The unique and enigmatic spirochete symbiont of latrunculid 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: Spirochete symbionts 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 rare in marine sponges, thus far only observed in *Clathrina clathrus* and various species within the Latrunculiidae 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 from eight latrunculid sponges revealed that these unusual

spirochetes are relatively recent symbionts and are phylogenetically distinct from other sponge-associated spirochetes. Functional comparative analysis suggests that the host sponge may have selected for these spirochetes due to their ability to produce terpenoids and/or possible structural contributions.

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 marine 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 of the latrunculid-associated spirochetes and found that these symbionts are phylogenetically distinct from all invertebrate-associated spirochetes. The symbiosis between the spirochetes and their sponge host appears to have been established relatively recently.

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 metagenome-assembled genomes (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) collected by Neulinger and colleagues from the Adriatic Sea off the coast of Croatia (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 key members in coral skeletal environment 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 distinct from all other sponge-associated spirochetes.

RESULTS AND DISCUSSION

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.

Survey of microbial communities in latrunculid sponges and other sponge species endemic to the South African coast

To survey the prevalence of spirochetes in sponge collected off the South African coastline, we clustered 16S rRNA gene fragment amplicons sourced from 155 marine sponges and 8 seawater samples into operational taxonomic units (OTUs) at a distance of 0.03 in mothur (45). These sponges were collected primarily from reefs within Algoa Bay, South Africa but also included samples from the Tsitsikamma National Park, the Amathole Marine Protected Area in the Indian Ocean, and the remote Bouvet Island in the Southern (Antarctic) Ocean (Table S1).

A total of 9711 OTUs were recovered from the 163 amplicon libraries. We identified spirochete OTUs with classifications from alignment of the OTUs against the SILVA and nr databases (Table S2). A total of 142 OTUs were classified within the Spirochaetota phylum, of which only 10 had an average abundance greater then 0.01% across all sponge specimens (Fig. 1A). OTU3 and OTU59 were most abundant in the *Tsitsikamma* and *Cyclacanthia* sponges. These OTUs were most closely related to spirochete 16S rRNA gene clones Sp02-3 and Sp02-15, previously identified in *T. favus* sponges (22). These two OTUs were present at low abundance in the *Latrunculia algoaensis* and *Latrunculia apicalis* sponge specimens (collected in Algoa Bay and the Southern Antarctic Ocean), as well as in some *Mycale* specimens and a single sympatric *Phorbus sp.* sponge (Fig. 1B).. As the *Mycale* specimens were found as encrusting species on the *Tsitsikamma favus* sponges, we cannot discount the possibility of contamination between

these two species. As we have only a single *Phorbus* sp. representative, additional specimens will be required to determine the significance of these spirochete OTUs in this genus or whether this was a result of contamination during collection. These two OTUs were otherwise absent in all other non-latrunculid sponges collected from sympatric regions. The presence, albeit low, of OTU3 and OTU59 in the *L. apicalis* sponges collected just off of Bouvet Island (~ 3000 km/ 1800 miles from Algoa Bay), and the presence of phylogenetically distinct spirochetes in sympatric non-latrunculid sponges of Algoa Bay would suggest that these Sp02-3 and Sp02-15 spirochetes are specifically associated with latrunculid sponges

Spirochete OTUs OTU105 and OTU128 were relatively abundant in other sponges collected from the South African coast, and absent in latrunculid sponges, appeared more sporadic in their distribution among sponge specimens (Fig. 1B). These OTUs were most closely related to spirochetes detected in *Spongia officinalis* (OY759747.1) and *Astrosclera willeyana* (HE985144.1) sponges, respectively (Table S2). Inspection of phylogeny of these ten OTUs (Fig. 1C) revealed that six of the ten spirochete OTUs formed a clade with spirochete clones previously cloned from *T. favus* sponges (22). Of the remaining four, OTU105 and OTU128 (which were more abundant in non-Latrunculid sponge specimens) were part of distant clades of other sponge associated spirochetes, while OTU581 and OTU399 belonged to a clade stemming from a variety of environments (Fig. 1C). Notably, a clone (Sp02sw36) isolated from the seawater extruded from *Tsitsikamma favus* sponges in 2012 (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).

Characterization of *Tsitsikamma* sponge-associated spirochete MAGs

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 S3). The 16S rRNA and 23S rRNA gene sequences from each MAG (if recovered) were aligned against the NR nucleotide database via online BLASTn (46).

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 (47), was previously shown to be the closest known relative of the conserved spirochete Sp02-3 clone (22, 29). Finally, all eight *Tsitsikamma*-associated spirochete MAGs were taxonomically classified, via GTDB-Tk (48), within the *Salinispira* genus (Table S3). 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 phylogeny from the 16S rRNA gene alignment showed that the *Tsitsikamma*-associated spirochete MAGs were distinct from all other invertebrate-associated spirochetes (Fig. S1). 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.

