Shaping chemodiversity: A unique spirochete symbiont of pyrroloiminoquinone-producing marine sponges

Samantha C. Waterworth[a,b], Gabriella M. Solomons[a], Jarmo-Charles J. Kalinski[a,c], Luthando S. Madonsela[a], Shirley Parker-Nance[a,d,e], Rosemary A. Dorrington* [a, e]

[a] Department of Biochemistry and Microbiology, Rhodes University, Makhanda, South Africa
[b] Current address: National Cancer Institute, Frederick, Maryland, USA
[c] Current address: Department of Biochemistry, University of California, Riverside, California, USA
[d] South African Environmental Observation Network, Elwandle Coastal Node, Port Elizabeth, South Africa
[e] South African Institute for Aquatic Biodiversity, Makhanda, South Africa

Running title: Characterization of spirochetes in latrunculid sponges

***Address correspondence to:** Rosemary A. Dorrington, r.dorrington@ru.ac.za

ABSTRACT

Bacterial symbionts are critical members of many marine sponge holobionts. Some sponge-associated bacterial lineages, such as Poribacteria, SAUL, and Tethybacterales appear to have broad host ranges and associate with a diversity of sponge species, while others are more species-specific, having adapted to the niche environment of their host. Host-associated spirochete symbionts that are numerically dominant have been documented in several invertebrates including termites, starfish, and corals. However, dominant spirochete populations are relatively rare in marine sponges, thus far only observed in *Clathrina clathrus* and various species within the Latrunculiidae sponge family, where they are co-dominant alongside Tethybacterales symbionts. This study aimed to characterize these spirochetes and their potential role in the host sponge.

Analysis of metagenome-assembled genomes (MAGs) from eight latrunculid sponges revealed that these unusual spirochetes are relatively recent symbionts currently undergoing genome erosion and are phylogenetically distinct from other sponge-associated spirochetes. Functional comparative analysis suggests that the host sponge may select for these spirochetes due to their ability to produce methionine and/or their ability to produce terpenoids. A dysbiosis of spirochete strains was correlated with a change in host chemotype suggesting that the spirochete populations may play a role in shaping the abundances of different pyrroloiminoquinone compounds characteristic of latrunculid sponges.

IMPORTANCE

South African latrunculid sponges are host to co-dominant Tethybacterales and Spirochete symbionts. While the Tethybacterales are broad-host range symbionts, the spirochetes have not been reported as abundant in any other sponge except *Clathrina clathrus*. However, spirochetes are regularly the most dominant populations in marine corals and terrestrial invertebrates where they are predicted to serve as beneficial symbionts. Here, we interrogated eight metagenome-assembled genomes (MAGs) of the latrunculid-associated spirochetes and found that these spirochetes are phylogenetically distinct from all invertebrate-associated spirochetes. The symbiosis between the spirochetes and their host appears to have been established relatively recently and may be based on the spirochetes' ability to produce methionine. Additionally, it was found that the dysbiosis of two spirochete strains is closely correlated with a switch between two

chemotypes exhibited by these sponges, suggesting that they may play a role in shaping the chemical profile of the sponges.

INTRODUCTION

The development of symbiotic relationships with prokaryotes likely predates the emergence of marine sponges (phylum Porifera) during the Cambrian explosion ~540 million years ago (1, 2) and these associations have played a critical role in the evolution of modern sponge taxa (3, 4). Bacterial symbionts have co-evolved with their host to perform specific, specialized services that promote the health and fitness of the host (5). The symbionts are involved in nitrogen, sulfur, and phosphorus cycling (6–9), carbon cycling, detoxification (10, 11) and in some cases, the production of bioactive secondary metabolites as chemical defenses against pathogens, predators, and competitors (12, 13). In return, the host provides its symbionts with a safe and nutrient-rich environment that promotes the fitness and survival of the symbiont (14). The taxonomic and functional diversity of sponge-associated microbiomes is generally host-specific, distinct from the surrounding water column, and acquired by recruitment and enrichment from the environment (5, 15, 16). However, there are a small number of specialized symbionts acquired by vertical inheritance from the parent sponge that are broadly distributed across phylogenetically distant sponge hosts (17, 18), including the Poribacteria, the “sponge-associated unclassified lineage” (SAUL), and the recently-discovered Tethybacterales symbionts (15, 19, 20).

The Tethybacterales represent a clade of cosmopolitan sponge-associated symbionts, comprising three families, namely the *Candidatus* Persebacteraceae, *Candidatus* Tethybacteraceae, and *Candidatus* Polydorabacteraceae (17, 20). As with the Poribacteria and Desulfobacteria, the Tethybacterales symbionts are present in phylogenetically diverse taxa that are primarily low-microbial abundance (LMA) sponge species but these bacteria have also been detected in some high-microbial abundance (HMA) species (17, 20). Characterization of MAGs of different species of the three Tethybacterales families and their associated hosts also indicates that there were multiple acquisition events and that host adaptation and co-evolution began after each acquisition event (17).

Sponges of the family Latrunculiidae (Demospongiae, Poecilosclerida) are known to be prolific producers of cytotoxic pyrroloiminoquinone alkaloid compounds (21–26) with pharmaceutical potential (Reviewed in Kalinksi et al., 2022 (27)). It has recently been discovered that there are two chemotypes present in the *Tsitsikamma favus* and *Tsitsikamma michaeli* latrunculid sponges (21, 28). Latrunculids are LMA sponges with highly conserved microbiomes that are dominated by Tethybacterales and Spirochete taxa (22, 29). The *Tsitsikamma favus* microbiome is dominated by two sponge-specific bacterial species defined by their 16S rRNA gene sequence, clones Sp02-1 and Sp02-3. The Sp02-1 symbiont has been recently characterized (17) and is classified as *Ca.* Ukwabelana africanus, a member of the *Ca.* Persebacteraceae family within the Tethybacterales (17). The *Ca.* U. africanus symbiont is phylogenetically related to

symbionts in sponges across multiple orders within the Demospongiae and may be involved in the reduction of nitrogen and sulfur in the sponge holobiont (17).

Unlike *Ca.* U. africanus (Sp02-1), the co-dominant spirochete (Sp02-3) is not representative of a globally distributed, broad-host range sponge symbiont. Spirochetes have been reported as minor members of several sponge microbiomes (30–32), but numerically dominant populations of spirochetes in sponges have only been reported in Latrunculiidae species endemic to the southeastern coast of South Africa, and the distantly related *Clathrina clathrus* (Calcarea, Clathrinida) (33). In addition, spirochetes, presumed to be symbionts, have been detected in the embryonic and larval cells of the marine sponge *Mycale laevis*, but their role is currently unknown (34, 35). Numerically dominant spirochete species are, however, present in several other marine invertebrates including sea anemones (36) and sea stars (37, 38) where decreased abundance of certain spirochete populations correlates with an increased incidence of disease (38). Spirochaeta symbionts are also commonly present as dominant populations in corals (39–42) and in termite guts (43), where they may be involved in the fixation of carbon or nitrogen (41). A recent study investigating the association between coral hosts and their associated microbiota found that Spirochaeta were most abundant in the coral skeleton, hypothesizing that they may be involved in coral skeletal development due to their ability to fix carbon and nitrogen (44).

The aim of the present study was to understand the relationship between latrunculid sponges and the Sp02-3 spirochete symbiont. Here we report the characterization of eight

spirochete MAGs from four *Tsitsikamma* sponge species and use comparative genomics to shed light on factors that may drive their conservation. Comparative analysis relative to publicly available genomes and MAGs of the Spirochaetaceae family suggests that the Sp02-3 spirochetes are relatively new symbionts that are distinct from all other sponge-associated spirochetes and a dysbiosis of two closely related spirochete strains correlates with a change in host sponge chemotype.

RESULTS

Characterization of *Tsitsikamma* sponge-associated spirochete MAGs

Previous studies identified two closely related spirochete species, Sp02-3 and Sp02-15, in the *T. favus* microbiome (22). Subsequently, the Sp02-3 symbiont was shown to be present in the microbiomes of other *Tsitsikamma* species and *Cyclacanthia bellae* (29). Our aim in this study was to characterize the genome of the Sp02-3 symbiont to better understand its role in the sponge holobiont. Eight sponges including five *T. favus* specimens (TIC2015-050A, TIC2015-050C, TIC2018-003B, TIC2018-003D, TIC2018-003M) and one each of *T. michaeli* (TIC2019-013N), *T. madiba* (TIC2022-009), and *T. pedunculata* (TIC2022-059) were selected for metagenomic analysis (Table S1). Following assembly, binning and taxonomic classification, eight spirochete MAGs were identified, one from each of the eight *Tsitsikamma* sponge metagenomes: MAGs 050A_2, 050C_7, 003B_7, 003D_7, 003M_1, 059_1, 013N_1, and 009_1 (Table 1, Table S2). The 16S rRNA and 23S rRNA gene sequences from each MAG (if recovered) were aligned against the NR nucleotide database via online BLASTn (45).

Table 1. Characteristics of putative representative genomes of *Tsitsikamma* sponge-associated spirochete symbiont MAGs

MAG	Size (Mbp)	Quality	16S rRNA (% ID)	23S rRNA (% ID)	Host	Sponge
003B_7	1.97	Medium	N/A	*Salinispira pacifica* L21-RPul-D2 (89.54%)	*T. favus*	TIC2018-003B
050A_2	2.73	Low	Uncultured marine clone Sp02-3 (99.52%)	*Salinispira pacifica* L21-RPul-D2 (89.54%)	*T. favus*	TIC2016_050A
003D_7	2.48	High	Uncultured marine clone Sp02-3 (99.52%)	*Salinispira pacifica* L21-RPul-D2 (89.58%)	*T. favus*	TIC2018-003D
003M_1	2.74	High	N/A	*Salinispira pacifica* L21-RPul-D2 (89.58%)	*T. favus*	TIC2018-003M
050C_7	1.72	Medium	Uncultured marine clone Sp02-3 (99.52%)	N/A	*T. favus*	TIC2016-050C
009_1	1.47	High	N/A	*Salinispira pacifica* L21-RPul-D2 (91.25%)	*T. madiba*	TIC2022-009
013N_1	2.33	High	N/A	*Salinispira pacifica* L21-RPul-D2 (89.48%)	*T. michaeli*	TIC2019-013N
059_1	2.04	Medium	N/A	N/A	*T. pedunculata*	TIC2022-059

The 16S rRNA gene sequences recovered from three MAGs all showed the greatest sequence identity with "Uncultured marine clone Sp02-3, " representing the conserved spirochete symbiont previously identified in *T. favus* sponges (22, 29). All recovered 23S rRNA sequences shared the greatest sequence similarity with *S. pacifica* L21-RPul-D2. This *S. pacifica* strain, isolated from a hypersaline microbial mat (46), was previously shown to be the closest known relative of the conserved spirochete Sp02-3 OTU (22, 29). Finally, all eight *Tsitsikamma*-associated spirochete MAGs were taxonomically classified, via GTDB-Tk (47), within the *Salinispira* genus (Table S2). Therefore, we were confident these MAGs represented the conserved spirochete symbiont (Sp02-3) previously reported in South African latrunculid sponges.

Phylogeny of *Tsitsikamma* sponge-associated spirochete MAGs

The 16S rRNA gene sequences recovered from three of the *Tsitsikamma*-associated spirochete MAGs were aligned against their closest matches in the NR database, and spirochetes from other marine invertebrates (37, 39), including the dominant spirochete present in the distantly related *Clathrina clathrus* sponges (33). Inferred maximum-likelihood hood phylogeny from the alignment which showed that the *Tsitsikamma*-associated spirochete MAGs were distinct from all other invertebrate-associated spirochetes (Fig. 1). Intriguingly, a clone (Sp02sw36) isolated from the seawater extruded from *Tsitsikamma favus* sponges (22), was a close relative of the spirochetes associated with crown-of-thorns starfish (37), and the dominant spirochete found in *C. clathrus* sponges (33). The *Tsitsikamma*-associated spirochete MAGs formed a distinct clade but were most closely related to spirochetes detected in non-host-associated environments including hypersaline microbial mats, seawater, estuary water, and volcanic mud.

