Unravelling microalgal-bacterial interactions in aquatic ecosystems through 16S rRNA gene-based co-occurrence networks

B.L.D. Uthpala Pushpakumara[*1], Kshitij Tandon[1], Anusuya Willis[2], Heroen Verbruggen[*1]

[1]School of Biosciences, University of Melbourne, Victoria, 3010, Australia

[2]Australian National Algae Culture Collection, CSIRO, Tasmania 7000, Australia

*Corresponding authors: e-mail: bpushpakumar@student.unimelb.edu.au, heroen@unimelb.edu.au

Abstract

Interactions between microalgae and bacteria can directly influence the global biogeochemical cycles but the majority of such interactions remain unknown. 16S rRNA gene-based co-occurrence networks have potential to help identify microalgal-bacterial interactions. Here, we used data from 10 Earth microbiome projects to identify potential microalgal-bacterial associations in aquatic ecosystems. A high degree of clustering was observed in microalgal-bacterial modules, indicating densely connected neighbourhoods. *Proteobacteria* and *Bacteroidetes* predominantly co-occurred with microalgae and represented hubs of most modules. Our results also indicated that species-specificity may be a global characteristic of microalgal associated microbiomes. Several previously known associations were recovered from our network modules, validating that biologically meaningful results can be inferred using this approach. A range of previously unknown associations were recognised such as co-occurrences of *Bacillariophyta* with uncultured *Planctomycetes OM190* and *Deltaproteobacteria* order *NB1-j*. *Planctomycetes* and *Verrucomicrobia* were identified as key associates of microalgae due to their frequent co-occurrences with several microalgal taxa. Despite no clear taxonomic pattern, bacterial associates appeared functionally similar across

different environments. To summarise, we demonstrated the potential of 16S rRNA gene-based co-occurrence networks as a hypothesis-generating framework to guide more focused research on microalgal-bacterial associations.

Keywords: microalgae, bacteria, microalgal-bacterial associations, co-occurrence networks, 16S rRNA gene

Introduction

Microalgae (photosynthetic micro-eukaryotes) and bacteria are the most widespread and dominant planktonic organisms in aquatic ecosystems. As primary producers, microalgae form the foundations of the food web and directly influence global carbon and nutrient cycling, as well as the energy flow in aquatic ecosystems. A wide spectrum of associations between microalgae and bacteria have been reported, predominantly related to nutrient exchange, with important bottom-up effects on primary production. Associations include increased bioavailability of vitamins (e.g.B12[1–3]), metals (e.g. iron[4,5]) and growth promoting hormones[6] by bacteria in exchange for organic carbon from the algae. Apart from many beneficial interactions, some bacteria can have negative influences on algae by being algicidal and opportunistic pathogens[7–9]. A recent study on the bloom-forming *Phaeocystis* algae revealed that its microbiomes can take both symbiotic and opportunistic modes[10]. Thus, algal-bacterial interactions can range from mutualistic to parasitic, and a complex array of interactions can be hypothesised. For instance, previous work suggests that strong associations may exist between communities of microalgae and bacteria over a large geographical range in the open ocean[11]. However, the knowledge we have so far about specific microalgal-bacterial interactions represents only a small fraction of what potentially occurs in nature. Microalgae are highly diverse and innumerable, therefore gathering knowledge on their associations with equally ubiquitous and diverse bacteria is challenging.

High-throughput DNA sequencing such as 16S and 18S rRNA gene metabarcoding has been used to deliver insights into bacterial and eukaryotic community composition of diverse ecosystems. Microbial community abundance data can be used to identify associations between community members via microbial association networks[12]. Furthermore, combining community data with environmental data can reveal novel connections between microbial communities and their environment[13]. A network consists of nodes representing taxa and edges representing the associations between taxa. Microbial association networks targeting positive correlations are termed co-occurrence networks. A positive relationship is presumed when taxonomically relevant units co-occur or exhibit similarity in their compositions over multiple samples[14] and biologically meaningful groups or communities of a network can be identified with network clustering[15]. Also, network metrics such as node degree, closeness centrality for example, can be used for quantitative description of communities and identify ecologically important taxa[16,17]. Various network analyses have reported co-occurring microbial taxa in different environments, identified clusters of microbes representing metabolic consortia, defined keystone species, documented recurrent microbial modules and helped elucidate microbial dark matter in microbial communities[18–21]. Ecological processes governing community structure, such as niche filtering and habitat preference, are thought to be reflected in co-occurrence network modules[22]. Overall, networks provide insights into the structure of a community and taxon co-occurrences, offering perspectives to use them to generate hypotheses about interactions that can be investigated with more focused studies. Thus, environmental sequencing data combined with network analysis can be a convenient addition to the toolkit for deriving potential interactions between organisms across myriads of environments.

Studies targeting 16S rRNA marker genes are generally focused on the bacterial communities and often disregard the chloroplast sequences that are also amplified. The shared origin of the chloroplasts of all oxygenic photosynthetic micro-eukaryotes and cyanobacteria[23] enable the use of 16S rRNA marker gene to concurrently estimate relative abundances of microalgae and bacteria across samples, facilitating the construction of correlation networks to identify co-occurring taxa and speculate on their potential interactions. Robust co-occurrences can be detected through analysis of a sufficient number of samples, ideally covering temporal and spatial gradients, as these provide adequate variability in taxon abundances[12]. This can be coupled with stringent network building steps such as filtration of low prevalence organisms to reveal significant co-occurrences[24]. The ever-expanding sequence databases can be a great starting point to investigate such correlations. Analysis of the Tara Ocean dataset based on interaction networks has shown that abiotic factors are incomplete predictors of community structure[22]. Similar observations have been shown in phytoplankton blooms where environmental parameters were insignificant in influencing the community structure of plankton[25]. These results indicate that biological interactions are more influential in determining the community structure than environmental factors. Thus, taxon-taxon co-occurrence networks built on taxon compositions alone could capture potential interactions in nature. Our work is based on the premise that previously unknown associations between microalgae and bacteria may be discovered at large scales from existing 16S rRNA gene-based metabarcoding datasets using co-occurrence networks. In this study, we generated taxon-taxon co-occurrence networks using ten 16S rRNA gene metabarcoding datasets from the Earth Microbiome Project (EMP) that have sampled aquatic environments (both marine and freshwaters). These datasets were individually analysed to create local networks and network modules were used to identify significant co-occurrences of microalgae and bacteria.

99 **Methods**

100 **Data Acquisition**

101 Publicly available EMP datasets[26] were screened using the Qiita portal[27] and 10 studies (4
102 marine and 6 Freshwater environments) targeting aquatic environments were chosen (Figure
103 S1). Since the goal of this study was to individually analyse samples representing a particular
104 environment to generate local networks and reveal co-occurrences, studies targeting any
105 aquatic samples (water and sediment samples) were selected. To our knowledge, based on
106 study and sample information provided on qiita portal for the selected studies, no size
107 fractionation was carried out on the samples (water samples) provided to EMP. The data
108 analysed in this study have amplified the same 16S rRNA gene V4 hypervariable region using
109 515F- 'GTGCCAGCMGCCGCGGTAA' and 806R- 'GGACTACHVGGGTWTCTAAT'
110 primer pair[28,29] and sequenced on Illumina Hiseq2000. Detailed information on the selected
111 EMP projects can be found in Supplementary material (Table S1) and can also be further
112 explored using Qiita ids provided using the Qiita portal.

113

114 **Sequencing Data Workflow**

115 Raw demultiplexed reads were downloaded using the EBI accessions (provided in Qiita portal)
116 from the European Nucleotide Archive and analysed individually employing a single
117 bioinformatic pipeline built on Qiime2 version 2019.10[30]. Primer sequences attached to all
118 reads were trimmed using cutadapt[31]. Sequence denoising, chimera checking and dereplication
119 was performed in DADA2 [32] to correct sequencing errors and remove low quality bases. Reads
120 were truncated based on a median quality score of 30. The final outputs of DADA2, an
121 abundance table of Amplicon Sequence Variants (ASVs) and fasta sequences of the ASVs were
122 further processed as described hereafter. Taxonomy was assigned against Silva v132 [30] 16S
123 rRNA gene sequences trained with a Naive Bayes classifier[33]. Chloroplast sequences were

124 filtered using the Silva taxonomy file by identifying sequences assigned as "chloroplast".
125 These chloroplast sequences were then classified again using a Qiime2-compatible version of
126 PhytoRef[34] database accessed at[35]. All the algal taxonomic identities with PhytoRef had a
127 confidence score of 0.7 or higher (Table S2 and S3). It is important to note that, each algal
128 node in the networks (see below for details) represents a single ASV, likely representing one
129 algal species or strain in the environment.
130
131 The DADA2 abundance table was filtered to create a bacteria-only abundance table (by
132 removing mitochondrial, Archaeal and chloroplast ASVs) and a chloroplast-only abundance
133 table (by retaining only the ASV ids assigned as chloroplast). The bacteria-only table was then
134 collapsed at the genus level to reduce weak links in downstream network building[36] and was
135 merged with the chloroplast-only table to create the final abundance table. ASVs identified as
136 microalgae were not collapsed since the 16S rRNA gene marker is highly conserved in algae
137 and ASVs likely correspond to species or higher-level taxa[34]. This final abundance table was
138 filtered to remove low prevalence organisms present in less than 10% of samples to prevent
139 them from introducing artefacts in network inference.
140
141 **Correlation analysis and network construction**
142 We used a compositional data analysis tool, FastSpar[37], which is a rapid and parallelizable
143 implementation of the SparCC algorithm[38] to compute correlations. FastSpar quantifies the
144 correlation between all ASVs and assesses the statistical significance (p-values) of the inferred
145 correlations using a bootstrap procedure. The correlation and p-value matrices created by
146 FastSpar were used to create a network of significant correlations for each study separately.
147 Co-occurrence networks were created using the igraph package[39] in R studio version 1.4.1106.
148 Undirected weighted networks were created using statistically ($p < 0.05$) significant

correlations > 0.5 or higher (0.5 for freshwater and 0.6 for marine datasets). Coefficient cut-off of 0.5 was selected for freshwater datasets as a value above that resulted in fewer algal nodes in networks compared to 0.5. The idea behind the coefficient cut-off selection was to preserve as many links as possible with microalgae in resulting networks while choosing a considerably higher coefficient cut off to report stronger links. Networks were visualised in Cytoscape version 3.8[40].

