RESEARCH ARTICLE

The impact of space, host dissimilitude, and environment on prokaryotic communities of golf ball sponges

Daniel F. R. Cleary[1] | Nicole J. de Voogd[2,3]

[1]Department of Biology & CESAM, University of Aveiro, Aveiro, Portugal

[2]Naturalis Biodiversity Center, Marine Biodiversity, Leiden, The Netherlands

[3]Institute of Biology (IBL), Leiden University, Leiden, Netherlands

Correspondence
Daniel F. R. Cleary, Department of Biology & CESAM, University of Aveiro, Campus de Santiago, 3810-193 Aveiro, Portugal.
Email: cleary@ua.pt

Nicole J. de Voogd, Naturalis Biodiversity Center, Marine Biodiversity, PO Box 9517, 2300 RA Leiden, The Netherlands.
Email: n.j.de.voogd@biology.leidenuniv.nl; nicole.devoogd@naturalis.nl

Funding information
Fundação para a Ciência e a Tecnologia; Nederlandse Organisatie voor Wetenschappelijk Onderzoek

Abstract

Golf ball sponges are small, sometimes inconspicuous, sponges. They can be found across a range of habitats varying from perturbed and pristine coral reefs to harbours and marine lakes and from the deep sea to shallow waters. They can be difficult to distinguish in the field and have presented some problems with taxonomists lumping and splitting species due to the difficulty in defining clear species boundaries. In the present study, we sampled golf ball sponges from Indo-Pacific and Caribbean locations and used 16S gene amplicon sequencing to study their prokaryotic communities. We show that golf ball sponges harbour a wide variety of prokaryotic communities. Among the most prevalent operational taxonomic units (OTUs), several belonged to a range of taxa, including the bacterial AqS1 and EC94 groups, which have been associated with genes known to facilitate interactions between hosts and microbes. Certain host taxa were enriched with OTUs classified to the SAR202 clade of Chloroflexi. Our findings show that prokaryotic dissimilarity varied as a function of space (geographical distance) and host dissimilitude. The importance of space and host dissimilitude, however, varied depending on the data transformation with host dissimilitude a more important predictor of untransformed data whereas space was a more important predictor of log-transformed data. Given that log-transformation downscales the influence of abundant taxa, we interpret these results by the tendency of closely related host organisms to host similar sets of abundant symbiotic microorganisms; distantly sampled specimens, in contrast, tend to harbour less abundant prokaryotic microorganisms found in the surrounding environment (e.g., seawater or sediment).

KEYWORDS
Cinachyrella, host, *Paratetilla*, porifera, spatial

1 | INTRODUCTION

Beta diversity plays a crucial role in unravelling the dynamics of ecosystems and is often explored by examining shifts in community dissimilarity as a function of distance, a phenomenon known as distance dependence (Cleary et al., 2004; Nekola & White, 1999; Tuomisto et al., 2003). The study of distance dependence in microbial communities holds particular pertinence because it challenges the conventional notion that microbes are ubiquitously distributed (Baas Becking, 1934). Given their exceptionally large population

sizes and apparent ease of dispersal, it was initially assumed that microbes would not be constrained by dispersal limitations; it is worth noting that some studies have indeed failed to establish a definitive link between geographical distance and community dissimilarity, as highlighted in Clark et al. (2021). Other studies, however, have shown microbial communities to exhibit significant distance dependence including communities associated with host organisms (Clark et al., 2021; Green et al., 2004; Hanson et al., 2012; Martiny et al., 2006, 2011; Papke et al., 2003; Whitaker et al., 2003).

Microbes not only inhabit various environmental niches such as soil, sediment and water but can also form intricate associations with multicellular host organisms. These microbial communities often develop into stable, host-specific assemblages, exerting a profound influence on host fitness and fulfilling vital ecological functions (Thompson et al., 2015). In previous studies of the poriferan *Xestospongia* species complex (Cleary et al., 2022; Swierts et al., 2018), we showed that variation in prokaryotic dissimilarity was primarily associated with geography with minor contributions of host dissimilitude and environmental parameters (sea surface temperature and chlorophyll a concentrations). Likewise, other studies have highlighted the importance of space in addition to environment in structuring microbial communities across a range of disparate habitats (Clark et al., 2021; Hanson et al., 2012 and references therein).

In the present study, we tested to what extent space, environment, and host dissimilitude structured prokaryotic communities associated with a golf ball sponge species complex; the name derives from the golf ball-like shape and appearance with numerous porocalices. Golf ball sponges tend to be highly adaptable and can be found across a wide range of habitats including coral reefs, marine lakes, mangroves, and rocky shores (Cleary et al., 2013, 2018b; de Voogd et al., 2009; de Voogd & Cleary, 2008, 2009; Ferreira & Cleary, 2022). It can also be particularly difficult to identify species in the field.

We employed a comprehensive approach combining 16S rRNA gene amplicon sequencing, 28S rRNA gene sequencing of sponge hosts, classical taxonomy, and the use of remotely sensed environmental parameters to investigate golf ball sponges (Family Tetillidae; de Voogd et al., 2022) and their prokaryotic communities. All specimens of golf ball sponges here were identified as belonging to the genera *Cinachyrella* and *Paratetilla*. Previous studies of the family Tetillidae used COI, 18S, and 28S markers (Carella et al., 2016; Szitenberg et al., 2010) to study phylogenetic relationships. Szitenberg et al. (2010) noted that the genus *Cinachyrella* was not monophyletic and Carella et al. (2016) suggested that species of *Cinachyrella* should be moved to the genus *Paratetilla*, as this is the older genus name (1905 vs. 1925).

In line with a previous study of the *Xestospongia* species complex (Cleary et al., 2022), we test the hypotheses that 1. golf ball sponge prokaryotic communities are structured by space, host dissimilitude, and environment and 2. log transformation of prokaryotic community data amplifies the relative importance of space at the expense of host dissimilitude. We believe this to be the case because log transformation increases the relative importance of less abundant community members. Much of the sponge prokaryotic community consists of rare, locally derived microorganisms of largely unknown importance. Giving greater weight to these community members should increase the relative importance of space. In contrast, highly abundant community members may be shared among closely related sponge host species, which should give greater importance to host dissimilitude. In addition to the above, we provide a description of the dominant prokaryotic taxa inhabiting golf ball sponge (morpho) species.

## 2 | MATERIALS AND METHODS

2.1 | Sampling

Samples were collected in various sites across the Indo-Pacific and Caribbean (Data S1). Site descriptions have been previously provided for the Berau (Cleary et al., 2018a; Cleary & Gomes, 2019) and Papua (Cleary et al., 2018b) regions of Indonesia, Taiwan (Cleary et al., 2019; Coelho et al., 2018), Mayotte (de Voogd et al., 2019), Martinique (Cleary, de Voogd, et al., 2023) and Rodrigues (Cleary, Oliveira, et al., 2023). Fragments of the sponge including the surface and interior, in order to sample as much of the prokaryotic community as possible, were stored in tubes containing 96% ethanol. All samples were kept cool (<4°C) immediately after collection and during transport. In the laboratory, samples were stored at −20°C until DNA extraction. Data for all samples including the sampling location and time of sampling can be found in Data S1.

A total of 46 sponge specimens were included in the present study. Voucher samples of all specimens have been deposited in the sponge collection of Naturalis Biodiversity Center (Data S1). Sponge samples were separated into distinct (morpho)-species based on classical taxonomical characters (Figure 1). For each specimen, the dimension and morphology of the silicious spicules were measured and compared with one another and with type material of various *Cinachyrella* species including *C. porosa*, *C. providentiae*, *C. australiensis* and *Paratetilla arcifera*. Using this, we identified 11 groups, that is (morpho)-species from different sampling locations. These included *Cinachyrella* spA from Berau (CaB) and Papua (CaP), *Cinachyrella* spB from Berau (CbB) and Papua (CbP), *Cinachyrella* spN from Taiwan (CnT), *Cinachyrella kuekenthali* (Uliczka, 1929) (CkM) from Martinique, *Cinachyrella alloclada* (Uliczka, 1929) (ClM) from Martinique, *Cinachyrella* spT from Rodrigues (CtR), *Cinachyrella* spU from Rodrigues (CuR), *Cinachyrella* spX (CxS) and spY (CyS) from Saudi Arabia, and *Paratetilla bacca* (Selenka, 1867) from Berau (PbB), Mayotte (PbM) and Taiwan (PbT). In addition to these specimens, we also sampled four specimens of *Cinachyrella* (CnZ) from a zoo aquarium in the Netherlands of unknown origin. Photographs of selected specimens are shown in Figure 1.

2.2 | Host dissimilitude

DNA was extracted from the sponge tissue of 28 specimens using the DNeasy Blood and Tissue kit (Qiagen, Germantown, MD)

FIGURE 1 Selected photographs of focal species. (a) *Cinachyrella alloclada* (CIM) from Martinique, (b) *Cinachyrella kuekenthali* (CkM) from Martinique, (c) *Cinachyrella* spY from Saudi Arabia (CyS), (d) *Cinachyrella* spB from Berau (CbB), (e) *Paratetilla bacca* from Mayotte (PbM), and (f) *Paratetilla bacca* Taiwan (PbT).
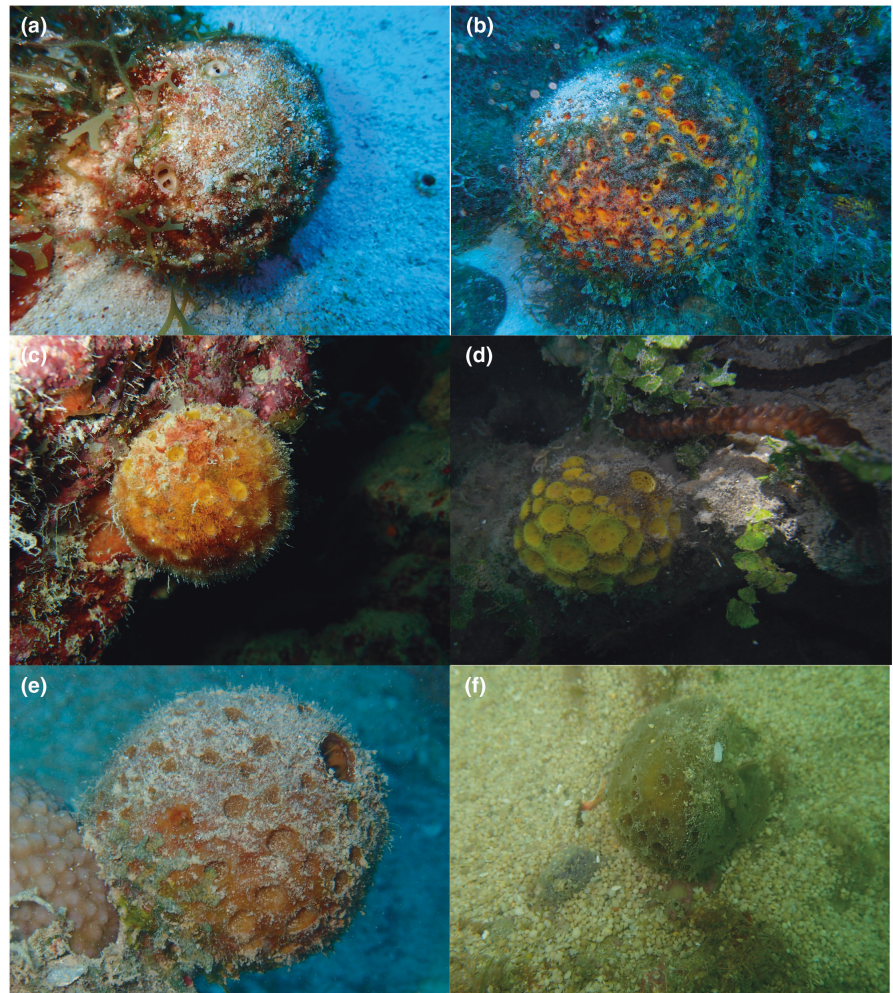
following the manufacturer's instructions. For the nuclear 28S marker, the D3-D5 region of the 28S rRNA gene was amplified using primers f RD3A (5′-GACCCGTCTTGAAACACGA-3′) and RD5B2 (5′-ACACACTCCTTAGCGGA-3′), which amplified a fragment of 596 bps. Amplifications were carried out in 25 μL reaction volumes containing 8.9 μL of MilliQ water, 5 μL 5× Phire® reaction buffer, 5 μL Q-solution, 2 μL dNTPs (2.5 mM), 1.3 μL of each primer (10 μM), 0.5 μL Phire® Hotstart II Taq polymerase, and 1 μL DNA. The PCR profile was as follows: an initial denaturing step of 98°C for 30 s, 35 cycles of denaturing at 98°C for 5 s, annealing at 50°C for 5 s, extension at 72°C for 15 s, and a final extension step at 72°C for 5 min. Sequence lengths of the 28S rRNA nucleotide data varied from 540 to 824 bps with an average of 705 bps.

