Microbial strategies for survival in the glass sponge *Vazella pourtalesii*

Kristina Bayer[1]*, Kathrin Busch[1]*, Ellen Kenchington[2], Lindsay Beazley[2], Sören Franzenburg[3], Jan Michels[4], Ute Hentschel[1,5], Beate M. Slaby[1#]

[1]GEOMAR Helmholtz Centre for Ocean Research Kiel, RD3 Marine Symbioses, Kiel, Germany

[2]Department of Fisheries and Oceans, Bedford Institute of Oceanography, Dartmouth, Nova Scotia, Canada

[3]Institute for Clinical Molecular Biology, Kiel University, Kiel, Germany

[4]Functional Morphology and Biomechanics, Institute of Zoology, Kiel University, Kiel, Germany

[5]Kiel University, Kiel, Germany

Running Head: Microbial strategies in *Vazella pourtalesii*

[#]Address correspondence to Beate M. Slaby, bslaby@geomar.de.

*shared first authorship: Kristina Bayer and Kathrin Busch contributed equally to this manuscript. As the main focus lies on the genomic interpretation, Kristina Bayer is named first.

Keywords: Glass sponge, Porifera, Hexactinellida, symbiosis, microbiome, microbial metabolism, metagenomic binning, SAR324, Crenarchaeota, Patescibacteria, Nanoarchaeota

Abstract

Few studies have thus far explored the microbiomes of glass sponges (Hexactinellida). The present study seeks to elucidate the composition of the microbiota associated with the glass sponge *Vazella pourtalesii* and the functional strategies of the main symbionts. We combined microscopic approaches with metagenome-guided microbial genome reconstruction and amplicon community profiling towards this goal. Microscopic imaging revealed that the host and microbial cells appeared within dense biomass patches that are presumably syncytial tissue aggregates. Based on abundances in amplicon libraries and metagenomic data, SAR324 bacteria, Crenarchaeota, Patescibacteria and Nanoarchaeota were identified as abundant members of the *V. pourtalesii* microbiome and their genomic potentials were thus analyzed in detail. A general pattern emerged in that the *V. pourtalesii* symbionts had very small genome sizes in the range of 0.5-2.2 Mb and low GC contents, even below those of seawater relatives. Based on functional analyses of metagenome-assembled genomes (MAGs), we propose two major microbial strategies: the “givers”, namely Crenarchaeota and SAR324, heterotrophs and facultative anaerobes, produce and partly secrete all required amino acids and vitamins. The “takers”, Nanoarchaeota and Patescibacteria, are anaerobes with reduced genomes that tap into the microbial community for resources, e.g., lipids and DNA, likely using pili-like structures. We posit that the existence of microbial cells in sponge syncytia together with the low-oxygen conditions in the seawater environment are factors that shape the unique compositional and functional properties of the microbial community associated with *V. pourtalesii*.

Importance

We investigated the microbial community of *V. pourtalesii* that forms globally unique, monospecific sponge grounds under low-oxygen conditions on the Scotian Shelf, where it plays a key role for its vulnerable ecosystem. The microbial community was found to be concentrated within biomass patches and is dominated by small cells (<1 µm). MAG

analyses showed consistently small genome sizes and low GC contents, which is unusual in comparison to known sponge symbionts. These properties as well as the (facultatively) anaerobic metabolism and a high degree of interdependence between the dominant symbionts regarding amino acid and vitamin synthesis are likely adaptations to the unique conditions within the syncytial tissue of their hexactinellid host and the low-oxygen environment.

Introduction

The fossil record shows that sponges (Porifera) have been essential members of reef communities in various phases of Earth's history, and even built biohermal reefs in the mid-Jurassic to early-Cretaceous (1). Today, extensive sponge aggregations, also known as 'sponge grounds', are found throughout the World's oceans from temperate to arctic regions along shelves, on ridges, and seamounts (2). They can be mono- to multispecific with a single or various sponge species dominating the benthic community, respectively. In sponge ground ecosystems, these basal animals play a crucial role in the provision of habitat, adding structural complexity to the environment and thereby attracting other organisms, ultimately causing an enhancement of local biodiversity (3–5).

Studies on demosponges have shown that they harbor distinct and diverse microbial communities that interact with each other, their host and the environment in various ways (6, 7). The microbial consortia of sponges are represented by diverse bacterial and archaeal communities with ≥63 prokaryotic phyla having been found in sponges so far (6, 8). These sponge microbiomes display host species-specific patterns that are distinctly different from those of seawater in terms of richness, diversity and community composition. Microbial symbionts contribute to holobiont metabolism (e.g., via nitrogen cycling and vitamin production) and defense (e.g., via secondary metabolite production), (reviewed in (7)). Sponges and their associated microbial communities (hereafter termed "holobionts") further contribute to fundamental biogeochemical cycles like nitrogen, phosphorous, and dissolved

76 organic matter in the ecosystem, but the relative contribution of microbial symbionts remains
77 mostly unresolved (1, 7, 9, 10).

78 Sponges of the class Hexactinellida (glass sponges) are largely present and abundant in the
79 mesopelagic realm below 400 feet. They can form extensive reefs of biohermal character
80 and can dominate the sponge ground ecosystems (1, 11). Glass sponges are characterized
81 by a skeleton of siliceous spicules that is six-rayed symmetrical with square axial
82 proteinaceous filament (12). Much of the body is composed of syncytial tissue which
83 represents extensive and continous regions of multinucleated cytoplasm (12, 13). Also
84 nutrients are transported via the cytoplasmic streams of these trabecular syncytia (12). Some
85 discrete cell types exist, including choanocytes and the pluripotent archaeocytes that are
86 likely non-motile and thus not involved in nutrient transport in Hexactinellida (12). While the
87 microbial symbiont diversity and functions are well studied in Demospongiae, much less is
88 known about the presence and function of microbes in glass sponges. In fact,
89 microorganisms have rarely been seen in glass sponges (12). A recent study of South Pacific
90 sponge microbial communities has however shown that general patterns seen previously in
91 shallow-water sponge microbiomes, such as host-specificity and LMA-HMA dichotomy, are
92 generally applicable for these deep-sea sponge microbiomes as well, including those of
93 glass sponges (14). Another study underlined the importance of ammonia-oxidizing archaea
94 (family Nitrosopumilaceae, phylum Thaumarchaeota) in the deep-sea hexactinellid
95 *Lophophysema eversa* using metagenomic data (15).

96 Here we investigate the microbial community of *Vazella pourtalesii* (16), a glass sponge
97 (class Hexactinellida) that forms globally unique, monospecific sponge grounds on the
98 Scotian Shelf off Nova Scotia, Canada. This ecosystem is characterized by relatively warm
99 and nutrient rich water with low oxygen concentrations (1, 17, 18). While there have been a
100 number of studies on the distribution, biomarkers, and possible functional roles of *V.*
101 *pourtalesii* in the ecosystem (1, 17, 19), little has been published to date on its associated
102 microbiota (20). According to phylogenetic and fossil studies, sponges (including glass

sponges) originate from Neoproterozoic times when oxygen was limited (21). Moreover, laboratory experiments have shown that sponges can cope with low oxygen levels for extended periods of time (22–24). Due to the observed low-oxygen conditions at the sampling location, we explored whether the *V. pourtalesii* microbiome contains compositional as well as functional adaptations to the low oxygen environment. Microscopy, metagenome-guided microbial genome reconstruction and amplicon community profiling were employed towards this goal.

Results

Site description

On the Scotian Shelf off Nova Scotia, eastern Canada, highest densities of *V. pourtalesii* were observed and/or predicted in the Emerald Basin and the Sambro Bank areas (Figure 1A). The water column of this region has a characteristic vertical structure with water masses of different temperatures and salinities gradually mixing and creating a distinct temperature-salinity profile (Figure 1B). Main water masses influencing the sampling sites are from surface to deep sea: Cabot Strait Subsurface Water (CBS), Inshore Labrador Current (InLC), Cold Intermediate Layer of Cabot Strait Subsurface Water (CBS-CIL), Labrador Slope Water (LSW), Warm Slope Water (WSW). All *V. pourtalesii* samples of this study originate from a relatively warm (>10 °C) and nutrient rich water mass called Warm Slope Water (WSW) which originates from the Gulf Stream (18). Relatively low oxygen concentrations (<4 mL/L) were measured at the sampling locations and depths, that lay in the range of a mild hypoxia (25).

Microscopic analyses of *V. pourtalesii*

In contrast to other sponges, Hexactinellidae mainly consist of a single syncytium, a fusion of eukaryotic cells forming multinucleate tissue, that permeates the whole sponge (12). By SEM we observed that the overall amount of sponge biomass in *V. pourtalesii* was low and its distribution within the spicule scaffolds was patchy (Figure 2A,B). Closer inspection of such

biomass patches by light microscopy and by TEM microscopy (Figure 2C,D) revealed numerous host cells with their characteristic nuclei as well as high densities of microbial cells of various morphologies and with a dominance of comparatively small cell sizes (<1 µm). In addition, we frequently observed microbial cells that were attached to each other (Figure 2E,F).

MAG selection

In total, 137 metagenome-assembled genomes (MAGs) of >50 % estimated completeness and <10 % redundancy were retrieved (Table S2). Proteobacteria followed by Patescibacteria were the dominant bacterial phyla according to amplicon analyses. In addition, LDA scores were obtained for MAGs based on their read abundance in the different metagenomic sample types (Figure S1): i) *V. pourtalesii* metagenomes vs. seawater metagenomes and ii) pristine *V. pourtalesii* vs. mooring *V. pourtalesii* vs. seawater metagenomes. Based on these assessments, we selected 13 representative *V. pourtalesii*-enriched MAGs for detailed analyses (Table 1). Five MAGs belonged to the candidate phylum Patescibacteria, three to the candidate phylum SAR324, four to the phylum Crenarchaeota, and one to the phylum Nanoarchaeota. The selected MAGs are evidently not redundant representations of the same microbial genomes, which is visualized by the comparatively long branches in the phylogenomic tree (Microscopy of *Vazella pourtalesii* tissue. A) Scanning electron microscopy overview of spicule scaffolds (scale bar: 75 µm). B) SEM close-up image of a biomass patch (scale bar: 3 µm). C) Light-microscopy image (scale bar: 5 µm) and D) TEM image of the same biomass patch (scale bar: 1 µm). E) SEM close-up presumably showing smaller microbes attached to larger ones by stalk- or filament-like structures (scale bar: 1 µm). F) TEM microscopy images of adjacent microbial cells (scale bars: 500 nm). Acronyms: sp= spicule, bp = biomass patch, hc = host cell, mic= microbes, p = potential pilus.

Figure 3). Additionally, the symbiont MAGs showed a maximum of 86.8 %, 93.2 %, and 88.3 % similarity to each other in the ANI analysis for SAR324, Crenarchaeota and Patescibacteria, respectively (Table S4).

Phylogenetic placement of the major players

Three of the selected MAGs belong to the candidate phylum SAR324. The SAR324 clade was recently moved to the level of candidate phylum along with the publication of the whole genome-based classification of microbial genomes (26). The three SAR324 MAGs clustered together outside of known orders with relatively long branches, showing that they are genomically distinct from published genomes of their closest relatives (Microscopy of *Vazella pourtalesii* tissue. A) Scanning electron microscopy overview of spicule scaffolds (scale bar: 75 µm). B) SEM close-up image of a biomass patch (scale bar: 3 µm). C) Light-microscopy image (scale bar: 5 µm) and D) TEM image of the same biomass patch (scale bar: 1 µm). E) SEM close-up presumably showing smaller microbes attached to larger ones by stalk- or filament-like structures (scale bar: 1 µm). F) TEM microscopy images of adjacent microbial cells (scale bars: 500 nm). Acronyms: sp= spicule, bp = biomass patch, hc = host cell, mic= microbes, p = potential pilus.

Figure 3). In the amplicon analysis, this taxon was placed within the class Deltaproteobacteria, of which it posed the most abundant order (Figure S1 Linear discriminant analysis (LDA) Effect Size (LEfSe) plots of MAG abundance based on the abundance table calculated from read coverage data by the metaWRAP quant_bins module. Two sets of groups were analyzed: A) *Vazella* metagenomes vs. seawater reference metagenomes, and B) pristine *Vazella*-derived metagenomes (vazella_p) vs. mooring *Vazella*-derived metagenomes (vazella_m) vs. seawater reference metagenomes. An LDA score of 2 was selected as cut-off.

Figure S2).

Three crenarchaeal MAGs clustered with *Cenarchaeum symbiosum* A (GCA000200715.1) associated with the sponge *Axinella mexicana* (27). One MAG (CrenArch_101) was closely

181 related with *Nitrosopumilus* spp. isolated from Arctic marine sediment (28) and the deep-sea
182 sponge *Neamphius huxleyi* (29). Crenarchaeota are not represented adequately in our
183 amplicon data due to the sequencing primers bias towards bacteria.

