GeoChip-based insights into the microbial functional gene repertoire of marine sponges (HMA, LMA) and seawater

Running title: Functional gene repertoire of sponge microorganisms

Kristina Bayer[1,#], Lucas Moitinho-Silva[1,#], Franz Brümmer[2], Carlo V. Cannistraci[3,4], Timothy Ravasi[3,*] and Ute Hentschel[1,*]

[#]authors contributed equally to this work

[1]Dept. of Botany II, Julius-von-Sachs Institute for Biological Sciences, University of Wuerzburg, Wuerzburg, Germany

[2]Biological Institute, Department of Zoology, University of Stuttgart, Stuttgart, Germany

[3]Division of Biological and Environmental Sciences & Engineering and Division of Applied Mathematics and Computer Science, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia

[4]Biomedical Cybernetics Group, Biotechnology Center (BIOTEC), Technische Universität Dresden, Dresden, Germany

*Please address correspondence to: U. Hentschel; Email: ute.hentschel@uni-wuerzburg.de; Tel.: +49-931-31-82581; Fax: +49-931-31-86235; T. Ravasi; Email: timothy.ravasi@kaust.edu.sa; Tel.: +966-2-808-2387; Fax: +966-2-8020127

23 Keywords: GeoChip, amplicon sequencing, marine sponge, Porifera, functional gene,
24 symbiosis

25 **Abstract**

26

27 The GeoChip 4 gene array was employed to interrogate the microbial functional gene
28 repertoire of sponges and seawater collected from the Red Sea and the Mediterranean.
29 Complementary amplicon sequencing confirmed the microbial community composition
30 characteristic of high microbial abundance (HMA) and low microbial abundance (LMA)
31 sponges. By use of GeoChip, altogether 20,273 probes encoding for 627 functional
32 genes and representing 16 gene categories were identified. Minimum curvilinear
33 embedding (MCE) analyses revealed a clear separation between the samples. The
34 HMA/LMA dichotomy was stronger than any possible geographic pattern, which is
35 shown here for the first time on the level of functional genes. However upon inspection
36 of individual genes, very few specific differences were discernible. Differences were
37 related to microbial ammonia oxidation, ammonification, and archaeal autotrophic
38 carbon fixation (higher gene abundance in sponges over seawater) as well as
39 denitrification and radiation-stress related genes (lower gene abundance in sponges
40 over seawater). Except for few documented specific differences, the functional gene
41 repertoire between the different sources appeared largely similar. This study expands
42 previous reports in that functional gene convergence is not only reported between HMA
43 and LMA sponges but also between sponges and seawater.

44

45

46

47

48 **Introduction**

49

50 Marine sponges (phylum Porifera) are important couplers of the benthic and pelagic
51 ecosystems by virtue of their immense filter-feeding capacities (Bell, 2008, de Goeij *et*
52 *al.*, 2013). Many marine sponges (hereafter referred to as high microbial abundance
53 sponges, HMA) host stable associations of symbiotic microbial consortia within their
54 mesohyl matrix that can contribute up to 35% of the animal's biomass (Hentschel *et al.*,
55 2012). Representatives of more than 28 bacterial phyla including new candidate phyla
56 such as *Poribacteria* and *Tectomicrobia* and two archaeal lineages were identified in
57 marine sponges so far (Lee *et al.*, 2011, Schmitt *et al.*, 2012a, Schmitt *et al.*, 2012b,
58 Simister *et al.*, 2012, Webster *et al.*, 2010, Wilson *et al.*, 2014). On the other hand, some
59 sponge species are essentially devoid of microorganisms (hereafter termed low
60 microbial abundance sponges, LMA) and the diversity is largely restricted to
61 *Proteobacteria* (*Alpha*-, *Gamma*-), *Cyanobacteria*, and *Archaea*. The classification into
62 these HMA or LMA sponges has long been recognized (Hentschel *et al.*, 2003, Reiswig,
63 1981, Vacelet & Donadey, 1977).

64

65 Our understanding of microbial functions in sponge symbiosis is still at the beginning
66 due to the fact that the vast majority of sponge-associated microorganisms have not
67 been cultured and due to the lack of an established model system for sponge-microbial
68 symbioses. However, as reviewed (Hentschel *et al.*, 2012, Taylor *et al.*, 2007, Webster
69 & Taylor, 2012), several publications have analyzed or inferred metabolic activities of
70 the sponge symbionts, for example as derived from pure-culture studies, analysis of

71 pathways by measurement of specific products and genes, and inference of metabolic
72 potentials based on 16S rRNA gene sequence analysis. A particular emphasis has been
73 placed on the metabolism of carbon, nitrogen, and sulfur as well as secondary
74 metabolism. Recently, the application of “omics” approaches has contributed to build an
75 integrated view of microbial function in the context of the host sponge. A recent
76 metagenomic study reported functional gene convergence in the microbiome of six
77 demosponge species (Fan *et al.*, 2012). Several metatranscriptomic/ metaproteomic
78 studies have provided first insights into the expressed functional gene repertoire of
79 sponge-associated microbial consortia (Liu *et al.*, 2012, Moitinho-Silva *et al.*, 2014b,
80 Radax *et al.*, 2012). Furthermore, single cell genomic studies have enabled to correlate
81 specific traits with specific symbiont lineages, thus providing a sought-after link between
82 phylogeny and function (Kamke *et al.*, 2013, Siegl *et al.*, 2011, Wilson *et al.*, 2014).
83 However, much remains to be learned about the functional basis of sponge-microbe
84 interactions.
85
86 Functional gene arrays, such as the GeoChip, are one efficient way to assess the
87 functional gene repertoire of environmental microbial communities (Nostrand *et al.*,
88 2012). Microarrays have been developed to probe a wide diversity of genes in various
89 ecological contexts. For example, functional gene arrays have been useful to establish
90 gene inventories (Chan *et al.*, 2013, Mason *et al.*, 2010), to assess differences between
91 field sites (Kang *et al.*, 2013), and to report shifts in response to environmental
92 parameters such as warming (Yergeau *et al.*, 2012), or following the deep-water Horizon
93 oil spill (Lu *et al.*, 2012). The GeoChip has so far been applied to natural soils and