We constructed a phylogenetic tree with the 28S rRNA nucleotide sequences, including specimens of all sponge (morpho)species, with the MEGAX program (http://www.megasoftware.net/; last checked 2019/09; Tamura et al., 2011). In addition to these sequences, we included 28S sequences of two other species belonging to the family Tetillidae, *Tetilla japonica* (GenBank #: JX177956) and *Acanthotetilla celebensis* (GenBank #: JX177920) as outgroups. The 28S rRNA gene sequences were first aligned using the ClustalW algorithm in MEGAX. A maximum-likelihood phylogenetic tree was subsequently built using the generalised time-reversible model

(Tavaré, 1986) with discrete Gamma distribution and invariant sites. Initial tree(s) for the heuristic search were automatically obtained with Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, followed by selection of the topology with superior log likelihood value. Branches with <50% bootstrap support were collapsed. In the results, we present a bootstrap consensus tree based on 100 replicates (Felsenstein, 1985). The bootstrap value is shown next to each branch when this exceeds 50%. This value represented the percentage of replicate trees in which the associated taxa clustered together.

2.3 | DNA extraction for next-generation sequencing analysis

PCR-ready genomic DNA was isolated from all samples using the FastDNA® SPIN soil Kit (MPbiomedicals) following the manufacturer's instructions. Briefly, ±500 mg of sponge tissue (including parts of the surface and interior) was added. The microbial cell lysis was performed in the FastPrep® Instrument (Q Biogene) for 80 s at 6.0 ms^{-1}. The extracted DNA was eluted into DNase/Pyrogen-Free water to a final volume of 50 μL and stored at −20°C until use. The 16S rRNA gene V3-V4

variable region PCR primers 341F 5′-CCTACGGGNGGCWGCAG-3′ and 785R 5′-GACTACHVGGGTATCTAATCC-3′ (Klindworth et al., 2013) with barcode on the forward primer were used in a 30-cycle PCR assay using the HotStarTaq Plus Master Mix Kit (Qiagen, USA) under the following conditions: 94°C for 3 min, followed by 28 cycles of 94°C for 30 s, 53°C for 40 s and 72°C for 1 min, after which a final elongation step at 72°C for 5 min was performed. After amplification, PCR products were checked on a 2% agarose gel to determine the success of amplification and the relative intensity of bands; the blank control did not yield any bands. PCR product was used to prepare the DNA library following the Illumina TruSeq DNA library preparation protocol. Next-generation, paired-end sequencing was performed at MrDNA (Molecular Research LP; http://www.mrdna lab.com/; last checked 18 November 2016) on an Illumina MiSeq device (Illumina Inc., San Diego, CA, USA) following the manufacturer's guidelines. Sequences from each end were joined following Q25 quality trimming of the ends followed by reorienting any 3′–5′ reads back into 5′–3′ and removal of short reads (<150 bp).

2.4 | 16S rRNA gene sequencing analysis

The 16S rRNA amplicon libraries were analysed using QIIME2 (version 2019.7; Bolyen et al., 2019). Raw data were imported yielding a demultiplexed 'qza' data file (artefact). The DADA2 plugin (Callahan et al., 2016) in QIIME 2 was subsequently used to trim sequences (final length 400 nt). The DADA2 analysis yielded output archives containing an OTU (at 100% sequence similarity, also known as amplicon sequence variant or 'ASV') table, denoising stats and a fasta file of representative sequences. The feature-classifier plugin with the extract-reads method was then used with the i-sequences argument set to silva-138-99-seqs.qza. This was followed by the feature-classifier plugin with the fit-classifier-naive-bayes method and the i-reference-taxonomy method set to silva-138-99-tax.qza. Both silva-138 files can be obtained from https://docs.qiime2.org/2020.8/data-resou rces/?highlight=silva; last checked 2024-02-20. The feature-classifier plugin was then used with the classify-sklearn method and the i-reads argument was set to the representative sequences file generated by the DADA2 analysis to produce a table with taxonomic assignment for all OTUs. Finally, mitochondria, chloroplasts, and Eukaryota were filtered out using the QIIME taxa plugin with the filter-table method. The OTU and taxonomy tables were later merged in R. After denoising, removal of chimera, mitochondria, chloroplasts, Eukaryota, and OTUs unclassified at Domain or Phylum level, the final dataset consisted of 3,975,174 sequences and 15,348 OTUs.

2.5 | Remotely sensed data

Global, remotely sensed data were downloaded from the National Oceanic and Atmospheric Administration OceanWatch website (https://oceanwatch.pifsc.noaa.gov/erddap/index.html; last checked 2020-12-11). This consisted of cumulative long-term mean (January 2003–February 2019) Aqua MODIS data for chlorophyll a (Chla) and Coral Reef Watch data (2003–2017) for Sea Surface Temperature (SST). All data were downloaded in netCDF format and imported into R (R Core Team, 2020) using the raster function from the raster package. Missing values were added using the knnImputation function from the DMwR package (Torgo, 2010) with a K of 2. A map of the study area showing long-term variation in SST, including the locations of the sampling sites, is shown in Figure 2.

2.6 | Statistical analyses

Tables containing OTU counts, geographic coordinates, SST, and Chla values were imported into R. Data S2 contains all operational taxonomic unit (OTU) counts per sample and taxonomic classifications of all OTUs. The 50 most abundant OTUs are presented in Data S3. The OTU counts matrix was used to compare diversity among host groups. Diversity indices were obtained using the rarefy and diversity functions from the vegan (Oksanen et al., 2020) package in R.

Variation in OTU composition among biotopes was visualised with principal coordinates analysis (PCO) using the Phyloseq package (McMurdie & Holmes, 2013). Prior to PCO analysis, the count data matrix was rarefied using the rarefy_even_depth function from the Phyloseq package with the sample.size argument set to the minimum sample size (12,398 in the present study). Subsequent to rarefaction, two matrices were generated consisting of an untransformed matrix and a $\log_{10}$ $(x+1)$ transformed matrix. In multidimensional scaling (MDS) and other ordination techniques, log transformation scales down the effect of abundant species, thus amplifying the affect of rarer species (Field et al., 1982; Ribera et al., 2001; van Tongeren, 1995).

In addition to data visualisation using PCO, we tested for significant associations between prokaryotic dissimilarity and space (geographic distance), host dissimilitude, and environment. For geographic distance, a distance matrix was obtained using the earth.dist function from the fossil package. This creates a distance matrix of pairwise distances in kms between all sample sites. The environmental data matrix (consisting of SST and Chla values) was obtained using the vegdist function in vegan with the method argument set to 'euclidean'. The 28S sequences were used to generate a matrix of host dissimilitude with the dist.ml function in the R package phangorn (Schliep, 2011).

Relationships between dissimilarity and geographic distance have been shown to follow power laws for a wide range of taxa (Clark et al., 2021). In the present study, we applied the Levenberg–Marquardt algorithm using the nlsLM function from the minpack.lm library in R to estimate the parameters of power law models of dissimilarity versus geographic distance. The MRM function from the ecodist library in R was used with 1000 permutations to carry out multiple regressions on distance matrices with the Bray–Curtis dissimilarity matrix as response variable and spatial (geographic distance), environmental (SST and Chla) and host dissimilitude distance matrices as predictor variables. Variance partitioning
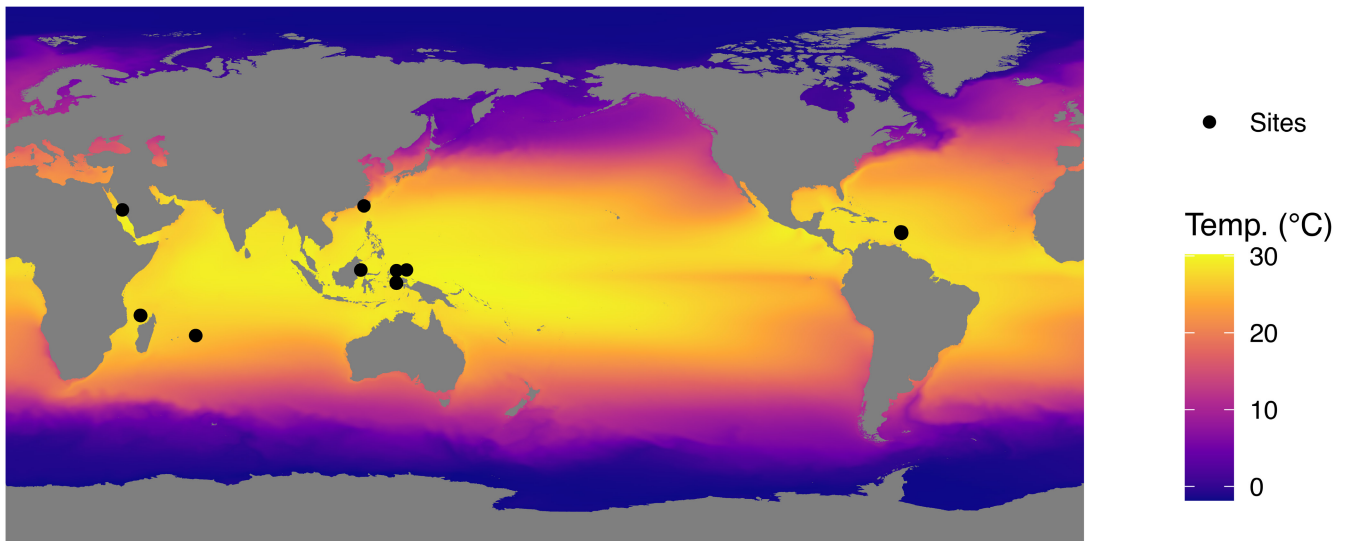
FIGURE 2 Large-scale variation in sea surface temperature (SST); the sampling locations are indicated by circles. Long-term mean values of SST for each sample site were obtained from satellite-derived data, which was downloaded from the National Oceanic and Atmospheric Administration OceanWatch website. These data were imported into R, and the extract function from the raster package was used to obtain values for each sample site. Missing values were added using the knnImputation function from the DMwR package with a K of 2.

(Borcard et al., 1992) was used to partition the variance in community dissimilarity attributable to space, host dissimilitude and environment in addition to the union (e.g., the spatially structured environmental component) of all data sets. Tests were performed using untransformed and log-transformed response dissimilarity matrices. Previous studies that have used a distance-based variance partitioning approach to disentangle the impact of spatial and other factors, for example environmental, on beta diversity include Duivenvoorden et al. (2002), Tuomisto et al. (2003), Cleary et al. (2004), Jones et al. (2006) and Lichstein (2007). More recent microbiome studies, which have used MRM to study beta diversity, include Chen et al. (2020), Díez-Vives et al. (2020), Hou et al. (2020), Locey et al. (2020) and Wang et al. (2020).

3 | RESULTS

Results of the phylogenetic analysis are presented in Figure 3 with six well-supported clusters (bootstrap values >90%) evident, namely cluster 1, consisting of *Cinachyrella* spB from Berau and Papua, *Cinachyrella* spT from Rodrigues, *Cinachyrella* spX from Saudi Arabia and the specimens of unknown origin from Burgers' Zoo; cluster 2, consisting of *C. alloclada*; cluster 3, consisting of *Cinachyrella* spA and *Cinachyrella* spN from Taiwan; cluster 4, consisting of *P. bacca*; cluster 5, consisting of *Cinachyrella* spY from Saudi Arabia and *Cinachyrella* spU from Rodrigues; and cluster 6, consisting of *C. kuekenthali*. These six clusters were included in two well-supported major clusters consisting of specimens from clusters 1, 2 and 3 in one major cluster and specimens from clusters 4, 5 and 6 in the other.

