and excluded any Sp0213 genes that were found to have the same annotation as a gene in one of the other bins. Unique spirochete Sp02-3 (putative spirochete symbiont) genes were identified as described for the Sp02-1 genes.

16S rRNA gene amplicon sequencing and analysis. 16S rRNA gene amplicons were prepared and analyzed as described in Kalinski et al., 2019[81]. 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 (See https://github.com/samche42/Family_matters for all scripts).

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 Ultimate 3000 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)[81]. The Principal Component Analysis was performed as part of a molecular networking job on the GNPS platform[82]. Visualization of the base peak chromatograms was done using MZMine2 (v. 2.5.3)[83].

RESULTS AND DISCUSSION

The microbiomes of sponges of the Latrunculiidae family are highly conserved and are dominated by populations of co-dominant betaproteobacteria and spirochetes symbionts. Based on their 16S rRNA gene sequence, the betaproteobacteria symbionts from different latrunculid sponges are closely related and are likely members of the newly described Tethybacterales. The spirochete symbiont has thus far only been found in *Tsitsikamma* and *Cyclacanthia* sponge species and unlike the betaproteobacterium symbiont, they are only distantly related to other sponge-associated spirochetes[53]. In this study we focussed on the microbiome of *T. favus*, which is endemic to Algoa Bay and has been well-characterized with respect to pyrroloiminoquinone production[55,56,81,84,85].

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 betaproteobacterium 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 betaproteobacterium 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 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[86]. An abundance of pseudogenes

and low coding density, is usually an indication that the genome in question may be undergoing genome reduction [87] similar to other betaproteobacteria in the proposed order of Tethybacterales[42].

The 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[8,88,89].

A total of 13 genes unique to the Sp02-1 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 Sp02-1 genome can acquire proline and convert it to glutamate[90] 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[91]. If

functionally active in the Sp02-1 Tethybacterales 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[92,93] and if active within the symbiont, they may help facilitate the symbiosis between the Sp02-1 betaproteobacterium and the sponge host.

Comparison of putative Sp02-1 with 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[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 these 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 2784132075
Tethybacter castellensis (ccb2r)	*Cymbastela concentrica*	1.69	81	0.3	Medium	85.71	Taylor et al., 2021 2784132054
Beroebacter blanensis (Crambe1)	*Crambe crambe*	2.25	80.38	1.81	Medium	73.8	Taylor et al., 2021 2814122900
Amphirhobacter heronislandensis (AqS2)	*Amphimedon queenslandica*	1.61	71.13	1.52	Medium	80.95	Gauthier et al., 2016 GCA_001750625.1
Calypsobacter congwongensis (B3)	*Scopalina sp.*	1.05	64.33	0.04	Medium	53.57	Taylor et al., 2021 2784132034
Telestobacter tawharauni (TSB1)	*Tethya stolonifera*	1.24	56.3	0	Medium	57.14	Taylor et al., 2021 2784132053
"*Candidatus* Hadiah malacca" (Csing_1)	*Coelocarteria singaporensis*	1.58	65.47	0.61	Medium	73.81	3300007741_3

"*Candidatus* Hadiah malacca" (Csing_2)	*Coelocarteria singaporensis*	1.36	58.12	0	Medium	72.62	3300007056_3
"*Candidatus* Hadiah malacca" (Csing_3)	*Coelocarteria singaporensis*	1.44	57.04	0.61	Medium	76.19	3300007046_3
"*Candidatus* Hadiah malacca" (Csing_5)	*Coelocarteria singaporensis*	1.15	55.82	0.61	Medium	75.00	3300021544_3
"*Candidatus* Hadiah malacca" (Csing_6)	*Coelocarteria singaporensis*	1.36	58.12	0	Medium	72.62	3300021545_3
"*Candidatus* Hadiah malacca" (Csing_7)	*Coelocarteria singaporensis*	0.98	54.76	1.93	Medium	69.05	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	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, 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[44,53,54,94]. 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)[95]. 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 [95]. 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.

We identified 4306 groups of orthologous genes between all seventeen Tethybacterales genomes, with only eighteen genes common to all the genomes. More common 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[96].

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[97–101]. Internal changes in oxygen levels[102], and temperature changes[103–105] 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.

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 are exclusively capable of transporting hydroxyproline which may imply a role in collagen degradation[106]. 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[107]. 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[108] and may provide the Tethybacterales with a competitive advantage over other sympatric microorganisms[109] and possibly allow the sponge hosts to regulate the symbioses via regulation of the quantity of amino acids available for symbiont uptake[110]. 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[111]. 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[103,112–114] 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 ubiquitous 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, sulphur metabolism, phosphate metabolism and several transporter systems (Figure 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), 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). 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. 3B), 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 which provides further evidence that these symbionts may serve different purposes within their sponge host.

We investigated the approximate divergence pattern of the Tethybacterales and the *Entoporibacteria* and whether their divergence followed that of their sponge host. 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[115]. We found that the estimated divergence pattern (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 or Poribacteria common ancestor from the environment at different time points throughout their evolution. Evidence of coevolution of betaproteobacteria symbionts within sponges families [46,52,53,116] 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). This earlier divergence of sponge-associated Tethybacterales relative to the *Entoporibacteria* may suggest that the Tethybacterales may have begun associating with sponge hosts before the *Entoporibacteria*, and may 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.

Characterization of putative spirochete Sp02-3 genome bin

Spirochetes have been noted as minor members of several sponge microbiomes [2,117,118], but have only been reported as dominant symbionts in several latrunculid sponge species[53,54] and *Clathrina clathrus[119]*. As there were no genomes available for these symbionts, their function was unknown. Spirochaetes are known to be dominant in bacterial communities associated with

several corals[120–122], where they may be involved in the fixation of carbon or nitrogen, as performed by spirochaete symbionts in termite guts[120,122,123].

Following assembly and binning of four *T. favus* metagenomic datasets, 16S rRNA and 23S rRNA gene sequences were extracted from individual bins and used to identify four genomes 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 3). Sp02-3 is one of the spirochetes previously identified in the *T. favus* microbiome[54].

Table 3. Characteristics of putative representative genomes of *T. favus* spirochete symbiont species

Bin	Size (Mbp)	Quality	16S rRNA (% ID)	23S rRNA (% ID)	Chemotype
003B_7	1.97	Medium	N/A	*Salinispira pacifica* L21-RPul-D2 (89.54%)	Chemotype I
003D_7	2.48	High	Uncultured marine clone Sp02-3 (99.52%)	*Salinispira pacifica* L21-RPul-D2 (89.54%)	Chemotype II
050A_2	2.73	Low	Uncultured marine clone Sp02-3 (99.52%)	*Salinispira pacifica* L21-RPul-D2 (89.54%)	Chemotype I
050C_7	1.72	Medium	Uncultured marine clone Sp02-3 (99.52%)	N/A	Chemotype II

Salinispira pacifica, previously identified as a close relative of Sp02-3 [6], was the closest characterized relative to the putative Sp02-3 genomes (Table 3). In both placement and *de novo*

mode, autoMLST assigned *S. pacifica* L21-RPul-D2 as the closest relative to the putative Sp02-3 genomes (67.1% ANI) (Fig. 5) and GTDB-Tk placed the bin within the *Salinispira* genus, confirming the taxonomic classification based on the 16S rRNA gene sequence. Validation using CheckM showed that genome bin 003D_7 was of the highest quality (Table 3, Table S2), and it was therefore used in subsequent analyses 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 genome coding density is 77.05%, which is lower than the average bacterial coding density of 85 - 90%[86]. Finally, the identification of 81% of essential genes[73] suggests that this genome is likely complete and may be reduced, however further study is required to confirm this. Given the divergence of the putative Sp02-3 genomes (Fig. 5) and low (<94%) shared identity[95] (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 rRNA gene (1524 bp), full-length 23S rRNA gene (2979 bp), two copies of the 5S rRNA genes (111 bp each), as well as 40 tRNAs. We propose the name "*Candidatus* Latrunculis umkhuseli" which acknowledges both the sponge family with which it is associated (Latrunculiidae) and "umkhuseli", which means "protector" in isiXhosa.

Annotation of the genes in the *Ca.* L. umkhuseli (Sp02-3) 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 present 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, *atpEIK* and *ntpABDE* were detected in representative Sp02-3 genome bins 003D_7, 003B_7 and 050C_7 and may serve either as a proton pump, or for the anaerobic production of ATP, as proposed for both *S. pacifica* and *Thermus thermophilus[124]*. The putative *Ca.* L. umkhuseli 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. umkhuseli uses sulfur for anaerobic respiration. It is possible that the produced sulfite may be exported and potentially used by other symbiotic bacteria, such as the conserved Tethybacterales. The *Ca.* L. umkhuseli genome carries a single putative terpene BGC that has a homologous gene cluster in present in *S. pacifica* L21-RPul-D2 (38 - 59% amino acid identity between genes). This gene cluster in *S. pacifica* L21-RPul-D2 is predicted to produce a red-orange pigment hypothesized to protect the bacterium against oxidative stress[124]. If a similar terpene is expressed in the *Ca.* L. umkhuseli genome, the compound may perform a similar protective function for *Ca.* L. umkhuseli, which appears to be a facultative anaerobe.

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

We identified a total of 23 genes that were functionally unique to the *Ca.* L. umkhuseli genomes, relative to the other bacteria of the *T. favus* metagenomes. Nine of these were common to all three *Ca.* L. umkhuseli genomes (Table S10). Eight of these unique genes encode proteins potentially involved in the transport and metabolism of sugars. As mentioned previously, sugar transport/uptake plays an important role in symbiotic interactions[103,112–114], therefore, the presence of these genes would suggest that sugars, as with the Tethybacterales symbionts, are

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 [125]), but no core genes involved in nitrogen metabolism were found in the *Ca.* L. umkhuseli spirochete genomes.

The putative *Ca.* L. umkhuseli symbionts also possess additional unique genes (i.e. not identified in the entire metagenome, including unbinned contigs) that are predicted to encode an unspecified methyltransferase, a tetratricopeptide repeat protein and *N*-acetylcysteine and pyridoxal 5'-phosphate synthases PdxS and PdxT. Pyridoxal 5′-phosphate, the only active form of Vitamin B6, can be produced *de novo* via two possible pathways: 1) The DXP-dependent pathway, which requires 5 enzymes (*pdxA*, *pdxB*, *serC*, *pdxJ*, or *pdxH* genes[126,127]) or 2) The DXP-independent pathway, which requires only two enzymes, PdxS and PdxT, that produce pyridoxal 5′-phosphate (active vitamin B6) from a pentose and a triose in the presence of glutamine[127]. This would imply, in the context of the *T. favus* metagenome, that the spirochete may be the only symbiont capable of producing the active form of vitamin B6, as genes for either pathway could not be detected in the rest of the metagenome (i.e. all assembled contigs, binned and unbinned, were checked). Vitamin B6 acts as a cofactor to a wide variety of enzymes across all kingdoms of life [126,128], and is an essential vitamin of marine sponges [20] and may potentially be used by spirochete symbionts as a competitive advantage in the sponge holobiont [129]. However, whether the vitamin is used by the host itself or sympatric microorganisms is not known at this point.

Comparison of putative *Ca.* L. umkhuseli 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 the *Ca.* L. umkhuseli genome bins (Table S2). We identified a total of 4037 orthologous gene groups (OGs) between the 14 genomes/bins and the hierarchical clustering of gene presence/absence showed that the *Ca.* L. umkhuseli symbionts did not cluster with *S. pacifica* and other free-living spirochetes (Fig. 6A). However, the *Ca.* L. umkhuseli bins shared the greatest abundance of orthologous genes with free-living spirochete *S. pacifica* (~21% average shared OGs) (Fig. 6B) and less than 14% with any of the host-associated spirochetes which would confirm *S. pacifica* as the closest relative and suggests that the difference in gene repertoire (Fig. 6A) may be due to gene loss in the *Ca.* L. umkhuseli genomes.

Only six genes were shared between the *Ca.* L. umkhuseli genomes and other 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[130]. The role of fucosylated glycans in the mammalian gut as points of adherence for various bacteria has been well documented[131–134] and it is possible that *Ca.* L. umkhuseli spirochetes are similarly capable of binding fucose in their respective hosts. Alternatively, *Ca.* L. umkhuseli spirochetes

may simply degrade the fucose of algal and bacterial cells that have been ingested by the sponge host, as is hypothesized for other sponge symbionts[135]. The remaining genes are likely involved in the uptake and utilization of sugars.

Conservation of spirochete symbiont populations in Latrunculiidae sponges

Previous studies identified two closely related spirochete species, Sp02-3 and Sp02-15, in the *T. favus* microbiome[54]. Subsequently, the Sp02-3 symbiont was identified as consistently conserved in the microbiomes of additional *Tsitsikamma* and *Cyclacanthia bellae* sponges species[53]. We aimed to expand these studies by investigating the microbial communities of 61 sponge specimens, representing seven genera within the Latrunculiidae family (Table S1). Using 16S rRNA gene amplicon sequencing we profiled the microbiomes of these sponges relative to the surrounding seawater and, in agreement with previous studies[53,54], found that different latrunculid species are associated with distinct bacterial communities which differ significantly from the surrounding water (Table S11). Each of the sponge species is associated with a distinct microbial community, where each sponge microbiome is dominated by a betaproteobacterium (Fig. S5). *T. favus*, *T. nguni* and *T. pedunculata* are all dominated by OTU1 and a closely-related betaproteobacterium, OTU2, which dominates *T. michaeli* and *C. bellae* sponges. The *L. algoaensis* sponge and *L. apicalis* sponges harbor their own dominant, and closely related betaproteobacteria (OTU21 and OTU7 respectively).

