

# Novel diversity within marine Mamiellophyceae (Chlorophyta) unveiled by metabarcoding

Margot Tragin<sup>1</sup> and Daniel Vaulot<sup>1,\*</sup>

<sup>1</sup>Sorbonne Université, CNRS, UMR 7144, Station Biologique, Place Georges Teissier, 29680 Roscoff, France

\*Corresponding author: (ORCID 0000-0002-0717-5685) vaulot@sb-roscott.fr

## ABSTRACT

The Ocean Sampling Day (OSD) project provided metabarcoding data for the V4 hyper-variable regions of the 18S rRNA gene from 157 samples collected at 143 mostly coastal stations. In this paper we focus on the class Mamiellophyceae, which was found at nearly all OSD stations and represented 55 % of the green microalgae (Chlorophyta) reads in the 2014 OSD dataset. We performed phylogenetic analyses of unique OSD metabarcodes (ASV, amplicon single variants) and reference GenBank sequences from cultures and from the environment, focusing on the four most represented genera: *Ostreococcus* (45 % of the Mamiellophyceae reads), *Micromonas* (34 %), *Bathycoccus* (10 %) and *Mantoniella* (8.7 %). These analyses uncovered novel diversity within each genus except *Bathycoccus*. In *Ostreococcus*, a new clade (E) with 2 very clear base pair differences compared to the oceanic clade B in the V4 region was the second most represented clade after the coastal *Ostreococcus "lucimarinus"*. Within *Micromonas*, ten clades were found exceeding the 4 species and 2 candidate species already described. Finally, we found 2 new environmental clades of *Mantoniella*. Each Mamiellophyceae clade had a specific distribution in the OSD dataset suggesting that they are adapted to different ecological niches.

Date: October 21, 2018

## 1 Introduction

2 In marine waters, the accepted paradigm is that the so-called "red" lineage (mainly Diatoms and  
3 Dinoflagellates) is dominant, while the "green" lineage (land plants) is dominating the terrestrial  
4 environments<sup>1</sup>. These two lineages differentiate by the evolutionary origin of their chloroplast:  
5 those of the "green" lineage are surrounded by two membranes, which is an evidence for a  
6 single endosymbiotic event and their major photosynthetic pigments are chlorophyll *a* and *b*<sup>1</sup>.  
7 Studies in coastal waters with both microscopic and molecular techniques found that the "green"  
8 lineage, which is mainly represented by Chlorophyta among unicellular protists, is dominant  
9 in these ecosystems<sup>2–4</sup>. Metabarcoding studies following the development of high throughput  
10 sequencing (HTS) have confirmed the importance of the Chlorophyta in marine waters<sup>5,6</sup>.

11 The European Ocean Sampling Day project (OSD) sampled the world coastal waters in 2014  
12 around the summer solstice (21 June) with the aim of analyzing the diversity and distribution of  
13 marine microorganisms<sup>7</sup> especially in coastal waters using 18S rRNA metabarcodes (V4 and V9  
14 hyper-variable regions)<sup>8</sup>. This dataset allowed to determine the distribution of fourteen classes  
15 of Chlorophyta<sup>9</sup>. Mamiellophyceae<sup>10</sup>, the most prevalent class in all coastal environments<sup>11</sup>, did  
16 not show any geographic distribution patterns or environmental preference<sup>9</sup>. We hypothesize  
17 that, in order to detect distribution patterns, this class should be investigated at lower taxonomic  
18 levels, in particular the species level.

19 Mamiellophyceae consist of four orders: Mamiellales, Bathycoccales, Dolichomastigales and  
20 Monomastigales<sup>10</sup>. Monomastigales are confined to freshwater environments while Dolichomasti-  
21 gales, although quite diversified in marine waters<sup>12,13</sup>, are a minor component of Mamiel-  
22 lophyceae in coastal waters. In contrast, Mamiellales and Bathycoccales host some of the most  
23 common Chlorophyta microalgae such as the ubiquitous *Micromonas*, the smallest known eu-  
24 karyote *Ostreococcus* or the coccoid *Bathycoccus*<sup>10</sup>. Within Mamiellales, *Micromonas pusilla*<sup>14</sup>  
25 was recently split into four species: *Micromonas bravo* (previously clade B.E.3), *Micromonas com-*  
26 *moda* (previous clade A.ABC.1-2), *Micromonas polaris* (previously clade B arctic), *Micromonas*  
27 *pusilla* (previously clade C.D.5) and two clades mentioned as candidate species 1 (clade B...4)

and candidate species 2 (clade B warm)<sup>15</sup>. Within the genus *Mantoniella*, only two species have been described: the ubiquitous *Mantoniella squamata*<sup>3,16</sup>, first described as *Micromonas squamata*<sup>17</sup>, and *Mantoniella antarctica*<sup>18</sup>. Within Bathycoccales, four *Ostreococcus* clades have been delineated<sup>19</sup>: *Ostreococcus tauri*<sup>20</sup>, *Ostreococcus mediterraneus*<sup>21</sup>, both of which have been formerly described, *Ostreococcus "lucimarinus"* (clade A) and clade B, which both lack formal taxonomic description. Analyses of pigment content and response to light levels allowed to distinguish two broad ecotypes: strains adapted to high (*O. tauri*, *O. mediterraneus*, *O. "lucimarinus"*) and low (*Ostreococcus* clade B) light<sup>22</sup>. The second genus within Bathycoccales hosts a single species, *Bathycoccus prasinos*<sup>23</sup>. No clades can be delineated inside this species based on 18S rRNA gene sequences from cultures and the environment. However, divergence in ITS sequences suggest that *Bathycoccus prasinos* probably consists of two different species<sup>24,25</sup>.

The present paper uses the OSD V4 dataset to analyze the taxonomic diversity and global distribution of four major Mamiellophyceae genera: *Ostreococcus*, *Micromonas*, *Bathycoccus* and the less studied *Mantoniella*. Our analyses reveal the existence of novel clades within *Ostreococcus* and *Micromonas* and allowed to determine that most species/clades have specific oceanic distributions.

## Material and Methods

The Ocean Sampling Day consortium provided 2 metabarcoding datasets for 2014 using the V4 region of the 18S rRNA gene: the LGC dataset consisting of 157 water samples from 143 stations (see Table S1) filtered on 0.22 µm pore size Sterivex without prefiltration and the Life Watch (LW) dataset consisting of a subset of 29 water samples filtered on 0.8 µm pore size polycarbonate membranes without pre-filtration. The extraction, PCR and sequencing protocols have been described previously<sup>8,9</sup>. In brief, the LGC and LW data originated from the same water samples but have been processed independently for filtration, DNA extraction, PCR amplification and sequencing (both Illumina 2x250 bp). LGC and LW provided respectively about 5 and 9 millions of V4 sequences, resulting in higher coverage for the LW dataset.

The LGC and LW datasets (<https://github.com/MicroB3-IS/osd-analysis/wiki/Guide-to-OSD>-

55 2014-data) were analyzed with the same pipeline using the Mothur software v. 1.35.1<sup>26</sup>. Reads  
56 were filtered to keep only sequences without ambiguities (N) and longer than 300 pb. Reads  
57 were aligned on SILVA seed release 123 alignment<sup>27</sup> corrected by hand to remove gaps at the  
58 beginning and at the end of alignments. The aligned datasets were filtered by removing columns  
59 containing only insertions. Chimeras were checked using Uchime v. 4.2.40<sup>28</sup> as implemented  
60 in Mothur. Unique sequences (ASVs or Amplicon Single Variants<sup>29</sup>) were assigned using  
61 the Wang classifier as implemented in Mothur and the PR<sup>2</sup> reference database<sup>30</sup> version 4.2  
62 ([https://figshare.com/articles/PR2\\_rRNA\\_gene\\_database/3803709/2](https://figshare.com/articles/PR2_rRNA_gene_database/3803709/2)) for which the Chlorophyta  
63 sequences had been recently curated<sup>6</sup>.