184 Three of the five *V. pourtalesii*-enriched Patescibacteria MAGs clustered together with
185 genome GCA002747955.1 from the oral metagenome of a dolphin (30). This clade is a sister
186 group to other families of the order UBA9983 in the class Paceibacteria. The other two
187 patescibacterial MAGs belonged to the family Kaiserbacteriaceae. They were placed
188 separate from each other and outside of known genera, where they cluster with two different
189 groundwater bacteria (GCA_000998045.1 and GCA_002773335.1). Patescibacteria were the
190 second most abundant phylum in the amplicon data with a dominance of the class
191 Parcubacteria which showed high abundances in *V. pourtalesii* compared to controls (Figure
192 S3). The majority of the Parcubacteria (73%) remain unclassified, while 27% are classified as
193 order Kaiserbacteria. As this phylum has not been noticed as particularily abundant in
194 sponge microbiomes before, we tested whether they are sponge specific by comparison with
195 the reference database of the Sponge Microbiome Project (SMP) (8), (data not shown). The
196 899 patescibacterial *V. pourtalesii* ASVs matched to 42 SMP OTUs (mostly listed as
197 'unclassified Bacteria' due to an older Silva version). We identified three SMP OTUs
198 matching *V. pourtalesii* Kaiserbacteria ASVs, namely OTU0005080, OTU0007201, and
199 OTU0159142. These OTUs occurred in 45, 26, and 7 sponge species, respectively, in the
200 SMP reference database showing a global distribution.

201 The one Nanoarchaeota MAG NanoArch_78 belongs phylogenomically to the class
202 Aenigmarchaeia, where it is placed together with MAG GCA_002254545.1 from a deep-sea
203 hydrothermal vent sediment metagenome and outside of known families. Despite the primer
204 bias towards bacteria in the amplicon analysis, the phylum Nanoarchaeota was among the
205 most abundant microbial phyla in this analysis (Figure S1 Linear discriminant analysis (LDA)
206 Effect Size (LEfSe) plots of MAG abundance based on the abundance table calculated from
207 read coverage data by the metaWRAP quant_bins module. Two sets of groups were

208 analyzed: A) *Vazella* metagenomes vs. seawater reference metagenomes, and B) pristine
209 *Vazella*-derived metagenomes (vazella_p) vs. mooring *Vazella*-derived metagenomes
210 (vazella_m) vs. seawater reference metagenomes. An LDA score of 2 was selected as cut-
211 off.

212 Figure S2).

213 **Genome sizes and GC contents**

214 With respect to genome size, the MAGs range from 0.46 Mb (Patesci_30) to 2.16 Mb
215 (SAR324_126), including completeness values into the genome size estimations, with N50
216 values between 3,080 (CrenArch_143) and 66,386 (Patesci_98) (Table 1). According to
217 CheckM, between 53.2 % and 79.4 % of the genomes are covered and redundancies range
218 from 0 % to 8.16 %. GC contents range from 24.3 % (NanoArch_78) to 40.38 %
219 (CrenArch_90). We compared the MAGs to genomes of symbionts from other sponge
220 species and of seawater-derived microbes of each respective phylum regarding their
221 estimated genome sizes and GC contents (Figure 4, Table S3). Due to the lack of published
222 genomes of SAR324 and Nanoarchaeota sponge symbionts, we included genome size and
223 GC content data of unpublished symbionts of *Phakellia* spp. and *Stryphnus fortis* in this
224 analysis. This comparison revealed that the genomes of *V. pourtalesii*-enriched microbes are
225 exceptionally small with very low GC percentages. This trend is especially striking for
226 Patescibacteria, SAR324 and Nanoarchaeota, as their values are not only low for sponge
227 symbionts, but even below the levels of the respective related seawater microbes. While this
228 is not the case for Crenarchaeota, the *V. pourtalesii* MAGs are, nevertheless, in the lower
229 ranges regarding size and GC content in comparison to other sponge symbionts.

230 **Predicted lifestyle of the major players**

231 ***SAR324***

232 Metagenomic analysis of the three SAR324 MAGs from *V. pourtalesii* (Figure 5A) revealed
233 the presence of a near-complete glycolysis pathway up to pyruvate (Pyr) along with the
234 genes for the tricarboxylate acid (TCA) cycle and for conversions of the pentose phosphate

pathway (PPP) (see Text S1 for in-depth analysis of more complex SAR324 and crenarchaeal MAGs). Pyruvate is converted aerobically by the pyruvate dehydrogenase enzyme complex into acetyl-CoA, which fuels the completely annotated (thiamin dependent) TCA cycle. While SAR324 have the genes for a near-complete respiratory chain, their lifestyle appears to be facultatively anaerobic. We detected enzymes of the glyoxylate-bypass (orange arrows within the TCA cycle in Figure 5A), which is required by bacteria to grow anaerobically on fatty acids and acetate (31). This is supported by the presence of a potential AMP-dependent acetyl-CoA synthetase to utilize acetate, whereas enzymes for fatty acids degradation were not found. SAR324 might gain additional energy by a cation-driven p-type ATPase and possibly also anaerobic respiration (furmarate, nitrite/sulfide respiration) (also see Text S1). There is evidence for assimilatory sulfate reduction, but the pathway was not fully resolved.

V. pourtalesii-associated SAR324 are able to take up di-and tricarboxylates using TRAP and TTT transporters, respectively. The imported substances can feed the TCA cycle under aerobic conditions or serve as energy source through fumarate respiration under anaerobic conditions (32, 33). The presence of lactate dehydrogenase (LHD) involved in fermentation is further supporting a facultatively anaerobic lifestyle (31).

SAR324 symbionts are capable of synthesizing diverse amino acids and b-vitamins (riboflavin, coenzyme F420, folate, panthotheonate and thiamin from pyridoxal), using precursors from glycolysis, PPP and TCA cycle (summarized in Figure 5A, see Text S1 for more details). Additionally, the genomes are well equipped with several transporters enabling the import and export of diverse substances (e.g., sugars, amino acids, petides, ions), (Text S1). These transporters are involved in, e.g. osmo-regulation and/or toxic ion reduction (cobalt, fluoride), and in multidrug resistance/import. A p-type ATPase was annotated that may aid in the export of cations or may use an electrochemical gradient for ATP synthesis. Proteins can be secreted by tat- and sec-transport as well as type 1 (T1SS) and type 2

(T2SS) secretion systems probably involved in excretion of symbiosis-relevant molecules. We further identified autoinducer-2 (AI-2) and DNA-T family transporter (Text S1).

Crenarchaeota

The Crenarchaeota of *V. pourtalesii* (Figure 5B) are capable of glycolysis from glucose-6-phosphate (G6P) to pyruvate (Pyr), which is likely anaerobically converted into acetyl-CoA using the enzyme pyruvate-ferredoxin oxidoreductase which suggests facultatively anaerobic metabolism. The TCA cycle was almost completely annotated. Genomic evidence for aerobic and anaerobic respiration (furmarate, nitrite/sulfide respiration) was detected (Text S1). Genes for autotrophic CO_2 fixation in *V. pourtalesii*-associated symbionts were lacking, but assimilatory sulfate reduction was annotated completely. A PPase was annotated which might deliver phosphates to feed the ATP synthase) and in one MAG, a transporter to import dicarboxylates (fumarate, malate, succinate) was annotated, a feature we found in SAR324 as well. These substances could feed the TCA cycle under aerobic conditions. Additionally supporting the hypothesis of a facultatively anaerobic lifestyle is the presence of the enzyme lactate dehydrogenase (LHD) that is involved in fermentation (31). The crenarchaeal genomes encode the synthesis of an even greater number b-vitamins than the SAR324 genomes including riboflavin, coenzyme F420, folate, panthotheonate, pyridoxal, nicotinate and cobalmin (anaerobically) using precursors of central metabolism. Interestingly, they can synthesize thiamin from 5-phosphoribosyl diphosphate (PRPP), while SAR324 partially encode thiamin synthesis from pyridoxal which would need to be imported from an external source (e.g. from other community members).

The crenarchaeal genomes are further well equipped with transporters that facilitate import and export of diverse substances, such as sugars, amino acids, petides, and ions among others (Text S1). These transporters are involved in e.g. osmo-regulation, reflecting the adaptation to a saline environment, and in multidrug resistance/import. A p-type ATPase was annotated in the genomes that may be involved in the export of cations (forced by ATP utilization). Protein secretion can be realized by tat- and sec-transport, which might be

288 involved in transport of proteins, such as those necessary for membrane formation and
289 maintenance (Text S1).

290 ***Patescibacteria***

291 The *V. pourtalesii*-associated Patescibacteria (Figure 5C) showed similar metabolic capacity
292 to published Patescibacteria from other environments: While we found several enzymes
293 involved in glycolysis, we could not resolve the pathway completely. The genomes lack
294 enzymes involved in oxidative phosphorylation (respiration) and the TCA cycle. The
295 synthesis pathways of the important precursor PRPP and subsequent synthesis of purines
296 and pyrimidines were only partially encoded. The biosynthesis of phenylalanine (Phe) from
297 phosphoenolpyruvate (PEP) and erythrose-4-phosphate (E4P) is encoded, but biosynthesis
298 pathways for other amino acids, co-factors or vitamins are missing. We found two p-type
299 ATPases which might export ions or provide energy (ATP) using cation and/or proton
300 gradients present in the environment (holobiont). An anaerobic lifestyle is likely due to the
301 presence of a lactate dehydrogenase (LDH), an enzyme involved in lactate fermentation and
302 the anaerobic acetyl-CoA synthesis using the enzyme pyruvate-ferredoxin oxidoreductase.
303 Patescibacterial MAGs also encode for some transporters, albeit in lower numbers compared
304 to above-described SAR324 and crenarchaeal genomes. These transporters may be
305 involved in osmo-regulation, multidrug in- and efflux, sugar-, amino acids-, and ion uptake.
306 Regarding further symbiosis-relevant features, we detected the autoinducer transporter AI-
307 2E. Additionally, we detected *ComEC/Rec2* and related proteins which are involved in the
308 uptake of single stranded DNA (34). This is supported by the presence of *PilT*, the motor
309 protein which is thought to drive pilin retraction prior to DNA uptake, and the pilus assembly
310 proteins *PilM*, *PilO* and *PilC*. Even though not the full machinery was annotated,
311 Patescibacteria may be able to take up foreign DNA via retraction of type IV-pilin-like
312 structures into the periplasm and via *ComEC* through the inner membrane.

313 ***Nanoarchaeota***

314 The nanoarchaeal MAG (Figure 5D) shows the genomic potential to convert glycerate-3-
315 phosphate (G3P) to Pyr, which represents a shortened glycolysis pathway and results in

reduced potential for energy production (ATP synthesis). It could use an anaerobic pyruvate-formate lyase (PFL) for acetyl-CoA production. Interestingly, an archaeal *type III* RuBisCo is encoded, which catalyzes light-independent CO_2 fixation using ribulose-1,5-bisposphate and CO_2 as substrates to synthesize G3P (35) to fill the only partially encoded glycolysis. All enzymes needed for respiration were absent supporting an anaerobic lifestyle. Energy might be gained using a cation-driven p-type ATPase. Like its published relatives (36, 37) the *V. pourtalesii*-associated nanorarchaeon lacks almost all known genes required for the *de novo* biosyntheses of amino acids, vitamins, nucleotides, cofactors, and lipids. The uptake of some amino acids and ions may be possible as few transporters were detected. Typical archaeal S-layer membrane proteins are encoded, which may be exported by an ABC-transporter (*LolCDE*) or by sec-transport. Other transporters are involved in osmo-regulation and in multidrug resistance and/or transport. We found a prepilin type IV leader peptidase encoded in the genome which is synthesized as a precursors before flagellin/pilin is incorporated into a filament (38).

Discussion

Microbial associations in *V. pourtalesii* syncytia

The glass sponge *V. pourtalesii* consists of a scaffold of spicules with cellular biomass concentrated in biomass patches that contain sponge as well as symbiont cells (Figure 2). We assume that the biomass patches were probably formed by dehydration of syncytial tissues during fixation resulting in higher biomass densities than in the *in vivo* situation (12). Surprisingly high numbers of microbial cells were found within the observed biomass patches, considering that microbes have rarely been noticed in glass sponges previously (12). In *V. pourtalesii*, the microbes appeared in various morphotypes indicating a taxonomically diverse microbial community. Microbial cells of strikingly small sizes (< 1µm) compared to those of shallow water demosponges (39) constituted a large fraction of the microbial community. Microbial cells were frequently seen in close association and even physically attached to each other (Figure 2E,F). These associations were observed between

343 equally sized cells, but also between cells of distinctly different sizes, where the smaller
344 microbes were attached to larger ones by stalk- or filament-like structures (Figure 2E).

345 **Main players in the *V. pourtalesii* microbial community with small, low GC genomes**
346 While previously published sponge metagenomes and symbiont MAGs tended towards high
347 GC contents (typically around 65-70%), the *V. pourtalesii* MAGs show lower GC levels in the
348 range of 24-40% that are more similar to those of seawater metagenomes (40–42), (Figure
349 4, Table S3). Also genome sizes are on the smaller side in comparison to previously
350 published sponge symbiont MAGs (40, 42). The large genome sizes of demosponge
351 symbionts may be attributed to the specific genomic toolbox they require to utilize the
352 mesohyl matrix, such as CAZy enzymes and arylsulfatases. These genes are frequently
353 found enriched in the sponge symbiont genomes compared to free-living relatives (42, 43).
354 This is, however, not the case in *V. pourtalesii* MAGs. On the contrary, here we see GC
355 contents and genome sizes similar to and even below the ones of free-living marine
356 microbes of the same respective phyla (Figure 4). Trophic specialization and avoidance of
357 DNA replication cost have been proposed as hypotheses for genome reduction in free-living
358 marine bacteria, e.g., of the genera *Idiomarina* (44) and *Pelagibacter* (45). For the *V.*
359 *pourtalesii*-associated microbial community, the Black Queen hypothesis may best explain
360 the apparent genomic streamlining: if some members carry out tasks that are beneficial to
361 the whole microbial community, most other members will lose the ability to carry out these
362 (often costly) tasks (46). The small sizes and low GC contents of the *V. pourtalesii* MAGs
363 could, thus, be a sign of adaptation to generally nutrient-limiting environmental conditions
364 (e.g. reviewed in (47)) and specialization on nutrient sources that are available within the
365 sponge host environment, such as ammonia.

