bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Marine sponges as Chloroflexi hot-spots: Genomic insights and high resolution visualization of an abundant and diverse symbiotic clade

Kristina Bayer[1], Martin T. Jahn[1,2], Beate M. Slaby[1], Lucas Moitinho-Silva[3], and Ute Hentschel[1,4]

[1]GEOMAR-Helmholtz Centre for Ocean Research, RD3-Marine Ecology, RU-Marine Microbiology, Duesternbrooker Weg 20, 24105 Kiel, Germany

[2]University of Wuerzburg, Imaging Core Facility at the Theodor-Boveri-Institute of Bioscience Am Hubland, 97074 Wuerzburg, Germany

[3]Centre for Marine Bio-Innovation, University of New South Wales, Biological Sciences Building D26, Sydney, NSW, Australia, 2052

[4]Christian-Albrechts University of Kiel, Christian-Albrechts-Platz 4, 24118 Kiel, Germany

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Chloroflexi represent a widespread, yet enigmatic bacterial phylum. Meta- and single cell genomics were performed to shed light on the functional gene repertoire of *Chloroflexi* symbionts from the HMA sponge *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, and amino acid utilization as nutrient source by SAR202. While Anaerolineae and Caldilineae import cofactors and vitamins, SAR202 genomes harbor genes encoding for co-factor biosynthesis. A number of features relevant to symbiosis were further identified, including CRISPRs-Cas systems, eukaryote-like repeat proteins and secondary metabolite gene clusters. *Chloroflexi* symbionts were visualized in the sponge extracellular matrix at ultrastructural resolution by FISH-CLEM method. *Chloroflexi* cells were generally rod-shaped and about 1 µm in length, albeit displayed different and characteristic cellular morphotypes per each class. The extensive potential for carbohydrate degradation has been reported previously for *Ca.* 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 also contribute to DOM re-cycling and primary productivity in reef ecosystems via a pathway termed the “sponge loop”.

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

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 matrix (mesohyl). 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 sponges” (HMA) harboring dense and diverse microbial consortia within their mesohyl tissues, and the “low microbial abundance sponges” (LMA) containing microbial numbers in the order of those found in seawater (11–13). While HMA sponges are enriched in *Chloroflexi*, *Acidobacteria*, *Ca.* Poribacteria, the LMA sponges are dominated by *Gamma*- and *Betaproteobacteria* as well as *Cyanobacteria*. 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 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–

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

21). However, the precise fluxes and mechanisms how DOM and POM are taken up and processed within the sponge holobiont remain unknown.

In the present study, we focused our metagenomic analyses on *Chloroflexi* as abundant and characteristic, yet understudied members of HMA sponge microbiomes. The phylum *Chloroflexi* comprises taxonomically and physiologically highly diverse lineages that populate a wide range of habitats (22–25) including the deep sea (26), uranium-contaminated aquifers (27) and the human oral cavity and gut (28, 29). *Chloroflexi* metabolism is very diverse, ranging from anxoxygenic photosynthesis, obligate aerobic/anaerobic heterotrophs, thermophiles, halophiles, clades capable of reductive halogenation, and even predators with gliding motility. Because only few *Chloroflexi* lineages have been cultivated (30) and because draft genomes are limited in number (26, 31), the specific functions of *Chloroflexi* within the ecosystem context remain frequently unknown.

Members of *Chloroflexi* are members of HMA sponge microbiomes, with representatives of classes SAR202, Anaerolineae, and Caldilineae, being most abundant (32). Visualization of *Chloroflexi* by fluorescence *in situ* hybridization (FISH) revealed bright and abundant signals (33, 34). Because *Chloroflexi* likely play an important role in the HMA sponge holobiont, we aimed (i) to assess the relative abundances and distribution in diverse HMA sponge species by using the largest dataset currently available (EMP sponge microbiome), (ii) to provide an updated phylogenetic analysis, (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-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.

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Materials and methods

Relative abundance of Chloroflexi in high microbial abundance (HMA) sponges

To investigate the abundance of the bacterial phylum *Chloroflexi* on a global scale, sponge species, microbiome data from HMA sponges, classified and predicted (Cluster 1), were obtained from Moitinho-Silva *et al.* (2017). This dataset is a rarefied operational taxonomic unit (OTU) abundance matrix (23,455) from the mothur processed data of the Sponge Microbiome Project (32). Abundance of *Chloroflexi* OTUs were grouped according to the class level based on SILVA taxonomy (36). Relative abundances were calculated and displayed using the superheat R package (37).

Sponge sampling, cell separation and handling

The *Chloroflexi* bins from an *Aplysina aerophoba* microbiome derived Illumina data set had been generated in a previous study (21). Briefly, a sponge individual was collected from Piran, Slovenia (45.5099 N; 13.5600 E), microbial cells were enriched from sponge tissues by differential centrifugation, The metagenomic DNA was isolated (33), sequenced (Illumina HiSeq2000 platform, 150 bp paired-end reads), and quality filtered by DOE Joint Genome Institute (JGI). Sequence data of metagenome bins are available from IMG under the Gold study ID Gs0099546 (Table 1). Data normalization, assembly and binning were conducted as described previously [Illumina-only assembly and binning published in (21)].

Single cells were sorted and their DNA amplified as described in Kamke *et al*. (19) and stored at -80°C in 96-well plates. Single amplified genomes (SAGs) were PCR screened using the universal primers 27f and 1492r to detect *Chloroflexi* 16S rRNA genes (38). SAGs 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 decontaminated using the IMG/EMR (Integrated microbial genomes & environmental samples) web tools following the single cell data decontamination protocol provided at JGI webpage (https://img.jgi.doe.gov/w/doc/SingleCellDataDecontamination.pdf). Contamination-free data

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of SAGs are available from IMG under the Gold study ID Gs0114494 (http://img.jgi.doe.gov/, for more details see Table 1).

Phylogenetic tree construction

16S rRNA genes from one metagenome bin and all 13 single amplified genomes (SAGs) were manually quality checked and aligned with several reference sequences obtained from Silva database (SSU release 132) using the SINA aligner (39). The program MEGA 7.0.4 (40) was used for the protein alignment based on amino acid sequences of nine ribosomal proteins (L2, L4, L14, L15, L22, L24, S3, S17, S19), determination of best tree construction model (GTR+G+I model for 16S rRNA genes and JTT model for proteins), and final tree construction (Neighbor Joining method). As references for the protein tree, protein sequences from all public available *Chloroflexi* genomes were included (Suppl. table 1). Due to low genome completeness, some of the used proteins were missing in the regarding genomes. The trees were visualized using iTOL (interactive tree of life; http://www.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 (41). Candidate probes were tested *in silico* for their specific hybridization conditions using different target and non-target reference sequences using mathFISH (http://mathfish.cee.wisc.edu/). Probes with best performance were tested for hybridization specificity on fixed (4% paraformaldehyde) *A. aerophoba* microbial cell preparations as described in Fieseler *et al.* (33) using formamide (FA) concentration gradients. Finally, we used for Caldilineae probe Cal825 (5'-[Cy3]-ACACCGCCCACACCTCGT-[Cy3]-3', *E.coli* binding position: 825-843) and for Anaerolineae probe Ana1005 (5'-[Alexa647]-TCCGCTTTCGCTTCCGTA-[Alexa647]-3', *E.coli* binding position: 1005-1023). Additionally, probe SAR202-104 (5'-[Alexa488]-

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

GTTACTCAGCCGTCTGCC-[Alexa488]-3', *E.coli* binding position: 104-122) was used to identify members of SAR202 group in sponges (42). All probes were double labelled at 5'- and 3'- ends (Sigma-Aldrich, Steinheim, Germany). To establish 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 co-localize using 10%, 20%, and 30% FA demonstrating specific binding of the probes to the *Chloroflexi* classes/ clades in standard FISH experiments (Figure S6).

For ultrastructural visualization of sponge *Chloroflexi*, we applied a recently established FISH-CLEM protocol (Fluorescence *in situ* Hybridization Correlated Light and Electron Microscopy (43). Briefly, freshly sampled *A. aerophoba* sponges were transported to the University of Wuerzburg where small mesohyl discs (2 mm diameter, 200 µm 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 sulphate, 20% dextran sulfate). All three class or clade specific probes were co-hybridized 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, SEM was carried out using a field emission scanning electron microscope JSM-7500F (JEOL, Japan) with LABE 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 as described in Jahn *et al.* (43).

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

196 *Functional genomic analysis*

197 Genomic data from single amplified genome sequences and the extracted metagenome bins
198 were loaded and analyzed in IMG (http://img.jgi.doe.gov/) using the KEGG Orthology (KO)
199 terms assigned to our data sets and metabolic pathways (KEGG) were analyzed. To identify
200 CRISPR related genes, CRISPRfinder (http://crispr.u-psud.fr/Server/) was used. For the
201 search of specific metabolite gene cluster, antiSMASH was used (44). The genomic potential
202 of investigated microbial symbionts to degrade and transform complex carbohydrates was
203 assessed by screening the IMG-predicted open reading frames (ORFs) of the genome data
204 against the dbCAN (45) and classified according to the carbohydrate-active enzymes
205 (CAZymes) database (46).