94 sediments, in particular in the context of contamination and bioremediation, and at least
95 once before in a marine invertebrate, the tropical coral *Montastraea faviolata* (Kimes *et*
96 *al.*, 2010). Since its invention in 2004, the GeoChip has been constantly improved and
97 updated. The newest version, GeoChip 4, contains 83,992 probes targeting 152,414
98 genes representing a total 410 gene categories (Tu *et al.*, 2014). In comparison to
99 previous versions, the categories stress response, antibiotic resistance, and
100 bacteriophage genes have been added (Nostrand *et al.*, 2012). In the present study, we
101 employed the GeoChip 4, to our knowledge for the first time, to assess the functional
102 gene repertoire of marine sponge-associated microbiomes.
103
104
105 **Methods**
106
107 *Sample collection*
108 Five individuals of the Mediterranean sponges *Aplysina aerophoba* (HMA) and *Dysidea*
109 *avara* (LMA) were collected in April 2012 by scuba diving at a depth of 10-15 m off the
110 coast at Rovinj, Croatia (45°08'N; 13°64'E). The sponges *Xestospongia testudinaria*
111 (HMA) and *Stylissa carteri* (LMA), (n=5 each) were collected by scuba diving at Fsar
112 reef (22°23'N; 39°03'E) at a depth of 13-15 m off the coast of Thuwal, Saudi Arabia in
113 March 2012. The animals were brought to the surface in plastic bags to avoid contact
114 with air. Small tissue pieces were removed with a sterile scalpel, rinsed in 0.22 µm
115 filtered seawater, immediately frozen in liquid nitrogen and stored at -80 °C until use.
116 Sponge pieces were taken such that portions of pinacoderm and mesohyl were included

117 in each sample. Seawater was sampled using folded plastic canisters (camping store),
118 which were filled during the dives in the vicinity of the sponges. Approximately 7 L of
119 Mediterranean seawater were filtered onto 0.22 µm PES bottle top filters (Merck-
120 Millipore, Germany) using a manually operated vacuum pump. The same volume of Red
121 Sea seawater was filtered onto 0.22 µm hydrophilic Durapore membrane filters
122 (Millipore, USA) using a Masterflex Easy-Load peristaltic pump (Cole-Parmer, USA).
123 The filters were stored at -80 °C until further processing.
124
125 *DNA extraction*
126 Frozen sponge samples were ground to fine powder using a sterile mortar and pistil.
127 Genomic DNA was extracted using the Allprep DNA/ RNA mini kit (Qiagen, Germany)
128 following the manufacturer's instructions. DNA quality and concentrations were
129 assessed with the NanoDrop 2000c spectrophotometer (Peqlab, Germany) as well as
130 using the Qubit 2.0 fluorometer and the Qubit dsDNA BR Assay Kit (Life Technologies,
131 Germany). DNA was freeze-dried using the ALPHA 1-2 LD lyophilizer (Christ, Germany)
132 and shipped to Glomics Inc. (Norman, OK, USA), where the DNA was processed.
133
134 *Amplicon sequencing*
135 The 16S rRNA genes were amplified according to Moitinho-Silva *et al.* (2014a) with
136 minor modifications. Briefly, PCRs were performed in triplicate using multiplex identifier
137 adaptor-ligated primers 533f (5'-GTGCCAGCAGCYGCGGTMA-3') and 907r (5'-
138 CCGTCAATTMMYTTGAGTTT-3'), (Simister *et al.*, 2012). The sequence region
139 between the 533f and 907r primers includes the hypervariable regions V4 and V5. PCR

140 amplification was performed using the Phusion Hot Start II High-Fidelity DNA

141 Polymerase (Thermo Scientific, Germany) with buffer CG. The PCR protocol was as

142 follows: initial denaturation at 98 °C for 3 min, followed by 30 cycles of denaturation (98

143 °C for 10 s), annealing (65 °C for 30 s), and extension (72 °C for 18 s). An additional

144 elongation step (72 °C for 5 min) was performed at the end of the protocol. Amplified

145 DNA was gel-purified, pooled at equimolar ratios and sequenced on a Roche 454 GS

146 FLX Titanium platform at the KAUST Genomics Core Lab. Raw pyrosequencing reads

147 were deposited under the NCBI SRA accession numbers SRP043983.

148

149 *Processing of 454 sequences and taxonomic assignment*

150 The 16S rRNA amplicon sequences were processed with the QIIME pipeline v 1.4

151 (Caporaso *et al.*, 2010) according to Moitinho-Silva *et al.* (2014a). Briefly, denoised,

152 quality-curated sequences from Mediterranean samples were pooled with publically

153 available DNA sequences from Red Sea seawater, *X. testudinaria* and *S. carteri*

154 (SRP017932, Moitinho-Silva *et al.* (2014a)). An additional, newly sequenced amplicon

155 dataset of *S. carteri* DNA replicate 2 (ASD2, collected in Moitinho-Silva *et al.* (2014a)

156 was included in the present study. Sequences with > 97% sequence similarity were

157 clustered into operational taxonomic units (OTUs), taxonomically classified and filtered,

158 excluding singletons, chimeras, and eukaryotic sequences.

159

160 *Ecological indices of amplicon data*

161 Ecological indices and estimators were calculated based on datasets rarefied to

162 minimum number of sequences recovered among all datasets. Richness was estimated

163 by chao1 and ACE (Chao, 1984, Chao & Lee, 1992). Diversity was assessed using
164 Simpson and Shannon indices (Shannon, 1948, Simpson, 1949). Ecological indices
165 were calculated using vegan package v 2.0-4 (Oksanen *et al.*, 2012) in R. UniFrac
166 analysis (Lozupone & Knight, 2005) was performed for amplicon sequences using the
167 QIIME pipeline (Caporaso *et al.*, 2010) and visualized as PCoA plot. Jackknife support
168 was obtained based on 100 repetitions at 10,000 sequences per sample.
169
170 *GeoChip 4 loading and data processing*
171 For each sample, 800 ng of DNA was used directly for labeling followed by a 16.5 hrs
172 hybridization step. Fluorescence-labeling, array hybridization and scanning were
173 conducted as described (Lu *et al.*, 2012). Data normalization and preliminary analysis
174 (including removal of poor-quality spots, normalization of spot signal intensity of each
175 spot by mean, removal of spots with low signal intensities based on the signal-to-noise
176 ratio (SNR) as well as removal of outliers) was conducted as described before (He *et*
177 *al.*, 2007, Wu *et al.*, 2006) and as provided by Glomics Inc. as customer service. Probes
178 were assigned as positive when at least three of five (or three of four for Mediterranean
179 seawater) biological replicates showed a positive signal. The normalized signal
180 intensities were calculated as the sum of all probes per gene, divided by the signal
181 intensity of all probes per category, and then averaged across all replicates per sample.
182 Statistical analysis was performed with the non-parametric Mann-Whitney U-test using
183 GraphPad Prism version 6.01 for Windows (GraphPad Software, USA). GeoChip data
184 were analyzed in the R environment v 2.15.3 (The R Core Team, 2012).
185