64 To check the assignation and explore the genetic diversity of Mamiellophyceae, ASVs from  
65 *Bathycoccus*, *Micromonas*, *Ostreococcus* and *Mantoniella* represented by more than 200 reads  
66 for either LW and LGC were aligned to GenBank reference sequences using MAFFT v. 7.017<sup>31</sup>.  
67 Maximum likelihood (ML) phylogeny was built using Fast Tree v. 1.0<sup>32,33</sup> as implemented in  
68 Geneious v. 7.1.9<sup>34</sup>. Bayesian phylogeny was built with Mr Bayes v. 3.2.6.<sup>35</sup> implemented in  
69 Geneious. Clades were defined both with the presence of clear signatures in the alignments and  
70 by phylogenetic features<sup>6,36,37</sup>): monophyletic assemblages of sequences in the phylogenetic  
71 trees, well supported (> 70 % bootstrap), and found with the two different construction methods  
72 (i.e. maximum likelihood and Bayesian).

73 Relative abundance of selected ASVs was computed using R software version 3.3.1 (<http://www.R-project.org/>) on the LGC dataset for a subset of 92 samples with more than 100 Mamiellophyceae  
74 reads. Graphics were performed using the R packages ggplot2, ComplexHeatmap<sup>38</sup> and  
75 treemapify.

## 77 Results and Discussion

78 We analyzed unique sequences (ASVs for Amplicon Single Variants<sup>29</sup>) of two separate OSD  
79 data sets: LGC and LW. We focused on the LGC dataset which encompasses a much larger  
80 number of samples than the LW dataset which corresponds to a subset of the OSD samples  
81 that have been processed in a completely independent manner compared to the LGC (different

82 filtration, DNA extraction, PCR and Illumina sequencing). The LW dataset was used mainly to  
83 confirm the fact that the LGC ASVs were not artefacts by verifying that to any major LGC ASV  
84 corresponded a LW ASV with a strictly identical sequence.

85 Mamiellophyceae represented 55 % of the Chlorophyta reads found in the OSD 2014 surface  
86 samples<sup>9</sup>. Seven described Mamiellophyceae genera were recovered among LGC ASVs while  
87 the other ASVs could only be classified at the family or order level. Four genera were clearly  
88 dominant: *Ostreococcus*, *Micromonas*, *Bathycoccus* and *Mantoniella*, the two former with almost  
89 equal contribution (Fig. 1). We decided to focus on the 23 major LGC ASVs for these four  
90 genera that were represented by at least than 200 reads (Table 1) corresponding 68% of the  
91 Mamiellophyceae reads. Among these ASVs, *B. prasinos* was clearly the most ubiquitous  
92 followed by 2 *Micromonas* clades and *O. "lucimarinus"* (Fig. S1).

93 ***Ostreococcus***

94 4,223 ASVs from the LGC dataset were assigned to the genus *Ostreococcus*, among which 10  
95 corresponded to more than 200 reads (Table 1). For each of these abundant LGC ASVs, we  
96 always found a corresponding ASV in the LW dataset. These sequences constituted five clades  
97 (Fig. 2A), four already described and a new one that we named clade E following the initial  
98 nomenclature of *Ostreococcus* clades<sup>19</sup>. The same tree topology was recovered with ML and  
99 Bayesian methods. Alignments (Fig. 2B) confirmed, that the V4 region exhibited clear signatures  
100 to delineate the five *Ostreococcus* clades. In this region, the genetic variation between clades  
101 (Table S2) is too low to discriminate the different clades when V4 OTUs are built at 99 % identity  
102 threshold, except for *O. mediterraneus* (98.3 %). In terms of overall distribution, *Ostreococcus*  
103 was completely absent from high latitudes beyond 60 °N (Fig. 3). *O. "lucimarinus"* was present  
104 in the largest number of samples (Fig. S1). In general none of the different *Ostreococcus*  
105 species/clades seemed to co-occur, with the exception of *O. tauri* and *O. "lucimarinus"* and to a  
106 lesser extent, clade B and E (Fig. S2).

107 Three ASVs were assigned to the first described species of *Ostreococcus*, *O. tauri*, with  
108 a number of reads ranging from 3,500 and 1,000 (Table 1). These unique sequences did not

109 form a monophyletic clade in either ML and Bayesian trees (Fig. 2A). The two major ASVs  
110 (ASV\_LGC\_00012 and ASV\_LGC\_00013) were close to the 18S rRNA sequence recovered from  
111 the genome of *O. tauri* (GQ426344). The third ASV (ASV\_LGC\_00018) was identical to sequence  
112 Y15814 which was obtained more than 20 years ago for the type strain of *O. tauri*. The V4  
113 alignment (Fig. 2B) reveals two signatures for *O. tauri*. *O. tauri* ASVs represented more than 1  
114 % of Mamiellophyceae reads at 35 stations and were abundant on the East Coast of the US, in  
115 the Baltic Sea, the Adriatic Sea (Venice lagoon), the Black Sea and in Uruguay lagoons (Fig. 3  
116 and S3). Some of these samples correspond to brackish environments (e.g. OSD35 – 8.97 psu,  
117 OSD36 – 7.42, OSD39 – 24.3 psu, OSD186 – 7.20 psu) which are similar to the Thau lagoon  
118 (highly variable salinity isfrom 24 to 38 psu), where *O. tauri* was initially isolated.<sup>20</sup> Surprisingly  
119 *O. tauri* was absent from the western Mediterranean Sea despite the fact that strains have been  
120 isolated from there<sup>21</sup>. The existence of several ASVs suggests that *O. tauri* might be a complex  
121 of species, that need to be better distinguished and that could be adapted to different ranges of  
122 salinity.

123 Two major ASVs were assigned to *O. "lucimarinus"*, which was initially described as a high  
124 light adapted clade<sup>22</sup>, corresponding to 23,119 reads (Table 1) and representing up to 64 % of  
125 the Mamiellophyceae reads off South Africa (Robben Island, OSD133). It dominated Atlantic and  
126 North Sea European coastal stations (Belgium OSD183 62 % and 184 44 %, Portugal OSD115  
127 53 %) and represented 40 % of Mamiellophyceae reads at one of the 3 Azores stations (OSD98,  
128 Fig. 3). In contrast, *O. "lucimarinus"* was almost totally absent from the Mediterranean Sea and  
129 tropical waters (Fig. 3 and S3). This distribution is coherent with observations by qPCR (clade  
130 OI according to Demir-Hilton *et al.* 2011<sup>39</sup>, see below) as a cold mesotrophic coastal clade.

131 A single ASV was assigned to *Ostreococcus* clade B, initially described as low light adapted  
132 clade<sup>22</sup>, representing 6,207 reads and reaching 62 % of Mamiellophyceae reads off Panama  
133 (OSD51, Fig. 3). Clade B contribution to Mamiellophyceae was higher than 10 % at 7 tropical  
134 and sub-tropical stations from a range of oceans (OSD60 South Carolina, OSD25, 37, 51 Florida,  
135 OSD95 Singapore, OSD122 Red Sea, OSD147 Sri Lanka, Fig. 3), which is consistent with  
136 previous results obtained by qPCR<sup>39,40</sup>.

137 The novel *Ostreococcus* clade E was represented by a single ASV in the LGC dataset  
138 (22,720 reads) and a similar sequence was also found in the LW dataset, suggesting that it  
139 was not an artefact. Clade E sequence is very close to that of clade B, with 2 clear differences  
140 (Fig. 2B) and is 100 % similar to a single Genbank sequence (MH008654) obtained also by  
141 Illumina sequencing from South China Sea waters. The V4 sequence of clade E is 99.4%  
142 similar to that of clade B, such that these two clades may have been lumped together in many  
143 metabarcoding studies which do not consider ASVs but OTUs. Clade E could be locally dominant  
144 in OSD samples representing up to 70 % of the Mamiellophyceae reads (OSD111, off Portugal).  
145 It dominated coastal warm temperate stations (Fig. 3) on both sides of the Atlantic Ocean  
146 (Southern USA, OSD39, 58 and 143; Portugal OSD81, 111, 117, 153; France OSD154) and the  
147 Mediterranean Sea (Adriatic Sea off Venice, OSD69). It is surprising that no culture has ever  
148 been obtained for this new clade but it could be due to the fact that it requires specific conditions  
149 to grow.