The most abundant higher taxa in the present study were Proteobacteria (6497 OTUs, 2,440,118 sequences), Actinobacteriota (783 OTUs, 453,771 sequences), Chloroflexi (709 OTUs, 236,175

sequences) and NB1-j (492 OTUs, 203,855 sequences). Plots of evenness (Pielou's J), richness, Shannon's H′, and Fisher's alpha diversity indices are shown in Figure 4. The mean relative abundances of the eight most abundant taxa are presented in Figure 5. These taxa accounted for 91.82% of all sequences and 66.74% of all OTUs. Proteobacteria and Actinobacteria were abundant in all groups with the highest abundance of Proteobacteria in *Cinachyrella* spA from Berau. Of note are the relatively high abundances of Chloroflexi in *P. bacca*, *Cinachyrella* spY from Saudi Arabia and *C. kuekenthali*, and NB1-j in *Cinachyrella* spZ from Burgers' Zoo, *Cinachyrella* spT from Rodrigues, and *C. alloclada*.

In the PCO analysis of untransformed data (Figure 6), the first axis separated samples of *P. bacca* from samples of *C. kuekenthali*. The second axis separated all samples of *P. bacca* and *C. kuekenthali* from samples of the remaining (morpho)species. Samples of *Cinachyrella* spB from Berau and Papua clustered with *Cinachyrella* spX from Saudi Arabia in line with the phylogenetic analysis of the host sponges. Specimens of *Cinachyrella* sp. from Burgers Zoo, however, clustered together with samples of *Cinachyrella* spU from Rodrigues in contrast to the phylogenetic analysis. The single sample of *Cinachyrella* sp. from Taiwan clustered with *Cinachyrella* spA in the phylogenetic analysis and the PCO ordination. In the PCO analysis of log-transformed data (Figure 7), the first axis separated samples of *Cinachyrella* species from Martinique and Rodrigues from the remaining samples. The second axis separated samples of *Cinachyrella* species from Martinique from samples of *Cinachyrella* species from Rodrigues. In Figure 7, samples tended to cluster according to geographical proximity (e.g. *C. kuekenthali* and *C. alloclada* from Martinique, and *Cinachyrella* SpT and spU from Rodrigues), in contrast to the results presented in Figure 6.

The 50 most abundant OTUs are presented in Figure 8 and listed in Data S3. Of these, the most frequent families were the
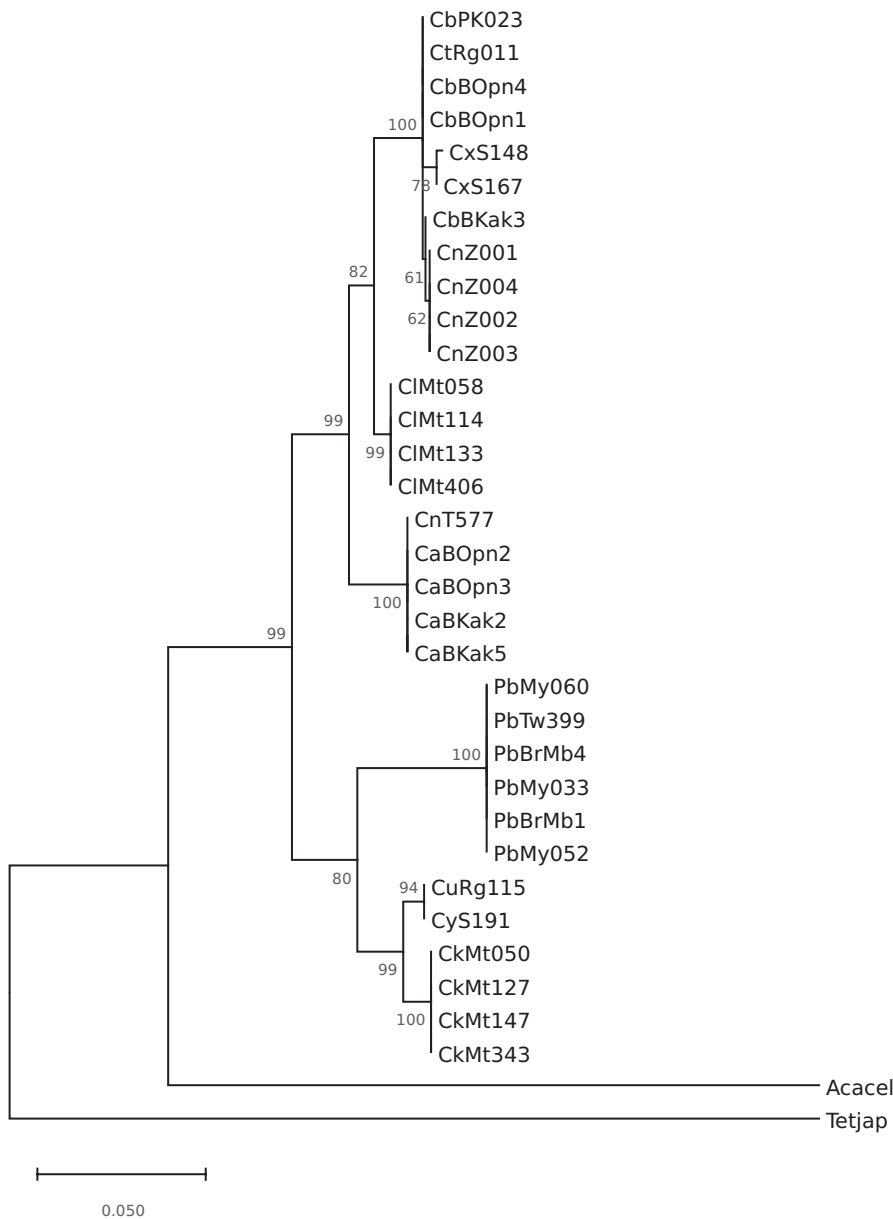
FIGURE 3 Phylogenetic analysis using Maximum Likelihood and 100 bootstraps obtained with MEGAX. Branches with <50% bootstrap support were collapsed. The bootstrap value is shown next to each branch when this exceeds 50%. The first three letters of the sample codes refer to: *Cinachyrella* spA from Berau (CaB) and Papua (CaP), *Cinachyrella* spB from Berau (CbB) and Papua (CbP), *Cinachyrella* sp. from Taiwan (CnT), *Cinachyrella kuekenthali* (CkM) from Martinique, *Cinachyrella alloclada* (ClM) from Martinique, *Cinachyrella* spT from Rodrigues (CtR), *Cinachyrella* spU from Rodrigues (CuR), CxS: *Cinachyrella* spX from the Red Sea, CyS: *Cinachyrella* spY from the red Sea, *Paratetilla bacca* from Berau (PbB), Mayotte (PbM), and Taiwan (PbT), Tetjap: *Tetilla japonica*, and Acacel: *Acanthotetilla celebensis*.

Nitrosococcaceae (9), Rhodobacteraceae (7), EC94 (7), NB1-j (4) and Microtrichaceae (4). Most (41 of 50) of these abundant OTUs were related to microorganisms previously obtained from a range of sponge species, including *Amphimedon queenslandica*, *Ancorina alata*, *Astrosclera willeyana*, *Cinachyra* sp., *Coelocarteria singaporensis*, *Corallistes* sp., *Cymbastella coralliophila*, *Holoxea* sp., *Ircinia variabilis*, *Isops phlegraei*, *Vaceletia crypta* and *Xestospongia muta* (Data S3). Furthermore, 19 of the 50 OTUs had very high (>99%) sequence similarities to sequences in GenBank, while 5 had relatively low (<93%) sequence similarities. Of these, one (OTU-14) was most dominant in *Cinachyrella* sp. from Taiwan, one (OTU-113) in *Cinachyrella* spB, one (OTU-136) in *C. alloclada*, and two (OTUs 76 and 138) in *P. bacca*. The OTUs shown in Figure 8 also revealed pronounced enrichment in particular host species. OTU-1, for example, was most abundant in samples of *P. bacca* and was closely related (sequence similarity >99%) to an organism obtained from a sponge identified as *Cinachyra* sp. from the Great Barrier Reef (Data S3). The

most abundant OTU in *C. kuekenthali*, OTU-6, was also closely related (sequence similarity = 100%) to an organism obtained from a sponge identified as *Cinachyra* sp. from the Gulf of Mannar, India. The most abundant OTU in *C. alloclada* (OTU-4) had 96% similarity to an organism obtained from a sponge identified as *Coelocarteria singaporensis*. OTU-16 was the most abundant OTU in *Cinachyrella* spB from Berau, *Cinachyrella* spT from Rodrigues and *Cinachyrella* spX (CxS) from Saudi Arabia, all of which clustered together in the phylogenetic analysis (Figure 3). It had 97% similarity to an organism obtained from a sponge identified as *Ircinia variabilis* (Data S3).

There were highly significant associations of space (MRM: Untransformed: $F = 93.34$, $R^2 = .199$, $p < .001$; Log-transformed: $F = 343.69$, $R^2 = .478$, $p < .001$), host dissimilitude (MRM: Untransformed: $F = 87.34$, $R^2 = .189$, $p < .001$; Log-transformed: $F = 87.34$, $R^2 = .189$, $p < .001$), environment (MRM: Untransformed: $F = 23.23$, $R^2 = .058$, $p < .001$; Log-transformed: $F = 73.73$, $R^2 = .164$, $p < .001$), and the full model, which included all of the previous three
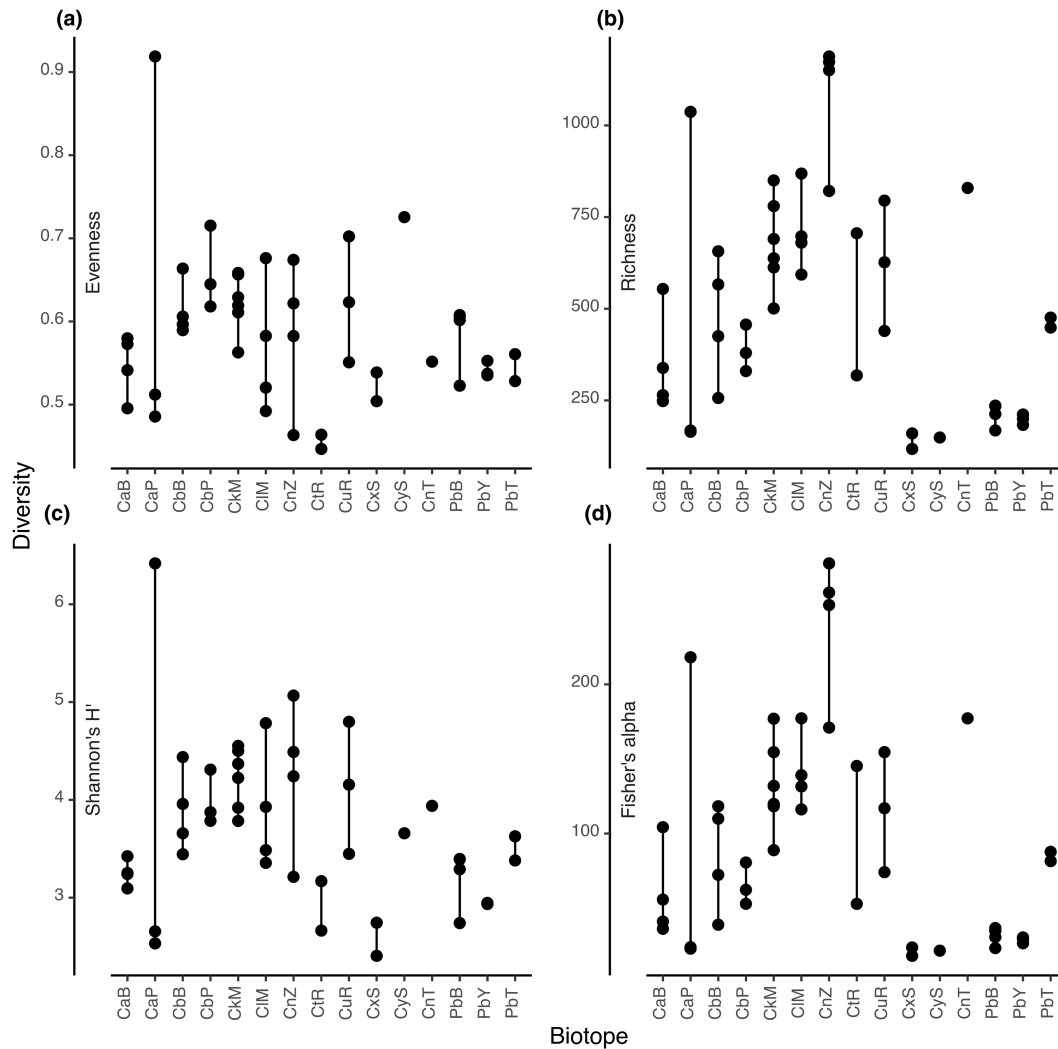
FIGURE 4 Values of (a) Evenness, (b) Richness, (c) Shannon's H′, and (d) Fisher's alpha diversity indices for CaB: *Cinachyrella* spA Berau, CaP: *Cinachyrella* spA Papua, CbB: *Cinachyrella* spB Berau, CbP: *Cinachyrella* spB Papua, CkM: *Cinachyrella kuekenthali*, ClM: *Cinachyrella alloclada*, CnT: *Cinachyrella* sp. Taiwan, CnZ: *Cinachyrella* sp. Burgers Zoo, CtR: *Cinachyrella* spT Rodrigues, CuR: *Cinachyrella* spU Rodrigues, CxS: *Cinachyrella* spX Red Sea, CyS: *Cinachyrella* spY Red Sea, PbB: *Paratetilla bacca* Berau, PbT: *Paratetilla bacca* Taiwan and PbY: *Paratetilla bacca* Mayotte.