366 **The “givers” and “takers” hypothesis**
367 Formerly placed within the Thaumarchaeota, the Crenarchaeota are well known and
368 widespread sponge symbionts (48–51). Different genera have recently been observed in
369 South Pacific Hexactinellida and Demospongiae (14). Also Nanoarchaeota and

Patescibacteria have recently been noticed as members of sponge microbial communities including glass sponges, but no sponge-associated nanoarchaeal genome has been studied to date and - likely due to their low abundance in other sponge species – patescibacterial symbiont genomes have not been studied in detail (14, 52, 53). No sponge-derived genomes are available for the phylum SAR324 so far. The genomes of *V. pourtalesii* symbionts lack a number of properties that we know from typical shallow-water sponge symbionts (7), such as the potential for the production of secondary metabolites and arylsulfatases, and they are not enriched in genes encoding carbohydrate-active enzymes (CAZy). This underlines the above-stated hypothesis that these sponge symbionts do not need a diverse toolbox of genes to make use of a complex mesohyl like symbionts of Demospongiae. On the contrary, they seem to possess streamlined genomes to save resources and likely rely on each other for essential substances, such as certain amino acids and vitamins.

Based on the functional genetic content of the four symbiont phyla that we analyzed in greater detail, we propose two major strategies: “the givers”, namely SAR324 and Crenarchaeota and with comparatively larger, more complex genomes, and likely bigger in cell size, and “the takers”, Patescibacteria and Nanoarchaeota and with reduced genomes and likely smaller cell sizes. We posit that the givers – being genomically well equipped – could be producing and partly secreting all required amino acids and vitamins drawing energy from various aerobic as well as anaerobic processes (Figure 6). Regarding their metabolic repertoire, the here described Crenarchaeota are rather similar to the SAR324 bacteria, namely in their facultatively anaerobic lifestyle, the reactions of the central metabolism, and their ability for amino acid and vitamin biosynthesis. At the same time, while published sponge-associated or free-living Crenarchaeota have the genomic repertoire to fix carbon (e.g., (37, 50, 54–56)), such pathways appear to be absent in *V. pourtalesii*-associated Crenarchaeota. These findings indicate that the symbionts are specifically adapted to the conditions within their respective host sponge and the surrounding environment, e.g. low-oxygen conditions in this case.

Supporting our hypothesis of genome streamlining in the sense of the Black Queen hypothesis, the two "givers" also seem to depend on each other metabolically: Crenarchaeota can produce several b-vitamins, which might be used by members of SAR324. One example is pyrioxidal provided by Crenarchaeota to SAR324, which would thus be able to produce thiamin. Their genomic similarity, their difference to close relatives from other environments, and their metabolic interdependence reinforces our hypothesis that they are, in fact, symbionts specifically adapted to life within their *V. pourtalesii* host. Beyond the scope of the microbial community, microbial vitamin production may also have an important role in the animal host metabolism, such as the respiratory chain, the synthesis of coenzyme A, protein, fatty acid, nucleic acid and carbohydrate metabolism, and co-factor synthesis. As previously hypothesized for Demospongiae symbionts (7), the capacity for vitamin synthesis by microbes associated to *V. pourtalesii* might be an important factor in maintaining the symbiosis with the animal host.

It is tempting to speculate that the takers – reduced in size and functional potential – would scavange from their neighbors. Marine Patescibacteria are known (and named) for their reduced genomes and metabolic capacities (57–59) and also Nanoarchaeota are known for their dependence on a crenarchaeal host, although in very different marine environments such as hydrothermal vents (60, 61). Regarding the exchange of substances between microbes, we propose that the nanoarchaeal symbionts "ride" on the crenarchaeal symbionts, analogous to what is described for *Nanoarchaeum equitans* and *Ignicoccus hospitalis* (37, 62), which use pili for attachment and possibly also for metabolite uptake. Jahn et al. (62) showed that *Nanoarchaeum equitans* may get large amounts of lipids (and possibly other substances) from its associated *Ignicoccus hospitalis*. We hypothesize that the nanoarchaeum in *V. pourtalesii* might be likewise directly associated with the abundant Crenarchaeota and receive, e.g., lipids and DNA via cell attachment and using pili-like structures to maintain cell-cell contact (see, e.g., small cells attached to larger cells in Figure 2E), (37, 60, 63). Patescibacteria could take up required nutrients from their

424 microenvironment, also utilizing pili equipped with a *pilT* motor protein and *comEC* enabling
425 the uptake of DNA for the recycling of nucleotides (64) that they cannot build themselves.
426 Interestingly, we detected copies of the *luxS* gene, the proposed autoinducer-2 (AI-2)
427 synthetase, in patescibacterial and SAR324 genomes. There is strong evidence that AI-2E
428 family homologues function as an AI-2 exporter in *E. coli* cells to control biofilm formation. AI-
429 2 is a proposed signaling molecule for interspecies communication in bacteria (reviewed in
430 (65)). Autoinducer production plays a crucial role in *Vibrio fischeri* colonization of (and
431 maintenance in) the light organ of the host squid *Euprymna scolopes* (66) and was recently
432 detected in sponge-associated *Vibrio* species (67). Microbial *quorom sensing* processes
433 (such as biofilm formation, bioluminescence, motility, virulence factor secretion, antibiotic
434 production, sporulation and competence for DNA uptake) (68) may display symbiosis-
435 relevant features additional or alternative to the ones described before for sponge-associated
436 microbes (e.g., arylsulfatases, TPRs, CRISPR-Cas).

437 ## Conclusions

438 The present study aimed to characterize the diversity and function of microbes residing in the
439 glass sponge *Vazella pourtalesii*. A general pattern emerged in that the *V. pourtalesii*
440 symbionts displayed smaller genome sizes and lower GC contents than bacterial relatives
441 from seawater or from demosponge symbionts. Genomic analysis revealed two putative
442 functional strategies: the “givers” (SAR324 and Crenarchaeota) producing and most likely
443 providing required amino acids and vitamins to the microbial community and the “takers”
444 (Patescibacteria and Nanoarchaeota) depending on the provision of compounds like lipids
445 and DNA that they likely take up via pili-like structures. Their localization within biomass
446 patches together with the environmental low-oxygen conditions could serve as explanation
447 for the unique compositional and functional properties of the microbial community of *V.*
448 *pourtalesii*.

449 ## Material and Methods

Sampling and assessment of microbial community composition

Sampling was performed on a cruise to the Scotian Shelf off Nova Scotia, eastern Canadain August-September 2017 onboard CCGS *Martha L. Black* (*MLB2017001*). Here, we selected a subset of all samples received during this cruise (Table S1) to study the lifestyle strategies of the dominant members of the microbial community (for details on sampling, DNA extraction, and amplicon sequencing see (20)), where we cover the complete dataset to study the microbial diversity inside *V. pourtalesii* in response to anthropogenic activities). Briefly, sponge individuals were collected for this study from pristine areas by the remotely operated vehicle ROPOS (Canadian Scientific Submersible Facility, Victoria, Canada) and tissue subsamples were taken, rinsed in sterile filtered seawater and frozen at -80°C. Samples were collected at an average sampling depth of 168 m (min=161 m, max=183 m) which coincides with the base of the euphotic zone. Oceanographic data such as temperature, salinity and oxygen were collected using CTD casts (sensors by Sea-Bird Electronics SBE 25). Water samples were taken during CTD casts and using Niskin bottles of the ROPOS ROV. Additionally, samples were collected from an Ocean Tracking Network (OTN) acoustic mooring located approximately 10 km northwest of the Sambro Bank Sponge Conservation Area on the Scotian Shelf (20). The mooring was anchored ~ 5 m above the seabed and was deployed for ~ 13 months (15th of August 2016 - 5th of September 2017) prior to its recovery.

DNA was extracted using the DNeasy Power Soil Kit (Qiagen). After quantification and quality assessment by NanoDrop spectrophotometer and by PCR, the V3V4 variable regions of the 16S rRNA gene were amplified in a dual-barcoding approach (69) using a one-step PCR with the primers 5'-CCTACGGGAGGCAGCAG-3' (70) and 5'-GGACTACHVGGGTWTCTAAT-3' (71). Samples were sequenced on a MiSeq platform (MiSeqFGx, Ilumina) with v3 chemistry. The raw sequences were quality-filtered using BBDUK (BBMAP version 37.75 (72)) with a Q20, a minimum length of 250 nt. Sequences were processed in QIIME2 (versions 2018.6 and 2018.8 (73)) implementing the DADA2 algorithm (74) to determine Amplicon Sequence Variants (ASVs). Sequences were denoised

and chimeras, chloroplast and mitochondrial sequences were removed. Taxonomy was assigned using a Naïve Bayes classifier (75) trained on the Silva 132 99% OTUs 16S database (76). An ASV-based phylogeny was generated using the FastTree2 plugin (77). The plots were produced with R (version 3.0.2 (78)), Inkscape (version 0.92.3 (79)), QGIS (version 2.18.4 (80)), and MATLAB (version R2016b including Gibbs Seawater toolbox (81)).

Scanning Electron Microscopy

Tissue subsamples of three sponge individuals (Table S1) were fixed for scanning electron microscopy (SEM) onboard ship in 6.25% GDA in PBS (FisherScientific) in two technical replicates each. Samples were then washed 3x 15 min in PBS, post-fixed for 2 h in 2% osmiumtetroxide (Carl Roth, Germany) and washed again 3x 15 min in PBS. Samples were dehydrated in an ascending ethanol series (ROTIPURAN® Carl Roth, Germany): 1x15 min 30% EtOH, 2x15 min 50% EtOH, 2x15 min 70% EtOH, 2x15 min 80% EtOH, 2x15 min 90% EtOH, 1x15 min 100% EtOH. Subsequent dehydration was continued with carbon dioxide in a Critical Point Dryer (BalzersCPD 030). After critical point drying, the samples were manually fractionated and sputter coated 3 min at 25 mA with gold/palladium (Balzers SCD 004). The preparations were visualized using a Hitachi S-4800 field emission scanning electron microscope (Hitachi High-Technologies Corporation, Tokyo, Japan) with a combination of upper and lower detector at an acceleration voltage of 3 kV and an emission current of 10 mA.

Transmission Electron Microscopy and Light Microscopy

Tissue samples of three sponge individuals (Table S1) were fixed onboard ship in 2.5% glutaraldehyde in 0.1 M natriumcacodylate buffer (pH 7.4; Science Services GmbH) for transmission electron microscopy (TEM) and light microscopy in two technical replicates each. Samples were then rinsed 3x with buffer at 4°C, post-fixed for 2 h in 2% osmiumtetroxide (Carl Roth) and washed with buffer 3x 15 min at 4°C. Samples were partially dehydrated with an ascending ethanol (ROTIPURAN® Carl Roth) series (2x15 min 30% EtOH, 1x15 min 50% EtOH, up to 70% ethanol). Samples were stored at 4°C overnight

505 before desilicification with 4% suprapure hydrofluoric acid (Merck; incubation of
506 approximately 5 hours). Afterwards, samples were washed 8x15 min in 70% EtOH (with an
507 overnight storage at 4°C in between washings). Dehydration was continued with a graded
508 ethanol series (1x15 min 90% EtOH and 2x15 min 100% EtOH) followed by gradual
509 infiltration with LR-White resin (Agar Scientific) at room temperature (1x1 h 2:1 Ethanol:LR-
510 White; 1x1 h 1:1 Ethanol:LR-White; 1x1 h 1:2 Ethanol:LR-White; 2x2 h pure LR-White).
511 Samples were incubated in pure LR-White resin at 4°C overnight before being transferred
512 into fresh resin and polymerized in embedding capsules at 57°C for 2 days.

513 Semithin sections (0.5 µm) were cut (in technical replicates) sing an ultramicrotome
514 (Reichert-Jung ULTRACUT E), equipped with a diamond knife (DIATOME, Switzerland), and
515 afterwards stained with Richardson solution (ingredients from Carl Roth; prepared as
516 described in (82)). Semithin sections were then mounted on SuperFrost Ultra Plus®
517 microscopy slices (Carl Roth; using Biomount medium produced by Plano) and visualized
518 with an Axio Observer.Z1 microscope (Zeiss, Germany). Ultrathin sections (70 nm) were cut
519 (in technical replicates) with the same ultramicrotome, mounted on pioloform coated copper
520 grids (75 mesh; Plano) and contrasted with uranyl acetate (Science Services; 20 min
521 incubation followed by washing steps with MilliQ water) and Reynold´s lead citrate
522 (ingredients from Carl Roth; 3 min incubation followed by washing steps with MilliQ water).
523 The ultrathin preparations were visualized at an acceleration voltage of 80 kV on a Tecnai
524 G2 Spirit Bio Twin transmission electron microscope (FEI Company).

