

1 **A genomic view of the sponge microbiome**

2 **Running title: A genomic view of the sponge microbiome**

3

4 Robbins, S.J.¹, Song, W.², Engelberts, J.P.¹, Glasl, B.³, Slaby, B.M.⁴, Boyd, J.¹, Marangon,
5 E.^{3,5}, Botté, E.S.¹, Laffy, P.³, Thomas, T.², Webster, N.S.^{1,3*}

6

7 ¹Australian Centre for Ecogenomics, University of Queensland, Brisbane 4072, Australia

8 ²Centre for Marine Science & Innovation, University of New South Wales, Kensington 2052,
9 Australia

10 ³Australian Institute of Marine Science, Townsville 4810, Australia

11 ⁴GEOMAR Helmholtz Centre for Ocean Research Kiel, Düsternbrooker Weg 20, 24105 Kiel,
12 Germany

13 ⁵College of Science and Engineering, James Cook University, Townsville 4810, Australia

14 *Corresponding Author

15

16

17

18

19

20

21

22

23

24

25

26

27 **Abstract**

28 Sponges underpin the productivity of coral reefs, yet few of their microbial symbionts
29 have been functionally characterized. Here we present an analysis of ~1200 metagenome-
30 assembled genomes (MAGs) spanning seven sponge species and 25 microbial phyla.
31 Compared to MAGs derived from reef seawater, sponge-associated MAGs were enriched in
32 glycosyl hydrolases targeting components of sponge tissue, coral mucus, and macroalgae,
33 revealing a critical role for sponge symbionts in cycling reef organic matter. Further,
34 visualisation of the distribution of these genes amongst symbiont taxa uncovered functional
35 guilds for reef organic matter degradation. Genes for the utilisation of sialic acids and
36 glycosaminoglycans present in sponge tissue were found in specific microbial lineages that
37 also encoded genes for attachment to sponge-derived fibronectins and cadherins, suggesting
38 these lineages can utilize specific structural elements of sponge tissue. Further, genes
39 encoding CRISPR and restriction modification systems used in defence against mobile
40 genetic elements were enriched in sponge symbionts, along with eukaryote-like gene motifs
41 thought to be involved in maintaining host association. Finally, we provide evidence that
42 many of these sponge-enriched genes are laterally transferred between microbial taxa,
43 suggesting they confer a selective advantage within the sponge niche and therefore play a
44 critical role in host ecology and evolution.

45

46 **Introduction**

47 Coral reefs are among the most productive ecosystems in the world and are frequently
48 referred to as ‘rainforests of the sea’ due to their immense biodiversity¹. However, despite
49 their exceptionally high primary productivity, nutrient levels in tropical reefs are typically

50 low, necessitating efficient mechanisms for nutrient retention. Marine sponges are therefore
51 an essential component of reef ecosystems because of their role in the “sponge loop,”
52 wherein they filter large volumes of seawater (up to thousands of liters per day²), retaining
53 organic matter and transforming it into biomass that can be consumed by detritivores and
54 recycled back into the reef system³. Reef-dwelling sponges harbor stable and diverse
55 microbial communities that can account for up to 35% of sponge biomass and are
56 hypothesized to carry out functions that support their host’s health and ecology, such as the
57 transformation of carbon (e.g. polysaccharides), nitrogen, and sulfur, as well as providing
58 essential vitamins and amino acids to the host⁴. Yet, despite their importance for host health,
59 few community-level functional investigations have been undertaken to capture the broad
60 range of microbial taxa found in sponges^{5,6}. Instead, gene-centric studies have identified a
61 number of interesting sponge symbiont traits but have been unable to link these features to
62 specific microbial taxa. In addition, nearly all genome-centric characterizations of sponge-
63 associated microbes have been restricted to a few lineages of interest^{7–9} or have focussed on
64 low-abundance microorganisms amenable to cultivation¹⁰, with the majority of lineages
65 remaining undescribed. This skew likely biases our understanding of the roles that each
66 symbiont lineage plays within the microbiome and hinders our ability to identify features that
67 underpin sponge-microbe symbiosis. To address this, we undertook an integrated analysis of
68 1188 metagenome-assembled genomes (MAGs) derived from seven marine sponge species,
69 spanning 25 microbial phyla and the vast majority of microbial taxa commonly found in
70 marine sponges.

71

72 **Materials and methods**

73 *Sample collection and enrichment of bacteria from sponges*

74 Two individuals of *Rhopaloides odorabile* were collected from Esk Island ($18^{\circ} 45.830'S$;
75 $146^{\circ} 31.159'E$) and one from Falcon Island ($18^{\circ} 46.116'S$; $146^{\circ} 32.201'E$) on the 25th of
76 October, 2018. Four individuals of *Coscinoderma matthewsi*, *Carteriospongia foliascens*,
77 *Styliosa flabelliformis*, *Ircinia ramosa*, and *Cliona orientalis* were collected from Davies Reef
78 ($18^{\circ} 49.948'S$; $147^{\circ} 37.995'E$) between the 22nd and the 23rd of December, 2015. All sponges
79 were rinsed in filter-sterilised seawater before being snap-frozen in liquid nitrogen and stored
80 at -80°C. Cell fractionation was performed according to methods described in Botte *et al*¹¹.
81 Briefly, sponges were rinsed twice in calcium-magnesium free seawater before being cut into
82 1 cm³ pieces and homogenized. Due to its bioeroding lifestyle, *C. orientalis* was crushed in
83 liquid nitrogen with a mortar and pestle to separate sponge tissue from the coral skeleton.
84 Sterile collagenase (Sigma Aldrich) was added at a concentration of 0.5g L⁻¹ and samples
85 were shaken at 150rpm for 30min. Samples were filtered through 100µm sterile cell strainers
86 and centrifuged at 100 x g for 1min before recovering the supernatant and centrifuging at 300
87 x g for 15min. This last step was repeated, and the resulting supernatant was filtered
88 sequentially through 8µm and 5µm filters. Microbial pellets were recovered by centrifuging
89 at 8000 x g for 20min, resuspended in Tris-HCl NaCl at pH 8.0 and kept at -20°C. To serve
90 as a point of comparison for statistical enrichment, seawater MAGs were also obtained from
91 the SeaSim at the Australian Institute for Marine Science (AIMS) in October 2016, as well as
92 two previous studies^{12,13}. All seawater samples were filtered onto 0.2um sterivexes before
93 extraction.

94

95 *Metagenomic sequencing*

96 Samples were extracted using the Qiagen MagAttract PowerSoil DNA kit as described by
97 Marotz *et al.*¹⁴. Metagenomic DNA from *R. odorabile* was sequenced at the Ramaciotti
98 Centre for Genomics (University of New South Wales, Sydney, Australia), while other

99 sponge species were sequenced at the University of California San Diego as part of the Earth
100 Microbiome Project (EMP). Metagenomic sequencing was performed over several iterations
101 as part of the EMP using the Kapa HyperPlus and Nextera XT library prep kits, with
102 sequencing performed on Illumina HiSeq 4000 and NovaSeq 600 machines (2 x 151 bp).

103

104 *Metagenome assembly, binning, and taxonomy assignment of bacterial and archaeal MAGs*

105 Reads belonging to the same sample and sequenced across multiple runs were
106 concatenated into a single set of forward and reverse fastq files. Adapter clipping of the reads
107 was performed using seqpurge v0.1-852-g5a7f2d2¹⁵ and read sets from each sample were
108 assembled using metaSPAdes v3.9.0¹⁶. Reads from each sponge species, or seawater source,
109 were mapped in an all-versus-all manner using BamM v1.7.3
110 (<https://github.com/Ecogenomics/BamM>), a wrapper script leveraging the BWA mapping
111 algorithm¹⁷, to obtain BAM files for differential coverage estimation. Binning was performed
112 using uniteM v0.0.16 (<https://github.com/dparks1134/UniteM>). In brief, UniteM uses the
113 binning algorithms maxbin v2 v2.2.4, metabat v1 v0.32.4 using all parameter sets (e.g. –
114 verysensitive, etc.), metabat v2 v2.12.1, and groopM2 v2 v2.0.0-1, and selects the best
115 MAGs by their CheckM¹⁸ quality scores. For *Aplysina aerophoba*, six previously published⁵
116 datasets (PRJNA366444-PRJNA366449 and PRJNA326328) were co-assembled using
117 Megahit v1.1.3¹⁹ and binning was performed with metaWRAP v1.0.1, which uses metabat,
118 metabat v2, and maxbin v2. Taxonomy was assigned to each MAG using GTDB-Tk²⁰, which
119 classifies MAGs based on placement in a reference tree inferred using a set of 120 bacterial
120 and 122 archaeal concatenated gene markers using a combination of FASTANI and
121 pplacer^{21,22}. GTDB-Tk annotation is based on the Genome Taxonomy Database (GTDB)
122 taxonomy²³ (<http://gtdb.ecogenomic.org>). The relative abundance of each MAG from
123 separate individuals was calculated with coverM v0.2.0 (<https://github.com/wwood/CoverM>),

124 after de-replicating all 1188 MAGs at 95% identity with dRep v2.2.4²⁴ to avoid arbitrary
125 mapping between representatives of highly similar genomes. Information regarding the
126 MAGs (taxonomy, MIMAG standard statistics, etc) can be found in **Table S1**. MAGs with an
127 overall abundance >5% in at least one sample were included in the heat map visualised with
128 R v3.5.1²⁵ (**Fig. S1**).

129

130 *Identification and analysis of CuMMO family genes*

131 GraftM²⁶ was used to recover CuMMO family genes (GraftM package ID: 7.22) from the
132 sponge-associated MAGs and published metagenomic dataset from sponges where bacterial
133 *amo* genes had previously been reported²⁷. Briefly, GraftM uses gene-specific (e.g. *amoA*)
134 hidden Markov models (HMMs) to identify and extract sequences from metagenomic reads
135 or assemblies and inserts them into a reference tree to assign them to pre-defined functional
136 clades. A phylogenetic tree was then inferred from these sequences with IQ-TREE v1.6.12²⁸,
137 including previously published CuMMO sequences^{29,30} after alignment with MAFFT
138 v7.221³¹. The tree was rooted between the domains Archaea and Bacteria, grouped into
139 functional clades based on the grouping in Alves *et al.* (**Fig. S2**)²⁹, and refined in iTOL
140 v4.2.3³².