predictors (Full additive MRM model: Untransformed: $F = 127.11$, $R^2 = .505$, $p < .001$; Log-transformed: $F = 179.28$, $R^2 = .590$, $p < .001$) with prokaryotic dissimilarity. The relationships of prokaryotic dissimilarity with space (geographic distance) and host dissimilitude are presented in Figure 9 (untransformed data) and 10 (log-transformed data). Histograms of Bray–Curtis dissimilarities exhibited pronounced left-skewed distributions for both untransformed and log-transformed data (Figures 9c and 10c). The masses of dissimilarities were thus concentrated at high values with a long tail of lower dissimilarities.

For the untransformed data, space, host dissimilitude, and environment explained >50% of the variation in Bray–Curtis dissimilarity (Figure 9d). The purely spatial component (geographic distance alone) explained 7.8%, the purely host dissimilitude component 30.48%, and the purely environmental component 0.27% of the variation in composition. The union of all three components (space, host dissimilitude, and environment) explained a further 1.61% of the variation in dissimilarity while the union of space and

host dissimilitude explained an additional 6.38% and environment and space an additional 4.09%. For the log-transformed data, the purely spatial component explained 25.5%, the purely host dissimilitude component 10.46%, and the purely environmental component 0.97% of the variation in dissimilarity. The union of all three components explained a further 1.96% of variation in dissimilarity while the union of space and host dissimilitude explained an additional 6.63% and environment and space an additional 13.67%.

4 | DISCUSSION

Golf ball sponges are small, morphologically similar, and difficult to identify to species level in the field. They are present across a range of habitat types, including reef flats, mangroves and perturbed reef areas and are often covered with sediment making it, sometimes, difficult to distinguish them from the surrounding environment.
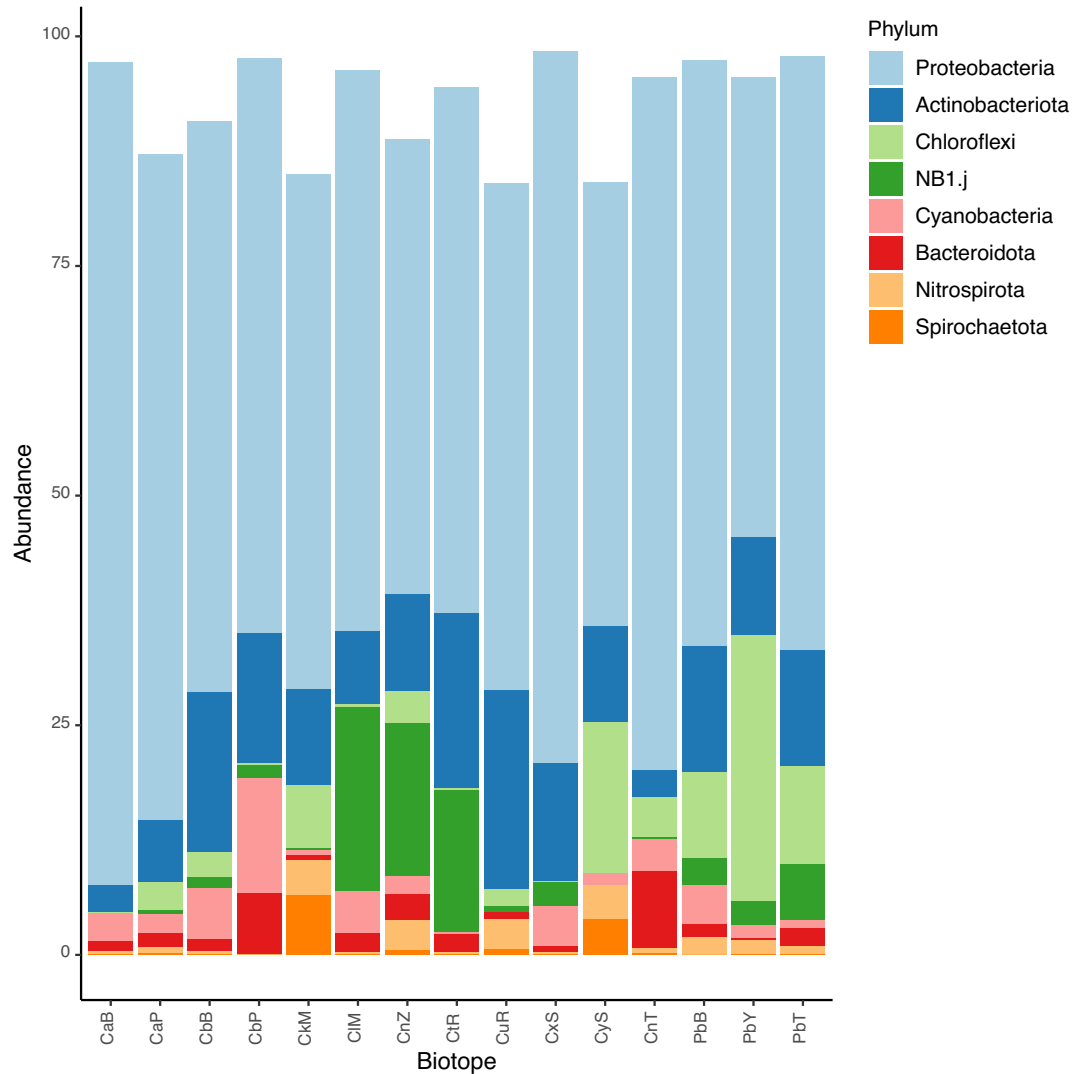
FIGURE 5 Stacked barplots showing the most abundant phyla in CaB: *Cinachyrella* spA Berau, CaP: *Cinachyrella* spA Papua, CbB: *Cinachyrella* spB Berau, CbP: *Cinachyrella* spB Papua, CkM: *Cinachyrella kuekenthali*, ClM: *Cinachyrella alloclada*, CnT: *Cinachyrella* sp. Taiwan, CnZ: *Cinachyrella* sp. Burgers Zoo, CtR: *Cinachyrella* spT Rodrigues, CuR: *Cinachyrella* spU Rodrigues, CxS: *Cinachyrella* spX Red Sea, CyS: *Cinachyrella* spY Red Sea, PbB: *Paratetilla bacca* Berau, PbT: *Paratetilla bacca* Taiwan and PbY: *Paratetilla bacca* Mayotte.

They are also present across a very large spatial range, including Indo-Pacific, Caribbean, and Antarctic waters (Carella et al., 2016).

4.1 | Dominant prokaryotic taxa inhabiting golf ball sponges

Several bacterial taxa were well represented in the list of most abundant OTUs including OTUs classified as belonging to the Rhodobacteraceae, Nitrosococcaceae, NB1-j, EC94, and Microtrichaceae families. Rhodobacteraceae are a major component of marine ecosystems, where they can be found in the bacterioplankton and as symbionts of multicellular organisms (Dogs et al., 2017; Elifantz et al., 2013; Simon et al., 2017).

Abundant Nitrosococcaceae OTUs, most of which were classified to the genus AqS1, were observed in *C. kuekenthali*, *Cinachyrella* spA, *Cinachyrella* spB, *Cinachyrella* spT, *Cinachyrella* spU from Rodrigues,

and *Cinachyrella* spX from Saudi Arabia. Nitrosococcaceae is a family of ammonia-oxidising bacteria within the Gammaproteobacteria. A draft genome of AqS1 was previously identified from the sponge *Amphimedon queenslandica* and indicated that it was capable of sulphur oxidation, carbon monoxide oxidation, and inorganic phosphate assimilation (Gauthier et al., 2016). In particular, the AqS1 genome was shown to encode for a Sox enzyme complex involved in the oxidation of reduced sulphur compounds in addition to a reverse dissimilatory sulphate reduction pathway. AqS1 has also been shown to be enriched in ankyrin repeat-containing proteins. These are putatively used to promote host-microbe interactions and facilitate a symbiotic lifestyle, for example, by inhibiting eukaryotic phagocytosis. In addition to ankyrin repeats, tetratricopeptide repeats, leucine-rich repeats, fibronectin domain IIIs, and cadherins putatively fulfill similar functions (references in Gauthier et al., 2016).

Abundant members of the EC94 clade were observed in *Cinachyrella* spA, *C. alloclada*, and *P. bacca*. Taylor et al. (2021)