206

207 **Results and discussion**

208

209 *Chloroflexi abundance in HMA sponges*

210 Recently, members of the phylum *Chloroflexi* were shown to be present in much higher
211 abundance and diversity in HMA sponges than in LMA sponges, which is why they were
212 termed “indicator species” for HMA sponges (35). Here, we provide further details into the
213 presence and abundance of *Chloroflexi* in sponges that were either classified or predicted by
214 machine learning as HMA (35), (Figure 1, tables S2A and S2B). The recently compiled
215 Sponge Microbiome Project (6, 32) was used as a reference database. In these 63
216 investigated sponge species, *Chloroflexi* abundances ranged from 4.39 % ± 3.02 %
217 (*Chondrilla caribensis*) to 31.89 % ± 5.27 % (*Aplysina* sp.), (Figure 1- right panel, suppl.
218 Table 2A). With respect to the *Chloroflexi* classes, the SAR202 clade was the most
219 abundant, contributing on average to 47.74% ± 22.00% of the phylum total abundance
220 (Figure 1 - top panel, suppl. table 2B). Members of the classes Caldilineae (22.35 % ±
221 17.93%) and Anaerolineae 11.64% ± 12.30% were also abundant. Unclassified OTUs at
222 class level represented 14.50% ± 10.77% of *Chloroflexi* sequences indicating that there is
223 phylogenetic novelty still to be discovered. Despite some variability (Figure 1, main panel),

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the classes SAR202, Caldilineae and Anaerolineae as well as diverse hitherto unclassified OTUs dominated the *Chloroflexi* population in the HMA sponges. The remaining classes amounted to a total of 3.78% of total phylum abundance.

Phylogeny of Chloroflexi metagenome bins and single amplified genomes (SAGs)

16S rRNA gene sequencing of 260 single amplified genomes (SAGs) obtained from sponges resulted in the identification of 13 SAGs that belonged to the phylum *Chloroflexi.* Phylogenetic analysis of 16S rRNA genes revealed that one (3D), four (1B, 1G, 1H, 4H) and eight (2D, 3B, 3H, 4A, 5H, 6B, 6C, 6F) SAGs belonged to the classes SAR202, Anaerolineae and Caldilineae, respectively, forming a well-supported sequence cluster (bootstrap: 100) with other sponge derived sequences (Figure S1). The binning of metagenomic sequence data (21) resulted in an 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 (Figure S1). To elucidate the 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 (Figure 2A). One bin (A154) was affiliated to the class Anaerolineae, two metagenome bins were associated to the Caldilineae (C141 and C174), and two were associated to 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. SCGC AAA240-N13, Gs0017605, (26) 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 phylum *Chloroflexi*. All three classes/clades were visualized in the *A. aerophoba* sponge mesohyl matrix by fluorescence in situ co-hybridization (FISH) on

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ultra-thin tissue sections using class/ clade-specific probes. *Chloroflexi* cell signal was abundant, especially for SAR202 and cells were metabolically active as judged by the brightness of the FISH probe (Figure 2B).

General description of genomes

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). Final genome assembly sizes for the metagenomic bins ranged from 3.3 - 6.3 Mbp, showing high estimated genome completeness between 82.8 - 93.4% (Table 1). The guanine-cytosine (GC) content ranged from 58.08 to 59.32%, from 58.36 to 62.98% and from 56.93 to 65.59% for Anaerolineae, Caldilineae and SAR202, respectively. The number of identified genes was 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, C174), two for SAR202 (S152, S156) and one for Anaerolineae (A154)) which had > 90% coverage were chosen for detailed metabolic analysis and inner-phylum comparison. Due to the lack of marine reference genomes for Anaerolineae and Caldilineae or incompleteness of SAR202 reference genomes (26) we did not compare sponge derived Chloroflexi with these references. The letters of the bins were chosen to reflect their phylogenetic identity (“A”= Anaerolineae, “C”= Caldilineae, “S”= SAR202). A gene or enzyme is considered present when identified in both bins of the corresponding clade. Additionally, the most complete Anaerolineae SAG 1B (55.76% genome completeness estimation) was included in the analysis. 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 (see supporting text and supplementary figures for details). Genes involved in

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

glycolysis and the tricarboxylic acid cycle (TCA) were almost completely identified in all metagenome bins (Figure S2A, B). The pentose phosphate pathway (PPP), including the oxidative and non-oxidative phase is largely present. Also, the Entner-Doudoroff-pathway was identified, but lacks the gene encoding for enzyme phosphogluconate dehydratase (EC: 4.2.1.12) in all clades. Furthermore, the enzyme 2-dehydro-3-deoxyphosphogluconate aldolase (EC: 4.2.1.14) is missing in SAR202 (Figure S2C). Interestingly, only the genomes of Anaerolineae and Caldilineae encode for enzymes involved in the ribulose monophosphate pathway (conversion of β-D-fructose-6P to D-ribulose-5P), which was originally found in methylotrophic bacteria but is now recognized as a widespread prokaryotic pathway involved in formaldehyde fixation and detoxification (47).

With respect to autotrophic carbon fixation, the reductive citrate acid cycle (Arnon-Buchanan cycle), is largely present, with the exception of ATP-citrate lyase (EC: 3.2.2.8), that is missing in all six genomes. A second pathway of autotrophic carbon fixation, the Wood-Ljungdahl-pathway was partially identified. While the genes encoding for carbon monoxide dehydrogenase (EC: 1.2.99.2) and formate dehydrogenase (EC: 1.2.1.43) are noticeably present in all six genomes, the rest of the Wood-Ljungdahl-pathway remains incomplete (see Figure S2D). Ammonia import and assimilation is encoded 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 (Figure S2E). The incorporation of sulfur (with thiosulfate or few other sulfur compounds 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 (Figure S2F).

Genes encoding for enzymes of the respiratory chain, including succinate dehydrogenase, cytochrome c oxidase, NADH dehydrogenase and an f-type ATPase, are largely represented on all genomes. These energy gaining processes additionally provide precursors for further metabolic pathways such as biosynthesis of purines and pyrimidines,

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

amino acids and co-factors, or structural compounds. Machinery for transcription and translation, purine and pyridimidine metabolism are largely present. Fatty acid (FA) biosynthesis and degradation pathways were detected in all six genomes. Genes involved in FA beta-oxidation were found almost completely (supporting text), but also the three key enzymes involved in the propionyl-CoA pathway for odd-length and methylated fatty acid degradation were found among the genomes. This includes propionyl-CoA carboxylase (EC: 6.1.4.3) which was annotated in all six genomes, methylmalonyl-CoA epimerase (EC 5.1.99.1), and methylmalonyl-CoA mutase (EC: 5.4.99.2) both of which were found in all genomes except in S152. All genomes encode a number of different ABC transporters to supplement for nutrition and cell growth related compounds (incl. oligopeptides, phosphate, L- and branched chain amino acids, minerals as iron (III) and molybdate, metal ions as zinc, manganese and iron (II)). Additionally, all six genomes largely encode enzymes needed for biosynthesis of most amino acids (see supporting text). We could not identify any of the typical phosphotransferase systems, as it was the case for Poribacteria described previously (19).

We found genomic potential for aromatic degradation in *Chloroflexi* genomes, but pathways remain incomplete (supporting text). Several genes encoding for phenylpropionate and cinnamate degradation, terephthalate degradation, catechol degradation, and xylene degradation were identified on *Chloroflexi* genomes. Also, genes encoding for 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. This finding is interesting in the context that many sponge species contain secondary metabolites that serve as a defense strategy against predators and biofouling (48). On the sponge genus level, highest (20-30% relative to the total microbiome) and most consistent presence of Chloroflexi within a sponge genus were found in the sponge genera Plakortis, Agelas (with the exception of *A. dispar*), *Aplysina* and sister taxon *Aiolochroia.* Interestingly, all of which contain characteristic natural products with aromatic ring structures that serve as chemotaxonomic markers (plakortolides, oroidins,

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

bromo tyrosine alkaloids, respectively). It is therefore tempting to speculate that *Chloroflexi* and SAR202 presences and abundances are shaped, at least to some extent, by the natural products chemistry of their corresponding host sponges.

With respect to cell wall structure, the Anaerolineae and Caldilineae genomes encode the gene repertoire for peptidoglycan biosynthesis. The noticeable lack of peptidoglycan biosynthesis genes in the SAR202 genomes (supporting text) is consistent with previous analyses of three *Chloroflexi* genomes derived from uranium-contaminated aquifers (27). Synthesis pathways encoding for lipopolysaccharides or biosynthesis pathways for other glycan-based membranes could also not be annotated. The synthesis of an S-layer was proposed for SAR202 bacteria (26) as well as a member of GIF09 clade of *Chloroflexi* (27) but genes involved in sialic acid formation (N-Acetylneuraminic acid - Neu5Ac) in the amino and nucleotide sugar metabolism pathway are incomplete (supporting text). Nevertheless, the SAR202 bin S152 encodes a type 2-ABC transporter (NodJI) to export lipo-oligosaccharides. These compounds were shown to play a role in nodulation process in rhizobium bacteria (49), but their potential role for sponge-associated bacteria remains unclear. Additionally, consistent with previous observations (50) none of the six genomes encoded flagellar and chemotaxis genes.

Metabolic specialization: extensive carbohydrate uptake and degradation in Anaerolineae and Caldilineae

The following features are metabolic specialities of Anaerolineae and Caldilineae and appear missing in SAR202, unless otherwise mentioned (Figure 3). We found a remarkable number of ABC transporters for the import of diverse mono- (ribose/ xylose, inositol, glycerol-3P and rhamnose) and oligosaccharides (sorbitol, raffinose/ stachyose/ melibiose, maltose, N-acetylglucosamine, arabinosaccharide) into Anaerolineae and Caldilineae cells. The pentoses xylose and ribose can also be imported into bacterial cells. Xylose can be processed to xylulose which may enter the pentose phosphate cycle finally leading into glycolysis. Ribose can be converted in 5-phosphoribosyl 1-pyrophosphate (PRPP) which is

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

364 precursor for the biosynthesis of the amino acid histidine or it may fuel into purine and
365 pyrimidine synthesis (supporting text). Both groups may also be able to import glycerol-3P
366 which is a phosphoric ester of glycerol (a component of glycerophospholipids) which can be
367 converted to fatty acids. Additionally, we found evidence for arabinose and rhamnose import
368 and degradation; however annotation was incomplete (supporting text).