Six OTUs were present in all sixty-one sponge specimens (Fig. 7A), three of which were closely related to the dominant betaproteobacterium Sp02-1 and two shared the greatest sequence identity to spirochete species Sp02-3 and Sp02-15, all of which had been previously identified in

T. favus[53,54]. 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. OTU17, is most closely related to uncultured bacteria isolated from seawater collected from the Santa Barbara channel (MG013048.1) and the North Yellow sea (FJ545595.1) and interestingly this OTU was also present in multiple samples of seawater collected over several years in Algoa Bay (Fig. S5). This OTU may represent a ubiquitous marine bacterium that may be present as a food bacterium.

In previous studies[56,81], we profiled the compounds present in the crude chemical extracts from these latrunculid sponge specimens and obtained distinct chemical profiles for each species (Fig. 7B and Fig. S6). In addition, it was found that *T. favus* and *T. michaeli* sponges exhibited two distinct chemotypes (Fig. 7B). For *T. favus*, Chemotype I, containing discorhabdins and tsitsikammamines, is considered “normal” and is characteristic of these sponges dating back to collections made since 1996[136]. The second chemotype has only very recently been discovered (2017) and was found to contain, almost exclusively, simple pyrroloiminoquinones such as makaluvamines. Regarding *T. michaeli*, both chemotypes appear to predominantly contain discorhabdins, with the chemical profile of chemotype I being closely related to *T. favus* chemotype I, whereas *T. michaeli* chemotype II is a source of new, yet unknown discorhabdins.

We performed an Indicator Species analysis (Fig. 7C), which showed that there was a statistically significant shift in the abundance of several OTUs between chemotypes (Table S12). Most notably, there was a significant increase in the abundance of OTU5 (Sp02-15 spirochete) in *T. favus* chemotype II (p = 0.005, R = 0.51) and *T. michaeli* chemotype II (p = 0.03, R = 0.87)

(Fig. 7C). Chemotype II is characterized by the presence of makaluvamines, which are the simplest form of pyrroloiminoquinone compounds and are hypothesized to be the starter compounds for more complex pyrroloiminoquinones (e.g., tsitsikammamines and discorhabdins). Although in no way conclusive evidence, the significant increased abundance of Sp02-15, suggests that it may play a role in preventing the formation of more complex pyrroloiminoquinone compounds. Extracting the genome of the Sp02-15 spirochete is the subject of ongoing studies in an effort to determine whether this symbiont plays a significant role in latrunculid chemotypes.

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, such as the spirochete, potential function is dictated by the sponge host. This work has expanded our understanding of the Tethybacterales and the possible functional specialization of the families within this new order. Here we found that Tethybacterale symbiont family rank is a determining factor of it's functional potential, rather than functional radiation in response to its sponge host. 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. 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

begun diverging 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.

Finally, the conserved spirochete Sp02-3 (*Ca.* L. umkhuseli) may be the exclusive producer of active vitamin B6 in the sponge metagenome, and therefore it is possible that this bacterium is consistently present within the *T. favus* sponges as a nutritional symbiont through the provision of this essential vitamin. Conversely, the closely related spirochete Sp02-15 may play a role in the production of complex cytotoxic pyrroloiminoquinone alkaloids that are characteristic of these sponges.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All 16S rRNA gene amplicon sequence data, sponge 28S rRNA gene sequences and raw shotgun metagenomic read data can be accessed from the NCBI website under BioProject PRJNA508092.

Competing Interests

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

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. 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.

Author Contributions

S.C.W and R.A.D. drafted the manuscript, which was substantively revised by all other co-authors. L.M. generated 16S amplicon data and J-C.J.K. generated and analyzed all chemical data. S.P.-N taxonomically identified all sponges using both gross morphology and spicule analysis. S.C.W. generated 16S rRNA gene amplicon data, 28S rRNA gene sequence data, all shotgun metagenomic data and was responsible for the analyses thereof. J.C.K aided in bioinformatic analysis.

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 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.

REFERENCES

1. Wilkins LGE, Leray M, O'Dea A, Yuen B, Peixoto RS, Pereira TJ, et al. Host-associated microbiomes drive structure and function of marine ecosystems. PLoS Biol. 2019;17:e3000533.

2. Taylor MW, Radax R, Steger D, Wagner M. Sponge-associated microorganisms: Evolution, ecology, and

biotechnological potential. Microbiol Mol Biol Rev. 2007;71:295–347.

3. Webster NS, Taylor MW. Marine sponges and their microbial symbionts: Love and other relationships. Environ Microbiol. 2012;14:335–46.

4. Zhang F, Jonas L, Lin H, Hill RT. Microbially mediated nutrient cycles in marine sponges. FEMS Microbiol Ecol [Internet]. 2019;95. Available from: http://dx.doi.org/10.1093/femsec/fiz155

5. Karimi E, Slaby BM, Soares AR, Blom J, Hentschel U, Costa R. Metagenomic binning reveals versatile nutrient cycling and distinct adaptive features in alphaproteobacterial symbionts of marine sponges. FEMS Microbiol Ecol [Internet]. 2018;94. Available from: http://dx.doi.org/10.1093/femsec/fiy074

6. Hoffmann F, Radax R, Woebken D, Holtappels M, Lavik G, Rapp HT, et al. Complex nitrogen cycling in the sponge *Geodia barretti*. Environ Microbiol. 2009;11:2228–43.

7. Bayer K, Schmitt S, Hentschel U. Microbial nitrification in Mediterranean sponges: possible involvement of ammonia-oxidizing betaproteobacteria. In: Custódio MR, Lôbo-Hajdu G, Hajdu E, Muricy G, editors. *Porifera research: biodiversity, innovation, sustainability*. Rio de Janeiro: Série Livros, Museu Naciona; 2007. pp. 165–171.

8. 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;11:1651–66.

9. Bayer K, Schmitt S, Hentschel U. Physiology, phylogeny and in situ evidence for bacterial and archaeal nitrifiers in the marine sponge *Aplysina aerophoba*. Environ Microbiol. 2008;10:2942–55.

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

11. Jensen S, Fortunato SAV, Hoffmann F, Hoem S, Rapp HT, Øvreås L, et al. The relative abundance and transcriptional activity of marine sponge-associated microorganisms emphasizing groups involved in sulfur cycle. Microb Ecol. 2017;73:668–76.

12. 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.

13. Zhang F, Blasiak LC, Karolin JO, Powell RJ, Geddes CD, Hill RT. Phosphorus sequestration in the form of polyphosphate by microbial symbionts in marine sponges. Proc Natl Acad Sci U S A. 2015;112:4381–6.

14. Colman AS. Sponge symbionts and the marine P cycle. Proc. Natl. Acad. Sci. U. S. A. 2015. p. 4191–2.

15. Wilkinson CR. Net primary productivity in coral reef sponges. Science. 1983;219:410–2.

16. Feng G, Li Z. Carbon and nitrogen metabolism of sponge microbiome. In: Li Z, editor. Symbiotic microbiomes of coral reefs sponges and corals. Dordrecht: Springer Netherlands; 2019. p. 145–69.

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

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

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

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

21. Song H, Hewitt OH, Degnan SM. Bacterial symbionts in animal development: Arginine biosynthesis complementation enables larval settlement in a marine sponge. Curr Biol. 2021 Jan 25;31(2):433-437.e3.

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

23. Lopanik NB. Chemical defensive symbioses in the marine environment. Funct Ecol. Wiley; 2014;28:328–40.

24. Mori T, Cahn JKB, Wilson MC, Meoded RA, Wiebach V, Martinez AFC, et al. Single-bacterial genomics validates rich and varied specialized metabolism of uncultivated *Entotheonella* sponge symbionts. Proc Natl Acad Sci U S A. 2018;115:1718–23.

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

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

27. Degnan SM. The surprisingly complex immune gene repertoire of a simple sponge, exemplified by the NLR genes: A capacity for specificity? Dev Comp Immunol. 2015;48:269–74.

28. Jahn MT, Arkhipova K, Markert SM, Stigloher C, Lachnit T, Pita L, et al. A phage protein aids bacterial symbionts in eukaryote immune evasion. Cell Host Microbe. 2019;26:542–50.e5.

29. Thomas T, Moitinho-Silva L, Lurgi M, Björk JR, Easson C, Astudillo-García C, et al. Diversity, structure and convergent evolution of the global sponge microbiome. Nat Commun. 2016;7:11870.

30. Astudillo-García C, Slaby BM, Waite DW, Bayer K, Hentschel U, Taylor MW. Phylogeny and genomics of SAUL, an enigmatic bacterial lineage frequently associated with marine sponges. Environ Microbiol. 2018;20:561–76.

31. Fieseler L, Horn M, Wagner M, Hentschel U. Discovery of the novel candidate phylum “Poribacteria” in marine sponges. Appl Environ Microbiol. 2004;70:3724–32.

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

33. 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;13:468–81.

34. Lafi FF, Fuerst JA, Fieseler L, Engels C, Goh WWL, Hentschel U. Widespread distribution of poribacteria in demospongiae. Appl Environ Microbiol. 2009;75:5695–9.

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

36. 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.

37. 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.

38. 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;7:2287–300.

39. 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.

40. Hill MS, Sacristán-Soriano O. Molecular and functional ecology of sponges and their microbial symbionts. In: Carballo JL, Bell JJ, editors. Climate change, ocean acidification and sponges: Impacts across multiple levels of organization. Cham: Springer International Publishing; 2017. p. 105–42.

41. Chen ML, Becraft ED, Pachiadaki M, Brown JM, Jarett JK, Gasol JM, et al. Hiding in plain sight: The globally distributed bacterial candidate phylum PAUC34f. Front Microbiol. 2020;11:376.

42. 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 2021, 15, 503–519

43. 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.

44. 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.

45. 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.

46. 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(4):95.

47. 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;88:516–27.

48. 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.

49. Webster NS, Wilson KJ, Blackall LL, Hill RT. Phylogenetic diversity of bacteria associated with the marine sponge *Rhopaloeides odorabile.* Appl Environ Microbiol. 2001;67:434–44.

50. 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;85:465–82.

51. 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.

52. 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.

53. 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. 2017;6(2):e00417.

54. 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;14:681–91.

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

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

57. Samaai T, Kelly M. Family Latrunculiidae Topsent, 1922. In: Hooper JNA, Van Soest RWM, Willenz P, editors. Systema Porifera: A Guide to the Classification of Sponges. Boston, MA: Springer US; 2002. p. 708–19.

58. 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;874:101–26.

59. Hamlyn-Harris R, Queensland Museum, Hamlyn-Harris R. Memoirs of the Queensland Museum. Queensland Museum,;40 (1996). Available from: https://www.biodiversitylibrary.org/item/123914

60. 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;19:455–77.

61. 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.

62. 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;25:1043–55.

63. 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;35:725–31.

64. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.

65. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: A toolkit to classify genomes with the Genome

Taxonomy Database. Bioinformatics. 2020;36;6;1925–1927.

66. 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.

67. 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:W276–82.

68. Seemann T. Prokka: Rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.

69. 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. Bioinformatics. 2020;36(7):2251-2252

70. 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:W81–7.

71. Rodriguez-R LM, Konstantinidis KT. The enveomics collection: A toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Preprints. 2016;4:e1900v1

72. Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci. Wiley; 2003;14:927–30.

73. 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(1):e02430-19.

74. 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;29:1152–63.

75. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.

76. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 2000;16:276–7.

77. Suyama M, Torrents D, Bork P. PAL2NAL: Robust conversion of protein sequence alignments into the

corresponding codon alignments. Nucleic Acids Res. 2006;34:W609–12.

78. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol. 2018;35:1547–9.

79. Sneath PHA, Sokal RR, Others. Numerical taxonomy. The principles and practice of numerical classification. Systematic Zoology. 1975;24;2:263-268

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

81. 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;1:60.

82. 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;34:828–37.

83. 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.

84. Davies-Coleman MT, Antunes EM, Beukes DR, Samaai T. The colourful chemistry of South African latrunculid sponges. S Afr J Sci. 2019;115:1-7.

85. Antunes EM, Beukes DR, Kelly M, Samaai T, Barrows LR, Marshall KM, et al. Cytotoxic pyrroloiminoquinones from four new species of South African latrunculid sponges. J Nat Prod. 2004;67:1268–76.

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

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

88. 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;116:16003–11.

89. 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;22:483–98.

90. 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;292:16360–7.

91. 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;1844:505–11.

92. 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;196:2775–88.

93. Al-Khodor S, Price CT, Kalia A, Abu Kwaik Y. Functional diversity of ankyrin repeats in microbial proteins. Trends Microbiol. 2010;18:132–9.

94. 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;9:1262.

95. Konstantinidis KT, Rosselló-Móra R, Amann R. Uncultivated microbes in need of their own taxonomy. ISME J. 2017;11:2399–406.

96. 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;55:149–58.

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

98. Gottesman S, Wickner S, Maurizi MR. Protein quality control: Triage by chaperones and proteases. Genes Dev. 1997;11:815–23.

