ⵣISME

ARTICLE

Phylogeny resolved, metabolism revealed: functional radiation within a widespread and divergent clade of sponge symbionts

Jessica A. Taylor[1,2] · Giorgia Palladino [1,3,4] · Bernd Wemheuer[1,5] · Georg Steinert [3] · Detmer Sipkema[3] · Timothy J. Williams [2] · Torsten Thomas [1,5]

Abstract

The symbiosis between bacteria and sponges has arguably the longest evolutionary history for any extant metazoan lineage, yet little is known about bacterial evolution or adaptation in this process. An example of often dominant and widespread bacterial symbionts of sponges is a clade of uncultured and uncharacterised Proteobacteria. Here we set out to characterise this group using metagenomics, in-depth phylogenetic analyses, metatranscriptomics, and fluorescence in situ hybridisation microscopy. We obtained five metagenome-assembled-genomes (MAGs) from different sponge species that, together with a previously published MAG (AqS2), comprise two families within a new gammaproteobacterial order that we named Ca. Tethybacterales. Members of this order share a heterotrophic lifestyle but vary in their predicted ability to use various carbon, nitrogen and sulfur sources, including taurine, spermidine and dimethylsulfoniopropionate. The deep branching of the Ca. Tethybacterales within the Gammaproteobacteria and their almost exclusive presence in sponges suggests they have entered a symbiosis with their host relatively early in evolutionary time and have subsequently functionally radiated. This is reflected in quite distinct lifestyles of various species of Ca. Tethybacterales, most notably their diverse morphologies, predicted substrate preferences, and localisation within the sponge tissue. This study provides new insight into the evolution of metazoan–bacteria symbiosis.

Introduction

Marine sponges (phylum Porifera) are diverse metazoans that arose in the Precambrian over 700 million years ago and are now globally distributed throughout marine and freshwater ecosystems in temperate, tropical and polar regions [1–3]. They possess a simple body plan divided into three layers of organisation: the outer pinacoderm, the inner choanoderm layer, and the mesohyl matrix in between [3]. As sessile filter-feeding organisms they are remarkably efficient in obtaining food from the surrounding water and thus draw down carbon, nitrogen and silicon from the pelagic environment into benthic biomass [4–6].

Marine sponges are well known for their wide range of interactions with microorganisms from all domains of life, from using them as food sources to housing them as symbionts [7]. Sponges are capable of differentiating between symbionts and food microorganisms [8], potentially through the recognition of specific eukaryote-like proteins (ELPs) presented by microbial symbionts [9–12] and/or immune receptors expressed by the sponge [13]. Sponges are often classified as either being of high microbial abundance (HMA), harbouring dense and often diverse microbial communities, or as being of low microbial abundance (LMA) [14, 15]. LMA sponges are typically dominated by only a few bacterial phyla, often being Proteobacteria [16]. Many bacteria found in sponges also fall into unique sponge-specific or sponge-enriched phylogenetic clusters [17, 18] as defined by being rarely found outside the sponge

✉ Torsten Thomas
   t.thomas@unsw.edu.au

1   Centre for Marine Science and Innovation, University of New South Wales, Sydney, NSW, Australia

2   School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia

3   Laboratory of Microbiology, Wageningen University & Research, Wageningen, The Netherlands

4   Department of Pharmacy and Biotechnology, Alma Mater Studiorum-University of Bologna, Bologna, Italy

5   School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW, Australia

environment [19]. These unique clades have been proposed to have likely undergone specific evolutionary adaptations that allow them to only (or preferentially) thrive in association with their sponge host [7].

One of these unique bacterial groups, which was initially assigned to the class Betaproteobacteria (now order Burkholderiales), contains dominant and persistent members of the microbiota in various species of LMA sponges, in some instances comprising over 70% of the sponge's microbial community [20–24]. First described in 2006 from the sponge *Tethya aurantium* [25], a few members of this bacterial group have been visualised via microscopy [20, 26, 27] and determined to be vertically transmitted from the adult sponge to the embryo [26, 28, 29]. It has also been proposed that this bacterial group co-evolved with their host [30] as each sponge species appears to harbour a distinct phylotype [20]. Despite the substantial interest in this group and its apparent broad involvement in sponge symbiosis, its phylogeny and taxonomy remain in disarray. Early studies placed this group in the family *Nitrosomonadaceae* (class Betaproteobacteria) based on the SILVA database 108 release [21] and assigned them to a sponge-specific cluster (SC112), which was related to other known betaproteobacterial clusters (SC110 and SC111) [22]. This classification was then superseded by placement in the betaproteobacterial order EC94 [31, 32], although in many studies these bacteria are simply described as unclassified betaproteobacteria [29, 33]. Furthermore, very little is known about this group's functional features or role in sponge-microbe interactions. One metagenome-assembled-genome (MAG) of this group, called AqS2, has recently been described from the Great Barrier Reef sponge *Amphimedon queenslandica* and showed a likely heterotrophic metabolism [33].

The aim of this study is to characterise this widespread and often dominant group of sponge-associated bacteria and define evolutionary and functional properties of symbiosis. For this we employed metagenomics, in-depth phylogenetic analyses, metatranscriptomics, and fluorescence in situ hybridisation microscopy across several sponge species (*Cymbastela concentrica*, *Crella incrustans*, *Crambe crambe*, *Scopalina* sp. and *Tethya stolonifera*) that are known to harbour this bacterial group.

Methods

Sampling, metagenomic sequencing, genome reconstruction and annotation

Sponges were collected via SCUBA or snorkel from Bare Island, NSW, Australia; the Mediterranean Coast of Spain; and from the north-east coast of New Zealand (Supplementary Table S1). Sponge-associated microorganisms were enriched in each sample following the cell separation protocol described in Thomas et al. [9]. High-molecular-weight DNA was extracted from microbial cell pellets using the most appropriate method based on sample type and sequenced with an array of next generation sequencing technologies (Table 1). Reads from each sample were quality trimmed, de novo assembled and contigs over 1000 bp binned into MAGs (Table 2). CheckM [34] was employed to estimate genome completeness and contamination using the taxon-specific workflow. Further details are given in the Supplementary Information (File 2).

MAGs, including AqS2, were submitted to the Joint Genome Institute (JGI) Gold database [35] and annotated via the Integrated Microbial Genome (IMG) pipeline [36]. IMG annotations were manually refined by using the IMG "Find Genes" BLAST tool against National Centre for Biotechnology Information (NCBI) protein reference sequences. Hits close to 35% protein sequence similarity were further manually refined using the ExPASy BLAST tool (UniProtKB/Swiss-Prot databases only). An additional Pfam (v31.0) search was conducted with HMMER [37] to investigate the presence and diversity of ELPs in each MAG. Hits with bit scores less than 20 were disregarded and duplicate hits per predicted protein across multiple Pfam families were removed. Gene functions were also

Table 1 Summary of DNA extraction, sequencing and analysis tools used to obtain MAGs and 16S rRNA gene sequences.

Sponge	DNA extraction	Sequencing technology	Assembler[a]	Binning[a]	rRNA reconstruction[a]
Cymbastela concentrica	Phenol–chloroform	Illumina GA II	metaSPAdes	MetaBAT/MyCC + binning refiner	SortMeRNA + rnaSPAdes
Crella incrustans	MoBio Powersoil	Illumina HiSeq	metaSPAdes	MyCC	Reago + rnaSPAdes
Crambe crambe	CTAB, chloroform	Illumina HiSeq	metaSPAdes	Cocacola	SortMeRNA + rnaSPAdes
Scopalina sp.	MoBio Powersoil	Illumina MiSeq	IDBA-UD	MetaBAT/MyCC + binning refiner	Reago, SortMeRNA + rnaSPAdes
Tethya stolonifera	CTAB, chloroform	454 GS-FLX	GS Assembler	MetaBAT	Metaxa + GS Assembler

[a]References and details are given in the Supplementary Information File 2.

compared to publicly available isolate genomes, MAGs and single-cell amplified genomes in the IMG database (July 2020). Only genomes assigned by the IMG taxonomy to the class Gammaproteobacteria and betaproteobacteria (i.e. now Burkholderiales) that had ecosystem information containing the keywords "marine" as well as "oceanic" or "pelagic" were selected for comparison. This resulted in 141 genomes (Supplementary File 6) that represent relatives that are likely free living. Differences in the abundance of gene functions between the MAGs and free-living relatives were assessed by t tests with Benjamini–Hochberg correction for multiple testing.

Phylogenetic analysis

Ribosomal RNA genes were reconstructed from the metagenomic reads from each sponge species using the tools listed in Table 1. Briefly, rRNA reads were identified and then assembled to form partial- or full-length 16S rRNA genes or rRNA operons. Reconstructed genes were searched (BLASTn) against the MAGs to aid in identification. Reconstructed 16S rRNA gene sequences were also aligned with the SINA web aligner [38] and imported into the ARB software package [39] containing the SILVA 132 SSU Ref Nr99 database [40]. Alignments were further refined manually in ARB and sequences were trimmed before insertion by parsimony into the phylogenetic tree. This identified neighbouring sequences, which were then aligned with SINA to sequences over 850 bp from the sponge-specific clusters SC112, SC110 and SC111 [18] as well as sequences from type strain examples of various orders of the Gammaproteobacteria and Alphaproteobacteria (out-group). A maximum-likelihood tree was constructed from this alignment with RAxML v. 8 [41] with 1000 bootstrap resampling and visualised using iTOL [42].

The MAGs from *C. concentrica*, *C. incrustans*, *C. crambe*, *Scopalina* sp., and *T. stolonifera* were analysed with the GTDB-Tk v0.2.1 Classify Workflow [43] containing the Genome Taxonomy Database (GTDB) [44] with 123,110 bacterial genomes. A genome tree and associated taxonomies were inferred using a set of 120 ubiquitous single-copy proteins and relative-evolutionary distance (RED) values were calculated. Relatedness of the MAGs was also assessed by average amino acid identity (AAI) [45] using the "AAI: Average Amino acid Identity calculator" [46].

Distribution analysis

As the group of bacteria investigated here has been reported to occur globally in a number of sponge species, we used the extensive dataset recently produced for the Sponge Earth Microbiome Project (SEMP) [16, 47] to explore its geographic distribution. For this, we first trimmed the V4 region of the 16S rRNA gene from each MAG to match the 100 bp regions of the SEMP dataset (https://github.com/amnona/SpongeEMP). These trimmed sequences were then BLASTn searched against the sequences used to construct the 16S rRNA gene phylogenetic tree shown in Fig. 1. From this, a 93% similarity threshold was found to encompass only sequences of the target group and exclude those from other orders. The trimmed V4 regions were then BLASTn searched against the operational taxonomic units (OTUs) of the SEMP database and hits with over 93% identity were matched with the relative abundance and metadata (host taxonomy and location) from Moitinho-Silva et al. [16]. As we were examining the group as a whole, relative abundances (RA%) of matched sequences were summed within samples and reported as the mean, standard deviation and range (min–max). Samples were retained for further analysis if the RA of the group was greater than 0.05%. Data were then subjected to statistical analysis (F test and one-way ANOVA).