141

142 *Gene annotation and statistics*

143 To ensure that the observed presence/absence of genes was not influenced by MAG
144 completeness, only a set of highly complete MAGs (>85% completeness) were annotated
145 with the “annotate” function of EnrichM v0.2.1 (<https://github.com/geronimp/enrichM>) using
146 the Pfam database to identify eukaryote-like repeat proteins and the Kyoto Encyclopedia of
147 Genes and Genomes (KEGG) Orthologies (KOs) to reconstruct metabolic pathways^{33,34}, and
148 the HMMs from dbCAN³⁵ to identify carbohydrate active enzymes (CAZY) like glycosyl

149 hydrolases and carbohydrates esterases³⁶. EnrichM’s “annotate” function was also used to
150 identify orthologous clusters between highly complete MAGs (**Fig. S3**). For amino acid
151 synthesis pathways, EnrichM’s “classify” function was used to calculate the completeness of
152 KEGG modules, which are groups of genes organised by steps in a metabolic pathway as
153 defined by KEGG. To identify genes and pathway enriched in sponge-associated MAGs and
154 therefore may represent functions important for sponge-microbe symbiosis, EnrichM’s
155 “enrichment” function (<https://github.com/geronimp/enrichM>) was used to perform statistical
156 tests for the enrichment of KEGG modules, CAZY genes, and Pfams between highly
157 complete sponge and seawater-associated MAGs (**Tables S2, S3, and S4**). Enrichment of
158 KOs within each module were calculated individually using a two-sided Mann-Whitney U-
159 test and only modules for which >70% of KOs were enriched were further considered. For
160 Pfam and CAZY enrichment calculations, EnrichM directly compares the number of proteins
161 per MAG matching that Pfam using a two-sided t-test. For both KEGG and Pfam
162 comparisons, a Benjamini–Hochberg correction was applied to control for the false discovery
163 for multiple comparisons. The ggplot package was used through EnrichM to make PCA plots
164 based on the gene content each MAG (KEGG, Pfam, orthologous clusters of genes).

165

166 *Tree building and visualisation*

167 To visualise the phylogenetic distribution of the recovered MAGs, phylogenetic trees were
168 constructed by supplying the bacterial and archaeal concatenated marker gene alignments
169 produced by GTDB-Tk to IQ-TREE using the LG+G model²⁸. Bacterial and archaeal trees
170 were initially constructed from all sponge-associated MAGs with >50% completeness and
171 <10% contamination, including those from the literature (**Fig. 1**). To determine how much
172 additional phylogenetic diversity the MAGs from this study added to the trees, phylogenetic
173 distance, defined as the “total branch length spanned by a set of taxa” and phylogenetic gain,

174 defined as the “additional branch length contributed by a set of taxa”²³, were calculated for
175 both the bacterial and archaeal tree using GenomeTreeTk v0.0.41
176 (<https://github.com/dparks1134/GenomeTreeTk>). A second tree constructed from all MAGs
177 with >85% completeness, including those from seawater, was used as a basis for visualising
178 the distribution of genes-of-interest across lineages using iTOL³² (**Figs. 2-3 and S4-S7**).

179

180 *Identification of lateral gene transfers*

181 To identify laterally transferred genes within the sponge-associated microbial communities,
182 MAGs associated with the same sponge species, as well as all seawater-derived MAGs, were
183 first dereplicated with dRep v2.3.2²⁴ at 99% average nucleotide identity (ANI). LGT analysis
184 was performed using MetaCHIP v1.7.5³⁷ on all dereplicated MAGs at phylum, class, order,
185 family and genus levels. In brief, MetaCHIP first clusters query genomes according to their
186 phylogenies, then performs an all-against-all BLASTN for all predicted genes. The BLASTN
187 matches for each gene are then compared among taxa and a gene is considered to be a
188 putative LGT, if the best-match comes from a non-self taxa. A phylogenetic approach is then
189 applied to putative LGTs for further validation using Ranger-DTL³⁸ and the direction of gene
190 flow is identified. False-positive LGTs, as could be introduced through MAG contamination,
191 were filtered out by removing LGTs marked by MetaCHIP as “full-length match” (LGT
192 makes up large proportion of contig) or “end match” (LGT falls at end of contig). For the
193 current dataset, LGTs on average made up only $12 \pm 14\%$ of the total length of their
194 respective contigs, ensuring that binning was based on compositional information (e.g. k-
195 mers as used by MetaBat) from the non-LGT portion of the contig. LGTs detected at all
196 taxonomic levels were combined and dereplicated. The genetic divergence of the dereplicated
197 LGTs, as well as the frequency of LGT transfer (the number of LGT per Mbp sequences) per

198 MAG, were summarized on the basis of taxon or host of MAGs and visualized using the
199 Circlize package in R (**Figs. S8 and S11**).

200

201 **Results and Discussion**

202 Six sponge species, *Rhopaloeides odorabile*, *Coscinoderma matthewsi*,
203 *Carteriospongia foliascens*, *Stylissa flabelliformis*, *Ircinia ramosa*, and *Cliona orientalis* (a
204 bioeroding sponge) were selected for metagenomic sequencing (7 ± 0.5 Gbp) as these species
205 represent dominant habitat forming taxa on tropical and temperate Australian reefs and
206 exhibit high intra-species similarity in their microbiomes. In addition, previously published
207 microbial MAGs from *Ircinia ramosa* and *Aplysina aerophoba* were analysed^{5,6}, including 62
208 additional unpublished MAGs from *A. aerophoba*. The recovered MAGs, averaging $86 \pm$
209 12% completeness and $2 \pm 2\%$ contamination, made up $72 \pm 21\%$ relative abundance of their
210 respective communities (by read mapping) on average and spanned the vast majority of
211 microbial lineages typically seen in marine sponges³⁹ (**Fig. S1**, **Table S1**), including the
212 bacterial phyla Proteobacteria (331 MAGs), Chloroflexota (242), Actinobacteriota (155),
213 Acidobacteriota (97), Gemmatimonadota (60), Latescibacterota (44; including lineages
214 Anck6, PAUC34, and SAUL), Cyanobacteria (43), Bacteroidota (38), Poribacteria (35
215 MAGs), Dadabacteria (22; including SBR1093), Nitrospirota (22), Planctomyctota (15),
216 UBP10 (14), Bdellovibrionota (13), Patescibacteria (9; includes Candidate Phylum
217 Radiation), Spirochaetota (8), Nitrospinota (7), Myxococcota (4), Entotheonella (2), and the
218 archaeal class Nitrososphaeria (21; phylum Crenarchaeota), hereafter referred to by their
219 historical name “Thaumarchaeota” for name recognition. Mapping of the metagenomic reads
220 to the recovered MAGs showed that the communities had high intraspecies similarity across
221 replicates, consistent with previous 16S rRNA gene-based analyses (**Fig. S1**). In general, taxa
222 present in *A. aerophoba*, *C. foliascens*, *C. orientalis*, and *S. flabelliformis* appeared unique to

223 those sponge species, with only one dominant lineage present in *C. orientalis* (order
224 Parvibaculales). In contrast, several Actinobacteriota, Acidobacteriota, and Cyanobacteria
225 populations were shared across *C. matthewsi*, *R. odorabile*, and *I. ramosa*. Further, members
226 of the Thaumarchaeota were detected in all sponge species and were particularly abundant in
227 *S. flabelliformis* at $12 \pm 4\%$ relative abundance (**Fig. S1**). Addition of these sponge MAGs to
228 genome trees comprising all publicly available sponge symbionts resulted in a phylogenetic
229 gain of 44 and 75% for Bacteria and Archaea, respectively, reflecting substantial novel
230 genomic diversity (**Fig. 1**).

231 Comparative genomic analysis of the sponge-derived MAGs provided unique insights
232 into the distribution of metabolic pathways across sponge symbiont taxa. For example,
233 microbial oxidation of ammonia benefits the sponge host by preventing ammonia from
234 accumulating to toxic levels⁴⁰, a process thought to be mediated by both symbiotic Bacteria
235 and Archaea (i.e. Thaumarchaeota)²⁷. Prior identification of ammonia oxidizers has been
236 based on functional inference from phylogeny (16S rRNA gene amplicon surveys) or
237 homology to specific Pfams (metagenomes). However, the copper-dependent membrane-
238 bound monooxygenase gene (CuMMO) family is diverse, encompassing functionally distinct
239 relatives that include ammonia monooxygenases (*amoA*), particulate methane
240 monooxygenases (*pmoA*), and hydrocarbon monooxygenases (*hmoA*) that cannot be
241 distinguished by homology alone²⁹. We used GraftM²⁶ to recover CuMMO genes from the
242 sponge MAGs and their metagenomic assemblies, as well as previously sequenced
243 metagenomic assemblies from six additional sponge microbiomes where bacterial *amoA* gene
244 sequences had been identified²⁷. Phylogenetic analysis of the recovered CuMMO genes
245 showed that all archaeal homologs came from Thaumarchaeota and fell within the archaeal
246 *amoA* clade. In contrast, bacterial CuMMO sequences were identified exclusively in MAGs
247 from the phylum UBP10 (formerly unclassified Deltaproteobacteria) and from an unknown

248 taxonomic group in the previous metagenomic assemblies²⁷. All recovered bacterial and
249 taxonomically unidentified CuMMO placed within the Deltaproteobacteria/Actinobacteria
250 *hmo* clade, indicating these genes are specific for hydrocarbons rather than ammonia (**Fig.**
251 **S2**). The finding that Thaumarchaeota are the only microbes within any of the surveyed
252 sponge species capable of oxidizing ammonia, and their ubiquity across sponges, suggests
253 they are a keystone species for this process.

