







Marine Sponges as *Chloroflexi* Hot Spots: Genomic Insights and High-Resolution Visualization of an Abundant and Diverse Symbiotic Clade

 Kristina Bayer,^a  Martin T. Jahn,^{a,b}  Beate M. Slaby,^a Lucas Moitinho-Silva,^c  Ute Hentschel^{a,d}

^aGEOMAR-Helmholtz Centre for Ocean Research, RD3-Marine Ecology, RU-Marine Microbiology, Kiel, Germany

^bUniversity of Wuerzburg, Imaging Core Facility at the Theodor Boveri Institute of Bioscience, Wuerzburg, Germany

^cCentre for Marine Bio-Innovation, University of New South Wales, Sydney, New South Wales, Australia

^dChristian-Albrechts University of Kiel, Kiel, Germany

ABSTRACT Members of the widespread bacterial phylum *Chloroflexi* can dominate high-microbial-abundance (HMA) sponge microbiomes. In the Sponge Microbiome Project, *Chloroflexi* sequences amounted to 20 to 30% of the total microbiome of certain HMA sponge genera with the classes/clades SAR202, *Caldilineae*, and *Anaerolineae* being the most prominent. We performed metagenomic and single-cell genomic analyses to elucidate the functional gene repertoire of *Chloroflexi* symbionts of *Aplysina aerophoba*. Eighteen draft genomes were reconstructed and placed into phylogenetic context of which six were investigated in detail. Common genomic features of *Chloroflexi* sponge symbionts were related to central energy and carbon converting pathways, amino acid and fatty acid metabolism, and respiration. Clade-specific metabolic features included a massively expanded genomic repertoire for carbohydrate degradation in *Anaerolineae* and *Caldilineae* genomes, but only amino acid utilization by SAR202. While *Anaerolineae* and *Caldilineae* import cofactors and vitamins, SAR202 genomes harbor genes encoding components involved in cofactor biosynthesis. A number of features relevant to symbiosis were further identified, including CRISPR-Cas systems, eukaryote-like repeat proteins, and secondary metabolite gene clusters. *Chloroflexi* symbionts were visualized in the sponge extracellular matrix at ultrastructural resolution by the fluorescence *in situ* hybridization-correlative light and electron microscopy (FISH-CLEM) method. Carbohydrate degradation potential was reported previously for “*Candidatus* Poribacteria” and SAUL, typical symbionts of HMA sponges, and we propose here that HMA sponge symbionts collectively engage in degradation of dissolved organic matter, both labile and recalcitrant. Thus, sponge microbes may not only provide nutrients to the sponge host, but they may also contribute to dissolved organic matter (DOM) recycling and primary productivity in reef ecosystems via a pathway termed the sponge loop.

IMPORTANCE *Chloroflexi* represent a widespread, yet enigmatic bacterial phylum with few cultivated members. We used metagenomic and single-cell genomic approaches to characterize the functional gene repertoire of *Chloroflexi* symbionts in marine sponges. The results of this study suggest clade-specific metabolic specialization and that *Chloroflexi* symbionts have the genomic potential for dissolved organic matter (DOM) degradation from seawater. Considering the abundance and dominance of sponges in many benthic environments, we predict that the role of sponge symbionts in biogeochemical cycles is larger than previously thought.

KEYWORDS *Chloroflexi*, DOM degradation, FISH-CLEM, metabolism, metagenomic binning, single-cell genomics, sponge symbiosis


Received 26 July 2018 **Accepted** 29 November 2018 **Published** 26 December 2018

Citation Bayer K, Jahn MT, Slaby BM, Moitinho-Silva L, Hentschel U. 2018. Marine sponges as *Chloroflexi* hot spots: genomic insights and high-resolution visualization of an abundant and diverse symbiotic clade. mSystems 3:e00150-18. <https://doi.org/10.1128/mSystems.00150-18>.

Editor Karen G. Lloyd, University of Tennessee at Knoxville

Copyright © 2018 Bayer et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Kristina Bayer, kbayer@geomar.de.

 Marine sponges as *Chloroflexi* hot spots: Genomic insights and high-resolution visualization of an abundant and diverse symbiotic clade. @GEOMAR_en @sponge_papers @SpongeLoop @Symbiosispapers #SpongeThursday

Sponges (*Porifera*) represent one of the oldest, still extant animal phyla. Fossil evidence shows their existence in the Precambrian long before the radiation of all other animal phyla (1, 2). Nowadays, sponges are globally distributed in all aquatic habitats from warm tropical reefs to the cold deep sea and are even present in freshwater lakes and streams (3). Sponges are increasingly recognized as important components of marine environments due to their immense filter-feeding capacities and consequent impacts upon coastal food webs and biogeochemical (e.g., carbon, nitrogen) cycles (4, 5). Many marine sponges contain dense and diverse microbial consortia within their extracellular mesohyl matrix. To date, 41 bacterial phyla (among them many candidate phyla) have been recorded from sponges, with recent amplicon sequencing studies suggesting up to 14,000 operational taxonomic units (OTUs) per sponge individual (6, 7). Sponges also constitute one of the most abundant natural sources of secondary metabolites, which are of commercial interest for the development of pharmaceuticals and new drugs (8) and are often produced by the microbial symbionts (9, 10).

Sponges can be classified into the so-called high-microbial-abundance (HMA) sponges harboring dense and diverse microbial consortia within their mesohyl tissues and the low-microbial-abundance (LMA) sponges containing microbial numbers on the order of those found in seawater (11–13). While HMA sponges are enriched in *Chloroflexi*, *Acidobacteria*, and “*Candidatus* Poribacteria,” the LMA sponges are dominated by *Gamma*- and *Betaproteobacteria* as well as *Cyanobacteria*, while *Chloroflexi* are typically absent. Differences have also been observed with respect to functional gene content (14), pumping rates (15), and exchange of carbon and nitrogen compounds (16). There is mounting evidence that HMA sponges are specialized to feed on dissolved organic matter (DOM), while the LMA sponges preferably feed on particulate organic matter (POM) (7, 17, 18). It is thus tempting to speculate that the symbiotic microbiota of HMA sponges is involved in DOM degradation, and indeed, the microbiomes analyzed so far encode a diverse repertoire for carbon metabolism pathways and transporters for low-molecular-weight compounds (10, 19–21). However, the precise fluxes and mechanisms how DOM and POM are taken up and processed within the sponge holobiont remain unknown. Recently, it was proposed that members of the phylum *Chloroflexi* are involved in recalcitrant DOM recycling in the water column (22, 23).

In the present study, we focused our metagenomic analyses on *Chloroflexi* as abundant and characteristic, yet understudied members of HMA sponge microbiota. The phylum *Chloroflexi* comprises taxonomically and physiologically highly diverse lineages that populate a wide range of habitats (24–27) including the deep sea (22), uranium-contaminated aquifers (28), and the human oral cavity and gut (29, 30). *Chloroflexi* metabolism is very diverse, ranging from anoxygenic photosynthesizers, obligate aerobic/anaerobic heterotrophs, thermophiles, halophiles, clades capable of reductive halogenation, and even predators with gliding motility. Because only a few *Chloroflexi* lineages have been cultivated (87) and because draft genomes are limited in number (22, 23, 32), the specific functions of *Chloroflexi* within the marine ecosystem but especially in the symbiont context remain largely unknown.

Chloroflexi are members of HMA sponge microbiota, with representatives of classes/clades SAR202, *Anaerolineae*, and *Caldilineae* being the most abundant (31). Visualization of *Chloroflexi* by fluorescence *in situ* hybridization (FISH) revealed bright and abundant signals (32, 33). Because *Chloroflexi* likely play an important role in the HMA sponge holobiont, we had the following aims: (i) to assess their relative abundances and distributions in diverse HMA sponge species by using the largest data set currently available (Earth Microbiome Project [EMP] sponge microbiome [31]), (ii) to provide their phylogenetic affiliation, (iii) to characterize the functional gene repertoire with a particular focus on carbon degradation and symbiotic lifestyle, and (iv) to visualize *Chloroflexi* in mesohyl tissues at ultrastructural resolution by FISH-correlative light and electron microscopy (CLEM) methodology. We applied a broad range of state-of-the-art methods, from global sponge surveys to single-cell genomics and microscopy, to acquire comprehensive insights into the lifestyle of *Chloroflexi* symbionts.

RESULTS AND DISCUSSION

Chloroflexi abundance in HMA sponges. Recently, members of the phylum *Chloroflexi* were shown to be present in much higher abundance and diversity in HMA sponges than in LMA sponges, which is why they were termed indicator species for HMA sponges (12). Here, we provide further details for the presence and abundance of *Chloroflexi* in HMA sponges (12) (Fig. 1; see also Tables S2A and S2B in the supplemental material). The recently compiled Sponge Microbiome Project (6, 31) was used as a reference database. In these 63 investigated sponge species, *Chloroflexi* abundances ranged from $4.39\% \pm 3.02\%$ (*Chondrilla caribensis*) to $31.89\% \pm 5.27\%$ (*Aplysina* spp.) (Fig. 1, right panel, and Table S2A). With respect to the *Chloroflexi* classes, the SAR202 clade was the most abundant, contributing on average to $47.74\% \pm 22.00\%$ of the phylum total abundance (Fig. 1, top panel, and Table S2B). Members of the classes *Caldilineae* ($22.35\% \pm 17.93\%$) and *Anaerolineae* ($11.64\% \pm 12.30\%$) were also abundant in some sponges but not others. Unclassified OTUs at the class level represented $14.50\% \pm 10.77\%$ of *Chloroflexi* sequences, indicating that there is phylogenetic novelty still to be discovered. Despite some variability (Fig. 1, heatmap), the classes/clades SAR202, *Caldilineae*, and *Anaerolineae* as well as diverse hitherto unclassified OTUs dominated the *Chloroflexi* population in the HMA sponges. The remaining classes/clades amounted to 3.78% of total phylum abundance. At the OTU level, there was a significant effect of geographical location [$F(15, 725) = 15.9, P = 0.001$] and of sponge taxonomy [$F(9, 725) = 18.6, P = 0.001$] on beta diversity (Table S2C). For example, a Mediterranean cluster and several Caribbean clusters become visible as shown by Bray-Curtis cluster analysis. In addition, a host taxonomic signature was revealed for example for the sponge genera *Aplysina* and *Agelas*. However, there were also exceptions (i.e., *Neopetrosia* species from the Caribbean did not cluster together, and *Stelletta maori* from New Zealand fell into the Mediterranean cluster). Altogether, this analysis showed the high abundance and consistency of the three above-mentioned main clades of *Chloroflexi* in HMA sponges and revealed a geographic and host taxonomic signature for the *Chloroflexi* community.