186 *Minimum curvilinear embedding (MCE)*

187 The Minimum Curvilinear embedding (MCE) (Cannistraci *et al.*, 2010) algorithm was
188 used for unsupervised exploration and discrimination of samples and gene patterns. The
189 same algorithm was already successfully adopted in environmental microbiology
190 essentially for a similar task (Moitinho-Silva *et al.*, 2014a). To explore the relation
191 between samples, MCE was computed starting from a table containing gene signal
192 intensities in each sample, which were calculated as the average of normalized probe
193 signal intensities. Normalized probe signal intensities were calculated by dividing the
194 signal intensity of each probe by the total signal intensity of the sample. To explore the
195 relation between genes, the table was transposed and used as input. Pearson
196 correlation-based distances between samples or genes were obtained using the
197 following expression (Cannistraci *et al.*, 2010):

198

$$correlation_based_distance(x, y) = 1 - Pearson_correlation(x, y)$$

199

200 These distances were used to construct the minimum spanning tree for each dataset.
201 The nonlinear distances stored in the Minimum Curvilinear matrices were finally
202 calculated as the traversal distances over the minimum spanning tree between the data
203 points in the multidimensional space. The calculated Minimum Curvilinear distance
204 matrices were adopted for MCE dimension reduction. MCE analysis was performed
205 using the Singular-Value-Decomposition-based algorithm published in Cannistraci *et al.*
206 (2013). MCE analysis was performed both in MATLAB and R, offering the same results.
207 Principal component analysis (PCA) was performed in MATLAB.

208 **Results and Discussion**

209

210 *Microbial diversity*

211 A total of 234,810 denoised, quality-controlled 16S rRNA gene sequences were
212 generated from three biological replicates each of Mediterranean seawater*, D. avara*, *A.*
213 *aerophoba*, and one biological replicate of *S. carteri* that had originally been collected
214 by Moitinho-Silva *et al*. (2014a). These sequences were combined with already
215 published 16S rRNA gene sequences from the Red Sea (seawater (n=3), *S. carteri*
216 (n=2), *X. testudinaria* (n=3), (Moitinho-Silva *et al.*, 2014a)).This effort resulted in 1,726
217 OTUs representing a total of 576,444 sequences of which 209,012 sequences were
218 produced in this study.

219

220 Rarefaction curves revealed that samples were sequenced with different sampling depth
221 (Suppl. Fig. 1A). Therefore, the datasets were rarefied to the minimum sampling effort
222 (8,925 sequences per sample). Seawater displayed the highest richness at both
223 locations. The term “richness” is based on the number of species in a given community
224 and is judged from the estimators ACE and chao1 (Suppl. Fig. 1B). However, the HMA
225 sponges showed the highest diversity. The term “diversity” takes the relative species
226 abundances into account and is judged from Shannon and Simpson indices. The LMA
227 sponges showed the lowest diversity at each location (Suppl. Fig. 1C).

228

229 The distance between the microbiomes under investigation were analyzed by UniFrac
230 and ordered by Principle coordinates analysis (PCoA) plot (Fig. 1A). PCoA of amplicon

231 sequences revealed a clear separation of the three sample types (Fig. 1A). According to
232 PC1 (which explained 43.65 % of the variation), samples were ordered according to
233 source (HMA, LMA, seawater) rather than to geographic location (Mediterranean, Red
234 Sea). The LMA sponge samples (*D. avara* and *S. carteri*) presented a larger variation
235 between sources and were placed between seawater and the HMA samples. PC2
236 (which explained 36.74 % of the variation) separated *D. avara* from the other sources.
237
238 A total of 28 bacterial, archaeal, and candidate phyla were detected in the present study
239 (Fig. 1B, Suppl. Table 1). The microbial diversity of the Red Sea sponges was published
240 in Moitinho-Silva *et al*. (2014a), but is included here for comparative purposes. The
241 phylum level composition of the HMA sponge species *A. aerophoba* and *X. testudinaria*
242 was complex and was composed of *Proteobacteria* (26.6 and 28.0%), *Chloroflexi* (22.9
243 and 34.5%), *Acidobacteria* (20.8 and 6.8%), *Deferribacteres* (8.7 and 11.7%),
244 *Nitrospirae* (5.1 and 2.9%)*, Gemmatimonadetes* (4.2 and 3.8%), *Actinobacteria* (2.9 and
245 4.6%), *Cyanobacteria* (2.1 and 0.27%), *Spirochaetes* (2.2 and 1.6%), the candidate
246 phylum *Poribacteria* (2.7 and 2.2%), *Bacteroidetes* (1.0 and 1.7%), and *Planctomycetes*
247 (0.02 and 1.4%). Each of the remaining 16 phyla contributed < 1% to the HMA sponge
248 dataset. The LMA sponge species *D. avara* and *S. carteri* were dominated by
249 *Proteobacteria* (91.9 and 74.1%), *Cyanobacteria* (3.8 and 12.8%), and *Bacteriodetes*
250 (2.9 and 6.6%) (Fig. 1B, Suppl. Table 1). It is noteworthy, that the proteobacterial
251 community of *D. avara* was dominated by *Beta*- and *Alphaproteobacteria*, while
252 *Gamma*- and *Alphaproteobacteria* were dominant in *S. carteri* (Suppl. Table 1). Each of
253 the remaining 25 phyla contributed < 1% to the LMA sponge dataset. The seawater

254 samples were dominated by *Proteobacteria* (64.3 in Mediterranean and 41.3% in Red
255 Sea seawater), followed by *Cyanobacteria* (15.3 and 30.4%), *Bacteriodetes* (15.5 and
256 12.7%), *Actinobacteria* (1.3 and 9.8%) and *Verrucomicrobia* (2.3 and 0.08%). Each of
257 the remaining 23 phyla contributed < 1% to the seawater dataset.