Network Analysis and module detection

The Network Analyzer plugin[41] in Cytoscape was used to compute global network properties of each network to get an overview of node specific and edge-specific attributes. Networks were checked for the scale free nature by identifying the presence of highly connected nodes coexisting with nodes with fewer links[42] using the node attributes. In order to identify communities, networks were clustered using the Cytoscape plugin clusterMaker[43] using the MCL clustering algorithm[44] with an inflation value of 2.0. Co-occurring microalgae and bacteria were identified using resulting modules (modules with nodes > or = 4). These co-occurring taxa (identified beyond the taxonomic rank Phylum) were recorded for each environment and summarised in heatmaps using Pheatmap v1.0.12[45] in R.

Results and Discussion

16S rRNA gene-based co-occurrence networks can recover microalgal-bacterial associations

We generated 10 co-occurrences networks representing aquatic environments using publicly available 16S rRNA gene datasets from the EMP. We reduced noise and false positives by including ASVs present in at least 10% of samples, filtering statistically insignificant and weaker correlations. We also used network modules that are indicative of ecological

174 communities to identify potential interactions. In each co-occurrence network generated, we
175 observed highly connected nodes coexisting with nodes with fewer links. This indicates the
176 scale-free nature of the networks, a characteristic in real world networks[46,47]. The clustered
177 networks (modules) comprised nodes representing microalgae or bacteria and the edges
178 between them were instances of significant co-occurrences.
179
180 Analysis of the microalgal-bacterial modules identified 40 algal nodes co-occurring with at
181 least one of 76 bacterial nodes in marine environments and 112 microalgal nodes with at least
182 one of 311 bacterial nodes in freshwater environments (refer Table S2 and S3 for recovered
183 co-occurrences with the correlation values). These significant co-occurrences inferred in
184 marine and freshwater environments are summarised in Figures 1 and 2, respectively. We
185 identified algal nodes at different taxonomic levels although many could not be classified at
186 lower taxonomic levels such as genus or species. To have consistency in algal node
187 annotations, we have only used class-level classification in figures. Tables S2 and S3 provide
188 full taxonomic affiliations of the nodes where possible. Even though taxonomic assignments
189 were made at higher rank, our ASVs represent taxa at lower taxonomic levels. As the breadth
190 of species included in reference databases such as PhytoRef increases, it is likely that more
191 algal ASVs can be assigned to the species-level.
192
193 Most modules exhibited high clustering coefficients (> 0.5) indicating that the neighbourhood
194 of microalgal-bacterial communities are generally densely connected. High clustering
195 coefficient values in networks were previously suggested as indicative of cross-feeding
196 relationships and enriched degradation pathways[48]. We identified potential interactions of
197 bacteria with diverse phyla such as *Cryptophyta*, *Ochrophyta*, *Haptophyta*, *Chlorophyta,*
198 *Streptophyta* and *Euglenophta* representing 17 different taxonomic classes. Bacterial nodes in

199 the marine modules were predominantly represented by *Proteobacteria* and *Bacteroidetes*
200 followed by *Planctomycetes and Verrucomicrobia.* Similar to marine environments
201 *Proteobacteria and Bacteroidetes* dominated the freshwater modules while *Actinobacteria*
202 and *Verrucomicrobia* were the 2nd and 3rd most common bacterial taxa associated with
203 microalgae (Figure S2). We also identified hub nodes with the highest node degree (number
204 of edges connected to a node). Hub nodes represent highly connected nodes and are usually
205 considered as keystone species[49]. Hub nodes (considering the top ten hubs) in both marine
206 and freshwater modules were mostly represented by *Alphaproteobacteria*,
207 *Gammaproteobacteria* (p_*Proteobacteria*) and *Bacteroidia* (p_*Bacteroidetes*). Other than
208 these, members of *Planctomycetes* (mostly belonging to c_*Planctomycetacia*) were
209 commonly found among the top hub nodes in freshwater modules. High prevalence and their
210 characteristic associations with microalgae may explain the presence of *Proteobacteria* and
211 *Bacteroidetes* as hub nodes in most modules[50,51]. The role of *Planctomycetes* in global
212 nitrogen, carbon and sulphur cycles is gaining attention[52,53]. Thus, microalgal-*Planctomycetes*
213 associations may be playing a crucial role in the global environmental cycles.

214

215 Based on the summarised significant co-occurrences, benthic diatoms exhibited different, and
216 fewer, associations compared to planktonic diatoms (Figures 1 and 2). For example, bacterial
217 taxa such as unclassified *Gammaproteobacteria*, *Caldilineaceae* (*Chloroflexi*), *Trichococcus*
218 (*Firmicutes*), *Hoeflea* (*Proteobacteria*), *Pseudorhodobacter* (*Proteobacteria*) and
219 *Cocleimonas* (*Proteobacteria)* exhibited associations only with marine benthic diatoms and
220 not with marine planktonic diatoms.

221

222 We investigated whether this network approach has recovered any known associations
223 identified and confirmed earlier. We recognized known bacterial associates of microalgae

224 such as *Flavobacteriales*[54,55], *Rhizobiales*[56,57], *Cytophagales*[58,59] and *Rhodobacterales*[60].
225 Previously reported associations of microalgal genera with specific bacterial groups were also
226 recovered in the analysis, including that between *Bacteroidetes* and the cosmopolitan
227 *Prymnesiophyceae* genus *Phaeocystis*[61–63] (Table S2) and that of diatoms (*Bacillariophyta*)
228 with the genus *Flavobacterium* (Table S3)[64–66]. The ability to identify previously known
229 associations suggests that the 16S rRNA gene-based network approach used in this study can
230 yield biologically meaningful results.

231

232 We also compared co-occurrence networks built in our study with previously published
233 networks of these data. Previous analysis of lake Mendota samples (EBI accession:
234 ERP016591) by Kara et al[67] described characteristics of bacterioplankton co-occurrence
235 networks across three seasons. The network properties of these bacterioplankton networks such
236 as clustering coefficient (0.167 - 0.256) and the characteristic path length (3.27 - 3.88) were
237 comparatively lower than those of our microalgal-bacterial network (clustering coefficient >
238 0.5, characteristic path length = 3.965). Similar to the observation by Gilbert et al[68], the co-
239 occurrence network built on Western English Channel samples (EBI accession: ERP016541)
240 featured many *Alphaproteobacteria and Rhodobacterales* nodes.

241

242 **Both Intrinsic and Extrinsic factors may influence microalgal-bacterial associations**
243 No clear taxonomic pattern was observed among the co-occurring microalgae and bacteria. In
244 accordance with previous work on phytoplankton-bacteria co-occurrences[67], taxonomically
245 diverse bacteria co-occurred with different microalgal groups. Moreover, except for a few,
246 there was no recurrence of bacterial genera (the lowest taxonomic identity of bacterial nodes
247 in a module) across modules generated in each project. Since each project represents a
248 specific environment, these observations indicate that each environment harbours its specific

249 co-occurrence relationships. Project specificity in inferred co-occurrences can clearly be seen
250 in the generated heatmaps (Figures 1 and 2). Interestingly, a recent study[68] has provided
251 evidence for biogeographic differentiation of algal microbiomes, showing ecological
252 boundaries driven by differences in environmental conditions altering the spatial scaling of
253 the algal microbiomes. Another plausible explanation for such observations is the species
254 specificity of algal microbiomes. Although some microalgal nodes could not be
255 taxonomically identified at species level, each microalgal node in a module is an ASV which
256 likely represents a single microalgal species or a strain. Species specificity of algal
257 microbiomes has been predominantly shown using algal cultures[50]. In addition to supporting
258 previous observations, our results show that species-specificity may be a global characteristic
259 of microalgal microbiomes.