369 Arabinooligosaccharides (such as α-L-arabinofuranosides, α-L-arabinans,
370 arabinoxylans, and arabinogalactans) result from degradation of plant-like cell material
371 entering the sponge by filtration. These substances may be imported by the almost
372 completely annotated AraNPQ and MsmX transporters and be utilized to L-arabian and L-
373 arabinose by the enzyme α-N-arabinofuranosidase (EC: 3.2.1.55, GH3). The enzyme L-
374 arabinose isomerase (EC: 5.3.1.4, AraA) is present in all four genomes of Anaerolineae and
375 Caldilineae and converts L-arabinose to L-ribulose which can further be converted by
376 reactions of PPP to glucose-6P suitable for entering gylcolysis. Additionally, other
377 oligosaccharides such as stachyose, raffinose, melibiose, and galactose can be imported
378 and used in central metabolism (figure 3, supporting text).

379 The utilization of myo-inositol as carbon source and possibly as a regulatory agent
380 was hypothesized previously for sponge-associated *Ca.* Poribacteria (19). Similarly, sponge-
381 associated Anaerolineae and Caldilineae encode the nearly complete inositol degradation
382 pathway (supporting text). Myo-inositol is likely degraded to glyceraldehyde-3-phosphate and
383 acetyl-CoA, which are further used in the central metabolism. Inositol phosphates are found
384 as part of eukaryotic and archaeal cell wall components (51). Phosphorylated inositol is a
385 precursor for several lipid molecules including sphingolipids, ceramides and
386 glycosylphosphatidylinositol anchors (52), as well as many stress-protective solutes of
387 eukaryotes (51) and might be part of the signal transduction in sponges (53). Therefore, the
388 sponge itself or eukaryotic microorganisms can probably provide inositol as a carbon source
389 or regulatory agent for the microbial symbionts.

390 Uronic acids are sugar acids that can be found in biopolymers of plants, animals and
391 bacteria (54, 55) and are known to occur in glycosaminoglycans (GAGs). GAGs in sponges

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

392 are mainly composed of fucose, glucuronic acid (glucoronate), mannose, galactose, N-
393 acetylglucosamine and sulfate (56–58). Enzymes involved in degradation of uronic acids
394 were found in Anaerolineae and Caldilineae genomes. The possibility of galacturonate and
395 glucuronate catabolism is supported by the conversion of 2-dehydro-3-deoxy-D-gluconate by
396 enzymes glucoronate isomerase (EC: 5.3.1.12), tagaturonate reductase (EC: 1.1.1.58) and
397 altronate hydrolase (EC: 4.2.1.7). Furthermore, the presence of genes encoding for
398 oligogalacturonide lyase (EC: 4.2.2.6), 2-deoxy-D-gluconate 3-dehydrogenase (EC:
399 1.1.1.125) and 2-dehydro-3-desoxy-D-glucokinase (EC: 2.7.1.45) supports possible 4(4-α-D-
400 gluc-4-enuronosyl)-D-galacturonate degradation activity. The products could then enter the
401 ED pathway via 2-dehydro-3-desoxyphophogluconate aldolase (EC: 4.1.2.14). Uronic acid
402 degradation could principally be connected to the inositol degradation pathway via D-
403 galacturonate even though additional genome evidence, such as genes encoding for the
404 enzyme inositol oxidase, (EC. 1.13.99.1) remain wanting (Figure 3, supporting text). A
405 number of transporters for N-acetylglucosamine, digalacturonate, mannose and galactose
406 (Figure 3, supporting text) were identified in Anaerolineae and Caldilineae genomes.
407 Digalacturonate can be utilized by uronic acid degradation pathway (supporting text) and N-
408 acetylglucosamine can be used directly in amino sugar and nucleotide sugar synthesis. The
409 presence of uronic acid degradation pathways provides strong support that Anaerolineae and
410 Caldilineae, similar to the previously described *Ca.* Poribacteria degrade glycosaminoglycan
411 chains of proteoglycans, which are important components of the sponge host matrix (19). In
412 that line, Anaerolineae and Caldilineae genomes were enriched in arylsulfatases A (Figure
413 S3) which were discussed to be involved in metabolisation of sulfated polysaccharides from
414 the sponge extracellular matrix (19, 21) and the heterotrophic ability of symbionts to use
415 sponge components fur nutritional purposes.
416
417 *Expanded carbohydrate-active enzyme (CAZymes) repertoire in Caldilinae and Anaerolinae*
418 In order to search for CAZymes, we screened the *Chloroflexi* genome data against dbCAN
419 (45) and classified the enzymes according to the CAZy database (46). Most *Chloroflexi* hits

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 higher amounts in Caldilineae and Anaerolineae than in SAR202 (Figure S4). 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 family GH74 is 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 third most abundant, but were restricted to Caldilineae bin C174. This enzyme family is annotated as sialidase (EC: 3.2.1.18), capable of hydrolysing glycosidic linkages of terminal sialic acid residues, which are present in sponge mesohyl (59). Altogether 17 glycosyl transferases were identified on *Chloroflexi* genomes, with families GT2, GT4, and GT83 being most abundant. Among the 11 CBM families identified on *Chloroflexi* genomes, with CBM50 as the most abundant, but 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 was reported previously for *Ca.* Poribacteria and the sponge associated unidentified lineage SAUL (19, 20).

Metabolic specialization: co-factor biosynthesis in SAR202 genomes

There is mounting evidence that vitamins and co-factors produced by diverse symbiont lineages could be beneficial to the sponge host (60–63). Parallel transcriptional activity profiling of the symbionts and the sponge showed that the symbionts had the capacity for

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

448 vitamin B biosynthesis whereas the host transcripts displayed capacity for vitamin catabolism
449 (64). It is thus tempting to speculate that the sponges' nutrition is augmented by symbiont
450 derived vitamins and cofactors. In the present study, at least two biosynthetic pathways for
451 co-factor biosynthesis were identified on SAR202 genomes, which were absent in
452 Anaerolineae and Caldilineae (Figure 4). Thiamine is an essential cofactor which is involved
453 in central metabolism. The biosynthesis of the biologically active form thiamine diphosphate
454 (TPP) from L-cysteine, glycine, pyruvate, and glyceraldehyde-3P is encoded on the SAR202
455 genomes (Figures 3, 4A). Although the pathway is incomplete, the data strongly suggest that
456 the synthesis of TPP is restricted to SAR202 bacteria. Instead, Anaerolineae and Caldilineae
457 appear to import thiamine via a ABC transporter (TbpA, ThiPQ) and convert it to TPP by use
458 of thiamine pyrophosphokinase (EC: 2.7.6.2).

459 Secondly, riboflavin (vitamin B2) is required by enzymes and proteins to perform
460 specific physiological functions. Specifically, the active forms, flavin mononucleotide (FMN)
461 and flavin adenine dinucleotide (FAD), serve as cofactors for a variety of flavoprotein enzyme
462 reactions. Most genomes encode for the enzymes FMN adenylyltransferase (EC: 2.7.7.2)
463 and FAD riboflavin kinase (EC: 2.7.1.26) which activate riboflavin into FMN. However, only
464 the SAR202 genomes encode the riboflavin biosynthesis enzymes which rely on GTP and
465 ribulose-5P (Figures 3, 4B). Both substrates can be provided by pathways of the central
466 metabolism pathways (purine metabolism and PPP).

467

468 *Potential for degradation of recalcitrant DOM in SAR202*

469 The possible participation of deep sea SAR202 bacteria in degradation of recalcitrant or
470 refractory DOM was recently postulated by Landry et al (26). Even though the exact
471 composition of DOM in the world's oceans remains to be elucidated, refractory DOM is an
472 important component of the global carbon budget in terms of sheer mass. Landry *et al.* (26)
473 argue that SAR202 genomes have an expanded repertoire of oxidative enzymes that may
474 help in the oxidation of recalcitrant compounds. Interestingly, some of the major enzymes
475 were also found to be enriched in SAR202 symbionts of sponges (Table S4). Among them

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

476 are genes encoding proteins from the CaiB/BaiF family as well as related family III
477 transferases. While the enrichment of CaiB was previously interpreted as carnitine being a
478 carbon and nitrogen source for sponge symbionts (21), an alternative explanation may be
479 that it serves to funnel substrates into degradation pathways without consumption of energy
480 by shuffling CoA, thus generating free electrons (supporting text). Even though the precise
481 function of CaiB/BaiF family proteins cannot be elucidated at the present time, the
482 enrichment in SAR202 genomes is noteworthy. Further, a total of 53 flavin-dependent, class
483 C oxidoreductases of the luciferase family (Flavin mononucleotide monooxygenases,
484 FMNOs, COG2141) which include alkanesulfonate monooxygenase SsuD and methylene
485 tetrahydromethanopterin reductase, were present and enriched in SAR202 genomes. These
486 enzymes are proposed to participate in the oxidation of (long-chain) aldehydes to carboxylic
487 acids, and/ or in cleavage of carbon-sulfur bonds in a variety of sulfonated alkanes (65). The
488 SAR202 genomes contained 22 genes encoding for short-chain alcohol dehydrogenases
489 (COG0300) which might be involved in canalization of ketone body derivate release. Some of
490 these genes from bin S152 showed homologies to cyclopentanol and 3-α (or 20-β) -
491 hydroxysteroid dehydrogenases which convert alicyclic-bound alcohol groups to ketones (26,
492 66). The combination of the enzymes described above could allow sponge associated
493 *Chloroflexi* the conversion of recalcitrant alicyclic ring structure to more labile carboxylic acid,
494 as proposed recently for SAR202 bacteria from deep sea (26).