150 Finally, two ASVs with a low number of reads (617 in the LGC dataset, Table 1) were assigned  
151 to the species *O. mediterraneus*, which was only found in significant proportion in a lagoon of  
152 the coast of Uruguay (OSD149 5 % and 150 6 %, Fig. 3). This distribution is coherent with the  
153 fact that almost all strains of *O. mediterraneus* have been isolated from coastal lagoons along  
154 the Mediterranean Sea coast<sup>21</sup>, suggesting that this species may be restricted to very specific  
155 environments with fluctuating salinity.

156 Two sets of qPCR primers and probes have been previously designed<sup>39</sup> based on available  
157 V4 sequences from strains in culture in order to discriminate two *Ostreococcus* groups (OI and  
158 OII). The OI set targets *O. "lucimarinus"* but also recognizes *O. tauri*<sup>39</sup> which has 2 mismatches  
159 (Fig. S4), while the OII group targets clade B. Interestingly the new clade E has four mismatches  
160 to set OI and two mismatches to set OII. Some of these mismatches are located on the 3' end of  
161 the forward primer which may prevent either of these sets to recognize *Ostreococcus* clade E,  
162 although this would have to be tested once clade E strains become available. If none of these  
163 sets recognizes clade E, it would be necessary to design new qPCR sets for clade E. If one  
164 or two of these sets recognize clade E, it will cast serious doubts on the validity of previous

165 analyses<sup>25,39,40</sup> since clade E seems to have a different distribution from both *O. "lucimarinus"*  
166 and clade B.

167 ***Micromonas***

168 4,285 unique OSD LGC sequences were assigned to the genus *Micromonas* with 10 correspond-  
169 ing to more than 200 reads. Phylogenetic analysis and V4 signatures allowed to divide these  
170 sequences into 10 major clades (Fig. 4) corresponding to the species, candidate species and  
171 clades recently described by Simon *et al.*<sup>15</sup> with the exception of a new clade not seen previously,  
172 named clade B sub-arctic. *M. commoda*, *M. bravo* and clade B warm (candidate species 2<sup>15</sup>)  
173 could be each divided further into sub-clades: *M. commoda* AB and C<sup>41,42</sup>, *M. bravo* I and II,  
174 and B warm I and II. The well supported tree topology was recovered with both ML and Bayesian  
175 methods. The genetic divergence between clades is larger than 1 % for almost all the clade pairs  
176 (Table S3) and allows to distinguish all clades when using a 99 % identity threshold except for *M.*  
177 *commoda* AB and C (99.2 % identity), *M. polaris* and the new clade B sub-arctic (99.2 %), and  
178 *M. commoda* and clade B warm. *M. commoda* C was present in the largest number of samples  
179 (Fig. S1). Most species/clades did not co-occur, with the exception of *M. bravo* I and II (Fig. S2).

180 The major *Micromonas* ASV represented by 28,810 reads, was assigned to *M. bravo* II  
181 (Fig. 4). *M. bravo* is a newly described species<sup>15</sup> which was previously part of the B clade<sup>19</sup>. This  
182 ASV represented up to 60 % of the Mamiellophyceae reads (Fig. 5 and S5) in the Black Sea  
183 (OSD131) and off Portugal (OSD102). It dominated most of Mediterranean Sea stations, some  
184 North European stations and to a lower extent some Pacific stations (off California OSD43 14 %  
185 and Hawaii OSD144 28 %). It was the only *Micromonas* ASV that represented more than 10 %  
186 of Mamiellophyceae reads off Japan (OSD124 29 %). *M. bravo* I ASV (5,459 reads) represented  
187 more than 10 % of Mamiellophyceae at 3 stations along the European coast (in particular 47 %  
188 in the English Channel off Plymouth, OSD1).

189 Two LGC ASVs with a single base pair difference (Fig. 4B) were closely related to the  
190 references sequences of the clade B-warm, corresponding to "candidate species 2" from Simon  
191 *et al.*<sup>15</sup> and to "clade VI" from Lin *et al.* 2017<sup>43</sup>. They are referred here as B warm I and B

warm II. Clade B warm I ASV (5,251 reads) which matched the original "candidate species 2" sequences (Fig. 4B) contributed to more than 10 % of the Mamiellophyceae reads at 5 stations from tropical or warm waters (Fig. 5: at the 3 Hawaii stations (OSD56 38 %, 57 20 %, 144 18 %), off Florida (OSD37 23 %), off Portugal (OSD101 23 %). It was also found in Singapore (OSD95 5 %). This distribution confirmed the data from available Genbank sequences<sup>15</sup>: one representative strain has been isolated in Mediterranean Sea in summer (<http://roscoff-culture-collection.org/rcc-strain-details/1109RCC1109>) and environmental clones have been recovered in the Red Sea<sup>44</sup>, South China Sea ("unknown clade"<sup>45</sup>) and off Taiwan (Micromonas clade VI<sup>43</sup>). The second ASV from B warm II did not match any GenBank sequence (Fig. 4B) and was recorded in 1,045 copies, representing more than 1 % of the Mamiellophyceae reads at three stations in Venice lagoon (OSD47, 69, 70 16), off Portugal (OSD81) and in Singapore strait (OSD95).

*M. commoda* C ASV (10,787 reads) represented more than 1% of the Mamiellophyceae at 57 stations (Fig. S1) and up to 35 % off the Atlantic coast of Canada (OSD152). It was found in the North Atlantic up to Iceland and as well off New Zealand (Fig. 5 and S5). In contrast, it was almost completely absent from the Mediterranean Sea as well as from tropical stations. *M. commoda* AB (5,465 reads) was above 1% at a much lower number of stations (16, Fig. 1) and contributed up to 40 % of the Mamiellophyceae off Sri Lanka (OSD147). It was distributed in tropical and subtropical waters (Fig. 5 and S5) in particular off Florida, Singapore and Hawaii as well as in the Eastern Mediterranean Sea off Israel. *M. commoda* was described by Van Baren *et al.*<sup>46</sup> who just mentioned that this species was not recorded in high latitudes yet (beyond 60 °North and South). The species was then revised by Simon *et al.*<sup>15</sup>, who described the distribution of this species as ubiquitous based on available Genbank sequences. The genetic variability within this species had already been highlighted previously<sup>41,47</sup> and is comforted by the OSD data since the two clades AB and C have clearly distinct distributions. Simon *et al.*<sup>15</sup> proposed the hypothesis that speciation events may ongoing within *M. commoda*.

Thz ASV (1,616 reads) corresponding to *Micromonas* clade B...4 ("candidate species 1" according to Simon *et al.* 2017<sup>15</sup>) reached 15 % off Florida (OSD29) and was found in rather

220 warm waters (off Hawaii, Israel, Morocco, Fig. 5) which matches the distribution of Genbank  
221 sequences that had been previously recovered from the Mediterranean and Red Seas as well  
222 from the Pacific Ocean<sup>15</sup>.

223 *M. polaris* ASV (2,200 reads) was the major Mamiellophyceae contributor (Fig. 5) at 2 stations  
224 in Arctic waters (73 % in Nunavut OSD105, 66 % in Fram Strait OSD146) and also present in the  
225 Gulf of Finland (OSD30). This consistent with the current knowledge on this species. *M. polaris*  
226 was first isolated from the Arctic Ocean<sup>48</sup> and shown to be the dominant pico-eukaryote in the  
227 Beaufort Sea in the summer<sup>49</sup>. It has also been recently recorded in the Southern Ocean<sup>50</sup>,  
228 although its presence appears less prevalent since it is absent from environmental Antarctic  
229 clone libraries<sup>6</sup>. A new *Micromonas* clade genetically close to *M. polaris* (B sub-arctic, 1,096  
230 reads) had maximum contribution off Canada (Bedford Basin OSD 152 32 %) and represented  
231 more than 10 % of Mamiellophyceae reads at 4 sub-arctic stations off Maine and Iceland as well  
232 as at a temperate location off UK coast in the North Sea (Fig. 5). This ASV is 100 % similar to  
233 Genbank sequences recently obtained in the White Sea<sup>51</sup>.