Since 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 (49) in *de novo* mode, with both concatenated alignment (Fig. 2A) and coalescent tree (Fig. 2B) approaches, using ten MAGs/genomes acquired from other sponge hosts, *Rhopaloides odorabile, Ircinia ramosa,* and *Aplysina aerophoba* (50–52), as references. 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 23S rRNA gene phylogeny. The *Tsitsikamma*-associated spirochetes appeared

phylogenetically clustered following their respective hosts, rather than geographically clustered. 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 (53) and RaxML (54) (Fig. S2). 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 MAG present in seawater (GCA 913043885.1) clustered with the other sponge-associated spirochetes. The origin of this particular genome, whether from a free-living spirochete or a sponge symbiont, remains uncertain due to potential annotation errors in the database. However, we have opted to follow the supplied annotation and presume that this MAG is likely representative of the closest free-living relative within the clade. Our phylogenetic analysis incorporated all publicly available genomes and MAGs of the Spirochaetaceae phylum, and therefore this presumption is limited by the existing dataset. We calculated pairwise average nucleotide identity (ANI) scores for all 363 spirochete genomes (Table S4). The *Tsitsikamma*-associated spirochetes shared between 93.9% to 98.2% ANI with each other (Table S5), 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) in orthologous genes present in all genomes. 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 may have begun diverging before the *Tsitsikamma*-associated spirochetes diverged from their free-living relative. The divergence pattern of the *Tsitsikamma*-associated spirochetes is congruent with the phylogeny of their sponge host and incongruent with geographic location, suggestive of phylosymbiosis. Finally, it appears that these spirochetes have only recently begun diverging from one another as they adapt to their sponge host and that their association with latrunculid sponges is more recent than that of the co-dominant Tethybacterales symbionts (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 (55). KO counts per genome were mapped back to associated pathways detailed in the KEGG database (56) (Table S6). Dimension reduction of these counts per genome revealed distinct clusters suggestive 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 S7) 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.

The biosynthetic potential of Sp02-3 spirochetes

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 (57) (Fig. 5A). Six of the eight *Tsitsikamma*-associated spirochetes had only a single predicted BGC. The remaining two MAGs, 003B_7 and 050A_2, which were of medium and low quality respectively, had no detected BGCs, likely due to incomplete coverage of the genomes. 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 performed 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 (terpenoid), which we assume is produced via the only terpene BGC present in the *S. pacifica* genome. Despite the low sequence and organizational similarities, the terpene, if produced in the latrunculid-associated spirochetes, may protect them or their host against oxidative stress, as hypothesized for the *S. pacifica* bacterium (47)

In our previous studies, we have reported the existence of two chemotypes that exist in the *T. favus* and *T. michaeli* sponge populations in Algoa Bay (21, 28). Chemotype I represents the majority of sponges, as the sponges appear visually healthy with turgid structure and their spicules are in the canonical form. Further, this Chemotype is defined by the presence of a variety of discorhabdins and tsitsikammamines (28). Conversely, the morphology of the Chemotype II sponges is considered abnormal where the tissues appear bruised, are soft to the touch (akin to rotten fruit), and many spicules are malformed (28, 58). This chemotype is further characterized by the presence of structurally simpler makaluvamines and brominated discorhabdins (21, 28).

Previous surveys of the microbial communities associated with ten *T. favus* sponges and found no correlation between any bacterial population and the chemotypes (28). We have

repeated the analysis with a larger cohort of *T. favus* and *T. michaeli* sponge specimens (N = 26). Using the same 16S rRNA gene amplicon datasets as presented in Figure 1, but instead including only data from the latrunculid sponges with associated chemical data, the analysis was repeated and OTUs were clustered at a maximum distance of 0.01 (Table S9) to disentangle the two spirochete strains previously identified in latrunculid sponges, Sp02-3 and Sp02-15 (22, 29). Using an Indicator Species Analysis (Table S10) we found that a decrease in Sp02-3 representative OTU abundance (OTU3) and an increase in Sp02-15 representative OTU abundance (OTU6) correlated with Chemotype II sponges, relative to Chemotype I specimens (Fig. S3 A – B, Table S10).