495 Additionally, a number of oxidative enzymes was identified on the Chloroflexi
496 genomes, but was not enriched in SAR202 (Table S4). These include a 2-
497 oxoglutarate:ferrodoxin oxidoreductase (EC: 1.2.7.11) which oxidizes acetyl-CoA, carbon
498 monoxide dehydrogenase (EC: 1.2.99.2) which might endow the bacteria to oxidize CO_2 as
499 described for some members of the Ktedonobacteria (67), CO- or xanthine dehydrogenases
500 (COG1529) which are possibly involved in oxidation of a broad range of complex substrates
501 (26), choline dehydrogenase (EC. 1.1.99.1) being possibly involved in the oxidation of
502 alcohols to aldehydes, sarcosine oxidase (EC. 1.5.3.1), the serine hydroxymethyltransferase
503 (2.1.2.1) with predicted function in choline degradation, formaldehyde dehydrogenase (EC:

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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-light benthic environments, it remains unclear whether the sponge symbionts encounter recalcitrant DOM derived from seawater sources. Alternatively, and similar to other high diversity microbiomes for example of ant, ruminant, and human guts (ref), the resident microbes are likely to specialize on 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, 50, 60, 61, 68–70). Particularly ankyrin (ANK), tetratricopeptide (TPR), and leucine-rich (LRR) repeat proteins are postulated to be involved in mediating host-microbe interactions (71, 72). Ankyrin and ankyrin repeat containing proteins were detected in all six genomes (Figure S5, Table S5A). It was recently proposed that the expression of sponge symbiont derived ankyrin protein prevents phagocytosis by amoeba (73), and it is tempting to speculate that they protect the symbionts from digestion by the sponge archaeocytes in vivo. TPRs, possibly functioning as module for protein-protein interaction involved in a variety of cellular functions, including those that participate in bacterial pathogenesis (74) were found in all six genomes. However, LRRs were only identified in Caldilineae and SAR202 genomes (Figure S5, Table S5A). Many LRR proteins are involved in protein-ligand interactions; these include plant immune response and the mammalian innate immune response (for review see (75), such as the detection of pathogen-associated molecular patterns by recognition receptors (76). Our findings are in good agreement with general patterns previously found in metagenomes of sponge symbionts (50, 61), in enriched (mini)-metagenomes of

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cyanobacterial sponge symbionts (69) and single amplified genomes from members of SAUL (20) and *Ca.* Poribacteria (68).

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, 50, 60, 61, 69). Here, the investigated Caldilineae genomes and the two SAR202 genomes showed an enrichment of CRISPR (Table 2). However, the two most complete Anaerolineae genomes did not contain any CRISPR-Cas systems. The presence CRISPRs can be explained the extensive filter feeding activity of sponge hosts that result in high exposure of sponge symbionts and phages and other sources of free DNA from ambient seawater.

The synthesis of secondary metabolites is an important defense mechanism of sessile microorganisms such as sponges to protect against predators or biofouling (48). Many of these compounds are in fact produced by the sponge microbiome (10, 48) Especially polyketide synthases (PKS), non-ribosomal peptide synthetases (NRPS), and halogenases are regularly enriched in sponge symbionts, often with new structures and putatively novel activities (38, 61, 77–81). Here, we assessed the genomic repertoire of sponge associated Chloroflexi for secondary metabolism using antiSMASH (44). 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, 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

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

559 The correlation of probe specific fluorescence with scanning electron microscopy
560 images (SEM) allowed the taxon-specific identification of *Chloroflexi* cells at ultrastructural
561 resolution. Overall, distributions of all three *Chloroflex*i clades in *Aplysina aerophoba* indicate
562 that SAR202 cells were more abundant than the other two *Chloroflexi* classes (Figure 5).
563 This is consistent with the relative abundances of *Chloroflexi* classes in HMA sponges
564 extracted from EMP data (Figure 1). Cells belonging to the SAR 202 clade (green signals in
565 Fig. 5A and 5B) are generally rod shaped (0.8 x 1-2 µm) with a regular distribution of cell
566 cytosol content. The Anaerolineae-specific probe targeted rod shaped cells (0.8 x 2.0µm)
567 (red signal, Figure 5A). The characteristic feature of Caldilineae positive cells (ca. 1 x 2 µm)
568 was the presence of electron dense capsules or mucus-like structures located at the cell
569 poles (orange signal, Fig. 5B). All cells which were stained positive with a corresponding
570 FISH probe showed a consistent morphology, which was taken as a measure of probe
571 specificity.
572
573 *Conclusions*
574 Owing to the lack of cultivation and difficult experimental access for the majority of
575 *Chloroflexi* clades, advancing knowledge has been limited to few lineages (26, 27, 31). The
576 present study provides a new experimental opportunity as HMA sponges were identified as
577 true *Chloroflexi* hotspots, both in terms of biomass and biodiversity. Meta- and single cell
578 genomic analyses revealed metabolic specialization in that Anaerolineae and Caldilineae
579 have an expanded gene repertoire for carbohydrate degradation while SAR202 specializes
580 on amino acids. Similarly, while Anaerolineae/ Caldilineae take up cofactors, SAR202 has
581 the genomic repertoire for their synthesis. A combination of FISH-CLEM allowed, for the first
582 time, to visualize *Chloroflexi* in the host context and to identify characteristic cellular
583 morphotypes. The results of this study suggest that *Chloroflexi* symbionts have the genomic
584 potential for DOM degradation from seawater, both labile and recalcitrant. These findings are
585 in line with previous reports that have shown extensive carbohydrate degradation potential in
586 other HMA sponge symbionts (10, 19, 20). Thus, we hypothesize that collectively sponge

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

587 microbes not only provide nutrients to the HMA sponge host, but also contribute to DOM
588 cycling and primary productivity in reef ecosystems via a pathway termed the “sponge loop”.
589 Considering the abundance and dominance of sponges in many benthic environments, we
590 predict that role of sponge symbionts in biogeochemical cycles is larger than previously
591 thought.
592
593 **Acknowledgements**
594 We gratefully acknowledge financial support from the DFG (CRC1182-TPB1) and from the
595 European Union’s Horizon 2020 research and innovation program under Grant Agreement
596 No. 679849 (‘SponGES’). B.M.S. and M.T.J. were supported by grants of the German
597 Excellence Initiative to the Graduate School of Life Sciences, University of Wuerzburg. We
598 thank Laura Rix and Lucia Pita Galan for many insightful discussions on sponge ecology and
599 Christian Stigloher at University of Wuerzburg for support in FISH-CLEM.
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

615 **Figure legend**

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

619

620 Table 2: Genomic characteristics of the six genomes investigated in detail. The absolute
621 numbers of CRISPR arrays – defined by CRISPRfinder and IMG, the number of Ankyrin
622 repeat containing Proteins (ANK), as well as the number of secondary metabolite
623 (antiSMASH) gene cluster per genome are shown.

624

625 Figure 1: Relative abundance of *Chloroflex*i classes in HMA sponges extracted from Earth
626 Microbiome Project (EMP) data (6). The central panel shows the mean relative abundance of
627 *Chloroflexi* classes per sponge species. Top panel shows mean relative abundance of
628 *Chloroflexi* classes in all HMA sponges. Right panel displays the mean relative abundance of
629 whole phylum *Chloroflexi* in HMA sponges (predicted and classified).

630

631 Figure 2A: Concatenated protein tree. Maximum Likelihood phylogenetic analysis of
632 *Chloroflexi* metagenome bins and SAGs (in red) derived from 1914 positions of 60
633 sequences. The percentage of replicate trees in which the associated taxa clustered together
634 in the bootstrap test (100 replicates) are shown as symbols (biggest closed circles bs 100,
635 symbol size represent values from 75 and above). Initial tree for the heuristic search were
636 obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of
637 pairwise distances estimated using a JTT model.