99. 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;73:7803–13.

100. 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;148:2727–33.

101. 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;528:409–12.

102. 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.

103. 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.

104. 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;14:3232–46.

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

106. 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;6:152.

107. 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;67:520–39.

108. 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.

109. 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;184:4071–80.

110. 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;106:12477–82.

111. Hogle SL, Barbeau KA, Gledhill M. Heme in the marine environment: From cells to the iron cycle. Metallomics. 2014;6:1107–20.

112. 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;161:1984–92.

113. 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.

114. 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;14;10:2554-2567

115. 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.

116. 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;94:5.

117. 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;78:243–56.

118. 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;11:758–70.

119. 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;48:411–8.

120. 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.

121. 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(5)

122. 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.

123. Lilburn TG, Kim KS, Ostrom NE, Byzek KR, Leadbetter JR, Breznak JA. Nitrogen fixation by symbiotic and free-living spirochetes. Science. 2001;292:2495–8.

124. 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 from a hypersaline microbial mat. Stand Genomic Sci. 2015;10:7.

125. Jeong HS, Jouanneau Y. Enhanced nitrogenase activity in strains of *Rhodobacter capsulatus* that overexpress the rnf genes. J Bacteriol. 2000;182:1208–14.

126. 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.

127. El Qaidi S, Yang J, Zhang JR, Metzger DW. The vitamin B6 biosynthesis pathway in *Streptococcus pneumoniae* is controlled by pyridoxal 5′-phosphate and the transcription factor PdxR and has an impact on ear infection. J Bacteriol. 2013;195;10:2187-96.

128. 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;23:1060–76.

129. 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.

130. 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;67:1524–30.

131. Pickard JM, Chervonsky AV. Intestinal fucose as a mediator of host-microbe symbiosis. J Immunol. 2015;194:5588–93.

132. Becerra JE, Yebra MJ, Monedero V. An L-fucose operon in the probiotic *Lactobacillus rhamnosu*s GG is involved in adaptation to gastrointestinal conditions. Appl Environ Microbiol. 2015;81:3880–8.

133. 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;384:307–16.

134. Coyne MJ, Reinap B, Lee MM, Comstock LE. Human symbionts use a host-like pathway for surface fucosylation. Science. 2005;307:1778–81.

135. Sizikov S, Burgsdorf I, Handley KM, Lahyani M, Haber M, Steindler L. Characterization of sponge-associated *Verrucomicrobia*: Microcompartment-based sugar utilization and enhanced toxin--antitoxin modules as features of host-associated Opitutales. Environ Microbiol. Wiley Online Library; 2020;22:4669–88.

136. Hooper GJ, Davies-Coleman MT, Kelly-Borges M, Coetzee PS. New alkaloids from a South African latrunculid sponge. Tetrahedron Lett. 1996;37:7135–8.

FIGURE LEGENDS

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.

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).

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).

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.

Figure 5. 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 6. 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 7. 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 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.

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. 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 S6. 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).

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

Table S2. Metadata and taxonomic classification of all bins/genomes used in this study.

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

Table S4. Accessions and metadata for all Poribacteria genomes analyzed in this study.

Table S5. Pairwise synonymous substitution rates (dS) between genomes within the sponge-associated Tethybacterales and *Entoporibacteria*, respectively.

Table S6. Genes unique to the putative betaproteobacteria symbiont Bin 003B4, relative to all other genes present in the four *T. favus* metagenomes.

Table S7. Pairwise average amino acid identity scores between Tethybacterales genomes

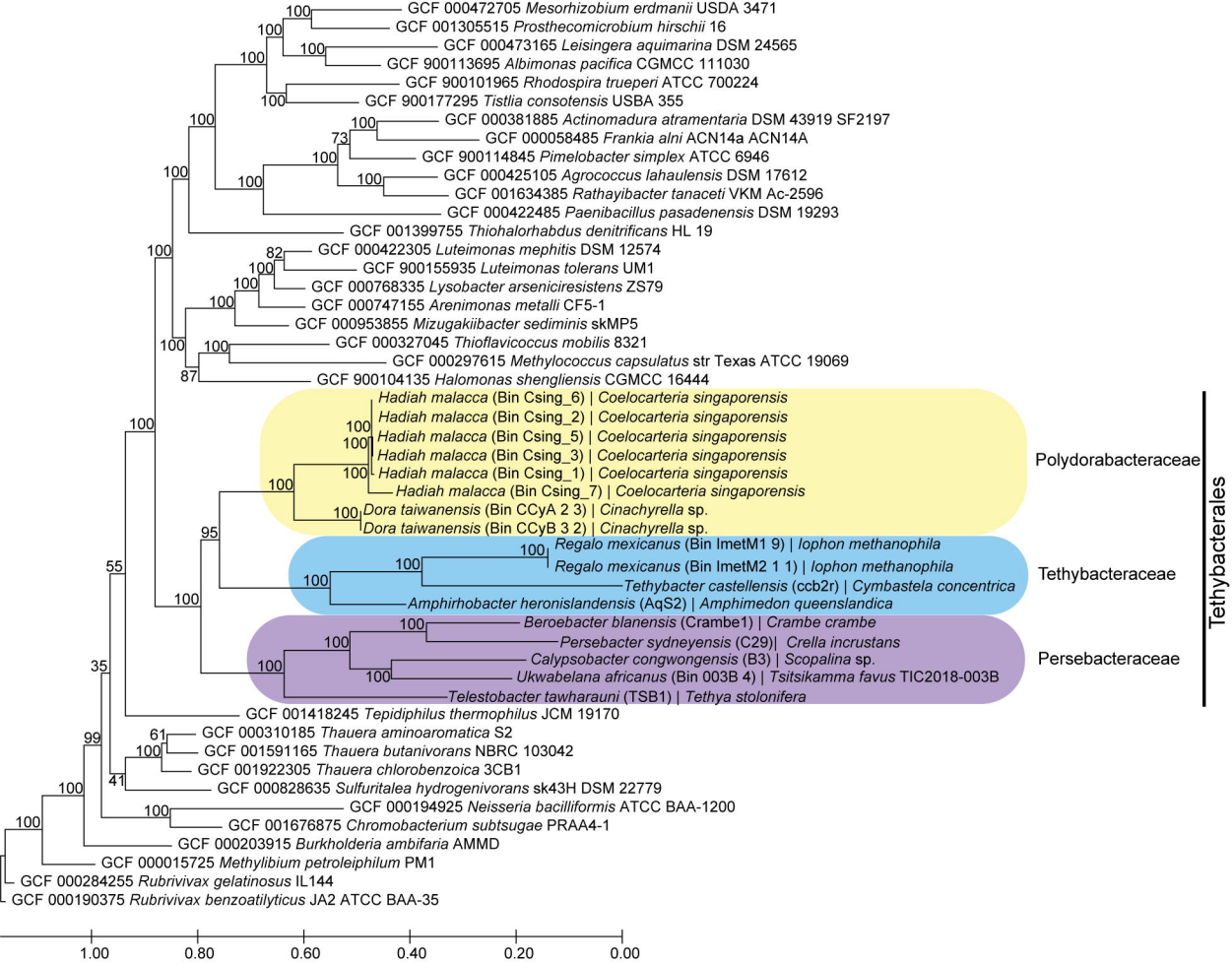
Table S8. Pairwise average 16S rRNA gene sequence identity scores between Tethybacterales genomes.

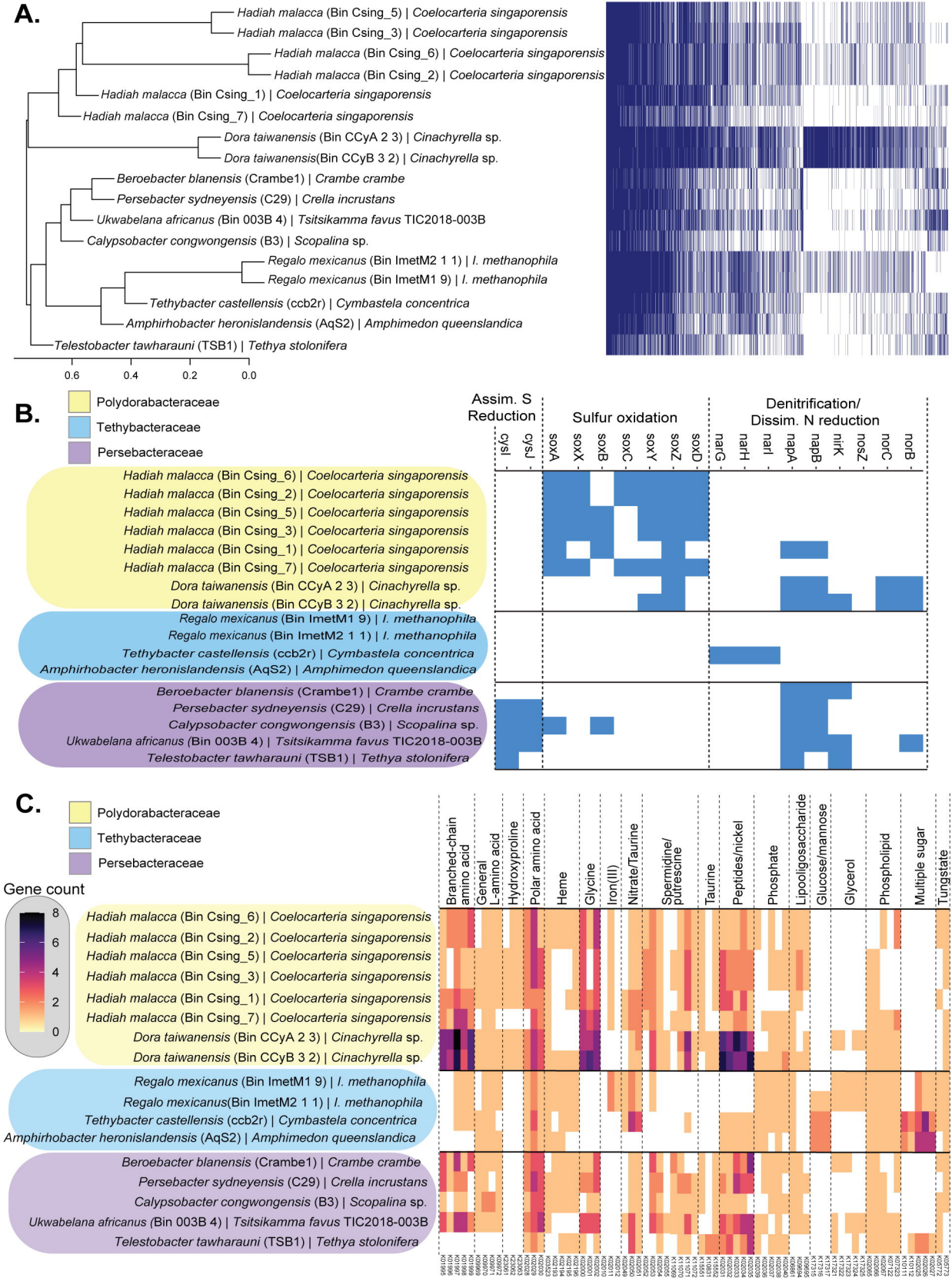
Table S9. KEGG annotation counts for all Tethybacterales and Poribacteria genomes grouped by metabolic pathway.

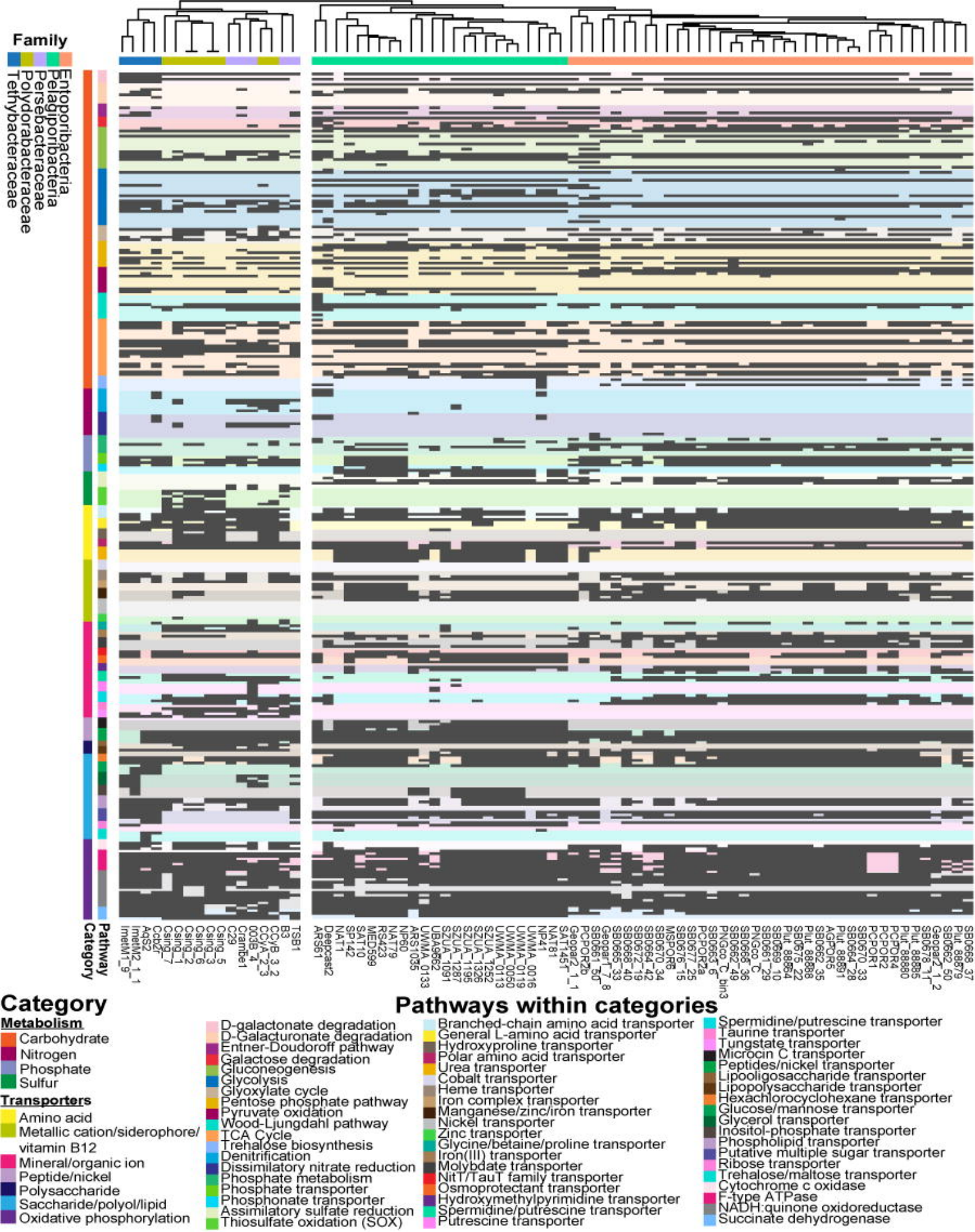
Table S10. Unique genes in Sp02-3 spirochete genome bins, relative to all other genes present in the four *T. favus* metagenomes.