We conducted a correlation analysis of the top 50 most abundant OTUs with relative pyrroloiminoquinone abundance per sponge sample (Fig. S4, Table S11). The Sp02-3 spirochetes (OTU3) were positively correlated with the increased abundance of Chemotype I pyrroloiminoquinones and negatively correlated with the presence of Chemotype II pyrroloiminoquinones. The converse was true of the Sp02-15 spirochetes (OTU6) (Fig. S4, Table S11). As there was no evidence of BGCs for the production of pyrroloiminoquinones in the spirochete MAGs, this result suggests that the switch from Chemotype I to Chemotype II (the cause of which has yet to be identified) appears to negatively impact the Sp02-3 spirochete and allows the Sp02-15 spirochete to thrive in place.

Since the decrease in Sp02-3 similarly correlated with the incidence of deformed spicules, we considered whether it may play a role in spicule formation. The most closely related

invertebrate-associated spirochete (Fig. 1 and Fig. 2) is a highly dominant and conserved spirochete in *Corallium rubrum* corals (39, 59) This spirochete is predicted to contribute to the coral's overall health of the coral (60) and to produce a pigmented carotenoid that influences the commercially prized color of this red coral, as the spirochete's presence correlates with the intensity of the observed red pigmentation (61). This spirochete was primarily found in the coenenchyme of the coral (61), which houses the sclerites (spicules) that are thought to act as initiation sites for the formation of the axial skeleton (62). Finally, the formation of spicules in a primary coral polyp is associated with a change in color from white to light pink (63). It is thus possible that the *C. rubrum*-associated spirochete may be involved in spicule formation as shown with the calcibacteria in *Hemimycale* sponges (pale orange to deep red in color) (64, 65), and hypothesized for the spirochetes in *Platygyra dadalea*, *Paragoniastrea australensis*, and Porites lutea sponges (44). While a speculative connection, as no MAG or genome is available for these spirochetes, this observation has prompted us to begin metatranscriptomic studies in conjunction with CARD-FISH experiments to determine the localization and potential structural role of spirochetes in latrunculid sponges from the South African coastline.

Conclusion: This study shows that the conserved Sp02-3 spirochete of latrunculid sponges is likely to be a relatively new symbiont that has begun co-evolving with its respective sponge hosts. The Sp02-3 symbiont 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 are functionally unique relative to other sponge-associated spirochetes. We found no

evidence that they are directly involved in the production of the pyrroloiminoquinones characteristic of their host sponges. The close phylogenetic relatedness of the latrunculid-associated spirochetes to a dominant, conserved coral-associated spirochete hints at a possibly structural role within the sponges. However, additional experiments will be necessary to test this hypothesis.

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, 58).

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) (45). 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 (66) 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 a distance of 0.03 and read counts thereof were converted to relative abundance (Table S2). Representative sequences of each OTU were aligned against the SILVA database (v138.1) in mothur and against the nt prokaryotic database using standalone blastn (67), 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 (68). Spirochete OTUs were subset out and aligned with reference sequences from the NCBI nucleotide database using MUSCLE (v. 5.1) (69, 70) and phylogeny was inferred from the alignment using the Maximum-likelihood method with 1000 bootstrap replicates in MEGA11 (71). Finally, the same analysis was repeated but using only the raw amplicon read data from

latrunculid sponges, and the OTUs were clustered at a distance of 0.01. in all other respects, the analyses were identical.

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 (72) 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.

Correlation of spirochete populations and sponge chemotypes

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

(73) 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) (69, 70), using the neighbor-joining approach with 1000 bootstraps in MEGA11 (71). The final tree was visualized in iTol (74) where the correlation matrix and the average OTU abundance per sponge species was visualized alongside the tree as datasets.

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 (50, 51). 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 S3).

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) (48) 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 S3.

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) (67). Resultant sequences were aligned using MUSCLE (v. 5.1) (69, 70) and phylogeny was inferred using the Maximum-likelihood method with 1000 bootstraps in MEGA11 (71). Phylogeny of the *Tsitsikamma*-associated spirochete MAGs was similarly inferred using whole genome data via autoMLST (49) and PhyloPhlan3 (53). Amino acid sequences and nucleotide sequences for all genes were found in all genomes using prokka (v 1.13) (78). 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) (79) to build a phylogenetic tree with 1000 bootstrap replicates using the PROTGAMMAAUTO model. The resultant tree was visualized in iTol (74). 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 Iatrunculid-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 (74). Finally, the pairwise average nucleotide identity (ANI) was calculated for all genomes using fastANI (v1.33)(80). If a pairwise alignment fraction (AF) was lower than 70% (81), 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 (53) and autoMLST(49, 53) 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) (82). A total of 11 orthologs common to all genomes were found using count_OGs.py and aligned using MUSCLE (v 5.1) (69, 70). 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) (69, 70). 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 (83) and aligned using PAL2NAL (84). 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 (85).