638

639 Figure 2B: Distribution of *Chloroflexi* clades in *Aplysina aerophoba* mesohyl using
640 fluorescence *in situ* hybridization (FISH): The picture show the overlay of all probes, SAR202
641 cells are displayed in green, Caldilineae cells in orange, Anaerolineae cells in red. The

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

642 nucleotide stain DAPI (white/ grey) served as reference for the localization of unstained cells.
643 Scale bar: 10 µm.
644
645 Figure 3: Reconstruction of class-specific metabolic features. On the left side (blueish color)
646 the transporter and utilization pathways found mainly in Anaerolineae and Caldilineae
647 genomes are summarized. SAR202-specific transporter and pathways (for co-factor
648 biosynthesis) are displayed on the right (reddish color). Common genomic features
649 (glycolysis, amino acid metabolism and transport, fatty acid metabolism, respiratory chain
650 etc.) are not shown.
651
652 Figure 4: Metabolic specialization in SAR202: Biosynthesis pathways of co-factors thiamine
653 (A) and riboflavin (B) in SAR202 genomes.
654
655 Figure 5: Visualization of sponge-associated *Chloroflexi* in *Aplysina aerophoba* mesohyl
656 using FISH-CLEM technique: SAR202 cells are displayed in green (panel A and B)
657 Caldilineae in orange (panel B) Anaerolineae in red (panel A). The nucleotide stain DAPI
658 (blue) served as reference for the localization of unstained cells. In both panels the right
659 picture is the overlay of all probes. Scale bars: 1 µm.
660
661 Supplemental tables
662 Supplemental figures
663 Supporting text
664
665
666
667
668
669

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Literature:

1. Li CW, Chen JY, Hua TE. 1998. Precambrian sponges with cellular structures. Science (80-) 279:879–882.
2. 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 201414577.
3. 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.
4. Bell JJ. 2008. The functional roles of marine sponges. Estuar Coast Shelf Sci 79:341–353.
5. Maldonado M, Ribes M, van Duyl FC. 2012. Nutrient Fluxes Through Sponges. Biology, Budgets, and Ecological ImplicationsAdvances in Marine Biology, 1st ed. Elsevier Ltd.
6. 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.
7. 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.
8. Blunt JW, Copp BR, Keyzers RA, Munro MHG, Prinsep MR. 2017. Marine natural products. Nat Prod Rep 34:235–294.
9. Wilson MC, Mori T, Rückert C, Uria AR, Helf MJ, Takada K, Gernert C, Steffens UAE, Heycke N, Schmitt S, Rinke C, Helfrich EJN, 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.

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

10. Lackner G, Peters EE, Helfrich EJN, Piel J. 2017. Insights into the lifestyle of uncultured bacterial natural product factories associated with marine sponges. Proc Natl Acad Sci 114:E347–E356.

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

12. Moitinho-Silva L, Steinert G, Nielsen S, Hardoim 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 8:1–14.

13. 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.

14. Bayer K, Moitinho-Silva L, Brümmer F, Cannistraci C V., 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.

15. Weisz JB, Lindquist N, Martens CS. 2008. Do associated microbial abundances impact marine demosponge pumping rates and tissue densities? Oecologia 155:367–376.

16. 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.

17. 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.

18. Maldonado M. 2016. Sponge waste that fuels marine oligotrophic food webs: a re-assessment of its origin and nature. Mar Ecol 37:477–491.

19. 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.

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

20. 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 00.

21. 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.

22. Woese CR. 1988. Bacterial evolution. Can J Microbiol 34:547–551.

23. Hugenholtz P, Goebel BM, Pace NR. 1998. Impact of culture independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol v:180p4765-4774.

24. Gupta RS, Chander P, George S. 2013. Phylogenetic framework and molecular signatures for the class Chloroflexi and its different clades; proposal for division of the class Chloroflexi class. nov. into the suborder Chloroflexineae subord. nov., consisting of the emended family Oscillochlorida. Antonie Van Leeuwenhoek 103:99–119.

25. Garrity G HJ. 2001. Bergey's Manual®of Systematic Bacteriology. Springer New York, New York, NY.

26. Landry Z, Swa 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 8:1–19.

27. 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 1:1–17.

28. 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.

29. Edmonds-Wilson SL, Nurinova NI, Zapka CA, Fierer N, Wilson M. 2015. Review of human hand microbiome research. J Dermatol Sci 80:3–12.

30. Yamada T, Sekiguchi Y. 2009. Cultivation of Uncultured Chloroflexi Subphyla:

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

754 Significance and Ecophysiology of Formerly Uncultured Chloroflexi “Subphylum I” with
755 Natural and Biotechnological Relevance. Microbes Environ 24:205–216.
756 31. Mehrshad M, Rodriguez-Valera F, Amoozegar MA, López-García P, Ghai R. 2017.
757 The enigmatic SAR202 cluster up close: shedding light on a globally distributed dark
758 ocean lineage involved in sulfur cycling. ISME J 1–14.
759 32. Moitinho-Silva L, Nielsen S, Amir A, Gonzalez A, Ackermann GL, Cerrano C, Astudillo-
760 Garcia C, Easson C, Sipkema D, Liu F, Steinert G, Kotoulas G, McCormack GP, Feng
761 G, Bell JJ, Vicente J, Björk JR, Montoya JM, Olson JB, Reveillaud J, Steindler L,
762 Pineda MC, Marra M V., Ilan M, Taylor MW, Polymenakou P, Erwin PM, Schupp PJ,
763 Simister RL, Knight R, Thacker RW, Costa R, Hill RT, Lopez-Legentil S, Dailianis T,
764 Ravasi T, Hentschel U, Li Z, Webster NS, Thomas T. 2017. The sponge microbiome
765 project. Gigascience 6:1–7.
766 33. Fieseler L, Horn M, Wagner M, Hentschel U. 2004. Discovery of the novel candidate
767 phylum “Poribacteria” in marine sponges.
768 34. Bayer K, Kamke J, Hentschel U. 2014. Quantification of bacterial and archaeal
769 symbionts in high and low microbial abundance sponges using real-time PCR. FEMS
770 Microbiol Ecol 89:679–690.
771 35. Moitinho-Silva L, Steinert G, Nielsen S, Hardoim CCP, Wu Y-C, McCormack GP,
772 López-Legentil S, Marchant R, Webster N, Thomas T, Hentschel U. 2017. Predicting
773 the HMA-LMA Status in Marine Sponges by Machine Learning. Front Microbiol 8.
774 36. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J,
775 Ludwig W, Glöckner FO. 2014. The SILVA and “All-species Living Tree Project (LTP)”
776 taxonomic frameworks. Nucleic Acids Res 42:D643–D648.
777 37. Rebecca Barter, Yu B. 2017. A Graphical Tool for Exploring Complex Datasets Using
778 Heatmaps.
779 38. Bayer K, Scheuermayer M, Fieseler L, Hentschel U. 2013. Genomic Mining for Novel
780 FADH2-Dependent Halogenases in Marine Sponge-Associated Microbial Consortia.
781 Mar Biotechnol 15:63–72.

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

782 39. Pruesse E, Peplies J, Glöckner FO. 2012. SINA: Accurate high-throughput multiple
783 sequence alignment of ribosomal RNA genes. Bioinformatics 28:1823–1829.
784 40. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics
785 Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33:1870–1874.
786 41. Ludwig W. 2004. ARB: a software environment for sequence data. Nucleic Acids Res
787 32:1363–1371.
788 42. Morris RM, Rappé MS, Urbach E, Connon S a, Rappe MS. 2004. Prevalence of the
789 Chloroflexi -Related SAR202 Bacterioplankton Cluster throughout the Mesopelagic
790 Zone and Deep Ocean Prevalence of the Chloroflexi -Related SAR202
791 Bacterioplankton Cluster throughout the Mesopelagic Zone and Deep Ocean †. Appl
792 Environ Microbiol 70:2836–2842.
793 43. Jahn MT, Markert SM, Ryu T, Ravasi T, Stigloher C, Hentschel U, Moitinho-Silva L.
794 2016. Shedding light on cell compartmentation in the candidate phylum Poribacteria
795 by high resolution visualisation and transcriptional profiling. Sci Rep 6.
796 44. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA,
797 Müller R, Wohlleben W, Breitling R, Takano E, Medema MH. 2015. antiSMASH 3.0—a
798 comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic
799 Acids Res 43:W237–W243.
800 45. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for
801 automated carbohydrate-active enzyme annotation. Nucleic Acids Res 40:W445–
802 W451.
803 46. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The
804 carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:490–
805 495.
806 47. Orita I, Sato T, Yurimoto H, Kato N, Atomi H, Imanaka T, Sakai Y. 2006. The ribulose
807 monophosphate pathway substitutes for the missing pentose phosphate pathway in
808 the archaeon Thermococcus kodakaraensis. J Bacteriol 188:4698–4704.
809 48. Pawlik JR. 2011. The Chemical Ecology of Sponges on Caribbean Reefs: Natural

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Products Shape Natural Systems. Bioscience 61:888–898.

49. Spaink HP, Wijfjes AHM, Lugtenberg BJJ. 1995. Rhizobium nodI and NodJ proteins play a role in the efficiency of secretion of lipochitin oligosaccharides. J Bacteriol 177:6276–6281.

50. Horn H, Slaby BM, Jahn MT, Bayer K, Moitinho-Silva L, Förster F, Abdelmohsen UR, Hentschel U. 2016. An Enrichment of CRISPR and other defense-related features in marine sponge-associated microbial metagenomes. Front Microbiol 7.

51. Michell RH. 2011. Inositol and its derivatives: Their evolution and functions. Adv Enzyme Regul 51:84–90.

52. 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.

53. 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.

54. Sutherland IW. 1985. Biosynthesis and Composition of Gram-Negative Bacterial Extracellular and Wall Polysaccharides. Annu Rev Microbiol 39:243–270.

55. Rehm BHA. 2010. Bacterial polymers: biosynthesis, modifications and applications. Nat Rev Microbiol 8:578–592.

56. 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.

57. Misevic GN, Burger MM. 1993. Carbohydrate-carbohydrate interactions of a novel acidic glycan can mediate sponge cell adhesion. J Biol Chem 268:4922–4929.

58. 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.