Phylogeny of *Chloroflexi* metagenome bins and single amplified genomes. In total, 260 single amplified genomes (SAGs) were screened for the presence of the phylogenetic marker. A total of 125 16S rRNA genes were identified in 112 SAGs, of which some were duplicates per well. Sequencing revealed that 39 genes (31.2%) were affiliated with the phylum *Chloroflexi* (33). We randomly chose 13 SAGs for genome sequencing. Phylogenetic analysis of 16S rRNA genes revealed that one SAG (3D), four SAGs (1B, 1G, 1H, and 4H), and eight SAGs (2D, 3B, 3H, 4A, 5H, 6B, 6C, and 6F) belonged to the classes/clades SAR202, *Anaerolineae*, and *Caldilineae*, respectively, forming a well-supported sequence cluster (bootstrap, 100) with other sponge-derived sequences (Fig. S2). The binning of metagenomic sequence data (21) resulted in additional five high-quality bins (Table 1). The only metagenome bin containing a 16S rRNA gene (S156) belonged to SAR202. The SAR202 sequences formed a well-supported cluster (bootstrap value, 98) with other sponge-derived 16S rRNA gene sequences (Fig. S1). To elucidate the clade affiliation of the remaining four metagenome bins lacking 16S rRNA genes of appropriate length, a concatenated genome tree based on nine ribosomal genes was calculated (Fig. 2). One bin (A154) was affiliated with the class *Anaerolineae*, two metagenome bins were associated with the *Caldilineae* (C141 and C174), and two bins were associated with the SAR202 clade (S152 and S156) within the phylum *Chloroflexi*. The phylogenetic affiliation of metagenome bin S156 was congruent with the 16S rRNA gene analysis. SAGs were included in the protein-based phylogenetic analysis when they encoded at least three of the nine ribosomal genes. Due to the lack of more-complete reference genomes from SAR202 microorganisms, the most complete one (ca. 25%, SAR202 cluster bacterium sp. strain SCGC AAA240-N13, IMG Gold Study ID [Gs0017605](https://www.ncbi.nlm.nih.gov/genbank/GS0017605) [22]) was included in this analysis, although only one ribosomal protein could be used for tree construction. Both analyses showed a stable phylogeny of all SAGs and metagenome bins to above-described classes or clades within the

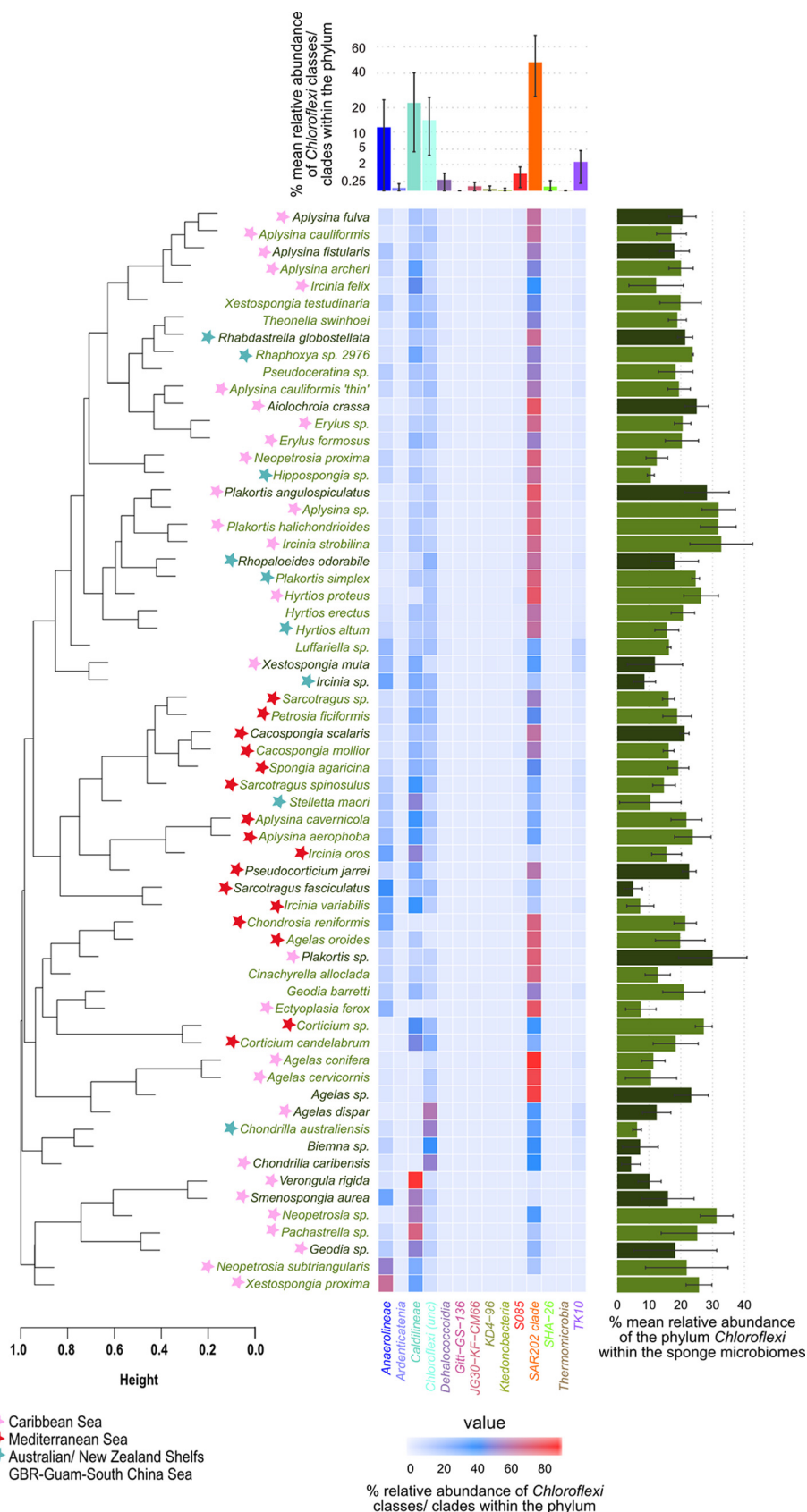


FIG 1 Heatmap showing the relative abundance of *Chloroflexi* classes/clades in 63 HMA sponges extracted from Earth Microbiome Project (EMP) data (31). The top panel shows the mean relative abundance of (Continued on next page)

TABLE 1 Genomic features overview of single amplified genomes (SAGs) and metagenome bins of *A. aerophoba* associated *Chloroflexi* and closest relative reference genomes analyzed in this study

Class/clade and SAG/bin ^a	Taxon ID	Genome size (Mbp)	No. of scaffolds	N50 (kbp)	GC (%)	Contamination (CheckM)	% completeness estimation (IMG)	No. of genes				
								Total	CDS	RNA	tRNA	Without function
<i>Anaerolineae</i>												
SAG 1B	2617270794	2.78	325	46.4	58.9	8.64	55.8	2,714	2,679	35	24	639
SAG 1G	2617270795	1.69	307	22.6	58.1	0.91	32.0	1,694	1,676	18	14	438
SAG 1H	2617270796	1.73	352	16.6	59.1	0.00	31.9	1,774	1,752	22	17	458
SAG 4H	2617270812	0.16	51	4.0	58.4	0.00	0.0	175	168	7	4	43
A154*	2619619053	3.73	107	63.9	59.3	1.82	91.8	3,358	3,307	51	47	613
<i>Caldilineae</i>												
SAG 2D	2617270806	3.51	489	17.8	58.4	0.91	65.5	3,207	3,177	30	24	855
SAG 3B	2617270807	2.02	310	15.3	59.2	0.00	38.7	1,844	1,823	21	17	444
SAG 3H	2617270810	2.47	290	31.1	58.9	0.00	50.9	2,271	2,242	29	24	641
SAG 4A	2617270811	3.35	437	31.3	59.4	0.00	64.8	3,024	2,994	30	26	777
SAG 5H	2617270814	1.26	183	17.5	58.4	0.00	17.2	1,132	1,112	20	17	288
SAG 6B	2617270816	4.25	685	14.6	58.7	3.18	66.8	3,943	3,901	42	35	1,154
SAG 6C	2617270818	2.75	414	18.6	58.5	1.82	45.9	2,579	2,544	35	28	703
SAG 6F	2617270820	3.66	839	8.4	58.4	0.91	56.5	3,594	3,561	33	25	1,063
C141*	2619619051	4.59	507	13.3	63.0	0.91	90.9	4,288	4,236	52	46	1,246
C174*	2619619055	6.36	647	13.0	58.4	10.20	96.3	5,662	5,601	61	55	1,472
SAR202												
SAG 3D	2617270809	0.58	106	22.8	60.5	0.00	25.4	622	608	14	13	145
S152*	2619619052	5.03	890	6.9	56.9	16.31	91.1	5,448	5,378	70	60	1,938
S156*	2619619054	3.35	334	15.2	65.6	2.97	98.0	3,463	3,402	61	52	964

^aIMG Gold Study IDs are [Gs0114494](#) and [Gs0099546](#) (marked with an asterisk). The letters of the bins reflect the phylogenetic identity of the bin (A for *Anaerolineae*, C for *Caldilineae*, and S for SAR202). The gray shaded bins/SAGs were used for further detailed metabolic analysis.

phylum *Chloroflexi*. All three classes/clades were visualized in the *A. aerophoba* sponge mesohyl matrix by fluorescence *in situ* cohybridization (FISH) on ultrathin tissue sections using class/clade-specific probes. The *Chloroflexi* cell signal was abundant as judged by the stained versus unstained bacterial signal, and SAR202 (green) seemed to dominate over either *Anaerolineae* (red) or *Caldilineae* (orange). The cells were metabolically active as judged by the brightness of the FISH probe (Fig. 3). These visual observations are consistent with the relative abundance of the phylum *Chloroflexi* within sponge microbiomes (24% ± 6% within *A. aerophoba* sponge microbiome; Fig. 1 and Table S2A) and the representation of *Chloroflexi* cells among SAGs (31%).

General description of genomes. The final genome assembly sizes for the sponge-associated *Chloroflexi* single cells (SAGs) ranged from 0.16 to 4.25 Mbp, representing up to 66.85% of genome completeness derived from IMG-based estimations (Table 1). The guanine-cytosine (GC) content ranged from 58.08 to 59.32%, 58.36 to 62.98%, and 56.93 to 65.59% for *Anaerolineae*, *Caldilineae*, and SAR202, respectively. The numbers of identified genes were highly variable, ranging from 3,358 genes for *Anaerolineae* genome bin A154 to 5,448 for the SAR202 bin S152 and 5,662 genes for the *Caldilineae* bin C174 (Table 1). The five metagenome bins (two for *Caldilineae* [C141 and C174], two for SAR202 [S152 and S156], and one for *Anaerolineae* [A154]) which had >90% coverage were chosen for detailed metabolic analysis and inner-phylum comparison. The letters of the bins were chosen to reflect their phylogenetic identity (A for *Anaerolineae*, C for *Caldilineae*, and S for SAR202). Additionally, the most complete *Anaerolineae* SAG 1B (55.76% genome completeness estimation) was included in the

FIG 1 Legend (Continued)

Chloroflexi classes/clades in all sponges (means ± standard deviations [error bars]). The right panel displays the mean relative abundance of the phylum *Chloroflexi* in predicted (light green names and medium green bars) and classified (dark green names and bars) HMA sponges (means ± standard deviations) determined by machine learning (12). Results of cluster analysis based on Bray-Curtis dissimilarities on mean relative abundances of OTUs within the phylum are presented on the left side. Sponges are marked with stars when all species samples came from one of the three major sample locations.

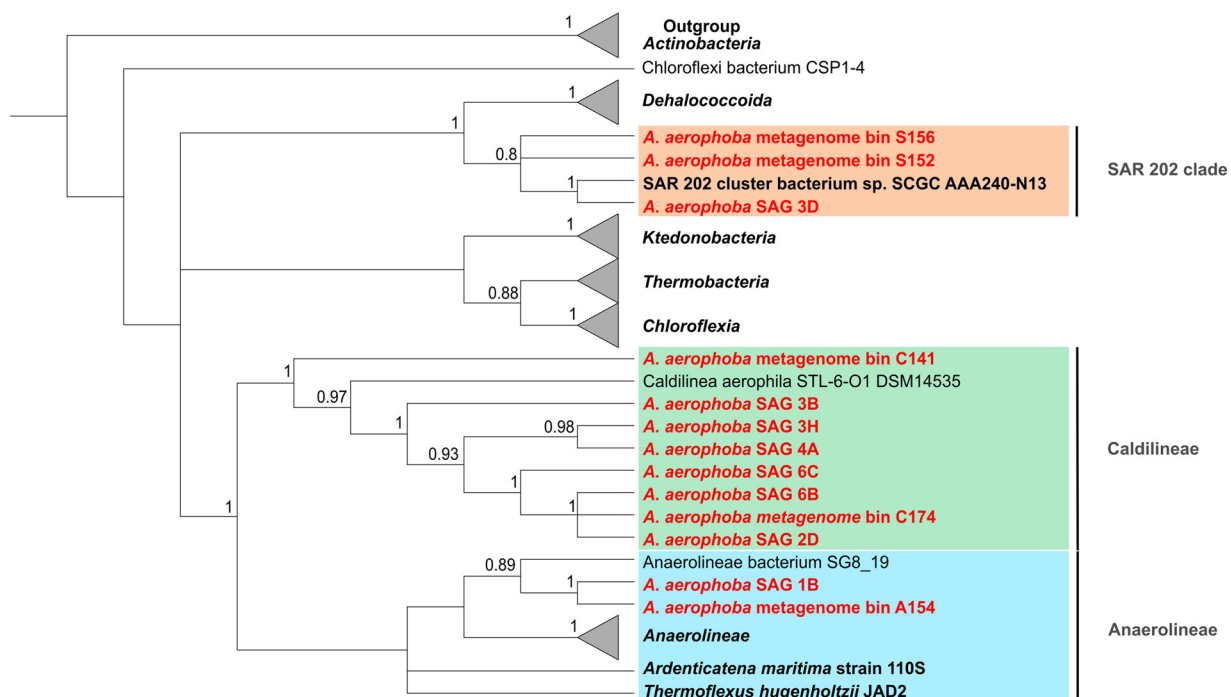


FIG 2 Concatenated protein tree. Maximum likelihood phylogenetic analysis of *Chloroflexi* metagenome bins and SAGs (in red) from 1,914 positions of 60 sequences using ribosomal proteins. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown. The initial tree for the heuristic search was obtained automatically by applying neighbor-joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model. Multiple sequences are included in the collapsed branches representing *Chloroflexi* classes/clades (bold). To root the tree, three representative genomes from the phylum *Actinobacteria* were used. Reference genomes with accession numbers can be found in Table S1 in the supplemental material.