Gene expression of MAG

Tissue of *C. concentrica*, *C. incrustans*, *C. crambe*, and *Scopalina* sp. individuals were collected via SCUBA from their respective sampling sites and brought to the surface within 20 min after being cut from the sponge individuals (Supplementary Table S2) and as part of previous studies involving *Scopalina* sp. [48] and *C. crambe* [49]. After surfacing, the sponge samples (~5–10 cm^3) were immediately washed three times (<1 min) in 50 ml of 0.22 μm filtered seawater to remove planktonic microorganisms before being snap frozen in liquid nitrogen within 5 min after surfacing of the SCUBA diver. Frozen samples were transported to the laboratory. Total RNA extraction and bacterial mRNA enrichment followed the protocol established by Diez et al. [48] and Öztürk et al. [49] (see Supplementary Information File 1 for details). Expression of the coding sequences for each MAG was estimated with RSEM v. 1.3.0 [50] using default parameters and implementing the Bowtie aligner [51]. Transcripts per million (TPM) measures were used as estimates of expression [52, 53] and only genes with an average TPM values over 30 across replicates were considered. Expression of genes is reported as TPM_{av} for those with multiple replicates. TPM_{sum} is the combined expression of all genes in a pathway, while $X - Y$ TPM signifies a range of expression values when multiple copies of a gene are present.

Bacterial visualisation with catalyzed reporter deposition fluorescence in situ hybridisation (CARD-FISH)

Sponges were collected from the original sampling sites (Supplementary Table S1). *T. stolonifera* was collected

via snorkel at low tide in October 2018 and *C. concentrica*, *C. incrustans* and *Scopalina* sp. were collected via SCUBA in November 2018. Sponges were dissected and soaked in 4% paraformaldehyde for 4 h then washed in PBS and stored in 1:1 PBS and ethanol at $-20\,°C$. Processing of sponge tissue was undertaken as described by Webster and Hill [54] and is described in more detail in the Supplementary Information File 1. CARD-FISH probes labelled with horseradish peroxidase (HRP) were designed for each target bacterium using the probe design tool in the ARB software package [39]. Probe specificity was assessed using the ARB probe match tool, the SILVA probe test tool and BLASTn searches against the NCBI database. Probe sequences can be found in the Supplementary Information (Table S3). Photo bleaching of *C. incrustans'* autofluorescence was conducted by applying light from a mercury lamp onto tissue sections for 1 min prior to the CARD-FISH procedure [54]. CARD-FISH was carried out on sponge tissue sections as described by Croue et al. [20] with several modification (see Supplementary Information File 1 for details) and slides were visualised on a Nikon A1 Spectral confocal microscope. Conditions for probe specificity and stringency were assessed experimentally in an instance of two mismatches between a probe and non-target bacteria (*Scopalina* sp. probe on *C. concentrica* tissue). Each CARD-FISH experiment was conducted with a negative control slide that underwent the exact same procedure, but without adding the probe.

Results and discussion

For each sponge species analysed here, we obtained a single MAG each that was affiliated initially with the EC94 cluster (see below). The MAGs range in size from 1.1 to 2.2 Mbp, have completenesses of 66.66–95.69%, contamination rates of <1%, and an estimated genome size of 1.65–2.41 Mbp (based on the bacteria-specific markers of CheckM [34]). The observed GC contents of the genomes vary over a wide range from 50 to 68% (Table 2). Of the protein-coding genes, $22.2 \pm 8.6\%$ are hypothetical, i.e. without a predicted function (Table 2). With the exception of Cram cc1, the relative proportion of hypothetical proteins is in the lower range of reported values for MAGs from sponge symbionts [33, 55, 56], possibly because of the relatively small genome sizes and because the phylum Proteobacteria is comparatively well studied and annotated [57].

Phylogenetic analyses revealed a novel order within the Gammaproteobacteria

We recovered near full-length 16S rRNA genes from the metagenomes of *C. concentrica* (1494 bp), *C. incrustans* (1494 bp), *C. crambe* (1527 bp), *Scopalina* sp. (1499 bp), and a partial sequence from *T. stolonifera* (941 bp) (see Supplementary File 1, Table S4 for sequences). The AqS2 16S rRNA gene sequence (1519 bp) was already present in the SILVA 132 release (MEIE01000018). The 16S rRNA gene sequences from the symbionts of *C. concentrica* and

Table 2 Characteristics for MAGs affiliated with the EC94.

	AqS2	Cym b2r	Crel c29	Cram cc1	Sco b3	Teth b1
Assembly size (Mbp)	1.6	1.7	1.5	2.2	1.1	1.2
Coverage	100X	5.5X	33X	6.4X	5.2X	2.3X
Number of contigs	239	71	239	941	180	252
Longest contig	82499	160261	46186	15980	24339	14808
Shortest contig	1017	2651	2022	1000	2502	2512
N50	12238	44178	8774	2758	7022	5107
Average GC content (%)	68.83	50.32	52.96	59	56.2	64.77
Genome completeness (%)	90.52	95.69	89.5	91.38	66.66	67.9
Estimated genome size (Mbp)	1.77	1.78	1.68	2.41	1.65	1.77
Contamination (%)	0	0	0	0.16	0.16	0
Number of CDS	1614	1622	1400	2455	1134	1340
Predicted function (%)	77.14	83.23	84.64	61.75	85.89	80.97
Hypothetical (%)	22.86	16.77	15.36	38.25	16.43	23.50
5S rRNA genes	1	1	1	1	1	1
16S rRNA genes	1	1	1	1	1	1
23S rRNA genes	1	1	1	2	0	1
tRNA genes	42	43	34	43	31	28

MAGs are named after their source sponge: Cym b2r from *Cymbastela concentrica*, Crel c29 from *Crella incrustans*, Cram cc1 from *Crambe crambe*, Sco b3 from *Scopalina* sp. and Teth b1 from *Tethya stolonifera*. Data for AqS2 (Gauthier et al. [33]) have also been included.

C. crambe were entirely assembled in their respective MAGs, while a 240 bp region flanking the reconstructed 16S rRNA gene from *Scopalina* sp. overlapped with 97% identity with the end of a MAG contig and was thus retrospectively added to it. The 16S rRNA gene from the symbionts of *C. incrustans* and *T. stolonifera* did not overlap with any MAG contig, but we still manually included them based on comparative 16S rRNA and genome phylogenetic analysis (see below).

Insertion of these reconstructed 16S rRNA gene sequences by parsimony into the SILVA 132 SSU Ref Nr99 tree places them into the Betaproteobacteriales group EC94 [58], among sequences from other uncultured bacteria from marine samples. The 16S rRNA phylogeny (Fig. 1) provides a taxonomic
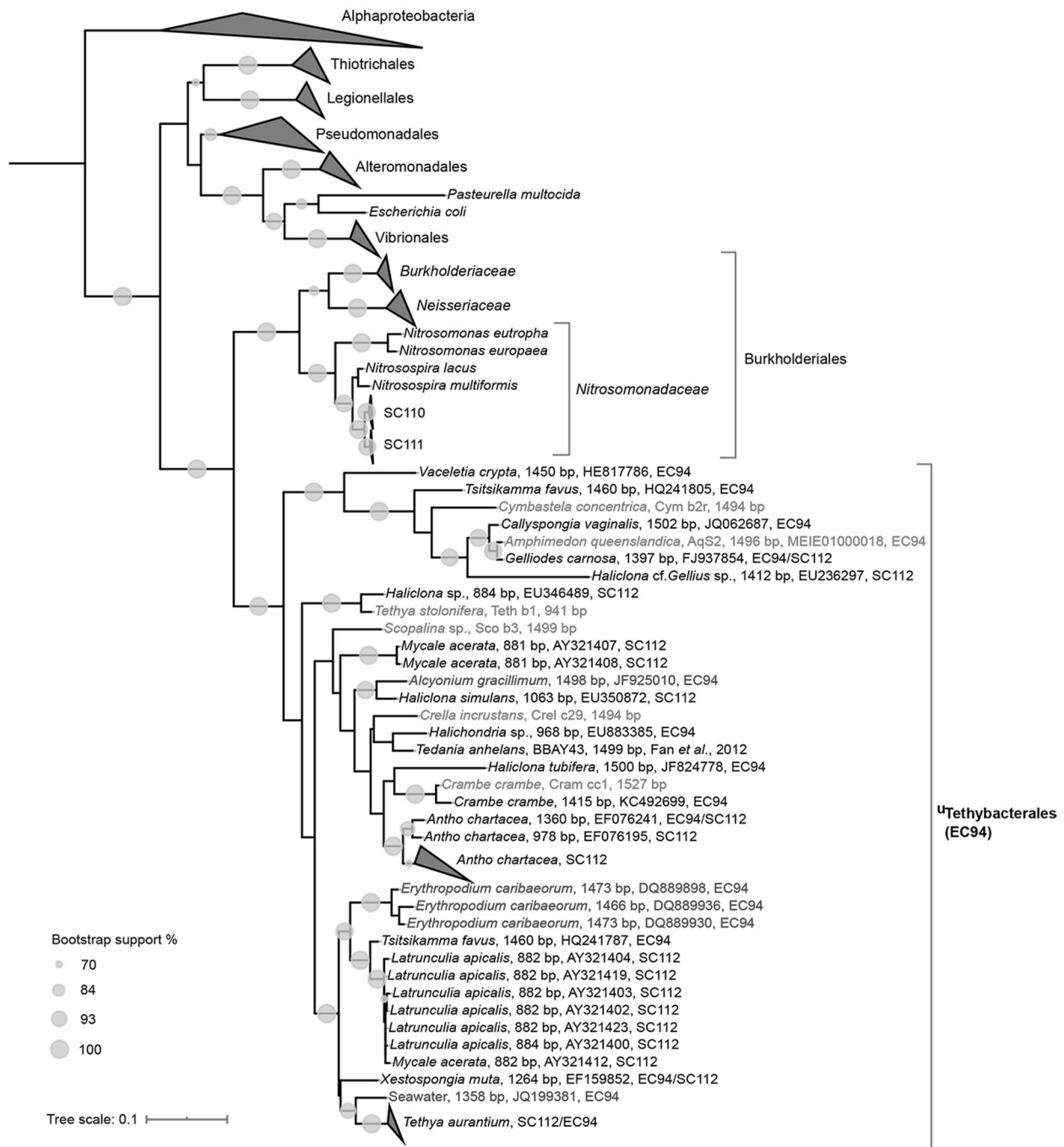

Fig. 1 16S rRNA phylogenetic tree of gammaproteobacterial orders with Alphaproteobacteria as an outgroup. The displayed tree is a maximum-likelihood tree constructed with sequences longer than 850 bp. Circles represent bootstrap support of 70–100% (1000 bootstraps). Bar represents 10% sequence divergence. Sequences are named as: host species or source, length bp, accession number, order/sponge-specific cluster. Names in red are reconstructed rRNA gene sequences from this study, names in blue represent EC94 members of non-sponge origin.

assignment for previously described sponge-specific cluster SC112 to within the group EC94. Contrary to the recent SILVA 132 release, EC94 appears to form a sister clade sharing a common ancestor with the recently re-ranked Betaproteobacteriales (now Burkholderiales), rather than a subgroup [44]. The reconstructed 16S rRNA gene sequences appear to form two families within this order based on previously proposed percentage similarity thresholds [59]. Specifically, the 16S rRNA gene from AqS2 and the symbiont of *C. concentrica* share 90% similarity (i.e. same family) and sequences from the symbionts of *C. incrustans*, *C. crambe*,

Scopalina sp., and *T. stolonifera* share between 89 and 93% similarity; however, the two families only share 83–85% sequence similarity. The order contains mostly 16S rRNA gene sequences derived from sponge samples, with the only exception being one sequence from a seawater sample and four sequences from two coral species (Fig. 1).