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

838 59. Garrone R, Thiney Y, Pavans de Ceccatty M. 1971. Electron microscopy of a
839 mucopolysaccharide cell coat in sponges. Experientia 27:1324–1326.
840 60. Thomas T, Rusch D, DeMaere MZ, Yung PY, Lewis M, Halpern A, Heidelberg KB,
841 Egan S, Steinberg PD, Kjelleberg S. 2010. Functional genomic signatures of sponge
842 bacteria reveal unique and shared features of symbiosis. ISME J 4:1557–1567.
843 61. Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS, Thomas T. 2012.
844 Functional equivalence and evolutionary convergence in complex communities of
845 microbial sponge symbionts. Proc Natl Acad Sci 109:E1878–E1887.
846 62. Hentschel U, Piel J, Degnan SM, Taylor MW. 2012. Genomic insights into the marine
847 sponge microbiome. Nat Rev Microbiol 10:641–654.
848 63. Radax R, Hoffmann F, Rapp HT, Leininger S, Schleper C. 2012. Ammonia-oxidizing
849 archaea as main drivers of nitrification in cold-water sponges. Environ Microbiol
850 14:909–923.
851 64. Fiore CL, Labrie M, Jarett JK, Lesser MP. 2015. Transcriptional activity of the giant
852 barrel sponge, Xestospongia muta Holobiont: molecular evidence for metabolic
853 interchange. Front Microbiol 6.
854 65. Eichhorn E, Van Der Ploeg JR, Leisinger T. 1999. Characterization of a two-
855 component alkanesulfonate monooxygenase from Escherichia coli. J Biol Chem
856 274:26639–26646.
857 66. Griffin M, Trudgill PW. 1972. The metabolism of cyclopentanol by Pseudomonas
858 N.C.I.B. 9872. Biochem J 129:595–603.
859 67. King CE, King GM. 2014. Description of Thermogemmatispora carboxidivorans sp.
860 nov., a carbon-monoxideoxidizing member of the class Ktedonobacteria isolated from
861 a geothermally heated biofilm, and analysis of carbon monoxide oxidation by members
862 of the class Ktedonobacteria. Int J Syst Evol Microbiol 64:1244–1251.
863 68. Kamke J, Rinke C, Schwientek P, Mavromatis K, Ivanova N, Sczyrba A, Woyke T,
864 Hentschel U. 2014. The candidate phylum Poribacteria by single-cell genomics: New
865 insights into phylogeny, cell-compartmentation, eukaryote-like repeat proteins, and

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

other genomic features. PLoS One 9.

69. 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.

70. Liu M, Fan L, Zhong L, Kjelleberg S, Thomas T. 2012. Metaproteogenomic analysis of a community of sponge symbionts. ISME J 6:1515–1525.

71. 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.

72. Reynolds D, Thomas T. 2016. Evolution and function of eukaryotic-like proteins from sponge symbionts. Mol Ecol 25:5242–5253.

73. Nguyen MTHD, Liu M, Thomas T. 2014. Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. Mol Ecol 23:1635–1645.

74. 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.

75. 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.

76. 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 108:4631–4638.

77. Hooper JA, Van Soest RM. 2002. Systema Porifera: A Guide to the Classification of Sponges. Springer US, Boston, MA.

78. Taylor MW, Radax R, Steger D, Wagner M. 2007. Sponge-Associated Microorganisms: Evolution, Ecology, and Biotechnological Potential. Microbiol Mol Biol Rev 71:295–347.

79. 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–

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

513.

80. 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.

81. 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.

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

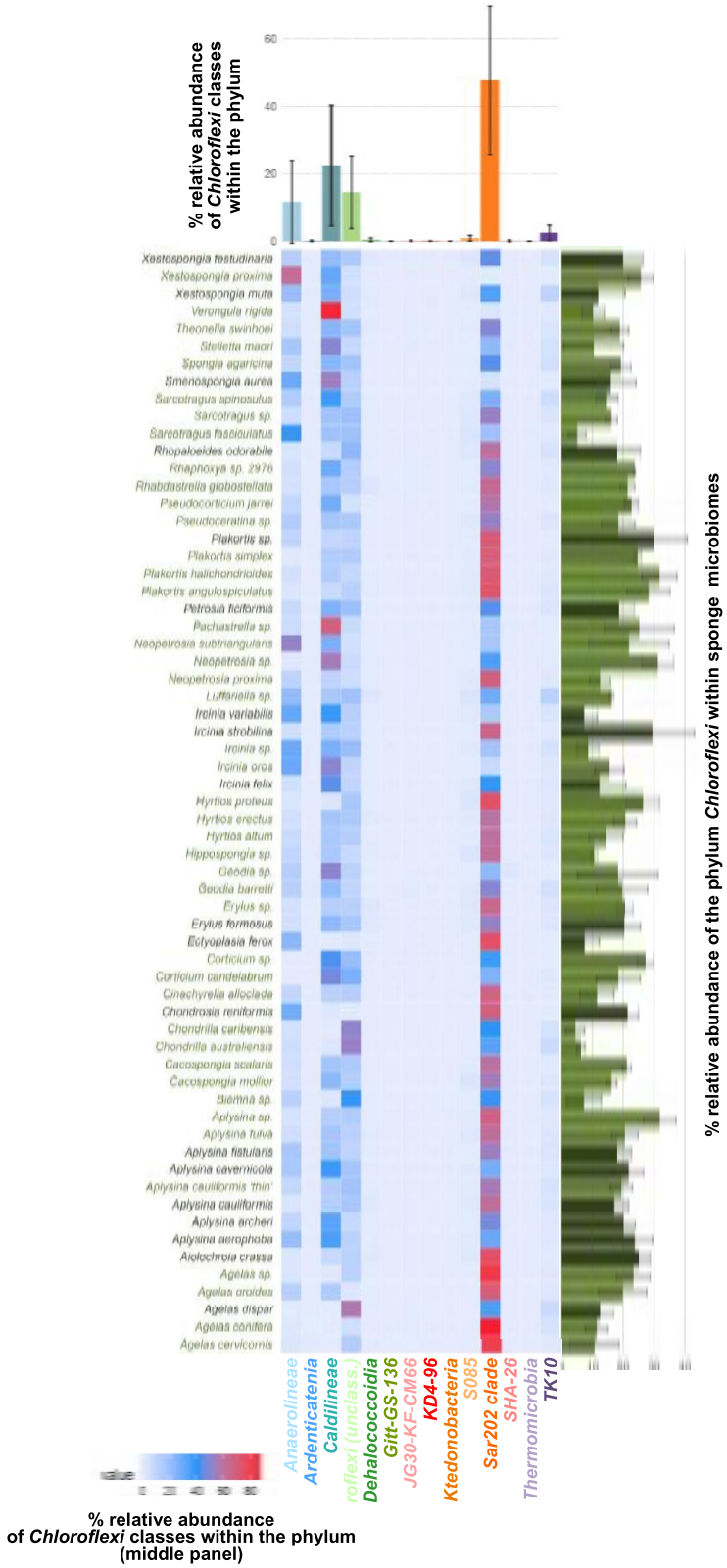

Figure 1: Relative abundance of *Chloroflexi* classes in HMA sponges extracted from Earth Microbiome Project (EMP) data (32). The central panel shows the mean relative abundance of classes per sponge species. Top panel shows mean relative abundance of *Chloroflexi* classes in all HMA sponges (± standard deviation). Right panel displays the mean relative abundance of the phylum *Chloroflexi* in predicted (light green bars) and classified (dark green bars) HMA sponges (± standard deviation).

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

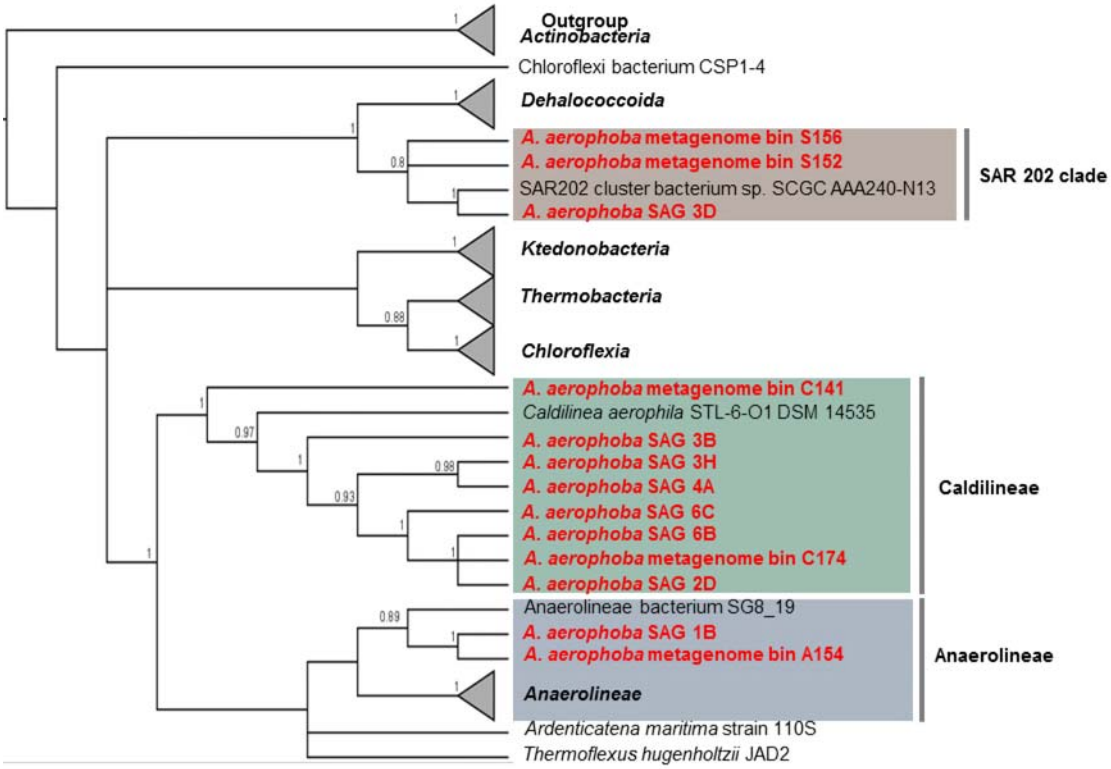

Figure 2A: Concatenated protein tree. Maximum Likelihood phylogenetic analysis of *Chloroflexi* metagenome bins and SAGs (in red) from 1914 positions of 60 sequences using ribosomal proteins. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown. Initial tree for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model. Used reference genomes with accession numbers can be found in Table S1.