analysis. Due to the lack of complete genomes of marine representatives of the *Anaerolineae*, *Caldilineae*, and SAR202 available at the time of analysis, sponge-derived *Chloroflexi* genomes could not be fully assessed. Taking the pitfalls inherent to metagenome sequencing into account (i.e., fragmented assemblies, unresolved ambiguities),

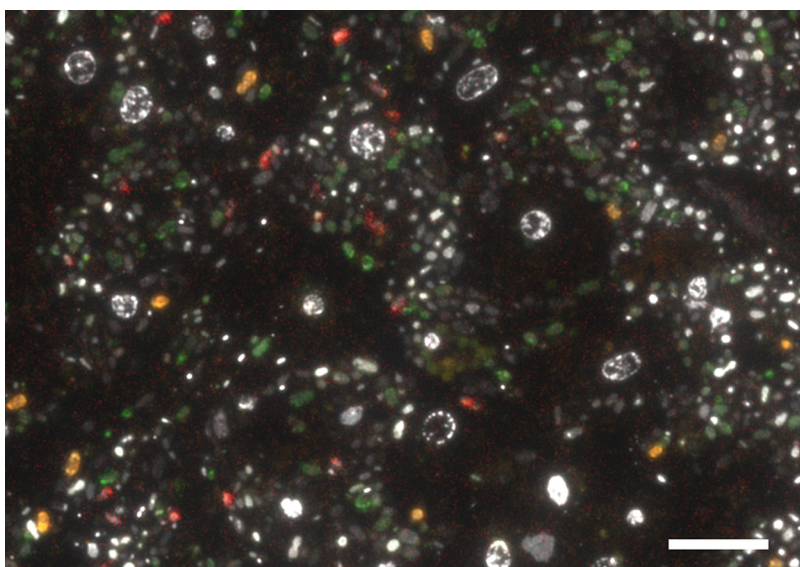


FIG 3 Distribution of *Chloroflexi* clades in *Aplysina aerophoba* mesohyl using fluorescence *in situ* hybridization (FISH). The image shows the overlay of all probes. SAR202 cells are displayed in green, *Caldilineae* cells in orange, and *Anaerolineae* cells in red. The nucleotide stain DAPI (white/gray) served as a reference for the localization of unstained cells. Bar, 10 μ m.

we have opted for a presence/absence approach: a gene or enzyme was considered present when it was identified in both bins of the corresponding clade. For *Anaerolineae*, we consider an enzyme or gene present when identified in bin A154, and the SAG 1B was taken as additional support.

Central metabolism of sponge-associated *Chloroflexi*. Metabolic reconstruction suggests that *Chloroflexi* are aerobic and heterotrophic bacteria including glycolysis, tricarboxylic acid (TCA) cycle, pentose phosphate pathway (PPP) (see Text S1 and Fig. S3A to C for details), and the respiratory chain as energy-producing pathways in all three clades.

With respect to autotrophic carbon fixation, the reductive citrate acid cycle (Arnon-Buchanan cycle), which is largely present, and the Wood-Ljungdahl pathway, which was partially identified (Fig. S2D) might be functional in times when sponges are not pumping and the mesohyl turns anoxic (34). Genes encoding components involved in ammonia import and assimilation are carried on all investigated genomes, but SAR202 and *Caldilineae* have additional genes for glutamate synthesis from glutamine and directly from ammonia. The transport of nitrite (and possibly also nitrate) is encoded on all investigated genomes, while the reduction to ammonia is encoded only by SAR202 (Fig. S3E). The incorporation of sulfur (with, e.g., thiosulfate as donor) into S-containing amino acids might be possible in all clades, whereas the assimilatory reduction of sulfate is restricted to *Anaerolineae* and *Caldilineae* genomes (Fig. S3F).

These processes additionally provide precursors for further metabolic pathways such as biosynthesis of purines and pyrimidines, amino acids, and cofactors, or structural compounds. Machinery for transcription, translation, and purine and pyrimidine metabolism are largely present. Fatty acid (FA) biosynthesis and degradation pathways were detected in all six genomes. All genomes encode a high number of different ABC transporters compared to genomes of free-living bacteria (22, 35) to supplement for nutrition and cell growth-related compounds [including oligopeptides, phosphate, L- and branched-chain amino acids, minerals such as iron(III) and molybdate, metal ions such as zinc, manganese, and iron(II)]. Additionally, all six genomes largely encode enzymes needed for biosynthesis of most amino acids (Text S1). We could not identify any of the typical phosphotransferase systems, as was the case for *Ca. Poribacteria* described previously (19).

We found genomic potential for aromatic degradation in *Chloroflexi* genomes, but pathways remain incomplete (Text S1). Several genes encoding components involved in phenylpropionate and cinnamate degradation, terephthalate degradation, catechol degradation, and xylene degradation were identified in *Chloroflexi* genomes. Also, genes encoding enzymes involved in ring cleavage by Baeyer-Villinger oxidation and beta oxidation as well as ring-hydroxylating dioxygenases and isomerases were identified which could be involved in degradation of aromatic compounds possibly synthesized by the sponge host or other microbes. This finding is interesting in the context that many sponge species contain secondary metabolites that often contain aromatic ring structures that serve as a defense strategy against predators and biofouling (36). Sponge symbionts may be able to degrade such substances, enabling them to a life within sponge hosts. These findings fit with the potential degradation of organic compounds which was suggested for *Chloroflexi* bacteria of the class SAR202 by (meta)genomic studies (22, 23, 35). The highest (20 to 30% relative to the total microbiome) and most consistent presence of *Chloroflexi* within a sponge genus was found in the sponge genera *Plakortis*, *Agelas* (with the exception of *Agelas dispar*), and *Aplysina* and sister taxon *Aiolochoira*. Interestingly, all of them contain characteristic natural products with aromatic ring structures. It is therefore tempting to speculate that the presence and abundance of *Chloroflexi* and especially SAR202 are shaped, at least to some extent, by the natural products present in the corresponding host sponges.

With respect to cell wall structure, the *Anaerolineae* and *Caldilineae* genomes carry the gene repertoire for peptidoglycan biosynthesis. The noticeable lack of peptidoglycan biosynthesis genes in the SAR202 genomes (Text S1) is consistent with previous

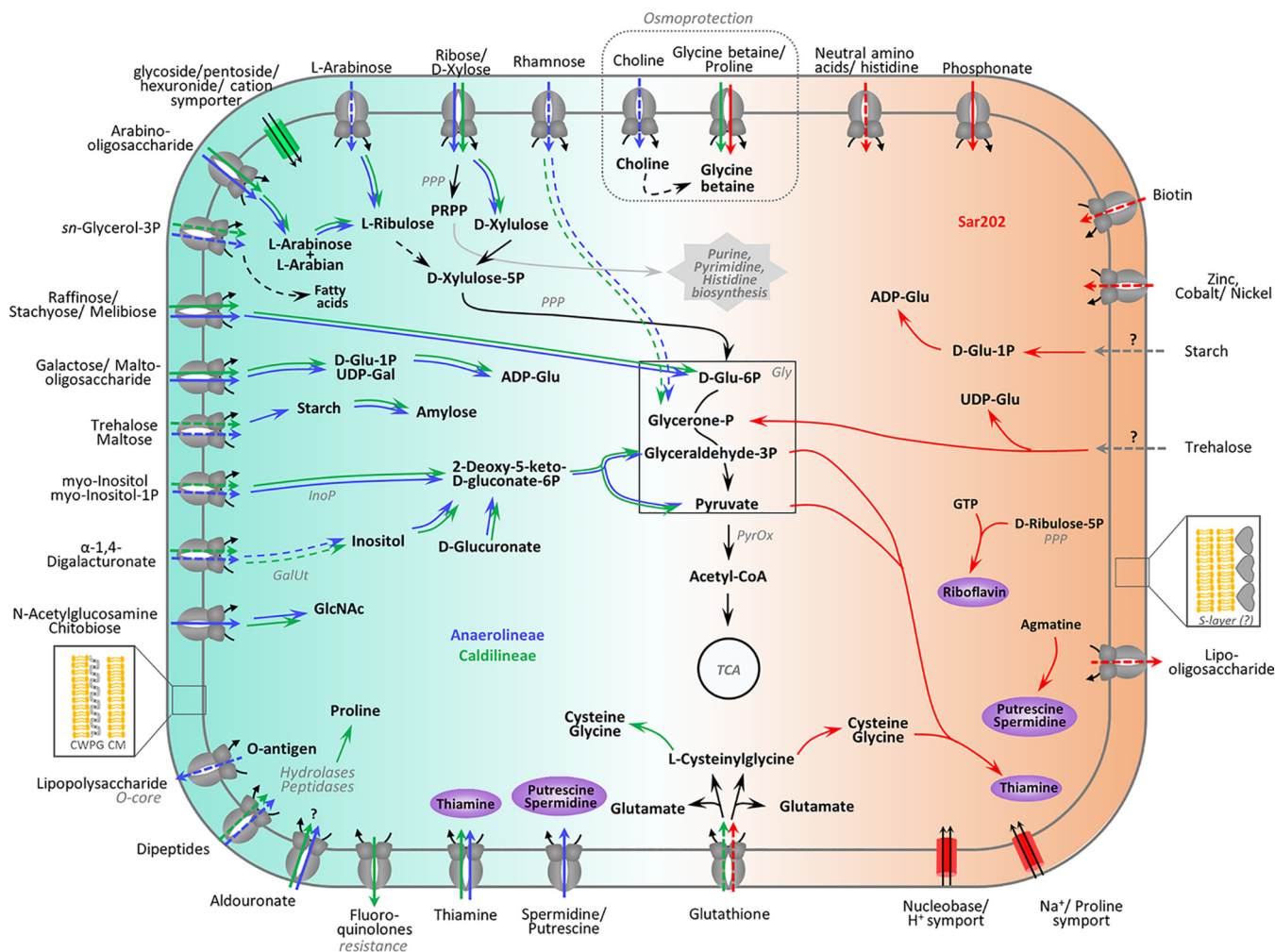


FIG 4 Summarized metabolic features which were found only in *Anaerolineae* and *Caldilineae* (left side, blue and green arrows, respectively) or in SAR202 genomes (right side, red arrows). The central metabolic pathways (glycolysis, TCA cycle, purine, pyrimidine histidine biosynthesis) located in the middle of the figure are general features found in all genomes. Lines are dashed when pathways or transporter could not be annotated completely (single enzymes of the pathway or single genes from the transporter were missing) or could not be annotated in both genomes of one clade. Gray dashed arrows indicate that those transporters were not identified.

analyses of three *Chloroflexi* genomes derived from uranium-contaminated aquifers (28) and hyperoxic zones from the Gulf of Mexico (37). Additionally, consistent with previous observations (38), none of the six genomes contained flagellar or chemotaxis genes.