As phylogeny inferred by a single marker gene can be limited, several orthogonal approaches were used to assess the phylogeny of the *Tsitsikamma* sponge-associated spirochete symbionts using whole genome data. Initially, we employed autoMLST (48) in *de novo* mode, with both concatenated alignment and coalescent tree approaches, using ten MAGs/genomes acquired from other sponge hosts: *Rhopaloides odorabile, Ircinia ramosa,* and *Aplysina aerophoba* from other studies (49–51) as reference. The resultant phylogenies from these two approaches had largely congruent topologies, with the *Tsitsikamma* sponge-associated Sp02-3 symbionts and other sponge-associated spirochetes forming two related, but distinct clades (Fig. 2). The closest relative of the

Tsitsikamma-associated spirochetes was *Salinispira pacifica*, in agreement with the 16S rRNA gene phylogeny. The *Tsitsikamma*-associated spirochetes appeared phylogenetically clustered following their respective hosts. This contrasted with other sponge-associated spirochetes that did not seem to follow any discernible pattern of possible co-phylogeny or phylosymbiosis (Fig. 2). As an orthogonal phylogenetic approach, we generated a phylogenetic tree using Phylophlan3 (52) and RaxML (53) (Fig. S1). Along with the eight *Tsitsikamma*-associated spirochete genomes and the ten genomes of spirochetes associated with other sponges, we included all Spirochaetaceae genomes from the NCBI database (N=300) and all host-associated spirochete MAGs from the JGI database (N=44). Again, the *Tsitsikamma*-associated spirochetes formed a clade distinct from all other sponge-associated spirochete genomes. Additionally, in this analysis, we found that a spirochete isolated from seawater (GCA 913043885.1) clustered with the other sponge-associated spirochetes and assumed this genome as the representative of the closest free-living relative of this clade. Finally, we calculated pairwise average nucleotide identity (ANI) scores for all 363 spirochete genomes (Table S3). The *Tsitsikamma*-associated spirochetes shared between 93.9% to 98.2% ANI with each other (Table S4), and less than 75% ANI with any other spirochete, including their closest relative *S. pacifica.*

Estimated evolutionary divergence patterns of sponge-associated spirochetes

The divergence pattern of all sponge-associated spirochetes and their closest known free-living relatives was estimated using their rate of synonymous substitutions (dS). Visualization of the pairwise dS revealed that the *Tsitsikamma*-associated spirochetes

are evolutionarily divergent from even their closest relative, *S. pacifica* (Fig. 3). It appears that the other sponge-associated spirochetes began diverging before the *Tsitsikamma*-associated spirochetes diverged from their free-living relative. The presence of a planktonic relative (GCA_913043885.1) within this clade suggests that either the spirochetes associated with the *A. aerophoba*, *R. odarabile,* and *I. ramosa* may be transient bacteria, or that these symbionts are acquired horizontally from the environment. In contrast, the divergence pattern of the *Tsitsikamma*-associated spirochetes is congruent with the phylogeny of their sponge host, suggestive of phylosymbiosis. Finally, it appears that these spirochetes have only recently begun diverging from one another as they adapt to their sponge host. Additionally, their association with latrunculid sponges is more recent than the sympatric and co-dominant Tethybacterales similarly conserved in these sponges (17).

Comparative analysis of functional potential in spirochete genomes

The functional potential for all 363 spirochete genomes was predicted by assigning KEGG Orthologs (KO) annotations using KofamScan (54). KO counts per genome were mapped back to associated pathways detailed in the KEGG database (55) (Table S5). Dimension reduction of these counts per genome revealed distinct clusters indicative of adaptation to the various environments from which these spirochetes were acquired (Fig. 4). The functional potential of the *Tsitsikamma*-associated spirochetes was distinct from spirochetes associated with other sponges and interestingly, was clustered more closely with the functional potential of spirochetes associated with oligochaete worms and spirochetes from hypersaline lake environments (Fig. 4).

An Analysis of Similarity (ANOSIM) of the same data (Table S6) showed that the functional gene repertoire of the *Tsitsikamma*-associated spirochetes and other sponge-associated spirochetes were significantly different ($p < 0.05$) from one another and from all other environments. However, when considering the associated R-values, the *Tsitsikamma*-associated spirochetes may exhibit some overlap in functional potential of spirochetes in hypersaline lakes (R = 0.26), sediment (R = 0.31), freshwater lakes (R = 0.38), termites (R = 0.47), and seawater (R=0.49). This suggests that the functional repertoire of *Tsitsikamma*-associated spirochetes may be more akin to free-living species than host-associated. This is in agreement with the divergence analysis (Fig. 3), in that it is suggestive of a relatively recent establishment of symbiosis.

Without transcriptomic data, it is impossible to definitively identify the characteristic roles of the spirochetes in their hosts. However, we looked to gain some insight using the data at hand by repurposing an Indicator Species Analysis using the gene counts of KEGG-annotated genes per genome (Table S7). Here we found that three KEGG-annotated genes were significantly associated with *Tsitsikamma* sponge-associated spirochetes: K21029 (*moeB*; molybdopterin-synthase adenylyltransferase), K00641 (metX; homoserine-O-acetyltransferase), and K00325 (pntB; proton-translocating NAD(P)+ transhydrogenase subunit beta). Both MoeB and PntB enzymes are involved in processes that handle oxidative stress (56, 57). The absence of oxidative phosphorylation genes in the genomes of sponge-associated spirochetes (Table S6) suggests that the spirochetes are likely either anaerobes or facultative anaerobes like their closest free-

living relative, *S. pacifica* (46). The significant association of *moeB* and *pntB* genes in the *Tsitsikamma*-associated spirochetes may reflect the pressure of maintaining favorable redox conditions for this putative anaerobe. The MetX enzyme is involved in the biosynthesis of methionine, where it serves as an alternative to MetA, where it activates homoserine via acetylation (via succinylation for MetA), which is the first step in the biosynthesis of methionine (58–60). A detailed inspection of the presence of this gene, in all spirochete genomes used in this study, revealed that metX was present and metA was absent in all sponge-associated spirochetes. MetA and MetX enzymes are used in trans- and direct-sulfurylation, respectively, for the assimilation of sulfur during methionine biosynthesis. During trans-sulfurylation (MetA) sulfur is cysteine-derived but during direct-sulfurylation (MetX) sulfur is hydrogen sulfide-derived (61). A key distinction between these two methionine biosynthesis routes is that the trans-route via MetA is strictly regulated and feedback-inhibited, whilst the direct route via MetX, present in sponge-associated spirochetes, lacks such regulation (58, 61, 62). While purely speculative at this point, the unregulated production of methionine by the sponge-associated spirochete may be of benefit to the host sponge (63, 64), which cannot produce this essential amino acid (65–67).

The biosynthetic potential of Sp02-3 spirochetes

We have previously speculated that the Sp02-3 symbiont may be involved in the production of the pyrroloiminoquinone compounds characteristic of latrunculid sponges (29) and therefore sought to investigate the biosynthetic potential of the spirochete MAGs. A total of 581 biosynthetic gene clusters (BGCs) were detected in all spirochete genomes

(N=363) (Table S8) and clustered into gene cluster families (GCFs) at a maximum distance of 0.3 with BiG-SCAPE (68) (Fig. 5A). Six of the eight *Tsitsikamma*-associated spirochetes had only a single predicted BGC. The remaining two MAGs, 003B_7 and 050A_2, had no detected BGCs. All six BGCs were predicted to encode a terpene product and were clustered into a single GCF (GCF1). Three other GCFs (GCFs 2, 3, and 4), consisting of terpene BGCs from other sponge-associated spirochetes, were identified but did not appear to have any homology with the terpene BGC in the *Tsitsikamma*-associated Sp02-3 spirochetes (Fig. 5B). Additional BiG-SCAPE analyses were run with less stringent cutoffs of 0.5 and 0.8, and no BGCs from other spirochete genomes or the MiBIG database were incorporated into a GCF with the terpene BGCs detected in the *Tsitsikamma*-associated spirochetes (Table S8), indicating that this BGC is likely novel. Nonetheless, the closest characterized relative of the *Tsitsikamma*-associated spirochetes, *S. pacifica*, produces an orange carotenoid-like pigment, which was hypothesized to protect against oxidative stress (46). The terpene BGC present in the *Tsitsikamma*-associated spirochetes may encode a compound with an analogous function.

Bacterial community profiles and sponge chemotypes

Previously, the Sp02-3 symbiont was shown to be the co-dominant OTU in sponge species from two latrunculid genera (*Tsitsikamma* and *Cyclacanthia*) but not detected in *Latrunculia* sponges (22, 28, 29). We profiled the microbiomes of additional latrunculid species, analyzing 16S rRNA gene amplicons from 75 sponges and 8 seawater samples (Table S1). Clustering the sequences at a distance of 0.03 resulted in 5909 Operational

Taxonomic Units (OTUs) (Table S9). Inspection of the OTU distribution showed that the microbiomes of the latrunculid sponges were distinct from the surrounding seawater and from one another, (Fig. S2), as observed in previous studies (29). Manual inspection of $OTU3_{0.03}$, classified as a spirochete, displayed equal sequence identity (98.45%) to spirochete 16S rRNA gene clones, Sp02-3 and Sp02-15. This spirochete OTU was present in all sponge specimens included in this analysis but was only dominant in *T. favus*, *T. nguni,* and *Cyclacanthia bellae* sponges (Fig. S2).

To disentangle the ambiguity of $OTU3_{0.03}$, we opted to investigate OTUs clustered at a more granular distance of 0.01. This resulted in 16338 OTUs (Table S10) of which the Tethybacterales populations were most dominant, i.e, $OTU1_{0.01}$ was the most dominant OTU in *T. favus*, *T. nguni,* and *T. pedunculata* sponges, $OTU2_{0.01}$ was most dominant in *T. madiba*, *T. michaeli,* and *C. bellae* sponges, and $OTU7_{0.01}$ was most dominant in *Latrunculia apicalis* sponges (Fig. 6A). Four spirochete populations, namely $OTU3_{0.01}$, $OTU6_{0.01}$, $OTU8_{0.01}$, and $OTU18_{0.01}$, were detected. Classification of the OTUs showed $OTU6_{0.01}$ was most closely related to the Sp02-15 clone and the other three were most closely related to the Sp02-3 clone. $OTU3_{0.01}$ was the dominant spirochete in *T. favus* and *T. nguni* sponges, while $OTU18_{0.01}$ was dominant in *C. bellae* sponges. $OTU8_{0.01}$ was detected in all latrunculid sponges (Table S10) but not dominant in any species.

Additionally, we know from previous studies (21, 27, 28) that the chemical profiles associated with latrunculid sponge species are distinct. Assessment of the chemical profiles of the sponges in this study showed that the distribution of pyrroloiminoquinone

alkaloids is species-specific (Fig. 6B). The exceptions to this are *T. favus* and *T. michaeli* Chemotype II sponges (Fig. 6B, indicated with asterisks), which show sponge-specific profiles that are distinct from the Chemotype I sponges within those sponge species.