260

261 To understand if there is any factor driving general patterns of interactions, some insights may
262 be gleaned from known functions of the co-occurring bacteria. Our observation was that despite
263 their taxonomic differences, bacterial associates in modules often shared functional
264 similarities. For instance, microalgae co-occurred with bacterial groups equipped with specific
265 metabolic functional potentials such as algal polysaccharide degradation and provision of
266 vitamins. A few examples of these are, co-occurrences with taxonomically diverse members
267 of *Bacteroidetes*[63] and *Verrucomicrobia*[69] with polysaccharide degradation ability and
268 *Rhizobiales*, *Rhodobacterales* [70] and *SAR116* [71–73] with Vitamin B12 synthesis. Therefore, our
269 results stemming from observations across multiple datasets suggest that, irrespective of the
270 environment, microalgae are associated with key functional types of bacteria. As most
271 microalgal-bacterial functional interactions remain unknown, it is difficult to identify global
272 trends in key functional types of bacteria associated with microalgal-bacterial communities.

This urges the need for more functional studies to improve our understanding of microalgal-bacterial communities across the globe.

Emerging microalgal-bacterial associations that can guide functional studies

An advantage of using a network approach is the ability to analyse large scale datasets to unravel previously unknown associations. We believe that some of the inferred co-occurrences in this study may help guide focused research to shed light on the functional nature of interactions. Therefore, we further explored these co-occurrences which were prominent due to their frequent observations and importance in modules based on topological properties.

For instance, uncultured *Deltaproteobacteria* order *NB1-j* was identified as consistently co-occurring with *Bacillariophyta* in marine environments. In one environment where the *NB1-j-Bacillariophyta* link was observed (Surface water samples from Western English Channel, EBI accession: ERP016541), *Bacillariophyta* was the only microalgal node directly interacting with *NB1-j*. In another environment (Seawater metagenome samples from Catlin Arctic survey (2010 expedition), EBI accession: ERP020022), *NB1-j* was identified as the hub node with the highest node degree (43), in a module consisting of 77 nodes and 817 edges. This *NB1-j* node also had the highest closeness centrality (0.69). Closeness centrality measures how close a node is to another node and helps to identify centrally positioned taxa in the network[24]. Other than this *NB1-j* node, there were 3 more *NB1-j* nodes in this community. All together these 4 *NB1-j* nodes were directly interacting (edges) with 57 nodes (out of 73 other nodes) of which 25 were representing microalgal nodes (Figure 3). Out of the 25 microalgal nodes, 13 represented the taxonomic class *Bacillariophyta*. Some of these *Bacillariophyta* were taxonomically identified at genus level as *Chaetoceros* and *Fragilariopsis*. Interestingly, except for one, all these Bacillariophyta nodes represented top hub nodes of the community with node degrees > 26. Most connected taxa are believed to have ecological relevance to the community as their removal causes the highest impact on many associations[74].

299

300 Functional roles of *NB1-j* in marine ecosystems are largely unknown. It is believed to be
301 involved in hydrocarbon degradation[75] and has been reported mostly from marine sediments
302 [70,76], mud volcanoes[75], sponges[77] and cyanobacterial mats[78]. To the best of our knowledge, this
303 *Deltaproteobacteria* group is not commonly reported in algal microbiomes. However, a recent
304 study found *NB1-j* ASVs associated with the coral skeleton algal symbiont *Ostreobium*[79]. A
305 predictive metagenomic approach based on sponge samples originating from reef sites in West
306 Java (Indonesia) suggest that *NB1-j* may be involved in nitrogen metabolism[77]. It was found
307 that *NB1-j* was responsible for the elevated predictive gene count corresponding to N-cycling
308 genes such as those encoding nitric oxide reductase (*norB)*, nitrogenase (*nifD*) and
309 hydroxylamine reductase. The available information indicates that the nitrogen cycling
310 capacity of *NB1-j* may underly its association with algae, potentially facilitating the nitrogen
311 needs of the algae while benefiting from algal organic carbon. However, a repertoire of
312 interactions, as often expected from a keystone species, can be hypothesised between *NB1-j*
313 and microalgae in a community.

314

315 *Bacillariophyta* (*Fragilariopsis, Chaetoceros* and many other diatoms unidentified at lower
316 taxonomic levels) also showed frequent co-occurrences with an uncultured clade of
317 *Planctomycetes*, *OM190* (SILVA taxonomy) in both freshwater and marine environments. In
318 the Catlin Arctic survey samples (ERP02022), *OM190* demonstrated frequent associations with
319 *Bacillariophyta*. Three *OM190* nodes were observed with high node degrees (29, 30 and 39)
320 and these were directly connected to 59 nodes out of a total of 74 other nodes in the community.
321 From the total of 35 microalgal nodes in this community, *OM190* were directly connected to
322 27 nodes out of which 15 nodes were represented by *Bacillariophyta* (Figure S3A). Similarly
323 in a freshwater environment (Lake Superior, Michigan, EBI accession: ERP016492), *OM190*

324 was identified as the 6th most connected node with a node degree of 89 and was directly
325 connected to 32 microalgal nodes including *Bacillariophyta* (Figure S3B). As mentioned
326 previously, highly connected nodes act as hub nodes in the community. Multiple *OM190* acting
327 as top hub nodes within their habitats indicate that they have high co-occurrences with both
328 microalgae and bacteria.
329
330 *OM190* shows deep branching within the *Planctomycetes* group and is usually found in
331 different environments such as soil and seawater. Members of this clade are usually considered
332 to be associated with macroalgae[80], such as red[81] and brown algae[82]. Since *OM190* is yet
333 uncultured, information on its metabolism is scarce. A metagenome-assembled-genome
334 (MAG) for *OM190* (likely *OM190*) with a rich diversity of secondary metabolite potential has
335 been reported[83]. The production of secondary metabolites by *OM190*, including antimicrobial
336 compounds may be one of the underlying reasons for their association with algae as they could
337 protect the alga from undesired microbes. In a recent study[84], it was shown that diatoms
338 produce fucose-containing sulfated polysaccharides (FCSP) which can be hypothesised as an
339 energy source for *OM190*. Many *Planctomycetes* have abundant sulfatases, but these were not
340 confirmed in the existing *OM190* MAG. In addition to their association with algae, *OM190*
341 and *NB1-j* seem to have a close relationship with one another. They were recently reported as
342 abundant co-occurring taxa in Beaufort Sea surface sediments[85]. The reasons for their direct
343 associations are not known and, along with their respective relationships with algae, this direct
344 interaction is an interesting topic for future investigation.
345
346 More generally speaking, *Planctomycetes* often associate with macroalgae and favour a biofilm
347 lifestyle[82]. Protein clusters that may be involved in *Planctomycetes* symbiosis or biofilm
348 maintenance have been reported[53], and production of sulfated polysaccharides by the alga that

349 serve as the substrate for the abundant sulfatases produced by the *Planctomycetes* contributes
350 to the reasons for successful associations between them[53,80]. Although successful associations
351 are known between macroalgae and *Planctomycetes*, interactions with microalgae are largely
352 unknown. Apart from their associations with diatoms, our network analysis identified taxa
353 representing *Planctomycetes* (c_*Phycisphaerae*, c_*Planctomycetacia*) co-occurring with a
354 range of microalgal groups. In a previous study, *Planctomycetes* closely related to *Pirellula*
355 were identified as one of the dominant lineages associated with diatom blooms[86]. Results of
356 our network analysis demonstrate that associations between *Planctomycetes* and microalgae
357 may be as common as those they maintain with macroalgae.
358
359 *Verrucomicrobia* were also consistently associated with an array of microalgae in fresh and
360 marine environments indicating that they may be common associates of microalgae. Here in
361 our study, all the Verrucomicrobial members co-occurring with microalgae were represented
362 by the taxonomic class *Verrucomicrobiae*. Among the genus-specific associations revealed in
363 our network analysis is that of the Verrucomicrobial genus *Lentimonas* and microalgal genera
364 *Pyramimonas* (c_*Prasinophyceae*), *Phaeocystis*, *Tisochrysis*, *Haptolina*, and
365 *Chrysochromulina* (c_*Prymnesiophyceae*). A recent study showed hundreds of *Lentimonas*
366 enzymes able to digest brown algal fucoidan[87]. Although fucose-containing sulphated
367 polysaccharides were regarded as a macroalgal polysaccharide, microalgae such as diatoms
368 have the potential to produce them[84]. We speculate that other microalgae might also produce
369 this sulfated polysaccharide, or that members of *Lentimonas* may have a broader palate than
370 just FCSP. Although widely distributed, the functional roles of *Verrucomicrobia* in aquatic
371 ecosystems are not well understood due to the lack of cultured strains[88]. Some members of
372 *Verrucomicrobia* have been shown to consume algal extracellular polymeric substances[69,89,90].
373 However, other factors contributing to their associations are not well understood.

Conclusion

In summary, our study illustrates the promise of using 16S rRNA gene-based co-occurrence networks as a hypothesis-generating framework to guide focused research and speculate on the functional nature of potential interactions. By studying multiple environments based on public datasets, we provided an overview of microalgal-bacterial communities in aquatic ecosystems from a network perspective. This identified a range of associations including previously unknown links that can set the stage for more focused research in the future.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

No data was generated in this study and all the data analysed are publicly available. The EMP datasets analysed are available in European Nucleotide Archive and EMP Qiita portal under the Accession numbers and Qiita ids, respectively, provided in Supplementary material TableS1. FastSpar script used for the correlation analysis, R script used in generating networks and the networks files generated in Cytoscape can be found at https://melbourne.figshare.com/projects/Unravelling_microalgal-bacterial_interactions_in_aquatic_ecosystems_through_16S_rRNA_gene-based_co-occurrence_networks/140071

397 **Competing interests**

398 The authors declare that they have no competing interests.

399 **Funding**

400 We acknowledge the funding from the Melbourne Research Scholarship and the

401 ResearchPlus Postgraduate Top-Up Scholarship Grants Program and National Research

402 Collections Australia, CSIRO.

403 **Authors' contribution**

404 UP, HV and AW formulated the project. UP analysed the data and wrote the manuscript with

405 inputs from all the authors. HV, AW and KT contributed ideas and helped with analysis. All

406 authors read and approved the final version of the manuscript.