525 **Microbial functional repertoire**

526 For metagenomic sequencing, DNA was extracted from the seven sponge samples (four
527 from natural *Vazella* grounds and three from the mooring to optimize for differential coverage
528 binning) and five seawater controls (Table S1) with the QiagenAllPrep DNA/RNA Mini Kit.
529 Two washing steps with buffer AW2 were employed during DNA extraction. DNase and
530 protease-free RNase A (Thermo Scientific) was used to remove remnant RNA from the DNA
531 extracts. For seawater controls, DNA was extracted from one half of a PVDF membrane filter

(seawater filter – SWF; see above). The DNA was concentrated by precipitation with 100 % ethanol and sodium acetate buffer and re-eluted in 50 µl water. For all extracts, DNA quantity and quality were assessed by NanoDrop measurements and Qubit assays, and 30 µl (diluted in water, if necessary) were sent for metagenomic Illumina Nextera sequencing (HiSeq 4000, 2x150 bp paired-end) at the Institute of Clinical Molecular Biology (IKMB) of Kiel University. Sequence quality of all read files was assessed with FastQC (83).

The raw reads were trimmed with Trimmomatic v0.36 (ILLUMINACLIP:NexteraPE-Pe.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36) and co-assembled with megahit v1.1.3 (84). The metaWRAP v1.0.2 pipeline was implemented for binning as follows (85): Initial binning was performed with metabat, metabat2, and maxbin2 (86–88) within metaWRAP. The bins were refined with the metaWRAP bin_refinement module and further improved where possible with the reassemble_bins module. This module uses the genome assembler SPAdes v3.12.0 (89) on two sets of reads mapped to the original bin with strict and more permissive settings and then compares the original bin with the two newly assembled genomes. Which of the three versions of the metagenome-assembled genome (MAG) was the best in each respective case and was, thus, used for further analyses, is indicated by the trailing letter in the names (‘o’ for ‘original’, ‘p’ for ‘permissive’, or ‘s’ for ‘strict’) in Table S2. The MAGs that were further analyzed in detail were renamed indicating their phylum-level affiliation and their bin number.

MAG taxonomy was determined by GTDB-Tk based on whole genome information and following the recently published, revised microbial taxonomy by Parks and colleagues (26, 90, 91). The phylogenomic trees produced by GTDB-Tk (77, 92, 93) were visualized on the Interactive Tree Of Life (iTOL) platform v4.3 (94). MAG abundance in the different metagenomic datasets was quantified with the metaWRAP quant_bins module and used to determine which MAGs were enriched in which sample type by calculating linear discriminant analysis (LDA) scores with LEfSe v1.0 (95) in two ways: i) “*V. pourtalesii*” vs. “water”, and ii) “pristine *V. pourtalesii*” vs. “mooring *V. pourtalesii*” vs. “water”. We identified MAGs belonging

559 to the bacterial candidate phyla Patescibacteria and SAR324 and the archaeal phyla
560 Crenarchaeota and Nanoarchaeota that were enriched in *V. pourtalesii* over seawater or in
561 one of the *V. pourtalesii* subsets. The MAGs were compared to each other within their
562 taxonomic groups using average nucleotide identity (ANI) of the pangenomic workflow of
563 anvi'o v5.2 (96, 97) and they were compared to seawater and other host sponge-derived
564 reference genomes (Table S3) regarding their genome sizes and GC contents. For functional
565 annotations, interproscan v5.30-69.0 including GO term and pathway annotations was used
566 (98, 99). The resulting EC numbers were converted to K terms to apply the online tool
567 'Reconstruction Pathway' in KEGG mapper (https://www.genome.jp/kegg/). Additionally,
568 manual search in the annotation tables (DOI: 10.6084/m9.figshare.12280313) allowed the
569 identification of several enzymes completing some pathways. Potential transporters were
570 identified in the above-described annotation and additionally using the online tool
571 TransportDB 2.0 (100).

572 **Data deposition**

573 Detailed sample metadata was deposited in the PANGAEA database:
574 https://doi.pangaea.de/10.1594/PANGAEA.917599. Amplicon and metagenomic raw read
575 data were deposited in the NCBI database under BioProject PRJNA613976. Individual
576 accession numbers for assembled MAGs are listed in Table S2. Interpro annotation output is
577 available on figshare under DOI 10.6084/m9.figshare.12280313.

578 Acknowledgments

579 The project "SponGES" has received funding from the European Union's Horizon 2020
580 research and innovation programme under grant agreement No 679849. We thank our
581 colleague Hans Tore Rapp for leading this great consortium for the last four years – you are
582 dearly missed. Shiptime and Canadian participation was enabled by Fisheries and Oceans
583 Canada's (DFO) International Governance Strategy Science Program through project
584 "Marine Biological Diversity Beyond Areas of National Jurisdiction (BBNJ): 3-Tiers of

585 Diversity (Genes-Species-Communities)" led by EK (2017-2019). We appreciated the
586 onboard support of the crew and scientific party of the expedition *MLB2017001*. Fred
587 Whoriskey provided the OTN mooring samples. Andrea Hethke, Ina Clefsen, and the
588 CRC1182 Z3 team (Katja Cloppenborg-Schmidt, Malte Rühlemann, John Baines) provided
589 valuable support with the amplicon pipeline. Further, we want to thank the following people
590 for microscopy related support: Yu-Chen Wu, Marie Sieberns, Anke Bleyer, Cay Kruse and
591 Julia-Vanessa Böge.

592 # Legends

593 **Figures**

594 Figure 1 Map of sampling region on the Canadian shelf (A) and TS diagram (B). In (A)
595 colours depict presence probability of *Vazella pourtalesii* based on data presented in Beazley
596 et al. (17), with yellow indicating areas of highest occurrence probability. In B) colouring
597 corresponds to oxygen concentrations measured during representative CTD casts at the
598 study area. Water masses (light grey dots, labels and square) were added according to
599 Dever et al. (101) and (102): CBS = Cabot Strait Subsurface Water; InLC= Inshore Labrador
600 Current, CBS-CIL= Cold Intermediate Layer of Cabot Strait Subsurface Water, LSW=
601 Labrador Slope Water, WSW=Warm Slope Water.

602 Figure 2 Microscopy of *Vazella pourtalesii* tissue. A) Scanning electron microscopy overview
603 of spicule scaffolds (scale bar: 75 µm). B) SEM close-up image of a biomass patch (scale
604 bar: 3 µm). C) Light-microscopy image (scale bar: 5 µm) and D) TEM image of the same
605 biomass patch (scale bar: 1 µm). E) SEM close-up presumably showing smaller microbes
606 attached to larger ones by stalk- or filament-like structures (scale bar: 1 µm). F) TEM
607 microscopy images of adjacent microbial cells (scale bars: 500 nm). Acronyms: sp= spicule,
608 bp = biomass patch, hc = host cell, mic= microbes, p = potential pilus.

609 Figure 3 Subtrees of the GTDB-Tk phylogenetic tree showing the setting of the MAGs of this
610 study (shown in bold) within the microbial phyla selected for detailed analysis. Class names

611 are indicated by a leading “c_”, order names by “o_”, family names by “f_”, and genus names
612 by “g_”. Vazella-enriched MAGs are marked with a red star.

613 Figure 4 Comparison of the MAGs retrieved in this study to published MAGs from sponge
614 and seawater metagenomes. From this study, only the MAGs enriched in either *Vazella*
615 *pourtalesii* or in water were considered – ‘neutral’ ones were excluded from this analysis.

616 Figure 5 Reconstruction of metabolic features found in the genomes of A) SAR324, B)
617 Crenarchaeota, C) Patescibacteria, and D) Nanoarchaeota. Solid lines indicate that
618 genes/enzymes, or < 50% of a given pathway were found, dashed lines indicate less than
619 50% of a pathway were found. Grey arrows, writing and lining indicate that the
620 genes/enzymes were found in less than 50% of genomes of the respective phylum. White
621 arrows and writing indicate missing genes/enzymes. Co-factor synthesis is indicated by
622 turquoise colour, amino acid production by magenta colour. Symport, antiport and direction
623 are indicated by number and direction of arrows.

624 Figure 6 Summary model of the main metabolic interactions between the four microbial taxa
625 studied herein. CoA: acetyl-CoA; CH: carbohydrates; FA: fatty acids; F420: coenzyme F420;
626 AS: amino acids; FAD: flavin adenine dinucleotide; FMN: flavin mononucleotide; NAD:
627 nicotinamide adenine dinucleotide; NADP: nicotinamide adenine dinucleotide phosphate.

628 **Tables**

629 Table 1 MAGs of the bacterial candidate phyla Patescibacteria and SAR324, and the
630 archaeal phyla Crenarchaeota and Nanoarchaeota selected for detailed functional analysis.
631 Genome properties were determined by QUAST, completeness and contamination
632 estimations were performed by CheckM implemented in the metaWRAP pipeline. Acronyms:
633 “Cov” – genome coverage, “Red” – redundancy.

634 **Supplementary material**

635 Figure S1 Linear discriminant analysis (LDA) Effect Size (LEfSe) plots of MAG abundance
636 based on the abundance table calculated from read coverage data by the metaWRAP

637 quant_bins module. Two sets of groups were analyzed: A) *Vazella* metagenomes vs.
638 seawater reference metagenomes, and B) pristine *Vazella*-derived metagenomes (vazella_p)
639 vs. mooring *Vazella*-derived metagenomes (vazella_m) vs. seawater reference
640 metagenomes. An LDA score of 2 was selected as cut-off.

641 Figure S2 The 15 most abundant microbial phyla (classes for Proteobacteria) in *V.*
642 *pourtalesii*, representing > 99 % of the total microbiome. Microbial taxa are sorted after
643 relative abundance in descending order from left to right. Symbols represent domain level
644 taxonomic classification: circles indicate Bacteria, squares indicate Archaea. Patescibacteria
645 (P), SAR324, and Nanoarchaeota (N) are particularly highlighted. The pie chart (grey)
646 around the bubble chart indicates percentages of highest achievable taxonomic resolution for
647 each taxonomically assignable Amplicon Sequence Variant (ASV). The asterisk highlights
648 the three small pies next to it, which represent percentages of ASVs with highest achievable
649 taxonomic resolution on phylum-level and kingdom-level, as well as sequences unassigned
650 on phylum-level.

651 Figure S3 Heatmap of relative abundances [%] in our dataset of Patescibacteria classes
652 within the phylum. Bubbles on the left side of the heatmap show richness (ASVs) of each
653 class (max.=Parcubacteria 223 ASVs). Purple bars above the heatmap indicated relative
654 abundance [%] of the phylum Patescibacteria within total microbiomes.

655 Table S1 Samples of this study and applied analyses.

656 Table S2 Overview of the MAGs binned from the *Vazella pourtalesii* metagenome in this
657 study. Enrichment in one of the groups or all *Vazella* vs. water was determined by LefSe,
658 classification is derived from GTDB-Tk, completeness and contamination were determined
659 wich CheckM.

660 Table S3 Reference genomes for size and GC comparison.

661 Table S4 ANI analysis.

662 Text S1 Detailed description of metabolic features detected in SAR324 and Crenarchaeota.

663 ## References

664 1. Maldonado M, Aguilar R, Bannister RJ, Bell JJ, Conway KW, Dayton PK, Díaz C, Gutt
665 J, Kelly M, Kenchington ELR, Leys SP, Pomponi SA, Rapp HT, Rützler K, Tendal OS,
666 Vacelet J, Young CM. 2015. Sponge grounds as key marine habitats: A synthetic
667 review of types, structure, functional roles, and conservation concerns, p. 1-39. *In*
668 Rossi, S, Bramanti, L, Gori, A, Orejas, C (eds.), Marine Animal Forests. Springer
669 International Publishing, Cham.

670 2. Howell K-L, Piechaud N, Downie A-L, Kenny A. 2016. The distribution of deep-sea
671 sponge aggregations in the North Atlantic and implications for their effective spatial
672 management. Deep Sea Res Part I Oceanogr Res Pap 115:309–320.

673 3. Beazley L, Kenchington E, Yashayaev I, Murillo FJ. 2015. Drivers of epibenthic
674 megafaunal composition in the sponge grounds of the Sackville Spur, northwest
675 Atlantic. Deep Res Part I Oceanogr Res Pap 98:102–114.

676 4. Hawkes N, Korabik M, Beazley L, Rapp HT, Xavier JR, Kenchington E. 2019. Glass
677 sponge grounds on the Scotian Shelf and their associated biodiversity. Mar Ecol Prog
678 Ser 614:91–109.

679 5. Murillo FJ, Kenchington E, Koen-Alonso M, Guijarro J, Kenchington TJ, Sacau M,
680 Beazley L, Rapp HT. 2020. Mapping benthic ecological diversity and interactions with
681 bottom-contact fishing on the Flemish Cap (northwest Atlantic). Ecol Indic 112:106135.