254 To further investigate the distribution of functions within the sponge microbiome, a
255 set of highly complete (>85%) sponge symbiont MAGs were grouped by principal
256 components analysis based on their KEGG and Pfam annotations, as well as orthologous
257 clusters that reflected all gene content. Similar analysis conducted on 37 MAGs from the
258 sponge *Aplysina aerophoba* suggested the presence of functional guilds, with MAGs from
259 disparate microbial phyla carrying out similar metabolic processes⁵ (e.g. carnitine
260 catabolism). Here, we find that MAGs clustered predominately by microbial taxonomy
261 (phylum) rather than function in all three analyses (**Fig. S3**). While functional guilds could
262 not be identified based on analysis of total genome content, this does not preclude the
263 existence of such guilds based on more specific metabolic pathways.

264 To identify pathways enriched within the sponge microbiome, sponge-associated
265 MAGs (N = 798) were compared with a set of coral-reef seawater MAGs (N= 86), 31 derived
266 from published datasets¹³ and 55 from this study (**Table S1**). Seawater MAGs with >85%
267 genome completeness (93 ± 4% completeness and 2 ± 2% contamination; **Table S1**) spanned
268 the bacterial phyla Proteobacteria (48 MAGs), Bacteroidota (13), Planctomycetota (5),
269 Myxococcota (5), Gemmatimonadota (3), Marinisomatota (3), Actinobacteriota (3),
270 Verrucomicrobiota (2), Cyanobacteriota (2), Bdellovibrionota (1), and the archaeal phylum
271 Nanoarchaeota (1). Comparative analysis revealed that sponge symbionts were enriched in
272 metabolic pathways for carbohydrate metabolism, defence against infection by mobile

273 genetic elements, amino acid synthesis, eukaryote-like gene repeat proteins (ELRs), and cell-
274 cell attachment (**Tables S2-S4**).

275 Genes belonging to glycosyl hydrolase (GH) and carbohydrate esterase (CE) families
276 (**Table S2**) acting on starch (GH77), arabinose (CAZY families GH127 and GH51), fucose
277 (GH95 and GH29), and xylan polymers (CE7 and CE15), were enriched in sponge-associated
278 lineages, likely reflecting the hosts critical role in catabolising dissolved organic matter
279 (DOM) present in reef seawater (**Fig. 2**). Microbial glycosyl hydrolases from the GH77
280 family target starch, the main sugar storage compound in marine algae⁴¹, whereas GHs from
281 families 51 and 127 are known to act on plant arabinosaccharides such as the hydroxyproline-
282 linked arabinosaccharides found in algal extensin glycoproteins^{42,43}. GH127 enzymes are also
283 required for microbial degradation of carrageenan, a complex heteropolysaccharide produced
284 by red algae⁴⁴. Members of the fucosidase GH95 and GH29 enzyme families are known to
285 degrade fucoidan, a complex fucosaccharide prominent in brown algae^{43,45}. Interestingly,
286 arabino- and fucopolysaccharides also make up a significant proportion of coral mucus, a
287 major component of DOM in coral reefs that sponges have been shown to utilise^{46,47}.
288 Supporting this observation, isotopic investigation of the fate of coral mucus and algal
289 polysaccharides in sponges showed that the microbiome participates in metabolism of these
290 compounds, particularly in sponges with high microbial abundance and diversity^{48,49}.
291 Enzymes from the carbohydrate esterase families 15 and 7 have been primarily characterised
292 in terrestrial plants where they act as glucuronyl esterases and acetyl-xylan esterases,
293 degrading lignocellulose and removing acetyl groups from hemicellulose⁵⁰ (e.g. xylyns).
294 Characterisation of CE15 and CE7 from marine microbes is rare, though activity on xylyns,
295 which are a structural component of marine algae, has previously been demonstrated⁵⁰⁻⁵².

296 GHs acting on sialic acids (GH33) and glycosaminoglycans (GH88) were also
297 enriched in the sponge-associated MAGs and may act on compounds found within sponge

tissue⁷ (**Fig. 2**). In contrast, no genes for the degradation of collagen (collagenases), one of the main structural components of the sponge skeleton were identified. Sialic acid-linked residues are found in the sponge mesohyl⁵³, and although the impact of cleavage on the host is unknown, analogy can be made to other symbioses. For example, sialidases are common in the commensal bacteria present in the human gut where they are used to cleave and metabolise the sialic acid-containing mucins lining the gut wall⁵⁴. Increased sialidase activity is associated with gut dysbiosis and inflammation⁵⁵ and careful control of sialidase-containing commensals is therefore necessary to maintain gut homeostasis⁵⁴. As glycosaminoglycans are also part of sponge tissue^{7,56}, the same may apply to microorganisms encoding GH88 family enzymes. However, these genes are also implicated in degradation of external sugar compounds such as ulvans, a major sugar storage compound found in green algae that can make up to 30% of their dry weight⁵⁷. Thus, the ecological role of GH88 family enzymes within the sponge microbiome requires further investigation.

Enrichment of GHs and CEs was largely restricted to the Poribacteria, Latescibacteria (class UBA2968), Spirochaetota, Chloroflexota (classes UBA2235 and Anaerolineae, but not Dehalococcoida), and Acidobacteriota (class Acidobacteriae). These findings corroborate previous targeted genomic characterisations of the Chloroflexota and Poribacteria^{7,8} but show that they are part of a larger set of polysaccharide-degrading lineages. Identification of disparately related microbial taxa across several sponge lineages (**Figs. 1 and 2**) that encode similar pathways for polysaccharide degradation, and therefore occupy a similar ecological niche, supports the existence of functional guilds within the sponge microbiome when viewed at the level of individual pathways. Given the fundamental role of marine sponges in recycling coral reef DOM, studies targeting these specific guilds are needed to quantify their contribution to reef DOM transformation.

Because sponges filter and retain biomass from an extensive range of reef taxa (eukaryotic algae, bacteria, archaea, etc), they are exposed to a greatly expanded variety of mobile genetic elements (MGE) from these organisms, including viruses, transposable elements, and plasmids^{27,58}. For this reason, sponge-associated microorganisms likely require a diverse toolbox of molecular mechanisms for resisting infection. Both restriction modification (RM) and clustered regularly interspaced short palindromic repeat (CRISPR) systems are capable of recognising and cleaving MGEs as part of the bacterial immune repertoire. RM systems are part of the innate immune system of bacteria and archaea and are encoded by a single (Type II) or multiple proteins (Type I, III, and IV) that recognise and cleave foreign DNA based on a defined target sequence. In contrast, CRISPR systems are part of the adaptive immune system of some bacteria and archaea and encode a target sequence derived from the genome of a previous infective agent that is used by a CRISPR-associated protein (CAS) to identify and cleave foreign DNA. RM (**Fig. S4**) and CAS (**Fig. S5**) genes were both enriched in the sponge-associated MAGs and relatively evenly distributed across taxa, with the exception of the Planctomycetota and Verrucomicrobiota, where they were largely absent. As these MAGS average 93 +/- 5% completeness, this result is not likely due to genome incompleteness. This finding contrasts with comparative investigations of Planctomycetota genomes from other environments⁵⁹ and additional research is required to ascertain the mechanisms used by sponge-associated Planctomycetota and Verrucomicrobiota to avoid infection. Although Type III RM genes were enriched in sponge MAGs, they were also present in all seawater MAGs. In contrast, Types I and II RM genes were present almost exclusively in the sponge-associated MAGs. In conjunction with an enrichment in CRISPR systems, this expanded repertoire of defence systems likely reflects the increased burden from MGEs associated with the hosts role in filtering and concentrating diverse sources of reef biomass. Supporting this hypothesis, metagenomic surveys of sponge-

347 associated viruses revealed a more diverse viral population than what could be recovered
348 from the surrounding seawater⁵⁸. Further, we found that genes encoding toxin-antitoxin (TA)
349 systems, which are present on MGEs such as plasmids, were also enriched in sponge-
350 associated MAGs. These observations suggest that RM and CRISPR systems are important
351 features of microbe-sponge symbiosis, allowing the symbionts to colonise and persist within
352 their host by avoiding viral infection or being overtaken by MGEs.

353 Pathways for the synthesis of amino acids were also enriched in the sponge
354 microbiome. The inability of animals to produce several essential amino acids has been
355 proposed as a primary reason that they harbor microbial symbionts^{60–63} and it has long been
356 thought that sponges acquire at least some of their essential amino acids from their
357 microbiome^{64,65}. Further, gene-centric characterisation of the *Xestospongia muta* and *R.*
358 *odorabile* microbiomes revealed pathways to synthesize and transport essential amino
359 acids^{27,65}. However, these same amino acid pathways are also used catabolically by the
360 microorganisms, and transporters could simply be importing amino acids into the microbial
361 cell. Further, as sponges are almost constantly filter feeding, essential amino acids could be
362 acquired through consumption of microorganisms present in seawater. Comparison of sponge
363 MAGs with those from seawater revealed enrichment of specific pathways for the synthesis
364 of lysine, arginine, histidine, threonine, valine, and isoleucine (**Table S4**). However,
365 visualisation of the distribution of these genes revealed that almost all MAGs in both sponges
366 and seawater produce all amino acids, though specific lineages may use different pathways to
367 achieve this (**Fig. S6**). The enrichment observed in the sponge MAGs was therefore ascribed
368 to differences in pathway completeness between sponge-associated and seawater microbes,
369 rather than an enhanced ability of sponge symbionts to produce any specific amino acid. In
370 contrast, compounds such as taurine, carnitine and creatine have also been proposed as
371 important host-derived carbon sources for symbionts⁶⁴, but pathways for their catabolism

372 were enriched in seawater rather than sponge-associated MAGs. While these findings do not
373 invalidate the possibility that microbial communities play a role in amino acid provisioning to
374 the host or that they utilise host-derived taurine, carnitine, or creatine, they suggest that these
375 are not key processes mediating microbe-sponge symbiosis.