Comparative analysis of functional potential in spirochete genomes

Genes were identified in all genomes/MAGs using Prokka (v 1.13) (78) and then annotated against the KEGG database using KOfamSCAN (55) 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 (86), 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 (87) 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 sponge-associated spirochetes

A total of 547 biosynthetic gene clusters (BGCs) were predicted from all spirochete genomes (N=363) using antiSMASH (v. 6.0.1) (88) 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)(57) at maximum distances of 0.3, 0.5, and 0.8. Network files of non-singleton gene cluster families (GCFs) were visualized in Cytoscape (89). Highlighted gene clusters of interest were visualized with clinker (90). Metadata for BGCs was extracted from individual GenBank files using antismash_summary.py.

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.

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. 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.

46. 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.

47. 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.

48. 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.

49. 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.

50. 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.

51. 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.

52. 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.

53. 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.

54. 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.

55. 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.

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

57. 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.

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

59. van de Water JAJM, Voolstra CR, Rottier C, Cocito S, Peirano A, Allemand D, Ferrier-Pagès C. 2018. Seasonal stability in the microbiomes of temperate gorgonians and the red coral Corallium rubrum across the Mediterranean sea. Microb Ecol 75:274–288.

60. Tignat-Perrier R, van de Water JAJM, Allemand D, Ferrier-Pagès C. 2023. Holobiont responses of mesophotic precious red coral Corallium rubrum to thermal anomalies. Environ Microbiome 18.

61. van de Water JAJM, Allemand D, Ferrier-Pagès C. 2024. Bacterial symbionts of the precious coral *Corallium rubrum* are differentially distributed across colony-specific compartments and differ among colormorphs. Environ Microbiol Rep 16.

62. Perrin J, Vielzeuf D, Ricolleau A, Dallaporta H, Valton S, Floquet N. 2015. Block-by-block and layer-by-layer growth modes in coral skeletons. Am Mineral 100:681–695.

63. Giordano B, Bramanti L, Perrin J, Kahramanoğulları O, Vielzeuf D. 2023. Early stages of development in Mediterranean red coral (Corallium rubrum): The key role of sclerites. Front Mar Sci 10.

64. Uriz MJ, Agell G, Blanquer A, Turon X, Casamayor EO. 2012. Endosymbiotic calcifying bacteria: a new cue to the origin of calcification in metazoa? Evolution 66:2993–2999.

65. Garate L, Sureda J, Agell G, Uriz MJ. 2017. Endosymbiotic calcifying bacteria across sponge species and oceans. Sci Rep 7:43674.

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

67. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

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

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

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

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

72. 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.

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

74. 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.

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

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

80. 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.

81. 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.

82. 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.

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

84. 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.

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

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

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

88. 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.

89. 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.