682 6. Thomas T, Moitinho-Silva L, Lurgi M, Björk JR, Easson C, Astudillo-García C, Olson
683 JB, Erwin PM, López-Legentil S, Luter H, Chaves-Fonnegra A, Costa R, Schupp PJ,
684 Steindler L, Erpenbeck D, Gilbert J, Knight R, Ackermann G, Lopez JV, Taylor MW,
685 Thacker RW, Montoya JM, Hentschel U, Webster NS. 2016. Diversity, structure and
686 convergent evolution of the global sponge microbiome. Nat Commun 7:11870.

687 7. Pita L, Rix L, Slaby BM, Franke A, Hentschel U. 2018. The sponge holobiont in a
688 changing ocean: from microbes to ecosystems. Microbiome 6:46.

689 8. Moitinho-Silva L, Nielsen S, Amir A, Gonzalez A, Ackermann GL, Cerrano C, Astudillo-
690 García C, Easson C, Sipkema D, Liu F, Steinert G, Kotoulas G, McCormack GP, Feng
691 G, Bell JJ, Vicente J, Björk JR, Montoya JM, Olson JB, Reveillaud J, Steindler L,
692 Pineda M-C, Marra MV, Ilan M, Taylor MW, Polymenakou P, Erwin PM, Schupp PJ,
693 Simister RL, Knight R, Thacker RW, Costa R, Hill RT, Lopez-Legentil S, Dailianis T,
694 Ravasi T, Hentschel U, Li Z, Webster NS, Thomas T. 2017. The sponge microbiome
695 project. Gigascience 6:gix077.

696 9. Maldonado M, Ribes M, van Duyl FC. 2012. Nutrient fluxes through sponges: Biology,
697 budgets, and ecological implications, p. 113-182. *In* Becerro M, Uriz M, Maldonado M,
698 Turon X (eds.), Advances in Marine Biology, Volume 62. Elsevier Ltd. Academic
699 Press, Amsterdam.

700 10. de Goeij JM, van Oevelen D, Vermeij MJA, Osinga R, Middelburg JJ, de Goeij AFPM,
701 Admiraal W. 2013. Surviving in a marine desert: The sponge loop retains resources
702 within coral reefs. Science 342:108–110.

703 11. Krautter M, Conway KW, Barrie JV, Neuweiler M. 2001. Discovery of a “Living
704 Dinosaur”: Globally unique modern hexactinellid sponge reefs off British Columbia,
705 Canada. Facies 44:265–282.

706 12. Leys SP, Mackie GO, Reiswig HM. 2007. The Biology of Glass Sponges, p. 1-145. *In*
707 Sims D (ed.), Advances in Marine Biology, Volume 52. Elsevier Ltd. Academic Press.

708 13. van Soest RWM, Boury-Esnault N, Vacelet J, Dohrmann M, Erpenbeck D, de Voogd
709 NJ, Santodomingo N, Vanhoorne B, Kelly M, Hooper JNA. 2012. Global diversity of
710 sponges (Porifera). PLoS One 7:e35105.

711 14. Steinert G, Busch K, Bayer K, Kodami S, Arbizu PM, Kelly M, Mills S, Erpenbeck D,

712 Dohrmann M, Wörheide G, Hentschel U, Schupp PJ. 2020. Compositional and
713 Quantitative Insights Into Bacterial and Archaeal Communities of South Pacific Deep-
714 Sea Sponges (Demospongiae and Hexactinellida). Front Microbiol 11:716.

715 15. Tian R-M, Sun J, Cai L, Zhang W-P, Zhou G-W, Qiu J-W, Qian P-Y. 2016. The deep-
716 sea glass sponge *Lophophysema eversa* harbours potential symbionts responsible for
717 the nutrient conversions of carbon, nitrogen and sulfur. Environ Microbiol 18:2481–
718 2494.

719 16. Schmidt O. 1870. Grundzüge einer Spongien-Fauna des atlantischen Gebietes.
720 Wilhelm Engelmann, Leipzig.

721 17. Beazley L, Wang Z, Kenchington E, Yashayaev I, Rapp HT, Xavier JR, Murillo FJ,
722 Fenton D, Fuller S. 2018. Predicted distribution of the glass sponge *Vazella pourtalesi*
723 on the Scotian Shelf and its persistence in the face of climatic variability. PLoS One
724 13:e0205505.

725 18. Townsend DW, Pettigrew NR, Thomas MA, Neary MG, McGillicuddy Jr. DJ, O'Donnell
726 J. 2015. Water masses and nutrient sources to the Gulf of Maine. J Mar Res 73:93–
727 122.

728 19. Hendry KR, Cassarino L, Bates SL, Culwick T, Frost M, Goodwin C, Howell KL. 2019.
729 Silicon isotopic systematics of deep-sea sponge grounds in the North Atlantic. Quat
730 Sci Rev 210:1–14.

731 20. Busch K, Beazley L, Kenchington E, Whoriskey F, Slaby B, Hentschel U. Microbial
732 diversity of the glass sponge *Vazella pourtalesii* in response to anthropogenic
733 activities. *In review* at Conservation Genetics. bioRxiv preprint DOI
734 10.1101/2020.05.19.102806.

735 21. Dohrmann M, Wörheide G. 2017. Dating early animal evolution using phylogenomic
736 data. Sci Rep 7:3599.

737 22. Mentel M, Röttger M, Leys S, Tielens AGM, Martin WF. 2014. Of early animals,
738 anaerobic mitochondria, and a modern sponge. BioEssays 36:924–932.

739 23. Mills DB, Francis WR, Vargas S, Larsen M, Elemans CP, Canfield DE, Wörheide G.
740 2018. The last common ancestor of animals lacked the HIF pathway and respired in
741 low-oxygen environments. eLife 7:e31176.

742 24. Leys SP, Kahn AS. 2018. Oxygen and the Energetic Requirements of the First
743 Multicellular Animals. Integr Comp Biol 58:666–676.

744 25. Hofmann AF, Peltzer ET, Walz PM, Brewer PG. 2011. Hypoxia by degrees:
745 Establishing definitions for a changing ocean. Deep Sea Res Part I Oceanogr Res Pap
746 58:1212–1226.

747 26. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A,
748 Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny
749 substantially revises the tree of life. Nat Biotechnol 36:996-1004.

750 27. Hallam SJ, Konstantinidis KT, Putnam N, Schleper C, Watanabe Y, Sugahara J,
751 Preston C, de la Torre J, Richardson PM, DeLong EF. 2006. Genomic analysis of the
752 uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. Proc Natl Acad Sci U S
753 A 103:18296–18301.

754 28. Park S-J, Kim J-G, Jung M-Y, Kim S-J, Cha I-T, Ghai R, Martín-Cuadrado A-B,
755 Rodríguez-Valera F, Rhee S-K. 2012. Draft Genome Sequence of an Ammonia-
756 Oxidizing Archaeon, "*Candidatus* Nitrosopumilus sediminis" AR2, from Svalbard in the
757 Arctic Circle. J Bacteriol 194:6948-6949.

758 29. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN,
759 Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled
760 genomes substantially expands the tree of life. Nat Microbiol 2:1533–1542.

761 30. Dudek NK, Sun CL, Burstein D, Kantor RS, Aliaga Goltsman DS, Bik EM, Thomas BC,

762 Banfield JF, Relman DA. 2017. Novel Microbial Diversity and Functional Potential in
763 the Marine Mammal Oral Microbiome. Curr Biol 27:3752-3762.

764 31. White D, Drummond J, Fuqua C. 2012. The Physiology and Biochemistry of
765 Prokaryotes. Oxford University Press.

766 32. Unden G, Strecker A, Kleefeld A, Kim O Bin. 2016. C4-Dicarboxylate Utilization in
767 Aerobic and Anaerobic Growth. Ecosal Plus 7:1-33.

768 33. Rosa LT, Bianconi ME, Thomas GH, Kelly DJ. 2018. Tripartite ATP-Independent
769 Periplasmic (TRAP) Transporters and Tripartite Tricarboxylate Transporters (TTT):
770 From Uptake to Pathogenicity. Front Cell Infect Microbiol 8:33.

771 34. Draskovic I, Dubnau D. 2005. Biogenesis of a putative channel protein, ComEC,
772 required for DNA uptake: membrane topology, oligomerization and formation of
773 disulphide bonds. Mol Microbiol 55:881–896.

774 35. Sato T, Atomi H, Imanaka T. 2007. Archaeal type III RuBisCOs function in a pathway
775 for AMP metabolism. Science 315:1003–1006.

776 36. Waters E, Hohn MJ, Ahel I, Graham DE, Adams MD, Barnstead M, Beeson KY, Bibbs
777 L, Bolanos R, Keller M, Kretz K, Lin X, Mathur E, Ni J, Podar M, Richardson T, Sutton
778 GG, Simon M, Söll D, Stetter KO, Short JM, Noordewier M. 2003. The genome of
779 *Nanoarchaeum equitans*: Insights into early archaeal evolution and derived parasitism.
780 Proc Natl Acad Sci 100:12984–12988.

781 37. Podar M, Anderson I, Makarova KS, Elkins JG, Ivanova N, Wall MA, Lykidis A,
782 Mavromatis K, Sun H, Hudson ME, Chen W, Deciu C, Hutchison D, Eads JR,
783 Anderson A, Fernandes F, Szeto E, Lapidus A, Kyrpides NC, Saier MH, Richardson
784 PM, Rachel R, Huber H, Eisen JA, Koonin EV, Keller M, Stetter KO. 2008. A genomic
785 analysis of the archaeal system *Ignicoccus hospitalis-Nanoarchaeum equitans*.
786 Genome Biol 9:R158.

787 38. Bardy SL, Jarrell KF. 2003. Cleavage of preflagellins by an aspartic acid signal
788 peptidase is essential for flagellation in the archaeon *Methanococcus voltae*. Mol
789 Microbiol 50:1339–1347.

790 39. Gloeckner V, Wehrl M, Moitinho-Silva L, Gernert C, Schupp P, Pawlik JR, Lindquist
791 NL, Erpenbeck D, Wörheide G, Hentschel U. 2014. The HMA-LMA dichotomy
792 revisited: An electron microscopical survey of 56 sponge species. Biol Bull 227:78–88.

793 40. Moreno-Pino M, Cristi A, Gillooly JF, Trefault N. 2020. Characterizing the microbiomes
794 of Antarctic sponges: a functional metagenomic approach. Sci Rep 10:645.

795 41. Horn H, Slaby BM, Jahn MT, Bayer K, Moitinho-Silva L, Förster F, Abdelmohsen UR,
796 Hentschel U. 2016. An enrichment of CRISPR and other defense-related features in
797 marine sponge-associated microbial metagenomes. Front Microbiol 7:1751.

798 42. Slaby BM, Hackl T, Horn H, Bayer K, Hentschel U. 2017. Metagenomic binning of a
799 marine sponge microbiome reveals unity in defense but metabolic specialization.
800 ISME J 11:2465–2478.

801 43. Kamke J, Sczyrba A, Ivanova N, Schwientek P, Rinke C, Mavromatis K, Woyke T,
802 Hentschel U. 2013. Single-cell genomics reveals complex carbohydrate degradation
803 patterns in poribacterial symbionts of marine sponges. ISME J 7:2287–2300.

804 44. Qin Q-L, Li Y, Sun L-L, Wang Z-B, Wang S, Chen X-L, Oren A, Zhang Y-Z. 2019.
805 Trophic Specialization Results in Genomic Reduction in Free-Living Marine *Idiomarina*
806 Bacteria. mBio 10:e02545-18.

807 45. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J,
808 Richardson TH, Noordewier M, Rappé MS, Short JM, Carrington JC, Mathur EJ. 2005.
809 Genome Streamlining in a Cosmopolitan Oceanic Bacterium. Science 309:1242–1245.

810 46. Morris JJ, Lenski RE, Zinser ER. 2012. The Black Queen Hypothesis: Evolution of
811 Dependencies through Adaptive Gene Loss. mBio 3:e00036-12.

812 47. Dutta C, Paul S. 2012. Microbial lifestyle and genome signatures. Curr Genomics
813 13:153–162.

814 48. Bayer K, Schmitt S, Hentschel U. 2008. Physiology, phylogeny and *in situ* evidence for
815 bacterial and archaeal nitrifiers in the marine sponge *Aplysina aerophoba*. Environ
816 Microbiol 10:2942–2955.

817 49. Moeller FU, Webster NS, Herbold CW, Behnam F, Domman D, Albertsen M,
818 Mooshammer M, Markert S, Turaev D, Becher D, Rattei T, Schweder T, Richter A,
819 Watzka M, Nielsen PH, Wagner M. 2019. Characterization of a thaumarchaeal
820 symbiont that drives incomplete nitrification in the tropical sponge *Ianthella basta*.
821 Environ Microbiol 21:3831-3854.

822 50. Moitinho-Silva L, Díez-Vives C, Batani G, Esteves AIS, Jahn MT, Thomas T. 2017.
823 Integrated metabolism in sponge–microbe symbiosis revealed by genome-centered
824 metatranscriptomics. ISME J 11:1–16.

825 51. Zhang S, Song W, Wemheuer B, Reveillaud J, Webster N, Thomas T. 2019.
826 Comparative Genomics Reveals Ecological and Evolutionary Insights into Sponge-
827 Associated Thaumarchaeota. mSystems 4:e00288-19.