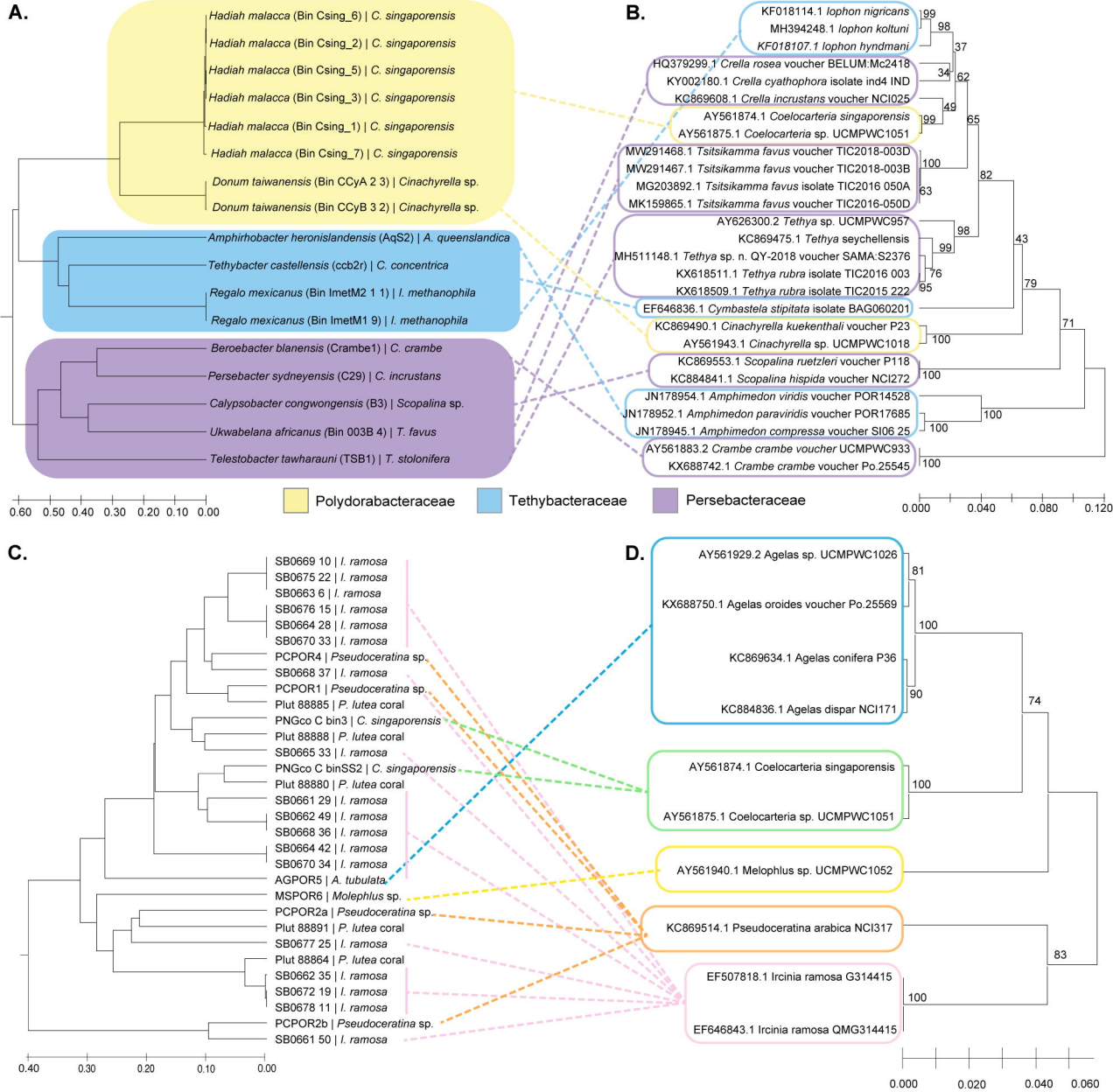
Table S11. Pairwise ANOSIM analysis of OTUs (distance of 0.03) showed that bacterial communities associated with different sponge genera are significantly different ($p < 0.05$).

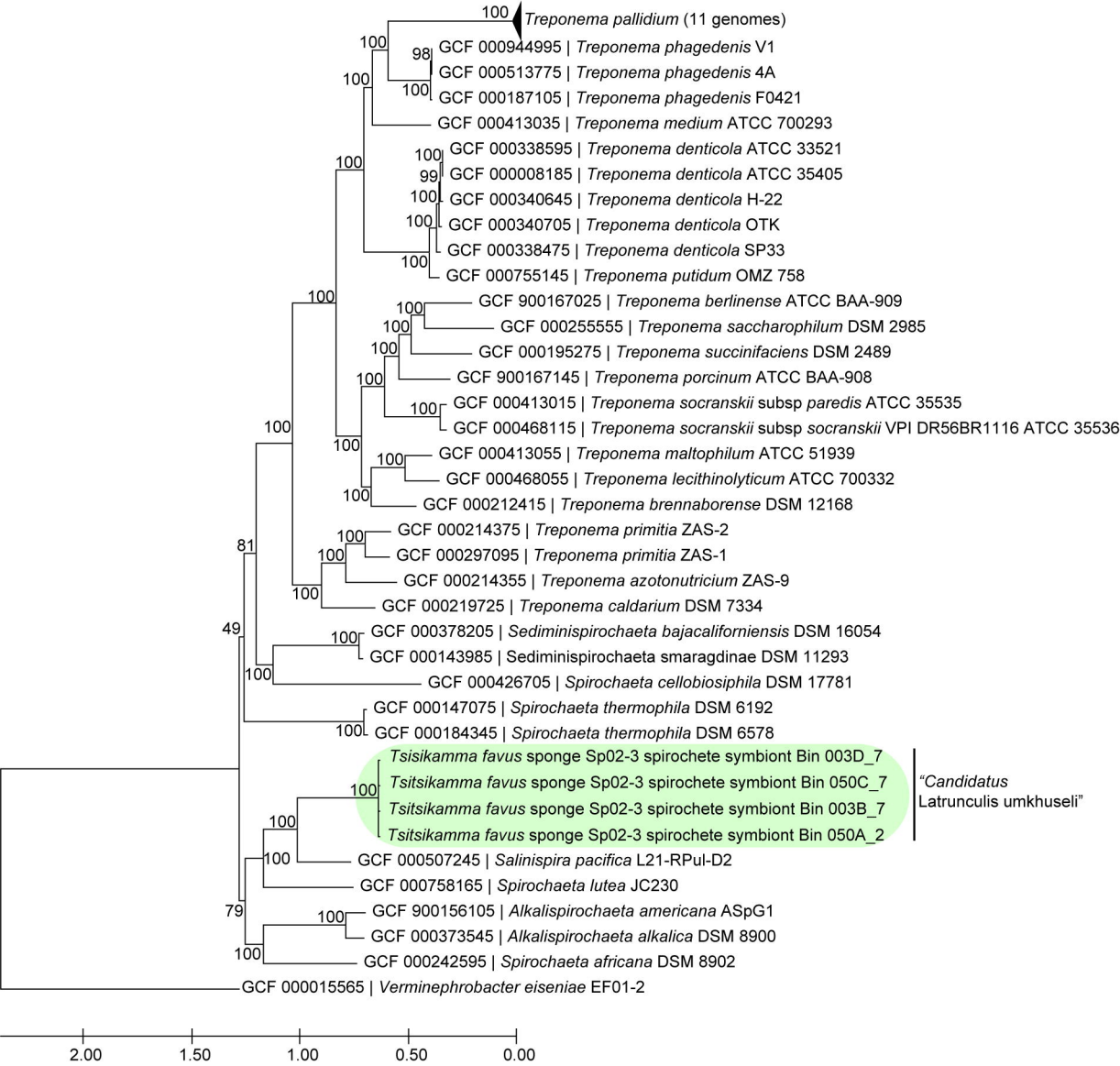
Table S12. Indicator Species Analysis found that there was a significant ($p < 0.05$) shift in the abundance of OTU5 between Type I and Type II *T. favus* and *T. michaeli* chemotypes.