We noted that the abundance of $OTU6_{0.01}$ (Sp02-15) appeared to be increased in Chemotype II sponges relative to Chemotype I sponges (Fig. 7A). We conducted an Indicator Species Analysis and found that $OTU6_{0.01}$ (Sp02-15), along with eight other OTUs most closely related to Sp02-15, had significantly increased ($p < 0.05$) abundances in Chemotype II sponges (Fig. 7B, Table S11). Additionally, we conducted a correlation analysis of the top 50 most abundant OTUs with relative pyrroloiminoquinone abundance per sponge sample (Fig. 7C, Table S12). The Sp02-3 spirochetes were primarily correlated with the presence of tsitsikammamines, except for $OTU18_{0.01}$ the dominant spirochete population in *C. bellae* sponges, which was correlated with the presence of discorhabdin I and makaluvamines. The Sp02-15 $OTU6_{0.01}$ population was positively correlated with the presence of makaluvamines, bromo-makaluvamines, bromo-orthoquinones, hydroxy-discorhabdins, and dibromo-discorhabdins. Makaluvamines, bromo-makaluvamines, and bromo-orthoquinones are characteristic of *T. favus* Chemotype II sponges, while hydroxy-discorhabdins and dibromo-discorhabdins are characteristic of *T. michaeli* Chemotype II sponges. Conversely, the presence of $OTU6_{0.01}$ is negatively correlated with the presence of tsitsikammamines, dihydro-discorhabdins, tribromo-disorhabdins, and A-series discorhabdins. The tribromo-disorhabdins and tsitsikammamines are characteristic of *T. favus* Chemotype I while *T. michaeli* Chemotype I sponges (Fig. 6) produce mostly tribromo-discorhabdins and no

tsitsikammamines. No other OTU exhibited a pattern of correlation that is associated with the change in chemotype. Simple makaluvamines were correlated with $OTU16_{0.01}$ (algae), $OTU31_{0.01}$ (Verrucomicrobia), $OTU35_{0.01}$ (Cyanobacteria), $OTU36_{0.01}$ (Planctomycete), $OTU39_{0.01}$ (Ilumatobacteraceae), and $OTU46_{0.01}$ (Chlamydiaceae). These six OTUs were generally present in low abundances but were most abundant in *C. bellae* sponges. These strains are of interest for further investigation to identify the biosynthetic gene cluster producing makaluvamine scaffolds in latrunculid sponges, which are considered to be biosynthetic precursors to the more complex tsitsikammamines and discorhabdins (27). Additionally, we noted the correlation of dominant Tethybacterales populations with various discorhabdin classes leading us to consider these broad-host range symbionts may also play a role in shaping the chemodiversity of latrunculid sponges. Metatranscriptomic studies are currently underway to gain further clarity on this ecological puzzle.

Conclusion: This study shows that the conserved Sp02-3 spirochete (*Ca.* Latrunculis umkhuseli) of latrunculid sponges appears to be a relatively new symbiont that has begun co-evolving with its respective sponge hosts and that genome erosion has begun. *Ca.* Latrunculis umkhuseli is distinct from all other invertebrate-associated spirochetes, including non-dominant spirochetes associated with other marine sponges. Assessment of their functional potential suggests that the Sp02-3 spirochetes may be maintained in the latrunculid hosts to produce methionine. We found no evidence that they are directly involved in the production of the pyrroloiminquinones, but the dysbiosis between the Sp02-3 and Sp02-15 spirochete strains is correlated with the change in chemotype

suggesting that the spirochete populations may either be responding to the change in pyrroloiminoquinone profile or are influencing the pyrroloiminoquinone profile through impact on either the sponge host itself or the yet undiscovered pyrroloiminoquinone producer. The Sp02-3 and Sp02-15 spirochete strains are likely closely related and therefore it is possible that the recovered Sp02-3 MAGs are a hybrid of these strains, a limitation imposed at both the assembly and binning steps of the workflow described here. We are currently moving forward with additional deep sequencing along with meta-transcriptomics and hope that we may be able to tease apart these two strains and their roles in the latrunculid sponges.

METHODS AND MATERIALS

Sponge Collection and taxonomic identification.

Sponges were collected by SCUBA or Remotely Operated Vehicle (ROV) from multiple locations within the Tsitsikamma Marine Protected Area, Algoa Bay (Port Elizabeth), the Amathole Marine Protected Area (East London), and the Garden Route National Park. 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 of 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; 2019: RES2019/13; 2020: RES2020/31; 2021: RES2021/81; 2022: RES2022/70. Collection metadata are provided in Table S1. Sponge specimens were stored on ice during collection and moved to -20 °C on return to the lab. Subsamples of each sponge, collected for DNA extraction, were

preserved in RNALater (Invitrogen) and stored at -20 °C. Sponge specimens were identified through inspection of gross morphology, spicule analysis, and molecular barcoding, as performed previously (21, 28, 29, 69).

Chemical Analysis and Chemotype Identification

Sponge extracts were prepared by extraction with methanol, drying *i. vac.* and resuspension in methanol at 1-10 mg/mL. LC-MS/MS data was acquired on a Bruker ESI-Q-TOF Compact (Bruker, Bremen) in positive ionization mode coupled to a Dionex Ultimate3000 Chromatograph (ThermoScientific, Sunnyvale, CA, USA) and using reversed-phase C18 columns and mobile phases consisting of water and acetonitrile with 0.1% formic acid each, using one of two methods (see Supplementary Methods for details). The data was converted to mzXML format and analyzed using MZmine3 (70) to assemble an aligned feature list (see Supplementary Methods for details). The feature list was filtered based on comparison of *m/z* values and MS/MS spectra to known or putative pyrroloiminoquinones. Peak area values were normalized to the overall pyrroloiminoquinone signal per sample and aggregated to the pyrroloiminoquinone class to summarize the latrunculid pyrroloiminoquinone profiles.

Bacterial community profiles in latrunculid sponges

The V4-V5 of the 16S rRNA gene was PCR amplified from 79 latrunculid sponges collected between 1994 and 2022 (See Table S1 for collection data). Amplicons were sequenced using the Illumina MiSeq platform and curated using mothur (v.1.48.0) (71). All raw amplicon read data can be accessed under accession number PRJNA508092.

Briefly, sequences that were shorter than 250 nt in length, longer than 350 nt in length, had homopolymeric runs of 7 nt or more, had ambiguous bases, or had a sliding window quality average lower than 20, were removed from the datasets. Chimeric sequences were detected using VSEARCH (72) and removed from the dataset. Sequences were then classified via alignment against the SILVA database (v138.1) and any sequences classified as “Chloroplast”, “Mitochondria", “unknown”, “Archaea”, or “Eukaryota” were removed. Sequences were clustered into Operational Taxonomic Units (OTUs) at distances of 0.03 and 0.01, and read counts thereof were converted to relative abundance (Tables S4 and S5 respectively). Representative sequences of each OTU were aligned against the nt prokaryotic database using standalone blastn (73), using parameters -max_hsps 1 -max_target_seqs 1 to return only the first match. Descriptions and isolation sources for each returned accession were retrieved using the esearch, efetch and xtract methods from the stand-alone entrez package (74).

Metagenomic sequencing and analysis of individual *T. favus* specimens.

The DNA extraction and metagenomic sequencing of four *Tsitsikamma favus* sponges that resulted in the recovery of four MAGs 050A_2, 050C_7, 003B_7, and 003D_7, classified as spirochetes, is described in Waterworth et al., 2021(17). In addition to these samples, four additional metagenomes of *Tsitsikamma* sponges (TIC2018-003M, TIC2019-013N, TIC2022-009, and TIC2022-059) were sequenced. These sponges were selected for sequencing based on the apparent abundance of spirochete OTUs found via 16S rRNA gene amplicon sequence.

Total genomic DNA (gDNA) was extracted using the Zymo Research Quick DNA Fecal/Soil Microbe Miniprep Kit (Catalog number: D6012) according to the manufacturer's specifications and stored at -4 °C. Shotgun metagenomic IonTorrent libraries of 200 bp reads were prepared and sequenced using an Ion P1.1.17 chip. All metagenomes were assembled, binned, and processed as described in Waterworth et al., 2021 (17). Four additional spirochete genome MAGs (003M_1, 059_1, 013N_1, and 009_1) were extracted from the new datasets. MAGs were named after the *Tsitsikamma* sponge specimen from which they were extracted (e.g. 050A_2 is the MAG from sponge specimen TIC2016-050A). The numbers associated with each MAG are an arbitrary artifact of the binning process.

Acquisition of reference genomes and MAGs

Four spirochete MAGs associated with *Aplysina aerophoba* and *Rhopaloeides odorabile sponges* from a study by Robbins and colleagues (75) were downloaded from https://data.ace.uq.edu.au/public/sponge_mags/, and five sponge-associated spirochete MAGs were acquired from the China National GeneBank DataBase (CNGBdb) from studies by O'Brien and colleagues (49, 50). One spirochete genome from an *Aplysina aerophoba* sponge was additionally downloaded from the NCBI database (GCA_002238925.1). Additionally, all other genomes classified within the Spirochaetaceae family were downloaded from the NCBI database (N=300) and all host-associated spirochete MAGs were downloaded from the JGI database (N=44). This resulted in a total of 354 reference genomes (Table S2).

Characterization of MAGs and genomes

All scripts used for bioinformatic analyses, and their associated inputs, used in the following methods can be found at https://github.com/samche42/Spirochete. All MAGs and genomes used in this study were assessed using CheckM (v1.1.3) (76) and taxonomically classified using GTDB-Tk (v2.3.2) (47) against the Release 214.1 reference database. Basic metrics such as size, number of contigs, and N50 were calculated using bin_summary.py. The number of genes, pseudogenes, and coding density per genome were calculated using all_included_genome_characteristics.py. All metadata per genome or MAG can be found in Table S2.

Phylogeny of spirochete genome MAGs extracted from individual *Tsitsikamma* sponges

Ribosomal sequences (23S rRNA, 16S rRNA, and 5S rRNA) were extracted from individual MAGs using barrnap (v 0.9) (77). The closest matches of recovered 16S sequences from sponge-associated MAGs were identified using BLASTn (v 2.7.1) (73). Resultant sequences were aligned using MUSCLE (v. 5.1) (78, 79) and phylogeny was inferred using the Maximum-likelihood method with 1000 bootstraps in MEGA11 (80). Phylogeny of the *Tsitsikamma*-associated spirochete MAGs was similarly inferred using whole genome data via autoMLST (48) and PhyloPhlan3 (52). Amino acid sequences and nucleotide sequences for all genes were found in all genomes using prokka (v 1.13) (81). The phylogeny of all 362 MAGs and genomes (8 *Tsitsikamma*-associated spirochete MAGs and 354 references) was inferred using Phylophlan3: Phylophlan3 was run with diversity set to medium, with default values in the supermatrix_aa configuration. The

resultant gene protein alignment was used in RaxML (v 8.2.12) (82) to build a phylogenetic tree with 1000 bootstrap replicates using the PROTGAMMAAUTO model. The resultant tree was visualized in iTol (83). Genomes from Myxococcota (GCA_002691025.1) and Deltaproteobacteria (GCA_020632655.1) were chosen as outgroups. These genomes had been downloaded from the NCBI database as their metadata indicated that they were classified within the Spirochaetaceae family. However, the taxonomic classification of these genomes with GTDB-Tk revealed that these genomes had likely been misclassified. These genomes were considered serendipitous choices for outgroups for the Phylophlan3 analysis. AutoMLST was deployed in *de novo* mode using concatenated alignments and coalescent trees of marker genes in two separate analyses. ModelFinder and IQ-TREE Ultrafast Bootstrap analysis were enabled in both analyses. All latrunculid-associated and other sponge-associated spirochete MAGs were included in this analysis. MAGs and genomes from JGI and NCBI were not used in this analysis as the number of query genomes is limited to 20 so we opted to include only sponge-associated spirochetes in this analysis. Resultant trees were downloaded in Newick format and revisualized in iTol (83). Finally, the pairwise average nucleotide identity (ANI) was calculated for all genomes using fastANI (v1.33)(84). If a pairwise alignment fraction (AF) was lower than 70% (85), the associated ANI score was nullified as the accuracy of the ANI score could not be trusted.