Metabolic specialization: extensive carbohydrate uptake and degradation in *Anaerolineae* and *Caldilineae*. The following features are metabolic specialties of *Anaerolineae* and *Caldilineae* and appear to be missing in SAR202, unless otherwise mentioned (Fig. 4). We found a number of ABC transporters for the import of diverse mono-saccharides (ribose/xylose, inositol, glycerol-3-phosphate [glycerol-3P], and rhamnose) and oligosaccharides (sorbitol, raffinose/stachyose/melibiose, maltose, *N*-acetylglucosamine, and arabinosaccharide) into *Anaerolineae* and *Caldilineae* cells. Xylose can be processed to xylulose which may enter the pentose phosphate cycle finally leading into glycolysis. Ribose can be converted to 5-phosphoribosyl 1-pyrophosphate (PRPP), which is a precursor for the biosynthesis of the amino acid histidine or it may fuel purine and pyrimidine synthesis (Text S1). Both groups may also be able to import glycerol-3P, which is a phosphoric ester of glycerol (a component of glycerophospholipids) which can be converted to fatty acids. Additionally, we found evidence for arabinose and rhamnose import and degradation; however, annotation was incomplete (Text S1).

Arabinooligosaccharides (such as α -L-arabinofuranosides, α -L-arabinans, arabinoxylans, and arabinogalactans) result from degradation of plant-like cell material entering

Downloaded from <http://msystems.asm.org/> on April 18, 2019 by guest

the sponge by filtration. These substances may be imported by the almost completely annotated AraNPQ and MsmX transporters and be utilized to L-arabian and L-arabinose by the enzyme α -N-arabinofuranosidase (EC 3.2.1.55, GH3). The enzyme L-arabinose isomerase (EC 5.3.1.4, AraA) is present in all four genomes of *Anaerolineae* and *Caldilineae* and converts L-arabinose to L-ribose, which can further be converted by reactions of PPP to glucose-6P suitable for entering glycolysis. Additionally, other oligosaccharides such as stachyose, raffinose, melibiose, and galactose can be imported and used in central metabolism (Fig. 4 and Text S1).

The utilization of myo-inositol as a carbon source and possibly as a regulatory agent was hypothesized previously for sponge-associated *Ca. Poribacteria* (19). Similarly, sponge-associated *Anaerolineae* and *Caldilineae* genes encode almost all the components involved in the inositol degradation pathway (Text S1). myo-Inositol is likely degraded to glyceraldehyde-3-phosphate and acetyl-CoA, which are further used in the central metabolism. Inositol phosphates are found as part of eukaryotic and archaeal cell wall components (39). Phosphorylated inositol is a precursor for several lipid molecules, including sphingolipids, ceramides, and glycosylphosphatidylinositol anchors (40), as well as many stress-protective solutes of eukaryotes (39), and it might be part of the signal transduction in sponges (41). Therefore, the sponge itself or eukaryotic microorganisms can probably provide inositol as a carbon source or regulatory agent for the microbial symbionts.

Uronic acids are sugar acids that can be found in biopolymers of plants, animals, and bacteria (42, 43) and are known to occur in glycosaminoglycans (GAGs). GAGs in sponges are mainly composed of fucose, glucuronic acid (glucuronate), mannose, galactose, N-acetylglucosamine, and sulfate (44–46). Genes encoding enzymes involved in degradation of uronic acids were found in *Anaerolineae* and *Caldilineae* genomes. The possibility of galacturonate and glucuronate catabolism is supported by the conversion of 2-dehydro-3-deoxy-D-glucuronate by the enzymes glucuronate isomerase (EC 5.3.1.12), tagaturonate reductase (EC 1.1.1.58), and altronate hydrolase (EC 4.2.1.7). Furthermore, the presence of genes encoding oligogalacturonide lyase (EC 4.2.2.6), 2-deoxy-D-glucuronate 3-dehydrogenase (EC 1.1.1.125), and 2-dehydro-3-desoxy-D-glucokinase (EC 2.7.1.45) supports possible 4(4- α -D-gluc-4-enuronosyl)-D-galacturonate degradation activity. The products could then enter the Entner-Doudoroff (ED) pathway via 2-dehydro-3-desoxyphosphogluconate aldolase (EC 4.1.2.14). Uronic acid degradation could principally be connected to the inositol degradation pathway via D-galacturonate even though additional genome evidence, such as genes encoding the enzyme inositol oxidase (EC 1.13.99.1), remain wanting (Fig. 4 and Text S1). A number of transporters for N-acetylglucosamine, digalacturonate, mannose, and galactose (Fig. 4 and Text S1) were identified in *Anaerolineae* and *Caldilineae* genomes. Digalacturonate can be utilized by the uronic acid degradation pathway (Text S1), and N-acetylglucosamine can be used directly in amino sugar and nucleotide sugar synthesis. The presence of uronic acid degradation pathways provides strong support that *Anaerolineae* and *Caldilineae*, similar to the previously described *Ca. Poribacteria*, degrade glycosaminoglycan chains of proteoglycans, which are important components of the sponge host matrix (19). In that line, *Anaerolineae* and *Caldilineae* genomes were enriched in arylsulfatases A (Fig. S4A), which are thought to be involved in metabolization of sulfated polysaccharides from the sponge extracellular matrix (19, 21) and in the heterotrophic ability of symbionts to use sponge components for nutritional purposes.

Expanded carbohydrate-active enzyme (CAZymes) repertoire in *Caldilineae* and *Anaerolineae*. In order to search for CAZymes, we screened the *Chloroflexi* genome data against dbCAN (47) and classified the enzymes according to the CAZy database (48). Most *Chloroflexi* hits were against glycosyl hydrolases (GH), glycosyl-transferases (GT), and carbohydrate-binding modules (CBM). Consistent with the above-described metabolic specializations, these enzyme classes were present in larger amounts in *Caldilineae* and *Anaerolineae* than in SAR202 (Fig. S4B and Table S3). Altogether, 40 GH families were identified in all *Chloroflexi* genomes (Table S3). Glycosyl hydrolase family 109 was the most abundant family of GHs and was identified in all six

genomes. GH109 family proteins are predicted as α -*N*-acetylgalactosaminidases (EC 3.2.1.49) with putative substrates such as glycolipids, glycopeptides, and glycoproteins, all of which are common constituents of sponge mesohyl as well as dissolved organic matter from seawater. The GH74 family is the second most abundant and is also present in all six genomes. These appear to be xyloglucan-hydrolyzing enzymes, that act on β -1,4 linkages and might help degrade various oligo- and polysaccharides. The previously reported glycosylhydrolases GH33 and GH32 (19, 20) were the third most abundant, but they were restricted to *Caldilineae* bin C174. This enzyme family is annotated as sialidase (EC 3.2.1.18), capable of hydrolyzing glycosidic linkages of terminal sialic acid residues, which are present in sponge mesohyl (49). Altogether, 17 glycosyltransferases were identified on *Chloroflexi* genomes with families GT2, GT4, and GT83 being the most abundant. Among the 11 CBM families identified on *Chloroflexi* genomes, CBM50 was the most abundant, but it was restricted to *Caldilineae* and *Anaerolineae*. CMB50 modules, also known as LysM domains, attach to various GH enzymes which are involved in the cleavage of chitin or peptidoglycan. The numbers of carbohydrate-active enzymes on *Chloroflexi* symbiont genomes reflect their extensive potential to degrade complex carbohydrates as reported previously for *Ca. Poribacteria* and the sponge-associated unidentified lineage SAUL (19, 20).

Metabolic specialization: cofactor biosynthesis in SAR202 genomes. There is mounting evidence that vitamins and cofactors produced by diverse symbiont lineages could be beneficial to the sponge host (50–53). Parallel transcriptional activity profiling of the symbionts and the sponge showed that the symbionts had the capacity for vitamin B biosynthesis, whereas the host transcripts displayed the capacity for vitamin catabolism (54). It is thus tempting to speculate that the sponges' nutrition is augmented by symbiont-derived vitamins and cofactors. In the present study, at least two biosynthetic pathways for cofactor biosynthesis were identified on SAR202 genomes, which were absent in *Anaerolineae* and *Caldilineae* (Fig. 5). Thiamine is an essential cofactor which is involved in central metabolism. The biosynthesis of the biologically active form thiamine diphosphate (TPP) from L-cysteine, glycine, pyruvate, and glyceraldehyde-3P is encoded on the SAR202 genomes (Fig. 4 and 5A). Although the pathway is incomplete, the data strongly suggest that the synthesis of TPP is restricted to SAR202 bacteria. Instead, *Anaerolineae* and *Caldilineae* appear to import thiamine via an ABC transporter (TbpA, ThiPQ) and convert it to TPP by using thiamine pyrophosphokinase (EC 2.7.6.2). Additionally, thiamine synthesis might be a symbiosis-related feature, since free-living SAR202 members lack the synthesis ability (22, 37).

Second, riboflavin (vitamin B₂) is required by enzymes and proteins to perform certain physiological functions. Specifically, the active forms, flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD), serve as cofactors for a variety of flavoprotein enzyme reactions. Most genomes carry genes encoding the enzymes FMN adenylyltransferase (EC 2.7.7.2) and FAD riboflavin kinase (EC 2.7.1.26) which activate riboflavin into FMN. However, only the SAR202 genomes contain genes encoding the riboflavin biosynthesis enzymes which rely on GTP and ribulose-5P (Fig. 4 and 5B). Both substrates can be provided by pathways of the central metabolism (purine metabolism and PPP). Only one free-living SAR202 bacterium (unclassified *Chloroflexi* bin 43) (37) has genes encoding components involved in riboflavin synthesis.

Potential for degradation of recalcitrant DOM in SAR202. The uptake of the amino acid L-Asp was shown before in subtropical Atlantic waters (35), and the possible participation of deep sea SAR202 bacteria in degradation of recalcitrant or refractory DOM was recently postulated by Landry et al. (22). Even though the exact composition of DOM in the world's oceans remains to be elucidated, refractory DOM is an important component of the global carbon budget in terms of sheer mass. Landry et al. (22) argue that SAR202 genomes have an expanded repertoire of oxidative enzymes that may help in the oxidation of recalcitrant compounds. Interestingly, some of the described enzymes were also found to be enriched in SAR202 symbionts of sponges (Table S4). Among them are genes encoding proteins from the CaiB/BaiF family as well as related