828 52. Engelberts JP, Robbins SJ, de Goeij JM, Aranda M, Bell SC, Webster NS. 2020.
829 Characterization of a sponge microbiome using an integrative genome-centric
830 approach. ISME J 14:1100-1110.

831 53. Turon M, Uriz MJ. 2020. New Insights Into the Archaeal Consortium of Tropical
832 Sponges. Front Mar Sci 6:789.

833 54. Berg IA. 2011. Ecological aspects of the distribution of different autotrophic CO_2
834 fixation pathways. Appl Environ Microbiol 77:1925–36.

835 55. Bayer B, Vojvoda J, Offre P, Alves RJE, Elisabeth NH, Garcia JAL, Volland J-M,
836 Srivastava A, Schleper C, Herndl GJ. 2016. Physiological and genomic

837 characterization of two novel marine thaumarchaeal strains indicates niche
838 differentiation. ISME J 10:1051–1063.

839 56. Hügler M, Sievert SM. 2011. Beyond the Calvin Cycle: Autotrophic Carbon Fixation in
840 the Ocean. Ann Rev Mar Sci 3:261–289.

841 57. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling AE,
842 Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu
843 W-T, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P,
844 Woyke T. 2013. Insights into the phylogeny and coding potential of microbial dark
845 matter. Nature 499:431–437.

846 58. Woyke T, Doud DFR, Eloe-Fadrosh EA. 2019. Genomes From Uncultivated
847 Microorganisms, p. 437-442. *In* Encyclopedia of Microbiology. Elsevier.

848 59. Castelle CJ, Brown CT, Anantharaman K, Probst AJ, Huang RH, Banfield JF. 2018.
849 Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN
850 radiations. Nat Rev Microbiol 16:629–645.

851 60. Forterre P, Gribaldo S, Brochier-Armanet C. 2009. Happy together: genomic insights
852 into the unique *Nanoarchaeum*/*Ignicoccus* association. J Biol 8:7.

853 61. Nicks T, Rahn-Lee L. 2017. Inside Out: Archaeal Ectosymbionts Suggest a Second
854 Model of Reduced-Genome Evolution. Front Microbiol 8:384.

855 62. Jahn U, Summons R, Sturt H, Grosjean E, Huber H. 2004. Composition of the lipids of
856 *Nanoarchaeum equitans* and their origin from its host *Ignicoccus* sp. strain KIN4/I.
857 Arch Microbiol 182:404–413.

858 63. Albers S-V, Meyer BH. 2011. The archaeal cell envelope. Nat Rev Microbiol 9:414–
859 426.

860 64. Mell JC, Redfield RJ. 2014. Natural Competence and the Evolution of DNA Uptake
861 Specificity. J Bacteriol 196:1471–1483.

862 65. Miller MB, Bassler BL. 2001. Quorum Sensing in Bacteria. Annu Rev Microbiol
863 55:165–199.

864 66. Chun CK, Troll JV, Koroleva I, Brown B, Manzella L, Snir E, Almabrazi H, Scheetz TE,
865 de Fatima Bonaldo M, Casavant TL, Soares MB, Ruby EG, McFall-Ngai MJ. 2008.
866 Effects of Colonization, Luminescence, and Autoinducer on Host Transcription During
867 Development of the Squid-*Vibrio* Association. Proc Natl Acad Sci 105:11323–11328.

868 67. Zan J, Fuqua C, Hill RT. 2011. Diversity and functional analysis of *luxS* genes in
869 Vibrios from marine sponges *Mycale laxissima* and *Ircinia strobilina*. ISME J 5:1505–
870 1516.

871 68. Ng W-L, Bassler BL. 2009. Bacterial Quorum-Sensing Network Architectures. Annu
872 Rev Genet 43:197–222.

873 69. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. 2013. Development
874 of a dual-index sequencing strategy and curation pipeline for analyzing amplicon
875 sequence data on the miseq illumina sequencing platform. Appl Environ Microbiol
876 79:5112–5120.

877 70. Muyzer G, de Waal EC, Uitterlinden AG. 1993. Profiling of complex microbial
878 populations by denaturing gradient gel electrophoresis analysis of polymerase chain
879 reaction-amplified genes coding for 16S rRNA. Appl Environ Microbiol 59:695–700.

880 71. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ,
881 Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of millions
882 of sequences per sample. Proc Natl Acad Sci 108:4516–4522.

883 72. Bushnell B. 2017. BBMap short read aligner, and other bioinformatic tools. Available
884 online at: https://sourceforge.net/projects/bbmap/.

885 73. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H,
886 Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn,

887 CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, DaSilva R,
888 Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst
889 M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K,
890 Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S,
891 Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights
892 D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu Y-X, Loftfield E,
893 Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV,
894 Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF,
895 Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen
896 LB, Rivers A, Robeson II MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R,
897 Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A,
898 Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y,
899 Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC,
900 Williamson CHD, Willis AD, Zech Xu Z, Zaneveld JR, Zhang Y, Zhu Q, Knight R,
901 Caporaso JG. 2019. Reproducible, interactive, scalable, and extensible microbiome
902 data science using QIIME 2. Nat Biotechnol 37:852-857.

903 74. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016.
904 DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods
905 13:581–583.

906 75. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA,
907 Caporaso JG. 2018. Optimizing taxonomic classification of marker-gene amplicon
908 sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6:90.

909 76. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner
910 FO. 2013. The SILVA ribosomal RNA gene database project: Improved data
911 processing and web-based tools. Nucleic Acids Res 41:590–596.

912 77. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-
913 Likelihood Trees for Large Alignments. PLoS One 5:e9490.

914 78. R Core Team. 2008. R: A language and environment for statistical computing. R
915 Foundation for Statistical Computing, Vienna, Austria.

916 79. Harrington B, Team and the D. 2005. Inkscape. Available online at:
917 http://www.inkscape.org/.

918 80. QGIS Development Team. 2017. Geographic Information System. Open Source
919 Geospatial Foundation Project. Available online at: http://qgis.osgeo.org.

920 81. McDougall TJ, Barker PM. 2011. Getting started with TEOS-10 and the Gibbs
921 Seawater (GSW) Oceanographic Toolbox. SCOR/IAPSO Working Group Rep.

922 82. Aescht E, Büchl-Zimmermann S, Burmester A, Dänhardt-Pfeiffer S, Desel C, Hamers
923 C, Jach G, Kässens M, Makovitzky J, Mulisch M, Nixdorf-Bergweiler B, Pütz D,
924 Riedelsheimer B, van den Boom F, Wegerhoff R, Welsch U. 2010. Romeis
925 Mikroskopische Technik. Springer Spektrum. Spektrum Akademischer Verlag,
926 Heidelberg.

927 83. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data.
928 Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

929 84. Li D, Luo R, Liu C-M, Leung C-M, Ting H-F, Sadakane K, Yamashita H, Lam T-W.
930 2016. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by
931 advanced methodologies and community practices. Methods 102:3–11.

932 85. Uritskiy GV, DiRuggiero J, Taylor J. 2018. MetaWRAP—a flexible pipeline for
933 genome-resolved metagenomic data analysis. Microbiome 6:158.

934 86. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately
935 reconstructing single genomes from complex microbial communities. PeerJ 3:e1165.

936 87. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an
937 adaptive binning algorithm for robust and efficient genome reconstruction from
938 metagenome assemblies. PeerJ 7:e7359.

939 88. Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm
940 to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605–607.

941 89. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,
942 Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G,
943 Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and
944 its applications to single-cell sequencing. J Comput Biol 19:455–477.

945 90. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal:
946 Prokaryotic gene recognition and translation initiation site identification. BMC
947 Bioinformatics 11:119.

948 91. Eddy SR. 2011. Accelerated Profile HMM Searches. PLoS Comput Biol 7:e1002195.

949 92. Matsen FA, Kodner RB, Armbrust EV. 2010. pplacer: linear time maximum-likelihood
950 and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC
951 Bioinformatics 11:538.

952 93. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High
953 throughput ANI analysis of 90K prokaryotic genomes reveals clear species
954 boundaries. Nat Commun 9:5114.

955 94. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display
956 and annotation of phylogenetic and other trees. Nucleic Acids Res 44:W242–W245.

957 95. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C.
958 2011. Metagenomic biomarker discovery and explanation. Genome Biol 12:R60.

959 96. Delmont TO, Eren AM. 2018. Linking pangenomes and metagenomes: the
960 Prochlorococcus metapangenome. PeerJ 6:e4320.

961 97. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015.
962 Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ
963 3:e1319.

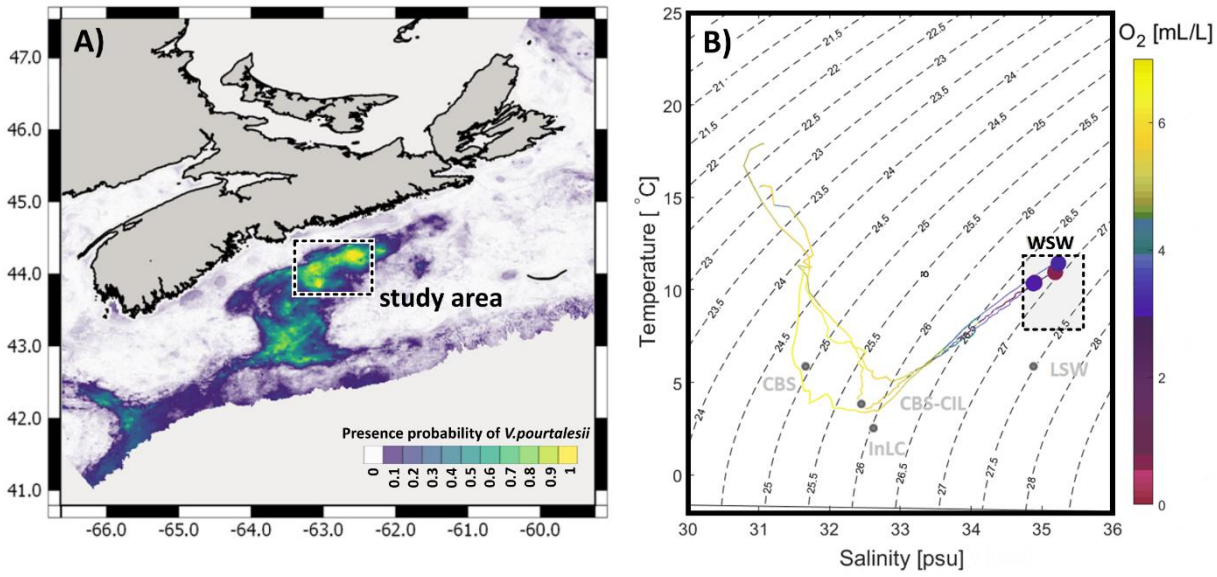
964 98. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J,
965 Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M,
966 Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: Genome-scale protein function
967 classification. Bioinformatics 30:1236–1240.

968 99. Sangrador-Vegas A, Mitchell AL, Chang H-Y, Yong S-Y, Finn RD. 2016. GO
969 annotation in InterPro: why stability does not indicate accuracy in a sea of changing
970 annotations. Database (Oxford) 2016:baw027.

971 100. Elbourne LDH, Tetu SG, Hassan KA, Paulsen IT. 2017. TransportDB 2.0: a database
972 for exploring membrane transporters in sequenced genomes from all domains of life.
973 Nucleic Acids Res 45:D320–D324.

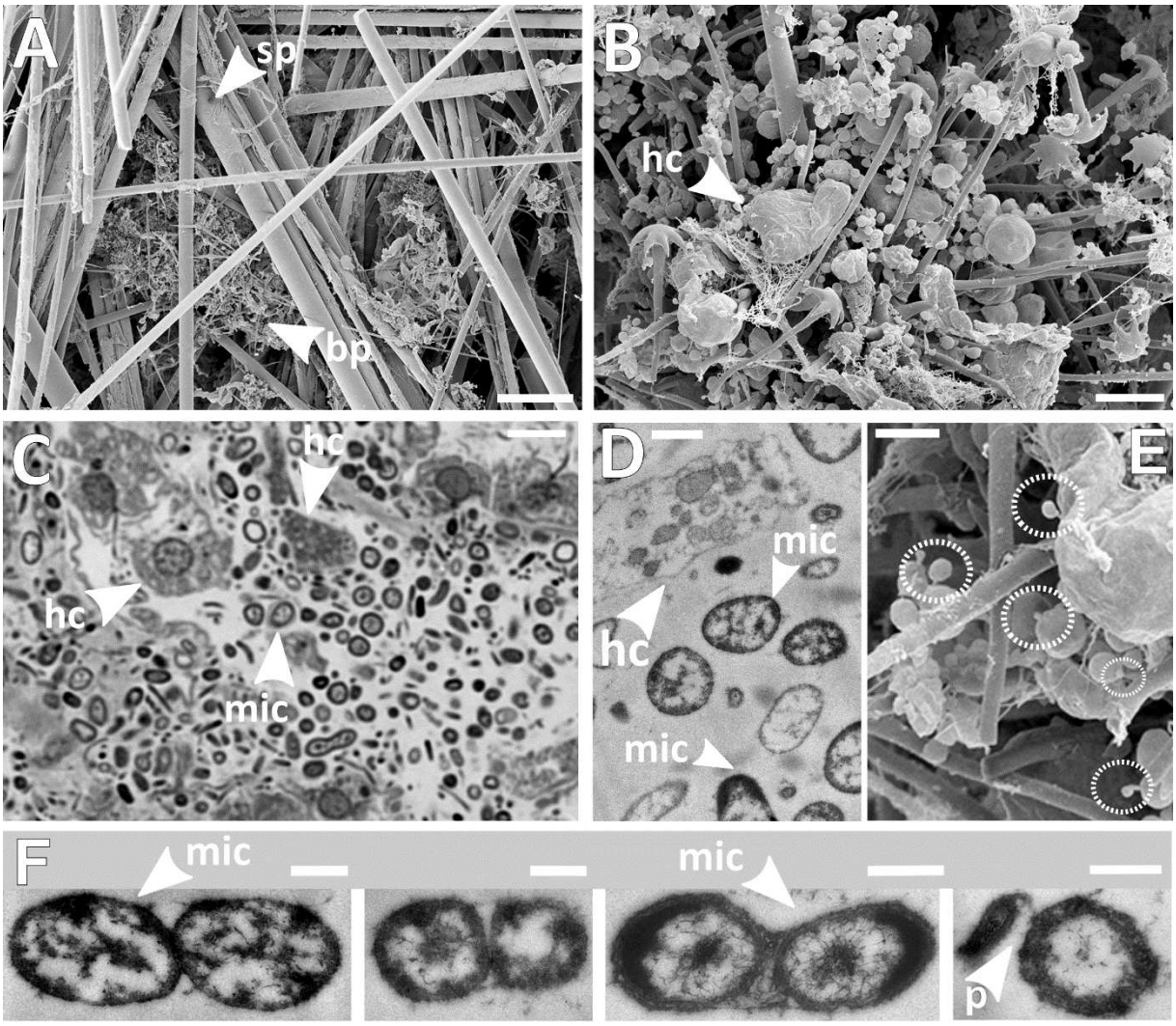
974 101. Dever M, Hebert D, Greenan BJW, Sheng J, Smith PC. 2016. Hydrography and
975 Coastal Circulation along the Halifax Line and the Connections with the Gulf of St.
976 Lawrence. Atmos - Ocean 54:199–217.