376 To form stable symbioses, bacteria must persist within the sponge tissue and avoid
377 phagocytosis by host cells. Microbial proteins containing eukaryote-like repeat (ELR) motifs
378 have been identified in a range of animal and plant-associated microbes and are thought to
379 modulate the host's intracellular processes to facilitate stable symbiotic associations^{66,67}. For
380 example, ELR-containing proteins from sponge-associated microbes have been shown to
381 confer the ability to evade host phagocytosis when experimentally expressed in *E. coli*^{68,69}.
382 ELR-containing proteins from the ankyrin (ARP), leucine-rich (LRR), tetratricopeptide
383 (TPR), and HEAT repeat families were enriched in the sponge-associated MAGs. In contrast,
384 WD40 repeats were not found to be enriched but are included here as they have previously
385 been reported as abundant in Poribacteria and symbionts of other marine animals^{7,13}. Most
386 ELRs were present across all taxa but were much more prevalent in specific lineages (**Fig. 3**).
387 For example, sponge-associated Poribacteria, Latescibacterota, and Acidobacteriota encoded
388 a high proportion of all ELR types, while other lineages such as the Gemmatimonadota
389 (average 0.25% coding genes per sponge-associated MAG versus 0.09% in seawater MAGs),
390 Verrucomicrobiota (2%), Deinococcota (0.85%), Acidobacteriota (0.20%; specifically class
391 Luteitaleia at 0.55%), Dadabacteria from *C. orientalis* (0.62%) encoded a comparatively high
392 percentage of ARPs and Nitrospirota encoded a high percentage of HEAT_2 family proteins
393 (0.55% versus 0.05% in seawater MAGs) relative to other taxa. In contrast, ELR abundances
394 were substantially lower, or absent, in the Actinobacteriota, the class Bacteroidia within the
395 phylum Bacteroidota, and the Thaumarchaeota, suggesting these microorganisms utilize
396 alternative mechanisms to maintain their stable associations with the host.

397 The mechanisms by which ELRs interact with sponge cells remains largely unknown,
398 although microbes in other host systems are known to deliver ELR-containing effector
399 proteins into host cells via needle-like secretion systems (types III, IV, and V) or extracellular
400 contractile injection systems (eCIS)^{70,71}, where they interact with the cellular machinery of
401 the host to modify its behaviour. In sponges, it is also possible that ELRs could be secreted
402 into the extracellular space by type I or II secretion systems. Interestingly, although most
403 sponge MAGs encoded eukaryote-like proteins (**Fig. 3**), few lineages encoded the necessary
404 genes to form secretion systems (**Fig. S7**). It is therefore unlikely that ELRs are introduced to
405 the sponge host via traditional secretion pathways used in other animal-symbiont systems.

406 Maintaining stable association with the sponge may also require mechanisms for
407 attachment to the host tissue. For example, cadherin domains are Ca²⁺ dependent cell-cell
408 adhesion proteins that are abundant in eukaryotes and have been found to serve the same
409 function in bacteria⁷². Similarly, fibronectin III domains mediate cell adhesion in eukaryotes,
410 but also occur in bacteria where they play various roles in carbohydrate binding and biofilm
411 formation^{73,74}. In addition, some bacterial pathogens utilise fibronectin-binding proteins to
412 gain entry into host tissue by binding to host fibronectin^{73,74}. Genes containing cadherin
413 domains were enriched in the sponge-associated MAGs and were identified in most bacterial
414 lineages, but were notably absent in the Cyanobacteriota and Verrucomicrobiota (**Fig. 4**).
415 Genes containing fibronectin III domains and those for fibronectin-binding proteins were also
416 enriched in sponge-associated MAGs and were distributed across most lineages, though were
417 particularly abundant in the Actinobacteriota and Chloroflexota. However, although
418 fibronectin III-containing genes were taxonomically widespread, those encoding fibronectin-
419 binding proteins were restricted to the phyla Poribacteria, Gemmatimonadota,
420 Latescibacterota, Cyanobacteriota, class Anaerolineae within the Chloroflexota (but not
421 Dehalococcoidia), class Rhodothermia within the Bacteroidota, Spirochaetota, Nitrospirota,

422 and the archaeal phylum Thaumarchaeota. Interestingly, the taxonomic distribution of these
423 genes shares significant overlap with lineages encoding the genes for sponge sialic acid and
424 glycosaminoglycans degradation, suggesting that attachment to the host may be necessary for
425 utilisation of these carbohydrates (**Fig. 2**). However, as all components of the sponge
426 holobiont (host, bacteria, archaea) have fibronectin domains, symbionts encoding fibronectin-
427 binding proteins may use these to adhere to the host tissue or potentially to form biofilms
428 (bacteria-bacteria attachment). In either case, the enrichment and wide distribution of
429 cadherins, fibronectins, and fibronectin-binding proteins in the sponge MAGs suggests that
430 cell-cell adhesion is critical for successful establishment in the sponge niche.

431 Distribution of genes encoding ELRs, polysaccharide-degrading enzymes (GHs and
432 CEs), cadherins, fibronectins, RMs, and CRISPRs across distantly related taxa suggests that
433 they were either acquired from a common ancestor or that they represent more recent lateral
434 gene transfer (LGT) events, potentially mediated by MGEs, which are enriched in sponge-
435 associated microbial communities⁶⁴. Here, we identify 4963 LGTs from five sponges for
436 which sufficient sequence data was available (>100Mbp total MAG sequence length), as well
437 as 136 LGTs from seawater MAGs, averaging 1.64 and 0.52 LGTs per Mbp sequences
438 respectively (**Fig. S8, Table S5**). Sequence similarity of LGTs from MAGs within a sponge
439 species was higher than between sponge species, indicating relatively recent gene transfers
440 (**Fig. S9**). A higher frequency (**Fig. S10**) and lower genetic divergence of LGTs among
441 MAGs derived from the same sponge species likely results from the close physical distance
442 between members of each microbiome, as has been observed in other host-symbiont
443 systems^{7d5}. LGTs included a subset of genes that were enriched within the sponge-associated
444 MAGs, such as GH33 (sialidases) and CE7 (acetyl-xylan esterases), attachment proteins
445 (cadherins and fibronectin III), RM and CAS proteins, and members of all ELR families
446 other than WD40 (**Fig. 5, Fig. S11**). The observation that a significant number of sponge-

447 enriched genes were laterally transferred between disparate microbial lineages suggests that
448 the processes they mediate provide a strong selective advantage within the sponge niche.

449 Sponges are important constituents of coral reef ecosystems because of their critical
450 role in DOM cycling and retention via the sponge-loop. Despite their importance, functional
451 characterisation of sponge symbiont communities has been restricted to just a few lineages of
452 interest, potentially biasing our view of sponge symbiosis. Here we present a comprehensive
453 characterisation of sponge symbiont MAGs spanning the complete range of taxa found in
454 marine sponges (**Fig. 6**), most of which were previously uncharacterised. We revealed
455 enrichment in glycolytic enzymes (GHs and CEs) reflecting specific functional guilds
456 capable of aiding the sponge in degradation of reef DOM. Further, we identified several
457 ELRs, CRISPRs, and RMs that likely facilitate stable association with the sponge host,
458 showing specificity of ELR types with individual microbial lineages. We also clarified the
459 role of Thaumarchaeota as a keystone taxon for ammonia oxidation and showed that
460 processes previously thought to be important, such as amino acid provisioning and taurine,
461 creatine, and carnitine metabolism are unlikely to be central mechanisms mediating sponge-
462 microbe symbiosis. Many of the enriched genes are laterally transferred between microbial
463 lineages, suggesting that LGT plays an important role in conferring a selective advantage to
464 specific sponge-associated microorganisms. Taken together, these data illustrate how
465 evolutionary processes have distributed and partitioned ecological functions across specific
466 sponge symbiont lineages, allowing them to occupy or share specific niches and live
467 symbiotically with their evolutionarily ancient hosts.

468

469 *Acknowledgments*

470 We are indebted to Illumina and the Earth Microbiome Project for metagenomic sequencing,
471 including these individuals: Luke Thompson, Jon Sanders, Rodolfo Salido Benitez, Karenina

472 Sanders, Caitriona Brennan, Jeremiah Minich, MacKenzie Bryant, Lindsay DeRight
473 Goldasich, Greg Humphrey and Rob Knight. Mari Carmen Pineda Torres and the AIMS
474 Seasmim staff are thanked for sample collection. Torsten Thomas and Wei Song were
475 additionally supported by an R&D project registered as ANP 21005-4, “PROBIO-DEEP -
476 Survey of potential impacts caused by oil and gas exploration on deep-sea marine holobionts
477 and selection of potential bioindicators and bioremediation processes for these ecosystems”
478 (UFRJ / Shell Brasil / ANP), sponsored by Shell Brasil under the ANP R&D levy as
479 “Compromisso de Investimentos com Pesquisa e Desenvolvimento”.

480

481 *Competing interests*

482 The authors declare that they do not have any competing financial interests in relation to the
483 work described.

484

485 *Data availability*

486 Metagenomic assemblies and MAGs from this study can be found under NCBI bioproject ID
487 PRJNA602572, accessible via a temporary reviewer link
488 (<https://dataview.ncbi.nlm.nih.gov/object/PRJNA602572?reviewer=oqh7hjopun20l8k82c5a5qm9jm>). MAGs are currently in the process of upload. Biosample accessions for each MAG
489 are listed in **Table S1** and will be amended to genome accessions before publication. All
490 GraftM packages can be found at <https://data.ace.uq.edu.au/public/graftm/7/>.

492

493 **References**

- 494 1. Knowlton, N. Coral Reef Biodiversity--Habitat Size Matters. *Science* (80-.). **292**,
495 1493 LP – 1495 (2001).
- 496 2. Hentschel, U., Usher, K. M. & Taylor, M. W. Marine sponges as microbial fermenters.