407 **Acknowledgements**

408 Not applicable

409 **References**

410 1. Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J. & Smith, A. G. Algae

411 acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* (2005)

412 doi:10.1038/nature04056.

413 2. Kazamia, E. *et al.* Mutualistic interactions between vitamin B12-dependent algae and

414 heterotrophic bacteria exhibit regulation. *Environ. Microbiol.* (2012)

415 doi:10.1111/j.1462-2920.2012.02733.x.

416 3. Bunbury, F. *et al.* Exploring the onset of B12-based mutualisms using a recently

417 evolved Chlamydomonas auxotroph and B12-producing bacteria. *Environ. Microbiol.*

418 **24**, 3134–3147 (2022).

419 4. Butler, A. Acquisition and utilization of transition metal ions by marine organisms.

420 *Science* (1998) doi:10.1126/science.281.5374.207.

5. Amin, S. A. *et al.* Photolysis of iron-siderophore chelates promotes bacterial-algal mutualism. *Proc. Natl. Acad. Sci. U. S. A.* (2009) doi:10.1073/pnas.0905512106.

6. Amin, S. A. *et al.* Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* (2015) doi:10.1038/nature14488.

7. Mayali, X. & Azam, F. Algicidal bacteria in the sea and their impact on algal blooms. in *Journal of Eukaryotic Microbiology* (2004). doi:10.1111/j.1550-7408.2004.tb00538.x.

8. Bagwell, C. E. *et al.* Discovery of bioactive metabolites in biofuel microalgae that offer protection against predatory bacteria. *Front. Microbiol.* (2016) doi:10.3389/fmicb.2016.00516.

9. Hoeger, A. L., Jehmlich, N., Kipping, L., Griehl, C. & Noll, M. Associated bacterial microbiome responds opportunistic once algal host Scenedesmus vacuolatus is attacked by endoparasite Amoeboaphelidium protococcarum. *Sci. Rep.* 1–11 (2022) doi:10.1038/s41598-022-17114-1.

10. Mars Brisbin, M., Mitarai, S., Saito, M. A. & Alexander, H. Microbiomes of bloom-forming Phaeocystis algae are stable and consistently recruited, with both symbiotic and opportunistic modes. *ISME J. 2022* 1–10 (2022) doi:10.1038/s41396-022-01263-2.

11. Milici, M. *et al.* Co-occurrence analysis of microbial taxa in the Atlantic ocean reveals high connectivity in the free-living bacterioplankton. *Front. Microbiol.* **7**, 1–20 (2016).

12. Barberán, A., Bates, S. T., Casamayor, E. O. & Fierer, N. Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* **6**, 343–351 (2012).

13. Tucker, A. E. & Brown, S. P. Sampling a gradient of red snow algae bloom density reveals novel connections between microbial communities and environmental features. *Sci. Rep.* **12**, 1–15 (2022).

14. Faust, K. & Raes, J. Microbial interactions: From networks to models. *Nat. Rev. Microbiol.* **10**, 538–550 (2012).

15. Chun, S. J. *et al.* Network analysis reveals succession of Microcystis genotypes accompanying distinctive microbial modules with recurrent patterns. *Water Res.* (2020) doi:10.1016/j.watres.2019.115326.

16. Huang, S. Back to the biology in systems biology: What can we learn from biomolecular networks? *Briefings Funct. Genomics Proteomics* (2004) doi:10.1093/bfgp/2.4.279.

17. Ma'ayan, A. Introduction to network analysis in systems biology. *Science Signaling* (2011) doi:10.1126/scisignal.2001965.

18. Williams, R. J., Howe, A. & Hofmockel, K. S. Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. *Front. Microbiol.* **5**, 1–10 (2014).

19. Fuhrman, J. A., Cram, J. A. & Needham, D. M. Marine microbial community dynamics and their ecological interpretation. *Nature Reviews Microbiology* (2015) doi:10.1038/nrmicro3417.

20. Chafee, M. *et al.* Recurrent patterns of microdiversity in a temperate coastal marine environment. *ISME J.* **12**, 237–252 (2018).

21. Zamkovaya, T., Foster, J. S., de Crécy-Lagard, V. & Conesa, A. A network approach to elucidate and prioritize microbial dark matter in microbial communities. *ISME J.* **15**, 228–244 (2021).

467 22. Lima-Mendez, G. *et al.* Determinants of community structure in the global plankton
468 interactome. *Science (80-.).* (2015) doi:10.1126/science.1262073.

469 23. Bennke, C. M. *et al.* The distribution of phytoplankton in the Baltic Sea assessed by a
470 prokaryotic 16S rRNA gene primer system. *J. Plankton Res.* **40**, 244–254 (2018).

471 24. Berry, D. & Widder, S. Deciphering microbial interactions and detecting keystone
472 species with co-occurrence networks. *Front. Microbiol.* (2014)
473 doi:10.3389/fmicb.2014.00219.

474 25. Needham, D. M. & Fuhrman, J. A. Pronounced daily succession of phytoplankton,
475 archaea and bacteria following a spring bloom. *Nat. Microbiol.* (2016)
476 doi:10.1038/nmicrobiol.2016.5.

477 26. Thompson, L. R. *et al.* A communal catalogue reveals Earth's multiscale microbial
478 diversity. *Nature* (2017) doi:10.1038/nature24621.

479 27. Gonzalez, A. *et al.* Qiita: rapid, web-enabled microbiome meta-analysis. *Nat. Methods*
480 (2018) doi:10.1038/s41592-018-0141-9.

481 28. Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of
482 sequences per sample. *Proc. Natl. Acad. Sci. U. S. A.* (2011)
483 doi:10.1073/pnas.1000080107.

484 29. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the
485 Illumina HiSeq and MiSeq platforms. *ISME J.* (2012) doi:10.1038/ismej.2012.8.

486 30. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data
487 science using QIIME 2. *Nature Biotechnology* (2019) doi:10.1038/s41587-019-0209-
488 9.

489 31. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing

490 reads. *EMBnet.journal* (2011) doi:10.14806/ej.17.1.200.

491 32. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina
492 amplicon data. *Nat. Methods* (2016) doi:10.1038/nmeth.3869.

493 33. Bokulich, N. A. *et al.* Optimizing taxonomic classification of marker-gene amplicon
494 sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* (2018)
495 doi:10.1186/s40168-018-0470-z.

496 34. Decelle, J. *et al.* PhytoREF: A reference database of the plastidial 16S rRNA gene of
497 photosynthetic eukaryotes with curated taxonomy. *Mol. Ecol. Resour.* **15**, 1435–1445
498 (2015).

499 35. Major oceanic 16S/18S databases in qiime2 format. https://github.com/ndu-
500 invitae/Oceanic_database/tree/master/PhytoRef.

501 36. Hemprich-Bennett, D. R., Oliveira, H. F. M., Le Comber, S. C., Rossiter, S. J. &
502 Clare, E. L. Assessing the impact of taxon resolution on network structure. *Ecology*
503 (2021) doi:10.1002/ecy.3256.

504 37. Watts, S. C., Ritchie, S. C., Inouye, M. & Holt, K. E. FastSpar: Rapid and scalable
505 correlation estimation for compositional data. *Bioinformatics* (2019)
506 doi:10.1093/bioinformatics/bty734.

507 38. Friedman, J. & Alm, E. J. Inferring Correlation Networks from Genomic Survey Data.
508 *PLoS Comput. Biol.* (2012) doi:10.1371/journal.pcbi.1002687.

509 39. Csardi, G., & Nepusz, T. The igraph software package for complex network research.
510 InterJournal, Complex Systems. *igraph Softw. Packag.* (2006).

511 40. Shannon, P. *et al.* Cytoscape: A software Environment for integrated models of
512 biomolecular interaction networks. *Genome Res.* (2003) doi:10.1101/gr.1239303.

513 41. Assenov, Y., Ramírez, F., Schelhorn, S. E. S. E., Lengauer, T. & Albrecht, M.
514 Computing topological parameters of biological networks. *Bioinformatics* (2008)
515 doi:10.1093/bioinformatics/btm554.

516 42. Barabási, A. L. Scale-free networks: A decade and beyond. *Science* (2009)
517 doi:10.1126/science.1173299.

518 43. Morris, J. H. *et al.* ClusterMaker: A multi-algorithm clustering plugin for Cytoscape.
519 *BMC Bioinformatics* (2011) doi:10.1186/1471-2105-12-436.

520 44. Van Dongen, S. & Abreu-Goodger, C. Using MCL to extract clusters from networks.
521 *Methods Mol. Biol.* (2012) doi:10.1007/978-1-61779-361-5_15.