Estimated evolutionary divergence patterns of sponge-associated spirochetes

Using the Phylophlan3 (52) and autoMLST(48, 52) trees as guidance, orthologous genes from the eight *Tsitsikamma*-associated spirochetes, the ten other sponge-associated

spirochetes, and their closest relatives were identified using OMA (v. 2.6.0) (86). A total of 11 orthologs common to all genomes were found using count_OGs.py and aligned using MUSCLE (v 5.1) (78, 79). The corresponding nucleotide sequence for each gene was retrieved using streamlined_seqretriever.py, all stop codons were removed using remove_stop_codons.py, and nucleotide sequences were aligned using MUSCLE (v 5.1) (78, 79). Ortholog gene sequences were grouped per genome using merge_fasta_for_dNdS.py. The nucleotide and amino acid sequences (per genome) were each concatenated union function from EMBOSS (87) and aligned using PAL2NAL (88). The alignment was used to estimate pairwise synonymous substitution rates (dS) and thereby infer the pattern of divergence between these genomes using codeml from the PAML package (89).

Comparative analysis of functional potential in spirochete genomes

Genes were identified in all genomes/MAGs using Prokka (v 1.13) (81) and then annotated against the KEGG database using KOfamSCAN (54) with detail-tsv as the output format. Reliable annotations were extracted from these results based on the criteria that the annotation score is greater than the estimated threshold, and then reliable annotations per MAG/genome were counted and summarized using the kegg_parser.py script. This produced a table of KO counts per genome that was used as input for both Analysis of Similarity (ANOSIM.R) processing and dimension reduction, via UMAP (90), for 3-dimensional and 2-dimensional visualizations (dimension_reduction.py). A Jupyter notebook is provided in the GitHub repository for easy reproduction and an interactive 3D figure. To find statistically significant KEGG-annotated drivers of the different samples,

we performed a re-purposed Indicator Species Analysis with the number of KEGG annotations per KO per genome in place of OTU abundance. This was performed using the multiplatt method from the “indicspecies” package in R (91) with 1000 permutations and specifying the point biserial correlation coefficient (“r.g”) as the association index as this both accounts for abundance data (rather than presence/absence data) and corrects for the different number of samples per host type.

The biosynthetic potential of *Tsitsikamma*-associated spirochetes

A total of 547 biosynthetic gene clusters (BGCs) were predicted from all spirochete genomes (N=363) using antiSMASH (v. 6.0.1) (92) with --cb-general --cb-knownclusters --cb-subclusters --asf --pfam2go --smcog-trees options enabled and genes found with prodigal. The resultant putative BGCs were clustered twice using BiG-SCAPE (v 1.1.5)(68) at maximum distances of 0.3, 0.5, and 0.8. Network files of non-singleton gene cluster families (GCFs) were visualized in Cytoscape (93). Highlighted gene clusters of interest were visualized with clinker (94). Metadata for BGCs was extracted from individual GenBank files using antismash_summary.py.

Bacterial community dysbiosis analyses

Analysis of Similarity (ANOSIM) of the bacterial communities associated with the different sponge species and the surrounding seawater was performed for OTU datasets clustered at distances of 0.03 and 0.01 respectively, using the anosim method from the vegan (95) package in R (16S_ANOSIM.R). An Indicator species analysis was performed using the OTUs clustered at a distance of 0.01 for all *T. favus* and *T. micheali* sponges for which a

chemotype had been assigned (16S_Chemotype_Indicator_Species_Analysis.R) to determine which OTUs, if any, were associated with the two chemotypes. The co-correlation analysis of the 50 most abundant OTUs (found as an average across all samples) was performed using the 'cor' function (96) native to R using dataframes of OTU and compound abundances as input. A 16S rRNA gene sequence phylogeny was built from the representative sequences of the top 50 OTUs, aligned with MUSCLE (v 5.1) (78, 79), using the neighbor-joining approach with 1000 bootstraps in MEGA11 (80). The final tree was visualized in iTol (83) where the correlation matrix and the average OTU abundance per sponge species was visualized alongside the tree as datasets.

DATA AVAILABILITY

All sequence data can be accessed under accession number PRJNA508092 in the NCBI SRA database. All scripts used for analysis and visualization can be accessed at https://github.com/samche42/Spirochete. Raw MS data can be accessed under the ID MSV000094809 at massive.ucsd.edu.

ACKNOWLEDGEMENTS

We would like to acknowledge Gwynneth Matcher (South African Institute for Aquatic Biodiversity, Aquatic Genomics Research Platform), Carel van Heerden and Alvera Vorster (Stellenbosch University Central Analytical Facility) for next-generation sequencing technical support. We thank Ryan Palmer and Koos Smith (African Coelacanth Ecosystem Programme) for logistics and technical support during sponge collections. We thank the South African Environmental Observation Network, Elwandle

Coastal Node, and the Shallow Marine and Coastal Research Infrastructure for the use of their research platforms and infrastructure for their assistance in SCUBA collections and logistical support. This research was supported by South African National Research Foundation grants to R.A.D., including the South Africa Research Chair Initiative (SARChI) grant (UID: 87583) and the SARChI-led Communities of Practice Programme (UID: 110612). S.C.W. was supported by an NRF Innovation and Rhodes University Henderson Ph.D. scholarships. G.M.S. and L.M. were supported by NRF Masters and PhD scholarships, respectively. S.P.-N. Was supported by an NRF PDP scholarship (UID: 101038). J.C.J.K. was supported by funding awarded to R.A.D. by the South African Medical Research Council as well as the UK Medical Research Council, with funds received from the UK Government's Newton Fund (Grant No.: 96185). We declare no competing interests, financial or otherwise, in relation to the work described here. The 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.

REFERENCES

1. Botting JP, Nettersheim BJ. 2018. Searching for sponge origins. Nat Ecol Evol.

2. Botting JP, Muir LA. 2018. Early sponge evolution: A review and phylogenetic framework. Palaeoworld 27:1–29.

3. Wilkinson CR, Smith DC. 1997. Immunological evidence for the Precambrian origin of bacterial symbioses in marine sponges. Proceedings of the Royal Society of London Series B Biological Sciences 220:509–518.

4. Hentschel U, Hopke J, Horn M, Friedrich AB, Wagner M, Hacker J, Moore BS. 2002. Molecular evidence for a uniform microbial community in sponges from different oceans. Appl Environ Microbiol

68:4431–4440.

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

6. Jensen S, Fortunato SAV, Hoffmann F, Hoem S, Rapp HT, Øvreås L, Torsvik VL. 2017. The Relative Abundance and Transcriptional Activity of Marine Sponge-Associated Microorganisms Emphasizing Groups Involved in Sulfur Cycle. Microb Ecol 73:668–676.

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

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

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

10. de Voogd NJ, Cleary DFR, Polónia ARM, Gomes NCM. 2015. Bacterial community composition and predicted functional ecology of sponges, sediment and seawater from the thousand islands reef complex, West Java, Indonesia. FEMS Microbiol Ecol 91.

11. Hentschel U, Usher KM, Taylor MW. 2006. Marine sponges as microbial fermenters. FEMS Microbiol Ecol 55:167–177.

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

13:e0197617.

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

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

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

16. Wang Y, Gong L, Gao Z, Wang Y, Zhao F, Fu L, Li X. 2023. Host-specific bacterial communities associated with six cold-seep sponge species in the South China Sea. Front Mar Sci 10.

17. Waterworth SC, Parker-Nance S, Kwan JC, Dorrington RA. 2021. Comparative genomics provides insight into the function of broad-host range sponge symbionts. MBio https://doi.org/10.1128/mbio.01577-21.

18. Fieth RA, Gauthier M-EA, Bayes J, Green KM, Degnan SM. 2016. Ontogenetic changes in the bacterial symbiont community of the tropical demosponge Amphimedon queenslandica: Metamorphosis is a new beginning. Front Mar Sci 3.

19. Fieseler Lars, Horn Matthias, Wagner Michael, Hentschel Ute. 2004. Discovery of the Novel Candidate Phylum “Poribacteria” in Marine Sponges. Appl Environ Microbiol 70:3724–3732.

20. Taylor JA, Palladino G, Wemheuer B, Steinert G, Sipkema D, Williams TJ, Thomas T. 2021. Phylogeny resolved, metabolism revealed: functional radiation within a widespread and divergent clade of sponge symbionts. ISME J 15:503–519.

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

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

23. Hooper GJ, Davies-Coleman MT, Kelly-Borges M, Coetzee PS. 1996. New alkaloids from a South African latrunculid sponge. Tetrahedron Lett 37:7135–7138.

24. Antunes EM, Copp BR, Davies-Coleman MT, Samaai T. 2005. Pyrroloiminoquinone and related metabolites from marine sponges. ChemInform 36.

25. Botić T, Defant A, Zanini P, Žužek MC, Frangež R, Janussen D, Kersken D, Knez Ž, Mancini I, Sepčić K. 2017. Discorhabdin alkaloids from Antarctic Latrunculia spp. sponges as a new class of cholinesterase inhibitors. Eur J Med Chem 136:294–304.

26. Ford J, Capon RJ. 2000. Discorhabdin R: a new antibacterial pyrroloiminoquinone from two latrunculiid marine sponges, Latrunculia sp. and Negombata sp. J Nat Prod 63:1527–1528.

27. Kalinski J-CJ, Polyzois A, Waterworth SC, Siwe Noundou X, Dorrington RA. 2022. Current Perspectives on Pyrroloiminoquinones: Distribution, Biosynthesis and Drug Discovery Potential. Molecules 27.

28. Kalinski J-CJ, Waterworth SC, Noundou XS, Jiwaji M, Parker-Nance S, Krause RWM, McPhail KL, Dorrington RA. 2019. Molecular Networking Reveals Two Distinct Chemotypes in Pyrroloiminoquinone-Producing Tsitsikamma favus Sponges. Mar Drugs 17.

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

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

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

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

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

34. Díez-Vives C, Koutsouveli V, Conejero M, Riesgo A. 2022. Global patterns in symbiont selection and transmission strategies in sponges. Front Ecol Evol 10.

35. Riesgo A, Taboada S, Sánchez-Vila L, Solà J, Bertran A, Avila C. 2015. Some like it fat: comparative ultrastructure of the embryo in two demosponges of the genus Mycale (order Poecilosclerida) from Antarctica and the Caribbean. PLoS One 10:e0118805.

36. Bonacolta AM, Connelly MT, Rosales SM, Del Campo J, Traylor-Knowles N. 2021. The starlet sea anemone, Nematostella vectensis, possesses body region-specific bacterial associations with spirochetes dominating the capitulum. FEMS Microbiol Lett 368.

37. Wada N, Yuasa H, Kajitani R, Gotoh Y, Ogura Y, Yoshimura D, Toyoda A, Tang S-L, Higashimura Y, Sweatman H, Forsman Z, Bronstein O, Eyal G, Thongtham N, Itoh T, Hayashi T, Yasuda N. 2020. A ubiquitous subcuticular bacterial symbiont of a coral predator, the crown-of-thorns starfish, in the Indo-Pacific. Microbiome 8:123.

38. Høj L, Levy N, Baillie BK, Clode PL, Strohmaier RC, Siboni N, Webster NS, Uthicke S, Bourne DG. 2018. Crown-of-thorns sea star Acanthaster cf. Solaris has tissue-characteristic microbiomes with potential roles in health and reproduction. Appl Environ Microbiol 84.