977 102. Fratantoni PS, Pickart RS. 2007. The Western North Atlantic Shelfbreak Current
978 System in Summer. J Phys Oceanogr 37:2509–2533.



979
980
981 Figure 1 Map of sampling region on the Canadian shelf (A) and TS diagram (B). In (A)
982 colours depict presence probability of *Vazella pourtalesii* based on data presented in Beazley

983 et al. (19), with yellow indicating areas of highest occurrence probability. In B) colouring
984 corresponds to oxygen concentrations measured during representative CTD casts at the
985 study area. Water masses (light grey dots, labels and square) were added according to
986 Dever et al. (103) and (104): CBS = Cabot Strait Subsurface Water; InLC= Inshore Labrador
987 Current, CBS-CIL= Cold Intermediate Layer of Cabot Strait Subsurface Water, LSW=
988 Labrador Slope Water, WSW=Warm Slope Water.



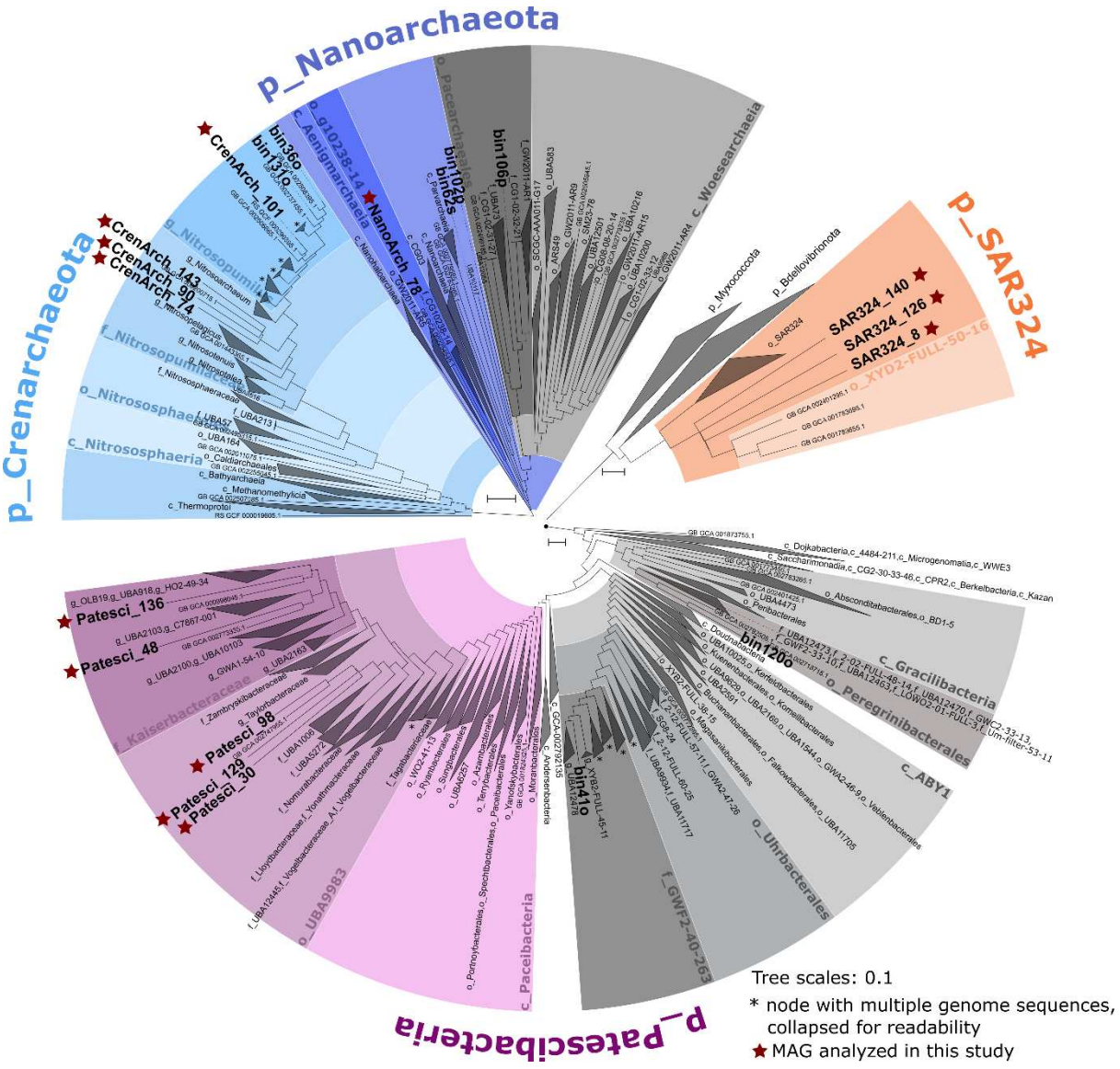
989

990 Figure 2 Microscopy of *Vazella pourtalesii* tissue. A) Scanning electron microscopy overview
991 of spicule scaffolds (scale bar: 75 µm). B) SEM close-up image of a biomass patch (scale
992 bar: 3 µm). C) Light-microscopy image (scale bar: 5 µm) and D) TEM image of the same
993 biomass patch (scale bar: 1 µm). E) SEM close-up presumably showing smaller microbes
994 attached to larger ones by stalk- or filament-like structures (scale bar: 1 µm). F) TEM

995 microscopy images of adjacent microbial cells (scale bars: 500 nm). Acronyms: sp= spicule,
996 bp = biomass patch, hc = host cell, mic= microbes, p = potential pilus.

997

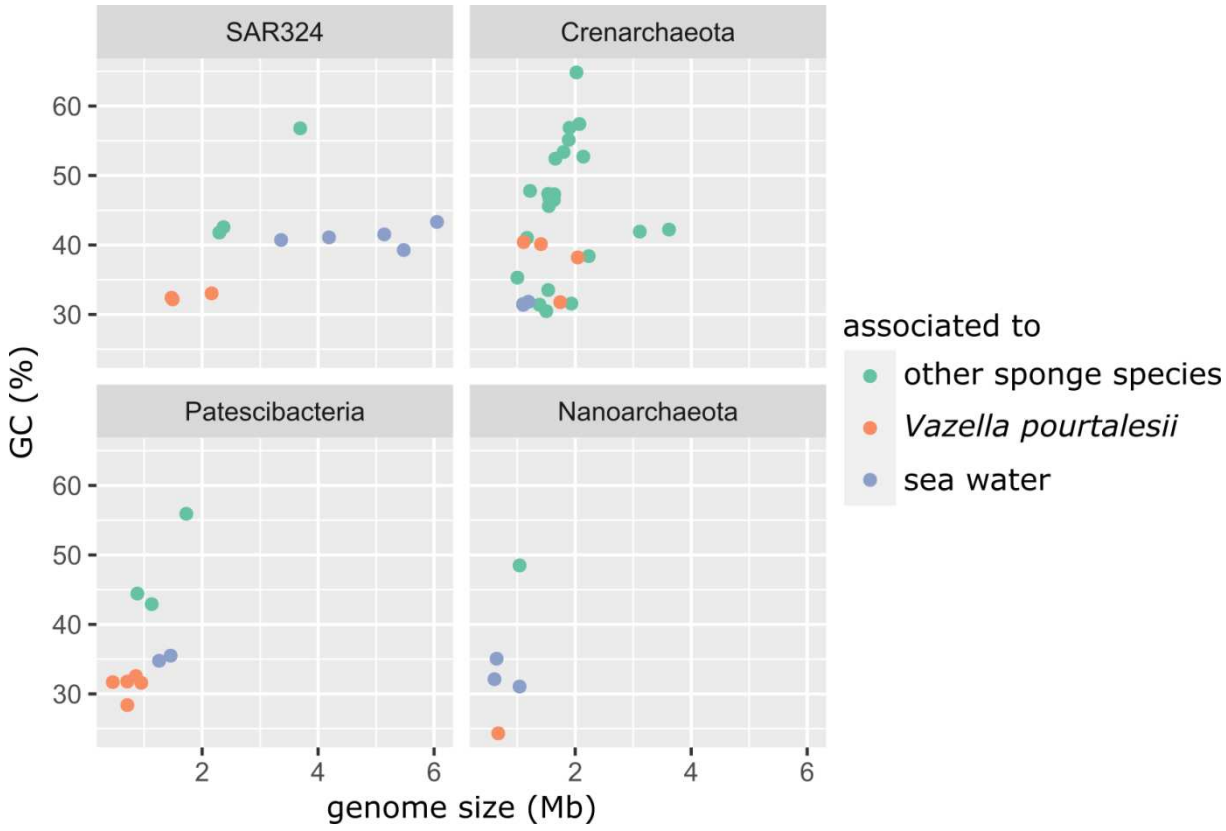
998



999

1000 Figure 3 Subtrees of the GTDB-Tk phylogenetic tree showing the setting of the MAGs of this
1001 study (shown in bold) within the microbial phyla selected for detailed analysis. Class names
1002 are indicated by a leading “c_”, order names by “o_”, family names by “f_”, and genus names
1003 by “g_”. Vazella-enriched MAGs are marked with a red star.

1004

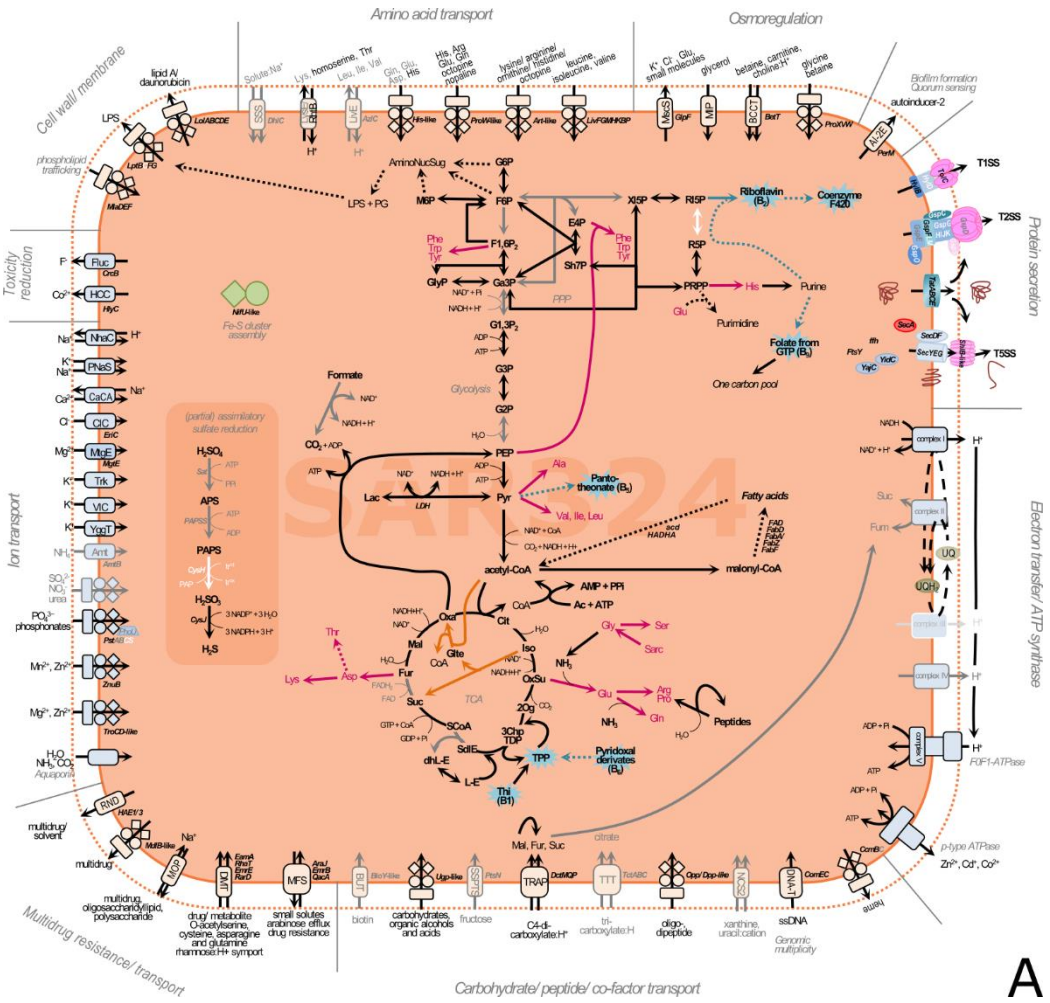


1005

1006 Figure 4 Comparison of the MAGs retrieved in this study to published MAGs from sponge
1007 and seawater metagenomes. From this study, only the MAGs enriched in either *Vazella*
1008 *pourtalesii* or in water were considered – 'neutral' ones were excluded from this analysis.