The genome tree based on the GTDB reflects a similar phylogenetic arrangement (Fig. 2). The MAGs found here are clustered within the AqS2 order, separate from the Burkholderiales. AqS2 is the only genome representing EC94 in the GTDB. The accompanying classification places
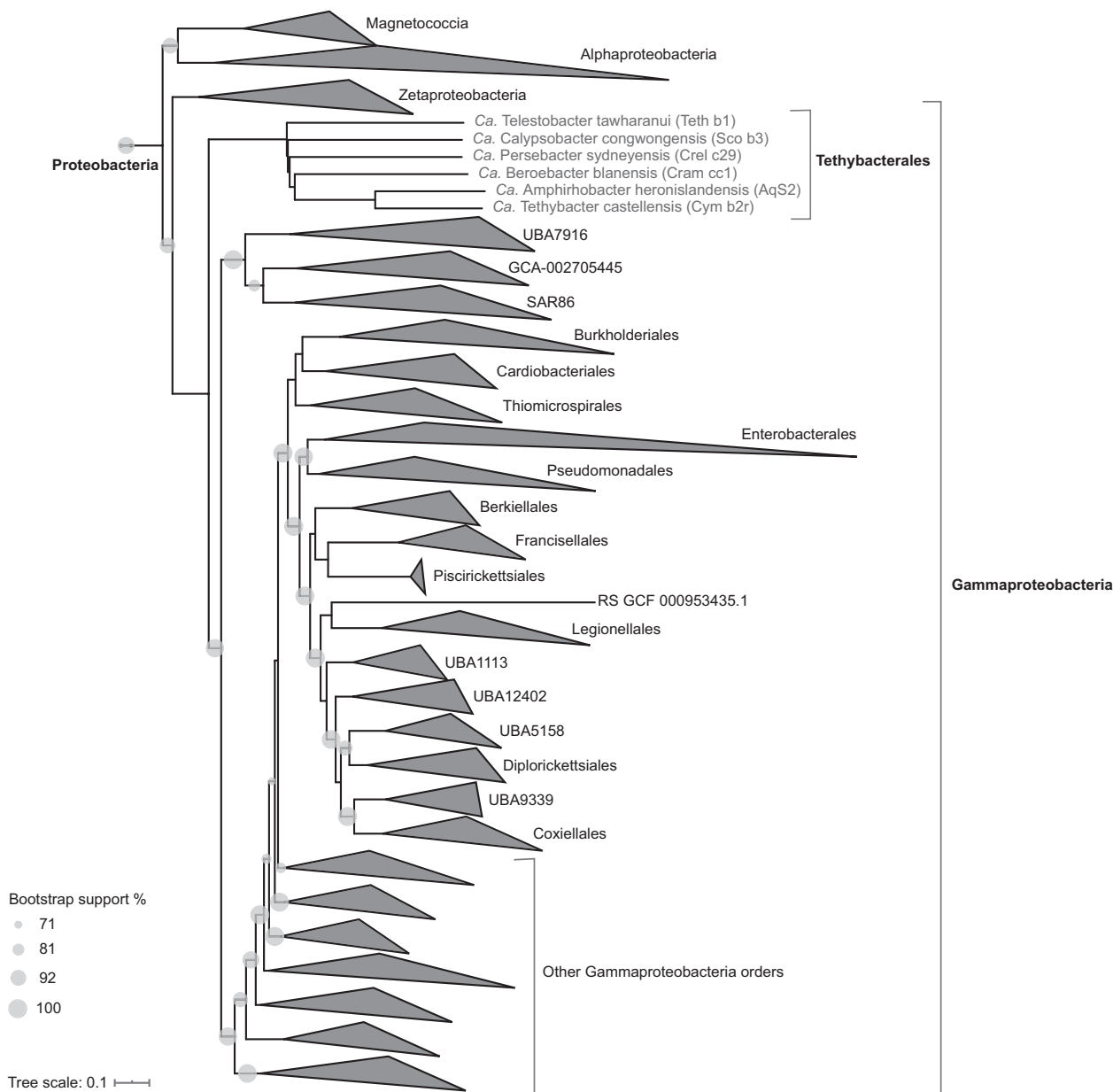

Fig. 2 Phylogenomic tree of the Proteobacteria based on GTDB. Names in red are draft genomes of this study. As AqS2 was already in the database (GCA 001750625.1) it was used in tree construction. Circles represent bootstrap support of 70–100%. Bar represents 10% sequence divergence. Genomes are named as species name (genome bin ID).

the genome Cym b2r in the same family as the AqS2 genome (RED value of 0.77), and the genomes Crel c29, Cram cc1, Sco b3, and Teth b1 (RED values of 0.58–0.59) in a separate family. The AAI analysis of the genomes supports this taxonomic discrimination. The genomes AqS2 and Cym b2r belong to the same family (AAI 53%), whilst the genomes Crel c29, Cram cc1, Sco b3, and Teth b1 belong to a second family (AAI 48.7–52.6%). Together both families belong to the same order (AAI 43–44.8%) using the thresholds proposed by Konstantinidis et al. [60].

Given these results and the refined placement of the order EC94, we propose taxonomic names following guidelines developed by the Genomics Standards Consortium [61], Konstantinidis et al. [60] and Chuvochina et al. [62]. As type material for this order we designate the MAG Cym b2r as it is highly complete (estimated to be 95%), contains a near full-length 16S rRNA gene (as well as 5S and 23S rRNA genes and 43 tRNA genes) and has the least number of contigs (Table 2) [62]. As all known members of EC94 are of marine origin, we chose to name them after the Oceanids (also known as sea nymphs), which are the 3000 daughters of the Titan Oceanus and Titaness Tethys in Greek mythology [63, 64]. In replacement of EC94 and the GTDB order-level placeholder AqS2, we propose the order name Ca. Tethybacterales, hailing from both Tethys and the sponge *Tethya stolonifera* studied here. The family Ca. *Tethybacteraceae* encompasses Ca. *Tethybacter castellensis* (represented by Cym b2r) and Ca. *Amphirhobacter heronislandensis* (represented by AqS2). The second family we propose is Ca. *Persebacteraceae*, encompassing Ca. *Persebacter sydneyensis* (represented by Crel c29), Ca. *Beroebacter blanensis* (represented by Cram cc1), Ca. *Calypsobacter congwongensis* (represented by Sco b3) and Ca. *Telestobacter tawharanui* (represented by Teth b1). A full etymological description can be found in the Supplementary Information (Table S5).

Ca. *T. castellensis* (Cym b2r), Ca. *A. heronislandensis* (AqS2), Ca. *P. sydneyensis* (Crel c29) and Ca. *B. blanensis* (Cram cc1) meet the minimum (>80% complete, <5% contamination, species genetic discreteness and ecological data; see below) and some additional (near full-length 16S rRNA gene and/or microscopy picture; see below) standards for descriptions of uncultured species as outlined by Konstantinidis et al. [60]. Ca. *C. congwongensis* (Sco b3) and Ca. *T. tawharanui* (Teth b1) only fall short on the genome completeness estimation; however, considering their symbiotic lifestyle (see below) and the fact that symbionts often have reduced genomes [62], it is possible that the draft genomes are more complete than CheckM predicts [34] Indeed, the Ca. Tethybacterales MAGs are all missing one CheckM bacterial single-copy marker gene (RecR). Konstantinidis et al. [60] noted that, for highly divergent species such as members of the proposed Ca. Tethybacterales, reliable matches may not be able to be retrieved against the universal protein-coding genes analysed by CheckM, which would result in an underestimated completeness. Given this, as well as the microscopic identification shown below, we feel confident in prescribing nomenclature.

Distribution analysis of Ca. Tethybacterales reveals global presence in both LMA and HMA sponges

In addition to world-wide reports of sponge-associated Ca. Tethybacterales, from Antarctic waters to tropical coral reefs [65, 66], our distribution analysis using the global data from the SEMP found Ca. Tethybacterales species belonging to the two described families to be present across a vast geographic range (Fig. 3). The BLASTn search found 157 OTUs in the SEMP belonging to the Ca. Tethybacterales, in 1077 samples of 159 identified sponge species (see Supplementary Information File 5).

Of these 159 sponge species, 77 are classified or predicted to be LMA sponges and had a RA of $13.29 \pm 23.23\%$
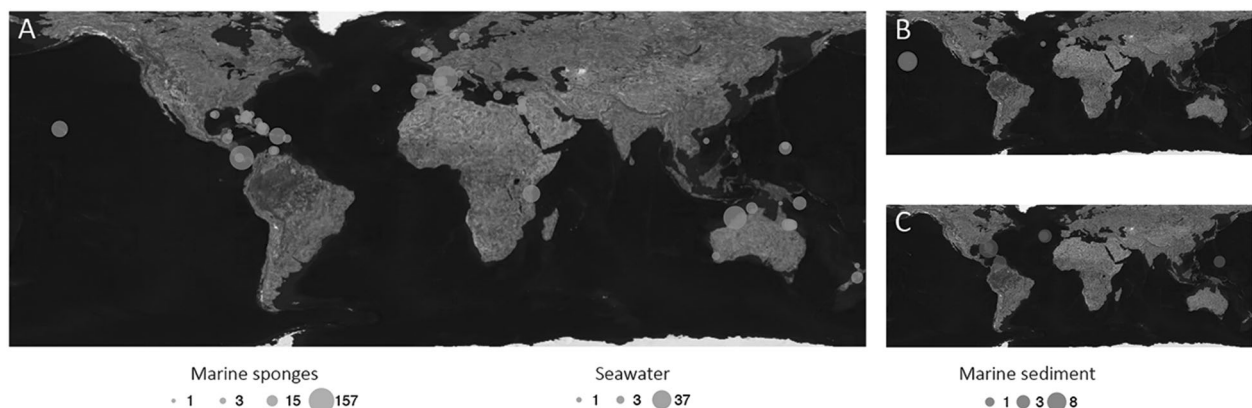

Fig. 3 Global distribution of Ca. Tethybacterales-containing samples collected for the Sponge Earth Microbiome Project. Bubbles indicate collection sites for (**a**) marine sponges, (**b**) seawater, and (**c**) marine sediment samples. Bubble sizes are proportional to number of samples as indicated.