FIGURE 6 Ordination of untransformed data showing the first two axes of the principal coordinates analysis (PCO) of OTU composition. Symbols are colour-coded and represent samples from different groups as shown in the legend on the right side of the figure. Grey symbols represent weighted averages scores for OTUs. The symbol size is proportional to group abundance (number of sequence reads). The legend symbols represent: CaB: *Cinachyrella* spA Berau, CaP: *Cinachyrella* spA Papua, CbB: *Cinachyrella* spB Berau, CbP: *Cinachyrella* spB Papua, CkM: *Cinachyrella kuekenthali*, ClM: *Cinachyrella alloclada*, CnT: *Cinachyrella* sp. Taiwan, CnZ: *Cinachyrella* sp. Burgers Zoo, CtR: *Cinachyrella* spT Rodrigues, CuR: *Cinachyrella* spU Rodrigues, CxS: *Cinachyrella* spX Red Sea, CyS: *Cinachyrella* spY Red Sea, PbB: *Paratetilla bacca* Berau, PbT: *Paratetilla bacca* Taiwan, and PbY: *Paratetilla bacca* Mayotte.
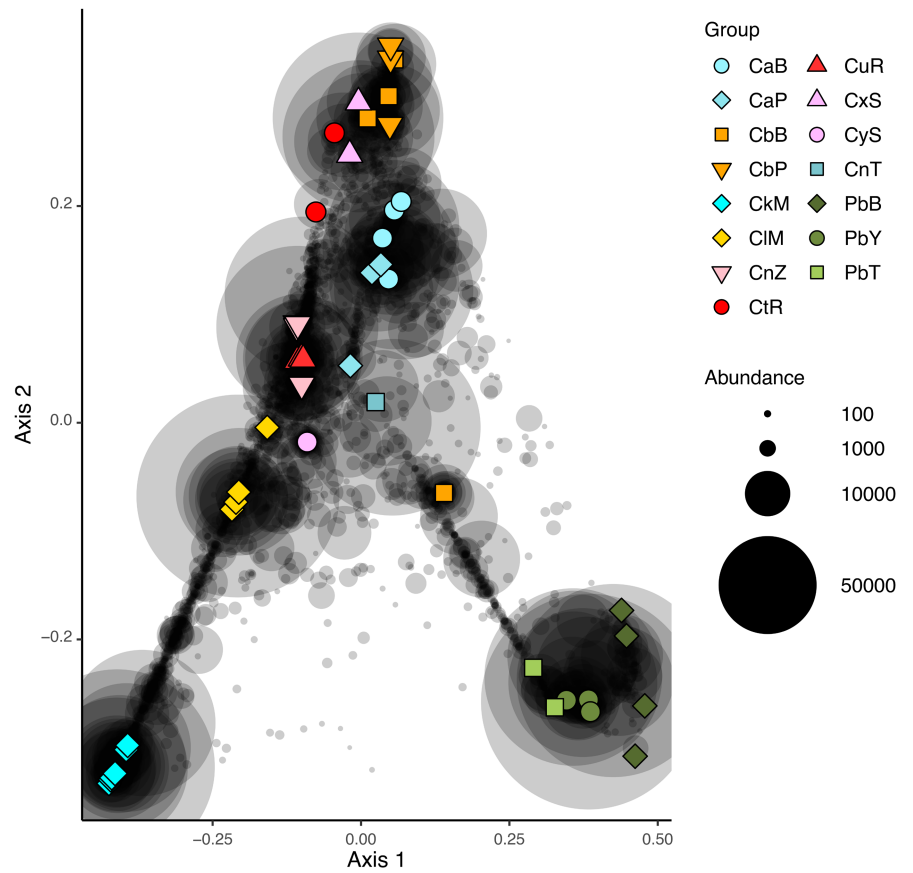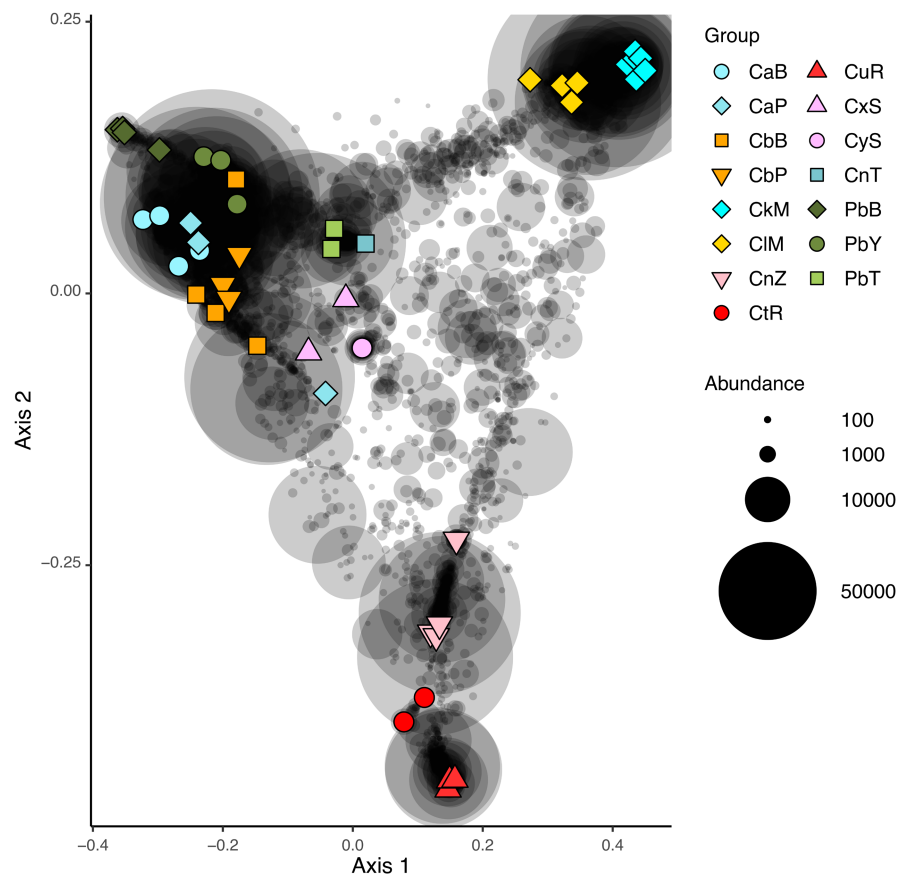

FIGURE 7 Ordination of log-transformed data showing the first two axes of the principal coordinates analysis (PCO) of OTU composition. Symbols are colour-coded and represent samples from different groups as shown in the legend on the right side of the figure. Grey symbols represent weighted averages scores for OTUs. The symbol size is proportional to group abundance (number of sequence reads). The legend symbols represent: CaB: *Cinachyrella* spA Berau, CaP: *Cinachyrella* spA Papua, CbB: *Cinachyrella* spB Berau, CbP: *Cinachyrella* spB Papua, CkM: *Cinachyrella kuekenthali*, ClM: *Cinachyrella alloclada*, CnT: *Cinachyrella* sp. Taiwan, CnZ: *Cinachyrella* sp. Burgers Zoo, CtR: *Cinachyrella* spT Rodrigues, CuR: *Cinachyrella* spU Rodrigues, CxS: *Cinachyrella* spX Red Sea, CyS: *Cinachyrella* spY Red Sea, PbB: *Paratetilla bacca* Berau, PbT: *Paratetilla bacca* Taiwan, and PbY: *Paratetilla bacca* Mayotte.
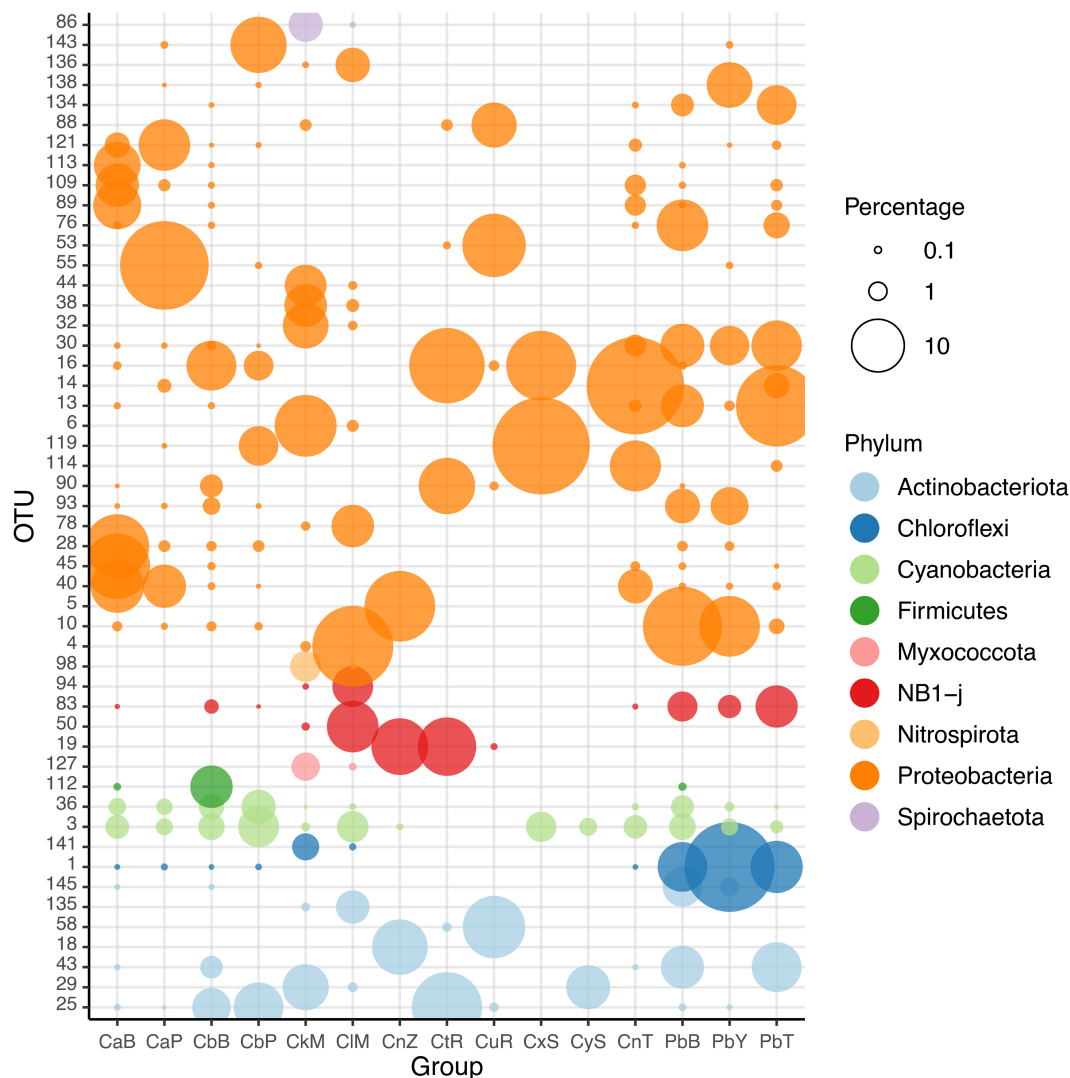
FIGURE 8 Relative abundances of the most abundant OTUs colour-coded according to prokaryotic phylum for: CaB: *Cinachyrella* spA Berau, CaP: *Cinachyrella* spA Papua, CbB: *Cinachyrella* spB Berau, CbP: *Cinachyrella* spB Papua, CkM: *Cinachyrella kuekenthali*, ClM: *Cinachyrella alloclada*, CnT: *Cinachyrella* sp. Taiwan, CnZ: *Cinachyrella* sp. Burgers Zoo, CtR: *Cinachyrella* spT Rodrigues, CuR: *Cinachyrella* spU Rodrigues, CxS: *Cinachyrella* spX Red Sea, CyS: *Cinachyrella* spY Red Sea, PbB: *Paratetilla bacca* Berau, PbT: *Paratetilla bacca* Taiwan, and PbY: *Paratetilla bacca* Mayotte. The circle size of the OTU is proportional to the mean percentage of sequences per biotope as indicated by the symbol legend in the upper right corner of the figure.

included the EC94 clade in their novel proposed order Tethybacterales and noted that its members were associated with sponges suggesting a symbiotic relationship with sponge hosts. As with AqS1, the genomes of Ca. Tethybacterales members were enriched with ankyrin repeat proteins highlighting their potential symbiotic nature. In addition to this, Ca. Tethybacterales genomes contained genes for lipoprotein release via the ABC-2 transporter LolABCDE complex, which have been reported to play roles in bacterial colonisation, host defence evasion, and immunomodulation (Taylor et al., 2021).

Abundant NB1-j members were observed in *Cinachyrella alloclada*, *Cinachyrella* sp. from Burgers Zoo, *Cinachyrella* spT from Rodrigues, and *P. bacca*. Relatively little is known about the NB1-j clade. It was previously identified as a dominant member of the alfalfa rhisosphere (Zhu et al., 2022), saltmarsh sediment (Cleary et al., 2016),

and the microbiomes of the LMA sponge species *Stylissa carteri* and *S. massa* (Cleary et al., 2015).

Abundant Microtrichaceae members, most of which belonged to the Sva0996 marine group, were observed in *C. kuekenthali*, *Cinachyrella* spY (CyS) from the Red Sea, *Cinachyrella* spU from Rodrigues, and *P. bacca*. Relatively little is known about the Sva0996 marine group, but members have been reported to be involved in nitrogen cycling and assimilation of dissolved protein (Orsi et al., 2016). Sva0996 was also reported to be an abundant component of the microbiomes of the brown alga *Laminaria digitata* (Ihua et al., 2020), and the carnivorous sponge *Chondrocladia grandis* (Verhoeven et al., 2016).