39. van de Water JAJM, Melkonian R, Junca H, Voolstra CR, Reynaud S, Allemand D, Ferrier-Pagès C. 2016. Spirochaetes dominate the microbial community associated with the red coral Corallium

rubrum on a broad geographic scale. Sci Rep 6:27277.

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

41. Lawler SN, Kellogg CA, France SC, Clostio RW, Brooke SD, Ross SW. 2016. Coral-Associated Bacterial Diversity Is Conserved across Two Deep-Sea Anthothela Species. Front Microbiol 7:458.

42. Park JS, Han J, Suh S-S, Kim H-J, Lee T-K, Jung SW. 2022. Characterization of bacterial community structure in two alcyonacean soft corals (Litophyton sp. and Sinularia sp.) from Chuuk, Micronesia. Coral Reefs 41:563–574.

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

44. Ricci F, Tandon K, Black JR, Lê Cao K-A, Blackall LL, Verbruggen H. 2022. Host traits and phylogeny contribute to shaping coral-bacterial symbioses. mSystems 7:e0004422.

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

46. Ben Hania W, Joseph M, Schumann P, Bunk B, Fiebig A, Spröer C, Klenk H-P, Fardeau M-L, Spring S. 2015. Complete genome sequence and description of Salinispira pacifica gen. nov., sp. nov., a novel spirochaete isolated form a hypersaline microbial mat. Stand Genomic Sci 10:7.

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

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

49. O'Brien PA, Andreakis N, Tan S, Miller DJ, Webster NS, Zhang G, Bourne DG. 2021. Testing

cophylogeny between coral reef invertebrates and their bacterial and archaeal symbionts. Mol Ecol 30:3768–3782.

50. O'Brien PA, Tan S, Frade PR, Robbins SJ, Engelberts JP, Bell SC, Vanwonterghem I, Miller DJ, Webster NS, Zhang G, Bourne DG. 2023. Validation of key sponge symbiont pathways using genome-centric metatranscriptomics. Environ Microbiol 25:3207–3224.

51. Slaby BM, Hackl T, Horn H, Bayer K, Hentschel U. 2017. Metagenomic binning of a marine sponge microbiome reveals unity in defense but metabolic specialization. ISME J 11:2465–2478.

52. Asnicar F, Thomas AM, Beghini F, Mengoni C, Manara S, Manghi P, Zhu Q, Bolzan M, Cumbo F, May U, Sanders JG, Zolfo M, Kopylova E, Pasolli E, Knight R, Mirarab S, Huttenhower C, Segata N. 2020. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nat Commun 11:2500.

53. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35:4453–4455.

54. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. 2020. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics 36:2251–2252.

55. Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30.

56. Mendel RR. 2013. The molybdenum cofactor. J Biol Chem 288:13165–13172.

57. Zhang Q, Padayatti PS, Leung JH. 2017. Proton-Translocating Nicotinamide Nucleotide Transhydrogenase: A Structural Perspective. Front Physiol 8:1089.

58. Bourhy P, Martel A, Margarita D, Saint Girons I, Belfaiza J. 1997. Homoserine O-acetyltransferase, involved in the Leptospira meyeri methionine biosynthetic pathway, is not feedback inhibited. J Bacteriol 179:4396–4398.

59. Leyn SA, Suvorova IA, Kholina TD, Sherstneva SS, Novichkov PS, Gelfand MS, Rodionov DA. 2014. Comparative genomics of transcriptional regulation of methionine metabolism in Proteobacteria. PLoS One 9:e113714.

60. Bastard K, Perret A, Mariage A, Bessonnet T, Pinet-Turpault A, Petit J-L, Darii E, Bazire P, Vergne-Vaxelaire C, Brewee C, Debard A, Pellouin V, Besnard-Gonnet M, Artiguenave F, Médigue C, Vallenet D, Danchin A, Zaparucha A, Weissenbach J, Salanoubat M, de Berardinis V. 2017. Parallel evolution of non-homologous isofunctional enzymes in methionine biosynthesis. Nat Chem Biol 13:858–866.

61. Gruzdev N, Hacham Y, Haviv H, Stern I, Gabay M, Bloch I, Amir R, Gal M, Yadid I. 2023. Conversion of methionine biosynthesis in Escherichia coli from trans- to direct-sulfurylation enhances extracellular methionine levels. Microb Cell Fact 22:151.

62. Ferla MP, Patrick WM. 2014. Bacterial methionine biosynthesis. Microbiology 160:1571–1584.

63. Shih JL, Selph KE, Wall CB, Wallsgrove NJ, Lesser MP, Popp BN. 2020. Trophic Ecology of the Tropical Pacific Sponge Mycale grandis Inferred from Amino Acid Compound-Specific Isotopic Analyses. Microb Ecol 79:495–510.

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

65. Guedes RLM, Prosdocimi F, Fernandes GR, Moura LK, Ribeiro HAL, Ortega JM. 2011. Amino acids biosynthesis and nitrogen assimilation pathways: a great genomic deletion during eukaryotes evolution. BMC Genomics 12 Suppl 4:S2.

66. Costa IR, Thompson JD, Ortega JM, Prosdocimi F. 2014. Metazoan remaining genes for essential amino acid biosynthesis: sequence conservation and evolutionary analyses. Nutrients 7:1–16.

67. Moeller FU, Herbold CW, Schintlmeister A, Mooshammer M, Motti C, Glasl B, Kitzinger K, Behnam F, Watzka M, Schweder T, Albertsen M, Richter A, Webster NS, Wagner M. 2023. Taurine as a key intermediate for host-symbiont interaction in the tropical sponge Ianthella basta. ISME J 17:1208–1223.

68. Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH, Parkinson EI, De Los Santos ELC, Yeong M, Cruz-Morales P, Abubucker S, Roeters A, Lokhorst W, Fernandez-Guerra A, Cappelini LTD, Goering AW, Thomson RJ, Metcalf WW, Kelleher NL, Barona-Gomez F, Medema MH. 2020. A computational framework to explore large-scale biosynthetic diversity. Nat Chem Biol 16:60–68.

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

70. Schmid R, Heuckeroth S, Korf A, Smirnov A, Myers O, Dyrlund TS, Bushuiev R, Murray KJ, Hoffmann N, Lu M, Sarvepalli A, Zhang Z, Fleischauer M, Dührkop K, Wesner M, Hoogstra SJ, Rudt E, Mokshyna O, Brungs C, Ponomarov K, Mutabdžija L, Damiani T, Pudney CJ, Earll M, Helmer PO, Fallon TR, Schulze T, Rivas-Ubach A, Bilbao A, Richter H, Nothias L-F, Wang M, Orešič M, Weng J-K, Böcker S, Jeibmann A, Hayen H, Karst U, Dorrestein PC, Petras D, Du X, Pluskal T. 2023. Integrative analysis of multimodal mass spectrometry data in MZmine 3. Nat Biotechnol 41:447–449.

71. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541.

72. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: A versatile open source tool for metagenomics. PeerJ 4:e2584.

73. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+:

architecture and applications. BMC Bioinformatics 10:421.

74. Maglott D, Ostell J, Pruitt KD, Tatusova T. 2005. Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res 33:D54-8.

75. Robbins SJ, Song W, Engelberts JP, Glasl B, Slaby BM, Boyd J, Marangon E, Botté ES, Laffy P, Thomas T, Webster NS. 2021. A genomic view of the microbiome of coral reef demosponges. ISME J 15:1641–1654.

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

77. Seemann T. 2013. barrnap 0.9: rapid ribosomal RNA prediction. Google Scholar.

78. Edgar RC. 2022. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. Nat Commun 13:6968.

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

80. Tamura K, Stecher G, Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol 38:3022–3027.

81. Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.

82. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

83. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296.

84. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114.

85. Gosselin S, Fullmer MS, Feng Y, Gogarten JP. 2022. Improving Phylogenies Based on Average Nucleotide Identity, Incorporating Saturation Correction and Nonparametric Bootstrap Support. Syst Biol 71:396–409.

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

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

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

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

90. McInnes L, Healy J, Saul N, Großberger L. 2018. UMAP: Uniform Manifold Approximation and Projection. J Open Source Softw 3:861.

91. De Cáceres M, Legendre P. 2009. Associations between species and groups of sites: indices and statistical inference. Ecology 90:3566–3574.

92. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res 49:W29–W35.

93. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504.

94. Gilchrist CLM, Chooi Y-H. 2021. Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. Bioinformatics https://doi.org/10.1093/bioinformatics/btab007.

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

96. Langfelder P, Horvath S. 2012. Fast R Functions for Robust Correlations and Hierarchical Clustering. J Stat Softw 46.

FIGURES

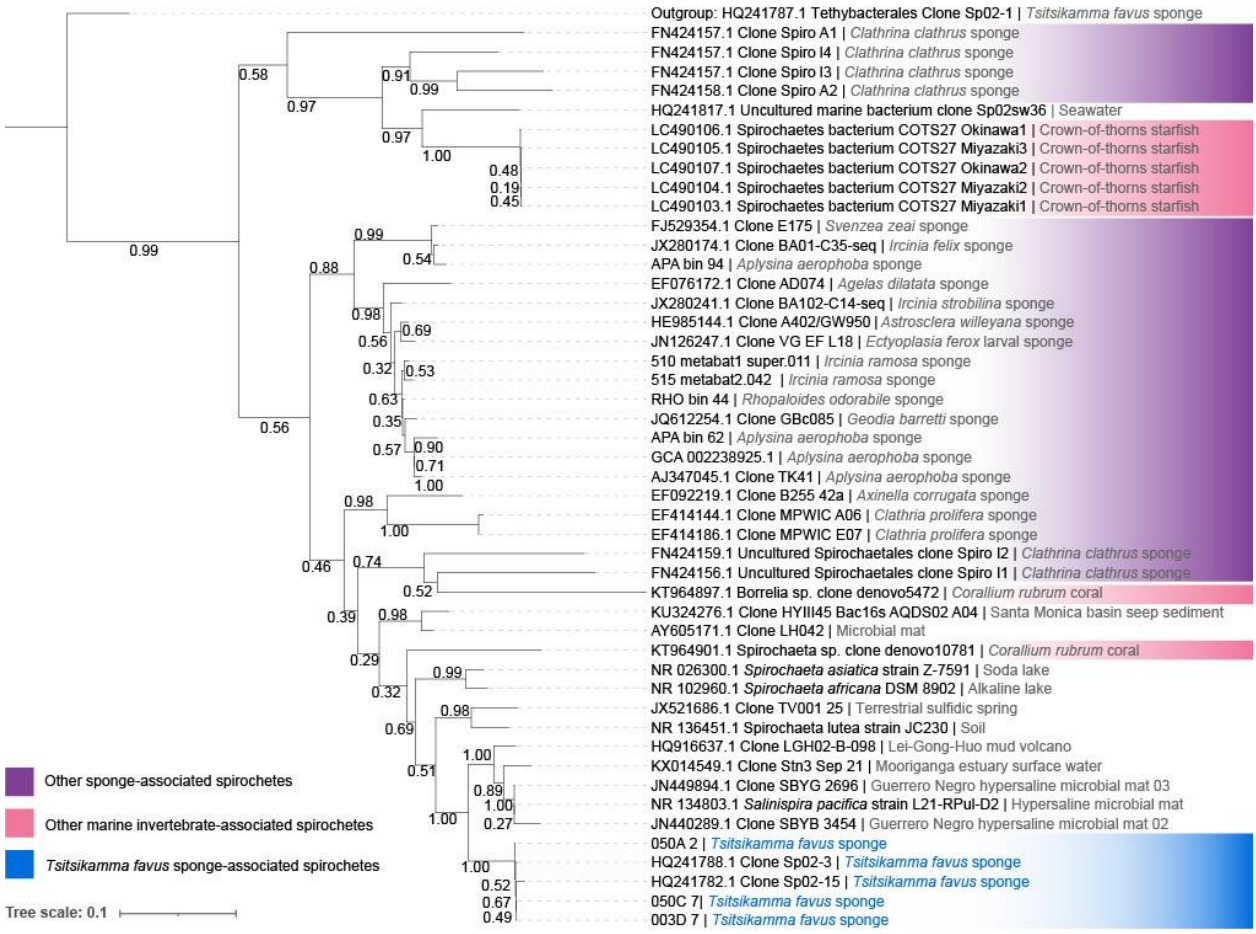


Figure 1. Maximum-likelihood phylogeny of spirochete 16S rRNA gene sequences inferred with 1000 bootstrap replicates. A 16S rRNA gene sequence from a sympatric *Tsitsikamma*-associated Tethybacterales clone was used as an outgroup. Bootstrap values are indicated on branch nodes. Sequences from latrunculid-associated spirochetes are indicated in blue, other sponge-associated spirochetes are indicated in purple, and spirochetes associated with other marine invertebrates are indicated in pink. For all sequences, accessions are provided where available, along with their description per the NCBI database and isolation source.