- 497 *FEMS Microbiol. Ecol.* **55**, 167–177 (2006).
- 498 3. De Goeij, J. M. *et al.* Surviving in a marine desert: The sponge loop retains resources
499 within coral reefs. *Science (80-.).* **342**, 108–110 (2013).
- 500 4. Hentschel, U., Piel, J., Degnan, S. M. & Taylor, M. W. Genomic insights into the
501 marine sponge microbiome. *Nat. Rev. Microbiol.* **10**, 641–654 (2012).
- 502 5. Slaby, B. M., Hackl, T., Horn, H., Bayer, K. & Hentschel, U. Metagenomic binning of
503 a marine sponge microbiome reveals unity in defense but metabolic specialization.
504 *ISME J.* **11**, 2465–2478 (2017).
- 505 6. Engelberts, J. P. & , Steven J. Robbins, Jasper M. de Goeij, Manuel Aranda, Sara C.
506 Bell, N. S. W. Characterisation of a sponge microbiome using an integrative genome-
507 centric approach. *ISME J*
- 508 7. Kamke, J. *et al.* Single-cell genomics reveals complex carbohydrate degradation
509 patterns in poribacterial symbionts of marine sponges. *ISME J* **7**, 2287–2300 (2013).
- 510 8. Bayer, K., Jahn, M. T., Slaby, B. M., Moitinho-Silva, L. & Hentschel, U. Marine
511 Sponges as Chloroflexi Hot Spots: Genomic Insights and High-Resolution
512 Visualization of an Abundant and Diverse Symbiotic Clade . *mSystems* **3**, 1–19 (2018).
- 513 9. Lackner, G., Peters, E. E., Helfrich, E. J. N. & Piel, J. Insights into the lifestyle of
514 uncultured bacterial natural product factories associated with marine sponges. *Proc.
515 Natl. Acad. Sci. U. S. A.* **114**, E347–E356 (2017).
- 516 10. Montalvo, N. F. *et al.* Integration of culture-based and molecular analysis of a complex
517 sponge-associated bacterial community. *PLoS One* **9**, 1–8 (2014).
- 518 11. Botté, E. S. *et al.* Changes in the metabolic potential of the sponge microbiome under
519 ocean acidification. *Nat. Commun.* **10**, 1–10 (2019).
- 520 12. Glasl, B. *et al.* Comparative genome-centric analysis reveals seasonal variation in the
521 function of coral reef microbiomes. *ISME J.* (2020). doi:10.1038/s41396-020-0622-6

- 522 13. Robbins, S. J. *et al.* A genomic view of the reef-building coral *Porites lutea* and its
523 microbial symbionts. *Nature Microbiology* **4**, 2090–2100 (2019).
- 524 14. Marotz, C., Amir, A., Humphrey, G., Gogul, G. & Knight, R. IN IT IS IO DNA
525 extraction for streamlined. **293**, 290–293 (2017).
- 526 15. Sturm, M., Schroeder, C. & Bauer, P. SeqPurge: Highly-sensitive adapter trimming for
527 paired-end NGS data. *BMC Bioinformatics* **17**, 1–7 (2016).
- 528 16. Nurk S, Meleshko D, Korobeynikov A, P. P. metaSPAdes: A New Versatile
529 Metagenomic Assembler. *Genome Res.* **1**, 30–47 (2017).
- 530 17. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler
531 transform. *Bioinformatics* **25**, 1754–1760 (2009).
- 532 18. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W.
533 CheckM: assessing the quality of microbial genomes recovered from isolates, single
534 cells, and metagenomes. *Genome Res* gr. 186072.114 (2015).
- 535 19. Li, D. *et al.* MEGAHIT v1.0: A fast and scalable metagenome assembler driven by
536 advanced methodologies and community practices. *Methods* **102**, 3–11 (2016).
- 537 20. Chaumeil, P., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes
538 with the Genome Taxonomy Database. *Prep.* 1–3 (2019).
- 539 doi:10.1093/bioinformatics/btz848
- 540 21. FA, M., RB, K. & EV, A. pplacer: linear time maximum-likelihood and Bayesian
541 phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*
542 **11**, 538 (2010).
- 543 22. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High
544 throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries.
545 *Nat. Commun.* **9**, 1–8 (2018).
- 546 23. Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny

- 547 substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996 (2018).
- 548 24. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. DRep: A tool for fast and
549 accurate genomic comparisons that enables improved genome recovery from
550 metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).
- 551 25. Team, R. C. A language and environment for statistical computing. Vienna, Austria: R
552 Foundation for Statistical Computing. (2014).
- 553 26. Boyd, J. A., Woodcroft, B. J. & Tyson, G. W. GraftM: a tool for scalable,
554 phylogenetically informed classification of genes within metagenomes. *Nucleic Acids
555 Res.* **46**, e59 (2018).
- 556 27. Fan, L. *et al.* Functional equivalence and evolutionary convergence in complex
557 communities of microbial sponge symbionts. *Proc. Natl. Acad. Sci. U. S. A.* **109**,
558 (2012).
- 559 28. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and
560 effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol.
561 Biol. Evol.* **32**, 268–274 (2015).
- 562 29. Alves, R. J. E., Minh, B. Q., Urich, T., Von Haeseler, A. & Schleper, C. Unifying the
563 global phylogeny and environmental distribution of ammonia-oxidising archaea based
564 on amoA genes. *Nat. Commun.* **9**, 1–17 (2018).
- 565 30. Singleton, C. M. Microbial community structural and metabolic dynamics in thawing
566 permafrost. (University of Queensland, 2018).
- 567 31. Katoh, K. & Toh, H. Recent developments in the MAFFT multiple sequence
568 alignment program. *Brief. Bioinform.* **9**, 286–298 (2008).
- 569 32. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display
570 and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245
571 (2016).

- 572 33. Finn, R. D. *et al.* Pfam: The protein families database. *Nucleic Acids Res.* **42**, 222–230
573 (2014).
- 574 34. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic
575 Acids Res.* **28**, 27–30 (2000).
- 576 35. Yin, Y. *et al.* dbCAN: a web resource for automated carbohydrate-active enzyme
577 annotation. *Nucleic Acids Res* **40**, W445–W451 (2012).
- 578 36. Cantarel, B. I. *et al.* The Carbohydrate-Active EnZymes database (CAZy): An expert
579 resource for glycogenomics. *Nucleic Acids Res.* **37**, 233–238 (2009).
- 580 37. Song, W., Wemheuer, B., Zhang, S., Steensen, K. & Thomas, T. MetaCHIP:
581 Community-level horizontal gene transfer identification through the combination of
582 best-match and phylogenetic approaches. *Microbiome* **7**, 1–14 (2019).
- 583 38. Bansal, M. S., Alm, E. J. & Kellis, M. Efficient algorithms for the reconciliation
584 problem with gene duplication , horizontal transfer and loss. *Bioinformatics* **28**, 283–
585 291 (2012).
- 586 39. Thomas, T. *et al.* Diversity, structure and convergent evolution of the global sponge
587 microbiome. *Nat. Commun.* **7**, (2016).
- 588 40. Zhang, F. *et al.* Symbiotic archaea in marine sponges show stability and host
589 specificity in community structure and ammonia oxidation functionality. *FEMS
590 Microbiol. Ecol.* **90**, 699–707 (2014).
- 591 41. Øverland, M., Mydland, L. T. & Skrede, A. Marine macroalgae as sources of protein
592 and bioactive compounds in feed for monogastric animals. *J. Sci. Food Agric.* **99**, 13–
593 24 (2019).
- 594 42. Stiger-Pouvreau, V., Bourgougnon, N. & Deslandes, E. *Carbohydrates from
595 Seaweeds. Seaweed in Health and Disease Prevention* (Elsevier Inc., 2016).
- 596 doi:10.1016/B978-0-12-802772-1.00008-7

- 597 43. Kappelmann, L. *et al.* Polysaccharide utilization loci of North Sea Flavobacteriia as
598 basis for using SusC/D-protein expression for predicting major phytoplankton glycans.
599 *ISME J.* **13**, 76–91 (2019).
- 600 44. Ficko-Blean, E. *et al.* Carrageenan catabolism is encoded by a complex regulon in
601 marine heterotrophic bacteria. *Nat. Commun.* **8**, (2017).
- 602 45. Dong, S., Chang, Y., Shen, J., Xue, C. & Chen, F. Purification, expression and
603 characterization of a novel α-L-fucosidase from a marine bacteria Wenyingzhuangia
604 fucanilytica. *Protein Expr. Purif.* **129**, 9–17 (2017).
- 605 46. Hadaidi, G., Gegner, H. M., Ziegler, M. & Voolstra, C. R. Carbohydrate composition
606 of mucus from scleractinian corals from the central Red Sea. *Coral Reefs* **38**, 21–27
607 (2019).
- 608 47. Meikle, P., Richards, G. N. & Yellowlees, D. Structural investigations on the mucus
609 from six species of coral. *Mar. Biol.* **99**, 187–193 (1988).
- 610 48. Rix, L. *et al.* Coral mucus fuels the sponge loop in warm-and cold-water coral reef
611 ecosystems. *Sci. Rep.* **6**, 1–11 (2016).
- 612 49. Rix, L. *et al.* Differential recycling of coral and algal dissolved organic matter via the
613 sponge loop. *Funct. Ecol.* **31**, 778–789 (2017).
- 614 50. De Santi, C., Willassen, N. P. & Williamson, A. Biochemical characterization of a
615 family 15 carbohydrate esterase from a bacterial marine Arctic metagenome. *PLoS*
616 *One* **11**, 1–22 (2016).
- 617 51. Baath, J. A. *et al.* Structure-function analyses reveal that a glucuronoyl esterase from
618 *Teredinibacter turnerae* interacts with carbohydrates and aromatic compounds. *J. Biol.*
619 *Chem.* **294**, 6635–6644 (2019).
- 620 52. Hettiarachchi, S. A. *et al.* Characterization of an acetyl xylan esterase from the marine
621 bacterium *Ochrovirga pacifica* and its synergism with xylanase on beechwood xylan.