(range: 0.05–93.40%), while 36 belonged to the HMA sponge group with RA of 1.66 ± 4.00% (range: 0.05–27.89%) [15], which were significantly different in RA (one-way ANOVA: $p = 5.48797E-14$). Ca. Tethybacterales dominance (e.g. >20% average relative sequence abundance) seems to be a feature predominately associated with some LMA sponges, as previously noted [20]. However, our analysis now clearly demonstrates that the Ca. Tethybacterales are also found in HMA sponges, contrasting with previous suggestions [20].

The global distribution patterns indicate a possible horizontal transmission of symbionts. Although there is also some strong evidence for vertical transmission of Ca. Tethybacterales symbionts in *T. rubra* [28], *Tedania* sp. [29], and of AqS2 in *A. queenslandica* (Ca. *A. heronislandensis*) [26], these two modes of transmission are not mutually exclusive. For example, Sipkema et al. [67] analysed the microbial communities of *Corticium candelabrum*, *C. crambe* and *Petrosia ficiformis* and identified vertical–horizontal transmission phylogenetic clusters (VHT clusters), where similar sponge-associated bacteria can be acquired either via vertical or horizontal transmission. It was proposed that successful horizontal transmission of bacteria could be facilitated through ELPs [67] (see further information below). Further, a recent study revealed that vertical transmission of symbiotic bacteria to sponge larva is fickle and the symbionts involved are often not as specific to a particular host species as previously assumed [68].

Members of the Ca. Tethybacterales are also found in marine invertebrates other than sponges (e.g. Fig. 1). Ca. Tethybacterales was found to make up 16–32% of the bacterial community of the coral *Erythropodium caribaeorum* [58, 69] and to be a dominant OTU in the coral *Scleronephthya gracillimum* [70]. Finding sponge-associated bacteria in corals is not uncommon and this has been defined by the concept of sponge/coral-specific clusters (SCCs), though SC112 was not originally considered one [18]. Shared and similar niches (e.g. biofilm-associated growth) in coral and sponge holobionts could facilitate the exchange of related bacteria (i.e. secondary symbiont transmission). Members of the Ca. Tethybacterales have also been recently reported in the healthy tissues of crown-of-thorns starfish *Acanthaster* cf. *solaris* at an abundance of 0.1–1% [71]. Since these starfish eat corals and sponges, it is possible that the presence of Ca. Tethybacterales is due to dietary uptake.

Our analyses also found Ca. Tethybacterales in 18% of seawater samples and 27% of marine sediment samples analysed in the SEMP (Fig. 3), although generally in very low abundances (RA%: 0.91 ± 2.36, range: 0.06–15.58 and RA%: 0.46 ± 0.71, range: 0.07–2.51, respectively). The discovery of sponge-associated bacteria in environmental samples is not surprising as many are thought to be widespread, but very rare. For example, Taylor et al. [19] investigated the prevalence of sponge-specific bacteria in

the environmental sample analysed by the International Census of Marine Microbes (ICoMM). They found a total of 23 reads assigned to SC112 (Ca. Tethybacterales) in 412 million 16S rRNA gene pyrotags (V6 region) generated by ICoMM in four sample types (Amazon-Guianas Water (AGW) sediment, Baltic Sea Proper (BPS) seawater, Salt Marsh (LSM) and New Zealand (NZS) marine sediment), which is congruent with our findings. Ca. Tethybacterales has also been detected in seawater samples by Gantt et al. [23] (<0.01% of all reads) and Matcher et al. [30] (<0.05% of all reads). Given the predicted metabolic repertoire (below), it is possible that these bacteria could persist in the seawater and sediments and this would in turn form reservoirs for horizontal symbiont transmission.

Ca. Tethybacterales exhibit diverse localisation patterns within the sponge tissue

Visualisation by CARD-FISH confirmed the existence of each proposed species within the tissue of its respective host sponge and revealed remarkable diversity in term of cell morphology and localisation (Fig. 4).

Ca. *T. tawharanui* cells appear large (2–2.5 μm) and clustered in structures that resemble bacteriocytes within the tissue of *T. stolonifera* (Fig. 4a). Individual spherical bacterial cells can be clearly seen surrounding an apparent sponge nuclei and probable vacuoles. *T. stolonifera* is considered a LMA sponge [72], though in this individual, bulging bacteriocytes are abundant throughout the mesohyl and in between megasclere tracts, forming a band from the inner pinacoderm to the outer cortex [73]. The intracellular lifestyle indicated here may also explain why the recovered Ca. *T. tawharanui* MAG had a relatively low read coverage compared to the other MAGs (Table 1), given that Ca. *T. tawharanui* comprises over 70% of *T. stolonifera*'s bacterial community based on 16S rRNA gene sequencing [22].

Ca. *P. sydneyensis* was difficult to localise due to the high autofluorescence of *C. incrustans*, even after photo bleaching the tissue. Figure 4b, however, shows a clear probe signal, where cells bear resemblance to Ca. *T. tawharanui* in shape, size and clustering, indicating a possible lifestyle associated with a bacteriocyte.

Ca. *T. castellensis* displays very large cell size (>3 μm) with a peculiar square shape (Fig. 4c). The cells are sparsely distributed on the pinacoderm side of the mesohyl in *C. concentrica*. The unusual shape and signal of these cells could be a result of active cell division and may be related to the minimum number of ribosomes required to emit a signal in both halves of a dividing cell [74].

Ca. *C. congwongensis* cells are comparatively small (1–1.75 μm) (Fig. 4d) and are more rod-like in shape. Ca. *C. congwongensis* is distributed abundantly and evenly throughout the mesohyl of *Scopalina* sp.. There is an

Fig. 4 Detection of Ca. Tethybacterales species by CARD-FISH in sponge hosts. **a** *Tethya stolonifera*, **b** *Crella incrustans*, **c** *Cymbastela concentrica* and **d** *Scopalina* sp.. Each sponge's target bacterium (Ca. *Telestobacter tawharanui*, Ca. *Persebacter sydneyensis* and Ca. *Tethybacter castellensis* Ca. *Calypsobacter congwongensis*, respectively) is labelled in green (Alexa Fluor™ 488) and counter-stained with blue (DAPI).
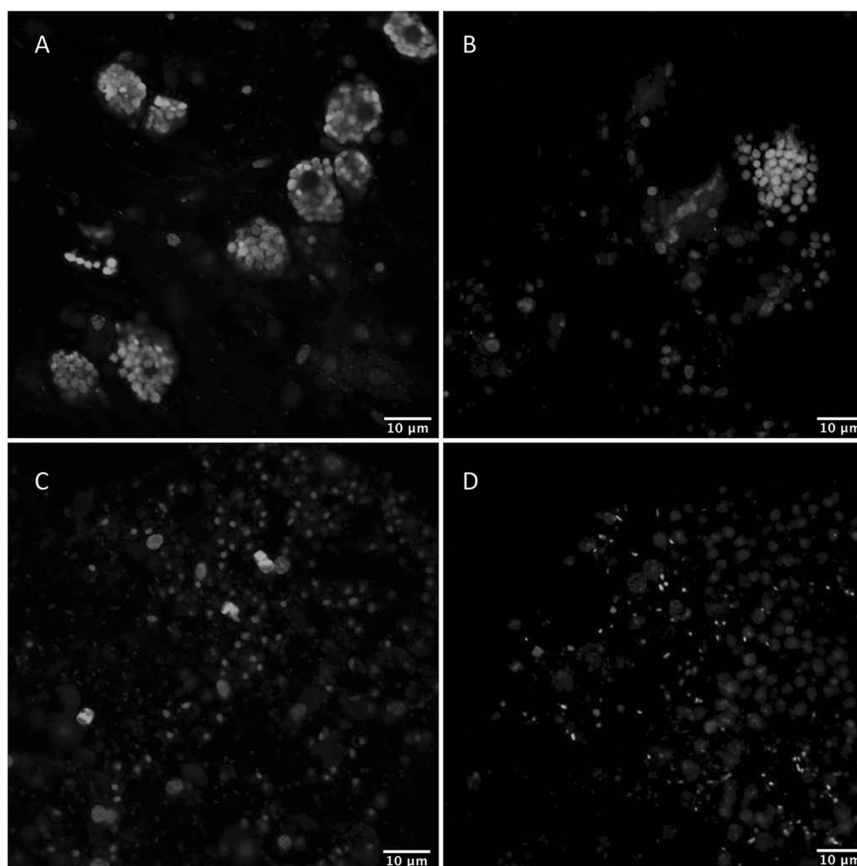
apparent increase in density of Ca. *C. congwongensis* cells at the pinacoderm edge of the tissue section (see Supplementary Information, Fig. S1).

Localisation experiments are useful to support functional predictions, thus helping to provide deeper insights into sponge-microbe interactions. Two studies have also employed FISH to gather morphological and localisation data on Ca. Tethybacterales from *C. crambe* [20] (i.e. Ca. *B. blanensis*) and *T. anhelans* [27]. Both studies describe a homogenous distribution of Ca. Tethybacterales cells throughout the sponge mesohyl, and in the case of *T. anhelans*, particularly near choanocyte chambers, similar to what we observed in *Scopalina* sp. Bacterial cells in close proximity to choanocyte chambers will be exposed to oxygenated seawater pumped in by the sponge, which could support aerobic respiration and an opportunity to promptly sequester seawater-derived nutrients. In both cases the cells appear oval shaped and vary in size from 0.5 to 1 μm in *C. crambe* and 1.01 to 1.6 μm in *T. anhelans*.

Our analysis reveals considerably more diversity in Ca. Tethybacterales localisation patterns than previously observed, likely involving an intracellular lifestyle in bacteriocytes. Bacteriocytes are specialised host cells that encase bacteria, which likely evolved from a mutualistic infection [75]. Bacteriocytes have also been reported in a

number of marine sponge species, mostly via transmission electron microscopy, although the taxonomic or phylogenetic nature of the intracellular bacteria is often undetermined [76–78]. The physiological reasons why sponges house bacteria within bacteriocytes is unclear, although various hypotheses have been proposed, such as that this type of cell structure can control the growth or metabolite production of bacteria that have been taken up from the environment [77]. Lastly, it could be the bacteria themselves require these specialised structures to survive and thus facilitate the specific metabolic or molecular interactions, such as those described in the following sections.

Metabolic features that unify the Ca. Tethybacterales

The five MAGs described in this study and Ca. *A. heronislandensis* share a number of predicted metabolic features, broadly summarised as being heterotrophic with the ability to use various carbon, nitrogen and sulfur sources, including taurine, spermidine and dimethylsulfoniopropionate (DMSP).