258

259 The microbial diversity of the investigated sponge species and seawater is fully
260 consistent with previous reports (Moitinho-Silva *et al.*, 2014a, Schmitt *et al.*, 2012a,
261 Schmitt *et al.*, 2012b). The HMA sponges show a more complex phylum-level
262 composition with *Proteobacteria* and *Chloroflexi* as most abundant phyla, followed by
263 *Acidobacteria*, *Deferribacteres*, *Nitrospirae*, *Gemmatimonadetes*, *Actinobacteria*, and
264 candidate phylum Poribacteria. The phylum level composition of the LMA sponge
265 species, *S. carteri*, was dominated by *Proteobacteria*, *Cyanobacteria*, and *Archaea*, as
266 has been reported previously (Giles *et al.*, 2013, Moitinho-Silva *et al.*, 2014a). Several
267 additional studies revealed a distinct and different microbial composition in HMA
268 sponges versus LMA sponges (Erwin *et al.*, 2011, Giles *et al.*, 2013, Gloeckner *et al.*,
269 2013, Kamke *et al.*, 2010, Moitinho-Silva *et al.*, 2014a, Poppell *et al.*, 2013, Schmitt *et*
270 *al.*, 2012b, Weisz *et al.*, 2007). However, even though the seawater samples resemble
271 the composition of LMA sponges on the phylum level, the LMA sponges still contain
272 their own characteristic symbiont guilds (Moitinho-Silva *et al.*, 2014a), as evidenced here
273 by a dominance of *Betaproteobacteria* in *D. avara* and *Gammaproteobacteria* in *S.*
274 *carteri*.

275

276 *GeoChip analysis of microbial functional genes*

277 Using the functional GeoChip array 4, a total of 20,273 probes were identified that
278 represented 627 functional genes, grouped altogether in 16 functional gene categories
279 in sponge and seawater samples (Suppl. Figure 1). The majority of positive gene probes
280 belonged to the categories Organic remediation (4577 probes), Stress (4178 probes),
281 Carbon cycling (2327 probes), Metal resistance (2297 probes), Nitrogen (1580 probes),
282 as well as in lesser amounts (< 850 probes) to the categories Antibiotic resistance,
283 Bacteria phage, Bioleaching, Energy Process, Fungi function, Others, Phosphorus, Soil
284 Benefit, Soil borne pathogen, Sulfur, and Virulence (Suppl. Table 2).
285
286 The exploration of the samples by MCE revealed a clear separation (Fig. 2A). The first
287 dimension (horizontal axis) was discriminative between the three sources (HMA, LMA,
288 seawater) and within that, between the locations (Red Sea, Mediterranean) for the
289 sponge samples. The second dimension separated the LMA sponges from HMA
290 sponges and seawater samples. The fact that the HMA LMA pattern was stronger than
291 any geographic bias is remarkable considering that the Mediterranean and Red Sea are
292 distinct locations. To our knowledge, this is the first time that the HMA LMA pattern was
293 identified on the level of microbial functional genes.
294
295 On the other hand, the exploration of the genes by MCE (Fig. 2B) unveiled very few
296 discernible patterns. Noticeably, the LMA sponges, in particular *D. avara*, lacked genes
297 as indicated by dark blue boxes. The absence of genes in *D. avara* resulted in higher
298 than average signal intensities after normalization, as indicated by the yellow color. In

conclusion, few specific gene patterns rather than a general overarching trend were revealed by MCE.

When the GeoChip data were analysed for the presence/absence of genes, the vast majority of genes (> 88% for the Mediterreanean site, > 91% for the Red Sea site) were shared between all three sources (Fig. 3). The sets of genes unique to a given source did not reveal any specific genes or pathways, thus the relevance of the unique genes remains unknown (data not shown). It is important to consider, that equal DNA concentrations for each sample were applied to the GeoChip. For example, a much larger volume of seawater was necessary to extract the same amount of DNA. Similarly, much more LMA sponge biomass was necessary to yield the same amount of DNA as from HMA sponges. The results show that the microbial functional gene repertoires of HMA and LMA sponges, as well as of seawater, are largely similar. Any differences appear to be due to microbial and/or gene abundances rather than to true differences in the functional gene repertoire. These findings are consistent with previous publications on the microbial communities of tropical corals and also of uranium-contaminated aquifers which also reported on a consistent functional gene repertoire in spite of a variable microbial diversity (Kimes *et al.*, 2010, Van Nostrand *et al.*, 2009) .

Functional convergence of microbial sponge symbionts has previously been reported based on comprehensive metagenomic analyses (Fan *et al.*, 2012) and based on nutrient fluxes in combination with phylogenetic analyses of selected sponge symbionts (Ribes *et al.*, 2012). Here we extend previous findings in that functional convergence is

322 also reported between sponges (HMA and LMA) and seawater. Several explanations
323 are possible for this observation. For one, the GeoChip may have only limited
324 applicability to sponge microbiomes in the sense that the truly unique microbial genes
325 may not have been covered. Secondly, owing to the presence of seawater bacteria in
326 sponge microbiomes, which has been shown to amount to 5-24% depending on the
327 HMA or LMA status (Moitinho-Silva *et al.*, 2014a), the extracted DNA may have
328 contained a sufficient amount of seawater bacterial DNA to mask true differences
329 between the sponge microbiomes. As a third possible explanation, it may be considered
330 that if seawater was the adaptive driving force, then the functional gene repertoire is
331 going to be the same whether the microorganisms reside inside or outside of a sponge.
332 In the latter hypothesis, it would thus not be surprising to observe functional gene
333 convergence in different sponge or seawater microbiomes.
334
335 The few differences that were found and that were of relevance to sponge symbioses
336 are discussed below. Only genes that resulted in statistically significantly different signal
337 intensities in either HMA or HMA sponges in comparison to seawater and where the
338 difference was present at both geographic locations are shown (Figs. 4-6). The gene
339 categories that were well covered by gene probes (resulting in normalized signal
340 intensity > 5%) are presented below. The unit signal intensity is taken hereafter as a
341 proxy for gene abundance.
342
343 *Nitrogen metabolism*