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

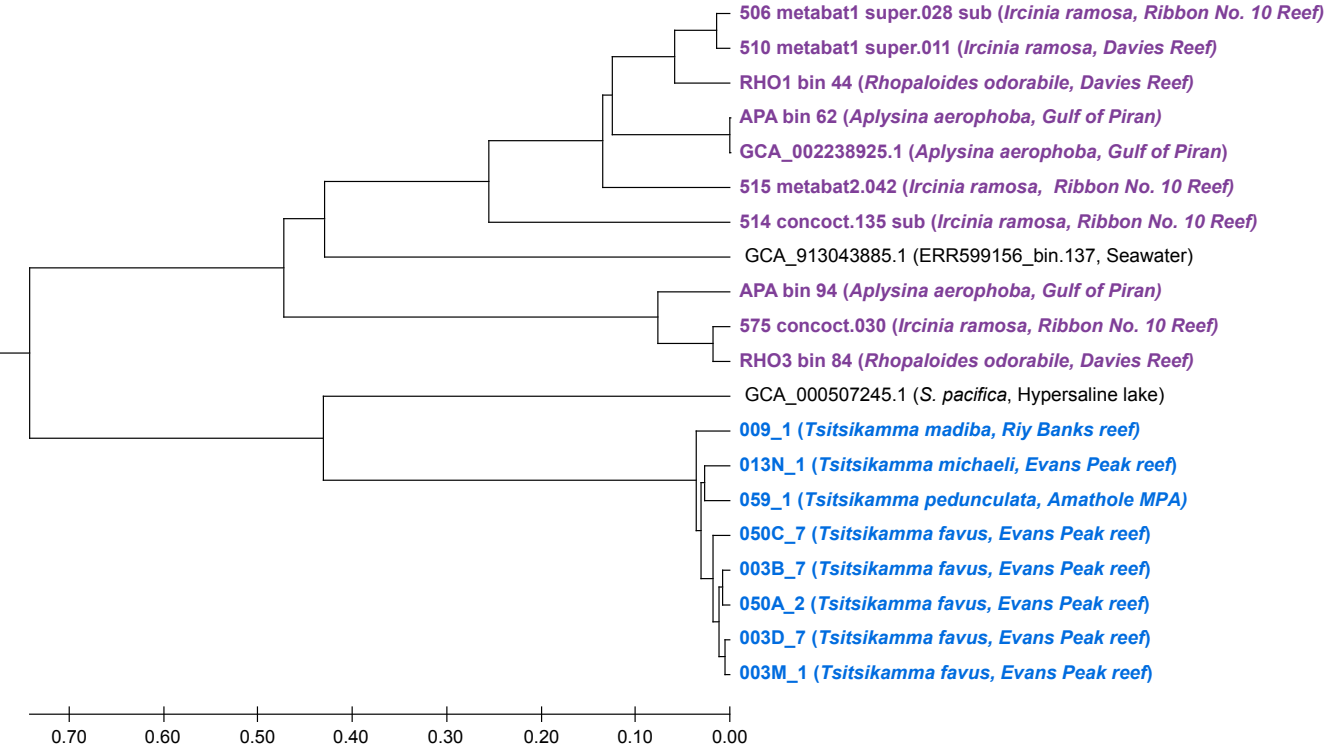

Figure 3. UPMGA representation of pairwise synonymous substitution rates (dS) of sponge-associated spirochete genomes, based on the alignment of 11 orthologous genes. PAL2NAL (88) and CodeML (89) from the PAML package were used to calculate pairwise dS values and the resultant matrix was visualized in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