The genomes contain mostly complete and expressed tricarboxylic acid cycles (TCA) and glycolysis pathways (Supplementary Information File 3). Of the six genomes analysed, Ca. *P. sydneyensis* and Ca. *C. congwongensis* also encode both proteins required for the glyoxylate bypass,

isocitrate lyase (Ca. *P. sydneyensis* 329 TPM$_{av}$) and malate synthase (Ca. *P. sydneyensis* 788 TPM$_{av}$, Ca. *C. congwongensis* 9069 TPM$_{av}$), allowing acetate to be used for growth. The Ca. *B. blanensis* genome contains only the malate synthase (148 TPM). It important to note though that the absence of specific proteins should be interpreted with caution due to potential genome incompleteness.

The genomes also display a mostly complete and expressed electron transport chain for respiration. Two types of heme- and copper-containing terminal oxidase (HCO) superfamily cytochrome *c* oxidases are found in the genomes: a low-O_2-affinity cytochrome aa_3-type oxidase is encoded in Ca. *A. heronislandensis*, Ca. *T. castellensis* (682 TPM$_{sum}$), Ca. *B. blanensis* (1360 TPM$_{sum}$), Ca. *C. congwongensis* (CtaC: 3231 TPM$_{av}$) and Ca. *T. tawharanui*; and a high-affinity cytochrome cbb_3-type oxidase is found in Ca. *B. blanensis* (6210 TPM$_{sum}$), Ca. *C. congwongensis* and Ca. *T. tawharanui*. The symbiont Ca. *P. sydneyensis* genome lacks one catalytic subunit (CtaD) but encodes and expresses subunit two (CtaC: 637 TPM$_{av}$) of the aa_3-type cytochrome complex IV, suggesting that the absence of a complete Complex IV is due to incompleteness of the genome. Ca. *T. castellensis* further encodes a non-HCO cytochrome *bd* oxidase (CydAB: 2188 TPM$_{sum}$), which has a high affinity for O_2 and is typically induced under O_2-limited conditions [79]. The presence and expression of both high- and low-affinity cytochrome *c* oxidases is likely a reflection for the differing oxygen concentrations experienced in the sponge microenvironment [80] and has been observed in other sponge symbionts [56].

Ca. *T. castellensis* displays the potential for nitrate respiration via the NarGHJI system (1898 TPM$_{sum}$) and also encodes a nitrate/nitrite transporter gene, NarK (Table 3); both were expressed. As with the presence of high-affinity cytochrome *c* oxidases, nitrate respiration would allow these bacteria to thrive under microaerobic conditions or when oxygen levels rapidly change, such as at the interface of choanocyte chambers and mesohyl tissues [81] where these bacteria have been localised (see Fig. 4). Nitrate respiration is also documented in an alphaproteobacterial symbiont of *C. concentrica* [56]. Ca. *P. sydneyensis* (2169 TPM$_{sum}$), Ca. *B. blanensis* (NapC: 188 TPM), Ca. *C. congwongensis* (NapA: 1420 TPM$_{av}$) and Ca. *T. tawharanui* encode a periplasmic NapABC nitrate reductase, which is possibly involved in dissimilatory nitrate reduction and redox balancing [82]. The Ca. *B. blanensis* (149 TPM) and Ca. *T. tawharanui* genomes also carry a copper-containing nitrite reductase gene (NirK) that can reduce nitrite to nitric oxide (NO). The enzymes required for subsequent reduction of NO are absent in all of the six genomes. In the absence of a mechanism for NO detoxification, NO accumulation would reach toxic levels [83], meaning that nitrite is unlikely to be used for respiration, and may solely be dissimilatory. The consistent lack of genes involved in NO reduction, despite the presence of nitrate and nitrite reductases, has been previously reported in the bacterial metagenomes from the marine sponges *Rhopaloeides odorabile* and *C. concentrica* [21].

Transporter systems and metabolic pathways indicate differential utilisation of environmentally- or sponge-derived nutrients in the Ca. Tethybacterales

The majority of the transporters present in the Ca. Tethybacterales MAGs belong to the ATP-binding cassette (ABC) transporter superfamily, which can target compounds at low concentrations [84], of which several are significantly enriched when compared to free-living, marine Gammaproteobacteria (Supplementary File 6). Other transporter families present are the tripartite ATP-independent periplasmic transporters (TRAP) [85], the solute:sodium symporter family of the amino acid–polyamine–organo–cation superfamily [86] the major facilitator superfamily [87], the betaine–choline–carnitine transporter family (BCCT) [88], and the ammonium transport proteins family [89].

The MAGs display a diverse, but variable, range of transporters reflecting the heterotrophic metabolism mentioned above (Table 3). Sugar transporters are one of the most notable differences between families. MAGs of the family Ca. *Tethybacteraceae* (Ca. *A. heronislandensis* and Ca. *T. castellensis*) encode more sugar transporters, which are almost entirely absent in the genomes of the Ca. *Persebacteraceae* (Ca. *P. sydneyensis*, Ca. *B. blanensis*, Ca. *C. congwongensis* and Ca. *T. tawharanui*). Sugar utilisation has been previously noted as an adaptation by members of the phylum Poribacteria to the sponge environment, possibly by benefitting from the recycling or scavenging of the carbohydrate matrix of the host [90].

Taurine is often found in crustaceans and sponges [91] and cultured alphaproteobacterial symbionts from sponges have been recently shown to have the genomic potential to degrade this organosulfonate [92]. Ca. Tethybacterales show a differential ability to utilise taurine as a potential source of carbon, nitrogen, and sulfur. ABC transporters for taurine import are encoded in the four members of the Ca. *Persebacteraceae* (Table 3), which are rarely found in free-living, marine Gammaproteobacteria (Supplementary File 6). These four members also encode and express enzymes to degrade taurine to acetyl-CoA (via sulfoacetaldehyde and acetyl phosphate: Xsc: Ca. *P. sydneyensis* 1322 TPM$_{av}$, Ca. *B. blanensis* 519 TPM, Ca. *C. congwongensis* 2048 TPM$_{av}$, Pta: Ca. *P. sydneyensis* 109 TPM$_{av}$ and Ca. *B. blanensis* 352 TPM), while these genes and taurine transporters were not detected in the genomes of Ca. *T. castellensis* and Ca. *A. heronislandensis*. Since Ca. *C. congwongensis*, Ca. *P. sydneyensis* and potentially Ca. *B.*

Table 3 Transport systems (TS) present in Ca. Tethybacterales.

			Ca. A. heronislandensis	Ca. T. castellensis	Ca. P. sydneyensis	Ca. B. blanensis	Ca. C. congwongensis	Ca. T. tawharanui
ABC		Phosphate		845	242	1145		
		Taurine			884	12591	8584	
		Spermidine/putrescine		562	5365	4631		
		Glycine betaine/proline			1628	1915		
		Tungstate			81	298		
		Zinc		645				
		Phospholipid		2806	1525	10328	11549	
	CUT1	Putative multiple sugar		4022				
		Maltose/Trehalose				1147		
		Polyol			19750	1013		
		Alpha-glucosides		1883				
		Glucose		4022				
	CUT2	Ribose		309				
		Xylose						
		Fructose		549				
	PAAT	General L-amino acid		17533	43936	32420	62968	
	HAAT	Branched-chain amino acid			665	818		
BCCT		Glycine betaine transporter				2442		
MFS	**NNP**	Nitrate/nitrite transporter		160				
SSSF		Cation/acetate symporter						
AMT		Ammonium transporter		1078		290		
ABC-2		Lipo-oligosaccharide			157	1080		
		Lipoprotein release system		164	272	1837		

Presence (shaded) of ABC transporters is determined by the genome having at minimum two of the TS genes, a breakdown of subunits can be found in the Supplementary Information (File 3). Numbers on the shading represent expression (TPM$_{sum}$) of all the genes in the TS.

blanensis encode a glyoxylate bypass (see above), taurine could potentially be used as a source of carbon for growth. Taurine degradation also yields sulfite, which can be reduced to sulfide for biosynthesis using an assimilatory sulfite reductase (CysJI), which is present in Ca. *B. blanensis* (521 TPM$_{sum}$), Ca. *P. sydneyensis* (1238 TPM$_{sum}$), Ca. *C. congwongensis* and Ca. *T. tawharanui*. In contrast, genes are present in all six symbionts for the synthesis of cysteine from serine (CysE: Ca. *T. castellensis* 93 TPM$_{av}$, Ca. *P. sydneyensis* 643 TPM$_{av}$, Ca. *B. blanensis* 169 TPM, CysM: ^{V}*T. castellensis* 90 TPM$_{av}$, Ca. *P. sydneyensis* 284 TPM$_{av}$). However, none of the six MAGs encode any identifiable genes for the reductive assimilation of sulfate, suggesting that Ca. *Persebacteraceae* are dependent on sulfur derived from taurine degradation, while Ca. *Tethybacteraceae* may require other organic sulfur sources to meet their cellular sulfur demands. In Ca. *B. blanensis*, sulfite generated by taurine desulfonation may also be detoxified by a Soe-type sulfite dehydrogenase, and oxidised to sulfate (2569 TPM) [93].

The Ca. *C. congwongensis* genome uniquely encodes sulfur oxidation proteins, including those involved in the oxidation of sulfide (SoxF) and thiosulfate (SoxA, SoxB) [94, 95], similar to some sponge-associated Chromatiales [96, 97]. The function of these proteins in Ca. *C.*

congwongensis may be to use sulfur compounds as energy sources, to supplement an organotrophic metabolism. Certain small organic substrates might also be oxidised as sources of reductant, including formate (FdhA: Ca. *P. sydneyensis* 463 TPM_{av}, Ca. *B. blanensis* 613 TPM Ca. *C. congwongensis*, Ca. *T. tawharanui*), methanol (MoxF: Ca. *T. tawharanui*), and di/trimethylamine (Tmd: Ca. *P. sydneyensis* 86 TPM_{av}). However, no symbiont exhibits an apparent genomic potential for methylotrophic growth. An aerobic carbon monoxide dehydrogenase (CoxLMS) is encoded in Ca. *A. heronislandensis* and Ca. *T. tawharanui*, suggesting CO oxidation can be used as an energy source. In these cases, it appears that oxidation is coupled to oxygen as the terminal electron acceptor (see above).