522 45. Raivo, K. Pheatmap: pretty heatmaps. *R Pacakage Version* (2012).

523 46. Albert, R., Jeong, H. & Barabási, A. L. Diameter of the world-wide web. *Nature*
524 (1999) doi:10.1038/43601.

525 47. Eisenberg, E. & Levanon, E. Y. Preferential attachment in the protein network
526 evolution. *Phys. Rev. Lett.* (2003) doi:10.1103/PhysRevLett.91.138701.

527 48. Ma, B. *et al.* Earth microbial co-occurrence network reveals interconnection pattern
528 across microbiomes. *Microbiome* **8**, 82 (2020).

529 49. Wang, H. *et al.* Combined use of network inference tools identifies ecologically
530 meaningful bacterial associations in a paddy soil. *Soil Biol. Biochem.* **105**, 227–235
531 (2017).

532 50. Goecke, F., Thiel, V., Wiese, J., Labes, A. & Imhoff, J. F. Algae as an important
533 environment for bacteria - Phylogenetic relationships among new bacterial species
534 isolated from algae. *Phycologia* (2013) doi:10.2216/12-24.1.

535 51. Krohn-Molt, I. *et al.* Metagenome survey of a multispecies and alga-associated biofilm
536 revealed key elements of bacterial-algal interactions in photobioreactors. *Appl.*
537 *Environ. Microbiol.* (2013) doi:10.1128/AEM.01641-13.

538 52. Woebken, D. *et al.* Fosmids of novel marine Planctomycetes from the Namibian and
539 Oregon coast upwelling systems and their cross-comparison with planctomycete
540 genomes. *ISME J.* (2007) doi:10.1038/ismej.2007.63.

541 53. Faria, M. *et al.* Planctomycetes attached to algal surfaces: Insight into their genomes.
542 *Genomics* (2018) doi:10.1016/j.ygeno.2017.10.007.

543 54. Barbeyron, T., L'Haridon, S., Corre, E., Kloareg, B. & Potin, P. Zobellia
544 galactanovorans gen. nov., sp. nov., a marine species of Flavobacteriaceae isolated
545 from a red alga, and classification of [Cytophaga] uliginosa (ZoBell and Upham 1944)
546 Reichenbach 1989 as Zobellia uliginosa gen. nov., comb. nov. *Int. J. Syst. Evol.*
547 *Microbiol.* (2001) doi:10.1099/00207713-51-3-985.

548 55. Nedashkovskaya, O. I. *et al.* Mesonia algae gen. nov., sp. nov., a novel marine
549 bacterium of the family Flavobacteriaceae isolated from the green alga Acrosiphonia
550 sonderi (Kütz) Kornm. *Int. J. Syst. Evol. Microbiol.* (2003) doi:10.1099/ijs.0.02626-0.

551 56. Kim, B. H., Ramanan, R., Cho, D. H., Oh, H. M. & Kim, H. S. Role of Rhizobium, a
552 plant growth promoting bacterium, in enhancing algal biomass through mutualistic
553 interaction. *Biomass and Bioenergy* (2014) doi:10.1016/j.biombioe.2014.07.015.

554 57. Rivas, M. O., Vargas, P. & Riquelme, C. E. Interactions of Botryococcus braunii
555 Cultures with Bacterial Biofilms. *Microb. Ecol.* (2010) doi:10.1007/s00248-010-9686-
556 6.

557 58. Cole, J. J. Interactions between bacteria and algae in aquatic ecosystems. *Annu. Rev.*

558 *Ecol. Syst. Vol. 13* (1982) doi:10.1146/annurev.es.13.110182.001451.

559 59. Fulbright, S. P. *et al.* Bacterial community changes in an industrial algae production
560 system. *Algal Res.* (2018) doi:10.1016/j.algal.2017.09.010.

561 60. Zhou, J. *et al.* Microbial community structure and associations during a marine
562 dinoflagellate bloom. *Front. Microbiol.* (2018) doi:10.3389/fmicb.2018.01201.

563 61. Simon, M., Glöckner, F. O. & Amann, R. Different community structure and
564 temperature optima of heterotrophic picoplankton in various regions of the Southern
565 Ocean. *Aquat. Microb. Ecol.* (1999) doi:10.3354/ame018275.

566 62. Brussaard, C. P. D., Mari, X., Van Bleijswijk, J. D. L. & Veldhuis, M. J. W. A
567 mesocosm study of Phaeocystis globosa (Prymnesiophyceae) population dynamics: II.
568 Significance for the microbial community. *Harmful Algae* (2005)
569 doi:10.1016/j.hal.2004.12.012.

570 63. Janse, I., Zwart, G., Van der Maarel, M. J. E. C. & Gottschal, J. C. Composition of the
571 bacterial community degrading Phaeocystis mucopolysaccharides in enrichment
572 cultures. *Aquat. Microb. Ecol.* (2000) doi:10.3354/ame022119.

573 64. Grossart, H. P., Levold, F., Allgaier, M., Simon, M. & Brinkhoff, T. Marine diatom
574 species harbour distinct bacterial communities. *Environ. Microbiol.* (2005)
575 doi:10.1111/j.1462-2920.2005.00759.x.

576 65. Sapp, M. *et al.* Species-specific bacterial communities in the phycosphere of
577 microalgae? *Microb. Ecol.* (2007) doi:10.1007/s00248-006-9162-5.

578 66. Sapp, M., Wichels, A. & Gerdts, G. Impacts of cultivation of marine diatoms on the
579 associated bacterial community. *Appl. Environ. Microbiol.* (2007)
580 doi:10.1128/AEM.02274-06.

581 67. Milici, M. *et al.* Bacterioplankton biogeography of the Atlantic ocean: A case study of
582 the distance-decay relationship. *Front. Microbiol.* **7**, 1–15 (2016).

583 68. Martin, K. *et al.* The biogeographic differentiation of algal microbiomes in the upper
584 ocean from pole to pole. *Nat. Commun.* (2021) doi:10.1038/s41467-021-25646-9.

585 69. Martinez-Garcia, M. *et al.* Capturing single cell genomes of active polysaccharide
586 degraders: An unexpected contribution of verrucomicrobia. *PLoS One* (2012)
587 doi:10.1371/journal.pone.0035314.

588 70. Dell'Anno, F. *et al.* Highly Contaminated Marine Sediments Can Host Rare Bacterial
589 Taxa Potentially Useful for Bioremediation. *Front. Microbiol.* (2021)
590 doi:10.3389/fmicb.2021.584850.

591 71. Newton, R. J. *et al.* Genome characteristics of a generalist marine bacterial lineage.
592 *ISME J.* (2010) doi:10.1038/ismej.2009.150.

593 72. Sañudo-Wilhelmy, S. A., Gómez-Consarnau, L., Suffridge, C. & Webb, E. A. The role
594 of B vitamins in marine biogeochemistry. *Ann. Rev. Mar. Sci.* (2014)
595 doi:10.1146/annurev-marine-120710-100912.

596 73. Durham, B. P. *et al.* Cryptic carbon and sulfur cycling between surface ocean
597 plankton. *Proc. Natl. Acad. Sci. U. S. A.* (2015) doi:10.1073/pnas.1413137112.

598 74. Proulx, S. R., Promislow, D. E. L. & Phillips, P. C. Network thinking in ecology and
599 evolution. *Trends in Ecology and Evolution* (2005) doi:10.1016/j.tree.2005.04.004.

600 75. Coelho, F. J. R. C. *et al.* Integrated analysis of bacterial and microeukaryotic
601 communities from differentially active mud volcanoes in the Gulf of Cadiz. *Sci. Rep.*
602 **6**, 1–10 (2016).

603 76. Queiroz, L. L. *et al.* Bacterial diversity in deep-sea sediments under influence of

604 asphalt seep at the São Paulo Plateau. *Antonie van Leeuwenhoek, Int. J. Gen. Mol.*
605 *Microbiol.* (2020) doi:10.1007/s10482-020-01384-8.

606 77. de Voogd, N. J., Cleary, D. F. R., Polónia, A. R. M. & Gomes, N. C. M. Bacterial
607 community composition and predicted functional ecology of sponges, sediment and
608 seawater from the thousand islands reef complex, West Java, Indonesia. *FEMS*
609 *Microbiol. Ecol.* (2015) doi:10.1093/femsec/fiv019.

610 78. Vigneron, A. *et al.* Multiple strategies for light-harvesting, photoprotection, and
611 carbon flow in high latitude microbial mats. *Front. Microbiol.* (2018)
612 doi:10.3389/fmicb.2018.02881.

613 79. Pushpakumara, B. L. D. U., Tandon, K., Willis, A., Verbruggen, H. The bacterial
614 microbiome of the coral skeleton algal symbiont Ostreobium shows preferential
615 associations and signatures of phylosymbiosis. *bioRxiv* (2022)
616 10.1101/2022.12.13.520198.

617 80. Lage, O. M. & Bondoso, J. Planctomycetes diversity associated with macroalgae.
618 *FEMS Microbiol. Ecol.* (2011) doi:10.1111/j.1574-6941.2011.01168.x.

619 81. Longford, S. R. *et al.* Comparisons of diversity of bacterial communities associated
620 with three sessile marine eukaryotes. *Aquat. Microb. Ecol.* (2007)
621 doi:10.3354/ame048217.