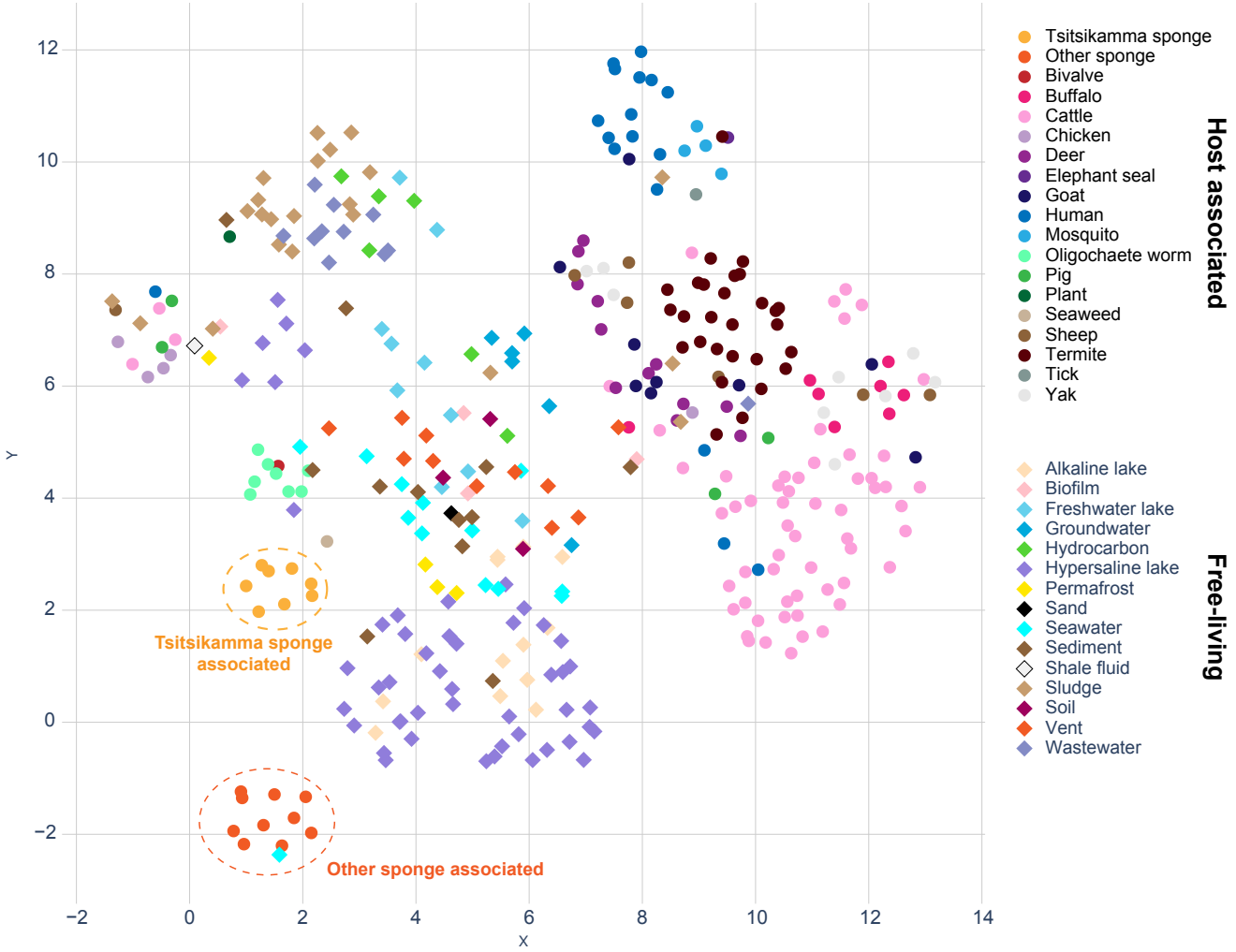

Figure 4. UMAP dimension reduction 2-dimensional representation of KEGG-annotated gene counts in all spirochete genomes. The isolation source of each genome is indicated by color, and shaped according to whether the isolation source is a living host (circles) or an abiotic environment (diamonds).

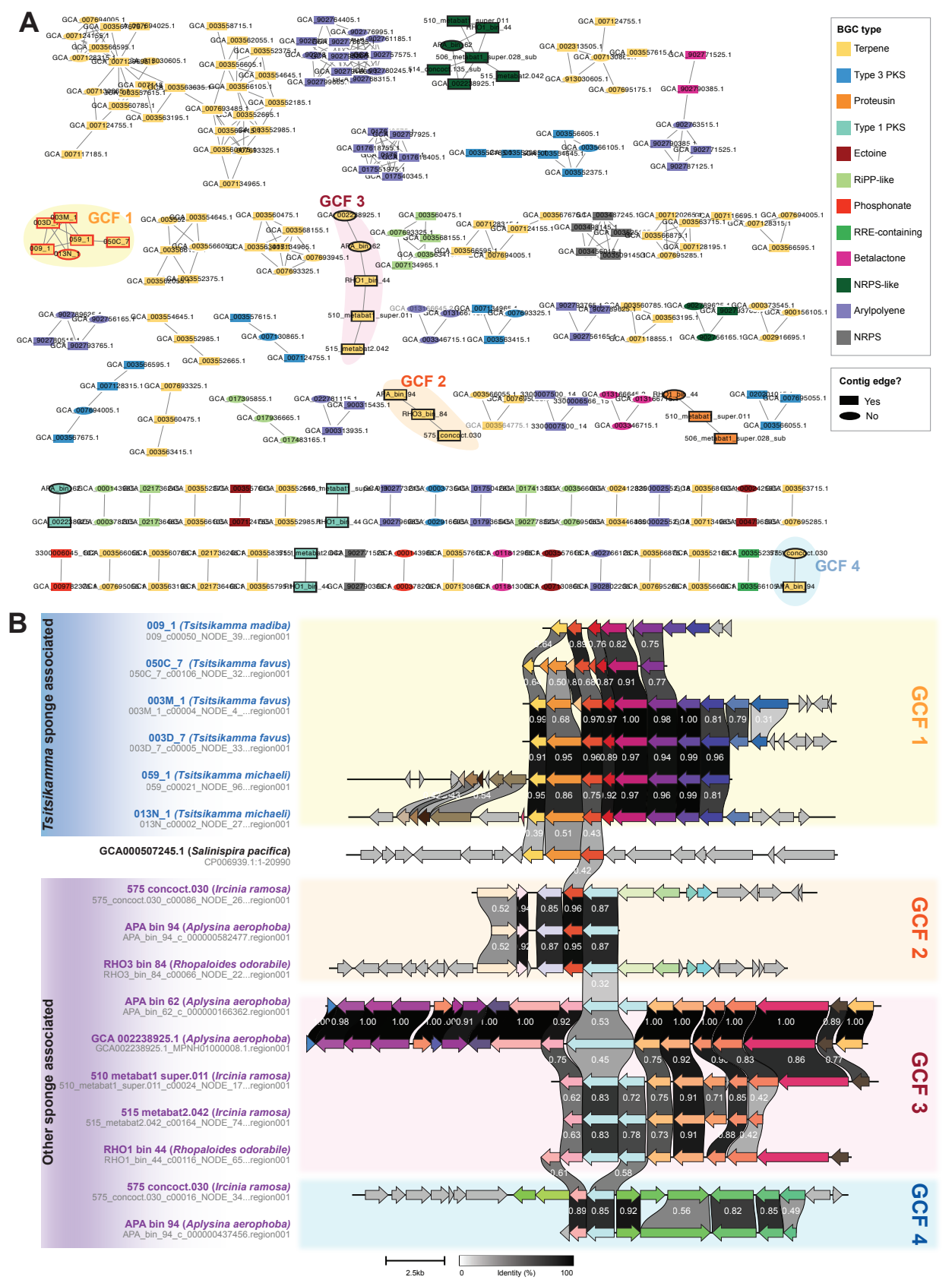

Figure 5. Assessment of biosynthetic potential in spirochetes. A) Network visualization of biosynthetic gene clusters from all spirochete genomes used in this study clustered into gene cluster families at a maximum distance of 0.3. BGCs from Tsitsikamma-associated spirochetes are highlighted with a red 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.