Marine phytoplankton often accumulate compatible solutes such as glycine betaine and DMSP in order to maintain favourable osmotic tensions and positive turgor [98]. These compounds are released into the marine environment upon cell lysis and are therefore also likely introduced into filter-feeding sponges. The symbionts Ca. *B. blanensis*, Ca. *P. sydneyensis*, Ca. *C. congwongensis* and Ca. *T. tawharanui* encode ABC transporters for the uptake of glycine betaine (Table 3), whereas Ca. *A. heronislandensis* and Ca. *B. blanensis* have BCCT family transporters for glycine betaine uptake [88]. In addition to being used as a potential osmolyte, glycine betaine may serve as a carbon and nitrogen source in some of these symbionts [99]. In support of this, Ca. *P. sydneyensis*, Ca. *B. blanensis*, Ca. *C. congwongensis* and Ca. *T. tawharanui* all encode enzymes for the degradation of glycine betaine to glycine through demethylation (Bhmt) via dimethylglycine (Dmgdh) and sarcosine (SoxBDAG) [100]. A glycine oxidase (ThiO) encoded in Ca. *P. sydneyensis* (240 TPM_{av}), Ca. *B. blanensis* and Ca. *C. congwongensis* could then convert glycine to glyoxyl imine and H_2O_2. Glyoxyl imine can then be used for the biosynthesis of the thiazole ring of thiamine or spontaneously hydrolyses to glyoxylate and ammonia [101]. Ca. *A. heronislandensis*, Ca. *T. castellensis*, Ca. *P. sydneyensis*, Ca. *B. blanensis*, and Ca. *T. tawharanui* also have genes for the synthesis of glycine betaine from choline (BetA: Ca. *T. castellensis* 40 TPM_{av}, Ca. *P. sydneyensis* 103–156 TPM_{av} Ca. *B. blanensis* 308–465 TPM, BetB: Ca. *P. sydneyensis* 195 TPM_{av}, Ca. *B. blanensis* 267 TPM), with Ca. *B. blanensis* additionally having choline sulfatase (310 TPM) for the conversion of choline sulfate to choline.

The sulfur analog of glycine betaine, DMSP [98], may also be imported by BCCT transporters [88]. DMSP can be demethylated by DMSP demethylase (DmdA: Ca. *P. sydneyensis* 192–17461 TPM_{av}, Ca. *B. blanensis* 543–3244 TPM, Ca. *C. congwongensis* and Ca. *T. tawharanui*) or cleaved by DMSP lyase (DddP: Ca. *P. sydneyensis* 310 TPM_{av}, Ca. *B. blanensis* 186 TPM and Ca. *T. tawharanui*).

In the latter case, the gas dimethylsulfide is released as a by-product. DMSP degradation seems to be widespread in sponge symbionts given recent evidence for the presence of relevant degradation genes in genomes of some alphaproteobacterial sponge symbionts [102] and the metagenomes of two Antarctic sponges [103].

Sulfopropanediol (2,3-dihydroxypropane-1-sulfonate; DHPS) is another common metabolite in the marine environment, which is produced and released by diatoms [104]. The members of the family Ca. *Persebacteraceae* exhibit the genetic capacity to utilise DHPS as a carbon source, while we could find no support for this in the Ca. *Tethybacteraceae*. While the method of uptake in unclear, the four genomes of the Ca. *Persebacteraceae* encode sulfopropanediol 3-dehydrogenases (HspN: Ca. *P. sydneyensis* 430 TPM_{av}, Ca. *B. blanensis* 342 TPM) and (2R)-sulfolactate sulfo-lyase (SuyAB: Ca. *P. sydneyensis* 997 TPM_{sum}, SuyB: Ca. *B. blanensis* 345 TPM).

The Ca. Tethybacterales MAGs also contain genes for the import and catabolism of spermidine and putrescine, which are polyamines commonly found in animals, including sponges [105, 106]. The primary transport system PotABCD (Table 3) can import both spermidine and putrescine, but has preference for the former [107, 108]. Again, the genomes of the Ca. *Persebacteraceae* apparently distinguish themselves from those of the Ca. *Tethybacteraceae* by encoding enzymes to degrade putrescine to succinate via the transamine pathway (SpuC: Ca. *P. sydneyensis* 450 TPM_{av}, Ca. *B. blanensis* 586 TPMGabD: Ca. *P. sydneyensis* 411 TPM_{av}, Ca. *B. blanensis* 215 TPM) and the glutamated putrescine pathway (PuuA/PuuB/PuuC: Ca. *P. sydneyensis* 594 TPM_{sum}, Ca. *B. blanensis* 683 TPM_{sum}, PuuC: Sco 4782 TPM_{av}) [109] yielding ammonia, alanine and glutamate. Succinate would then be available to enter the TCA cycle for energy conservation and biosynthesis [110]. Even though PotABCD prefers spermidine import over putrescine, genes for spermidine-specific degradation (e.g. SpdH) are absent [111]. This somewhat contradictory instance has also been observed in the marine alphaproteobacterium *Ruegeria pomeroyi* [110], where it was concluded that given the expression of homologues for putrescine degradation, exogenous spermidine is imported and converted to putrescine (then ultimately succinate) through a mechanism yet to be determined. Indeed, spermidine to putrescine conversion has been documented in other systems [112] and therefore could also be occurring here.

All six Ca. Tethybacterales also encode proteins for the synthesis of polyhydroxyalkanoate, a carbon and energy storage polymer produced from acetyl-CoA (PhaA/PhaB/PhaC: Ca. *T. castellensis* 3650 TPM_{sum}, Ca. *P. sydneyensis* 2870 TPM_{sum}, Ca. *B. blanensis* 2744 TPM_{sum} and Ca. *C. congwongensis* 13418 TPM_{sum}). Symbionts Ca. *C. congwongensis* and Ca. *T. tawharanui* also encode polyphosphate kinase, allowing for the accumulation of inorganic

Table 4 Number of genes encoding ELP classes in MAGs.

	Ca. *A. heronislandensis*	Ca. *T. castellensis*	Ca. *P. sydneyensis*	Ca. *B. blanensis*	Ca. *C. congwongensis*	Ca. *T. tawharanui*
ANK	2	2	2	10	1	4
TPR	10	9	14	15	12	13
PQQ	1	1	1	1	1	3
SEL1	1	2	0	0	1	1
FN3	0	4	0	0	0	0
Cadherin	0	0	0	1	0	0

polyphosphate, for energy and/or phosphate storage [113]. Cyanobacterial symbionts of sponges have been previously found to sequester phosphate in this fashion, thought to be driven, in part, by oxic–anoxic perturbations within the sponge tissue, which ultimately impacts phosphate cycling within the benthos [114]. Polyphosphate accumulation has also been associated with bacterial environmental durability and virulence [115, 116] and while many species of endo-symbiotic bacteria appear to have lost this ability [115], retaining this function in sponge symbionts may be related to the horizontal transmission (see above) and the requirement to persist outside of their host. Overall, these data indicate that members of the Ca. Tethybacterales have strategies to cope with carbon, nitrogen (see Supplementary Information File 1) and, for some species, phosphate imbalances.

Symbiosis effectors in the Ca. Tethybacterales

Given their intimate interaction with and within sponge cells (see Fig. 4), the Ca. Tethybacterales must contain and express effector molecules to mediate and control their symbiosis, especially as sponges feed on bacteria and nutrient particles through phagocytosis [8]. ELPs are pre-dicted to be involved in mediating sponge-bacteria inter-actions and have been shown to inhibit eukaryotic phagocytosis, which would be an important feature for a symbiont to persist inside the sponge [10–12]. The Ca. Tethybacterales genomes contain ELPs belonging to seven classes: ankyrin repeats (ANK), tetratricopeptide repeats (TPR), pyrrolo-quinoline quinone repeats (PQQ), Sel1-like repeats (SEL1) fibronectin domain III repeats (FN3), and cadherin (a calcium-dependent cell adhesion molecule) (Table 4). Of these, ANK-type ELPs are more abundant in the Ca. Tethybacterales than in free-living, marine Gam-maproteobacteria, while the other ELP types do not appear to be enriched in the Ca. Tethybacterales. The distribution and abundance of these proteins is generally consistent across all genomes (Table 4), with the exception of FN3 only being present in Ca. *T. castellensis* and cadherin only being found in Ca. *B. blanensis*. Expression (>40 TPM) of all ELPs is observed in Ca. *T. castellensis* and Ca. *B. blanensis*, whereas expression of only TPR and PQQ is observed in Ca. *P. sydneyensis* and only TPR has

detectable expression in Ca. *C. congwongensis* (Supple-mentary Information File 3). Such subtle differences in the presence and/or expression of certain ELPs might deter-mine whether a bacterium can reside extracellularly (e.g. by avoiding an initial uptake through a phagocytic cells) or end up in intracellular localisation (e.g. vacuole) by blocking the maturation of a phagosome [10, 11]. Toti-potent archaeocytes are the main phagocytes in the sponge mesohyl [117] and an ELP-mediated accumulation of intracellular bacteria could contribute to the formation of bacteriocytes (e.g. structure in Fig. 4).

Often considered a feature of pathogenicity, by definition another form of symbiosis, all Ca. Tethybacterales genomes also contain genes for the release of lipoproteins via the ABC-2 transporter LolABCDE complex (Table 3). The release of lipoproteins is reported to play a direct role in bacterial colonisation/invasion, evasion of host defences and immuno-modulation [118] and strategies advantageous to a symbiont in the host environment. The genomes Ca. *P. sydneyensis*, Ca. *B. blanensis* and Ca. *C. congwongensis* also contain lipo-oligosaccharide exporting nodulation genes, NodIJ (Table 3). Better known from legume-rhizobia symbiosis, the NodIJ transport system secrets host-specific Nod factors important for the recognition of rhizobia during legume nodulation [119]. The possible role of nodulation genes in bacterial symbionts of sponges is completely unexplored but could represent another mechanism for mediating symbiosis.

Conclusion

Our analysis reveals that Ca. Tethybacterales are a large, widespread, and divergent clade of (almost exclusively) sponge-associated symbionts, which suggests that they have adapted to the sponge environment reflecting a degree of co-evolution, as also proposed by Matcher et al. [30]. We propose that the Ca. Tethybacterales entered a symbiosis with sponges relatively early in evolutionary time, as indi-cated by their deep branching within the Gammaproteo-bacteria. This early symbiosis is reflected in the common features that the Ca. Tethybacterales share, such as their heterotrophic metabolism and respiratory pathways that have allowed them to occupy specific niches in the sponge

and derive energy and carbon either from their host or compounds in seawater. However, we also find that members of the Ca. Tethybacterales have subsequently functionally radiated, reflected in their quite distinct lifestyles, most notably their diverse morphologies, predicted substrate preferences and localisation within the sponge tissue. These differences may be the result of a more specialised co-evolution with their present-day sponge hosts, similar to the diverse characteristics observed in the Rickettsiales, which have evolved to have mutualistic, commensal, and parasitic relationships with an extensive array of hosts [120]. To expand on these hypotheses, further acquisition and analysis of additional Ca. Tethybacterales genomes are required, complemented with the inference of a molecular clock analysis [121].

Data availability

Raw sequencing data are publicly available at the NCBI Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/Traces/sra) under Project Accession number PRJNA589708.

RNA-Seq data from *Cymbastela concentrica* and *Crella incrustans* is available on request through Bioplatforms Australia (https://data.bioplatforms.com/organisation/about/australian-microbiome). Sample IDs are found in Supplementary File 1, Table S2.