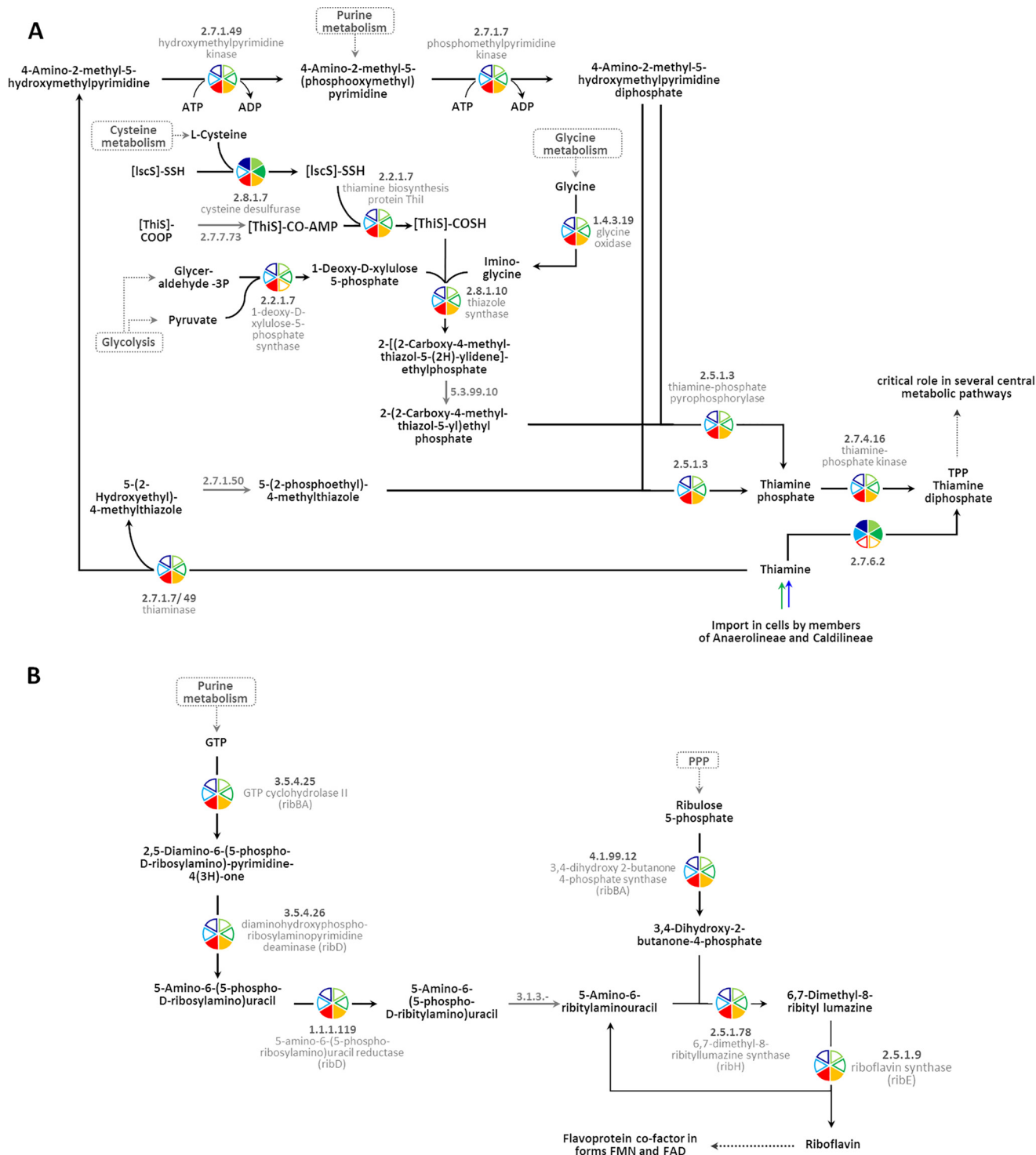


FIG 5 Pathway for synthesis of thiamine (A) and riboflavin (B) in both SAR202 genomes. (A) Members of classes *Anaerolineae* and *Caldilineae* carry genes encoding components involved in the import of thiamine (see also Fig. 4). (B) The conversion of riboflavin into the biologically active forms (flavin mononucleotide [FMN] and flavin adenine dinucleotide [FAD]) was encoded by genes in the genomes of all three classes (filled pies compared to empty pies). The colors represent the genomes. The numbers are KEGG identifiers. PPP, pentose phosphate pathway. Gray arrows and numbers indicate unidentified enzymes.

family III transferases. While the enrichment of CaiB was previously interpreted as carnitine being a carbon and nitrogen source for sponge symbionts (21), an alternative explanation may be that it serves to funnel substrates into degradation pathways without consumption of energy by shuffling CoA, thus generating free electrons (Text

S1). Even though the precise function of CaiB/BaiF family proteins cannot be elucidated at this time, the enrichment in SAR202 genomes is noteworthy. Further, a total of 53 flavin-dependent, class C oxidoreductases of the luciferase family (flavin mononucleotide monooxygenases [FMNOs], COG2141) which include alkanesulfonate monooxygenase SsuD and methylene tetrahydromethanopterin reductase, were present and enriched in SAR202 genomes compared to the genomes of *Anaerolineae* and *Caldilineae* bacteria (Table S4). These enzymes are proposed to participate in the oxidation of (long-chain) aldehydes to carboxylic acids and/or cleavage of carbon-sulfur bonds in a variety of sulfonated alkanes (55). The SAR202 genomes contained 22 genes encoding short-chain alcohol dehydrogenases (COG0300) which might be involved in canalization of ketone body derivate release. Some of these genes from bin S152 showed homologies to cyclopentanol and 3- α (or 20- β)-hydroxysteroid dehydrogenases which convert alicyclic-bound alcohol groups to ketones (22, 56). The combination of the enzymes described above could allow sponge-associated *Chloroflexi* to convert recalcitrant alicyclic ring structures to more labile carboxylic acid, as proposed recently for SAR202 bacteria from deep sea (22).

Additionally, a number of genes encoding oxidative enzymes were identified on the *Chloroflexi* genomes but were not enriched in SAR202 (Table S4). These enzymes include a 2-oxoglutarate:ferredoxin oxidoreductase (EC 1.2.7.11) which oxidizes acetyl-CoA, carbon monoxide dehydrogenase (EC 1.2.99.2) which might allow the bacteria to oxidize CO as described for some members of the Ktedonobacteria (57), CO- or xanthine dehydrogenases (COG1529) which are possibly involved in oxidation of a broad range of complex substrates (22), choline dehydrogenase (EC 1.1.99.1) being possibly involved in the oxidation of alcohols to aldehydes, sarcosine oxidase (EC 1.5.3.1), the serine hydroxymethyltransferase (2.1.2.1) with predicted function in choline degradation, formaldehyde dehydrogenase (EC 1.2.1.46), and subunits of formate dehydrogenase (EC 1.2.1.2/43) which oxidize formaldehyde and formate and might be involved in demethylation of various compounds. The overall presence and frequent enrichment of enzymes with oxidative capacity in SAR202 would be consistent with gene functions in degradation of recalcitrant DOM. However, owing to the sponges' existence in shallow-water sun-lit benthic environments, it remains unclear whether the sponge symbionts encounter recalcitrant DOM derived from seawater sources. Interestingly, Colatrisano et al. recently proposed the degradation of terrestrial DOM (tDOM) by members of the SAR202 clade (23). Shallow-water sponges and associated symbionts could be faced by tDOM via freshwater inflow in ocean waters they inhabit. Alternatively, and similar to other high-diversity microbiota, for example, of ant, ruminant, and human guts, the resident microbes were likely to specialize in certain substrates, thus promoting maximum nutrient exploitation and also securing their individual niche in the holobiont ecosystem.

Symbiosis-related features. Eukaryotic-like proteins (ELPs) seem to be a general genomic feature of sponge symbionts (20, 38, 50, 51, 58–60). Ankyrin (ANK), tetratricopeptide (TPR), and leucine-rich (LRR) repeat proteins are postulated to be involved in mediating host-microbe interactions (61, 62). Ankyrin and ankyrin repeat-containing proteins were detected in all six genomes of sponge symbionts (Fig. S4C and Table S5A) in higher numbers than in free-living bacteria (22, 37). It was recently proposed that the expression of sponge symbiont-derived ankyrin protein prevents phagocytosis by amoeba (63), and it is tempting to speculate that they protect the symbionts from digestion by the sponge archaeocytes *in vivo*. TPRs, possibly functioning as a module for protein-protein interaction involved in a variety of cellular functions, including those that participate in bacterial pathogenesis (64) were found in all six genomes. However, LRR genes were identified only in *Caldilineae* and SAR202 genomes (Fig. S4C and Table S5A). Many LRR proteins are involved in protein-ligand interactions; these interactions include plant immune response and the mammalian innate immune response (for a review, see reference 65), such as the detection of pathogen-associated molecular patterns by recogni-

TABLE 2 Genomic characteristics of the six genomes investigated in detail^a

Class/clade and SAG/bin ^b	Taxon ID	CRISPR		Secondary metabolite gene cluster			
		CRISPRfinder total (no. of repeats per spacer)	IMG total	ANK	Type 1 PKS	Terpene	Other
<i>Anaerolineae</i>							
SAG 1B	2617270794	-	-	1	-	-	-
A154*	2619619053	-	-	2	-	-	-
<i>Caldilineae</i>							
C141*	2619619051	7 (27, 24, 13, 7, 3, 30, 4)	9	9	2	-	-
C174*	2619619055	5 (21, 24, 8, 32, 27)	8	1	-	1	-
SAR202							
S152*	2619619052	5 (7, 34, 26, 7, 4)	9	5	3	4	1
S156*	2619619054	1 (13)	3	2	1	2	1

^aThe absolute numbers of CRISPR arrays defined by CRISPRfinder and IMG, the number of ankyrins and ankyrin repeat-containing proteins (ANK), as well as the number of secondary metabolite (antiSMASH) gene clusters per genome are shown. The IMG Gold Study IDs are [Gs0114494](https://doi.org/10.1093/igs/gaa011) and [Gs0099546](https://doi.org/10.1093/igs/gaa012). Secondary metabolite clusters were found using antiSMASH 3.0; the values are total numbers of genes per genome.

^bAsterisks on bins indicate extracted metagenome bins from IMG Gold Study ID [Gs0099546](https://doi.org/10.1093/igs/gaa012). The letters of the bins reflect the phylogenetic identity of the bin (A for *Anaerolineae*, C for *Caldilineae*, and S for SAR202).

tion receptors (66). Our findings are in good agreement with general patterns previously found in the metagenomes of sponge symbionts (38, 51), in enriched (mini)metagenomes of cyanobacterial sponge symbionts (59) and single amplified genomes from members of SAUL (20) and *Ca. Poribacteria* (58).

Another example of sponge-symbiont enriched features are the clustered, regularly interspaced, short, palindromic repeats (CRISPRs) and their associated proteins (Cas) that have recently been reported from the genomes of sponge symbionts (20, 38, 50, 51, 59). Here, the investigated *Caldilineae* and SAR202 genomes contained CRISPR-Cas systems, while the two most complete *Anaerolineae* genomes did not (Table 2). The presence of CRISPRs can be explained by the extensive filter-feeding activity of sponge hosts that result in high exposure of sponge symbionts to phages and other sources of free DNA from ambient seawater.

The synthesis of secondary metabolites is an important defense mechanism of sessile organisms such as sponges to protect themselves against predators or biofouling (36). Many of these compounds are in fact produced by the sponge microbiome (10, 36). In particular, genes for polyketide synthases (PKS), nonribosomal peptide synthetases (NRPS), and halogenases are regularly enriched in sponge symbionts, often with new structures and putatively novel activities (51, 67–72). Here, we assessed the genomic repertoire of sponge-associated *Chloroflexi* for secondary metabolism using antiSMASH (73). In both SAR202 genomes and in *Caldilineae* C141, we found up to three polyketide synthase (PKS) gene clusters, all of which showed homologies to the previously reported type I PKS gene cluster from other sponge symbionts. Additional gene clusters for the production of terpenes and other yet to be identified substances were identified in the two SAR202 genomes and in *Caldilineae* C174 (Table 2 and Table S5B). Both *Anaerolineae* genomes did not contain any gene clusters for the biosynthesis of secondary metabolites. While the exact functions of these gene clusters putatively involved in defense remain unknown, it appears that at least SAR202 bacteria and *Caldilineae* have the genomic repertoire for chemical defense within the sponge holobiont.

Ultrastructural identification of sponge-specific *Chloroflexi*. The correlation of probe-specific fluorescence with scanning electron microscopy (SEM) images allowed taxon-specific identification of *Chloroflexi* cells at ultrastructural resolution. Overall, distributions of all three *Chloroflexi* clades in *Aplysina aerophoba* indicate that SAR202 cells were more abundant than the other two *Chloroflexi* classes (Fig. 6). This is consistent with the relative abundances of *Chloroflexi* classes in HMA sponges extracted from EMP data (Fig. 1). Cells belonging to the SAR202 clade (green signals in Fig. 6A and

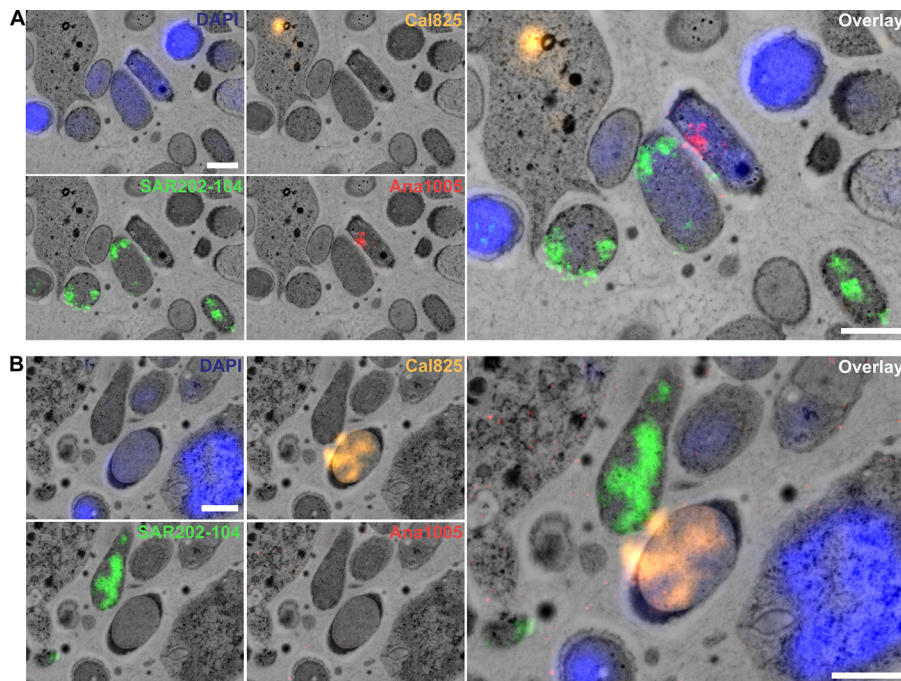


FIG 6 Visualization of sponge-associated *Chloroflexi* in *Aplysina aerophoba* mesohyl using FISH-CLEM. (A and B) SAR202 cells are displayed in green, *Anaerolineae* in red (A), and *Caldilineae* in orange (B). The nucleotide stain DAPI (blue) served as reference for the localization of unstained cells. In both panels, the picture on the right is the overlay of all probes and DAPI. Bars, 1 μm .