344 Microbial nitrogen metabolism is a major theme to have emerged out of the past decade
345 of sponge microbiology (Han *et al.*, 2013, Hentschel *et al.*, 2012, Taylor *et al.*, 2007,
346 Webster & Taylor, 2012, Zhang *et al.*, 2014). Sponges excrete ammonia as a metabolic
347 waste product, which makes them a particularly attractive niche for microorganisms in
348 an otherwise nitrogen-poor marine environment. Accordingly, genes corresponding to
349 nearly all major pathways (nitrification, denitrification, ammonification, assimilatory and
350 dissimilatory nitrogen reduction, nitrogen fixation) were identified by GeoChip analyses.
351 Genes involved in anaerobic ammonia oxidation (anammox) were noticeably absent.
352
353 With respect to nitrification, the abundance of the *amoA* gene, as an indicator for
354 bacterial ammonia oxidation was higher in both HMA and LMA sponges than in
355 seawater (Fig. 4). The archaeal *amoA* was however slightly reduced in HMA sponges
356 and so was the *hao* gene encoding for hydroxylamine oxidase. In spite of the higher
357 presence of bacterial *amoA* over archaeal *amoA* gene, metatranscriptome analyses
358 revealed archaeal ammonia monooxygenase to be among the most highly expressed
359 microbial features in *Geodia barretti* and *S. carteri* (Moitinho-Silva *et al.*, 2014b, Radax
360 *et al.*, 2012). Archaea appear to be the main drivers of ammonia oxidation in sponges
361 (Zhang *et al.*, 2014), but also in many other ecosystems such as seawater (Shi *et al.*,
362 2011) and soil (Leininger *et al.*, 2006).
363
364 Several genes encoding for denitrification (*narG*, *nirS*, *nosZ)* were reduced in sponges
365 over seawater (Fig. 4). This observation is consistent with the current perception that
366 sponges have generally aerobic metabolism and that the mesohyl matrix is well

367 oxygenated from the intense pumping activities of the animal. Somewhat contradictory,
368 the *nrfA* gene abundance, encoding for the dissimilatory reduction of nitrate to ammonia,
369 was slightly increased in LMA sponges over seawater. The sponge tissues were shown
370 to turn anaerobic during periods of non-pumping, so denitrification may still be a likely, if
371 only temporary, scenario (Schlappy *et al.*, 2010).
372
373 The symbionts may further cover their nitrogen needs from urea hydrolysation. Urea can
374 originate from a variety of sources (Crandall & Teece, 2012), such as by bacterial
375 degradation of nucleic and amino acids and is therefore a likely product to be
376 encountered in the sponge mesohyl. In fact, the phylogenetic diversity of the *ureC* gene
377 as well as its transcriptional activity has been demonstrated in the microbial symbionts
378 *Xestospongia testudinaria* (Su *et al.*, 2013).GeoChip analyses revealed higher *ureC*
379 gene abundance in sponges than in seawater (Fig. 4). This gene encoding for a urease
380 subunit and endows the bacteria with the capacity to hydrolyze urea. Urease-encoding
381 gene clusters, urea transporters, and accessory genes were previously identified in
382 sponge symbiont genomes (Hallam *et al.*, 2006, Siegl *et al.*, 2011). Furthermore,
383 phylogenetic diversity and transcriptional activities of the *ureC* gene was recently
384 documented for the microbial symbionts of *X. testudinaria*, providing strong support for
385 an *in vivo* relevance of microbial urea utilization in sponge symbioses (Su *et al.*, 2013).
386
387 *Carbon metabolism*
388 Autotrophic carbon fixation is a hallmark of cyanobacteria, including cyanobacterial
389 symbionts of sponges (Taylor *et al.*, 2007). Gene copy numbers for RuBisCO, the key

390 enzyme of cyanobacterial carbon fixation, were however lower in the sponges than in
391 seawater (Fig. 5). On the other hand, the key gene, *pcc*, encoding for propionyl-CoA-
392 carboxylase, and representing the recently discovered autotrophic carbon fixation
393 pathway in archaea (Berg *et al.*, 2007) was higher in HMA sponges than in seawater. It
394 appears likely that the wide-spread sponge symbiont *Crenarchaeum symbiosum* is
395 responsible for autotrophic carbon fixation via the 3-hydroxypropionate/4-
396 hydroxybutyrate cycle in sponges (Hallam *et al.*, 2006).
397
398 With respect to carbon degradation, only two genes were identified as being statistically
399 different between sponges and seawater (Fig. 5). Higher exochitinase gene abundance
400 underlines the potential of sponge symbionts for chitin degradation. In this context, a
401 glycoside hydrolase (GH74) with potential for chitin deacetylation was recently identified
402 on poribacterial genomes by single cell genomics (Kamke *et al.*, 2013). Alpha-chitin was
403 previously identified as a structural component of the sponge skeleton and may thus
404 likely be encountered in the sponge mesohyl (Ehrlich *et al.*, 2007). The lower abundance
405 the *xylA* gene indicates that hemicellulose degradation is probably not relevant in
406 sponge microbiomes. Similarly, the lower gene abundance of *mcrA*, an indicator gene
407 for anaerobic, methanogenic archaea, was lower in sponges than in seawater, indicating
408 that the potential for methanogenesis in sponges is probably limited.
409
410 *Stress-related genes*
411 A consistently lower abundance of genes involved in stress, particularly with respect to
412 oxygen (*ahpC*, *katE*) and radiation stress (*obgE*) was observed in sponges over

413 seawater (Fig. 6). The finding of reduced *obgE* gene levels was observed for HMA and
414 LMA sponges at both locations. A microbial existence within animals is frequently
415 correlated with an increased repertoire of stress protection proteins such as chaperonins
416 (Fan *et al.*, 2013, Liu *et al.*, 2012). Whether the symbionts of sponges are more or less
417 stressed than their planktonic counterparts continues to be an interesting topic for future
418 investigations.
419
420 The *pstS* gene, which is involved in phosphate transport, was slightly increased in LMA
421 sponges over seawater, possibly indicating the sponge symbionts are subject to
422 phosphate limitation. In this context it is interesting to note that the proteobacterial
423 glycerol-3-phosphate ABC transporter was among the highly expressed genes in the *S.*
424 *carteri* transcriptome (Moitinho-Silva *et al.*, 2014b). Furthermore, a number of glycerol-3-
425 P ABC-type transporter proteins, termed UgpB, were identified in the
426 metaproteogenome of *C. concentrica* (Liu *et al.*, 2012). The Ugp system is thought to be
427 involved in scavenging phosphate-containing compounds (Boos, 1998). Furthermore,
428 single cell genomic analyses revealed an abundance of *phyH* genes in poribacterial
429 symbionts of sponges that should endow the bacteria with the ability to utilize 2-
430 aminoethylphosphonate (2-AEPn) as dissolved organic posphorus source (Kamke *et al.*,
431 2013). However, in spite of several predictions resulting from omics approaches,
432 functional experimental evidence for phosphorous metabolism is still lacking for sponge
433 symbionts.
434
435 *Conclusions*