234 Finally, the ASV corresponding to the first described *Micromonas* species, *M. pusilla*, was  
235 found in low abundance (1,259 reads) mostly in temperate locations (Fig. 5) corresponding to  
236 the environment from which it was initially described and isolated (e.g. CCMP490 isolated from  
237 Woods Hole, USA, and CCMP491 from the English Channel, UK<sup>41</sup>).

### 238 ***Bathycoccus***

239 The ASV corresponding to *Bathycoccus* was both the most abundant (24,391 reads) as well  
240 as the most prevalent, exceeding 1 % of the Mamiellophyceae reads at 72 stations (Fig. S1)  
241 distributed all over the coastal ocean from tropical to polar waters (Fig. 6 and S6). This very  
242 global distribution of the genus matches what has been observed based on the TARA Oceans  
243 dataset<sup>24</sup> where the metagenome of *Bathycoccus* was recovered at a wide rang of stations.  
244 *Bathycoccus* is now known to be composed of two cryptic species with identical 18S rRNA  
245 sequences (that therefore cannot be distinguished in the OSD dataset) but differences in the  
246 ITS gene as well as at the genomic level<sup>24,52</sup>. The distribution of these two cryptic species

247 (BI-genome RCC1105 and BII-genome TOSAG39-1) determined by metagenomic analysis  
248 and qPCR suggest that BI could be coastal, while BII could be adapted to warmer oceanic  
249 waters<sup>24, 25, 53</sup>.

250 **Mantoniella**

251 8,570 reads corresponding to two ASVs with more than 200 reads could be assigned to the  
252 genus *Mantoniella*. Besides the morphological species *M. squamata*<sup>3, 16</sup> and *M. antarctica*<sup>18</sup>, no  
253 clades based on 18S rRNA gene sequences have been yet described<sup>10, 19</sup>. However, the V4  
254 hyper-variable region of the 18S rRNA gene using publicly available reference sequences as  
255 well as OSD ASVs highlights two new *Mantoniella* clades that we named A and B, both with well  
256 supported phylogenies (Fig. 7A) and with clear sequence signatures (Fig. 7B).

257 The most abundant ASV corresponded to the species *M. squamata* (4,500 reads) represented  
258 up to 70 % of the Mamiellophyceae reads off Greenland (OSD80, Fig. 8 and S7) while at all  
259 other stations it contributed always less than 10 %. However it was found at stations with very  
260 different environmental conditions, including off Hawaii and in a lagoon on the coast of Uruguay  
261 (Fig. 8). This is consistent with previous descriptions of *M. squamata* as a cosmopolitan species<sup>3</sup>  
262 and several studies reported in particular its presence in Northern high latitudes<sup>49, 54</sup>.

263 The ASV corresponding to *Mantoniella* clade B (4,070 reads) was more widespread, rep-  
264 resenting more than 1 % of the Mamiellophyceae reads at 16 stations (Fig. S1) especially in  
265 the Venice lagoon (OSD47 53 %, OSD70 36 % and OSD69 12 %) but also in other moderately  
266 warm waters (Eastern Mediterranean Sea, Gibraltar, California, Hawaii, Fig. 8 and S7), matching  
267 Genbank sequences previously found in the Mediterranean Sea<sup>12</sup>.

268 No ASV corresponding to *Mantoniella* clade A was found in the LGC dataset but one ASV  
269 from the LW dataset (841 reads) matched this clade but its only present off Greenland (OSD80)  
270 which matched the environment where the other Genbank sequences have been obtained (sea  
271 ice from the Arctic Ocean<sup>55</sup>, White Sea<sup>51</sup> and Baltic Sea<sup>56</sup>) pointing to *Mantoniella* clade A being  
272 potentially an ice alga.

273 Finally no ASV corresponded to the species *M. antarctica* was found either in the LGC or

274 LW dataset. This suggests that this species is probably restricted to Antarctica and was not  
275 present at the single station sampled there (OSD187) in the mid of the austral winter where only  
276 85 Mamiellophyceae reads were recorded.

277 **Limitations of the OSD dataset**

278 The OSD metabarcoding dataset is invaluable to determine the distribution of many phytoplankton  
279 groups in coastal waters<sup>9,11,57</sup>. However it must be always emphasized that metabarcoding is not  
280 a quantitative method because of biases in the PCR reaction and of the variation in the number of  
281 18S rRNA gene copies per organism<sup>58</sup> However in the case of Mamiellophyceae, because their  
282 size is small, the number of copies is low<sup>58</sup> and does not vary too much between the different  
283 genera (e.g. 2 and 4, copies in *Bathycoccus* and *Ostreococcus*, respectively<sup>59,60</sup>). This supports  
284 the use of relative number of reads as a semi-quantitative proxy of Mamiellophyceae contributions.  
285 Another potential problem of the OSD dataset is the fact that sampling was done everywhere  
286 almost on the same date (around June 21). It has the advantage of providing a snapshot of  
287 the world coastal ocean, which allows to work on spatial distribution without the impact of the  
288 seasonality, in contrast to oceanographic expeditions such as Tara Oceans<sup>5</sup> which sampled  
289 different ecosystems at different time of the year. One OSD limitation however is that the northern  
290 and hemisphere were sampled at opposite sides of the yearly cycle, at the end of the spring  
291 and of the fall, respectively. However since sampling was mostly in the Northern hemisphere,  
292 this has a low impact on the data interpretation. The existence of phytoplankton cycles driven  
293 by temporal changes and nutrient cycles in coastal environment<sup>61–63</sup> may explain why some  
294 species were not found in the OSD dataset in regions where strains or clones corresponding to  
295 the same species have been isolated before. As an example, *Bathycoccus* initially isolated from  
296 the Gulf of Naples in the Mediterranean Sea<sup>23</sup> was not recovered at the Naples OSD station  
297 (OSD4), where only 2 Mamiellophyceae reads were obtained. Metabarcoding analyses at the  
298 Long Term Ecological Research station in the Gulf of Naples show that Mamiellophyceae were  
299 absent in June<sup>64</sup>, which explain why they are absent from the OSD dataset. In contrast, analysis  
300 of several time series<sup>65</sup> led to the conclusion that *M. bravo* (previously non arctic B.E.3 clade)

301 dominated the *Micromonas* community in summer and should be adapted to warm well lighted  
302 coastal waters which is consistent with what we observed in the OSD dataset for which sampling  
303 was done in June.

## 304 Conclusion

305 Although Mamiellophyceae are clearly the dominant group of green algae in coastal waters<sup>9,11</sup>,  
306 previous analysis revealed that their relative abundance could not be related to any environmental  
307 variable<sup>9</sup>. Even analysis at the genus level is not sufficient to detect bio-geographical patterns.  
308 For example, the genus *Micromonas* is found at virtually all OSD stations (Fig. 5). The only  
309 genus which is not ubiquitous is *Ostreococcus* which is not found in polar regions. In contrast,  
310 analysis at the species/clade levels allows detecting some very clear patterns in particular with  
311 respect to latitudinal distribution (Table 2. Some species/clade have quite restricted latitudinal  
312 ranges, e.g. *M. polaris* only found in Arctic samples, *M. pusilla* only found in temperate waters  
313 or *Micromonas* clade B warm I only in tropical waters. Others are much more wide spread, for  
314 example *Ostreococcus* clade B which extends from temperate to tropical waters. Some taxa  
315 seem to be restricted to specific habitats, in particular *O. mediterraneus* to coastal lagoons. The  
316 case of the Mediterranean Sea is also interesting. It has been previously shown to harbour  
317 specific taxonomic groups such as Chlorodendrophyceae<sup>9</sup>, clade A6 of Chloropicophyceae<sup>11</sup> or  
318 the Bolidophyceae *Triparma mediterranea*<sup>66</sup>. Here, we found that some temperate species such  
319 as *M. pusilla* or *Ostreococcus* clade E were completely absent from this region but we did not  
320 find any species/clade restricted to it.