B) are generally rod shaped (0.8 μm by 1 to 2 μm) with a regular distribution of cell cytosol content. The *Anaerolineae*-specific probe targeted rod-shaped cells (0.8 by 2.0 μm) (red signal in Fig. 6A). The characteristic feature of *Caldilineae*-positive cells (ca. 1 by 2 μm) was the presence of electron-dense capsules or mucus-like structures located at the cell poles (orange signal in Fig. 6B). All cells that were stained positive with a corresponding FISH probe showed a consistent morphology, which was taken as a measure of probe specificity.

Conclusions. Owing to the lack of cultivation and difficult experimental access for the majority of *Chloroflexi* clades, advancing knowledge has been limited to few lineages (22, 28, 74). The present study provides a new experimental opportunity, as HMA sponges were identified as true *Chloroflexi* hot spots, both in terms of biomass and biodiversity. Metagenomic and single-cell genomic analyses revealed metabolic specialization in that *Anaerolineae* and *Caldilineae* have an expanded gene repertoire for carbohydrate degradation. SAR202 genomes lack transporter and degradation pathways for carbohydrates, and we therefore speculate that they may gain the energy needed by degradation of amino and fatty acids. Similarly, while *Anaerolineae/Caldilineae* take up cofactors, SAR202 has the genomic repertoire for their synthesis. A combination of FISH-CLEM allowed us, for the first time, to visualize *Chloroflexi* in the host context and to identify characteristic cellular morphotypes. The results of this study suggest that *Chloroflexi* symbionts have the genomic potential for DOM degradation from seawater, both labile and recalcitrant. These findings are in line with previous reports that have shown extensive carbohydrate degradation potential in other HMA sponge symbionts (10, 19, 20). Thus, we hypothesize that collectively, sponge microbes not only provide nutrients to the HMA sponge host but also contribute to DOM cycling and primary productivity in reef ecosystems via a pathway termed the “sponge loop.” Considering the abundance and dominance of sponges in many benthic environments, we predict that the role of sponge symbionts in biogeochemical cycles is larger than previously thought.

MATERIALS AND METHODS

Relative abundance of *Chloroflexi* in high-microbial-abundance sponges. To investigate the abundance of the bacterial phylum *Chloroflexi* on a global scale, microbiome data from HMA sponges, classified and predicted (cluster 1), were obtained from Moitinho-Silva et al. (31). This data set is a rarefied operational taxonomic unit (OTU) abundance matrix (23,455) from the mothur processed data of the Sponge Microbiome Project (31). The abundance of *Chloroflexi* OTUs was grouped according to the class level based on SILVA taxonomy (75). Relative abundances were calculated and displayed using the R packages ggplot2 version 3.0.0 (76) and ggpubr version 0.1.8 (<https://CRAN.R-project.org/package=ggpubr>), and complete-linkage hierarchical clustering was performed. For this purpose, Bray-Curtis dissimilarities were calculated on relative abundance values of *Chloroflexi* classes/clades within the phylum using the R package vegan version 2.4-5 (<https://cran.r-project.org/package=vegan>). To test the effect of geography and sponge phylogenetic on the beta diversity of *Chloroflexi* communities, Bray-Curtis dissimilarities were calculated on *Chloroflexi* abundances. Here, the rarefied OTU abundance matrix that contained only *Chloroflexi* OTUs from classified and predicted HMA species was used. Samples with less than 100 sequences were excluded from the analysis. Type II permutation MANOVA using distance matrices was performed with RVAideMemoire package version 0.9-69-3 (<https://CRAN.R-project.org/package=RVAideMemoire>), using 999 permutations and an alpha level of 5%. For this test, each sample was assigned to the geographic region according to their collection sites. Sponge taxonomic order following NCBI taxonomy was used as a proxy of sponge phylogeny. Graphs and tests were performed in R environment version 3.4.3 (<https://www.R-project.org/>).

Sponge sampling and cell separation and handling. *Aplysina aerophoba* was collected from Piran, Slovenia (45.5099 N; 13.5600 E) in May 2013 and transported to the laboratory in Würzburg, Germany, in ambient seawater. Sponge-associated prokaryotes (SAPs) were enriched from fresh sponge tissues within 1 week of collection, for mesohyl and pinacoderm separately, by differential centrifugation as described previously (32). DNA was extracted from frozen SAP aliquots either from pinacoderm or mesohyl tissue (three replicates each) using the FastDNA SPIN kit for Soil (MP Biomedicals, Illklich, France) by the method of Slaby et al. (21). Briefly, different cell lysis protocols were applied for each triplicate to obtain differential sequencing coverage: (i) bead beating, following the manufacturer's protocol, (ii) freeze-thaw cycling (three cycles of 20 min at -80°C and 20 min at 42°C), (iii) proteinase K digestion for 1 h at 37°C (TE buffer with 0.5% SDS and proteinase K at a final concentration of 100 ng/ml). The quantity and quality of the extracted DNA were assessed by Nanodrop, Qubit high-sensitivity assay, and agarose gel electrophoresis. The DNA from two extraction rounds was pooled for each extraction approach separately, and the six sets of metagenomic DNA were sequenced on an Illumina HiSeq2000 platform (150-bp paired-end reads), quality filtered, and assembled at the DOE Joint Genome Institute (Walnut Creek, CA, USA) within the JGI sequencing and data processing pipeline (77). Differential coverage binning was performed with CONCOCT v. 0.2.1 (78) at default settings utilizing the coverage values from the six metagenomic data sets differing in tissue type and/or cell lysis method. A fasta file for each bin was created with the in-house python script mkBinFasta.py (<https://github.com/bslaby/scripts/>). Assembly statistics were obtained from QUAST v. 3.1 (79).

Single cells from cell preparations which were freshly prepared from a sponge from Rovinj/Croatia were sorted and their DNA was amplified by the method of Kamke et al. (19) and stored in 96-well plates at -80°C . Single amplified genomes (SAGs) were PCR screened using the universal primers 27f and 1492r to detect *Chloroflexi* 16S rRNA genes (67). SAGs that tested positive for the presence of a single *Chloroflexi* 16S rRNA gene were sequenced at GATC GmbH (Konstanz, Germany) on an Illumina MiSeq personal sequencer (300 bp; paired end). Sequences were trimmed with Trimmomatic-0.32 (minlen 150, avgqual 25, slidingwindow 4:25) (80) and filtered against eukaryotic, archaeal, and *Delftia* reads (known betaproteobacterial contaminant of the single-cell amplification kit) using blastn (nt database). The SAGs were assembled with SPAdes 3.5.0 ($-sc$, $--careful$, keep contigs >1000 bp) (81) decontaminated using the IMG/MER (Integrated microbial genomes & environmental samples) web tools following the single-cell data decontamination protocol provided at the JGI webpage (<https://img.jgi.doe.gov/w/doc/SingleCellDataDecontamination.pdf>). Only contigs showing clearly different GC/kmer frequency profiles from those of the bulk and that were not identified as *Chloroflexi* derived were filtered out. The SAGs were named according to the columns and rows of the 96-well plate they were identified.

Phylogenetic tree construction. 16S rRNA genes from one metagenome bin (S165) and all 13 single amplified genomes (SAGs) were manually quality checked and aligned with closely related sponge- and non-sponge-derived environmental reference sequences obtained from the Silva database (SSU release 132) using the SINA aligner (82). The program MEGA 7.0.4 (83) was used to align the amino acid sequences of 60 genomes, in total 1,914 positions from nine ribosomal proteins (L2, L4, L14, L15, L22, L24, S3, S17, and S19). The determination of best tree construction model (JTT model), and final tree construction (neighbor-joining method) was conducted in MEGA. As references for the protein tree, sequences from publicly available genomes, basically from cultured *Chloroflexi* were included (see Table S1 in the supplemental material). Due to low genome completeness, some of the proteins used were missing in the relevant genomes. For the 16S rRNA gene-based tree, the neighbor-joining method (GTR+G+I model) was also applied using 1,787 positions from 182 sequences. The trees were visualized using iTOL (interactive tree of life; <https://itol.embl.de/>).

Fluorescence *in situ* hybridization and FISH-CLEM. FISH probes were designed based on the 16S rRNA gene alignment for sponge-specific clades within the classes *Anaerolineae* and *Caldilineae* using the probe design tool implemented in ARB (84). Candidate probes were tested *in silico* for their specific hybridization conditions using different target and nontarget reference sequences using mathFISH (<http://mathfish.cce.wisc.edu/>). The probes with the best performance were tested for hybridization

specificity on fixed (4% paraformaldehyde) *A. aerophoba* microbial cell preparations by the method of Fieseler et al. (32) using formaldehyde (FA) concentration gradients. Finally, we used for the *Caldilineae* probe Cal825 (5'-[Cy3]-ACACCGCCACACCTCGT-[Cy3]-3'; *E. coli* binding positions, 825 to 843) and for the *Anaerolineae* probe Ana1005 (5'-[Alexa Fluor 647]-TCCGCTTTCGCTCCGTA-[Alexa Fluor 647]-3'; *E. coli* binding positions, 1005 to 1023). Additionally, probe SAR202-104 (5'-[Alexa Fluor 488]-GTTACTCAGCCG TCTGCC-[Alexa Fluor 488]-3'; *E. coli* binding positions, 104 to 122) was used to identify members of the SAR202 group in sponges (85). All probes were double labeled at 5' and 3' ends (Sigma-Aldrich, Steinheim, Germany). To test the efficiencies of the newly designed sponge-specific *Chloroflexi* probes and the previously published SAR202-104R probe for the sponge microbiome, FISH conditions were optimized using microbial cell preparations from *A. aerophoba*. The three probes did not colocalize using 10%, 20%, and 30% FA, demonstrating specific binding of the probes to the *Chloroflexi* classes/clades in standard FISH experiments (Fig. S1).

For ultrastructural visualization of sponge *Chloroflexi*, we applied a recently established FISH-CLEM (fluorescence *in situ* hybridization-correlative light and electron microscopy) protocol (86). Briefly, freshly sampled *A. aerophoba* sponges were transported to the University of Wuerzburg where small mesohyl discs (2-mm diameter, 200- μ m thickness) were subjected to high-pressure freezing (HPF) and freeze substitution. Samples were embedded in LR white, and 100-nm ultrathin sections were cut using a Histo Jumbo Diamond knife (Diatome AG, Biel, Switzerland) on a Leica EM UC7 ultramicrotome (Leica Microsystems, Wetzlar, Germany). The sections were placed on poly-L-lysine-coated slides and subjected to fluorescence *in situ* hybridization with the *Chloroflexi* clade-specific probes at 10% FA concentration (900 mM NaCl, 20 mM Tris-HCl [pH 7.4], 0.01% sodium dodecyl sulfate, 20% dextran sulfate). All three class- or clade-specific probes were cohybridized, and fluorescence signals were detected using an Axio Observer.Z1 microscope equipped with AxioCam 506 and Zen 2 version 2.0.0.0 software (Carl Zeiss Microscopy GmbH, Göttingen, Germany). On the same sections that were used for fluorescence microscopy, scanning electron microscopy (SEM) was carried out using a field emission scanning electron microscope JSM-7500F (JEOL, Japan) with LAGE detector (for back scattered electron imaging at extremely low acceleration voltages) directly on the microscope slides. FISH and SEM images of same regions were computer correlated based on sponge heterochromatin pattern by the method of Jahn et al. (86).