- 622 *Microb. Cell Fact.* **18**, 1–10 (2019).
- 623 53. Garrone, R., Thiney, Y., Pavans de Ceccatty, M. Electron microscopy of a
624 mucopolysaccharide cell coat in sponges. *Experientia* **27**, 1324–1326 (1971).
- 625 54. Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: A mini-review.
626 *Biochem. Soc. Trans.* **44**, 166–175 (2016).
- 627 55. Huang, Y. L., Chassard, C., Hausmann, M., Von Itzstein, M. & Hennet, T. Sialic acid
628 catabolism drives intestinal inflammation and microbial dysbiosis in mice. *Nat.*
629 *Commun.* **6**, (2015).
- 630 56. Fernàndez-Busquets, X. & Burger, M. M. Circular proteoglycans from sponges: First
631 members of the spongican family. *Cell. Mol. Life Sci.* **60**, 88–112 (2003).
- 632 57. Reisky, L. *et al.* A marine bacterial enzymatic cascade degrades the algal
633 polysaccharide ulvan. *Nat. Chem. Biol.* **15**, 803–812 (2019).
- 634 58. Laffy, P. W. *et al.* Reef invertebrate viromics: diversity, host specificity and functional
635 capacity. *Environ. Microbiol.* **20**, 2125–2141 (2018).
- 636 59. Koonin, E. V., Makarova, K. S. & Wolf, Y. I. Evolutionary Genomics of Defense
637 Systems in Archaea and Bacteria. *Annu. Rev. Microbiol.* **71**, 233–261 (2017).
- 638 60. Gordon, B. R. & Leggat, W. Symbiodinium - Invertebrate symbioses and the role of
639 metabolomics. *Mar. Drugs* **8**, 2546–2568 (2010).
- 640 61. Hurst, G. D. D. Extended genomes: Symbiosis and evolution. *Interface Focus* **7**,
641 (2017).
- 642 62. Ma, N. & Ma, X. Dietary Amino Acids and the Gut-Microbiome-Immune Axis:
643 Physiological Metabolism and Therapeutic Prospects. *Compr. Rev. Food Sci. Food*
644 *Saf.* **18**, 221–242 (2019).
- 645 63. Douglas, A. E. Nutritional Interactions in Insect-Microbial Symbioses: Aphids and
646 Their Symbiotic Bacteria Buchnera. *Annu. Rev. Entomol.* **43**, 17–37 (1998).

- 647 64. Pita, L., Rix, L., Slaby, B. M., Franke, A. & Hentschel, U. The sponge holobiont in a
648 changing ocean : from microbes to ecosystems. **6**, 1–18 (2018).
- 649 65. Fiore, C. L., Labrie, M., Jarett, J. K. & Lesser, M. P. Transcriptional activity of the
650 giant barrel sponge , *Xestospongia muta* Holobiont : molecular evidence for metabolic
651 interchange. *Front. Microbiol.* **6**, (2015).
- 652 66. Carolin Frank, A. Molecular host mimicry and manipulation in bacterial symbionts.
653 *FEMS Microbiol. Lett.* **366**, 1–10 (2019).
- 654 67. Reynolds, D. & Thomas, T. Evolution and function of eukaryotic-like proteins from
655 sponge symbionts. *Mol. Ecol.* **25**, 5242–5253 (2016).
- 656 68. Jahn, M. T. *et al.* A Phage Protein Aids Bacterial Symbionts in Eukaryote Immune
657 Evasion. *Cell Host Microbe* **26**, 542-550.e5 (2019).
- 658 69. Nguyen, M. T. H. D., Liu, M. & Thomas, T. Ankyrin-repeat proteins from sponge
659 symbionts modulate amoebal phagocytosis. *Mol. Ecol.* **23**, 1635–1645 (2014).
- 660 70. Costa, T. R. D. *et al.* Secretion systems in Gram-negative bacteria: structural and
661 mechanistic insights. *Nat. Rev. Microbiol.* **13**, 343–359 (2015).
- 662 71. Chen, L. *et al.* Genome-wide Identification and Characterization of a Superfamily of
663 Bacterial Extracellular Contractile Injection Systems. *Cell Rep.* **29**, 511-521.e2 (2019).
- 664 72. Fraiberg, M., Borovok, I., Weiner, R. M. & Lamed, R. Discovery and characterization
665 of cadherin domains in *Saccharophagus degradans* 2-40. *J. Bacteriol.* **192**, 1066–1074
666 (2010).
- 667 73. Schwarz-Linek, U. *et al.* Pathogenic bacteria attach to human fibronectin through a
668 tandem β-zipper. *Nature* **423**, 177–181 (2003).
- 669 74. Hymes, J. P. & Klaenhammer, T. R. Stuck in the middle: Fibronectin-binding proteins
670 in gram-positive bacteria. *Front. Microbiol.* **7**, 1–9 (2016).
- 671 75. Smillie, C. S. *et al.* Ecology drives a global network of gene exchange connecting the

672 human microbiome. *Nature* **480**, 241–244 (2011).

673

674

675 **Figure 1.** Phylogenetic tree showing all publicly available bacterial and archaeal MAGs and
676 genomes (N = 1253 genomes) with >50% completeness and <10% contamination, recovered
677 from 30 sponge species (**Table S1**). Inner clade colour denotes phylum affiliation except for
678 the Proteobacteria, which is split by class into alpha and gamma-proteobacteria. Outer tree
679 colour strip identifies the sponge species from which the genome or MAG originated. Red stars
680 indicate MAGs produced in this study.

681

682 **Figure 2.** Phylogenetic tree showing the distribution of glycosyl hydrolases and esterases
683 across MAGS with >85% completeness (N = 884). Values represent the copy number of each
684 gene per MAG. Internal branches of the tree are coloured by phylum, while the outer strip is
685 coloured by class. Both are listed clockwise in the order in which they appear. Seawater
686 MAGs are denoted by grey labels with red text.

687

688 **Figure 3.** Phylogenetic tree showing the distribution of eukaryote-like repeat proteins—
689 ankyrin (ARP), leucin-rich (LRR), tetratricopeptide (TPR), HEAT, and WD40—across
690 MAGS with >85% completeness (N = 884). Values represent the percentage of coding genes
691 per MAG devoted to each gene class. Internal branches of the tree are coloured by phylum,
692 while the outer strip is coloured by class, and both are listed clockwise in the order in which
693 they appear. MAGs from seawater are denoted by grey labels with red text.

694

695 **Figure 4.** Phylogenetic tree showing the distribution of cadherins, fibronectins, and
696 fibronectin-binding proteins across MAGS with >85% completeness (N = 884). Values

697 represent the copy number of each gene per MAG. Internal branches of the tree are coloured
698 by phylum while the outer strip is coloured by class. Both are listed clockwise in the order in
699 which they appear. Seawater MAGs are denoted by grey labels with red text.

700

701 **Figure 5.** Visualization of gene flow among microbial phyla for gene families enriched in
702 sponge-associated MAGs. The inner ring and band connecting donor and recipient is
703 coloured by protein family of the gene being transferred, with the width of the band
704 correlating to the number of LGTs. Recipient MAGs are shown in grey. The outer ring is
705 coloured by microbial phylum. Representation of all LGTs can be found in Figure S8 and of
706 RMs and CAS genes in Figure S11.

707

708 **Figure 6.** Schematic overview of microbial interactions with the host as inferred from the
709 functional potential encoded by the sponge-associated microbial MAGs. Abbreviations:
710 fibronectin (Fbn), cadherins (cdh), restriction-modification (RM) systems, CRISPR-
711 associated proteins (CAS), eukaryotic-like repeat proteins (ELP), carbohydrate-esterase
712 family 7 (CE7), and glycosyl hydrolase family 33 (GH33).

713

714

715

716

Figure 1

Sponge species (outer strip)

- *Agelas tubulata*
- *Amphilectus fucorum*
- *Amphimedon queenslandica*
- *Aplysina aerophoba*
- *Arenosclera brasiliensis*
- *Carteriospongia foliascens*
- *Cliona orientalis*
- *Coscinoderma matthewsi*
- *Crambe crambe*
- *Cymbastela concentrica*
- *Dysidea avara*
- *Halichondria oshoro*
- *Haliclona cymaeformis*
- *Ircinia ramosa*
- *Ircinia variabilis*
- *Lophophysema eversa*
- *Melophlus sarasinorum*
- *Mycale laxissima*
- *Opheliaspongia papilla*
- *Petrosia ficiformis*
- *Polymastia penicillus*
- *Pseudoceratina sp.*
- *Rhopaloides odorabile*
- *Spongia officinalis*
- *Styliosa flabelliformis*
- *Suberites sp.*
- *Tedania sp.*
- *Theonella swinhonis*
- *unknown sponge*
- *Axinella mexicana*

Phyla (clades)

- Poribacteria
- Gemmatimonadota
- Bacteroidota
- TA06
- UBP14
- Latescibacterota
- Spirochaetota
- Verrucomicrobiota
- Planctomycetota
- Firmicutes
- Cyanobacteriota
- Deinococcota
- Actinobacteriota
- Patescibacteria
- Chloroflexota
- Acidobacteriota
- Myxococcota
- UBP10
- Entotheonellota
- Nitrospinota
- Nitrospirota
- Bdellovibrionota
- Dababacteria
- Alphaproteobacteria
- Gammaproteobacteria
- Thaumarchaeota