436 In conclusion, the following specific insights were obtained: (i) GeoChip analyses by
437 MCE showed, for the first time, that the HMA/LMA dichotomy exists also on the
438 functional gene level. MCE revealed further nonlinear relations between the samples
439 that are only poorly discriminated by conventional methods like principal component
440 analysis (PCA; see Suppl. Fig. 2). Future efforts will be directed at developing
441 algorithms that identify which genes in particular account for the sample separation. (ii)
442 Microbial nitrification and ammonification genes were increased in sponges, while
443 denitrification genes were reduced. A higher abundance of archaeal autotrophic carbon
444 fixation genes was noted in sponges than in seawater. Stress genes were found at lower
445 abundances in sponge microbiomes than in seawater. (iii) While methodological
446 limitations, such as the applicability of the GeoChip outside of its original "soil" context,
447 cannot be ruled out, it appears nonetheless conceivable, that sponge-associated and
448 seawater microorganisms have most of their functional gene repertoire in common.
449
450
451 **Acknowledgements**
452
453 We thank Martin Pfannkuchen at the Institute Ruder Boskovic (Rovinj, Croatia) as well
454 as the Coastal and Marine Resources Core Lab and the Biosciences Core Laboratory
455 (KAUST, Saudi Arabia) for expert help in sample collection. We thank Zhili He
456 (University of Oklahoma, USA) and Joy D. van Noystrand at Glomics, Inc. for advice.
457 LM-S was supported by a grant of the German *Excellence Initiative* to the Graduate

School of Life Sciences, University of Wuerzburg. Additional financial support was generously provided by KAUST baseline funds to TR.

Figure legends

Figure 1: Microbial composition of HMA and LMA sponges as well as of seawater from the Mediterranean and the Red Sea collection site as obtained by 454 pyrosequencing. Relation among samples as performed by principle coordinates analysis (PCoA) based on UniFrac (A), taxon composition of samples (B). Phyla and major classes of Proteobacteria are shown for each source. Taxa which are present at > 1% abundance are shown. See Suppl. Table 1 for details of taxon composition.

Figure 2: Exploration and visualization of the GeoChip 4 by Minimum Curvilinear Embedding (MCE). The reciprocal relations between the samples are mapped in a two-dimensional space by nonlinear dimension reduction performed by the MCE algorithm. Each spot represents a sample (A). The label “Red” indicates samples collected in the Red Sea; the label “Med” indicates samples collected in the Mediterranean Sea. Heatmap of functional genes' normalized signal intensities. The x-axis plots the index of genes which are ordered according to first dimension, D1, of MCE analysis (B).

Figure 3: Presence/absence analysis of functional genes identified by the GeoChip. A total of 524 genes (88%) were shared between sources for the Mediterranean site and 557 genes (91%) were shared between sources of the Red Sea site.

481

482 Figure 4: Normalized average signal intensities of genes involved in nitrogen cycling.
483 The microbial processes and corresponding genes are as follows: nitrification (archaeal
484 and bacterial *amoA* encoding ammonia monooxygenase, *hao* for hydroxylamine
485 oxidoreductase); denitrification (*narG* for nitrate reductase, *nirS* for nitrite reductase,
486 *nosZ* for nitrate reductase); dissimilatory N reduction to ammonium (*nrfA* for c-type
487 cytochrome nitrite reductase); ammonification (*ureC* for urease). Data are presented as
488 the mean ± SE. **: $P<0.01$, *: $P<0.05$. “Med Sw” and “RS Sw” stand for Mediterranean
489 and Red Sea seawater.

490

491 Figure 5: Normalized average signal intensities of gene categories involved in carbon
492 cycling. The microbial processes and corresponding genes are as follows: methane
493 production (*mcrA*, encoding for methyl coenzyme M reductase A); carbon fixation in the
494 citrate cycle (*pcc* for propionyl-CoA carboxylase) and in the Calvin cycle (*RuBisCO* for
495 Ribulose-1, 5-bisphosphate carboxylase/oxygenase); degradation of chitin (with genes
496 encoding for exochitinase) and of hemicellulose (*xylA* for xylanase). The standard
497 nomenclature of the GeoChip 4 was maintained (ref). Data are presented as the mean ±
498 SE. **: $P<0.01$, *: $P<0.05$. “Med Sw” and “RS Sw” stand for Mediterranean and Red Sea
499 seawater.

500

501 Figure 6: Normalized average signal intensities of gene categories involved in stress.
502 The microbial genes involved in stress response are as follows: *ahpC* encoding for alkyl
503 hydroperoxide reductase and katE involved in oxygen stress; *pstS* encoding for a

phosphate binding protein involved in phosphate limitation; *obgE* encoding for a GTPase involved in radiation stress. Data are presented as the mean ± SE. **: $P<0.01$, *: $P<0.05$. "Med Sw" and "RS Sw" stand for Mediterranean and Red Sea seawater.

Supplementary Figure 1: Rarefaction curves based on a 97% sequence similarity threshold for marine sponge- and seawater-derived 16S rRNA amplicon pyrosequences from the Mediterranean and the Red Sea (A). *Aplysina aerophoba* (AA1-3), *Dysidea avara* (DA1-3), Mediterranean seawater (MWT1-3), *Xestospongia testudinaria* (AXD1-3), *Stylissa carteri* (ASD1-3), and Red Sea seawater (WTD1-3). Richness (B) and diversity (C) statistics.

Supplementary Figure 2: Principal component analysis (PCA) of the GeoChip 4. Relation among samples were explored by PCA. Samples of HMA (green) and LMA (red) sponges as well as seawater (blue) collected in the Mediterranean or Red Sea are shown.

References

Bell JJ (2008) The functional roles of marine sponges. *Estuarine Coastal and Shelf Science* **79**: 341-353.

Berg IA, Kockelkorn D, Buckel W & Fuchs G (2007) A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science* **318**: 1782-1786.

Boos W (1998) Binding protein-dependent ABC transport system for glycerol 3-phosphate of *Escherichia coli*. *Methods in Enzymology* **292**: 40-51.