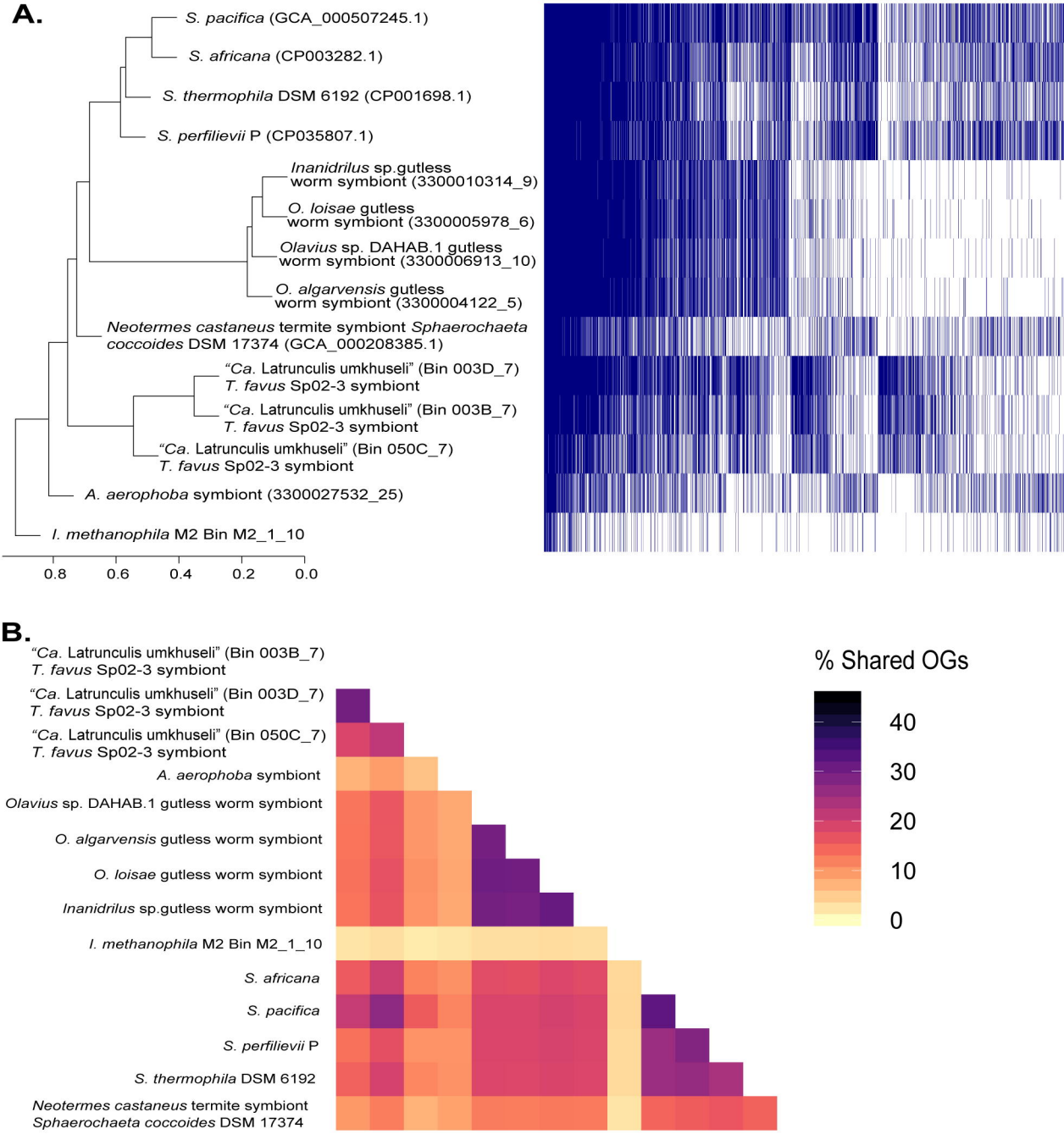

GCF 000472705 Mesorhizobium erdmanii USDA 3471
GCF 001305515 Prosthecomicrobium hirschii 16
GCF 000473165 Leisingera aquimarina DSM 24565
GCF 900113695 Albimonas pacifica CGMCC 111030
GCF 900101965 Rhodospira trueperi ATCC 700224
GCF 900177295 Tistlia consotensis USBA 355
GCF 000381885 Actinomadura atramentaria DSM 43919 SF2197
GCF 000058485 Frankia alni ACN14a ACN14A
GCF 900114845 Pimelobacter simplex ATCC 6946
GCF 000425105 Agrococcus lahaulensis DSM 17612
GCF 001634385 Rathayibacter tanaceti VKM Ac-2596
GCF 000422485 Paenibacillus pasadenensis DSM 19293
GCF 001399755 Thiohalorhabdus denitrificans HL 19
GCF 000422305 Luteimonas mephitis DSM 12574
GCF 900155935 Luteimonas tolerans UM1
GCF 000768335 Lysobacter arseniciresistens ZS79
GCF 000747155 Arenimonas metalli CF5-1
GCF 000953855 Mizugakiibacter sediminis skMP5
GCF 000327045 Thioflavicoccus mobilis 8321
GCF 000297615 Methylococcus capsulatus str Texas ATCC 19069
GCF 900104135 Halomonas shengliensis CGMCC 16444
Hadiah malacca (Bin Csing_6) | Coelocarteria singaporensis
Hadiah malacca (Bin Csing_2) | Coelocarteria singaporensis
Hadiah malacca (Bin Csing_5) | Coelocarteria singaporensis
Hadiah malacca (Bin Csing_3) | Coelocarteria singaporensis
Hadiah malacca (Bin Csing_1) | Coelocarteria singaporensis
Hadiah malacca (Bin Csing_7) | Coelocarteria singaporensis
Dora taiwanensis (Bin CCyA 2 3) | Cinachyrella sp.
Dora taiwanensis (Bin CCyB 3 2) | Cinachyrella sp.
Polydorabacteraceae
Regalo mexicanus (Bin ImetM1 9) | Iophon methanophila
Regalo mexicanus (Bin ImetM2 1 1) | Iophon methanophila
Tethybacter castellensis (ccb2r) | Cymbastela concentrica
Amphirhobacter heronislandensis (AqS2) | Amphimedon queenslandica
Tethybacteraceae
Beroebacter blanensis (Crambe1) | Crambe crambe
Persebacter sydneyensis (C29)| Crella incrustans
Calypsobacter congwongensis (B3) | Scopalina sp.
Ukwabelana africanus (Bin 003B 4) | Tsitsikamma favus TIC2018-003B
Telestobacter tawharauni (TSB1) | Tethya stolonifera
Persebacteraceae
Tethybacterales
GCF 001418245 Tepidiphilus thermophilus JCM 19170
GCF 000310185 Thauera aminoaromatica S2
GCF 001591165 Thauera butanivorans NBRC 103042
GCF 001922305 Thauera chlorobenzoica 3CB1
GCF 000828635 Sulfuritalea hydrogenivorans sk43H DSM 22779
GCF 000194925 Neisseria bacilliformis ATCC BAA-1200
GCF 001676875 Chromobacterium subtsugae PRAA4-1
GCF 000203915 Burkholderia ambifaria AMMD
GCF 000015725 Methylibium petroleiphilum PM1
GCF 000284255 Rubrivivax gelatinosus IL144
GCF 000190375 Rubrivivax benzoatilyticus JA2 ATCC BAA-35
1.00
0.80
0.60
0.40
0.20
0.00

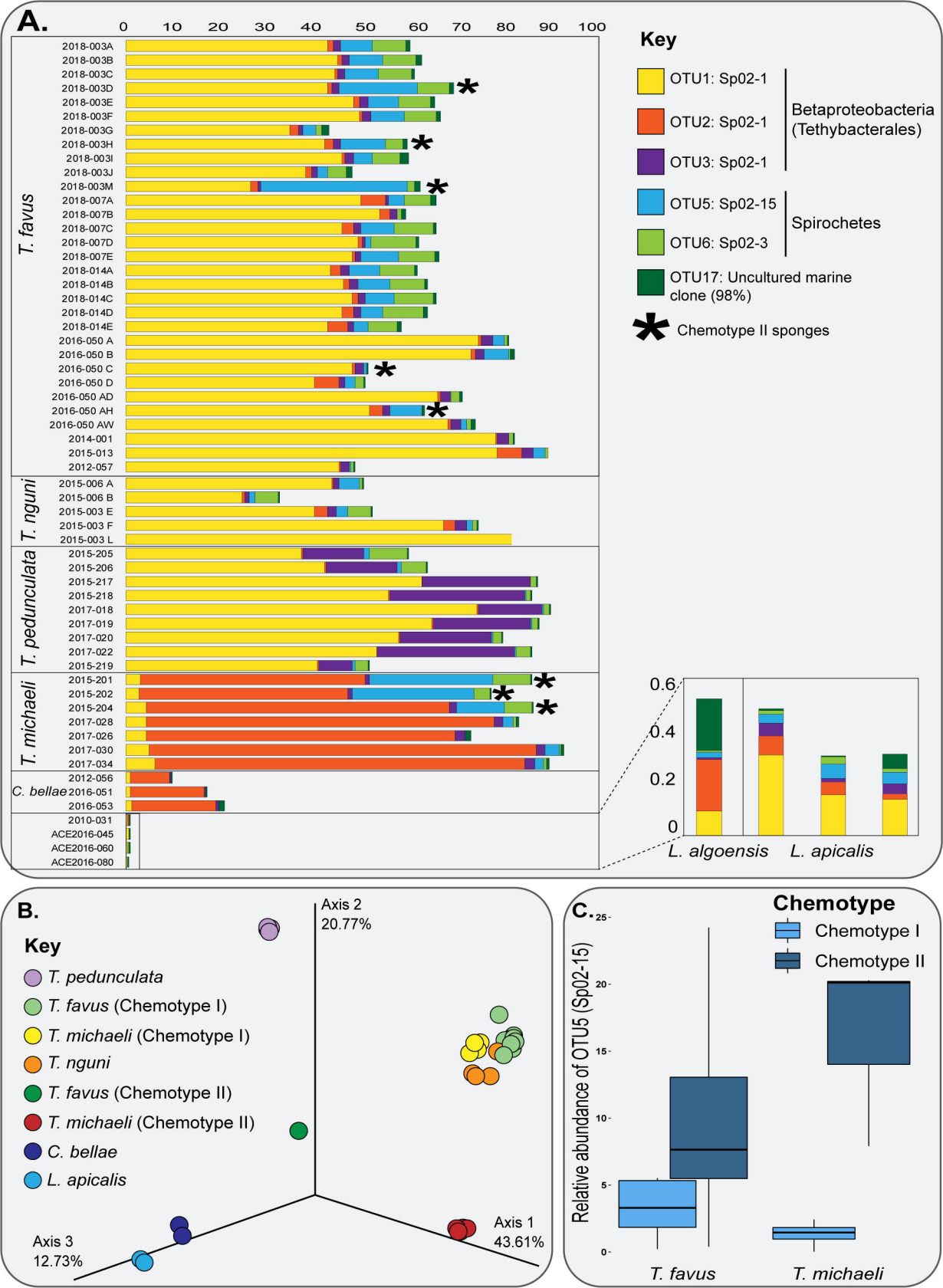

A.
Hadiah malacca (Bin Csing_5) | Coelocarteria singaporensis
Hadiah malacca (Bin Csing_3) | Coelocarteria singaporensis
Hadiah malacca (Bin Csing_6) | Coelocarteria singaporensis
Hadiah malacca (Bin Csing_2) | Coelocarteria singaporensis
Hadiah malacca (Bin Csing_1) | Coelocarteria singaporensis
Hadiah malacca (Bin Csing_7) | Coelocarteria singaporensis
Dora taiwanensis (Bin CCyA 2 3) | Cinachyrella sp.
Dora taiwanensis(Bin CCyB 3 2) | Cinachyrella sp.
Beroebacter blanensis (Crambe1) | Crambe crambe
Persebacter sydneyensis (C29) | Crella incrustans
Ukwabelana africanus (Bin 003B 4) | Tsitsikamma favus TIC2018-003B
Calypsobacter congwongensis (B3) | Scopalina sp.
Regalo mexicanus (Bin ImetM2 1 1) | I. methanophila
Regalo mexicanus (Bin ImetM1 9) | I. methanophila
Tethybacter castellensis (ccb2r) | Cymbastela concentrica
Amphirhobacter heronislandensis (AqS2) | Amphimedon queenslandica
Telestobacter tawharauni (TSB1) | Tethya stolonifera
0.6
0.4
0.2
0.0
B.
Polydorabacteraceae
Tethybacteraceae
Persebacteraceae
Assim. S Reduction
Sulfur oxidation
Denitrification/ Dissim. N reduction
cysI
cysJ
soxA
soxX
soxB
soxC
soxY
soxZ
soxD
narG
narH
narI
napA
napB
nirK
nosZ
norC
norB
C.
Gene count
8
6
4
2
0
Branched-chain amino acid
General L-amino acid
Hydroxyproline
Polar amino acid
Heme
Glycine
Iron(III)
Nitrate/Taurine
Spermidine/ putrescine
Taurine
Peptides/nickel
Phosphate
Lipooligosaccharide
Glucose/mannose
Glycerol
Phospholipid
Multiple sugar
Tungstate