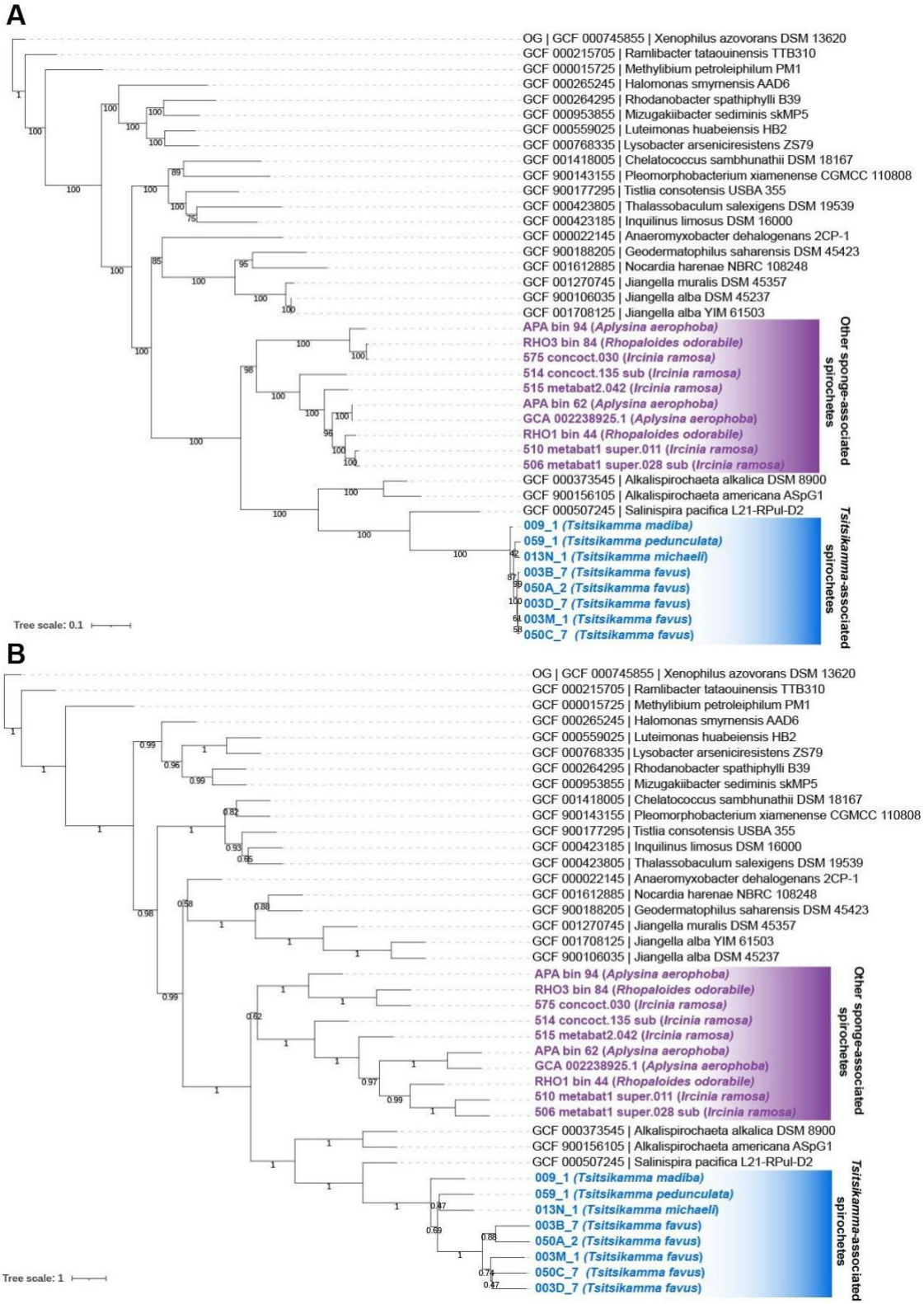


Figure 2. Phylogeny of sponge-associated spirochetes inferred with autoMLST in de novo mode using A) concatenated alignment and B) coalescent tree approaches. *Tsitsikamma*-associated

spirochetes are highlighted in blue with their respective hosts. Other sponge-associated spirochetes are highlighted in purple with their associated hosts. All other reference spirochete genomes are listed in the format of "Accession number | Scientific name".

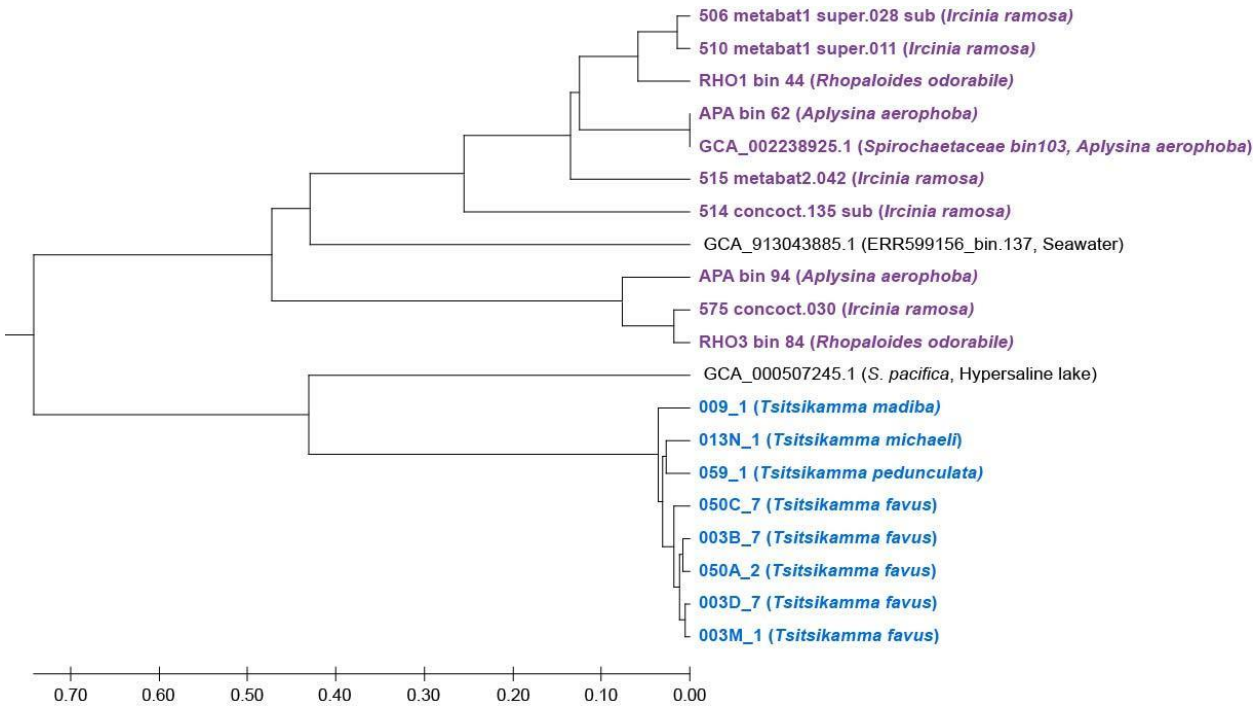


Figure 3. UPMGA representation of pairwise synonymous substitution rates (dS) of sponge-associated spirochete genomes, based on the alignment of 11 orthologous genes. PAL2NAL (88) and CodeML (89) from the PAML package were used to calculate pairwise dS values and the resultant matrix was visualized in MEGA11 (80). The *Tsitsikamma*-associated spirochetes are colored in blue and other sponge-associated spirochetes are colored in purple.

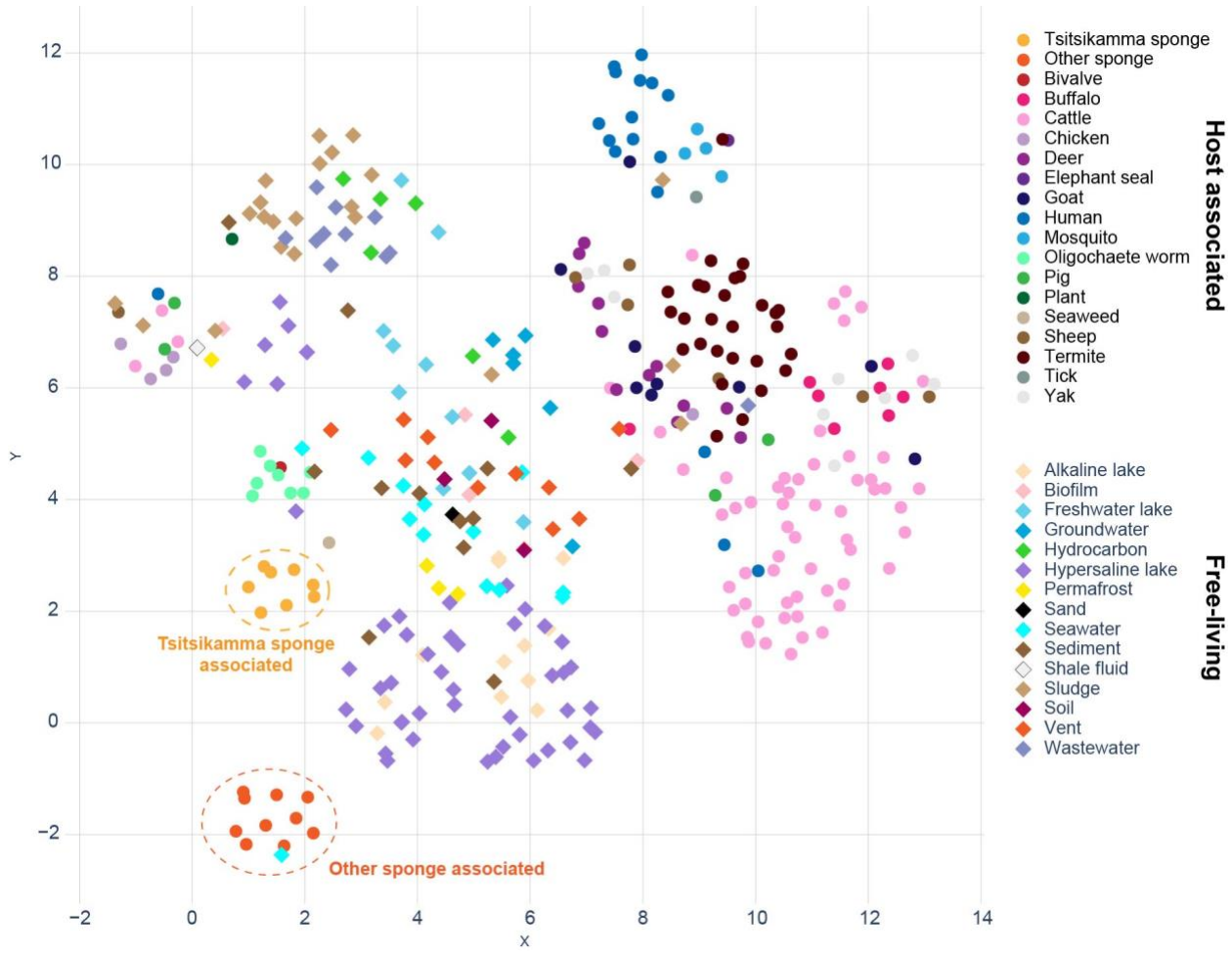


Figure 4. UMAP dimension reduction 2-dimensional representation of KEGG-annotated gene counts in all spirochete genomes. The isolation source of each genome is indicated by color, and shaped according to whether the isolation source is a living host (circles) or an abiotic environment (diamonds).