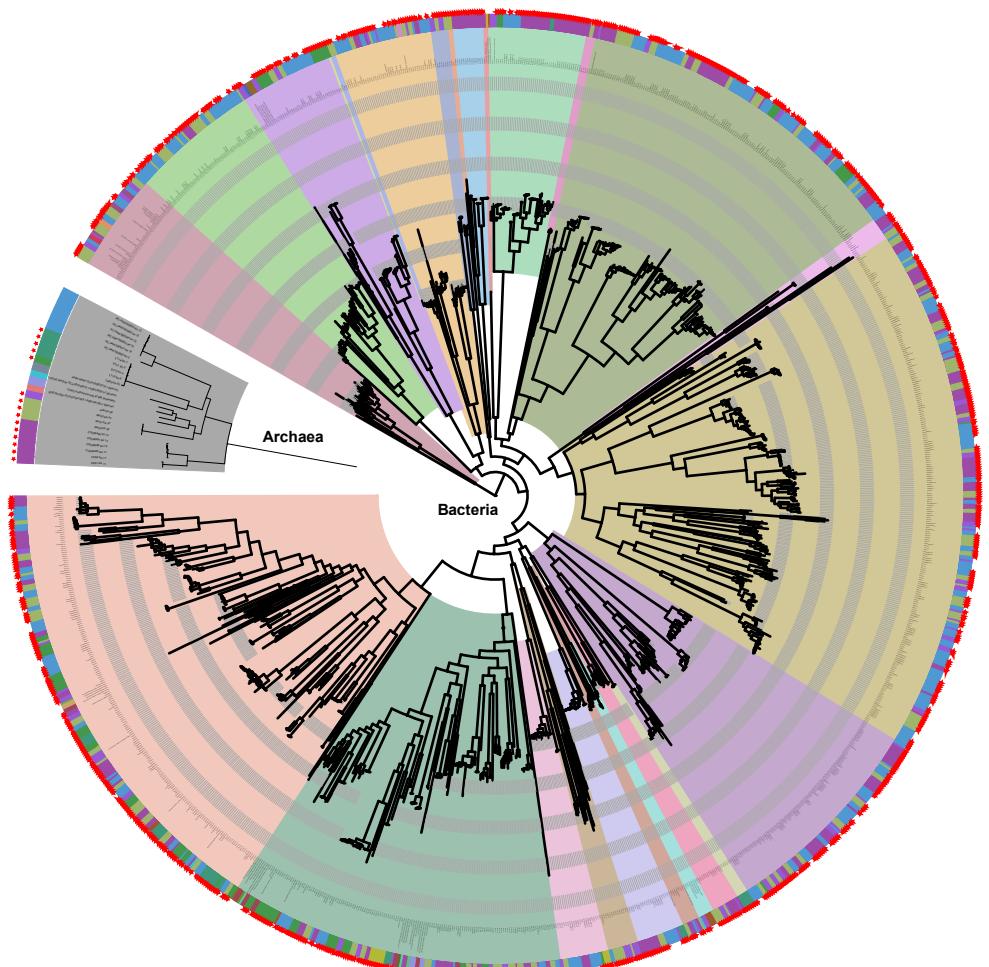


Figure 2

Phylum

Poribacteria	Chloroflexota
Gemmimonadota	Acidobacteriota
Latescibacterota	Bdellovibrionota
Marinisomatota	Myxococcota
Bacteroidota	Nitrospinota
Spirochaetota	Nitrospirota
Verrucomicrobiota	UBP10
Planctomycetota	Dababacteria
Deinococcota	Proteobacteria
Cyanobacteriota	Nanoarchaeota
Actinobacteriota	Crenarchaeota

Class

WGA-4E	Anaerolineae
Gemmimonadetes	Thermoanaerobaculida
unknown	bin61
UBA2968	Luteitaleia
Marinisomatia	unknown
Bacteroidia	Acidobacteriae
Rhodothermia	Bacteriovoracia
Spirochaeta	Bdellovibrionia
Chlamydia	Oligoflexia
Verrucomicrobiae	Polyangia
UBA8108	UBA9160
Planctomycetes	UBA8248
Deinococci	UBA8248
Vampirovibrionia	Nitrospira
Cyanobacteria	GR-WP33-30
Thermoleophilia	UBA1144
Actinobacteria	Gammaproteobacteria
Acidimicrobia	Alphaproteobacteria
UBA2235	Woesearchaeia
Dehalococcoidia	Nitrosphaeria

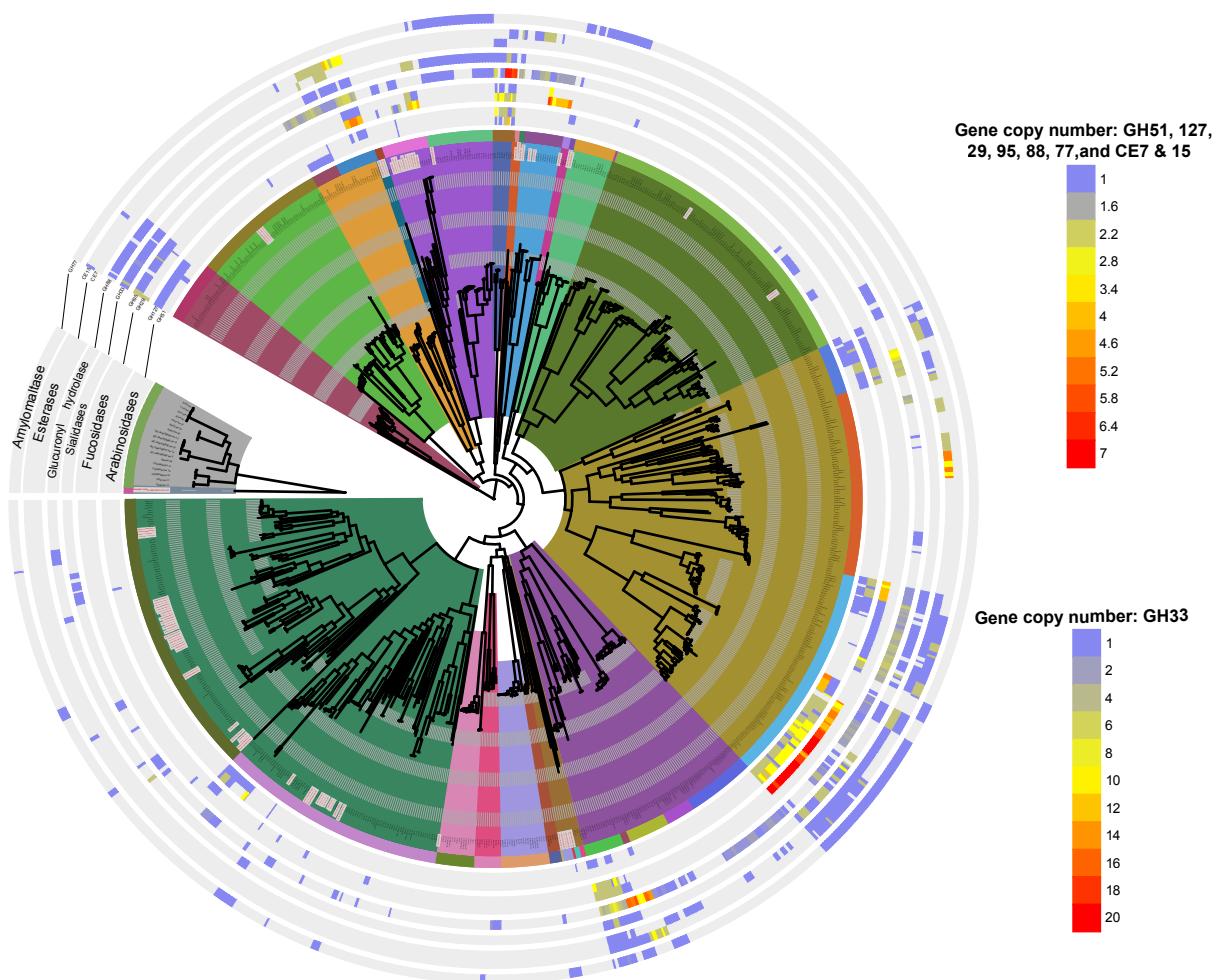


Figure 3

Phylum

Poribacteria	Chloroflexota
Gemmimatimonadota	Acidobacteriota
Latescibacterota	Bdellovibrionota
Marinisomatota	Myxococcota
Bacteroidota	Nitrospinota
Spirochaetota	Nitrospirota
Verrucomicrobiota	UBP10
Planctomycetota	Dadarbacteria
Deinococcota	Proteobacteria
Cyanobacteriota	Nanoarchaeota
Actinobacteriota	Crenarchaeota

Class

WGA-4E	Aanaerolineae
Gemmatimonades	Thermoanaerobaculia
unknown	bin61
UBA2968	Luteitaleia
Marinisomatia	unknown
Bacteroidia	Acidobacteriae
Rhodothermia	Bacteriovoracia
Spirochaeta	Bdellovibrionia
Chlamydia	Oligoflexia
Verrucomicrobiae	Polyangia
UBA8108	UBA9160
Planctomycetes	UBA8248
Deinococci	UBA8248
Vampirovibrionia	Nitrospira
Cyanobacteria	GR-WP33-30
Thermoleophilia	UBA1144
Actinobacteria	Gammaproteobacteria
Acidimicrobia	Alphaproteobacteria
UBA2235	Woesearchaeia
Dehalococcoidia	Nitrosphaeria