In addition to the above-mentioned taxa, certain host sponges were enriched with abundant Chloroflexi members, mostly belonging to the SAR202 clade, including *Cinachyrella kuekenthali* (OTU-141), *Cinachyrella* spY from Saudi Arabia (OTU-427), and *P. bacca*
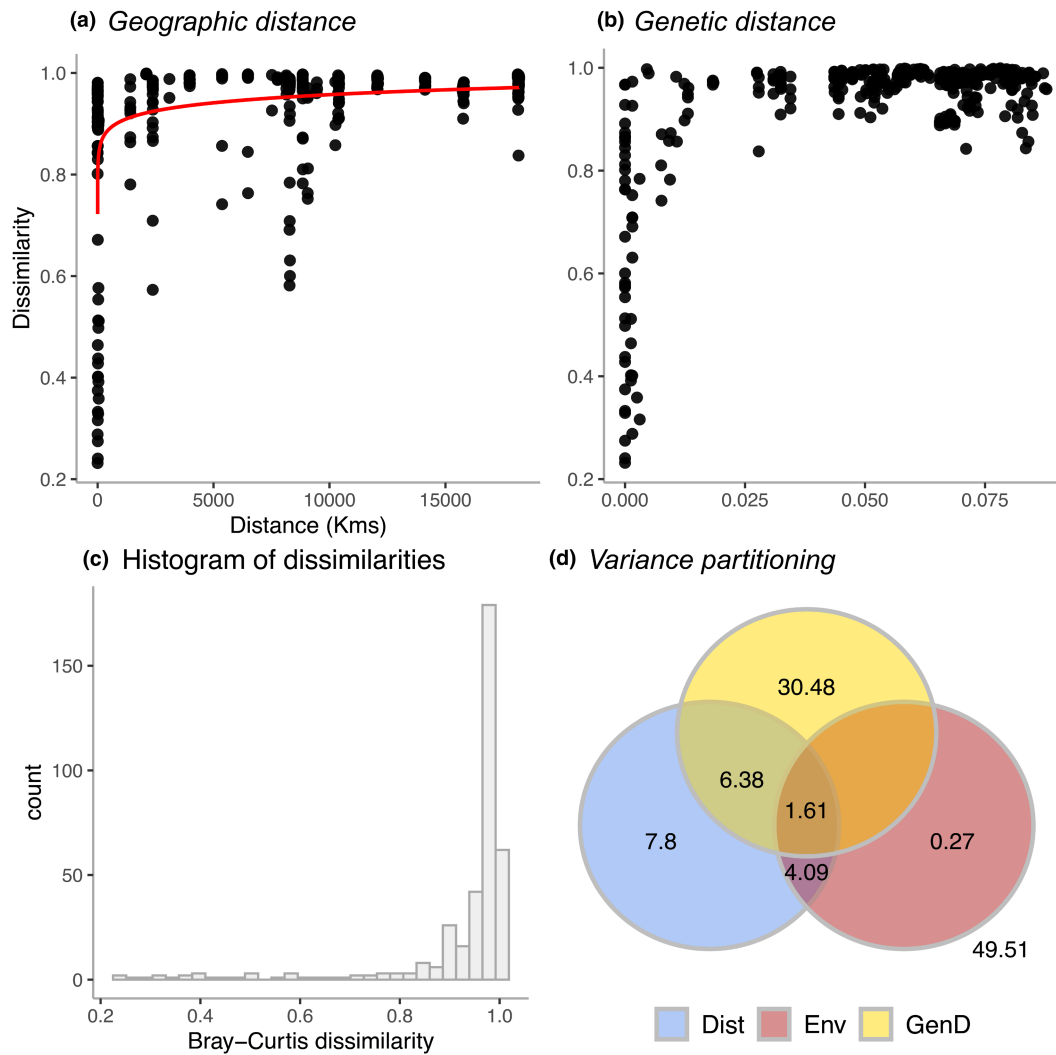
FIGURE 9 Variation in untranformed community dissimilarity as a function of (a) distance between sampling sites and (b) maximum likelihood distance. Each point represents a single pairwise comparison of Bray–Curtis similarity between a pair of sites. The fitted lines in (a) were obtained using power law functions. (c) Histogram of the Bray–Curtis dissimilarities between pairs of samples. (d) Venn diagrams showing the results of variance partitioning analysis based on MRM. Nonoverlapping areas indicate purely spatial (Dist), purely host dissimilitude (GenD) and purely environmental (Env) components. Overlapping areas represent the union of components.

(OTU-1). OTU-1 was, furthermore, the most abundant OTU over-all. Previous studies have suggested that SAR202 members play a role in the degradation of recalcitrant dissolved organic matter such as humic substances (Colatriano et al., 2018; Landry et al., 2017). Importantly, *P. bacca* and other *Paratetilla* species are often found in coastal areas, in proximity to riverine outflow. They also appear to tolerate high sediment loads and are often covered with sediment (pers. obs. both authors).

4.2 | Role of space, host dissimilitude and environment in structuring prokaryotic communities inhabiting golf ball sponges

The sponge taxa (clusters) identified using the phylogenetic analy-sis housed compositionally distinct prokaryotic communities. The

samples from Burgers' Zoo, the Netherlands, however, proved to be an exception. They all belonged to the cluster, including *Cinachyrella* spB, but housed prokaryotic communities compositionally similar to samples of *Cinachyrella* spU from Rodrigues. It is unclear why this should be the case. It is noteworthy that the *Cinachyrella* specimens from Burgers' Zoo maintained a prokaryotic community similar to wild specimens, but living in an artificial environment probably had an impact on their microbiomes.

Prokaryotic dissimilarity varied as a function of geographic distance and host dissimilitude with the most pronounced shifts in dissimilarity at the smallest spatial scales and degree of host dissimilitude. There was, furthermore, considerable variation at the smallest scales with dissimilarity ranging from 0.2 (0.4 for log-transformed data) to >0.9 as shown in Figures 9 and 10. The purely spatial component (space alone) varied from 7.8% for untrans-formed data to 25.5% for log-transformed data. In the barrel sponge
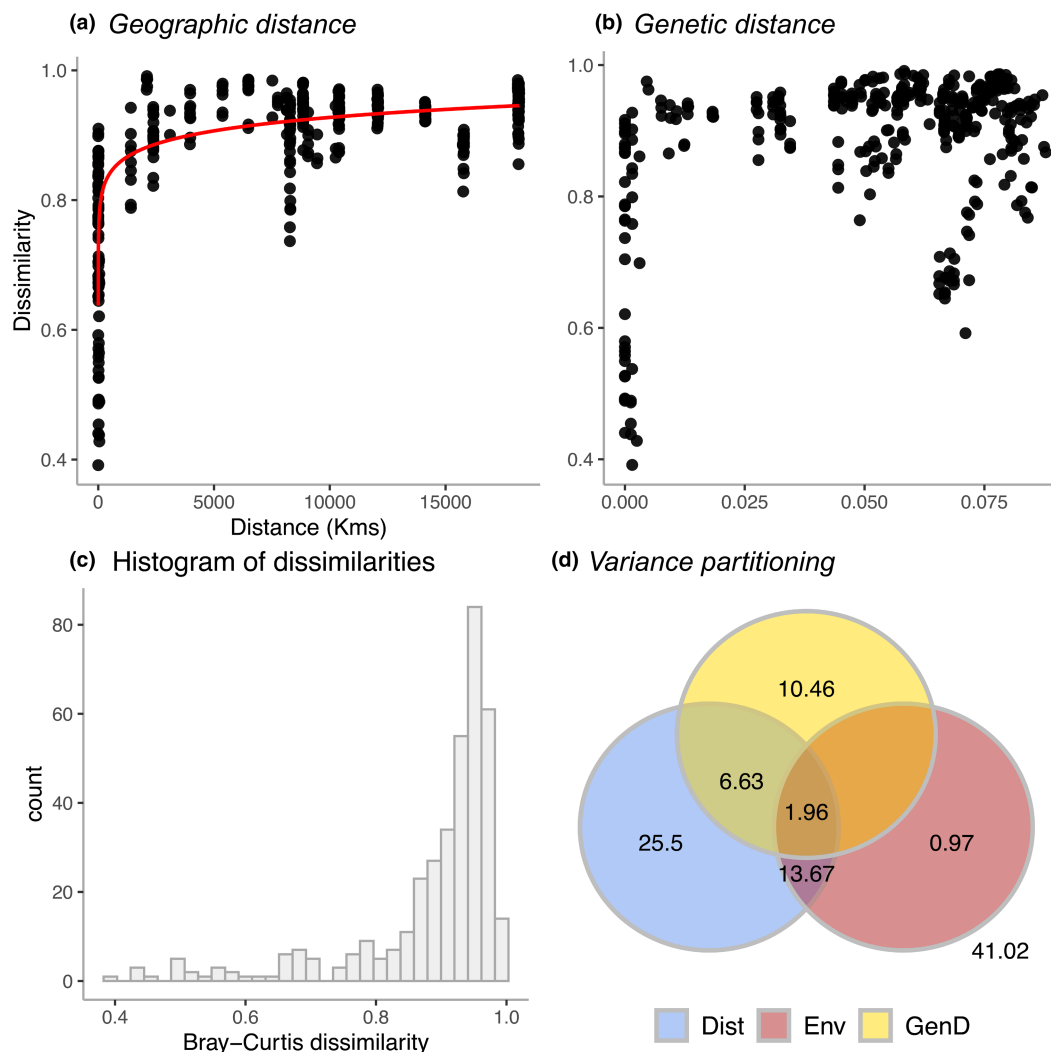
FIGURE 10 Variation in log-transformed community dissimilarity as a function of (a) distance between sampling sites and (b) maximum likelihood distance. Each point represents a single pairwise comparison of Bray–Curtis similarity between a pair of sites. The fitted lines in (a) were obtained using power law functions. (c) Histogram of the Bray–Curtis dissimilarities between pairs of samples. (d) Venn diagrams showing the results of variance partitioning analysis based on MRM. Nonoverlapping areas indicate purely spatial (Dist), purely host dissimilitude (GenD) and purely environmental (Env) components. Overlapping areas represent the union of components.

species complex (*Xestospongia* spp.), the purely spatial component explained 39% of total variation (Cleary et al., 2022). The importance of space in structuring microbial communities in general, and host-associated microbial communities in particular, aligns with recent microbial studies (Hanson et al., 2012 and references therein; Chen et al., 2020; Clark et al., 2021; Díez-Vives et al., 2020; Hou et al., 2020; Locey et al., 2020; Wang et al., 2020) and earlier studies of multicellular communities of organisms (Cleary et al., 2004, 2005, 2008; Condit et al., 2002; Nekola & White, 1999; Tuomisto et al., 2003). The importance of space has been attributed to dispersal limitation, which may be intrinsic and depend on dispersal ability or due to the habitat configuration. Microorganisms, which strictly depend on host sponges for their survival, for example, may vary in their ability to disperse depending on host density. This will probably be less important in the cases where the microorganisms are horizontally derived from the 'rare biosphere' (Pascoal et al., 2021).

In addition to this, other factors such as mass and founder (initial colonisers) effects may influence the spatial structure of host-associated microbial communities (Guélat et al., 2008; Leibold et al., 2004; Maas et al., 2018). The importance of space in structuring microbial communities may, thus, be related to a range of disparate processes including intrinsic limitations of the organisms in question and processes collectively known as historical processes (Hanson et al., 2012).

The variation explained by host dissimilitude alone varied from 10.46% for log-transformed data to 30.48% for untransformed data. In barrel sponges (*Xestospongia* spp.), this component explained 3% of the variation in prokaryotic composition using log-transformed data (Cleary et al., 2022). There did, however, appear to be scale dependence in *Xestospongia* spp. where host dissimilitude explained >24% of variation at a single location (Cleary et al., 2022). Our results suggest an interplay between space and host dissimilitude with

much lower mean dissimilarities at smaller scales, but with considerable variation. Taken together, space and host dissimilitude explained >50% of the variation in the prokaryotic communities of golf ball sponges.

The relative importance of space versus host dissimilitude, however, shifted depending on whether the response matrix of prokaryotic dissimilarity was untransformed or log-transformed. Host dissimilitude was a more important explanatory factor for untransformed data. This suggests that host dissimilitude plays a more important role for more abundant prokaryotic taxa, whereas space is more important for less abundant taxa. We attribute this to the tendency of closely related host taxa to share abundant symbionts across large spatial scales. Distantly sampled specimens, in contrast, housed less abundant prokaryotic taxa reflective of their distant environments, thus increasing the importance of spatial distance when using log-transformed data, which scales down the effect of abundant taxa. This effect is highlighted by the spatially structured environmental component, which was more important for log-transformed (13.67%) than untransformed data (4.09%).

Previous studies have identified genetic dissimilarity among hosts (host dissimilitude) as a significant determinant of microbiome beta diversity across a wide range of host taxa including mice, humans, black soldier flies, plants, oysters, and sponges (Bonder et al., 2016; Easson et al., 2020; Greenwood et al., 2021; Griffiths et al., 2019; Wegner et al., 2013; Zhang et al., 2023). Sponge studies have also shown that host identity is one of the most important determinants of microbiome variation (Cleary et al., 2019, 2020, 2021; Thomas et al., 2016). Thomas et al. (2016), however, noted that sponge phylogeny only explained a minor amount of the variation in dissimilarity of sponge-associated prokaryotic communities.

In contrast to space and host dissimilitude, the purely environmental component only explained a very small amount of variation in composition, 0.3%. Almost 40%–50% of the variation in dissimilarity also remained unexplained. The unexplained variation may be attributable to stochastic processes, historical factors such as past perturbations or unmeasured environmental variables. Potentially important environmental factors that should be considered in future studies include pH, nutrient concentrations, sediment load and the concentration and composition of dissolved organic matter in addition to others. Future research should attempt to ascertain to what extent these and other environmental parameters may be important in structuring prokaryotic community composition in golf ball sponges.