321 Genetic diversity is quite different between the four genera that we examined. While *Bathy-*  
322 *coccus* is composed of a single clade at least based on the 18S rRNA gene, while *Micromonas*  
323 seems extremely diversified with a large number of clades. Although most of the clades found in  
324 OSD have been observed before, we uncovered some new diversity. One species, *M. bravo* was  
325 split into two sub-clades, I and II, that seem to have distinct distribution. Two new clades were  
326 uncovered. *Micromonas* B subarctic seems to be restricted to a specific latitudinal band (roughly  
327 45 to 65 °N). *Ostreococcus* clade E seems also very interesting since it is very abundant and has

328 a distinct distribution from the closely related *Ostreococcus* clade B by being restricted to tem-  
329 perate waters while clade B is found tropical waters. Since some oceanic regions have not been  
330 covered by the OSD dataset, it is possible that yet undiscovered Mamiellophyceae clade/species  
331 may exist in particular in the Southern hemisphere. In order to obtain more information on  
332 the new clades reported here and to determine whether they correspond to potentially novel  
333 species, several strategies are possible. First, now that their geographical distribution begins to  
334 be known, we may target specific environments and obtain isolates. This will allow analyzing  
335 finer resolution markers such as the ITS<sup>15</sup> and determine physiological preferences, for example  
336 in terms of temperature<sup>67</sup>. Another strategy for clades that are hard to isolate in culture would be  
337 to determine longer sequences of the ribosomal operon including in particular the ITS region  
338 directly from the environment either by long amplicon PCR using novel sequencing technologies  
339 such as Nanopore<sup>68</sup> or by extracting them from existing metagenomics datasets such as those  
340 obtained during the Tara Oceans project<sup>24</sup>.

## 341 References

- 342 1. Falkowski, P. G., Schofield, O., Katz, M. E., Van de Shootbrugge, B. & Knoll, A. H. Why is the  
343 land green and the ocean red? In Thierstein, H. R. & Young, J. R. (eds.) *Coccolithophores:*  
344 *from Molecular processes to global impact*, 427–453 (Springer, Berlin, 2004).
- 345 2. Throndsen, J. Occurrence and productivity of small marine flagellates. *Nor. J. Bot.* **23**,  
346 269–293 (1976).
- 347 3. Thomsen, H. A. & Buck, K. R. Nanoflagellates of the central California waters: tax-  
348 onomy, biogeography and abundance of primitive, green flagellates (Pedinophyceae,  
349 Prasinophyceae). *Deep. Sea Res. Part II: Top. Stud. Oceanogr.* **45**, 1687–1707, DOI:  
350 10.1016/S0967-0645(98)80013-1 (1998).
- 351 4. Not, F., Simon, N., Biegala, I. & Vaultot, D. Application of fluorescent in situ hybridization  
352 coupled with tyramide signal amplification (FISH-TSA) to assess eukaryotic picoplankton  
353 composition. *Aquatic Microb. Ecol.* **28**, 157–166, DOI: 10.3354/ame028157 (2002).

- 354 5. de Vargas, C. *et al.* Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**, 1261605–  
355 1261605, DOI: 10.1126/science.1261605 (2015).
- 356 6. Tragin, M., Lopes dos Santos, A., Christen, R. & Vaultot, D. Diversity and ecology of  
357 green microalgae in marine systems: an overview based on 18S rRNA gene sequences.  
358 *Perspectives Phycol.* **3**, 141–154, DOI: 10.1127/pip/2016/0059 (2016).
- 359 7. Kopf, A. *et al.* The ocean sampling day consortium. *GigaScience* **4**, 27, DOI: 10.1186/  
360 s13742-015-0066-5 (2015).
- 361 8. Tragin, M., Zingone, A. & Vaultot, D. Comparison of coastal phytoplankton composition  
362 estimated from the V4 and V9 regions of the 18S rRNA gene with a focus on photosynthetic  
363 groups and especially Chlorophyta. *Environ. Microbiol.* **20**, 506–520, DOI: 10.1111/1462-2920.  
364 13952 (2018).
- 365 9. Tragin, M. & Vaultot, D. Green microalgae in marine coastal waters: The Ocean Sampling  
366 Day (OSD) dataset. *Sci. Reports* **8**, 14020, DOI: 10.1038/s41598-018-32338-w (2018).
- 367 10. Marin, B. & Melkonian, M. Molecular phylogeny and classification of the Mamiellophyceae  
368 class. nov. (Chlorophyta) based on sequence comparisons of the nuclear- and plastid-  
369 encoded rRNA operons. *Protist* **161**, 304–336, DOI: 10.1016/j.protis.2009.10.002 (2010).
- 370 11. Lopes dos Santos, A. *et al.* Diversity and oceanic distribution of prasinophytes clade VII,  
371 the dominant group of green algae in oceanic waters. *The ISME J.* **11**, 512–528, DOI:  
372 10.1038/ismej.2016.120 (2016).
- 373 12. Viprey, M., Guillou, L., Ferréol, M. & Vaultot, D. Wide genetic diversity of picoplanktonic  
374 green algae (Chloroplastida) in the Mediterranean Sea uncovered by a phylum-biased PCR  
375 approach. *Environ. Microbiol.* **10**, 1804–1822, DOI: 10.1111/j.1462-2920.2008.01602.x (2008).
- 376 13. Monier, A., Worden, A. Z. & Richards, T. A. Phylogenetic diversity and biogeography of the  
377 Mamiellophyceae lineage of eukaryotic phytoplankton across the oceans. *Environ. Microbiol.*  
378 *Reports* **8**, 461–469, DOI: 10.1111/1758-2229.12390 (2016).

- 379 14. Butcher, R. W. Contributions to our knowledge of the smaller marine algae. *J. Mar. Biol. Assoc. United Kingd.* **31**, 175, DOI: 10.1017/S0025315400003751 (1952).
- 380  
381 15. Simon, N. et al. Revision of the genus *Micromonas* Manton et Parke (Chlorophyta, Mamiel-  
382 lophyceae), of the species *M. pusilla* (Butcher) Manton et Parke, of the species *M. commoda*  
383 van Baren, Bachy et Worden and description of two new species, based on the genetic and  
384 phenotypic characterization of cultured isolates. *Protist* **168**, 612–635 (2017).
- 385 16. Desikachary, T. V. Notes on Volvocales. I. *Curr. Sci.* **41**, 445 (1972).
- 386 17. Manton, I. & Parke, M. Further observations on small green flagellates with special reference  
387 to possible relatives of *Chromulina pusilla* Butcher. *J. Mar. Biol. Assoc. U.K.* **39**, 275–298  
388 (1960).
- 389 18. Marchant, H. J., Buck, K. R., Garrison, D. L. & Thomsen, H. A. *Mantoniella* in Antarctic  
390 waters including the description of *M. antarctica* sp. nov. (Prasinophyceae). *J. Phycol.* **25**,  
391 167–174, DOI: 10.1111/j.0022-3646.1989.00167.x (1989).
- 392 19. Guillou, L. et al. Diversity of picoplanktonic prasinophytes assessed by direct nuclear SSU  
393 rDNA sequencing of environmental samples and novel isolates retrieved from oceanic and  
394 coastal marine ecosystems. *Protist* **155**, 193–214, DOI: 10.1078/143446104774199592 (2004).
- 395 20. Chrétiennot-Dinet, M.-J. et al. A new marine picoeucaryote: *Ostreococcus tauri* gen.  
396 et sp. nov. (Chlorophyta, Prasinophyceae). *Phycologia* **34**, 285–292, DOI: 10.2216/  
397 i0031-8884-34-4-285.1 (1995).
- 398 21. Subirana, L. et al. Morphology, genome plasticity, and phylogeny in the genus *Ostreococcus*  
399 reveal a cryptic species, *O. mediterraneus* sp. nov. (Mamiellales, Mamiellophyceae). *Protist*  
400 **164**, 643–659, DOI: 10.1016/j.protis.2013.06.002 (2013).
- 401 22. Rodríguez, F. et al. Ecotype diversity in the marine picoeukaryote *Ostreococcus* (Chloro-  
402 phyta, Prasinophyceae). *Environ. Microbiol.* **7**, 853–859 (2005).
- 403 23. Eikrem, W. & Throndsen, J. The ultrastructure of *Bathycoccus* gen. nov. and *B. prasi-*  
404 *nos* sp. nov., a non-motile picoplanktonic alga (Chlorophyta, Prasinophyceae) from the