622 82. Bengtsson, M. M. & Øvreås, L. Planctomycetes dominate biofilms on surfaces of the
623 kelp Laminaria hyperborea. *BMC Microbiol.* (2010) doi:10.1186/1471-2180-10-261.

624 83. Ludington, W. B. *et al.* Assessing biosynthetic potential of agricultural groundwater
625 through metagenomic sequencing: A diverse anammox community dominates nitrate-
626 rich groundwater. *PLoS One* (2017) doi:10.1371/journal.pone.0174930.

627 84. Vidal-Melgosa, S. *et al.* Diatom fucan polysaccharide precipitates carbon during algal
628 blooms. *Nat. Commun.* **12**, 1–13 (2021).

629 85. Walker, A. M., Leigh, M. B. & Mincks, S. L. Patterns in Benthic Microbial
630 Community Structure Across Environmental Gradients in the Beaufort Sea Shelf and
631 Slope. *Front. Microbiol.* (2021) doi:10.3389/fmicb.2021.581124.

632 86. Morris, R. M., Longnecker, K. & Giovannoni, S. J. Pirellula and OM43 are among the
633 dominant lineages identified in an Oregon coast diatom bloom. *Environ. Microbiol.*
634 (2006) doi:10.1111/j.1462-2920.2006.01029.x.

635 87. Sichert, A. *et al.* Verrucomicrobia use hundreds of enzymes to digest the algal
636 polysaccharide fucoidan. *Nat. Microbiol.* (2020) doi:10.1038/s41564-020-0720-2.

637 88. Bohórquez, J. *et al.* Different types of diatom-derived extracellular polymeric
638 substances drive changes in heterotrophic bacterial communities from intertidal
639 sediments. *Front. Microbiol.* (2017) doi:10.3389/fmicb.2017.00245.

640 89. Landa, M. *et al.* Phylogenetic and structural response of heterotrophic bacteria to
641 dissolved organic matter of different chemical composition in a continuous culture
642 study. *Environ. Microbiol.* (2014) doi:10.1111/1462-2920.12242.

643 90. Orsi, W. D. *et al.* Diverse, uncultivated bacteria and archaea underlying the cycling of
644 dissolved protein in the ocean. *ISME J.* (2016) doi:10.1038/ismej.2016.20.

645

646 **Figure legends**

647 **Figure 1: Summarised interactions in marine environments**

648 Correlation heatmap of co-occurring microalgae (columns) and bacteria (rows) in marine
649 environments. “Taxonomy” represents the taxonomic affiliations of microalgal nodes at class

level. “Taxonomy2” represents the taxonomic affiliations of bacteria (p_, c_, o_, f_ and g_ represent Phylum, Class, Order, Family and Genus, respectively). “Project” represents the accession numbers of EMP projects and indicate which project the microalgal and bacterial nodes were recovered from. “Sample_origin” indicates if each sample is planktonic or benthic. Heatmap colour gradient indicates correlation coefficients. Refer Table S2 for raw data used to generate the heatmap. Heatmap was generated using Pheatmap v1.0.12 (https://rdrr.io/cran/pheatmap/).

Figure 2: Summarised interactions in Freshwater environments

Correlation heatmap of co-occurring microalgae (columns) and bacteria (rows) in freshwater environments**.** “Taxonomy” represents the taxonomic affiliations of microalgal nodes at class level. “Taxonomy2” represents the taxonomic affiliations of bacteria (p_, c_, o_, f_ and g_ represent Phylum, Class, Order, Family and Genus, respectively). “Project” represents the accessions of EMP projects and indicate which project the microalgal and bacterial nodes were recovered from. “Sample_origin” indicates if each sample is planktonic or benthic. Heatmap colour gradient indicates correlation coefficients. Refer Table S3 for raw data used to generate the heatmap. Heatmap was generated using Pheatmap v1.0.12 (https://rdrr.io/cran/pheatmap/).

Figure 3: Interactions of the Deltaproteobacterial order *NB1-j* with microalgae in a marine environment.

Figure shows the first neighbours (directly interacting taxa) including microalgal taxa co-occurring with Deltaproteobacterial order *NB1-j*. Nodes are labelled at the order level (If unclassified at the order level, higher taxonomic affiliations are provided). The edge width and node size are continuously mapped to edge weight (correlation strength) and node degree, respectively. Network image was generated using Cytoscape v3.8 (https://cytoscape.org/)

675 **Supplementary Data**

676 **Figure S1: Map indicating the sample collection sites for each EMP project.** Sample
677 collection locations for each EMP project are shown on the world map. Geographic
678 coordinates were drawn in R studio (v1.4.1106) using packages rworldmap v1.3-6
679 (https://rdrr.io/cran/rworldmap/).

680 **Figure S2: Major bacterial Phyla associated with microalgae in freshwater and marine**
681 **environments**

682 Bars represent the number of bacterial nodes affiliated with each phylum. The total number
683 of bacterial nodes identified to be co-occurring with microalgae in marine environments were
684 76 and freshwater environments were 311.

685 **Figure S3: *Planctomycetes* class *OM190* interacting with microalgal taxa in a marine (A)**
686 **and freshwater (B) environment.**

687 Networks (A, B) created from a microalgal-bacterial module by selecting the first neighbours
688 of the *OM190* to showcase the complex interactions with the neighbouring microalgal taxa.
689 The edge width and node size are continuously mapped to edge weight and node degree.
690 Marine and Freshwater environments represent EMP project ERP020022 and ERP016492,
691 respectively. Nodes labelled at order level (If unclassified at the order level, higher
692 taxonomic affiliations are provided). Network image was generated using Cytoscape v3.8
693 (https://cytoscape.org/).

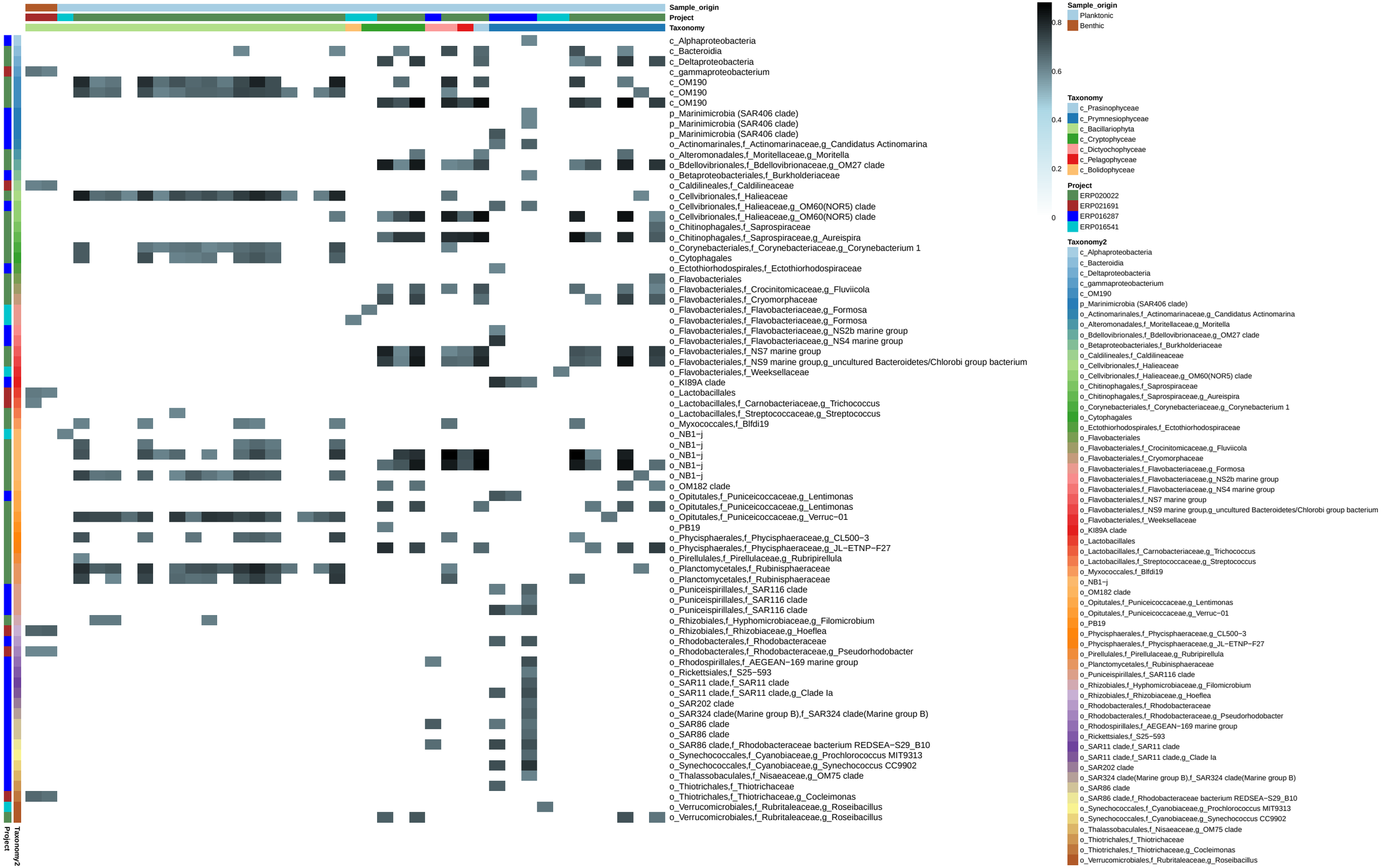
694 **Table S1: Information on selected EMP studies.** Details on Qiita study title, study ids and
695 ENA project accession number.