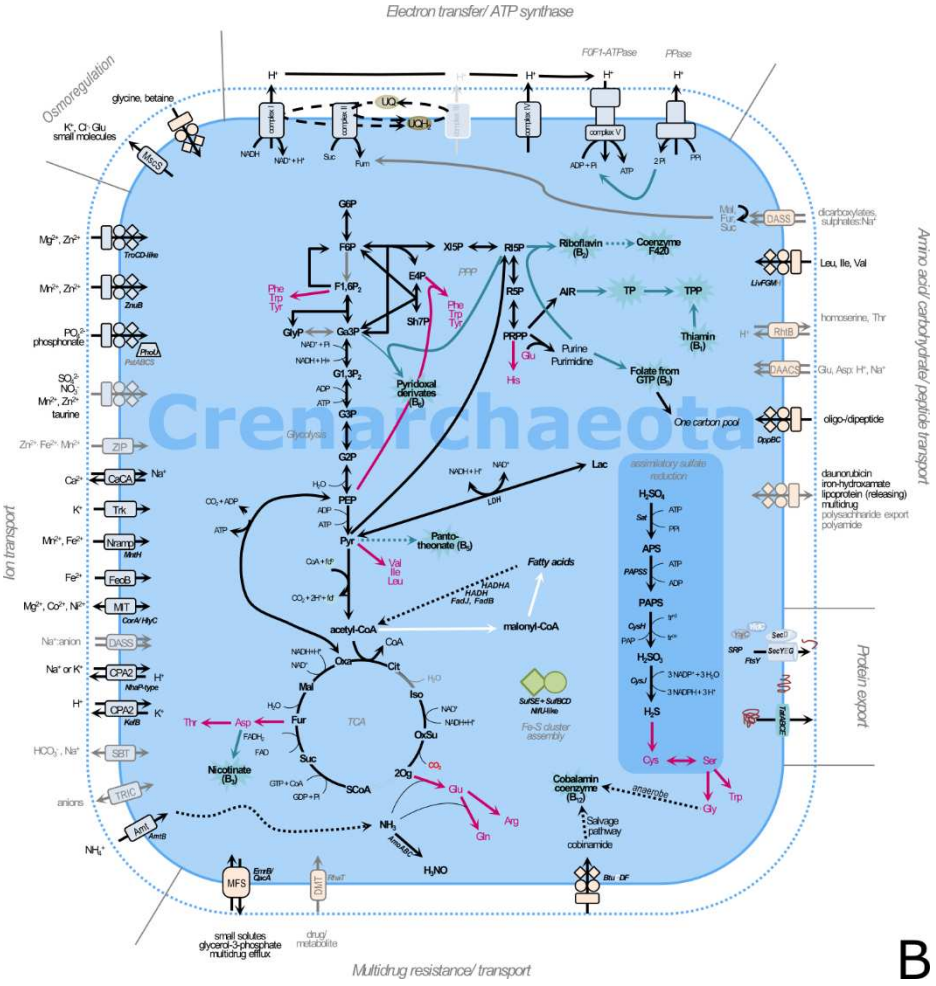
1009 Table 1 MAGs of the bacterial candidate phyla Patescibacteria and SAR324, and the archaeal phyla Crenarchaeota and Nanoarchaeota selected
1010 for detailed functional analysis. Genome properties were determined by QUAST, completeness and contamination estimations were performed by
1011 CheckM implemented in the metaWRAP pipeline. Acronyms: “Cov” – genome coverage, “Red” – redundancy.

Phylum-level affiliation	MAG	# contigs	estim. Genome size (Mb)	GC (%)	N50	Cov (%)	Red (%)
Patescibacteria	Patesci_129	109	0.71	31.8	5,408	67.37	0
	Patesci_30	95	0.46	31.7	5,089	66.19	1.18
	Patesci_136	206	0.95	31.6	4,840	61.21	0.47
	Patesci_48	163	0.86	32.6	8,405	71.52	0
	Patesci_98	15	0.71	28.4	66,386	79.44	0
SAR324 (Deltaproteo-bacteria)	SAR324_126	562	2.16	33.0	8,540	75.63	5.39
	SAR324_140	257	1.49	32.2	4,131	54.53	0.75
	SAR324_8	270	1.47	32.4	3,229	54.81	2.45
Crenarchaeota	CrenArch_143	507	2.04	38.2	3,080	63.44	1.46
	CrenArch_74	366	1.41	40.1	3,432	53.2	8.16
	CrenArch_90	115	1.11	40.4	8,261	59.77	2.59
	CrenArch_101	336	1.74	31.8	3,376	60.75	3.40
Nanoarchaeota	NanoArch_78	39	0.67	24.3	39,559	76.63	0

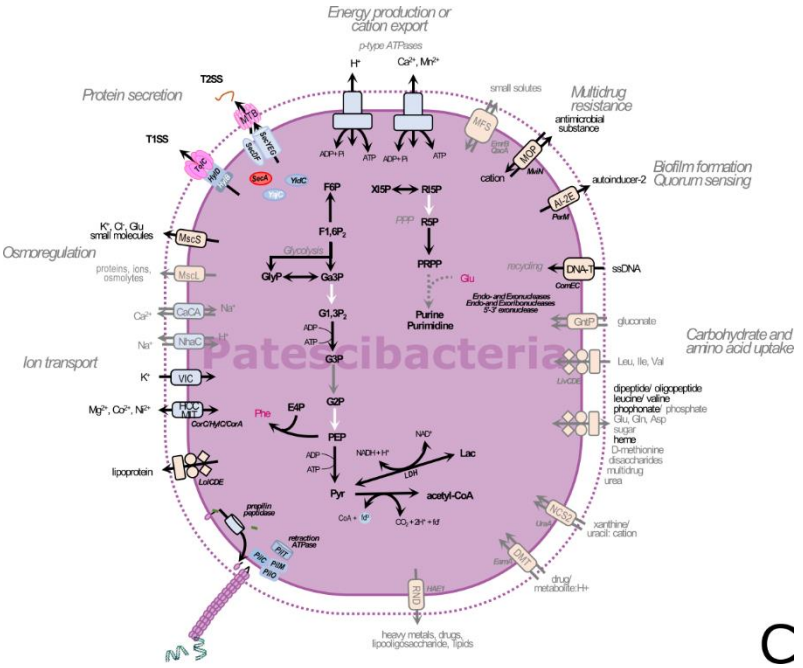


1012

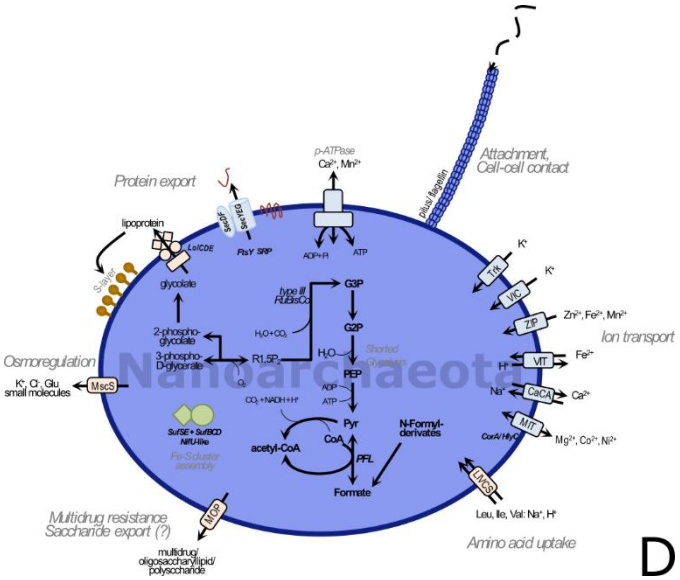
A



1013 B



1014 C

1015 

D

1016 Figure 5 Reconstruction of metabolic features found in the genomes of A) SAR324, B)
1017 Crenarchaeota, C) Patescibacteria, and D) Nanoarchaeota. Solid lines indicate that
1018 genes/enzymes, or < 50% of a given pathway were found, dashed lines indicate less than
1019 50% of a pathway were found. Grey arrows, writing and lining indicate that the
1020 genes/enzymes were found in less than 50% of genomes of the respective phylum. White
1021 arrows and writing indicate missing genes/enzymes. Co-factor synthesis is indicated by
1022 turquoise colour, amino acid production by magenta colour. Symport, antiport and direction
1023 are indicated by number and direction of arrows.

1024

1025

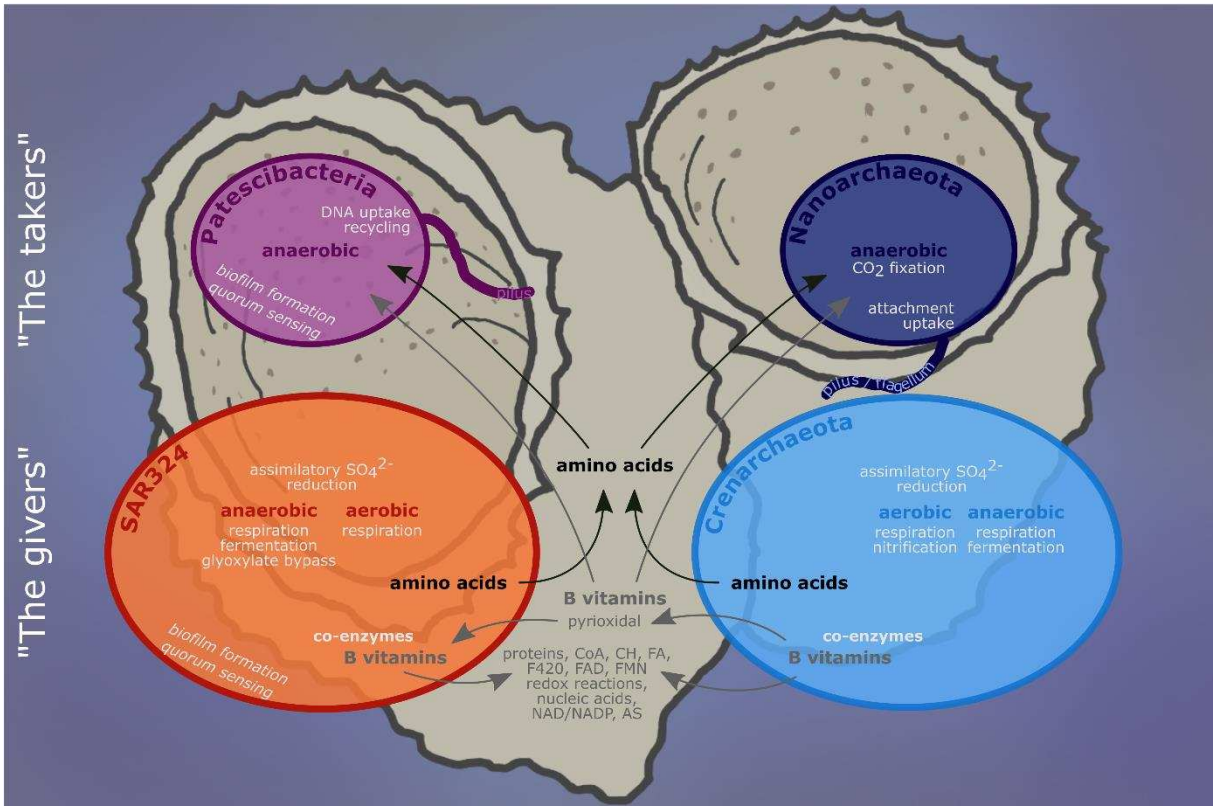


Figure 6 Summary model of the main metabolic interactions between the four microbial taxa studied herein. CoA: acetyl-CoA; CH: carbohydrates; FA: fatty acids; F420: coenzyme F420; AS: amino acids; FAD: flavin adenine dinucleotide; FMN: flavin mononucleotide; NAD: nicotinamide adenine dinucleotide; NADP: nicotinamide adenine dinucleotide phosphate.