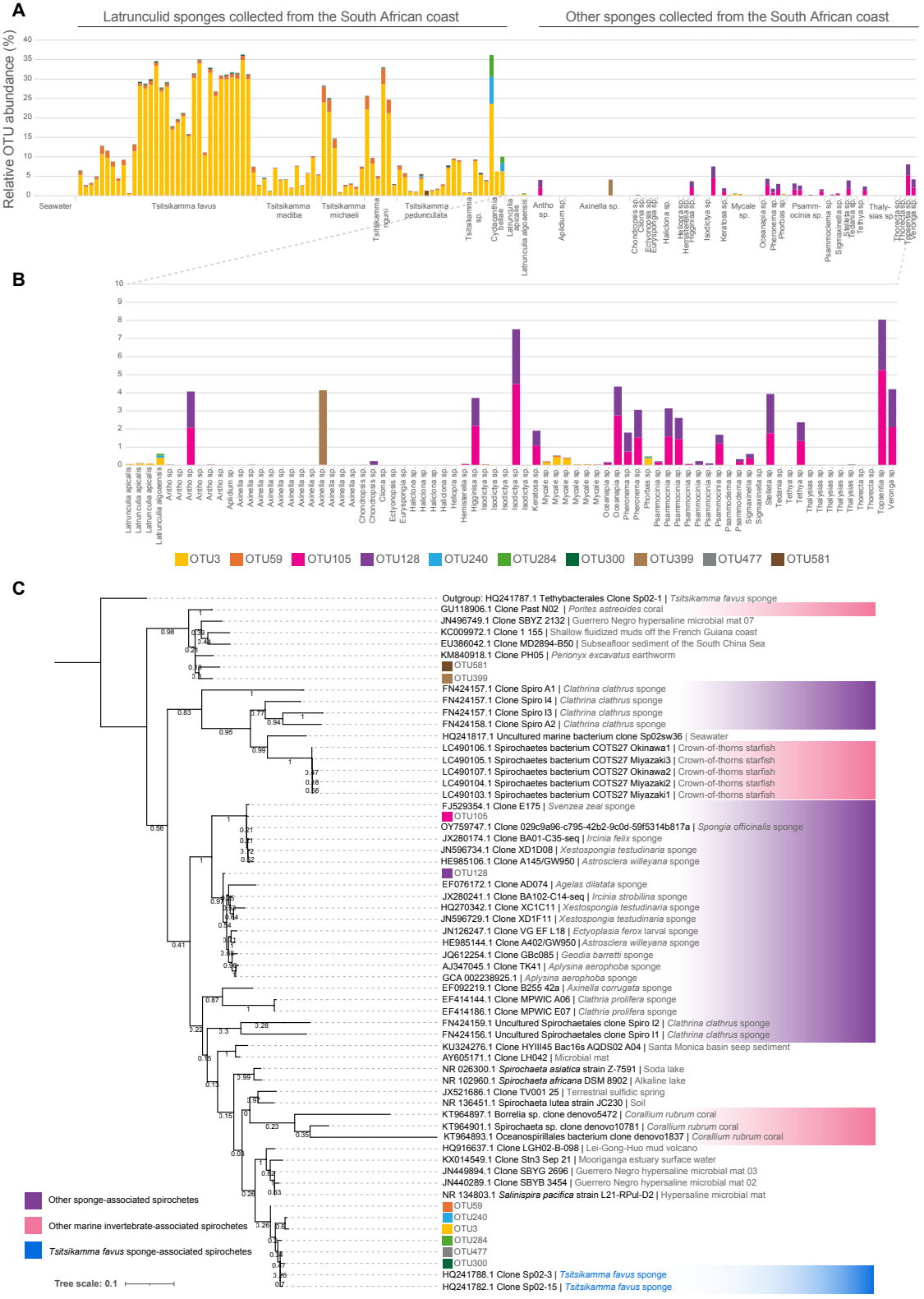

Figure 1. Spirochete population distribution in sponges collected from the South African coast and the Antarctic Southern Ocean. A) The relative abundance of OTUs clustered at a distance of 0.03 and classified as spirochetes, B) a magnified view of the spirochete OTUs present in non-latrunculid sponges collected from the south eastern coast of South Africa, three *L. apicalis* sponges collected from the Southern Ocean, and one sympatric *L. algoaensis* sponge. C) Maximum-likelihood phylogeny (with 1000 bootstraps) of the top ten most abundant spirochete OTUs recovered from the sponges included in this study.