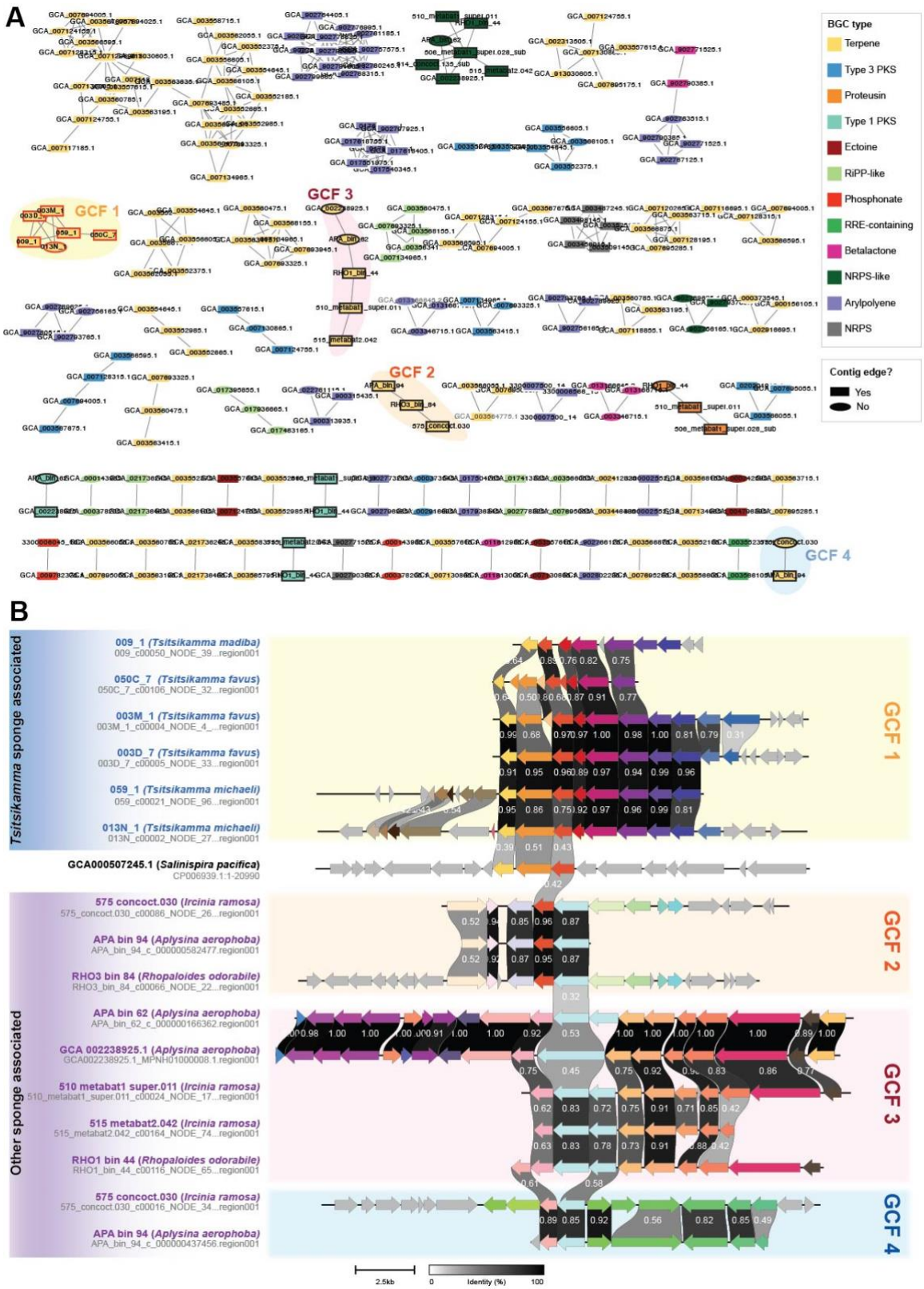


Figure 5. Assessment of biosynthetic potential in spirochetes. A) Network visualization of biosynthetic gene clusters from all spirochete genomes used in this study clustered into gene

cluster families at a maximum distance of 0.3. BGCs from *Tsitsikamma*-associated spirochetes are highlighted with a red outline. BGCs from all other sponge-associated spirochetes are highlighted with a black outline. Gene cluster families (GCFs) of interest are highlighted. B) Pairwise comparison of amino-acid sequence identity of terpene biosynthetic gene clusters from sponge-associated spirochetes. The pairwise similarity between genes is indicated between genes, and genes are colored according to their predicted function. The GCFs to which the BGCs belong have been indicated.

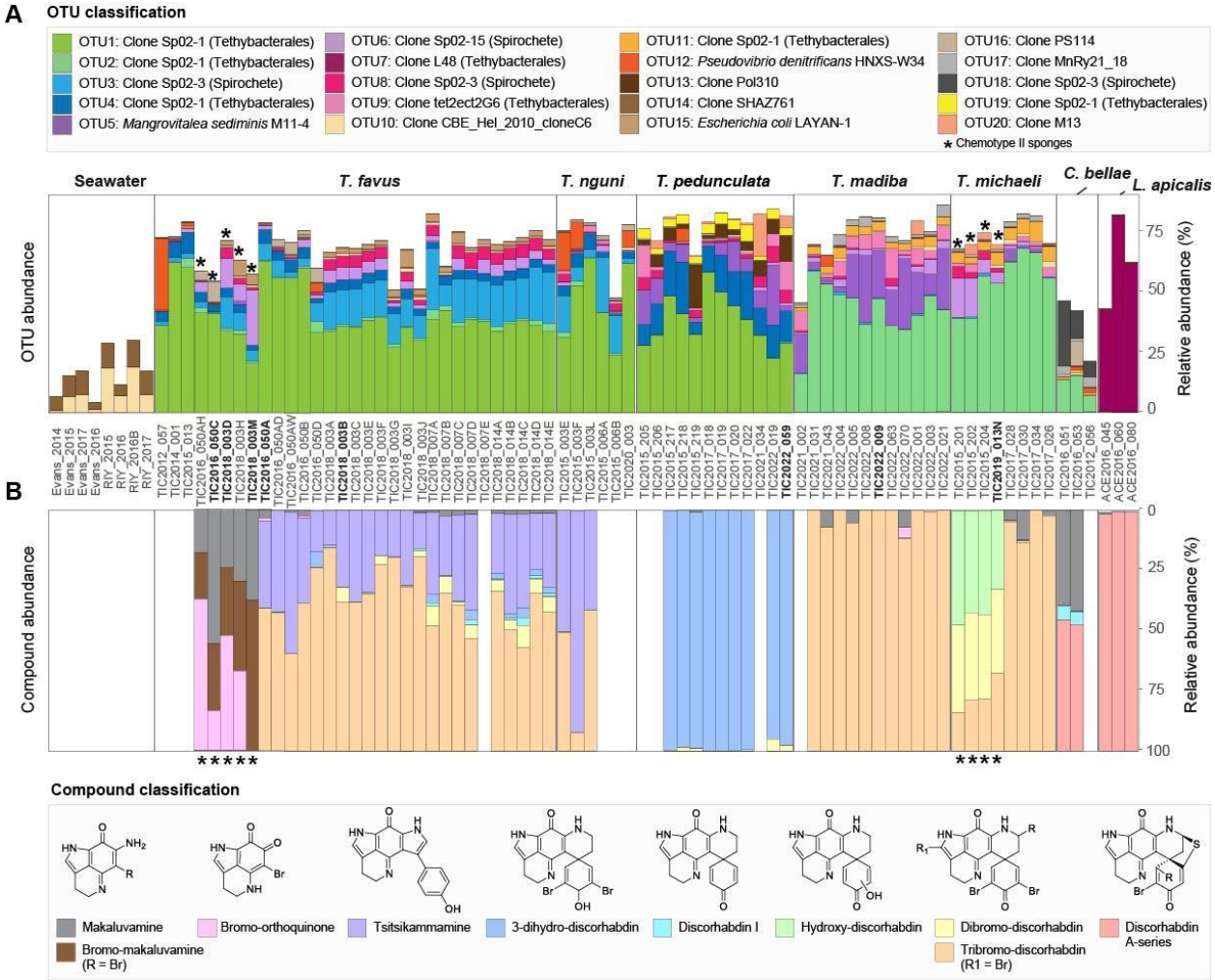
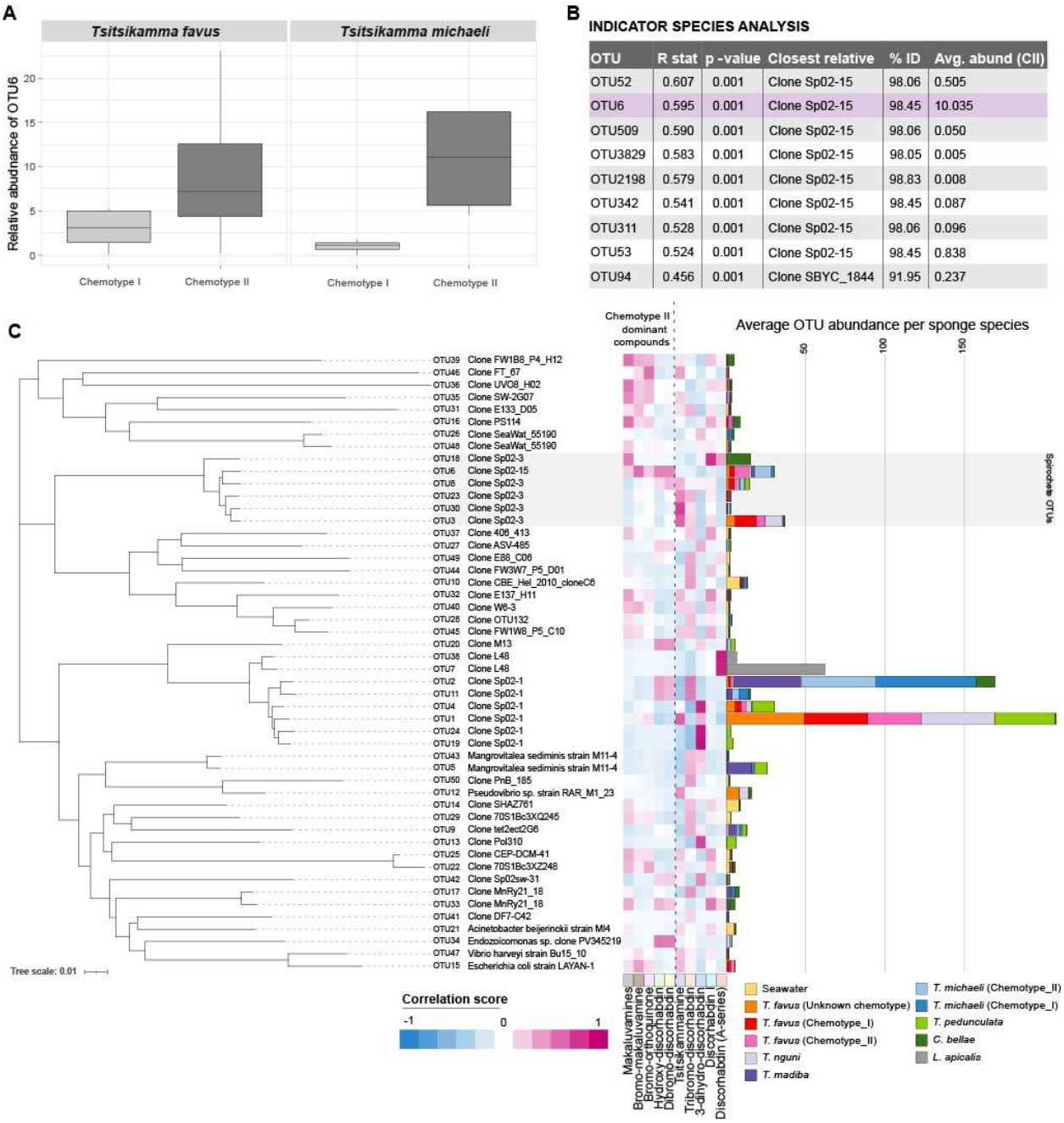