Functional genomic analysis. Genomic data from SAG sequences and the extracted metagenome bins were loaded and analyzed in IMG (<https://img.jgi.doe.gov/>) using the KEGG Orthology (KO) terms assigned to our data sets, and metabolic pathways (KEGG) were analyzed. To identify CRISPR-related genes, CRISPRfinder (<http://crispr.i2bc.paris-saclay.fr/Server/>) was used. For the search of specific metabolite gene clusters, antiSMASH was used (73). The genomic potential of investigated microbial symbionts to degrade and transform complex carbohydrates was assessed by screening the IMG-predicted open reading frames (ORFs) of the genome data against the dbCAN (47) and classified according to the carbohydrate-active enzymes (CAZymes) database (48).

Data availability. Data sets for SAGs and the metagenomic bins are available at the NCBI Sequence Read Archive under the BioProject accession numbers or identifiers (IDs) PRJNA506133 and PRJNA366444 to PRJNA366449, respectively. Complete assembled and annotated data are available from IMG (<https://img.jgi.doe.gov/>) under the Gold Study IDs Gs0114494 and Gs0099546 (for more details, see Table 1).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mSystems.00150-18>.

TEXT S1, PDF file, 2 MB.

FIG S1, TIF file, 0.1 MB.

FIG S2, TIF file, 0.8 MB.

FIG S3, TIF file, 0.6 MB.

FIG S4, TIF file, 0.1 MB.

TABLE S1, DOCX file, 0.02 MB.

TABLE S2, DOCX file, 0.02 MB.

TABLE S3, DOCX file, 0.04 MB.

TABLE S4, DOCX file, 0.02 MB.

TABLE S5, DOCX file, 0.01 MB.

ACKNOWLEDGMENTS

We acknowledge Laura Rix and Lucia Pita Galan for many insightful discussions on sponge ecology and Sebastian M. Markert and Christian Stigloher at University of Wuerzburg for help with FISH-CLEM.

This study was supported by DFG (CRC1182-TPB1) and European Union's Horizon 2020 research and innovation program grants (679849 SponGES). M.T.J. was supported by a grant of the German Excellence Initiative to the Graduate School of Life Sciences, University of Wuerzburg. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

- Li CW, Chen JY, Hua TE. 1998. Precambrian sponges with cellular structures. *Science* 279:879–882.
- Yin Z, Zhu M, Davidson EH, Bottjer DJ, Zhao F, Tafforeau P. 2015. Sponge grade body fossil with cellular resolution dating 60 Myr before the Cambrian. *Proc Natl Acad Sci U S A* 112:E1453–E1460. <https://doi.org/10.1073/pnas.1414577112>.
- Van Soest RWM, Boury-Esnault N, Vacelet J, Dohrmann M, Erpenbeck D, De Voogd NJ, Santodomingo N, Vanhoorne B, Kelly M, Hooper JNA. 2012. Global diversity of sponges (Porifera). *PLoS One* 7:e35105. <https://doi.org/10.1371/journal.pone.0035105>.
- Bell JJ. 2008. The functional roles of marine sponges. *Estuar Coast Shelf Sci* 79:341–353. <https://doi.org/10.1016/j.ecss.2008.05.002>.
- Maldonado M, Ribes M, van Duyl FC. 2012. Nutrient fluxes through sponges: biology, budgets, and ecological implications. *Adv Mar Biol* 62:113–182. <https://doi.org/10.1016/B978-0-12-394283-8.00003-5>.
- Thomas T, Moitinho-Silva L, Lurgi M, Björk JR, Easson C, Astudillo-García C, Olson JB, Erwin PM, López-Legentil S, Luter H, Chaves-Fonnegra A, Costa R, Schupp PJ, Steindler L, Erpenbeck D, Gilbert J, Knight R, Ackermann G, Victor Lopez J, Taylor MW, Thacker RW, Montoya JM, Hentschel U, Webster NS. 2016. Diversity, structure and convergent evolution of the global sponge microbiome. *Nat Commun* 7:11870. <https://doi.org/10.1038/ncomms11870>.
- McMurray S, Stubler A, Erwin P, Finelli C, Pawlik J. 2018. A test of the sponge-loop hypothesis for emergent Caribbean reef sponges. *Mar Ecol Prog Ser* 588:1–14. <https://doi.org/10.3354/meps12466>.
- Blunt JW, Copp BR, Keyzers RA, Munro MHG, Prinsep MR. 2017. Marine natural products. *Nat Prod Rep* 34:235–294. <https://doi.org/10.1039/c6np00124f>.
- Wilson MC, Mori T, Rückert C, Uria AR, Helf MJ, Takada K, Gernert C, Steffens UAE, Heycke N, Schmitt S, Rinke C, Helfrich EJM, Brachmann AO, Gurgui C, Wakimoto T, Kracht M, Crüsemann M, Hentschel U, Abe I, Matsunaga S, Kalinowski J, Takeyama H, Piel J. 2014. An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* 506:58–62. <https://doi.org/10.1038/nature12959>.
- Lackner G, Peters EE, Helfrich EJM, Piel J. 2017. Insights into the lifestyle of uncultured bacterial natural product factories associated with marine sponges. *Proc Natl Acad Sci U S A* 114:E347–E356. <https://doi.org/10.1073/pnas.1616234114>.
- Hentschel U, Usher KM, Taylor MW. 2006. Marine sponges as microbial fermenters. *FEMS Microbiol Ecol* 55:167–177. <https://doi.org/10.1111/j.1574-6941.2005.00046.x>.
- Moitinho-Silva L, Steinert G, Nielsen S, Haridoim CCP, Wu YC, McCormack GP, López-Legentil S, Marchant R, Webster N, Thomas T, Hentschel U. 2017. Predicting the HMA-LMA status in marine sponges by machine learning. *Front Microbiol* 8:752. <https://doi.org/10.3389/fmicb.2017.00752>.
- Pita L, Rix L, Slaby BM, Franke A, Hentschel U. 2018. The sponge holobiont in a changing ocean: from microbes to ecosystems. *Microbiome* 6:46. <https://doi.org/10.1186/s40168-018-0428-1>.
- Bayer K, Moitinho-Silva L, Brümmer F, Cannistraci CV, Ravasi T, Hentschel U. 2014. GeoChip-based insights into the microbial functional gene repertoire of marine sponges (high microbial abundance, low microbial abundance) and seawater. *FEMS Microbiol Ecol* 90:832–843. <https://doi.org/10.1111/1574-6941.12441>.
- Weisz JB, Lindquist N, Martens CS. 2008. Do associated microbial abundances impact marine demosponge pumping rates and tissue densities? *Oecologia* 155:367–376. <https://doi.org/10.1007/s00442-007-0910-0>.
- Ribes M, Jiménez E, Yahel G, López-Sendino P, Diez B, Massana R, Sharp JH, Coma R. 2012. Functional convergence of microbes associated with temperate marine sponges. *Environ Microbiol* 14:1224–1239. <https://doi.org/10.1111/j.1462-2920.2012.02701.x>.
- Hoer DR, Gibson PJ, Tommerdahl JP, Lindquist NL, Martens CS. 2018. Consumption of dissolved organic carbon by Caribbean reef sponges. *Limnol Oceanogr* 63:337–351. <https://doi.org/10.1002/lno.10634>.
- Maldonado M. 2016. Sponge waste that fuels marine oligotrophic food webs: a re-assessment of its origin and nature. *Mar Ecol* 37:477–491. <https://doi.org/10.1111/maec.12256>.
- Kamke J, Sczyrba A, Ivanova N, Schwientek P, Rinke C, Mavromatis K, Woyke T, Hentschel U. 2013. Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *ISME J* 7:2287–2300. <https://doi.org/10.1038/ismej.2013.111>.
- Astudillo-García C, Slaby BM, Waite DW, Bayer K, Hentschel U, Taylor MW. 2017. Phylogeny and genomics of SAUL, an enigmatic bacterial lineage frequently associated with marine sponges. *Environ Microbiol* 20:561–576. <https://doi.org/10.1111/1462-2920.13965>.
- Slaby BM, Hackl T, Horn H, Bayer K, Hentschel U. 2017. Metagenomic binning of a marine sponge microbiome reveals unity in defense but metabolic specialization. *ISME J* 11:2465–2478. <https://doi.org/10.1038/ismej.2017.101>.
- Landry Z, Swan BK, Herndl GJ, Stepanauskas R, Giovannoni SJ. 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>.
- Colatriano D, Tran PQ, Guéguen C, Williams WJ, Lovejoy C, Walsh DA. 2018. Genomic evidence for the degradation of terrestrial organic matter by pelagic Arctic Ocean *Chloroflexi* bacteria. *Commun Biol* 1:90. <https://doi.org/10.1038/s42003-018-0086-7>.
- Woese CR. 1988. Bacterial evolution. *Can J Microbiol* 34:547–551.
- Hugenholtz P, Goebel BM, Pace NR. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* 180:4765–4774.
- Gupta RS, Chander P, George S. 2013. Phylogenetic framework and molecular signatures for the class *Chloroflexi* class. nov. into the suborder *Chloroflexineae* subord. nov., consisting of the emended family *Oscillochlorida*. *Antonie Van Leeuwenhoek* 103:99–119. <https://doi.org/10.1007/s10482-012-9790-3>.
- Garrity G, Boone DR, Castenholz RW (ed). 2001. *Bergey's manual of systematic bacteriology*. Springer, New York, NY.
- Hug LA, Castelle CJ, Wrighton KC, Thomas BC, Sharon I, Frischkorn KR, Williams KH, Tringe SG, Banfield JF. 2013. Community genomic analyses constrain the distribution of metabolic traits across the *Chloroflexi* phylum and indicate roles in sediment carbon cycling. *Microbiome* 1:22. <https://doi.org/10.1186/2049-2618-1-22>.
- Campbell AG, Schwientek P, Vishnivetskaya T, Woyke T, Levy S, Beall CJ, Griffen A, Leys E, Podar M. 2014. Diversity and genomic insights into the uncultured *Chloroflexi* from the human microbiota. *Environ Microbiol* 16:2635–2643. <https://doi.org/10.1111/1462-2920.12461>.
- Edmonds-Wilson SL, Nurinova NI, Zapka CA, Fierer N, Wilson M. 2015. Review of human hand microbiome research. *J Dermatol Sci* 80:3–12. <https://doi.org/10.1016/j.jdermsci.2015.07.006>.
- Moitinho-Silva L, Nielsen S, Amir A, Gonzalez A, Ackermann GL, Cerrano C, Astudillo-García C, Easson C, Sipkema D, Liu F, Steinert G, Kotoulas G, McCormack GP, Feng G, Bell JJ, Vicente J, Björk JR, Montoya JM, Olson JB, Reveillaud J, Steindler L, Pineda MC, Marra MV, Ilan M, Taylor MW, Polymenakou P, Erwin PM, Schupp PJ, Simister RL, Knight R, Thacker RW, Costa R, Hill RT, Lopez-Legentil S, Dailianis T, Ravasi T, Hentschel U, Li Z, Webster NS, Thomas T. 2017. The sponge microbiome project. *Gigascience* 6:1–7. <https://doi.org/10.1093/gigascience/gix077>.
- Fieseler L, Horn H, Wagner M, Hentschel U. 2004. Discovery of the novel candidate phylum “Poribacteria” in marine sponges. *Appl Environ Microbiol* 70:3724–3732. <https://doi.org/10.1128/AEM.70.6.3724-3732.2004>.
- Bayer K, Kamke J, Hentschel U. 2014. Quantification of bacterial and archaeal symbionts in high and low microbial abundance sponges using real-time PCR. *FEMS Microbiol Ecol* 89:679–690. <https://doi.org/10.1111/1574-6941.12369>.
- Hoffmann F, Røy H, Bayer K, Hentschel U, Pfannkuchen M, Brümmer F, De Beer D. 2008. Oxygen dynamics and transport in the Mediterranean sponge *Aplysina aerophoba*. *Mar Biol* 153:1257–1264. <https://doi.org/10.1007/s00227-008-0905-3>.
- Varela MM, van Aken HM, Herndl GJ. 2008. Abundance and activity of *Chloroflexi*-type SAR202 bacterioplankton in the meso- and bathypelagic waters of the (sub)tropical Atlantic. *Environ Microbiol* 10:1903–1911. <https://doi.org/10.1111/j.1462-2920.2008.01627.x>.
- Pawlik JR. 2011. The chemical ecology of sponges on Caribbean reefs: natural products shape natural systems. *Bioscience* 61:888–898. <https://doi.org/10.1525/bio.2011.61.11.8>.
- Thrash JC, Seitz KW, Baker BJ, Temperton B, Gillies LE, Rabalais NN, Henrissat B, Mason OU. 2017. Metabolic roles of uncultivated bacterioplankton lineages in the northern Gulf of Mexico “dead zone.” *mBio* 8:e01017-17. <https://doi.org/10.1128/mBio.01017-17>.
- Horn H, Slaby BM, Jahn MT, Bayer K, Moitinho-Silva L, Förster F, Abdel-