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

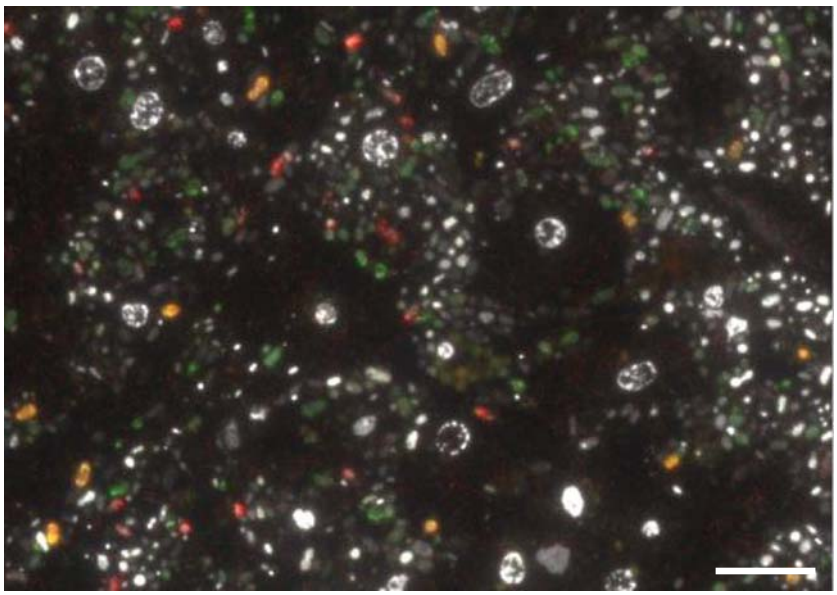

Figure 2B: Distribution of *Chloroflexi* classes/ clades in *Aplysina aerophoba* mesohyl using FISH: SAR202 cells are displayed in green, Caldilineae in orange, and Anaerolineae cells are red. DAPI signal is shown in white/ grey for better contrast. Scale bar: 10 µm

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

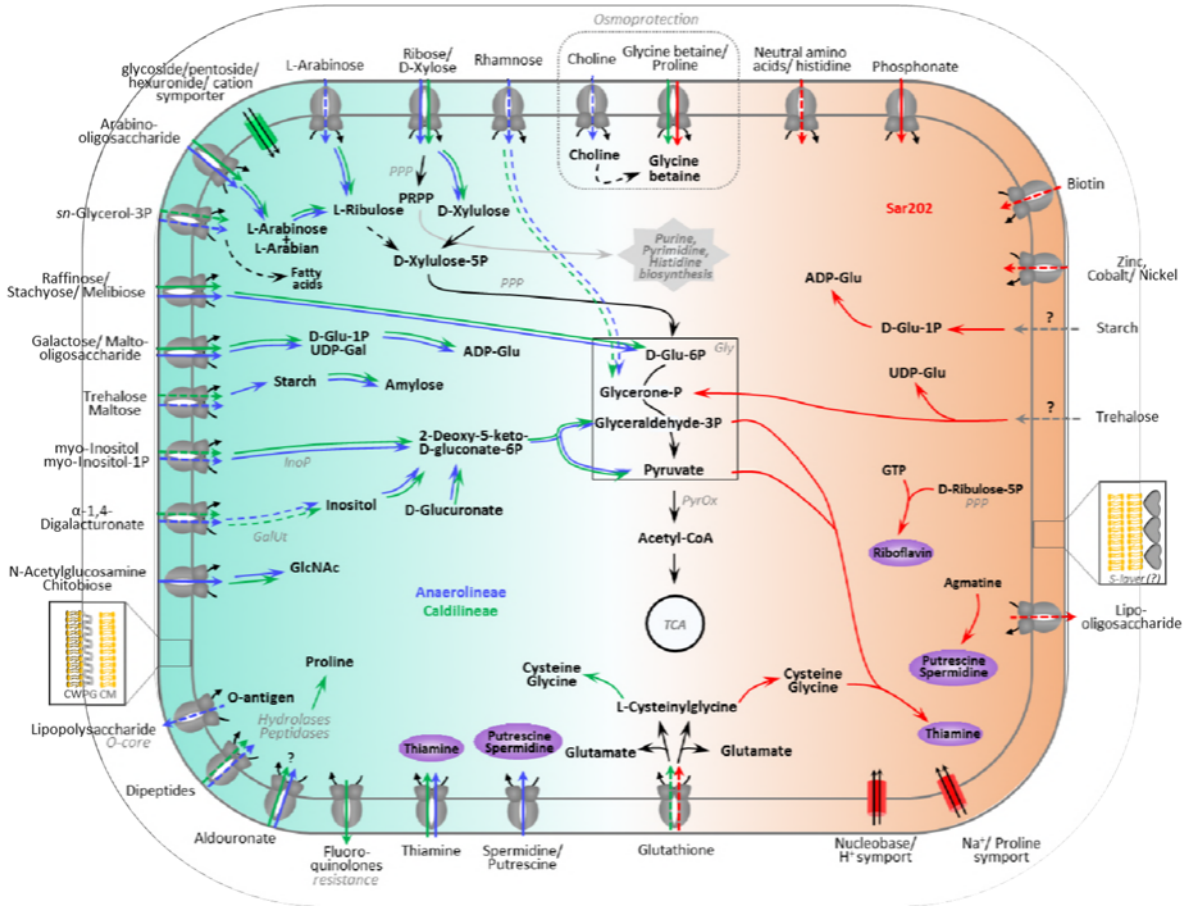

Figure 3: Summarized metabolic features which where found only in Anaerolineae and Caldilineae (left side, blue and green arrows) or in SAR202 genomes (right side, red arrows). The central metabolic pathways (glycolysis, TCA cycle, purine, pyrimidine histidine biosynthesis) located in the figure center 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 genes from the transporter where missing) or could not be annotated in both genomes of one clade.

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

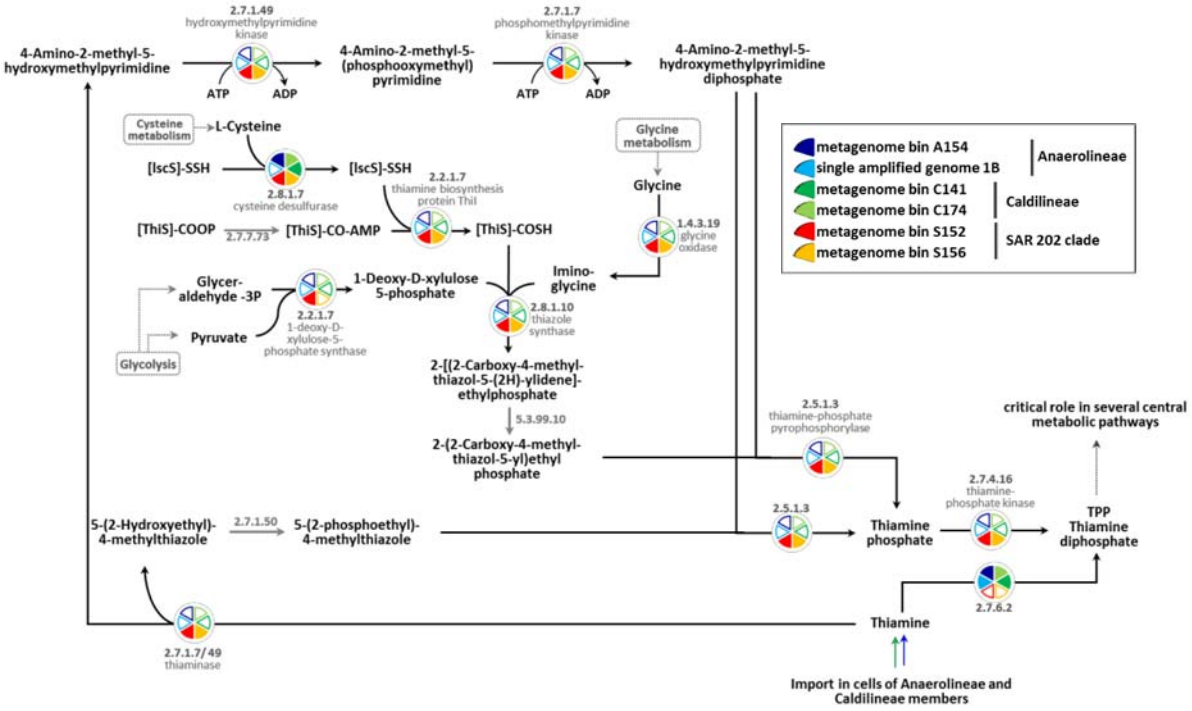

Figure 4A: Pathway encoding for synthesis of thiamine in both SAR202 genomes (yellow and red pie symbols) could be annotated almost complete. Members of classes Anaerolineae and Caldilineae encode for the import of thiamine (blue and green arrow indicating import).