Figure 6. Distribution of microbial OTUs and pyrroloiminoquinone classes in latrunculid sponges A) Bacterial community profiles associated with latrunculid sponges and surrounding seawater represented by OTUs clustered at a distance of 0.01. The relative abundance of the top 20 OTUs with the highest average abundance across all samples is shown. The closest relative determined by alignment against the NCBI nt database is indicated within the colored key legend. B) The relative abundance of different pyrroloiminoquinone classes per sponge sample as indicated by the colored key. Scaffold structures are provided for identified chemical groups. Blank regions indicated sponges that have not had their chemistry assessed. Sponges classified as Chemotype II have been indicated with an asterisk. Sponge samples from which metagenomic data was retrieved and analyzed are indicated with bold lettering.



OTU	R stat	p -value	Closest relative	% ID	Avg. abund (CII)
OTU52	0.607	0.001	Clone Sp02-15	98.06	0.505
OTU6	0.595	0.001	Clone Sp02-15	98.45	10.035
OTU509	0.590	0.001	Clone Sp02-15	98.06	0.050
OTU3829	0.583	0.001	Clone Sp02-15	98.05	0.005
OTU2198	0.579	0.001	Clone Sp02-15	98.83	0.008
OTU342	0.541	0.001	Clone Sp02-15	98.45	0.087
OTU311	0.528	0.001	Clone Sp02-15	98.06	0.096
OTU53	0.524	0.001	Clone Sp02-15	98.45	0.838
OTU94	0.456	0.001	Clone SBYC_1844	91.95	0.237

Figure 7. Populations of Sp02-15 spirochetes, primarily represented by OTU6, have A) an increased abundance in Chemotype II sponges, B) and are significantly associated with Chemotype II sponges per an Indicator Species Analysis. C) Co-correlation analysis of OTU and pyrroloiminoquinone compound abundance shows that OTU6 is exclusively correlated with all compounds that are characteristic of *T. favus* Chemotype II and *T. michaeli* Chemotype II sponges. The correlation score per OTU-compound combination is indicated by a color scale ranging from -1 to 1. A negative score is indicative of a presence-absence relationship and a positive score is indicative of a presence-presence relationship.

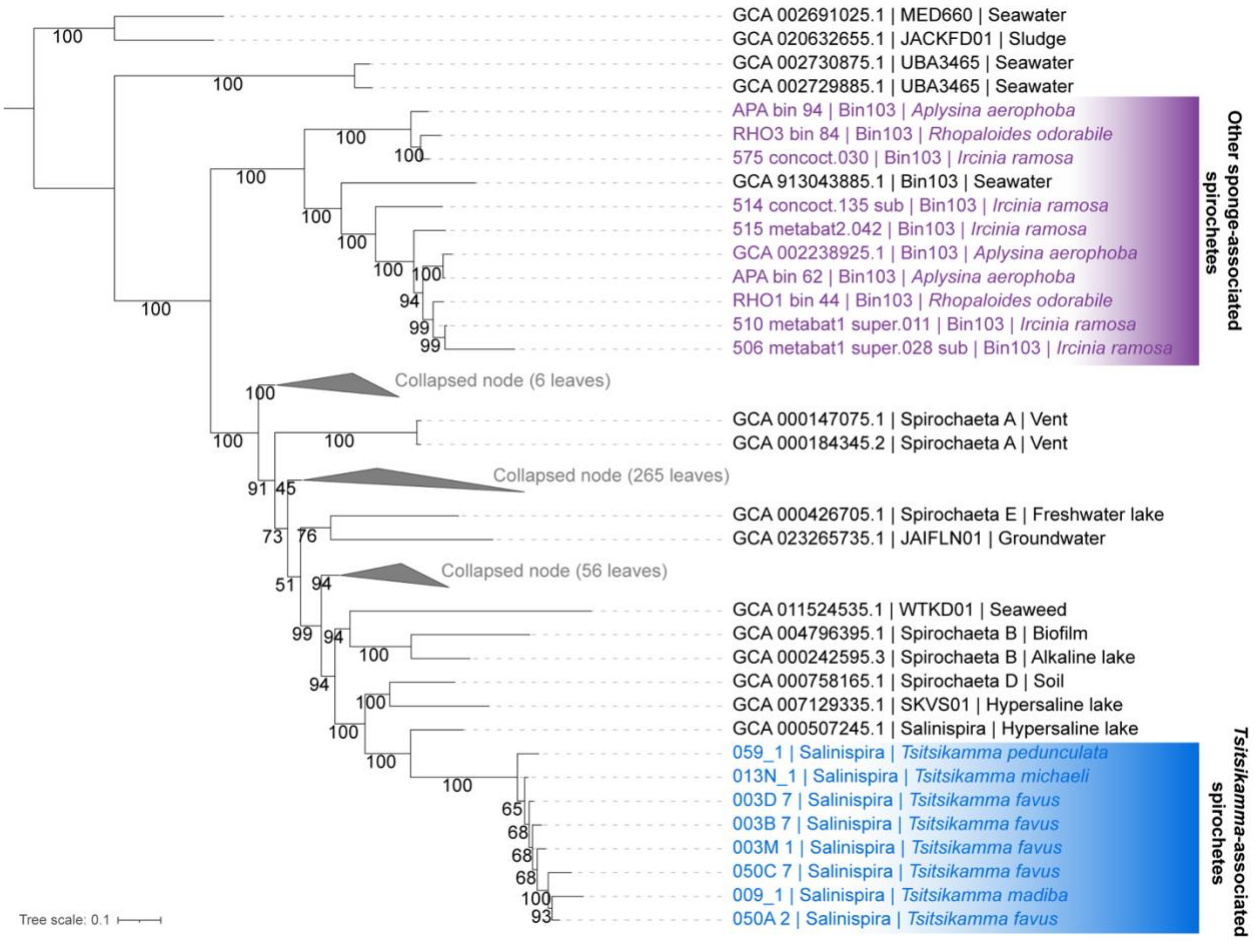


Figure S1. Phylogeny of all Spirochateacea genomes as inferred using an alignment generated by Phylophlan3 and RaxML with 1000 bootstrap replicates. *Tsitsikamma*-associated spirochetes are highlighted in blue, and other sponge-associated spirochetes are highlighted in purple. Bootstrap values and branch length are indicated.

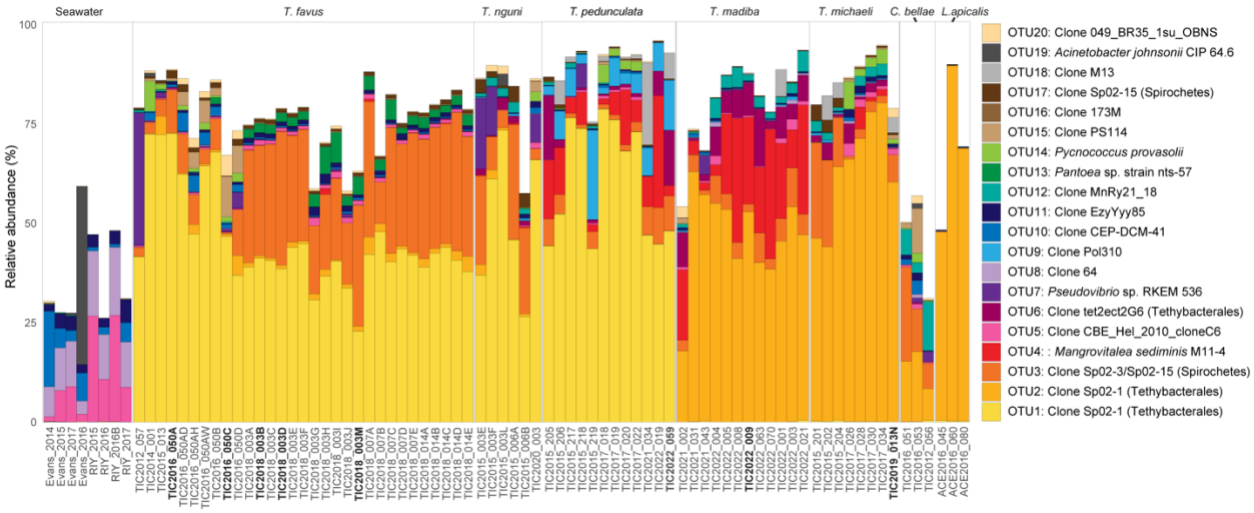


Figure S2. The distribution of the top 20 most abundant OTUs clustered at a distance of 0.03 in latrunculid sponges and seawater samples. OTUs are indicated by a colored key along with their closest match as determined by blastn analysis against the NCBI nt database

Table 1. Characteristics of putative representative genomes of Tsitsikamma sponge-associated spirochete symbiont MAGs

MAG	Size (Mbp)	Quality	16S rRNA (% ID)	23S rRNA (% ID)	Host	Chemotype	Sample
003B_7	1.97	Medium	N/A	Salinispira pacifica L21-RPul-D2 (89.54%)	T. favus	I	TIC2018-003B
050A_2	2.73	Low	Uncultured marine clone Sp02-3 (99.52%)	Salinispira pacifica L21-RPul-D2 (89.54%)	T. favus	I	TIC2016_050A
003D_7	2.48	High	Uncultured marine clone Sp02-3 (99.52%)	Salinispira pacifica L21-RPul-D2 (89.58%)	T. favus	II	TIC2018-003D
003M_1	2.74	High	N/A	Salinispira pacifica L21-RPul-D2 (89.58%)	T. favus	II	TIC2018-003M
050C_7	1.72	Medium	Uncultured marine clone Sp02-3 (99.52%)	N/A	T. favus	II	TIC2016-050C
009_1	1.47	High	N/A	Salinispira pacifica L21-RPul-D2 (91.25%)	T. madiba	N/A	TIC2022-009
013N_1	2.33	High	N/A	Salinispira pacifica L21-RPul-D2 (89.48%)	T. michaeli	II	TIC2019-013N
059_1	2.04	Medium	N/A	N/A	T. pedunculata	N/A	TIC2022-05