- mohsen UR, Hentschel U. 2016. An enrichment of CRISPR and other defense-related features in marine sponge-associated microbial metagenomes. *Front Microbiol* 7:1751. <https://doi.org/10.3389/fmicb.2016.01751>.
39. Michell RH. 2011. Inositol and its derivatives: their evolution and functions. *Adv Enzyme Regul* 51:84–90. <https://doi.org/10.1016/j.advvenzreg.2010.10.002>.
40. Reynolds TB. 2009. Strategies for acquiring the phospholipid metabolite inositol in pathogenic bacteria, fungi and protozoa: making it and taking it. *Microbiology* 155:1386–1396. <https://doi.org/10.1099/mic.0.025718-0>.
41. Müller WE, Rottmann M, Diehl-Seifert B, Kurelec B, Uhlenbruck G, Schröder HC. 1987. Role of the aggregation factor in the regulation of phosphoinositide metabolism in sponges. Possible consequences on calcium efflux and on mitogenesis. *J Biol Chem* 262:9850–9858.
42. Sutherland IW. 1985. Biosynthesis and composition of Gram-negative bacterial extracellular and wall polysaccharides. *Annu Rev Microbiol* 39:243–270. <https://doi.org/10.1146/annurev.mi.39.100185.001331>.
43. Rehm BHA. 2010. Bacterial polymers: biosynthesis, modifications and applications. *Nat Rev Microbiol* 8:578–592. <https://doi.org/10.1038/nrmicro2354>.
44. Misevic GN, Burger MM. 1986. Reconstitution of high cell binding affinity of a marine sponge aggregation factor by cross-linking of small low affinity fragments into a large polyvalent polymer. *J Biol Chem* 261:2853–2859.
45. Misevic GN, Burger MM. 1993. Carbohydrate-carbohydrate interactions of a novel acidic glycan can mediate sponge cell adhesion. *J Biol Chem* 268:4922–4929.
46. Misevic GN, Finne J, Burger MM. 1987. Involvement of carbohydrates as multiple low affinity interaction sites in the self-association of the aggregation factor from the marine sponge *Microciona prolifera*. *J Biol Chem* 262:5870–5877.
47. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 40:W445–W451. <https://doi.org/10.1093/nar/gks479>.
48. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42:D490–D495. <https://doi.org/10.1093/nar/gkt1178>.
49. Garrone R, Thiney Y, Pavans de Ceccatty M. 1971. Electron microscopy of a mucopolysaccharide cell coat in sponges. *Experientia* 27:1324–1326. <https://doi.org/10.1007/BF02136717>.
50. Thomas T, Rusch D, DeMaere MZ, Yung PY, Lewis M, Halpern A, Heidelberg KB, Egan S, Steinberg PD, Kjelleberg S. 2010. Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *ISME J* 4:1557–1567. <https://doi.org/10.1038/ismej.2010.74>.
51. Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS, Thomas T. 2012. Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proc Natl Acad Sci U S A* 109:E1878–E1887. <https://doi.org/10.1073/pnas.1203287109>.
52. Hentschel U, Piel J, Degnan SM, Taylor MW. 2012. Genomic insights into the marine sponge microbiome. *Nat Rev Microbiol* 10:641–654. <https://doi.org/10.1038/nrmicro2839>.
53. Radax R, Hoffmann F, Rapp HT, Leininger S, Schleper C. 2012. Ammonia-oxidizing archaea as main drivers of nitrification in cold-water sponges. *Environ Microbiol* 14:909–923. <https://doi.org/10.1111/j.1462-2920.2011.02661.x>.
54. Fiore CL, Labrie M, Jarett JK, Lesser MP. 2015. Transcriptional activity of the giant barrel sponge, *Xestospongia muta* holobiont: molecular evidence for metabolic interchange. *Front Microbiol* 6:364. <https://doi.org/10.3389/fmicb.2015.00364>.
55. Eichhorn E, Van Der Ploeg JR, Leisinger T. 1999. Characterization of a two-component alkanesulfonate monooxygenase from *Escherichia coli*. *J Biol Chem* 274:26639–26646.
56. Griffin M, Trudgill PW. 1972. The metabolism of cyclopentanol by *Pseudomonas N.C.I.B.* 9872. *Biochem J* 129:595–603.
57. King CE, King GM. 2014. Description of *Thermogemmatipora carboxidivorans* sp. nov., a carbon-monoxide-oxidizing member of the class Ktedonobacteria isolated from a geothermally heated biofilm, and analysis of carbon monoxide oxidation by members of the class Ktedonobacteria. *Int J Syst Evol Microbiol* 64:1244–1251. <https://doi.org/10.1099/ij.s.0.059675-0>.
58. Kamke J, Rinke C, Schwientek P, Mavromatis K, Ivanova N, Szczyrba A, Woyke T, Hentschel U. 2014. The candidate phylum Poribacteria by single-cell genomics: new insights into phylogeny, cell-compartmentation, eukaryote-like repeat proteins, and other genomic features. *PLoS One* 9:e87353. <https://doi.org/10.1371/journal.pone.0087353>.
59. Burgsdorf I, Slaby BM, Handley KM, Haber M, Blom J, Marshall CW, Gilbert JA, Hentschel U, Steindler L. 2015. Lifestyle evolution in cyanobacterial symbionts of sponges. *mBio* 6:e00391-15. <https://doi.org/10.1128/mBio.00391-15>.
60. Liu M, Fan L, Zhong L, Kjelleberg S, Thomas T. 2012. Metaproteogenomic analysis of a community of sponge symbionts. *ISME J* 6:1515–1525. <https://doi.org/10.1038/ismej.2012.1>.
61. Díez-Vives C, Moitinho-Silva L, Nielsen S, Reynolds D, Thomas T. 2017. Expression of eukaryotic-like protein in the microbiome of sponges. *Mol Ecol* 26:1432–1451. <https://doi.org/10.1111/mec.14003>.
62. Reynolds D, Thomas T. 2016. Evolution and function of eukaryotic-like proteins from sponge symbionts. *Mol Ecol* 25:5242–5253. <https://doi.org/10.1111/mec.13812>.
63. Nguyen MTHD, Liu M, Thomas T. 2014. Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. *Mol Ecol* 23:1635–1645. <https://doi.org/10.1111/mec.12384>.
64. Cerveny L, Straskova A, Dankova V, Hartlova A, Ceckova M, Staud F, Stulik J. 2013. Tetratricopeptide repeat motifs in the world of bacterial pathogens: role in virulence mechanisms. *Infect Immun* 81:629–635. <https://doi.org/10.1128/IAI.01035-12>.
65. Matsushima N, Enkhbayar P, Kamiya M, Osaki M, Kretsinger R. 2005. Leucine-rich repeats (LRRs): structure, function, evolution and interaction with ligands. *Drug Des Rev - Online* 2:305–322. <https://doi.org/10.2174/1567269054087613>.
66. Ng ACY, Eisenberg JM, Heath RJW, Huett A, Robinson CM, Nau GJ, Xavier RJ. 2011. Human leucine-rich repeat proteins: a genome-wide bioinformatic categorization and functional analysis in innate immunity. *Proc Natl Acad Sci U S A* 108:4631–4638. <https://doi.org/10.1073/pnas.1000093107>.
67. Bayer K, Scheuermayer M, Fieseler L, Hentschel U. 2013. Genomic mining for novel FADH₂-dependent halogenases in marine sponge-associated microbial consortia. *Mar Biotechnol* 15:63–72. <https://doi.org/10.1007/s10216-012-9455-2>.
68. Hooper JA, Van Soest RM. 2002. *Systema Porifera: a guide to the classification of sponges*. Springer US, Boston, MA.
69. Taylor MW, Radax R, Steger D, Wagner M. 2007. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol Mol Biol Rev* 71:295–347. <https://doi.org/10.1128/MMBR.00040-06>.
70. Siegl A, Hentschel U. 2010. PKS and NRPS gene clusters from microbial symbiont cells of marine sponges by whole genome amplification. *Environ Microbiol Rep* 2:507–513. <https://doi.org/10.1111/j.1758-2229.2009.00057.x>.
71. Pimentel-Elardo SM, Grozdanov L, Proksch S, Hentschel U. 2012. Diversity of nonribosomal peptide synthetase genes in the microbial metagenomes of marine sponges. *Mar Drugs* 10:1192–1202. <https://doi.org/10.3390/md10061192>.
72. Rua CPJ, de Oliveira LS, Froes A, Tschoeke DA, Soares AC, Leomil L, Gregoracci GB, Coutinho R, Hajdu E, Thompson CC, Berlinck RGS, Thompson FL. 2018. Microbial and functional biodiversity patterns in sponges that accumulate bromopyrrole alkaloids suggest horizontal gene transfer of halogenase genes. *Microb Ecol* 76:825–838. <https://doi.org/10.1007/s00248-018-1172-6>.
73. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W, Breitling R, Takano E, Medema MH. 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res* 43:W237–W243. <https://doi.org/10.1093/nar/gkv437>.
74. Mehrshad M, Rodriguez-Valera F, Amoozegar MA, López-García P, Ghai R. 2018. The enigmatic SAR202 cluster up close: shedding light on a globally distributed dark ocean lineage involved in sulfur cycling. *ISME J* 12:655–668. <https://doi.org/10.1038/s41396-017-0009-5>.
75. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO. 2014. The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. *Nucleic Acids Res* 42:D643–D648. <https://doi.org/10.1093/nar/gkt1209>.
76. Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer-Verlag, New York, NY.
77. Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC. 2012. IMG: the Integrated

- Microbial Genomes database and comparative analysis system. *Nucleic Acids Res* 40:D115–D122. <https://doi.org/10.1093/nar/gkr1044>.
78. Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. *Nat Methods* 11:1144–1146. <https://doi.org/10.1038/nmeth.3103>.
 79. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
 80. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
 81. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
 82. Pruesse E, Peplies J, Glöckner FO. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28:1823–1829. <https://doi.org/10.1093/bioinformatics/bts252>.
 83. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874. <https://doi.org/10.1093/molbev/msw054>.
 84. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadukumar, Buchner A, Lai T, Steppi S, Jobb G, et al. 2004. ARB: a software environment for sequence data. *Nucleic Acids Res* 32:1363–1371. <https://doi.org/10.1093/nar/gkh293>.
 85. Morris RM, Rappé MS, Urbach E, Connon SA, Giovannoni SJ. 2004. Prevalence of the *Chloroflexi*-related SAR202 bacterioplankton cluster throughout the mesopelagic zone and deep ocean. *Appl Environ Microbiol* 70:2836–2842. <https://doi.org/10.1128/AEM.70.5.2836-2842.2004>.
 86. Jahn MT, Markert SM, Ryu T, Ravasi T, Stigloher C, Hentschel U, Moitinho-Silva L. 2016. Shedding light on cell compartmentation in the candidate phylum Poribacteria by high resolution visualisation and transcriptional profiling. *Sci Rep* 6:35860. <https://doi.org/10.1038/srep35860>.
 87. Yamada T, Sekiguchi Y. 2009. Cultivation of uncultured *Chloroflexi* subphyla: significance and ecophysiology of formerly uncultured *Chloroflexi* “subphylum I” with natural and biotechnological relevance. *Microbes Environ* 24:205–216. <https://doi.org/10.1264/jsme2.ME091515>.