MAGs can be accessed publicly via JGI GOLD (https://gold.jgi.doe.gov/), GOLD Analysis Project IDs can be found in Supplementary File 1, Table S6.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Bergquist PR. Sponges. London, United Kingdom: Hutchinson and Co. Ltd; 1978.
2. Brain CKB, Prave AR, Hoffmann KH, Fallick AE, Botha A, Herd DA, et al. The first animals: Ca. 760-million-year-old sponge-like fossils from Namibia. S Afr J Sci. 2012;108:1–8.
3. Hooper JNA, van Soest RWM. Systema Porifera: a guide to the classification of sponges. New York: Kluwer Academic/Plenum Publishers; 2002.
4. Webster NS. Sponge disease: a global threat? Environ Microbiol. 2007;9:1363–75.
5. Bell JJ. The functional roles of marine sponges. Estuar Coast Shelf Sci. 2008;79:341–53.
6. de Goeij JM, van Oevelen D, Vermeij MJA, Middelburg JJ, Osinga R, de Goeij AFPM, et al. Surviving in a Marine Desert: the sponge loop retains resources within coral reefs. Science. 2013;342:108–10.
7. Taylor MW, Radax R, Steger D, Wagner M. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. Microbiol Mol Biol Rev. 2007;71:295–347.
8. Wehrl M, Steinert M, Hentschel U. Bacterial uptake by the marine sponge *Aplysina aerophoba*. Microb Ecol. 2007;53:355–65.
9. Thomas T, Rusch D, DeMaere MZ, Yung PY, Lewis M, Halpern A, et al. Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. ISME J. 2010;4:1557–67.
10. Nguyen MTHD, Liu M, Thomas T. Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. Mol Ecol. 2013;23:1635–45.
11. Reynolds D, Thomas T. Evolution and function of eukaryotic-like proteins from sponge symbionts. Mol Ecol. 2016;25:5242–53.
12. Jahn MT, Arkhipova K, Markert SM, Stigloher C, Lachnit T, Pita L, et al. A phage protein aids bacterial symbionts in eukaryote immune evasion. Cell Host Microbe. 2019;26:542–50.e5.
13. Degnan SM. The surprisingly complex immune gene repertoire of a simple sponge, exemplified by the NLR genes: a capacity for specificity? Dev Comp Immunol. 2015;48:269–74.
14. Hentschel U, Usher KM, Taylor MW. Marine sponges as microbial fermenters. FEMS Microbiol Ecol. 2006;55:167–77.
15. Moitinho-Silva L, Steinert G, Nielsen S, Hardoim CCP, Wu YC, McCormack GP, et al. Predicting the HMA-LMA status in marine sponges by machine learning. Front Microbiol. 2017;8:1–14.
16. Moitinho-Silva L, Nielsen S, Amir A, Gonzalez A, Ackermann GL, Cerrano C, et al. The sponge microbiome project. Gigascience. 2017;6:1–7.
17. Hentschel U, Hopke J, Horn M, Friedrich AB, Wagner M, Hacker J, et al. Molecular evidence for a uniform microbial community in sponges from different oceans. Appl Environ Microbiol. 2002;68:4431–40.
18. Simister RL, Deines P, Botté ES, Webster NS, Taylor MW. Sponge-specific clusters revisited: a comprehensive phylogeny of sponge-associated microorganisms. Environ Microbiol. 2012;14:517–24.
19. Taylor MW, Tsai P, Simister RL, Deines P, Botte E, Ericson G, et al. "Sponge-specific" bacteria are widespread (but rare) in diverse marine environments. ISME J. 2013;7:438–43.
20. Croué J, West NJ, Escande M-L, Intertaglia L, Lebaron P, Suzuki MT. A single betaproteobacterium dominates the microbial community of the crambescidine-containing sponge *Crambe crambe*. Sci Rep. 2013;3:2583.
21. Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS, et al. Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. PNAS. 2012;109:E1878–87.
22. Simister RL, Taylor MW, Rogers KM, Schupp PJ, Deines P. Temporal molecular and isotopic analysis of active bacterial

communities in two New Zealand sponges. FEMS Microbiol Ecol. 2013;85:195–205.

23. Gantt SE, López-Legentil S, Erwin PM. Stable microbial communities in the sponge *Crambe crambe* from inside and outside a polluted Mediterranean harbor. FEMS Microbiol Lett. 2017; 364:1–7.

24. Bin JeongJ, Kim KH, Park JS. Sponge-specific unknown bacterial groups detected in marine sponges collected from Korea through barcoded pyrosequencing. J Microbiol Biotechnol. 2015;25:1–10.

25. Thiel V, Neulinger SC, Staufenberger T, Schmaljohann R, Imhoff JF. Spatial distribution of sponge-associated bacteria in the Mediterranean sponge *Tethya aurantium*. FEMS Microbiol Ecol. 2006;59:47–63.

26. Fieth RA, Gauthier M-EA, Bayes J, Green KM, Degnan SM. Ontogenetic changes in the bacterial symbiont community of the tropical demosponge *Amphimedon queenslandica*: metamorphosis is a new beginning. Front Mar Sci. 2016;3:1–20.

27. Batani G. Fluorescence in situ hybridisation for the localisation and culturing of marine bacteria: co-localisation of symbionts in sponges (Unpublished PhD thesis chapter). UNSW Sydney. 2018.

28. Waterworth SC, Jiwaji M, Kalinski JCJ, Parker-Nance S, Dorrington RA. A place to call home: an analysis of the bacterial communities in two *Tethya rubra* Samaai and Gibbons 2005 populations in algoa bay, South Africa. Mar Drugs. 2017;15:95.

29. Wu S, Ou H, Liu T, Wang D, Zhao J. Structure and dynamics of microbiomes associated with the marine sponge *Tedania* sp. during its life cycle. FEMS Microbiol Ecol. 2018;94:1–9.

30. Matcher GF, Waterworth SC, Walmsley TA, Matsatsa T, Parker-Nance S, Davies-Coleman MT, et al. Keeping it in the family: coevolution of latrunculid sponges and their dominant bacterial symbionts. Microbiologyopen. 2017;6:1–13.

31. Jackson SA, Flemer B, McCann A, Kennedy J, Morrissey JP, O'Gara F, et al. Archaea appear to dominate the microbiome of *Inflatella pellicula* deep sea sponges. PLoS ONE. 2013;8:1–8.

32. Steinert G, Taylor MW, Deines P, Simister RL, de Voogd NJ, Hoggard M, et al. In four shallow and mesophotic tropical reef sponges from Guam the microbial community largely depends on host identity. PeerJ. 2016;4:e1936.

33. Gauthier M-EA, Watson JR, Degnan SM. Draft genomes shed light on the dual bacterial symbiosis that dominates the microbiome of the coral reef sponge *Amphimedon queenslandica*. Front Mar Sci. 2016;3:1–18.

34. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Cold Spring Harb Lab Press. 2014;25:1043–55.

35. Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Verezemska O, Isbandi M, et al. Genomes OnLine database (GOLD) v.6: data updates and feature enhancements. Nucleic Acids Res. 2017;45:D446–56.

36. Huntemann M, Ivanova NN, Mavromatis K, James Tripp H, Paez-Espino D, Palaniappan K, et al. The standard operating procedure of the DOE-JGI microbial genome annotation pipeline (MGAP v.4). Stand Genom Sci. 2015;10:4–9.

37. Eddy S. Profile hidden Markov models. Bioinformatics. 1998;14: 755–63.

38. Pruesse E, Peplies J, Glöckner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 2012;28:1823–9.

39. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, et al. ARB: a software environment for sequence data. Nucleic Acids Res. 2004;32:1363–71.

40. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project:

Improved data processing and web-based tools. Nucleic Acids Res. 2013;41:590–6.

41. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.

42. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44:W242–5.

43. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2020;36:1925–7.

44. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 2018;36:996–1004.

45. Konstantinidis KT, Tiedje JM. Towards a genome-based taxonomy for prokaryotes. J Bacteriol. 2005;187:6258–64.

46. Rodriguez-R LM, Konstantinidis KT. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. Peer J Prepr. 2016;4:e1900v1.

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

48. Díez-Vives C, Moitinho-Silva L, Nielsen S, Reynolds D, Thomas T. Expression of eukaryotic-like protein in the microbiome of sponges. Mol Ecol. 2017;26:1432–51.

49. Öztürk B, De Jaeger L, Smidt H, Sipkema D. Culture-dependent and independent approaches for identifying novel halogenases encoded by *Crambe crambe* (marine sponge) microbiota. Sci Rep. 2013;3:1–9.

50. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:1–16.

51. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25.

52. Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN. RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics. 2010;26:493–500.

53. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. 2012;131:281–5.

54. Webster NS, Hill RT. The culturable microbial community of the great barrier reef sponge *Rhopaloeides odorabile* is dominated by an α-Proteobacterium. Mar Biol. 2001;138:843–51.

55. Engelberts JP, Robbins SJ, de Goeij JM, Aranda M, Bell SC, Webster NS. Characterization of a sponge microbiome using an integrative genome-centric approach. ISME J. 2020;14: 1100–10.

56. Moitinho-Silva L, Díez-Vives C, Batani G, Esteves AIS, Jahn MT, Thomas T. Integrated metabolism in sponge-microbe symbiosis revealed by genome-centered metatranscriptomics. ISME J. 2017;11:1651–66.

57. Lobb B, Tremblay BJM, Moreno-Hagelsieb G, Doxey AC. An assessment of genome annotation coverage across the bacterial tree of life. Microb Genom. 2020;6:e000341.

58. Lopez JV, Ranzer LK, Ledger A, Schoch B, Duckworth A, Mccarthy PJ, et al. Comparison of bacterial diversity within the coral reef sponge, axinella corrugata, the encrusting coral erythropodium caribaeorum. Proc. 11 Int. Coral Reef Symp. 2008;2:1362–6.

59. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer KH, et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Microbiol. 2014;12:635–45.

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

61. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-smith M, Doud D, Jarett J, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol. 2017;35:725–31.

62. Chuvochina M, Rinke C, Parks DH, Rappé MS, Tyson GW, Yilmaz P, et al. The importance of designating type material for uncultured taxa. Syst Appl Microbiol. 2019;42:15–21.

63. Tripp E. Crowell's handbook of classical mythology. New York: Thomas Y. Crowell Company; 1970.

64. Athanassakis AN. Hesiod: Theogony, Works and Days, Shield. 2nd ed. Baltimore and London: The John Hopkins University Press; 2005.

65. Webster NS, Negri AP, Munro MMHG, Battershill CN. Diverse microbial communities inhabit Antarctic sponges. Environ Microbiol. 2004;6:288–300.

66. Coelho FJRC, Cleary DFR, Gomes NCM, Pólonia ARM, Huang YM, Liu LL, et al. Sponge prokaryote communities in Taiwanese coral reef and shallow hydrothermal vent ecosystems. Microb Ecol. 2018;75:239–54.

67. Sipkema D, de Caralt S, Morillo JA, Al-Soud WAB, Sørensen SJ, Smidt H, et al. Similar sponge-associated bacteria can be acquired via both vertical and horizontal transmission. Environ Microbiol. 2015;17:3807–21.