- 405 Mediterranean and Atlantic. *Phycologia* **29**, 344–350, DOI: 10.2216/i0031-8884-29-3-344.1  
406 (1990).
- 407 **24.** Vannier, T. *et al.* Survey of the green picoalga *Bathycoccus* genomes in the global ocean.  
408 *Sci. reports* **6**, 37900, DOI: 10.1038/srep37900 (2016).
- 409 **25.** Limardo, A. J. *et al.* Quantitative biogeography of picoprasinophytes establishes ecotype  
410 distributions and significant contributions to marine phytoplankton. *Environ. Microbiol.* **19**,  
411 3219–3234, DOI: 10.1111/1462-2920.13812 (2017).
- 412 **26.** Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-  
413 supported software for describing and comparing microbial communities. *Appl. environmental  
414 microbiology* **75**, 7537–41 (2009).
- 415 **27.** Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned  
416 ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188–7196,  
417 DOI: 10.1093/nar/gkm864 (2007).
- 418 **28.** Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves  
419 sensitivity and speed of chimera detection. *Bioinforma. (Oxford, England)* **27**, 2194–200  
420 (2011).
- 421 **29.** Callahan, B. J., McMurdie, P. J. & Holmes, S. P. Exact sequence variants should replace  
422 operational taxonomic units in marker-gene data analysis. *The ISME J.* **11**, 2639–2643,  
423 DOI: 10.1038/ismej.2017.119 (2017).
- 424 **30.** Guillou, L. *et al.* The Protist Ribosomal Reference database (PR2): A catalog of unicellular  
425 eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* **41**,  
426 597–604, DOI: 10.1093/nar/gks1160 (2013).
- 427 **31.** Katoh, K. MAFFT: a novel method for rapid multiple sequence alignment based on fast  
428 Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066, DOI: 10.1093/nar/gkf436 (2002).

- 429 **32.** Price, M. N., Dehal, P. S. & Arkin, A. P. Fasttree: Computing large minimum evolution  
430 trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650, DOI:  
431 10.1093/molbev/msp077 (2009).
- 432 **33.** Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees  
433 for large alignments. *PLoS one* **5**, e9490, DOI: 10.1371/journal.pone.0009490 (2010).
- 434 **34.** Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform  
435 for the organization and analysis of sequence data. *Bioinforma. (Oxford, England)* **28**,  
436 1647–9, DOI: 10.1093/bioinformatics/bts199 (2012).
- 437 **35.** Ronquist, F. *et al.* MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice  
438 Across a Large Model Space. *Syst. Biol.* **61**, 539–542, DOI: 10.1093/sysbio/sys029 (2012).
- 439 **36.** Groisillier, A., Massana, R., Valentin, K., Vaulot, D. & Guillou, L. Genetic diversity and  
440 habitats of two enigmatic marine alveolate lineages. *Aquatic Microb. Ecol.* **42**, 277–291,  
441 DOI: 10.3354/ame042277 (2006).
- 442 **37.** Guillou, L. *et al.* Widespread occurrence and genetic diversity of marine parasitoids belong-  
443 ing to Syndiniales (Alveolata). *Environ. Microbiol.* **10**, 3349–3365, DOI: 10.1111/j.1462-2920.  
444 2008.01731.x (2008).
- 445 **38.** Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in  
446 multidimensional genomic data. *Bioinformatics* **32**, 2847–2849, DOI: 10.1093/bioinformatics/  
447 btw313 (2016).
- 448 **39.** Demir-Hilton, E. *et al.* Global distribution patterns of distinct clades of the photosynthetic  
449 picoeukaryote *Ostreococcus*. *The ISME journal* **5**, 1095–1107, DOI: 10.1038/ismej.2010.209  
450 (2011).
- 451 **40.** Clayton, S., Lin, Y.-C., Follows, M. J. & Worden, A. Z. Co-existence of distinct *Ostreococcus*  
452 ecotypes at an oceanic front. *Limnol. Oceanogr.* **62**, 75–88, DOI: 10.1002/lno.10373 (2017).
- 453 **41.** Slapeta, J. *et al.* Global dispersal and ancient cryptic species in the smallest marine  
454 eukaryotes. *Mol. Biol. Evol.* **23**, 23–29, DOI: 10.1093/molbev/msj001 (2006).

- 455 42. Worden, A. Picoeukaryote diversity in coastal waters of the Pacific Ocean. *Aquatic Microb. Ecol.* **43**, 165–175, DOI: 10.3354/ame043165 (2006).
- 457 43. Lin, Y.-C. *et al.* Community composition of photosynthetic picoeukaryotes in a subtropical  
458 coastal ecosystem, with particular emphasis on *Micromonas*. *J. Eukaryot. Microbiol.* **64**,  
459 349–359, DOI: 10.1111/jeu.12370 (2017).
- 460 44. Acosta, F., Ngugi, D. K. & Stingl, U. Diversity of picoeukaryotes at an oligotrophic site off the  
461 Northeastern Red Sea Coast. *Aquatic Biosyst.* **9**, 1:16, DOI: 10.1186/2046-9063-9-16 (2013).
- 462 45. Wu, W., Huang, B., Liao, Y. & Sun, P. Picoeukaryotic diversity and distribution in the  
463 subtropical-tropical South China Sea. *FEMS Microbiol. Ecol.* **89**, 563–579, DOI: 10.1111/  
464 1574-6941.12357 (2014).
- 465 46. Van Baren, M. J. *et al.* Evidence-based green algal genomics reveals marine diversity and an-  
466cestral characteristics of land plants. *BMC Genomics* **17**, 22, DOI: 10.1186/s12864-016-2585-6  
467 (2016).
- 468 47. Worden, A. Z. *et al.* Green Evolution and Dynamic Adaptations Revealed by Genomes of  
469 the Marine Picoeukaryotes *Micromonas*. *Science* **324** (2009).
- 470 48. Lovejoy, C. *et al.* Distribution, phylogeny, and growth of cold-adapted picoprasinophytes in  
471 arctic seas. *J. Phycol.* **43**, 78–89, DOI: 10.1111/j.1529-8817.2006.00310.x (2007).
- 472 49. Balzano, S., Marie, D., Gourvil, P. & Vault, D. Composition of the summer photosynthetic  
473 pico and nanoplankton communities in the Beaufort Sea assessed by T-RFLP and sequences  
474 of the 18S rRNA gene from flow cytometry sorted samples. *The ISME J.* **6**, 1480–1498,  
475 DOI: 10.1038/ismej.2011.213 (2012).
- 476 50. Simmons, M. P. *et al.* Intron Invasions Trace Algal Speciation and Reveal Nearly Identical  
477 Arctic and Antarctic *Micromonas* Populations. *Mol. Biol. Evol.* **32**, 2219–2235, DOI: 10.1093/  
478 molbev/msv122 (2015).
- 479 51. Belevich, T. A. *et al.* Photosynthetic Picoeukaryotes in the Land-Fast Ice of the White Sea,  
480 Russia. *Microb. Ecol.* 1–16, DOI: 10.1007/s00248-017-1076-x (2017).