In summary, the present study points to an important role of space and host dissimilitude in structuring prokaryotic communities of golf ball sponges. Our study, however, has considerable limitations due to the small sample size. That being said, the results largely align with a previous study of the sponge species complex *Xestospongia* spp. sampled over a similarly large spatial scale encompassing IndoPacific and Caribbean sites (Cleary et al., 2022). Future studies of other sponge species complexes should help to shed greater light on the importance of space, host dissimilitude and environment in structuring prokaryotic communities.

5 | CONCLUSION

In the present study, we identified putative host sponge taxa using a phylogenetic analysis of the 28S rRNA gene and dominant prokaryotic taxa associated with these host sponges. The most abundant OTUs were classified to a range of taxa including the AqS1 and EC94 groups, which have been shown to carry genes that are known to promote host–microbe interactions. In addition to the above, certain sponge hosts were enriched in OTUs classified to the SAR202 clade of Chloroflexi. We, furthermore, showed that prokaryotic dissimilarity was significantly structured by space and host dissimilitude.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest.

DATA AVAILABILITY STATEMENT

The DNA sequences generated in this study can be downloaded from NCBI BioProject Ids: PRJNA382576, PRJNA479568, PRJNA715755, PRJNA715750, and PRJNA715763.

ORCID

Daniel F. R. Cleary https://orcid.org/0000-0002-6143-3390
Nicole J. de Voogd https://orcid.org/0000-0002-7985-5604

REFERENCES

Baas Becking, L. G. M. (1934). *Geobiologie of inleiding tot de milieukunde, den Haag*. W.P. van Stockum & Zoon.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, 37, 852–857. https://doi.org/10.1038/s41587-019-0209-9

Bonder, M. J., Kurilshikov, A., Tigchelaar, E. F., Mujagic, Z., Imhann, F., Vila, A. V., Deelen, P., Vatanen, T., Schirmer, M., Smeekens, S. P., Zhernakova, D. V., Jankipersadsing, S. A., Jaeger, M., Oosting, M., Cenit, M. C., Masclee, A. A., Swertz, M. A., Li, Y., Kumar, V., … Zhernakova, A. (2016). The effect of host genetics on the gut microbiome. *Nature Genetics*, 48(11), 1407–1412. https://doi.org/10.1038/ng.3663

Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, 73, 1045–1055.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581–583. https://doi.org/10.1038/nmeth.3869

Carella, M., Agell, G., Cárdenas, P., & Uriz, M. J. (2016). Phylogenetic reassessment of Antarctic Tetillidae (Demospongiae, Tetractinellida) reveals new genera and genetic similarity among morphologically distinct species. *PLoS One*, 11(8), e0160718. https://doi.org/10.1371/journal.pone.0160718

Chen, L., Liu, S., Chen, Q., Zhu, G., Wu, X., Wang, J., Li, X., Hou, L., & Ni, J. (2020). Dispersal limitation drives biogeographical patterns of anammox bacterial communities across the Yangtze River. *Applied Microbiology and Biotechnology*, 104(5), 535–5546. https://doi.org/10.1007/s00253-020-10511-4

Clark, D. R., Underwood, G. J., McGenity, T. J., & Dumbrell, A. J. (2021). What drives study dependent differences in distance–decay relationships of microbial communities? *Global Ecology and Biogeography*, 30(4), 811–825. https://doi.org/10.1111/geb.13266

Cleary, D. F. R., Becking, L. E., de Voogd, N. J., Renema, W., de Beer, M., van Soest, R. W., & Hoeksema, B. W. (2005). Variation in the diversity and composition of benthic taxa as a function of distance offshore, depth and exposure in the Spermonde Archipelago, Indonesia. *Estuarine, Coastal and Shelf Science*, 65(3), 557–570.

Cleary, D. F. R., Becking, L. E., Pires, A. C. C., de Voogd, N. J., Egas, C., & Gomes, N. C. M. (2013). Habitat and host related variation in sponge bacterial communities in Indonesian coral reefs and marine lakes. *FEMS Microbiology Ecology*, 85, 465–482.

Cleary, D. F. R., DeVantier, L., Giyanto, Vail, L., Manto, P., de Voogd, N. J., Rachello-Dolmen, P. G., Tuti, Y., Budiyanto, A., Wolstenholme, J., Hoeksema, B. W., & Suharsono. (2008). Relating variation in species composition to environmental variables: A multi-taxon study in an Indonesian coral reef complex. *Aquatic Sciences*, 70, 419–431.

Cleary, D. F. R., de Voogd, N. J., Polónia, A. R. M., Freitas, R., & Gomes, N. C. M. (2015). Composition and predictive functional analysis of bacterial communities in seawater, sediment and sponges in an Indonesian coral reef environment. *Microbial Ecology*, 70, 889–903. https://doi.org/10.1007/s00248-015-0632-5

Cleary, D. F. R., de Voogd, N. J., Stuij, T., Swierts, T., Oliveira, V., Polónia, A. R. M., Louvado, A., Gomes, N. C. M., & Coelho, F. J. R. C. (2023). A study of sponge symbionts from different light habitats. *Microbial Ecology*, 86, 2819–2837. https://doi.org/10.1007/s00248-023-02267-x

Cleary, D. F. R., Mooers, A. Ø., Eichhorn, K. A. O., van Tol, J., de Jong, R., & Menken, S. B. J. (2004). Diversity and community composition of butterflies and odonates in an ENSO-induced fire affected habitat mosaic: A case study from East Kalimantan, Indonesia. *Oikos*, 105, 426–446.

Cleary, D. F. R., & Gomes, N. C. M. (2019). Bacterial, archaeal and microeukaryote plankton communities inside and outside of Indonesian marine lakes. *Journal of Sea Research*, 148–149, 23–32. https://doi.org/10.1016/j.seares.2019.04.001

Cleary, D. F. R., Oliveira, V., Gomes, N. C. M., Bialecki, A., & de Voogd, N. J. (2023). A comparison of free-living and sponge-associated prokaryotic communities with a focus on calcareous sponges. *FEMS Microbiology Ecology*, 99(3), fiad014. https://doi.org/10.1093/femsec/fiad014

Cleary, D. F. R., Polónia, A. R. M., & de Voogd, N. J. (2018a). Bacterial communities inhabiting the sponge *Biemna fortis*, sediment and water in marine lakes and the open sea. *Microbial Ecology*, 76, 610–624. https://doi.org/10.1007/s00248-018-1156-6

Cleary, D. F. R., Polónia, A. R. M., & de Voogd, N. J. (2018b). A comparison of prokaryote community composition in two *Cinachyrella* morphospecies and water from west Papuan marine lakes. *FEMS Microbiology Ecology*, 94(2), fix175. https://doi.org/10.1093/femsec/fix175

Cleary, D. F. R., Polónia, A. R. M., & de Voogd, N. J. (2021). Composition and diversity of prokaryotic communities sampled from sponges and soft corals in Maldivian waters. *Marine Ecology*, 42, e12638. https://doi.org/10.1111/maec.12638

Cleary, D. F. R., Polónia, A. R. M., Reijnen, B. T., Berumen, M. L., & de Voogd, N. J. (2020). Prokaryote communities inhabiting endemic and newly discovered sponges and octocorals from the Red Sea. *Microbial Ecology*, 80, 103–119. https://doi.org/10.1007/s00248-019-01465-w

Cleary, D. F. R., Polónia, A. R. M., Sousa, A., Lillebø, A. I., Queiroga, H., & Gomes, N. C. M. (2016). A comparison of the rhisosphere-associated bacterial communities of two saltmarsh plant species, *Juncus maritimus* and *Spartina maritima*. *Plant Biology*, 18, 824–834. https://doi.org/10.1111/plb.12459

Cleary, D. F. R., Polónia, A. R. M., Swierts, T., Coelho, F. J. R., de Voogd, N. J., & Gomes, N. C. M. (2022). Spatial and environmental variables structure sponge symbiont communities. *Molecular Ecology*, 31, 4932–4948. https://doi.org/10.1111/mec.16631

Cleary, D. F. R., Swierts, T., Coelho, F. J. R. C., Polónia, A. R. M., Huang, Y. M., Ferreira, M. R. S., Putchakarn, S., Carvalheiro, L., van der Ent, E., Ueng, J. P., Gomes, N. C. M., & de Voogd, N. J. (2019). The sponge microbiome within the greater coral reef microbial metacommunity. *Nature Communications*, 10, 1644. https://doi.org/10.1038/s41467-019-09537-8

Coelho, F. J. R. C., Cleary, D. F. R., Gomes, N. C. M., Pólonia, A. R. M., Huang, Y. M., Liu, L. L., & de Voogd, N. J. (2018). Sponge prokaryote communities in Taiwanese coral reef and shallow hydrothermal vent ecosystems. *Microbial Ecology*, 75, 239–254. https://doi.org/10.1007/s00248-017-1023-x

Colatriano, D., Tran, P. Q., Guéguen, C., Williams, W. J., Lovejoy, C., & Walsh, D. A. (2018). Genomic evidence for the degradation of terrestrial organic matter by pelagic Arctic Ocean Chloroflexi bacteria. *Communications Biology*, 1, 90. https://doi.org/10.1038/s42003-018-0086-7

Condit, R., Pitman, N., Leigh, E. G., Chave, J., Terborgh, J., Foster, R. B., Núñez, P., Aguilar, S., Valencia, R., Villa, G., Muller-Landau, H. C., Losos, E., & Hubbell, S. P. (2002). Beta-diversity in tropical forest trees. *Science*, 295, 666–669. https://doi.org/10.1126/science.1066854

de Voogd, N. J., Alvarez, B., Boury-Esnault, N., Carballo, J. L., Cárdenas, P., Díaz, M.-C., Dohrmann, M., Downey, R., Hajdu, E., Hooper, J. N. A., Kelly, M., Klautau, M., Manconi, R., Morrow, C. C., Pisera, A. B., Ríos, P., Rützler, K., Schönberg, C., Vacelet, J., & van Soest, R. W. M. (2022). *World porifera database*. https://www.marinespecies.org/porifera

de Voogd, N. J., & Cleary, D. F. R. (2009). Variation in sponge composition among Singapore reefs. *Raffles Bulletin of Zoology*, 22(Suppl), 59–67.

de Voogd, N. J., Gauvin-Bialecki, A., Polónia, A. R. M., & Cleary, D. F. R. (2019). Assessing the bacterial communities of sponges inhabiting the remote western Indian Ocean Island of Mayotte. *Marine Ecology*, 39(6), e12517. https://doi.org/10.1111/maec.12517

de Voogd, N. J., Becking, L. E., & Cleary, D. F. R. (2009). Sponge community composition in the Derawan Islands, NE Kalimantan, Indonesia. *Marine Ecology Progress Series*, 396, 219–230. https://doi.org/10.3354/meps08349

de Voogd, N. J., & Cleary, D. F. R. (2008). An analysis of sponge diversity and distribution at three taxonomic levels in the Thousand Islands/Jakarta Bay reef complex, West-Java, Indonesia. *Marine Ecology*, 29, 205–215.

Díez-Vives, C., Taboada, S., Leiva, C., Busch, K., Hentschel, U., & Riesgo, A. (2020). On the way to specificity – microbiome reflects sponge genetic cluster primarily in highly structured populations. *Molecular Ecology*, 29(22), 4412–4427. https://doi.org/10.1111/mec.15635

Dogs, M., Wemheuer, B., Wolter, L., Bergen, N., Daniel, R., Simon, M., & Brinkhoff, T. (2017). Rhodobacteraceae on the marine brown alga *Fucus spiralis* are abundant and show physiological adaptation to an epiphytic lifestyle. *Systematic and Applied Microbiology*, 40(6), 370–382. https://doi.org/10.1016/j.syapm.2017.05.006

Duivenvoorden, J. F., Svenning, J.-C., & Wright, S. J. (2002). Beta diversity in tropical forests. *Science*, 295, 636–637. https://doi.org/10.1126/science.295.5555.636

Easson, C. G., Chaves-Fonnegra, A., Thacker, R. W., & Lopez, J. V. (2020). Host population genetics and biogeography structure the microbiome of the sponge *Cliona delitrix*. *Ecology and Evolution*, 10(4), 2007–2020. https://doi.org/10.1002/ece3.6033

Elifantz, H., Horn, G., Ayon, M., Cohen, Y., & Minz, D. (2013). Rhodobacteraceae are the key members of the microbial community of the initial biofilm formed in eastern Mediterranean coastal seawater. *FEMS Microbiology Ecology*, 85(2), 348–357. https://doi.org/10.1111/1574-6941.12122

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, 39, 783–791.