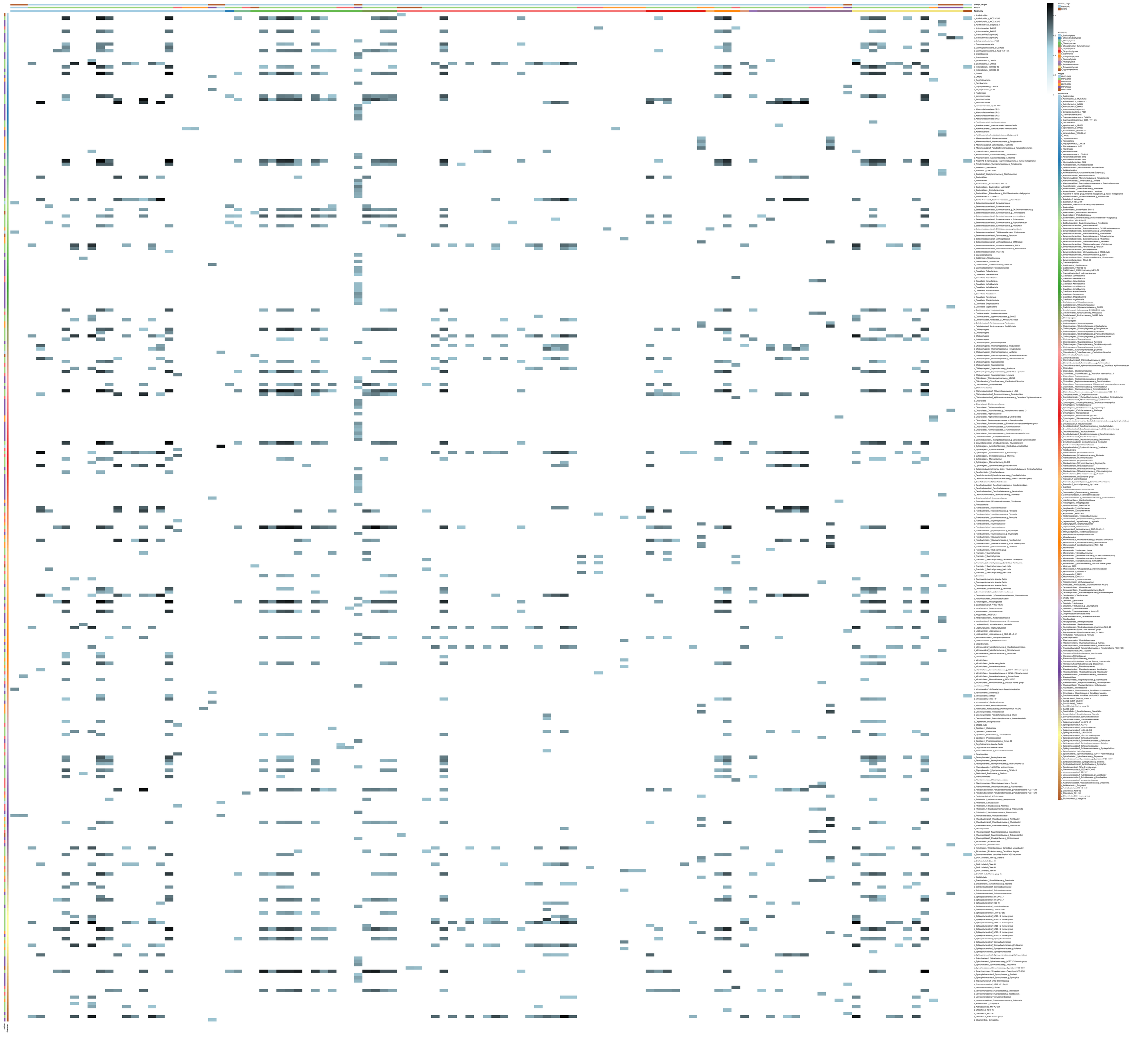
696 **Table S2: Correlations recovered in marine environments between microalgae and**
697 **bacteria.** Summary of significant co-occurrences recovered from the marine environments.
698 Full taxonomic affiliations of the microalgal and bacterial nodes are provided.

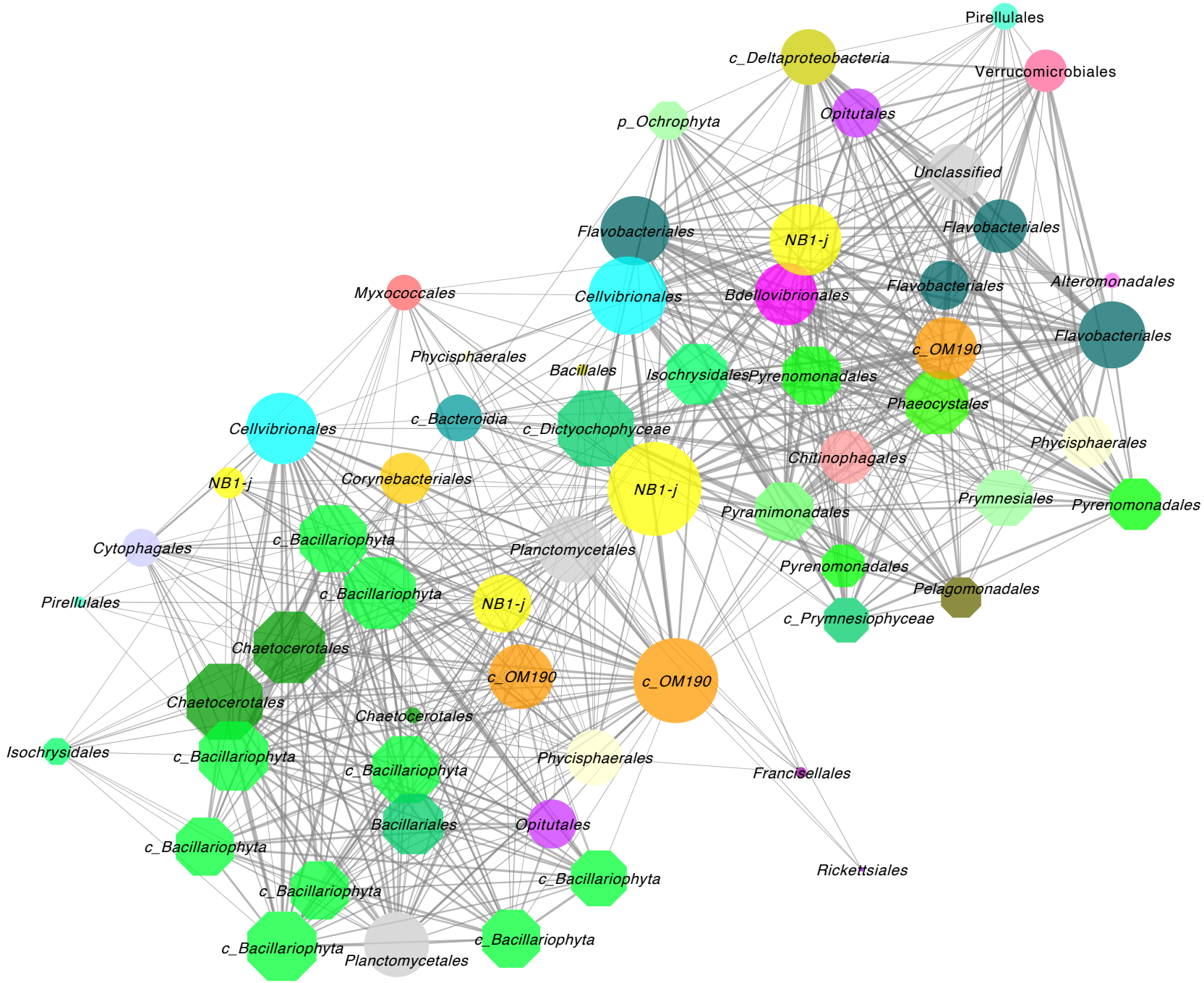
699 **Table S3: Correlations recovered in freshwater environments between microalgae and**

700 **bacteria.** Summary of significant co-occurrences recovered from the freshwater

701 environments. Full taxonomic affiliations of the microalgal and bacterial nodes are provided.