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

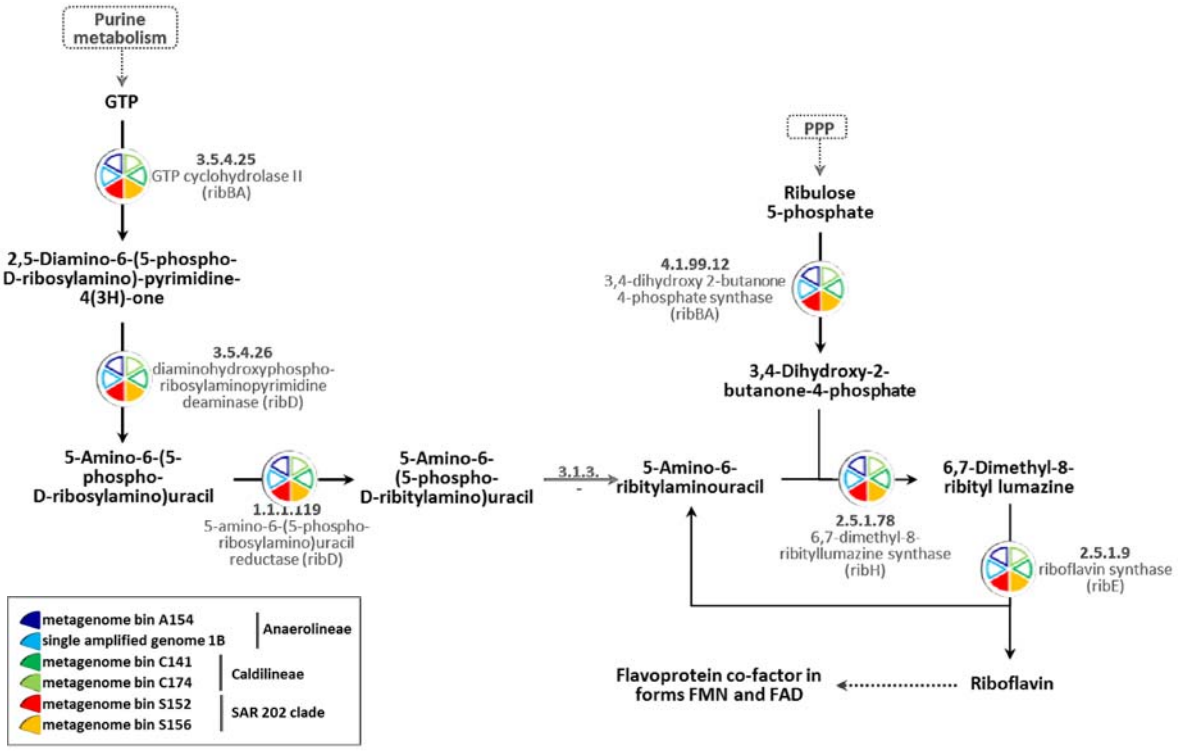

Figure 4B: Pathway encoding for synthesis of riboflavin in both SAR202 genomes (yellow and red pie symbols) could be annotated almost complete. The conversion of Riboflavin in the biological active forms (FMN and FAD) was encoded in the genomes of all three classes.

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

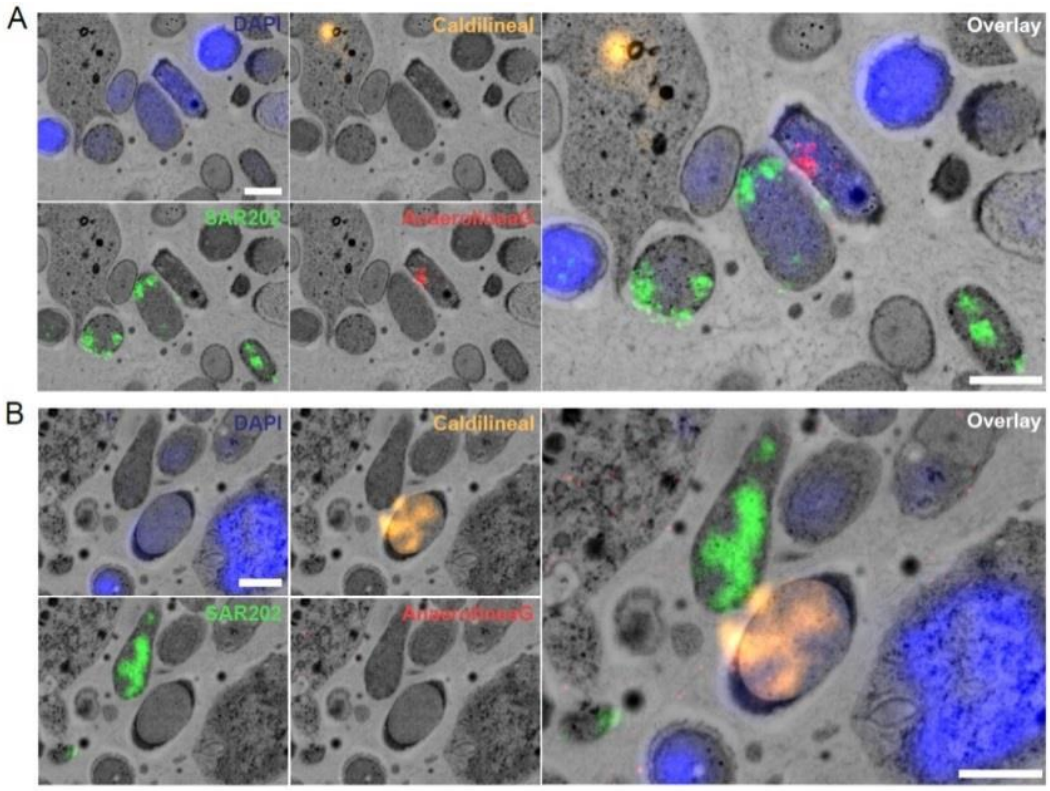

Figure 5: Visualization of sponge-associated *Chloroflexi* in *Aplysina aerophoba* mesohyl using FISH-CLEM: SAR202 cells are displayed in green (panel A and B) Caldilineae in orange (panel B) Anaerolineae in red (panel A). The nucleotide stain DAPI (blue) served as reference for the localization of unstained cells. In both panels the picture on the right is the overlay of all probes and DAPI. scale bars 1 µm.

Table1: Overview of genomes analyzed in this study

	Name	Taxon ID	Genome Size (Mbp)	Scaffold Count	GC (%)	Gene count					Completeness estimation (IMG)
						Total	CDS	RNA	tRNA	w/o func.	
Anaerolineae	SAG 1B	2617270794	2.78	325	58.9	2714	2679	35	24	639	55.8
	SAG 1G	2617270795	1.69	307	58.1	1694	1676	18	14	438	32.0
	SAG 1H	2617270796	1.73	352	59.1	1774	1752	22	17	458	31.9
	SAG 4H	2617270812	0.16	51	58.4	175	168	7	4	43	0
	A154*	2619619053	3.73	107	59.3	3358	3307	51	47	613	91.8
Caldilineae	SAG 2D	2617270806	3.51	489	58.4	3207	3177	30	24	855	65.5
	SAG 3B	2617270807	2.02	310	59.2	1844	1823	21	17	444	38.7
	SAG 3H	2617270810	2.47	290	58.9	2271	2242	29	24	641	50.9
	SAG 4A	2617270811	3.35	437	59.4	3024	2994	30	26	777	64.8
	SAG 5H	2617270814	1.26	183	58.4	1132	1112	20	17	288	17.2
	SAG 6B	2617270816	4.25	685	58.7	3943	3901	42	35	1154	66.8
	SAG 6C	2617270818	2.75	414	58.5	2579	2544	35	28	703	45.9
	SAG 6F	2617270820	3.66	839	58.4	3594	3561	33	25	1063	56.5
	C141*	2619619051	4.59	507	63.0	4288	4236	52	46	1246	90.9
	C174*	2619619055	6.36	647	58.4	5662	5601	61	55	1472	96.3
SAR 202	SAG 3D	2617270809	0.58	106	60.5	622	608	14	13	145	25.4
	S152*	2619619052	5.03	890	56.9	5448	5378	70	60	1938	91.1
	S156*	2619619054	3.35	334	65.6	3463	3402	61	52	964	98.0

IMG Gold Study IDs: Gs0114494; *Gs0099546

The letters of the bins reflect the phylogenetic identity of the bin (“A”= Anaerolineae, “C”= Caldilineae, “S”= SAR202).

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table2: Sponge specific features of selected genomes.

Class/ clade	Name	Taxon ID	CRISPR		ANK	secondary metabolite gene cluster		
			CRISPRfinder Total (No of repeats per spacer)	IMG Total		type 1 PKS	terpene	other
Anaerolineae	SAG 1B	2617270794	-	-	1	-	-	-
	A154*	2619619053	-	-	2	-	-	-
Caldilineae	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

IMG Gold Study IDs: Gs0114494; *Gs0099546 - extracted metagenome bins; ANK: Ankyrins and Ankyrin-repeat containing proteins; secondary metabolite cluster found using antiSMASH 3.0; values are total numbers of genes per genome. The letters of the bins reflect the phylogenetic identity of the bin ("A"= Anaerolineae, "C"= Caldilineae, "S"= SAR202).

bioRxiv preprint doi: https://doi.org/10.1101/328013; this version posted May 22, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.