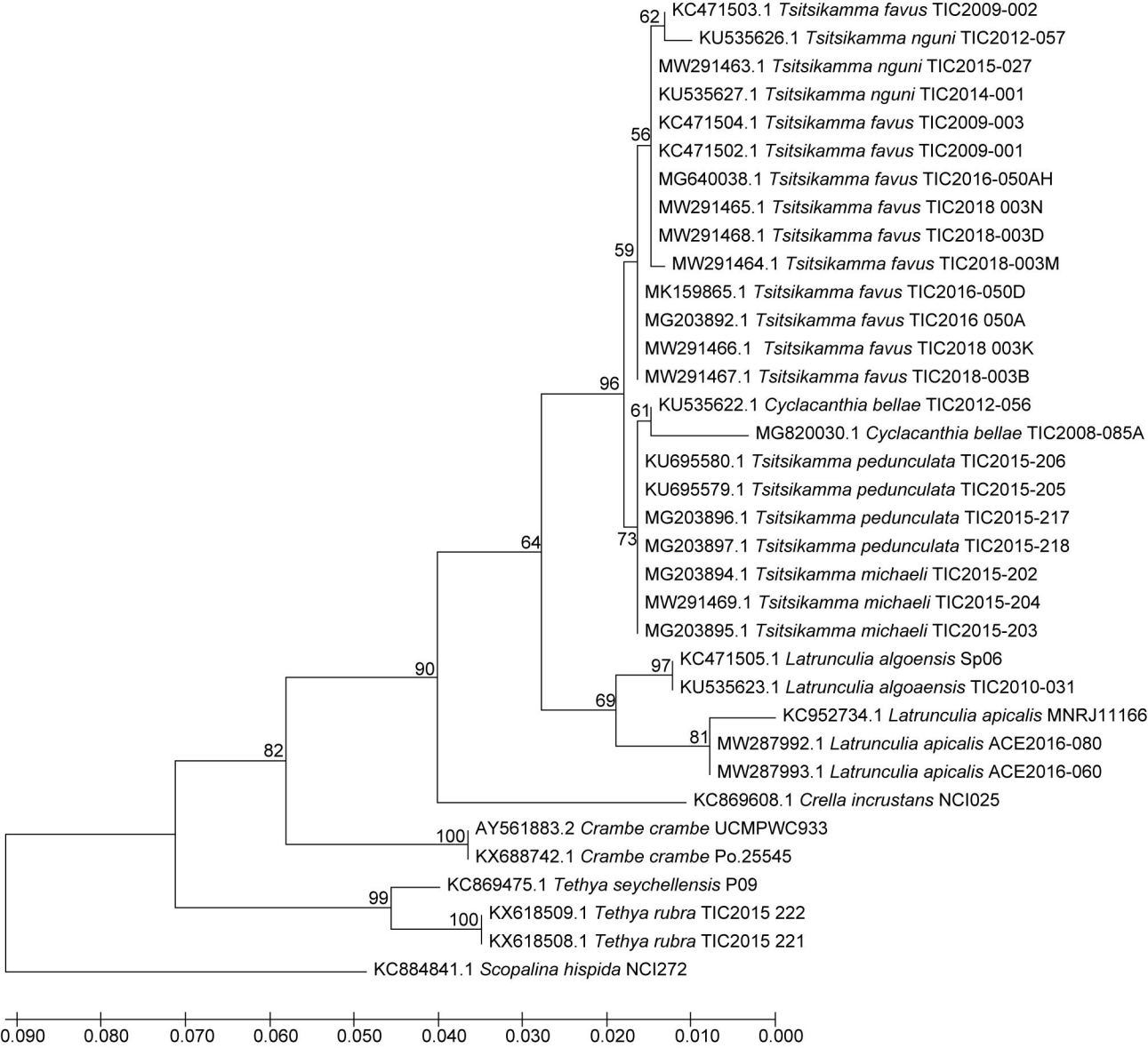

Family
Entoporibacteria
Pelagiporibacteria
Persebacteraceae
Polydorabacteraceae
Tethybacteraceae
Category
Pathway
Pathways within categories
Category
Metabolism
Carbohydrate
Nitrogen
Phosphate
Sulfur
Transporters
Amino acid
Metallic cation/siderophore/ vitamin B12
Mineral/organic ion
Peptide/nickel
Polysaccharide
Saccharide/polyol/lipid
Oxidative phosphorylation
D-galactonate degradation
D-Galacturonate degradation
Entner-Doudoroff pathway
Galactose degradation
Gluconeogenesis
Glycolysis
Glyoxylate cycle
Pentose phosphate pathway
Pyruvate oxidation
Wood-Ljungdahl pathway
TCA Cycle
Trehalose biosynthesis
Denitrification
Dissimilatory nitrate reduction
Phosphate metabolism
Phosphate transporter
Phosphonate transporter
Assimilatory sulfate reduction
Thiosulfate oxidation (SOX)
Branched-chain amino acid transporter
General L-amino acid transporter
Hydroxyproline transporter
Polar amino acid transporter
Urea transporter
Cobalt transporter
Heme transporter
Iron complex transporter
Manganese/zinc/iron transporter
Nickel transporter
Zinc transporter
Glycine/betaine/proline transporter
Iron(III) transporter
Molybdate transporter
NitT/TauT family transporter
Osmoprotectant transporter
Hydroxymethylpyrimidine transporter
Spermidine/putrescine transporter
Putrescine transporter
Spermidine/putrescine transporter
Taurine transporter
Tungstate transporter
Microcin C transporter
Peptides/nickel transporter
Lipooligosaccharide transporter
Lipopolysaccharide transporter
Hexachlorocyclohexane transporter
Glucose/mannose transporter
Glycerol transporter
Inositol-phosphate transporter
Phospholipid transporter
Putative multiple sugar transporter
Ribose transporter
Trehalose/maltose transporter
Cytochrome c oxidase
F-type ATPase
NADH:quinone oxidoreductase
Succinate dehydrogenase

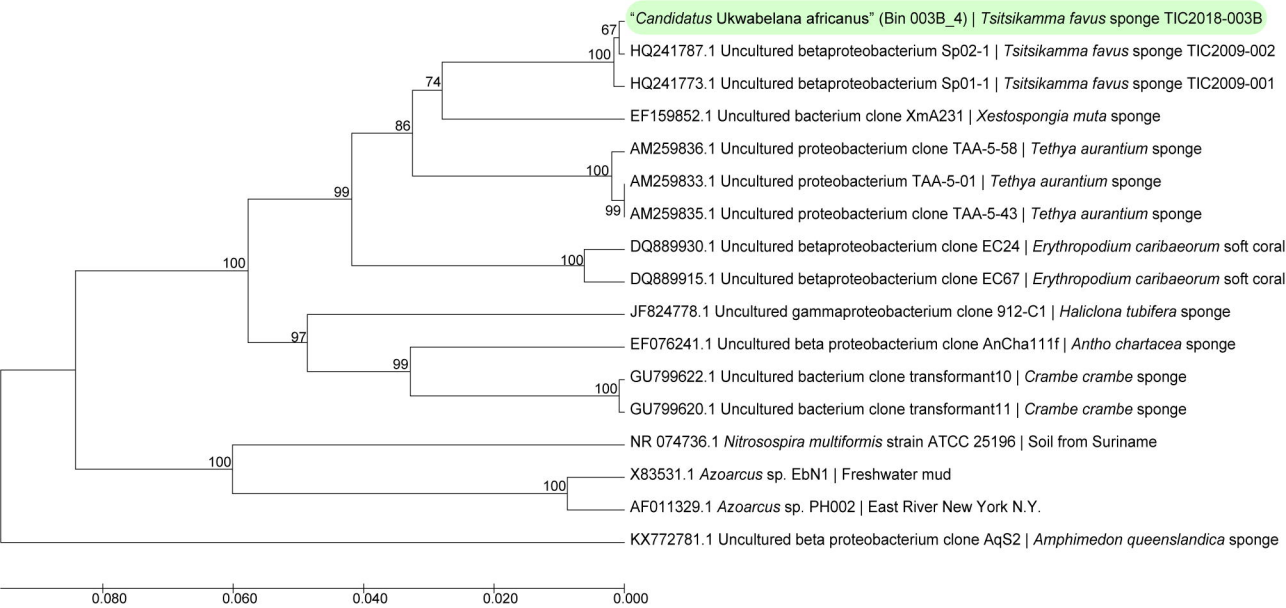

A.
Hadiah malacca (Bin Csing_6) | C. singaporensis
Hadiah malacca (Bin Csing_2) | C. singaporensis
Hadiah malacca (Bin Csing_5) | C. singaporensis
Hadiah malacca (Bin Csing_3) | C. singaporensis
Hadiah malacca (Bin Csing_1) | C. singaporensis
Hadiah malacca (Bin Csing_7) | C. singaporensis
Donum taiwanensis (Bin CCyA 2 3) | Cinachyrella sp.
Donum taiwanensis (Bin CCyB 3 2) | Cinachyrella sp.
Amphirhobacter heronislandensis (AqS2) | A. queenslandica
Tethybacter castellensis (ccb2r) | C. concentrica
Regalo mexicanus (Bin ImetM2 1 1) | I. methanophila
Regalo mexicanus (Bin ImetM1 9) | I. methanophila
Beroebacter blanensis (Crambe1) | C. crambe
Persebacter sydneyensis (C29) | C. incrustans
Calypsobacter congwongensis (B3) | Scopalina sp.
Ukwabelana africanus (Bin 003B 4) | T. favus
Telestobacter tawharauni (TSB1) | T. stolonifera
0.60 0.50 0.40 0.30 0.20 0.10 0.00
B.
KF018114.1 lophon nigricans
MH394248.1 lophon koltuni
KF018107.1 lophon hyndmani
HQ379299.1 Crella rosea voucher BELUM:Mc2418
KY002180.1 Crella cyathophora isolate ind4 IND
KC869608.1 Crella incrustans voucher NCI025
AY561874.1 Coelocarteria singaporensis
AY561875.1 Coelocarteria sp. UCMPWC1051
MW291468.1 Tsitsikamma favus voucher TIC2018-003D
MW291467.1 Tsitsikamma favus voucher TIC2018-003B
MG203892.1 Tsitsikamma favus isolate TIC2016 050A
MK159865.1 Tsitsikamma favus voucher TIC2016-050D
AY626300.2 Tethya sp. UCMPWC957
KC869475.1 Tethya seychellensis
MH511148.1 Tethya sp. n. QY-2018 voucher SAMA:S2376
KX618511.1 Tethya rubra isolate TIC2016 003
KX618509.1 Tethya rubra isolate TIC2015 222
EF646836.1 Cymbastela stipitata isolate BAG060201
KC869490.1 Cinachyrella kuekenthali voucher P23
AY561943.1 Cinachyrella sp. UCMPWC1018
KC869553.1 Scopalina ruetzleri voucher P118
KC884841.1 Scopalina hispida voucher NCI272
JN178954.1 Amphimedon viridis voucher POR14528
JN178952.1 Amphimedon paraviridis voucher POR17685
JN178945.1 Amphimedon compressa voucher SI06 25
AY561883.2 Crambe crambe voucher UCMPWC933
KX688742.1 Crambe crambe voucher Po.25545
99 98 37 34 62 49 99 65 100 82 63 98 43 99 76 95 79 71 100 100 100 100 100
0.000 0.040 0.080 0.120
Polydorabacteraceae
Tethybacteraceae
Persebacteraceae
C.
SB0669 10 | I. ramosa
SB0675 22 | I. ramosa
SB0663 6 | I. ramosa
SB0676 15 | I. ramosa
SB0664 28 | I. ramosa
SB0670 33 | I. ramosa
PCPOR4 | Pseudoceratina sp.
SB0668 37 | I. ramosa
PCPOR1 | Pseudoceratina sp.
Plut 88885 | P. lutea coral
PNGco C bin3 | C. singaporensis
Plut 88888 | P. lutea coral
SB0665 33 | I. ramosa
PNGco C binSS2 | C. singaporensis
Plut 88880 | P. lutea coral
SB0661 29 | I. ramosa
SB0662 49 | I. ramosa
SB0668 36 | I. ramosa
SB0664 42 | I. ramosa
SB0670 34 | I. ramosa
AGPOR5 | A. tubulata
MSPOR6 | Molephlus sp.
PCPOR2a | Pseudoceratina sp.
Plut 88891 | P. lutea coral
SB0677 25 | I. ramosa
Plut 88864 | P. lutea coral
SB0662 35 | I. ramosa
SB0672 19 | I. ramosa
SB0678 11 | I. ramosa
PCPOR2b | Pseudoceratina sp.
SB0661 50 | I. ramosa
0.40 0.30 0.20 0.10 0.00
D.
AY561929.2 Agelas sp. UCMPWC1026
KX688750.1 Agelas oroides voucher Po.25569
KC869634.1 Agelas conifera P36
KC884836.1 Agelas dispar NCI171
AY561874.1 Coelocarteria singaporensis
AY561875.1 Coelocarteria sp. UCMPWC1051
AY561940.1 Melophlus sp. UCMPWC1052
KC869514.1 Pseudoceratina arabica NCI317
EF507818.1 Ircinia ramosa G314415
EF646843.1 Ircinia ramosa QMG314415
81 100 90 100 74 83 100
0.000 0.020 0.040 0.060