Ferreira, M. R. S., & Cleary, D. F. R. (2022). Microeukaryotic communities of golf ball sponges inside and outside of marine lakes. *Journal of Sea Research*, 180, 102151. https://doi.org/10.1016/j.seares.2021.102151

Field, J. G., Clarke, K. R., & Warwick, R. M. (1982). A practical strategy for analysing multispecies distribution patterns. *Marine Ecology Progress Series*, 8, 37–52.

Gauthier, M.-E. A., Watson, J. R., & Degnan, S. M. (2016). Draft genomes shed light on the dual bacterial symbiosis that dominates the microbiome of the coral reef sponge *Amphimedon queenslandica*. *Frontiers in Marine Science*, 3, 196. https://doi.org/10.3389/fmars.2016.00196

Green, J. L., Holmes, A. J., Westoby, M., Oliver, I., Briscoe, D., Dangerfield, M., Gillings, M., & Beattie, A. J. (2004). Spatial scaling of microbial eukaryote diversity. *Nature*, 432(7018), 747–750. https://doi.org/10.1038/nature03034

Greenwood, M. P., Hull, K. L., Brink-Hull, M., Lloyd, M., & Rhode, C. (2021). Feed and host genetics drive microbiome diversity with resultant consequences for production traits in mass-reared black soldier Fly (*Hermetia illucens*) larvae. *Insects*, 12(12), 1082. https://doi.org/10.3390/insects12121082

Griffiths, S. M., Antwis, R. E., Lenzi, L., Lucaci, A., Behringer, D. C., Butler, M. J., & Preziosi, R. F. (2019). Host genetics and geography influence microbiome composition in the sponge *Ircinia campana*. *Journal of Animal Ecology*, 88(11), 1684–1695. https://doi.org/10.1111/1365-2656.13065

Guélat, J., Jaquiéry, J., Berset-Brändli, L., Pellegrini, E., Moresi, R., Broquet, T., Hirzel, A. H., & Perrin, N. (2008). Mass effects mediate coexistence in competing shrews. *Ecology*, 89(7), 2033–2042. https://doi.org/10.1890/07-0905.1

Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C., & Martiny, J. B. (2012). Beyond biogeographic patterns: Processes shaping the microbial landscape. *Nature Reviews Microbiology*, 10(7), 497–506. https://doi.org/10.1038/nrmicro2795

Hou, J., Wu, L., Liu, W., Ge, Y., Mu, T., Zhou, T., Li, Z., Zhou, J., Sun, X., Luo, Y., & Christie, P. (2020). Biogeography and diversity patterns of abundant and rare bacterial communities in rice paddy soils across China. *Science of the Total Environment*, 730, 139116. https://doi.org/10.1016/j.scitotenv.2020.139116

Ihua, M. W., FitzGerald, J. A., Guihéneuf, F., Jackson, S. A., Claesson, M. J., Stengel, D. B., & Dobson, A. D. W. (2020). Diversity of bacteria populations associated with different thallus regions of the brown alga *Laminaria digitata*. *PLoS One*, 15(11), e0242675. https://doi.org/10.1371/journal.pone.0242675

Jones, M. M., Tuomisto, H., Clark, D. B., & Olivas, P. (2006). Effects of mesoscale environmental heterogeneity and dispersal limitation on floristic variation rain forest ferns. *Journal of Ecology*, 94, 181–195.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glöckner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, 41(1), e1. https://doi.org/10.1093/nar/gks808

Landry, Z., Swan, B. K., Herndl, G. J., Stepanauskas, R., & Giovannoni, S. J. (2017). SAR202 genomes from the dark ocean predict pathways for the oxidation of recalcitrant dissolved organic matter. *MBio*, 8, e00413-17. https://doi.org/10.1128/mBio.00413-17

Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., Holt, R. D., Shurin, J. B., Law, R., Tilman, D., & Loreau, M. (2004). The metacommunity concept: A framework for multi-scale community ecology. *Ecology Letters*, 7(7), 601–613.

Lichstein, J. W. (2007). Multiple regression on distance matrices: A multivariate spatial analysis tool. *Plant Ecology*, 188, 117–131.

Locey, K. J., Muscarella, M. E., Larsen, M. L., Bray, S. R., Jones, S. E., & Lennon, J. T. (2020). Dormancy dampens the microbial distance-decay relationship. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 375(1798), 20190243. https://doi.org/10.1098/rstb.2019.0243

Maas, D. L., Prost, S., Bi, K., Smith, L. L., Armstrong, E. E., Aji, L. P., Toha, A. H. A., Gillespie, R. G., & Becking, L. E. (2018). Rapid divergence of mussel populations despite incomplete barriers to dispersal. *Molecular Ecology*, 27(7), 1556–1571. https://doi.org/10.1111/mec.14556

Martiny, J. B., Bohannan, B. J., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., Horner-Devine, M. C., Kane, M., Krumins, J. A., Kuske, C. R., Morin, P. J., Naeem, S., Ovreås, L., Reysenbach, A. L., Smith, V. H., & Staley, J. T. (2006). Microbial biogeography: Putting microorganisms on the map. *Nature Reviews Microbiology*, 4(2), 102–112. https://doi.org/10.1038/nrmicro1341

Martiny, J. B., Eisen, J. A., Penn, K., Allison, S. D., & Horner-Devine, M. C. (2011). Drivers of bacterial beta-diversity depend on spatial scale. *Proceedings of the National Academy of Sciences of the United States of America*, 108(19), 7850–7854. https://doi.org/10.1073/pnas.1016308108

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, 8(4), e61217. https://doi.org/10.1371/journal.pone.0061217

Nekola, J. C., & White, P. S. (1999). The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, 26(4), 867–878.

Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P. M., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2020). *Vegan: Community ecology package*. R Package Version 2.5-7. https://CRAN.R-project.org/package=vegan

Orsi, W. D., Smith, J. M., Liu, S., Liu, Z., Sakamoto, C. M., Wilken, S., Poirier, C., Richards, T. A., Keeling, P. J., Worden, A. Z., & Santoro, A. E. (2016). Diverse, uncultivated bacteria and archaea underlying the cycling of dissolved protein in the ocean. *The ISME Journal*, 10(9), 2158–2173. https://doi.org/10.1038/ismej.2016.20

Papke, R. T., Ramsing, N. B., Bateson, M. M., & Ward, D. M. (2003). Geographical isolation in hot spring cyanobacteria. *Environmental Microbiology*, 5(8), 650–659. https://doi.org/10.1046/j.1462-2920.2003.00460.x

Pascoal, F., Costa, R., & Magalhães, C. (2021). The microbial rare biosphere: Current concepts, methods and ecological principles. *FEMS Microbiology Ecology*, 97(1), fiaa227. https://doi.org/10.1093/femsec/fiaa227

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ribera, I., Dolédec, S., Downie, I., & Foster, G. N. (2001). Effect of land disturbance and stress on species traits of ground beetle assemblages. *Ecology*, 82, 1112–1129.

Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in R. *Bioinformatics*, 27(4), 592–593.

Simon, M., Scheuner, C., Meier-Kolthoff, J. P., Brinkhoff, T., Wagner-Döbler, I., Ulbrich, M., Klenk, H. P., Schomburg, D., Petersen, J., & Göker, M. (2017). Phylogenomics of Rhodobacteraceae reveals evolutionary adaptation to marine and non-marine habitats. *The*

ISME Journal, *11*(6), 1483–1499. https://doi.org/10.1038/ismej.2016.198

Swierts, T., Cleary, D. F. R., & de Voogd, N. J. (2018). Biogeography of prokaryote communities in closely related giant barrel sponges across the Indo-Pacific. *FEMS Microbiology Ecology*, *94*(12), fiy194. https://doi.org/10.1093/femsec/fiy194

Szitenberg, A., Rot, C., Ilan, M., & Huchon, D. (2010). Diversity of sponge mitochondrial introns revealed by cox 1 sequences of Tetillidae. *Evolutionary Biology*, *10*, 288.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, *28*, 2731–2739. https://doi.org/10.1093/molbev/msr121

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of, DNA sequences. In R. M. Miura (Ed.), *Lectures on mathematics in the life sciences* (pp. 57–86). American Mathematical Society.

Taylor, J. A., Palladino, G., Wemheuer, B., Steinert, G., Sipkema, D., Williams, T. J., & Thomas, T. (2021). Phylogeny resolved, metabolism revealed: Functional radiation within a widespread and divergent clade of sponge symbionts. *The ISME Journal*, *15*(2), 503–519. https://doi.org/10.1038/s41396-020-00791-z

Thomas, T., Moitinho-Silva, L., Lurgi, M., Björk, J. R., Easson, C., Astudillo-García, C., Olson, J. B., Erwin, P. M., López-Legentil, S., Luter, H., Chaves-Fonnegra, A., Costa, R., Schupp, P. J., Steindler, L., Erpenbeck, D., Gilbert, J., Knight, R., Ackermann, G., Victor Lopez, J., … Webster, N. S. (2016). Diversity, structure and convergent evolution of the global sponge microbiome. *Nature Communications*, *7*, 11870. https://doi.org/10.1038/ncomms11870

Thompson, J. R., Rivera, H. E., Closek, C. J., & Medina, M. (2015). Microbes in the coral holobiont: Partners through evolution, development, and ecological interactions. *Frontiers in Cellular and Infection Microbiology*, *4*, 176. https://doi.org/10.3389/fcimb.2014.00176

Torgo, L. (2016). *Data mining with R: learning with case studies*. Chapman and Hall/CRC. http://www.dcc.fc.up.pt/~ltorgo/DataMiningWithR

Tuomisto, H., Ruokolainen, K., & Yli-Halla, M. (2003). Dispersal, environment, and floristic variation of western Amazonian forests. *Science*, *299*(5604), 241–244. https://doi.org/10.1126/science.1078037

van Tongeren, O. F. R. (1995). Cluster analysis. In R. H. G. Jongman, C. J. F. Ter Braak, & O. F. R. van Tongeren (Eds.), *Data analysis in community and landscape ecology* (pp. 174–212). Cambridge University Press.

Verhoeven, J. T. P., Kavanagh, A. N., & Dufour, S. C. (2016). Microbiome analysis shows enrichment for specific bacteria in separate anatomical regions of the deep-sea carnivorous sponge *Chondrocladia grandis*. *FEMS Microbiology Ecology*, *93*, fiw214. https://doi.org/10.1093/femsec/fiw214

Wang, Z. B., Sun, Y. Y., Li, Y., Chen, X. L., Wang, P., Ding, H. T., Chen, B., Zhang, X. Y., Song, X. Y., Wang, M., McMinn, A., Zhang, Y. Z., & Qin, Q. L. (2020). Significant bacterial distance-decay relationship in continuous, well-connected Southern Ocean surface water. *Microbial Ecology*, *80*(1), 73–80. https://doi.org/10.1007/s00248-019-01472-x

Wegner, K. M., Volkenborn, N., Peter, H., & Eiler, A. (2013). Disturbance induced decoupling between host genetics and composition of the associated microbiome. *BMC Microbiology*, *13*, 252. https://doi.org/10.1186/1471-2180-13-252

Whitaker, R. J., Grogan, D. W., & Taylor, J. W. (2003). Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science*, *301*(5635), 976–978.

Zhang, J., Liu, W., Bu, J., Lin, Y., & Bai, Y. (2023). Host genetics regulate the plant microbiome. *Current Opinion in Microbiology*, *72*, 102268. https://doi.org/10.1016/j.mib.2023.102268

Zhu, R., Liu, C., Xu, Y. D., He, W., Liu, J., Chen, J., An, Y., & Shi, S. (2022). Ratio of carbon and nitrogen in fertilizer treatment drives distinct rhizosphere microbial community composition and co-occurrence networks. *Frontiers in Microbiology*, *13*, 968551. https://doi.org/10.3389/fmicb.2022.968551

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.