Sample_origin
Project
Taxonomy
c_Alphaproteobacteria
c_Bacteroidia
c_Deltaproteobacteria
c_gammaproteobacterium
c_OM190
c_OM190
c_OM190
p_Marinimicrobia (SAR406 clade)
p_Marinimicrobia (SAR406 clade)
p_Marinimicrobia (SAR406 clade)
o_Actinomarinales,f_Actinomarinaceae,g_Candidatus Actinomarina
o_Alteromonadales,f_Moritellaceae,g_Moritella
o_Bdellovibrionales,f_Bdellovibrionaceae,g_OM27 clade
o_Betaproteobacteriales,f_Burkholderiaceae
o_Caldilineales,f_Caldilineaceae
o_Cellvibrionales,f_Halieaceae
o_Cellvibrionales,f_Halieaceae,g_OM60(NOR5) clade
o_Cellvibrionales,f_Halieaceae,g_OM60(NOR5) clade
o_Chitinophagales,f_Saprospiraceae
o_Chitinophagales,f_Saprospiraceae,g_Aureispira
o_Corynebacteriales,f_Corynebacteriaceae,g_Corynebacterium 1
o_Cytophagales
o_Ectothiorhodospirales,f_Ectothiorhodospiraceae
o_Flavobacteriales
o_Flavobacteriales,f_Crocinitomicaceae,g_Fluviicola
o_Flavobacteriales,f_Cryomorphaceae
o_Flavobacteriales,f_Flavobacteriaceae,g_Formosa
o_Flavobacteriales,f_Flavobacteriaceae,g_Formosa
o_Flavobacteriales,f_Flavobacteriaceae,g_NS2b marine group
o_Flavobacteriales,f_Flavobacteriaceae,g_NS4 marine group
o_Flavobacteriales,f_NS7 marine group
o_Flavobacteriales,f_NS9 marine group,g_uncultured Bacteroidetes/Chlorobi group bacterium
o_Flavobacteriales,f_Weeksellaceae
o_KI89A clade
o_Lactobacillales
o_Lactobacillales,f_Carnobacteriaceae,g_Trichococcus
o_Lactobacillales,f_Streptococcaceae,g_Streptococcus
o_Myxococcales,f_Blfdi19
o_NB1−j
o_NB1−j
o_NB1−j
o_NB1−j
o_NB1−j
o_OM182 clade
o_Opitutales,f_Puniceicoccaceae,g_Lentimonas
o_Opitutales,f_Puniceicoccaceae,g_Lentimonas
o_Opitutales,f_Puniceicoccaceae,g_Verruc−01
o_PB19
o_Phycisphaerales,f_Phycisphaeraceae,g_CL500−3
o_Phycisphaerales,f_Phycisphaeraceae,g_JL−ETNP−F27
o_Pirellulales,f_Pirellulaceae,g_Rubripirellula
o_Planctomycetales,f_Rubinisphaeraceae
o_Planctomycetales,f_Rubinisphaeraceae
o_Puniceispirillales,f_SAR116 clade
o_Puniceispirillales,f_SAR116 clade
o_Puniceispirillales,f_SAR116 clade
o_Rhizobiales,f_Hyphomicrobiaceae,g_Filomicrobium
o_Rhizobiales,f_Rhizobiaceae,g_Hoeflea
o_Rhodobacterales,f_Rhodobacteraceae
o_Rhodobacterales,f_Rhodobacteraceae,g_Pseudorhodobacter
o_Rhodospirillales,f_AEGEAN−169 marine group
o_Rickettsiales,f_S25−593
o_SAR11 clade,f_SAR11 clade
o_SAR11 clade,f_SAR11 clade,g_Clade Ia
o_SAR202 clade
o_SAR324 clade(Marine group B),f_SAR324 clade(Marine group B)
o_SAR86 clade
o_SAR86 clade
o_SAR86 clade,f_Rhodobacteraceae bacterium REDSEA−S29_B10
o_Synechococcales,f_Cyanobiaceae,g_Prochlorococcus MIT9313
o_Synechococcales,f_Cyanobiaceae,g_Synechococcus CC9902
o_Thalassobaculales,f_Nisaeaceae,g_OM75 clade
o_Thiotrichales,f_Thiotrichaceae
o_Thiotrichales,f_Thiotrichaceae,g_Cocleimonas
o_Verrucomicrobiales,f_Rubritaleaceae,g_Roseibacillus
o_Verrucomicrobiales,f_Rubritaleaceae,g_Roseibacillus
Project
Taxonomy2
0.8
0.6
0.4
0.2
0
Sample_origin
Planktonic
Benthic
Taxonomy
c_Prasinophyceae
c_Prymnesiophyceae
c_Bacillariophyta
c_Cryptophyceae
c_Dictyochophyceae
c_Pelagophyceae
c_Bolidophyceae
Project
ERP020022
ERP021691
ERP016287
ERP016541
Taxonomy2
c_Alphaproteobacteria
c_Bacteroidia
c_Deltaproteobacteria
c_gammaproteobacterium
c_OM190
p_Marinimicrobia (SAR406 clade)
o_Actinomarinales,f_Actinomarinaceae,g_Candidatus Actinomarina
o_Alteromonadales,f_Moritellaceae,g_Moritella
o_Bdellovibrionales,f_Bdellovibrionaceae,g_OM27 clade
o_Betaproteobacteriales,f_Burkholderiaceae
o_Caldilineales,f_Caldilineaceae
o_Cellvibrionales,f_Halieaceae
o_Cellvibrionales,f_Halieaceae,g_OM60(NOR5) clade
o_Chitinophagales,f_Saprospiraceae
o_Chitinophagales,f_Saprospiraceae,g_Aureispira
o_Corynebacteriales,f_Corynebacteriaceae,g_Corynebacterium 1
o_Cytophagales
o_Ectothiorhodospirales,f_Ectothiorhodospiraceae
o_Flavobacteriales
o_Flavobacteriales,f_Crocinitomicaceae,g_Fluviicola
o_Flavobacteriales,f_Cryomorphaceae
o_Flavobacteriales,f_Flavobacteriaceae,g_Formosa
o_Flavobacteriales,f_Flavobacteriaceae,g_NS2b marine group
o_Flavobacteriales,f_Flavobacteriaceae,g_NS4 marine group
o_Flavobacteriales,f_NS7 marine group
o_Flavobacteriales,f_NS9 marine group,g_uncultured Bacteroidetes/Chlorobi group bacterium
o_Flavobacteriales,f_Weeksellaceae
o_KI89A clade
o_Lactobacillales
o_Lactobacillales,f_Carnobacteriaceae,g_Trichococcus
o_Lactobacillales,f_Streptococcaceae,g_Streptococcus
o_Myxococcales,f_Blfdi19
o_NB1−j
o_OM182 clade
o_Opitutales,f_Puniceicoccaceae,g_Lentimonas
o_Opitutales,f_Puniceicoccaceae,g_Verruc−01
o_PB19
o_Phycisphaerales,f_Phycisphaeraceae,g_CL500−3
o_Phycisphaerales,f_Phycisphaeraceae,g_JL−ETNP−F27
o_Pirellulales,f_Pirellulaceae,g_Rubripirellula
o_Planctomycetales,f_Rubinisphaeraceae
o_Puniceispirillales,f_SAR116 clade
o_Rhizobiales,f_Hyphomicrobiaceae,g_Filomicrobium
o_Rhizobiales,f_Rhizobiaceae,g_Hoeflea
o_Rhodobacterales,f_Rhodobacteraceae
o_Rhodobacterales,f_Rhodobacteraceae,g_Pseudorhodobacter
o_Rhodospirillales,f_AEGEAN−169 marine group
o_Rickettsiales,f_S25−593
o_SAR11 clade,f_SAR11 clade
o_SAR11 clade,f_SAR11 clade,g_Clade Ia
o_SAR202 clade
o_SAR324 clade(Marine group B),f_SAR324 clade(Marine group B)
o_SAR86 clade
o_SAR86 clade,f_Rhodobacteraceae bacterium REDSEA−S29_B10
o_Synechococcales,f_Cyanobiaceae,g_Prochlorococcus MIT9313
o_Synechococcales,f_Cyanobiaceae,g_Synechococcus CC9902
o_Thalassobaculales,f_Nisaeaceae,g_OM75 clade
o_Thiotrichales,f_Thiotrichaceae
o_Thiotrichales,f_Thiotrichaceae,g_Cocleimonas
o_Verrucomicrobiales,f_Rubritaleaceae,g_Roseibacillus


Sample_origin
Project
Taxonomy
Sample_origin
Pharaonic
Benthic
Taxonomy
Project
Taxonomy


Pirellulales
c_Deltaproteobacteria
Verrucomicrobiales
p_Ochrophyta
Opitutales
Unclassified
Flavobacteriales
NB1-j
Flavobacteriales
Myxococcales
Cellvibrionales
Bdellovibrionales
Flavobacteriales
Alteromonadales
Flavobacteriales
c_OM190
Phycisphaerales
Bacillales
Isochrysidales
Pyrenomonadales
Phaeocystales
Cellvibrionales
c_Bacteroidia
c_Dictyochophyceae
Phycisphaerales
Chitinophagales
NB1-j
Corynebacteriales
NB1-j
Prymnesiales
Pyrenomonadales
Pyramimonadales
Cytophagales
c_Bacillariophyta
Planctomycetales
Pyrenomonadales
Pelagomonadales
Pirellulales
c_Bacillariophyta
NB1-j
c_Prymnesiophyceae
Chaetocerotales
c_OM190
c_OM190
Chaetocerotales
Chaetocerotales
Isochrysidales
c_Bacillariophyta
c_Bacillariophyta
Phycisphaerales
Francisellales
Bacillariales
Opitutales
c_Bacillariophyta
Rickettsiales
c_Bacillariophyta
c_Bacillariophyta
c_Bacillariophyta
c_Bacillariophyta
Planctomycetales