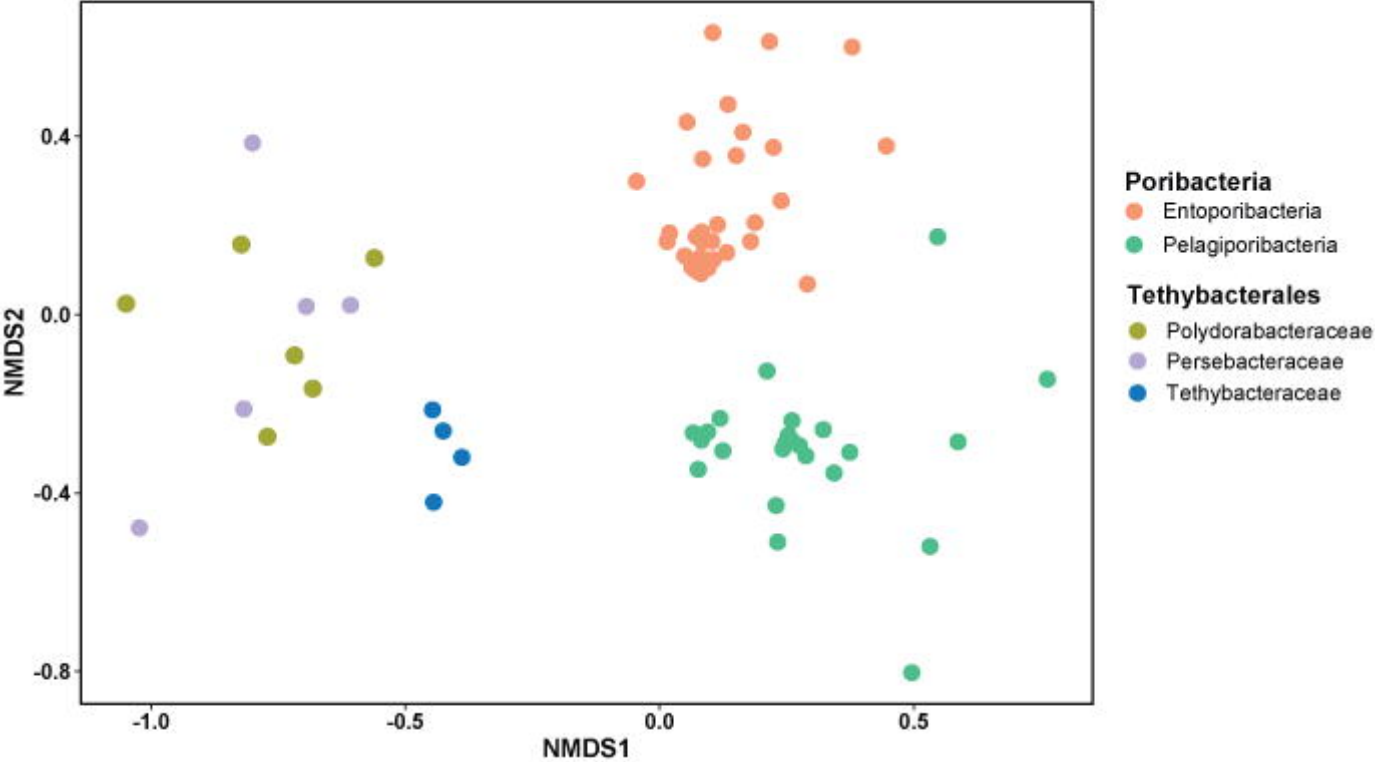

Treponema pallidium (11 genomes)
GCF 000944995 | Treponema phagedenis V1
GCF 000513775 | Treponema phagedenis 4A
GCF 000187105 | Treponema phagedenis F0421
GCF 000413035 | Treponema medium ATCC 700293
GCF 000338595 | Treponema denticola ATCC 33521
GCF 000008185 | Treponema denticola ATCC 35405
GCF 000340645 | Treponema denticola H-22
GCF 000340705 | Treponema denticola OTK
GCF 000338475 | Treponema denticola SP33
GCF 000755145 | Treponema putidum OMZ 758
GCF 900167025 | Treponema berlinense ATCC BAA-909
GCF 000255555 | Treponema saccharophilum DSM 2985
GCF 000195275 | Treponema succinifaciens DSM 2489
GCF 900167145 | Treponema porcinum ATCC BAA-908
GCF 000413015 | Treponema socranskii subsp paredis ATCC 35535
GCF 000468115 | Treponema socranskii subsp socranskii VPI DR56BR1116 ATCC 35536
GCF 000413055 | Treponema maltophilum ATCC 51939
GCF 000468055 | Treponema lecithinolyticum ATCC 700332
GCF 000212415 | Treponema brennaborense DSM 12168
GCF 000214375 | Treponema primitia ZAS-2
GCF 000297095 | Treponema primitia ZAS-1
GCF 000214355 | Treponema azotonutricium ZAS-9
GCF 000219725 | Treponema caldarium DSM 7334
GCF 000378205 | Sediminispirochaeta bajacaliforniensis DSM 16054
GCF 000143985 | Sediminispirochaeta smaragdinae DSM 11293
GCF 000426705 | Spirochaeta cellobiosiphila DSM 17781
GCF 000147075 | Spirochaeta thermophila DSM 6192
GCF 000184345 | Spirochaeta thermophila DSM 6578
Tsisikamma favus sponge Sp02-3 spirochete symbiont Bin 003D_7
Tsitsikamma favus sponge Sp02-3 spirochete symbiont Bin 050C_7
Tsitsikamma favus sponge Sp02-3 spirochete symbiont Bin 003B_7
Tsitsikamma favus sponge Sp02-3 spirochete symbiont Bin 050A_2
"Candidatus Latrunculis umkhuseli"
GCF 000507245 | Salinispira pacifica L21-RPul-D2
GCF 000758165 | Spirochaeta lutea JC230
GCF 900156105 | Alkalispirochaeta americana ASpG1
GCF 000373545 | Alkalispirochaeta alkalica DSM 8900
GCF 000242595 | Spirochaeta africana DSM 8902
GCF 000015565 | Verminephrobacter eiseniae EF01-2
2.00
1.50
1.00
0.50
0.00

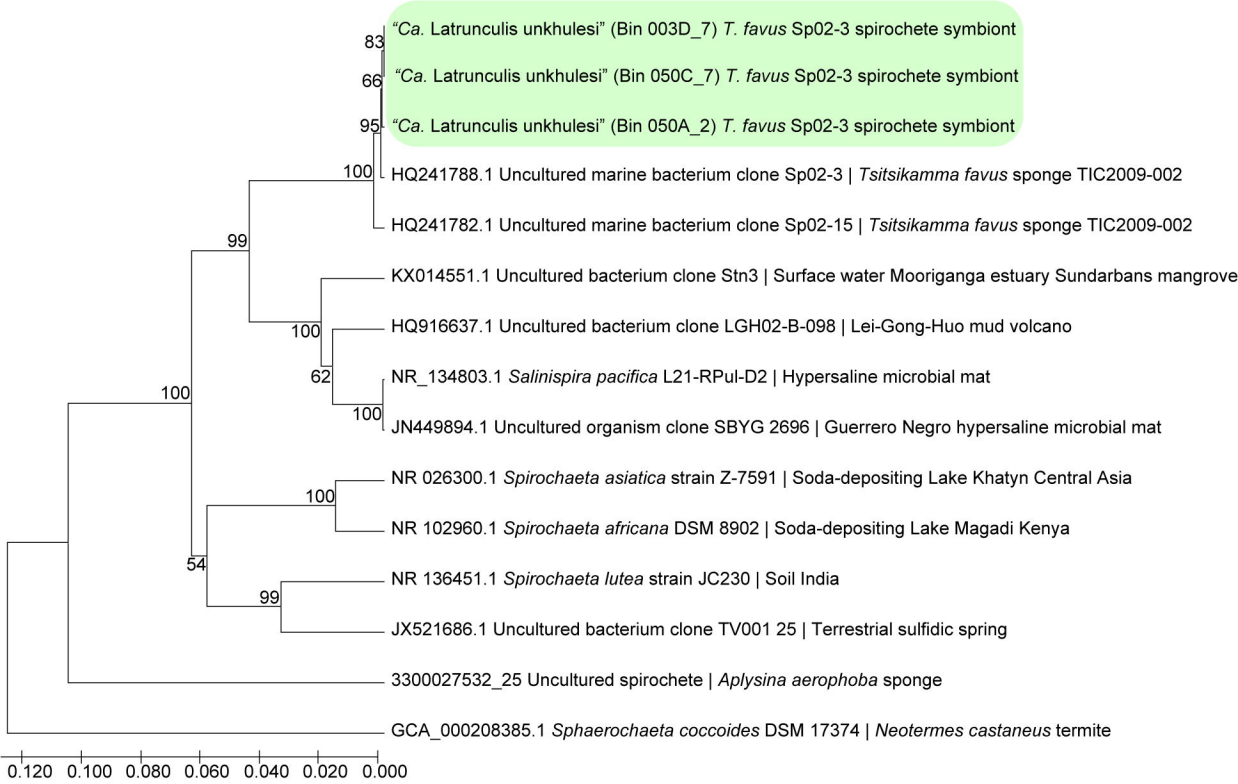

A.
S. pacifica (GCA_000507245.1)
S. africana (CP003282.1)
S. thermophila DSM 6192 (CP001698.1)
S. perfilievii P (CP035807.1)
Inanidrilus sp.gutless worm symbiont (3300010314_9)
O. loisae gutless worm symbiont (3300005978_6)
Olavius sp. DAHAB.1 gutless worm symbiont (3300006913_10)
O. algarvensis gutless worm symbiont (3300004122_5)
Neotermes castaneus termite symbiont Sphaerochaeta coccoides DSM 17374 (GCA_000208385.1)
"Ca. Latrunculis umkhuseli" (Bin 003D_7) T. favus Sp02-3 symbiont
"Ca. Latrunculis umkhuseli" (Bin 003B_7) T. favus Sp02-3 symbiont
"Ca. Latrunculis umkhuseli" (Bin 050C_7) T. favus Sp02-3 symbiont
A. aerophoba symbiont (3300027532_25)
I. methanophila M2 Bin M2_1_10
0.8
0.6
0.4
0.2
0.0
B.
"Ca. Latrunculis umkhuseli" (Bin 003B_7) T. favus Sp02-3 symbiont
"Ca. Latrunculis umkhuseli" (Bin 003D_7) T. favus Sp02-3 symbiont
"Ca. Latrunculis umkhuseli" (Bin 050C_7) T. favus Sp02-3 symbiont
A. aerophoba symbiont
Olavius sp. DAHAB.1 gutless worm symbiont
O. algarvensis gutless worm symbiont
O. loisae gutless worm symbiont
Inanidrilus sp.gutless worm symbiont
I. methanophila M2 Bin M2_1_10
S. africana
S. pacifica
S. perfilievii P
S. thermophila DSM 6192
Neotermes castaneus termite symbiont Sphaerochaeta coccoides DSM 17374
% Shared OGs
40
30
20
10
0

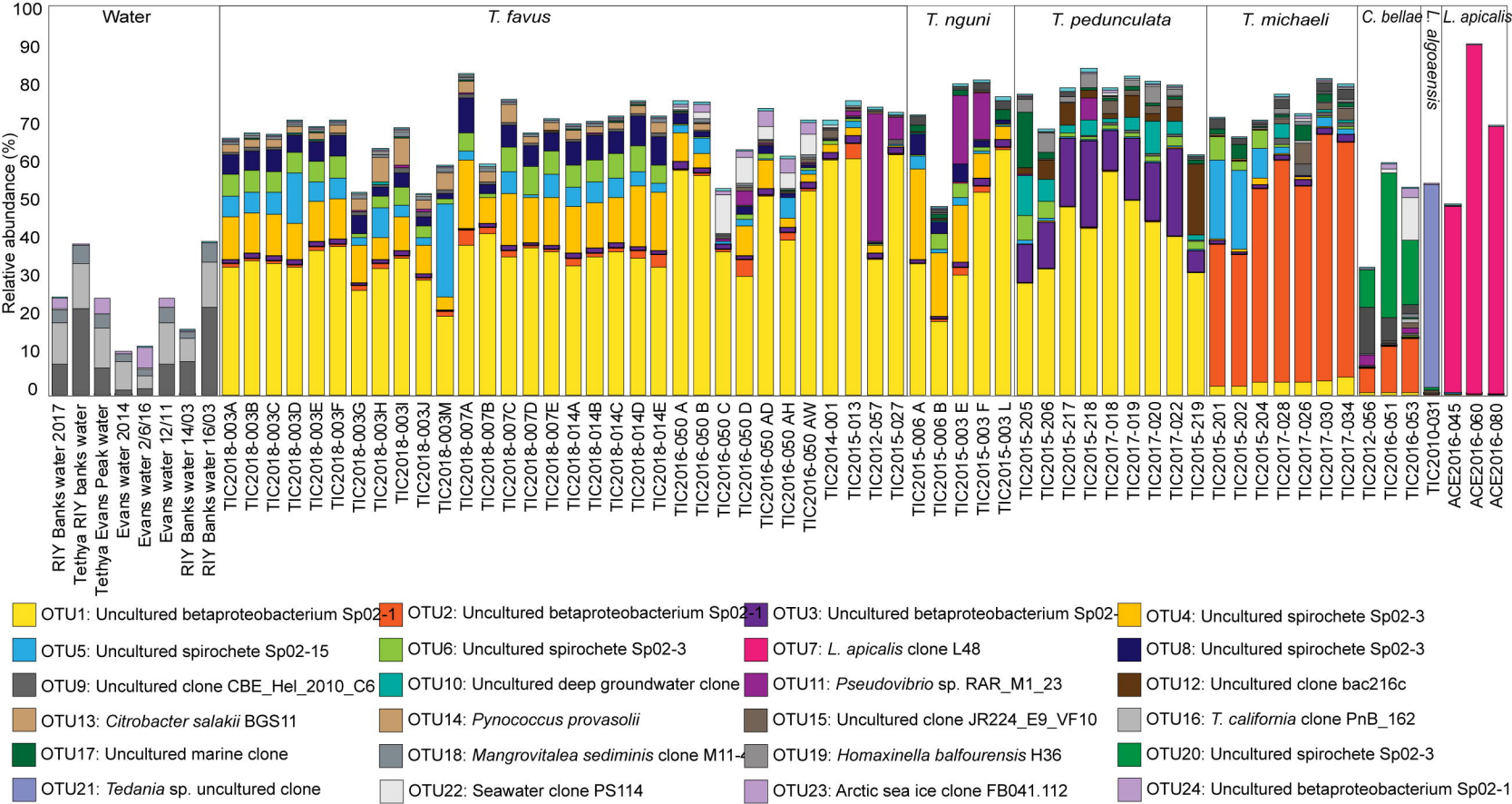

A.
0 10 20 30 40 50 60 70 80 90 100
T. favus
2018-003A
2018-003B
2018-003C
2018-003D
2018-003E
2018-003F
2018-003G
2018-003H
2018-003I
2018-003J
2018-003M
2018-007A
2018-007B
2018-007C
2018-007D
2018-007E
2018-014A
2018-014B
2018-014C
2018-014D
2018-014E
2016-050 A
2016-050 B
2016-050 C
2016-050 D
2016-050 AD
2016-050 AH
2016-050 AW
2014-001
2015-013
2012-057
T. nguni
2015-006 A
2015-006 B
2015-003 E
2015-003 F
2015-003 L
T. pedunculata
2015-205
2015-206
2015-217
2015-218
2017-018
2017-019
2017-020
2017-022
2015-219
T. michaeli
2015-201
2015-202
2015-204
2017-028
2017-026
2017-030
2017-034
C. bellae
2012-056
2016-051
2016-053
2010-031
ACE2016-045
ACE2016-060
ACE2016-080
Key
OTU1: Sp02-1
OTU2: Sp02-1
OTU3: Sp02-1
Betaproteobacteria (Tethybacterales)
OTU5: Sp02-15
OTU6: Sp02-3
Spirochetes
OTU17: Uncultured marine clone (98%)
Chemotype II sponges
0.6
0.4
0.2
0
L. algoensis
L. apicalis
B.
Key
T. pedunculata
T. favus (Chemotype I)
T. michaeli (Chemotype I)
T. nguni
T. favus (Chemotype II)
T. michaeli (Chemotype II)
C. bellae
L. apicalis
Axis 2
20.77%
Axis 1
43.61%
Axis 3
12.73%
C.
Chemotype
Chemotype I
Chemotype II
Relative abundance of OTU5 (Sp02-15)
25
20
15
10
5
0
T. favus
T. michaeli