68. Bjork JR, Diez-Vives C, Astudillo-Garcia C, Archie E, Montoya JM. Vertical transmission of sponge microbiota is inconsistent and unfaithful. Nat Ecol Evol. 2019;3:1172–83.

69. Gonzalez-Zapata FL, Bongaerts P, Ramírez-Portilla C, Adu-Oppong B, Walljasper G, Reyes A, et al. Holobiont diversity in a reef-building coral over its entire depth range in the mesophotic zone. Front Mar Sci. 2018;5:1–13.

70. Yang S, Sun W, Zhang F, Li Z. Phylogenetically diverse denitrifying and ammonia-oxidizing bacteria in corals *Alcyonium gracillimum* and *Tubastraea coccinea*. Mar Biotechnol. 2013;15:540–51.

71. Høj L, Levy N, Baillie BK, Clode PL, Strohmaier RC, Siboni N, et al. Crown-of-thorns sea star *Acanthaster* cf. *solaris* has tissue—characteristic microbiomes with potential roles in health and reproduction. Appl Environ Microbiol. 2018;84:1–18.

72. Schmitt S, Deines P, Behnam F, Wagner M, Taylor MW. Chloroflexi bacteria are more diverse, abundant, and similar in high than in low microbial abundance sponges. FEMS Microbiol Ecol. 2011;78:497–510.

73. Bergquist PR, Kelly-Borges M. An evaluation of the genus *Tethya* (Porifera: Demospongiae: Hadromerjda) with descriptions of new species from the southwest Pacific. Beagle Rec Mus Art Galleries North Territ. 1991;8:37–72.

74. Hoshino T, Yilmaz LS, Noguera DR, Daims H, Wagner M. Quantification of target molecules needed to detect microorganisms by fluorescence in situ hybridization (FISH) and catalyzed reporter deposition-FISH. Appl Environ Microbiol. 2008;74:5068–77.

75. Moran NA, Wernegreen JJ. Lifestyle evolution in symbiotic bacteria: insights from genomics. Trends Ecol Evol. 2000;15:321–6.

76. Vacelet J, Donadey C. Electron microscope study of the association between some sponges and bacteria. J Exp Mar Bio Ecol. 1977;30:301–14.

77. Maldonado M. Intergenerational transmission of symbiotic bacteria in oviparous and viviparous demosponges, with emphasis on intracytoplasmically-compartmented bacterial types. J Mar Biol Assoc UK. 2007;87:1701–13.

78. Ilan M, Abelson A. The life of a sponge in a sandy lagoon. Biol Bull. 1995;189:363–9.

79. Borisov VB, Gennis RB, Hemp J, Verkhovsky MI. The cytochrome bd respiratory oxygen reductases. Biochim Biophys Acta Bioenerg. 2011;1807:1398–413.

80. Lavy A, Keren R, Yahel G, Ilan M. Intermittent hypoxia and prolonged suboxia measured in situ in a marine sponge. Front Mar Sci. 2016;3:1–11.

81. Hoffmann F, Larsen O, Thiel V, Rapp HT, Pape T, Michaelis W, et al. An anaerobic world in sponges. Geomicrobiol J. 2005;22:1–10.

82. Moreno-Vivian C, Cabello C, Martinez-Luque M, Blasco R, Castillo F. Prokaryotic nitrate reduction: molecular properties and functional distinction among bacterial nitrate reductases. J Bacteriol. 1999;181:6573–84.

83. Toffanin A, Wu Q, Maskus M, Casella S, Abruña HD, Shapleigh JP. Characterization of the gene encoding nitrite reductase and the physiological consequences of its expression in the non-denitrifying Rhizobium "hedysari" strain HCNT1. Appl Environ Microbiol. 1996;62:4019–25.

84. Jones PM, George AM. The ABC transporter structure and mechanism: perspectives on recent research. Cell Mol Life Sci. 2004;61:682–99.

85. Mulligan C, Fischer M, Thomas GH. Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. FEMS Microbiol Rev. 2011;35:68–86.

86. Vastermark A, Wollwage S, Houle ME, Rio R, Saier MH. Expansion of the APC superfamily of secondary carriers. Proteins Struct Funct Bioinform. 2014;82:2797–811.

87. Pao SS, Paulsen IT, Saier MH. Major facilitator superfamily. Microbiol Mol Biol Rev. 1998;62:1–34.

88. Ziegler C, Bremer E, Krämer R. The BCCT family of carriers: from physiology to crystal structure. Mol Microbiol. 2010;78:13–34.

89. Andrade SLA, Einsle O. The Amt/Mep/Rh family of ammonium transport proteins (Review). Mol Membr Biol. 2007;24:357–65.

90. Kamke J, Sczyrba A, Ivanova N, Schwientek P, Rinke C, Mavromatis K, et al. Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. ISME J. 2013;7:2287–300.

91. Clifford EL, Hansell DA, Varela MM, Nieto-Cid M, Herndl GJ, Sintes E. Crustacean zooplankton release copious amounts of dissolved organic matter as taurine in the ocean. Limnol Oceanogr. 2017;62:2745–58.

92. Karimi E, Keller-Costa T, Slaby BM, Cox CJ, da Rocha UN, Hentschel U, et al. Genomic blueprints of sponge-prokaryote symbiosis are shared by low abundant and cultivatable Alphaproteobacteria. Sci Rep. 2019;9:1–15.

93. Dahl C, Franz B, Hensen D, Kesselheim A, Zigann R. Sulfite oxidation in the purple sulfur bacterium *Allochromatium vinosum*: Identification of SoeABC as a major player and relevance of SoxYZ in the process. Microbiology. 2013;159:2626–38.

94. Bardischewsky F, Quentmeier A, Friedrich CG. The flavoprotein SoxF functions in chemotrophic thiosulfate oxidation of Paracoccus pantotrophus in vivo and in vitro. FEMS Microbiol Lett. 2006;258:121–6.

95. Gregersen LH, Bryant DA, Frigaard NU. Mechanisms and evolution of oxidative sulfur metabolism in green sulfur bacteria. Front Microbiol. 2011;2:1–14.

96. Lavy A, Keren R, Yu K, Thomas BC, Alvarez-Cohen L, Banfield JF, et al. A novel Chromatiales bacterium is a potential sulfide oxidizer in multiple orders of marine sponges. Environ Microbiol. 2018;20:800–14.

97. Tian R-M, Wang Y, Bougouffa S, Gao ZM, Cai L, Bajic V, et al. Genomic analysis reveals versatile heterotrophic capacity of a potentially symbiotic sulfur-oxidizing bacterium in sponge. Environ Microbiol. 2014;16:3548–61.

98. Keller MD, Kiene RP, Matrai PA, Bellows WK. Production of glycine betaine and dimethylsulfoniopropionate in marine

phytoplankton. II. N-limited chemostat cultures. Mar Biol. 1999; 135:249–57.

99. Diaz MR, Visscher PT, Taylor BF. Metabolism of dimethylsulfoniopropionate and glycine betaine by a marine bacterium. FEMS Microbiol Lett. 1992;96:61–5.

100. Sun J, Steindler L, Thrash JC, Halsey KH, Smith DP, Carter AE, et al. One carbon metabolism in SAR11 pelagic marine bacteria. PLoS ONE. 2011;6:e23973.

101. Equar MY, Tani Y, Mihara H. Purification and properties of glycine oxidase from pseudomonas putida KT2440. J Nutr Sci Vitaminol. 2015;61:506–10.

102. Karimi E, Slaby BM, Soares AR, Blom J, Hentschel U, Costa R. Metagenomic binning reveals versatile nutrient cycling and distinct adaptive features in alphaproteobacterial symbionts of marine sponges. FEMS Microbiol Ecol. 2018;94:1–18.

103. Moreno-Pino M, Cristi A, Gillooly JF, Trefault N. Characterizing the microbiomes of Antarctic sponges: a functional metagenomic approach. Sci Rep. 2020;10:1–12.

104. Mayer J, Huhn T, Habeck M, Denger K, Hollemeyer K, Cook AM. 2,3-Dihydroxypropane-1-sulfonate degraded by *Cupriavidus pinatubonensis* JMP134: purification of dihydroxypropanesulfonate 3-dehydrogenase. Microbiology. 2010;156:1556–64.

105. Michael AJ. Polyamines in eukaryotes, bacteria, and archaea. J Biol Chem. 2016;291:14896–903.

106. Tsukamoto S, Kato H, Hirota H, Fusetani N. Pseudoceratidine: a new antifouling spermidine derivative from the marine sponge *Pseudoceratina purpurea*. Tetrahedron Lett. 1996;37:1439–40.

107. Igarashi K, Kashiwagi K. Polyamine transport in bacteria and yeast. Biochem J. 1999;344:633–42.

108. Shah P, Swiatlo E. A multifaceted role for polyamines in bacterial pathogens. Mol Microbiol. 2008;68:4–16.

109. Schneider BL, Reitzer L. Pathway and enzyme redundancy in putrescine catabolism in *Escherichia coli*. J Bacteriol. 2012; 194:4080–8.

110. Mou X, Sun S, Rayapati P, Moran MA. Genes for transport and metabolism of spermidine in *Ruegeria pomeroyi* DSS-3 and other marine bacteria. Aquat Microb Ecol. 2010; 58:311–21.

111. Dasu VV, Nakada Y, Ohnishi-Kameyama M, Kimura K, Itoh Y. Characterization and a role of *Pseudomonas aeruginosa* spermidine dehydrogenase in polyamine catabolism. Microbiology. 2006;152:2265–72.

112. Tofalo R, Cocchi S, Suzzi G. Polyamines and gut microbiota. Front Nutr. 2019;6:1–5.

113. Sharfstein ST, Keasling JD. Polyphosphate metabolism in *Escherichia coli*. Ann N Y Acad Sci. 1994;745:77–91.

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

115. Wang L, Yan J, Wise MJ, Liu Q, Asenso J, Huang Y, et al. Distribution patterns of polyphosphate metabolism pathway and its relationships with bacterial durability and virulence. Front Microbiol. 2018;9:1–10.

116. Romano S, Schulz-Vogt HN, González JM, Bondarev V. Phosphate limitation induces drastic physiological changes, virulence-related gene expression, and secondary metabolite production in *Pseudovibrio* sp. strain FO-BEG1. Appl Environ Microbiol. 2015;81:3518–28.

117. Bergquist PR. In: eLS (ed) Porifera (Sponges). John Wiley & Sons, Ltd. 2001.

118. Kovacs-Simon A, Titball RW, Michell SL. Lipoproteins of bacterial pathogens. Infect Immun. 2011;79:548–61.

119. Aoki S, Ito M, Iwasaki W. From β- To α-proteobacteria: The origin and evolution of rhizobial nodulation genes nodij. Mol Biol Evol. 2013;30:2494–508.

120. Darby AC, Cho NH, Fuxelius HH, Westberg J, Andersson SGE. Intracellular pathogens go extreme: genome evolution in the Rickettsiales. Trends Genet. 2007;23:511–20.

121. Kuo CH, Ochman H. Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. Biol Direct. 2009;4:35.