- 481 52. Vaulot, D. *et al.* Metagenomes of the Picoalga *Bathycoccus* from the Chile coastal upwelling.  
482 *PLoS ONE* **7**, e39648, DOI: 10.1371/journal.pone.0039648 (2012).
- 483 53. Simmons, M. P. *et al.* Abundance and biogeography of picoprasinophyte ecotypes and other  
484 phytoplankton in the Eastern North Pacific Ocean. *Appl. Environ. Microbiol.* **82**, 1693–705,  
485 DOI: 10.1128/AEM.02730-15 (2016).
- 486 54. Throndsen, J. & Kristiansen, S. *Micromonas pusilla* (Prasinophyceae) as part of pico-and  
487 nanoplankton communities of the Barents Sea. *Polar Res.* **10**, 201–208, DOI: 10.3402/polar.  
488 v10i1.6739 (1991).
- 489 55. Bachy, C., López-García, P., Vereshchaka, A. & Moreira, D. Diversity and Vertical Distribution  
490 of Microbial Eukaryotes in the Snow, Sea Ice and Seawater Near the North Pole at the End  
491 of the Polar Night. *Front. Microbiol.* **2**, 106, DOI: 10.3389/fmicb.2011.00106 (2011).
- 492 56. Majaneva, M., Rintala, J.-M. M., Piisilä, M., Fewer, D. P. & Blomster, J. Comparison of  
493 wintertime eukaryotic community from sea ice and open water in the Baltic Sea, based on  
494 sequencing of the 18S rRNA gene. *Polar Biol.* **35**, 875–889, DOI: 10.1007/s00300-011-1132-9  
495 (2012).
- 496 57. Kuwata, A. *et al.* Bolidophyceae, a sister picoplanktonic group of diatoms - a review. *Front.*  
497 *Mar. Sci.* **in press** (2018).
- 498 58. Zhu, F., Massana, R., Not, F., Marie, D. & Vaulot, D. Mapping of picoeucaryotes in marine  
499 ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol. Ecol.* **52**, 79–92,  
500 DOI: 10.1016/j.femsec.2004.10.006 (2005).
- 501 59. Derelle, E. *et al.* Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri*  
502 unveils many unique features. *Proc. Natl. Acad. Sci. United States Am.* **103**, 11647–52, DOI:  
503 10.1073/pnas.0604795103 (2006).
- 504 60. Moreau, H. *et al.* Gene functionalities and genome structure in *Bathycoccus prasinos* reflect  
505 cellular specializations at the base of the green lineage. *Genome Biol.* **13**, R74 (2012).

- 506 61. Cloern, J. E. & Jassby, A. D. Patterns and Scales of Phytoplankton Variability in Estuar-  
507 ine–Coastal Ecosystems. *Estuaries Coasts* **33**, 230–241, DOI: 10.1007/s12237-009-9195-3  
508 (2010).
- 509 62. Treusch, A. H. *et al.* Phytoplankton distribution patterns in the northwestern Sargasso Sea  
510 revealed by small subunit rRNA genes from plastids. *The ISME journal* **6**, 481–92, DOI:  
511 10.1038/ismej.2011.117 (2012).
- 512 63. Lambert, S. *et al.* Rhythmicity of coastal marine picoeukaryotes, bacteria and archaea  
513 despite irregular environmental perturbations. *The ISME J. in press*, DOI: 10.1038/  
514 s41396-018-0281-z (2018).
- 515 64. Piredda, R. *et al.* Diversity and temporal patterns of planktonic protist assemblages at a  
516 Mediterranean Long Term Ecological Research site. *FEMS Microbiol. Ecol.* **93**, fiw200, DOI:  
517 10.1093/femsec/fiw200 (2017).
- 518 65. Foulon, E. *et al.* Ecological niche partitioning in the picoplanktonic green alga *Micromonas*  
519 *pusilla*: Evidence from environmental surveys using phylogenetic probes. *Environ. Microbiol.*  
520 **10**, 2433–2443, DOI: 10.1111/j.1462-2920.2008.01673.x (2008).
- 521 66. Ichinomiya, M. *et al.* Diversity and oceanic distribution of the Parmales (Bolidophyceae),  
522 a picoplanktonic group closely related to diatoms. *The ISME J.* **10**, 2419–2434, DOI:  
523 10.1038/ismej.2016.38 (2016).
- 524 67. Demory, D. *et al.* Picoeukaryotes of the *Micromonas* genus: sentinels of a warming ocean.  
525 *The ISME J.* DOI: 10.1038/s41396-018-0248-0 (2018).
- 526 68. Krehenwinkel, H. *et al.* Nanopore sequencing of long ribosomal DNA amplicons enables  
527 portable and simple biodiversity assessments with high phylogenetic resolution across broad  
528 taxonomic scale. *bioRxiv* DOI: 10.1101/358572 (2018).

529 **Acknowledgements**

530 Financial support for this work was provided by the European Union projects MicroB3 (UE-  
531 contract-287589), the ANR PhytoPol (ANR-15-CE02-0007) and TaxMArc (Research Council

532 of Norway, 268286/E40). MT was supported by a PhD fellowship from the Université Pierre et  
533 Marie Curie and the Région Bretagne (ARED GreenPhy). We would like to thank the Ocean  
534 Sampling Day consortium for providing sequence data and the ABIMS platform in Roscoff for  
535 access to bioinformatics resources.

## 536 **Author contributions statement**

537 D.V. and M.T. conceived the study. D.V. produced some of the figures and edited the paper.  
538 M.T. analyzed the data, produced some of the figures and wrote the initial draft of the paper. All  
539 authors reviewed the manuscript.

## 540 **Additional information**

### 541 **Data availability**

542 Scripts for mothur, Mamiellophyceae OTU sequences, alignments, assignation and abundance for  
543 the LGC and LW datasets are provided as supplementary files on Figshare: <https://figshare.com/s/6e074ed24>  
544 R processing script is detailed at [https://vaulot.github.io/papers/OSD\\_Mamiello.html](https://vaulot.github.io/papers/OSD_Mamiello.html) and provided  
545 at [https://github.com/vaulot/Paper-2018\\_Tragin\\_Mamiellophyceae-R-scripts](https://github.com/vaulot/Paper-2018_Tragin_Mamiellophyceae-R-scripts)

### 546 **Competing interests**

547 The authors declare no competing interests.

## 548 List of Tables

- 549      Table 1     Major Mamiellophyceae ASVs from the LGC and LW datasets: assignation, total  
550                  abundance and representative sequences name.  
551      Table 2     Summary of the coastal distribution of Mamiellophyceae species and clades. The  
552                  column indicate the number of samples where the species/clade represented more  
553                  than 1% of Mamiellophyceae

## 554 List of Figures

- 555      Fig. 1 Treemap of the Mamiellophyceae genera and clade contribution for the OSD2014 LGC  
556                  dataset. Only selected ASVs with more than 200 reads and stations with more than 100  
557                  reads were taken into account.  
558      Fig. 2 Phylogenetic diversity within the genus *Ostreococcus*. A. Phylogenetic tree of 26  
559                  *Ostreococcus* V4 regions of the 18S rRNA gene. The tree was rooted with *B. prasinos*  
560                  and only ML bootstrap values higher than 70 % are represented and Bayesian posterior  
561                  probabilities are in bold. Reference sequences from GenBank are in bold, ASVs from  
562                  the LW dataset (starting with M) are in grey and ASVs from LGC in black (starting  
563                  with H). Numbers in brackets correspond to the number of reads for each ASV. Only  
564                  ASVs represented by more than 200 reads were taken into account. B. Alignment of 26  
565                  *Ostreococcus* V4 regions, the alignment is 341 bp long, but only the main signatures are  
566                  shown (between positions 20 and 140 of the original alignment).  
567      Fig. 3 Distribution of the major *Ostreococcus* ASVs for OSD2014 (LGC). The three major *O.  
568                  tauri* ASVs have been pooled together. Circle surface corresponds to the percentage of  
569                  ASV reads relative to the total number of Mamiellophyceae reads. Samples for which  
570                  the ASV contribution was lower than 1 % are represented by blue crosses. A zoom on  
571                  European waters is provided in Fig. S3.  
572      Fig. 4 Phylogenetic diversity within the genus *Micromonas*. A. Phylogenetic tree of 39 *Mi-  
573                  cromonas* V4 regions of the 18S rRNA gene (Fasttree). The tree was rooted with  
574                  Mamiellales (RCC391, AY425321 and *Mamiella gilva*, FN562450). Legend as Fig. 2.  
575                  B. Alignment of 39 *Micromonas* V4 regions, the alignment is 327 bp long, but only the  
576                  main signatures are shown (between positions 50 and 150 of the original alignment).  
577      Fig. 5 Distribution of the major *Micromonas* ASVs for OSD2014 (LGC). Legend as Fig. 3. A  
578                  zoom on European waters is provided in Fig. S5.