529 Cannistraci CV, Alanis-Lobato G & Ravasi T (2013) Minimum curvilinearity to enhance
530 topological prediction of protein interactions by network embedding. *Bioinformatics* **29**: i199-209.
531 Cannistraci CV, Ravasi T, Montevecchi FM, Ideker T & Alessio M (2010) Nonlinear dimension
532 reduction and clustering by Minimum Curvilinearity unfold neuropathic pain and tissue
533 embryological classes. *Bioinformatics* **26**: i531-539.
534 Caporaso JG, Kuczynski J, Stombaugh J*, et al.* (2010) QIIME allows analysis of high-throughput
535 community sequencing data. *Nat Methods* **7**: 335-336.
536 Chan Y, Van Nostrand JD, Zhou J, Pointing SB & Farrell RL (2013) Functional ecology of an
537 Antarctic Dry Valley. *Proceedings of the National Academy of Sciences of the United States of*
538 *America* **110**: 8990-8995.
539 Chao A (1984) Nonparametric estimation of the number of classes in a population *Scandinavian*
540 *Journal of Statistics* **11**: 265-270.
541 Chao A & Lee SM (1992) Estimating the number of classes via sample coverage. *Journal of the*
542 *American Statistical Association* **87**: 210-217.
543 Crandall JB & Teece MA (2012) Urea is a dynamic pool of bioavailable nitrogen in coral reefs.
544 *Coral Reefs* **31**: 207-214.
545 de Goeij JM, van Oevelen D, Vermeij MJ, Osinga R, Middelburg JJ, de Goeij AF & Admiraal W
546 (2013) Surviving in a marine desert: the sponge loop retains resources within coral reefs.
547 *Science* **342**: 108-110.
548 Ehrlich H, Maldonado M, Spindler KD, Eckert C, Hanke T, Born R, Goebel C, Simon P,
549 Heinemann S & Worch H (2007) First evidence of chitin as a component of the skeletal fibers of
550 marine sponges. Part I. Verongidae (demospongia: Porifera). *Journal of Experimental Zoology*
551 *Part B, Molecular and Developmental Evolution* **308**: 347-356.
552 Erwin PM, Olson JB & Thacker RW (2011) Phylogenetic diversity, host-specificity and
553 community profiling of sponge-associated bacteria in the northern Gulf of Mexico. *PloS One* **6**:
554 e26806.
555 Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS & Thomas T (2012) Functional
556 equivalence and evolutionary convergence in complex communities of microbial sponge
557 symbionts. *Proceedings of the National Academy of Sciences of the United States of America*
558 **109**: E1878-1887.
559 Fan Y, Thompson JW, Dubois LG, Moseley MA & Wernegreen JJ (2013) Proteomic analysis of
560 an unculturable bacterial endosymbiont (Blochmannia) reveals high abundance of chaperonins
561 and biosynthetic enzymes. *Journal of Proteome Research* **12**: 704-718.
562 Giles EC, Kamke J, Moitinho-Silva L, Taylor MW, Hentschel U, Ravasi T & Schmitt S (2013)
563 Bacterial community profiles in low microbial abundance sponges. *FEMS Microbiology Ecology*
564 **83**: 232-241.
565 Gloeckner V, Hentschel U, Ereskovsky AV & Schmitt S (2013) Unique and species-specific
566 microbial communities in *Oscarella lobularis* and other Mediterranean *Oscarella* species
567 (Porifera: Homoscleromorpha). *Marine Biology* **160**: 781-791.
568 Hallam SJ, Mincer TJ, Schleper C, Preston CM, Roberts K, Richardson PM & DeLong EF (2006)
569 Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic
570 analyses of marine Crenarchaeota. *PLoS Biology* **4**: e95.
571 Han M, Li Z & Zhang F (2013) The ammonia oxidizing and denitrifying prokaryotes associated
572 with sponges from different sea areas. *Microbial Ecology* **66**: 427-436.
573 He Z, Gentry TJ, Schadt CW*, et al.* (2007) GeoChip: a comprehensive microarray for
574 investigating biogeochemical, ecological and environmental processes. *The ISME Journal* **1**: 67-
575 77.
576 Hentschel U, Fieseler L, Wehrl M, Gernert C, Steinert M, Hacker J & Horn M (2003) Microbial
577 diversity of marine sponges. *Prog Mol Subcell Biol* **37**: 59-88.

578 Hentschel U, Piel J, Degnan SM & Taylor MW (2012) Genomic insights into the marine sponge
579 microbiome. *Nature Reviews Microbiology* **10**: 641-654.
580 Kamke J, Sczyrba A, Ivanova N, Schwientek P, Rinke C, Mavromatis K, Woyke T & Hentschel U
581 (2013) Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial
582 symbionts of marine sponges. *The ISME Journal* **7**: 1-14.
583 Kamke J, Taylor MW & Schmitt S (2010) Activity profiles for marine sponge-associated bacteria
584 obtained by 16S rRNA vs 16S rRNA gene comparisons. *The ISME Journal* **4**: 498-508.
585 Kang S, Van Nostrand JD, Gough HL, He Z, Hazen TC, Stahl DA & Zhou J (2013) Functional
586 gene array-based analysis of microbial communities in heavy metals-contaminated lake
587 sediments. *FEMS Microbiology Ecology* **86**: 200-214.
588 Kimes NE, Van Nostrand JD, Weil E, Zhou J & Morris PJ (2010) Microbial functional structure of
589 *Montastraea faveolata*, an important Caribbean reef-building coral, differs between healthy and
590 yellow-band diseased colonies. *Environmental Microbiology* **12**: 541-556.
591 Lee OO, Wang Y, Yang J, Lafi FF, Al-Suwailem A & Qian PY (2011) Pyrosequencing reveals
592 highly diverse and species-specific microbial communities in sponges from the Red Sea. *The*
593 *ISME Journal* **5**: 650-664.
594 Leininger S, Urich T, Schloter M, Schwark L, Qi J, Nicol GW, Prosser JI, Schuster SC &
595 Schleper C (2006) Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature*
596 **442**: 806-809.
597 Liu M, Fan L, Zhong L, Kjelleberg S & Thomas T (2012) Metaproteogenomic analysis of a
598 community of sponge symbionts. *The ISME Journal* **6**: 1515-1525.
599 Lozupone C & Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial
600 communities. *Appl Environ Microbiol* **71**: 8228-8235.
601 Lu Z, Deng Y, Van Nostrand JD*, et al.* (2012) Microbial gene functions enriched in the
602 Deepwater Horizon deep-sea oil plume. *The ISME Journal* **6**: 451-460.
603 Mason OU, Nakagawa T, Rosner M, Van Nostrand JD, Zhou J, Maruyama A, Fisk MR &
604 Giovannoni SJ (2010) First investigation of the microbiology of the deepest layer of ocean crust.
605 *PloS One* **5**: e15399.
606 Moitinho-Silva L, Bayer K, Cannistraci CV, Giles EC, Ryu T, Seridi L, Ravasi T & Hentschel U
607 (2014a) Specificity and transcriptional activity of microbiota associated with low and high
608 microbial abundance sponges from the Red Sea. *Molecular Ecology* **23**: 1348-1363.
609 Moitinho-Silva L, Seridi L, Ryu T, Voolstra CR, Ravasi T & Hentschel U (2014b) Revealing
610 microbial functional activities in the Red Sea sponge *Stylissa carteri* by metatranscriptomics.
611 *Environmental Microbiology* DOI: 10.1111/1462-2920.12533.
612 Nostrand JD, He Z & Zhou J (2012) Use of functional gene arrays for elucidating in situ
613 biodegradation. *Frontiers in Microbiology* **3**: 339.
614 Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos
615 P, Stevens MHH & Wagner H (2012) vegan: Community Ecology Package.
616 Poppell E, Weisz JB, Spicer L, Massaro A, Hill A & Hill M (2013) Sponge heterotrophic capacity
617 and bacterial community structure in high- and low-microbial abundance sponges. *Marine*
618 *Ecology* doi: 10.1111/maec.12098.
619 Radax R, Rattei T, Lanzen A, Bayer C, Rapp HT, Urich T & Schleper C (2012)
620 Metatranscriptomics of the marine sponge *Geodia barretti*: tackling phylogeny and function of its
621 microbial community. *Environmental Microbiology* **14**: 1308-1324.
622 Reiswig HM (1981) Partial carbon and energy budgets of the bacteriosponge *Verongia fistularis*
623 (Porifera: Demospongiae) in Barbados West-Indies *Marine Biology* **2**: 273-294.
624 Ribes M, Jimenez E, Yahel G, Lopez-Sendino P, Diez B, Massana R, Sharp JH & Coma R
625 (2012) Functional convergence of microbes associated with temperate marine sponges.
626 *Environmental Microbiology* **14**: 1224-1239.