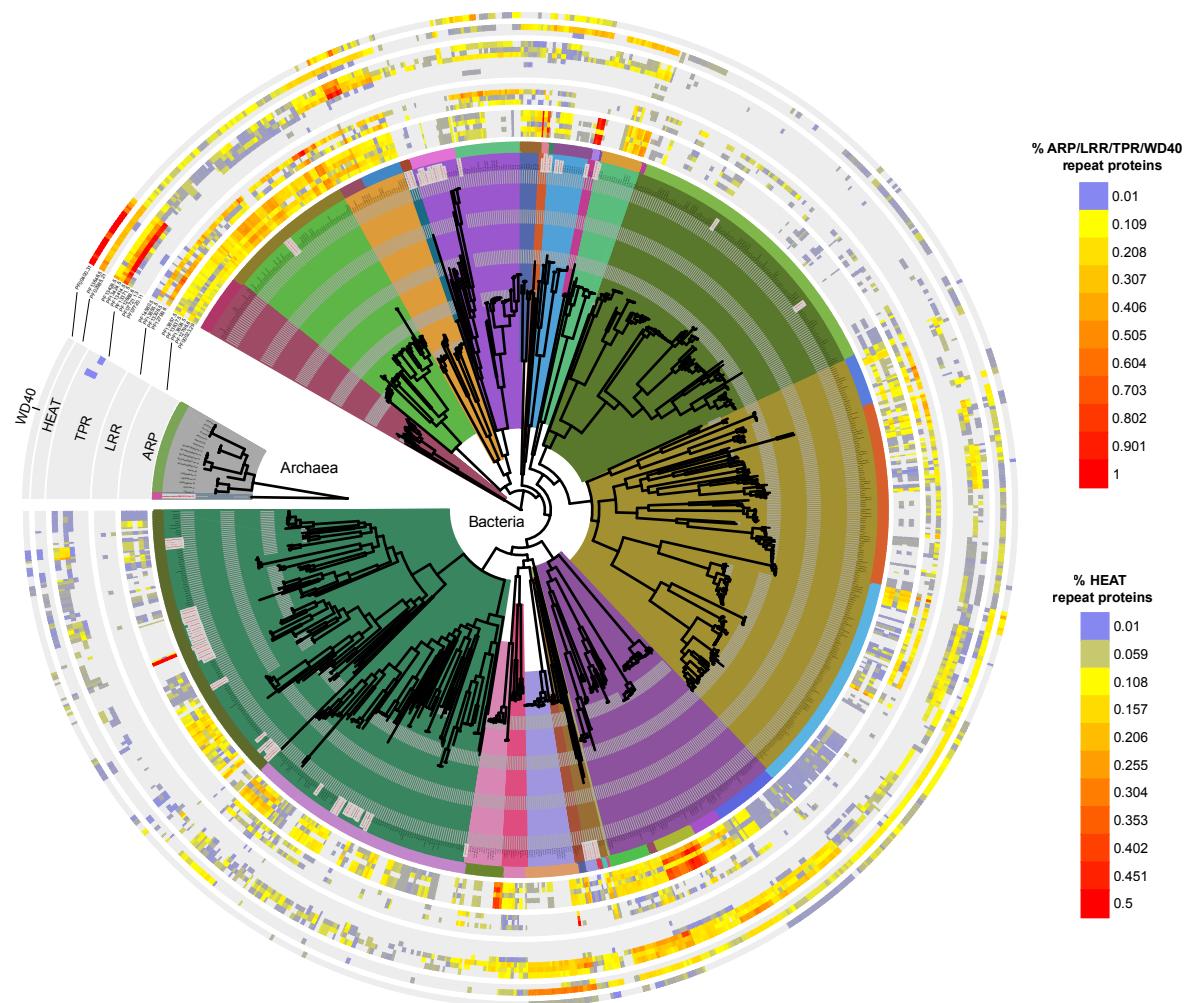
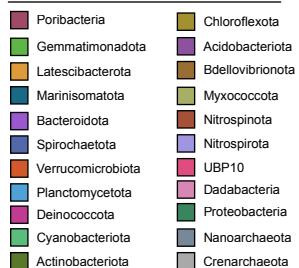


Figure 4

Phylum



Class

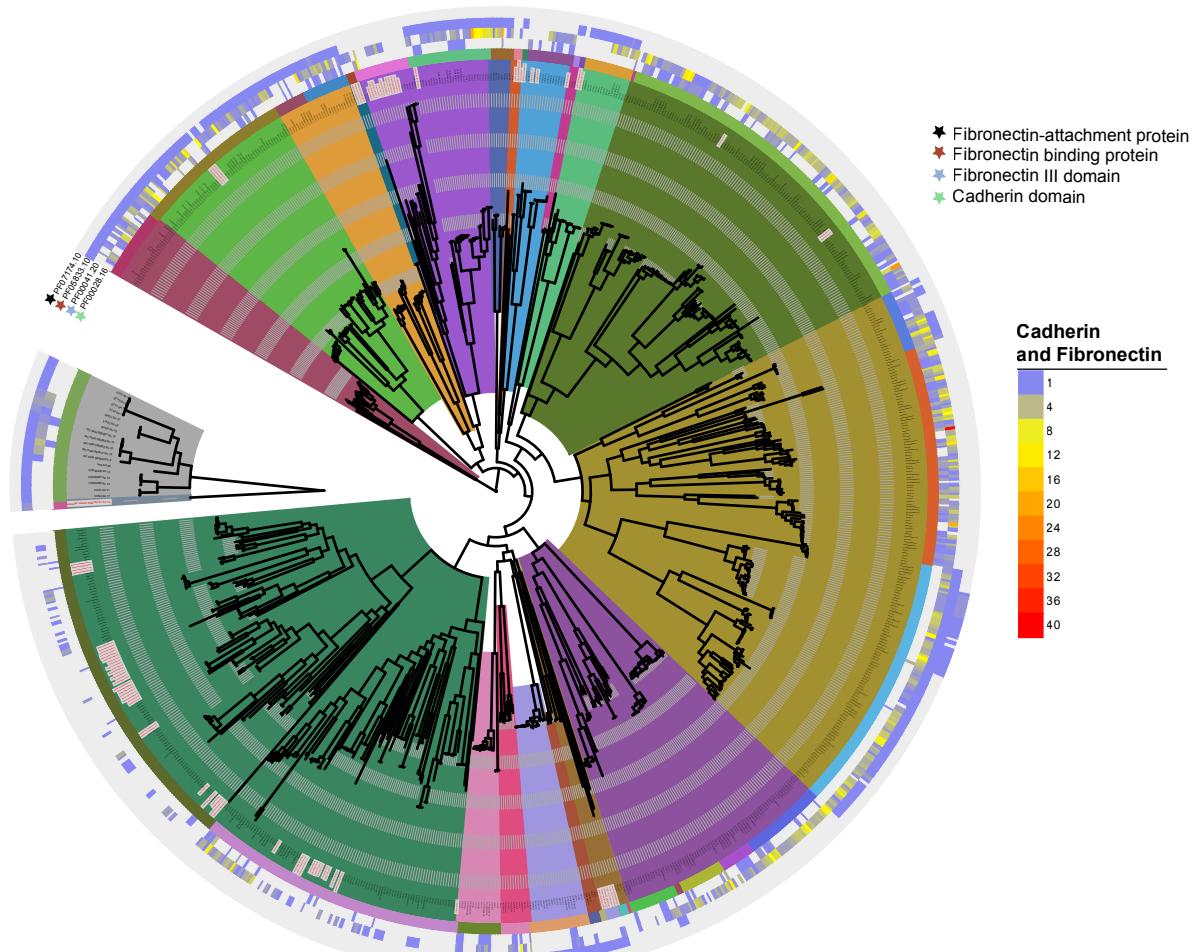
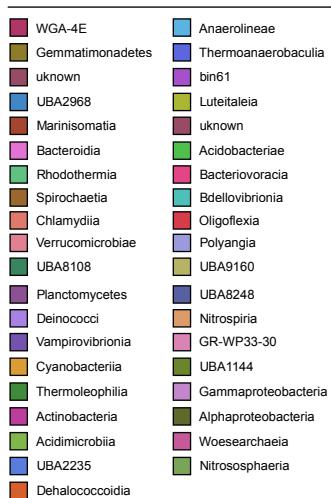


Figure 5

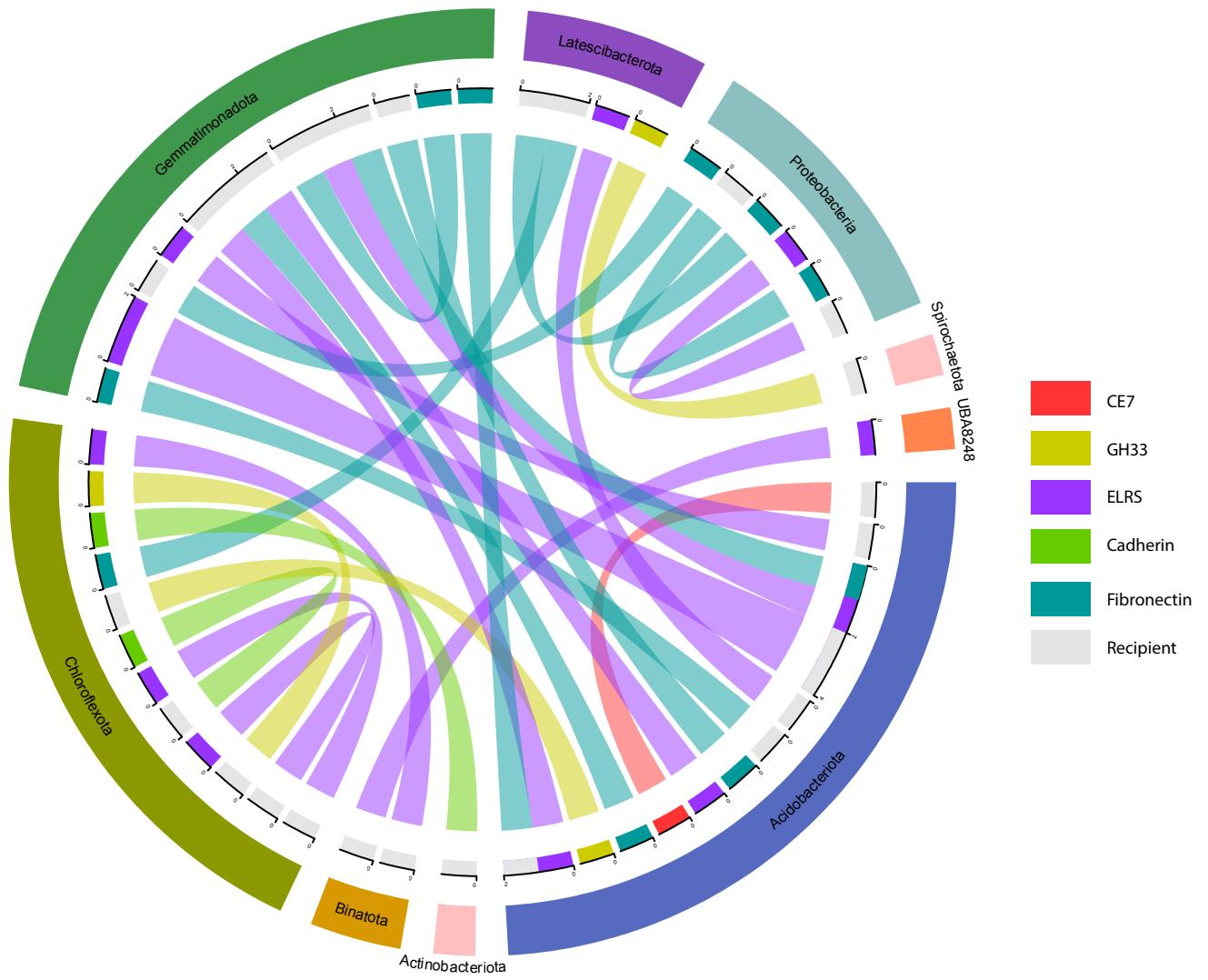


Figure 6