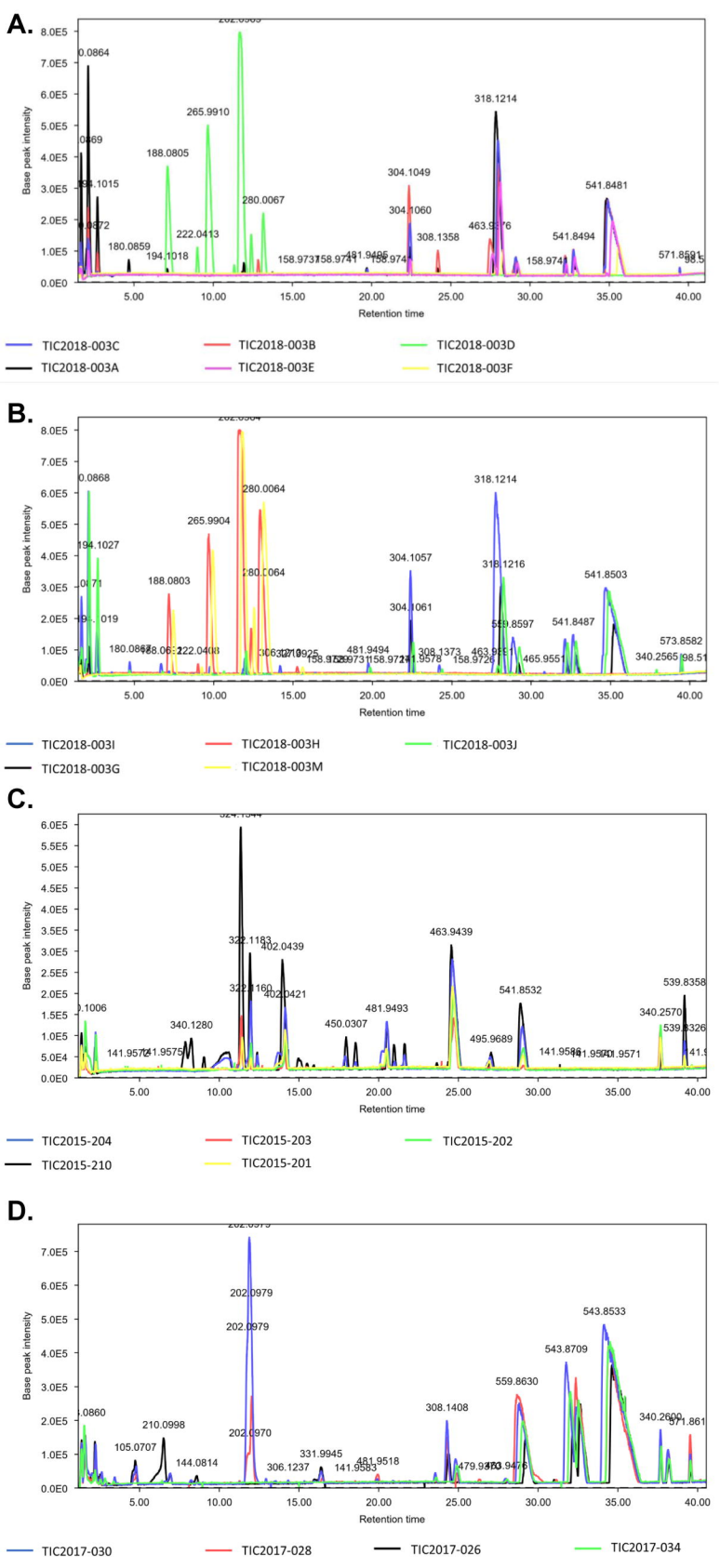

KC471503.1 Tsitsikamma favus TIC2009-002
KU535626.1 Tsitsikamma nguni TIC2012-057
MW291463.1 Tsitsikamma nguni TIC2015-027
KU535627.1 Tsitsikamma nguni TIC2014-001
KC471504.1 Tsitsikamma favus TIC2009-003
KC471502.1 Tsitsikamma favus TIC2009-001
MG640038.1 Tsitsikamma favus TIC2016-050AH
MW291465.1 Tsitsikamma favus TIC2018 003N
MW291468.1 Tsitsikamma favus TIC2018-003D
MW291464.1 Tsitsikamma favus TIC2018-003M
MK159865.1 Tsitsikamma favus TIC2016-050D
MG203892.1 Tsitsikamma favus TIC2016 050A
MW291466.1 Tsitsikamma favus TIC2018 003K
MW291467.1 Tsitsikamma favus TIC2018-003B
KU535622.1 Cyclacanthia bellae TIC2012-056
MG820030.1 Cyclacanthia bellae TIC2008-085A
KU695580.1 Tsitsikamma pedunculata TIC2015-206
KU695579.1 Tsitsikamma pedunculata TIC2015-205
MG203896.1 Tsitsikamma pedunculata TIC2015-217
MG203897.1 Tsitsikamma pedunculata TIC2015-218
MG203894.1 Tsitsikamma michaeli TIC2015-202
MW291469.1 Tsitsikamma michaeli TIC2015-204
MG203895.1 Tsitsikamma michaeli TIC2015-203
KC471505.1 Latrunculia algoensis Sp06
KU535623.1 Latrunculia algoaensis TIC2010-031
KC952734.1 Latrunculia apicalis MNRJ11166
MW287992.1 Latrunculia apicalis ACE2016-080
MW287993.1 Latrunculia apicalis ACE2016-060
KC869608.1 Crella incrustans NCI025
AY561883.2 Crambe crambe UCMPWC933
KX688742.1 Crambe crambe Po.25545
KC869475.1 Tethya seychellensis P09
KX618509.1 Tethya rubra TIC2015 222
KX618508.1 Tethya rubra TIC2015 221
KC884841.1 Scopalina hispida NCI272
62
56
59
96
61
73
64
97
69
81
90
82
100
99
100
0.090
0.080
0.070
0.060
0.050
0.040
0.030
0.020
0.010
0.000


"Candidatus Ukwabelana africanus" (Bin 003B_4) | Tsitsikamma favus sponge TIC2018-003B
HQ241787.1 Uncultured betaproteobacterium Sp02-1 | Tsitsikamma favus sponge TIC2009-002
HQ241773.1 Uncultured betaproteobacterium Sp01-1 | Tsitsikamma favus sponge TIC2009-001
EF159852.1 Uncultured bacterium clone XmA231 | Xestospongia muta sponge
AM259836.1 Uncultured proteobacterium clone TAA-5-58 | Tethya aurantium sponge
AM259833.1 Uncultured proteobacterium TAA-5-01 | Tethya aurantium sponge
AM259835.1 Uncultured proteobacterium clone TAA-5-43 | Tethya aurantium sponge
DQ889930.1 Uncultured betaproteobacterium clone EC24 | Erythropodium caribaeorum soft coral
DQ889915.1 Uncultured betaproteobacterium clone EC67 | Erythropodium caribaeorum soft coral
JF824778.1 Uncultured gammaproteobacterium clone 912-C1 | Haliclona tubifera sponge
EF076241.1 Uncultured beta proteobacterium clone AnCha111f | Antho chartacea sponge
GU799622.1 Uncultured bacterium clone transformant10 | Crambe crambe sponge
GU799620.1 Uncultured bacterium clone transformant11 | Crambe crambe sponge
NR 074736.1 Nitrosospira multiformis strain ATCC 25196 | Soil from Suriname
X83531.1 Azoarcus sp. EbN1 | Freshwater mud
AF011329.1 Azoarcus sp. PH002 | East River New York N.Y.
KX772781.1 Uncultured beta proteobacterium clone AqS2 | Amphimedon queenslandica sponge
67
100
74
86
100
99
99
100
100
97
99
100
100
100
0.080
0.060
0.040
0.020
0.000


0.4
0.0
-0.4
-0.8
NMDS2
-1.0
-0.5
0.0
0.5
NMDS1
Poribacteria
Entoporibacteria
Pelagiporibacteria
Tethybacterales
Polydorabacteraceae
Persebacteraceae
Tethybacteraceae


"Ca. Latrunculis unkhulesi" (Bin 003D_7) T. favus Sp02-3 spirochete symbiont
"Ca. Latrunculis unkhulesi" (Bin 050C_7) T. favus Sp02-3 spirochete symbiont
"Ca. Latrunculis unkhulesi" (Bin 050A_2) T. favus Sp02-3 spirochete symbiont
HQ241788.1 Uncultured marine bacterium clone Sp02-3 | Tsitsikamma favus sponge TIC2009-002
HQ241782.1 Uncultured marine bacterium clone Sp02-15 | Tsitsikamma favus sponge TIC2009-002
KX014551.1 Uncultured bacterium clone Stn3 | Surface water Mooriganga estuary Sundarbans mangrove
HQ916637.1 Uncultured bacterium clone LGH02-B-098 | Lei-Gong-Huo mud volcano
NR_134803.1 Salinispira pacifica L21-RPul-D2 | Hypersaline microbial mat
JN449894.1 Uncultured organism clone SBYG 2696 | Guerrero Negro hypersaline microbial mat
NR 026300.1 Spirochaeta asiatica strain Z-7591 | Soda-depositing Lake Khatyn Central Asia
NR 102960.1 Spirochaeta africana DSM 8902 | Soda-depositing Lake Magadi Kenya
NR 136451.1 Spirochaeta lutea strain JC230 | Soil India
JX521686.1 Uncultured bacterium clone TV001 25 | Terrestrial sulfidic spring
3300027532_25 Uncultured spirochete | Aplysina aerophoba sponge
GCA_000208385.1 Sphaerochaeta coccoides DSM 17374 | Neotermes castaneus termite
83
66
95
100
99
100
62
100
100
100
54
99
0.120 0.100 0.080 0.060 0.040 0.020 0.000


Water
T. favus
T. nguni
T. pedunculata
T. michaeli
C. bellae ?
L. apicalis
Relative abundance (%)
0
10
20
30
40
50
60
70
80
90
100
RIY Banks water 2017
Tethya RIY banks water
Tethya Evans Peak water
Evans water 2014
Evans water 2/6/16
Evans water 12/11
RIY Banks water 14/03
RIY Banks water 16/03
TIC2018-003A
TIC2018-003B
TIC2018-003C
TIC2018-003D
TIC2018-003E
TIC2018-003F
TIC2018-003G
TIC2018-003H
TIC2018-003I
TIC2018-003J
TIC2018-003M
TIC2018-007A
TIC2018-007B
TIC2018-007C
TIC2018-007D
TIC2018-007E
TIC2018-014A
TIC2018-014B
TIC2018-014C
TIC2018-014D
TIC2018-014E
TIC2016-050 A
TIC2016-050 B
TIC2016-050 C
TIC2016-050 D
TIC2016-050 AD
TIC2016-050 AH
TIC2016-050 AW
TIC2014-001
TIC2015-013
TIC2012-057
TIC2015-027
TIC2015-006 A
TIC2015-006 B
TIC2015-003 E
TIC2015-003 F
TIC2015-003 L
TIC2015-205
TIC2015-206
TIC2015-217
TIC2015-218
TIC2017-018
TIC2017-019
TIC2017-020
TIC2017-022
TIC2015-219
TIC2015-201
TIC2015-202
TIC2015-204
TIC2017-028
TIC2017-026
TIC2017-030
TIC2017-034
TIC2012-056
TIC2016-051
TIC2016-053
TIC2010-031
ACE2016-045
ACE2016-060
ACE2016-080
OTU1: Uncultured betaproteobacterium Sp02-1
OTU2: Uncultured betaproteobacterium Sp02-1
OTU3: Uncultured betaproteobacterium Sp02-
OTU4: Uncultured spirochete Sp02-3
OTU5: Uncultured spirochete Sp02-15
OTU6: Uncultured spirochete Sp02-3
OTU7: L. apicalis clone L48
OTU8: Uncultured spirochete Sp02-3
OTU9: Uncultured clone CBE_Hel_2010_C6
OTU10: Uncultured deep groundwater clone
OTU11: Pseudovibrio sp. RAR_M1_23
OTU12: Uncultured clone bac216c
OTU13: Citrobacter salakii BGS11
OTU14: Pynococcus provasolii
OTU15: Uncultured clone JR224_E9_VF10
OTU16: T. california clone PnB_162
OTU17: Uncultured marine clone
OTU18: Mangrovitalea sediminis clone M11-4
OTU19: Homaxinella balfourensis H36
OTU20: Uncultured spirochete Sp02-3
OTU21: Tedania sp. uncultured clone
OTU22: Seawater clone PS114
OTU23: Arctic sea ice clone FB041.112
OTU24: Uncultured betaproteobacterium Sp02-1


A.
Base peak intensity
Retention time
TIC2018-003C
TIC2018-003B
TIC2018-003D
TIC2018-003A
TIC2018-003E
TIC2018-003F
B.
Base peak intensity
Retention time
TIC2018-003I
TIC2018-003H
TIC2018-003J
TIC2018-003G
TIC2018-003M
C.
Base peak intensity
Retention time
TIC2015-204
TIC2015-203
TIC2015-202
TIC2015-210
TIC2015-201
D.
Base peak intensity
Retention time
TIC2017-030
TIC2017-028
TIC2017-026
TIC2017-034