627 Schlappy ML, Schottner SI, Lavik G, Kuypers MM, de Beer D & Hoffmann F (2010) Evidence of
628 nitrification and denitrification in high and low microbial abundance sponges. *Marine Biology*
629 **157**: 593-602.
630 Schmitt S, Hentschel U & Taylor M (2012a) Deep sequencing reveals diversity and community
631 structure of complex microbiota in five Mediterranean sponges. *Hydrobiologia* **687**: 341-351.
632 Schmitt S, Tsai P, Bell J*, et al.* (2012b) Assessing the complex sponge microbiota: core, variable
633 and species-specific bacterial communities in marine sponges. *The ISME Journal* **6**: 564-576.
634 Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* **27**: 379–423.
635 Shi Y, Tyson GW, Eppley JM & DeLong EF (2011) Integrated metatranscriptomic and
636 metagenomic analyses of stratified microbial assemblages in the open ocean. *The ISME Journal*
637 **5**: 999-1013.
638 Siegl A, Kamke J, Hochmuth T, Piel J, Richter M, Liang C, Dandekar T & Hentschel U (2011)
639 Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically
640 associated with marine sponges. *The ISME Journal* **5**: 61-70.
641 Simister RL, Deines P, Botte ES, Webster NS & Taylor MW (2012) Sponge-specific clusters
642 revisited: a comprehensive phylogeny of sponge-associated microorganisms. *Environmental*
643 *Microbiology* **14**: 517-524.
644 Simpson EH (1949) Measurement of Diversity. *Nature* **163**: 688-688.
645 Su J, Jin L, Jiang Q, Sun W, Zhang F & Li Z (2013) Phylogenetically diverse ureC genes and
646 their expression suggest the urea utilization by bacterial symbionts in marine sponge
647 *Xestospongia testudinaria. PloS One* **8**: e64848.
648 Taylor MW, Radax R, Steger D & Wagner M (2007) Sponge-associated microorganisms:
649 evolution, ecology, and biotechnological potential. *Microbiology and Molecular Biology Reviews*
650 **71**: 295-347.
651 Tu Q, Yu H, He Z*, et al.* (2014) GeoChip 4: a functional gene-array-based high-throughput
652 environmental technology for microbial community analysis. *Molecular Ecology Resources.*
653 Vacelet J & Donadey C (1977) Electron microscope study of the association between some
654 sponges and bacteria. *Journal of Experimental Marine Biology and Ecology* **30**: 301-314.
655 Van Nostrand JD, Wu WM, Wu L*, et al.* (2009) GeoChip-based analysis of functional microbial
656 communities during the reoxidation of a bioreduced uranium-contaminated aquifer.
657 *Environmental Microbiology* **11**: 2611-2626.
658 Webster NS & Taylor MW (2012) Marine sponges and their microbial symbionts: love and other
659 relationships. *Environmental Microbiology* **14**: 335-346.
660 Webster NS, Taylor MW, Behnam F, Lucker S, Rattei T, Whalan S, Horn M & Wagner M (2010)
661 Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge
662 symbionts. *Environ Microbiol* **12**: 2070-2082.
663 Weisz JB, Hentschel U, Lindquist N & Martens CS (2007) Linking abundance and diversity of
664 sponge-associated microbial communities to metabolic differences in host sponges. *Marine*
665 *Biology* **152**: 475-483.
666 Wilson MC, Mori T, Ruckert C*, et al.* (2014) An environmental bacterial taxon with a large and
667 distinct metabolic repertoire. *Nature* **506**: 58-62.
668 Wu L, Liu X, Schadt CW & Zhou J (2006) Microarray-based analysis of subnanogram quantities
669 of microbial community DNAs by using whole-community genome amplification. *Applied and*
670 *Environmental Microbiology* **72**: 4931-4941.
671 Yergeau E, Bokhorst S, Kang S, Zhou J, Greer CW, Aerts R & Kowalchuk GA (2012) Shifts in
672 soil microorganisms in response to warming are consistent across a range of Antarctic
673 environments. *The ISME Journal* **6**: 692-702.

674 Zhang F, Pita L, Erwin PM, Abaid S, Lopez-Legentil S & Hill RT (2014) Symbiotic archaea in
675 marine sponges show stability and host specificity in community structure and ammonia
676 oxidation functionality. *FEMS Microbiology Ecology*.

677

678

679