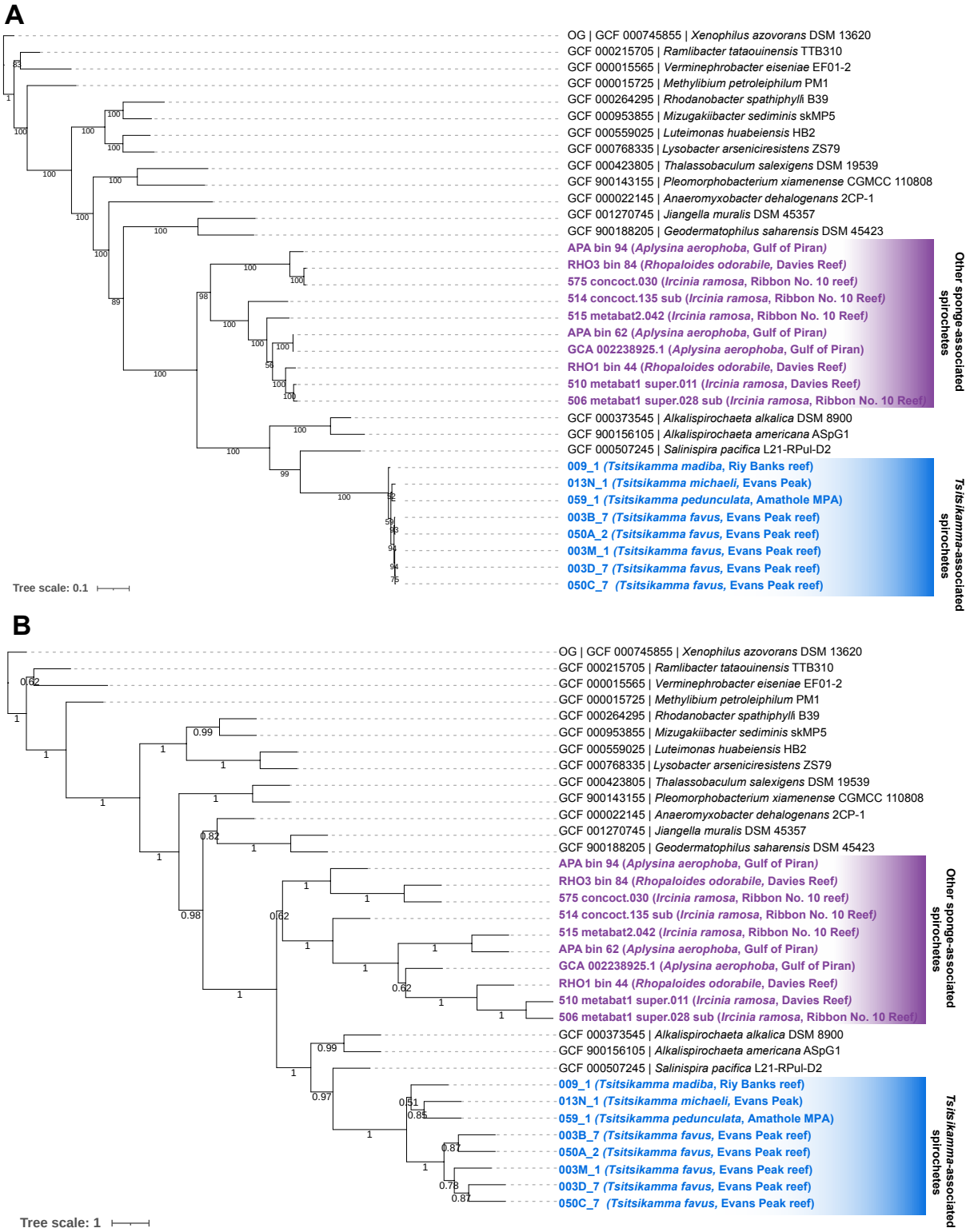

Figure 2. Phylogeny of sponge-associated spirochetes inferred with autoMLST in de novo mode using A) concatenated alignment and B) coalescent tree approaches. Tsitsikamma-associated spirochetes are highlighted in blue with their respective hosts. Other sponge-associated spirochetes are highlighted in purple with their associated hosts. All other reference spirochete genomes are listed in the format of "Accession number | Scientific name".