579 Fig. 6 Distribution of the major *Bathycoccus* ASV for OSD2014 (LGC). Legend as Fig. 3. A  
580 zoom on European waters is provided in Fig. S6.

581 Fig. 7 Phylogenetic diversity within the genus *Mantoniella*. A. Phylogenetic tree of 19 *Mantoniella*  
582 V4 regions of the 18S rRNA gene (Fasttree). The tree was rooted with the 3  
583 *Micromonas* sequences (AB183589, HM191693, JF698749) and only bootstrap values  
584 higher than 70 % are represented. Legend as Fig. 2. B. Alignment of 19 *Mantoniella* V4  
585 regions, the alignment was 368 bp long, but only the main signatures are shown (around  
586 the 20th and 141th position of the original alignment).

587 Fig. 8 Distribution of the major *Mantoniella* ASVs for OSD2014 (LGC). Legend as Fig. 3. A  
588 zoom on European waters is provided in Fig. S7.

**Table 1.** Major Mamiellophyceae ASVs from the LGC and LW datasets: assignation, total abundance and representative sequences name.

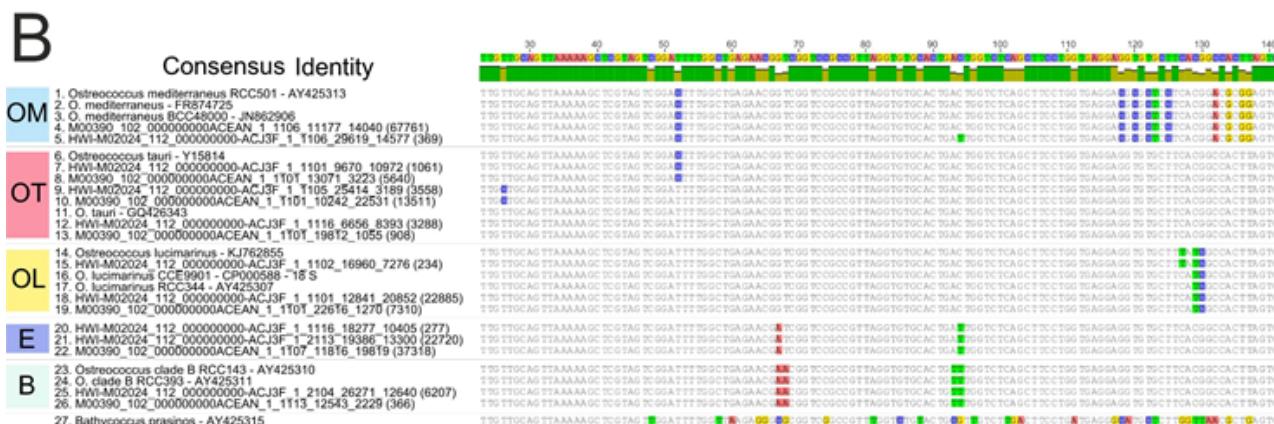
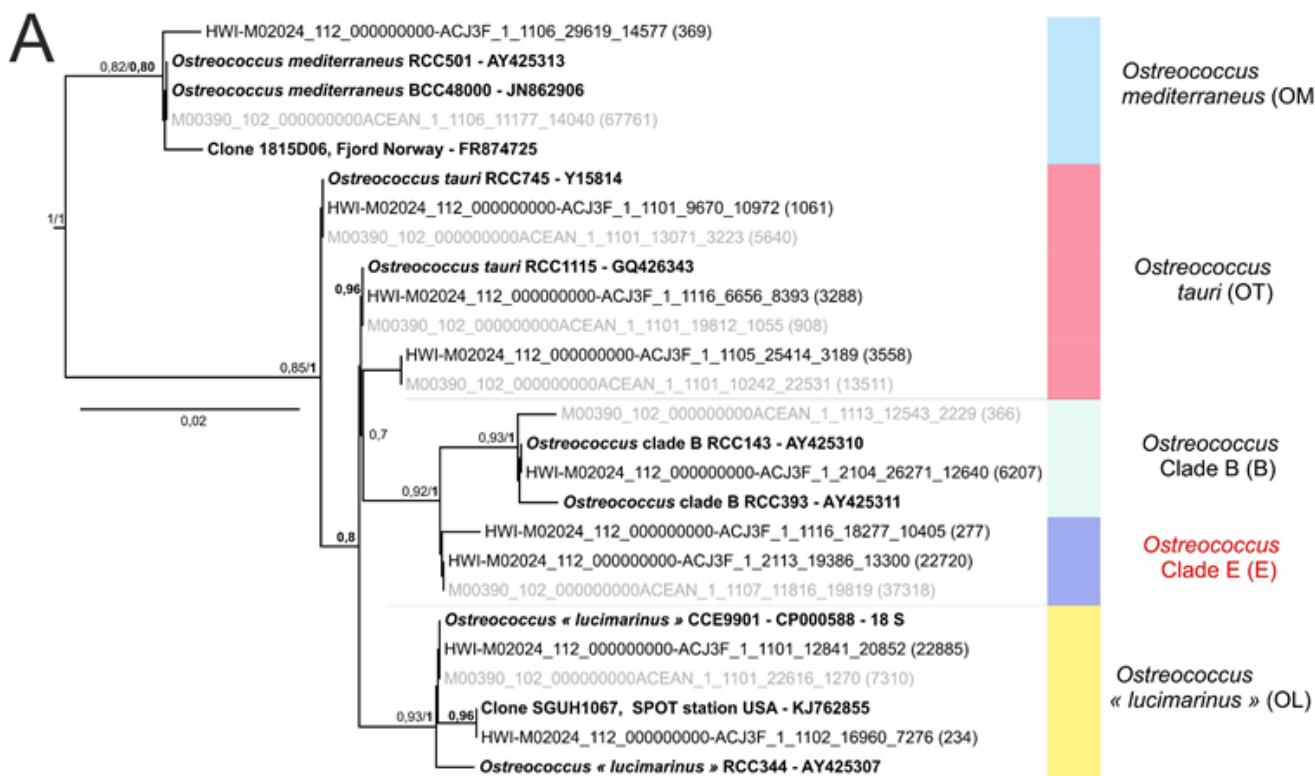
| ASV code      | Species                           | Reads | Representative_Sequence                           |
|---------------|-----------------------------------|-------|---------------------------------------------------|
| ASV_LGC_00001 | <i>Micromonas bravo</i> II        | 28810 | HWI-M02024_112_000000000-ACJ3F_1_2114_15858_18370 |
| ASV_LGC_00002 | <i>Ostreococcus lucimarinus</i>   | 22885 | HWI-M02024_112_000000000-ACJ3F_1_1101_12841_20852 |
| ASV_LGC_00003 | <i>Ostreococcus</i> clade E       | 22720 | HWI-M02024_112_000000000-ACJ3F_1_2113_19386_13300 |
| ASV_LGC_00004 | <i>Bathycoccus prasinos</i>       | 16750 | HWIM02024_112_000000000ACJ3F_1_1110_9700_15401    |
| ASV_LGC_00005 | <i>Micromonas commoda</i> C       | 10787 | HWI-M02024_112_000000000-ACJ3F_1_1105_17215_10467 |
| ASV_LGC_00006 | <i>Ostreococcus</i> clade B       | 6207  | HWI-M02024_112_000000000-ACJ3F_1_2104_26271_12640 |
| ASV_LGC_00007 | <i>Micromonas commoda</i> AB      | 5465  | HWI-M02024_112_000000000-ACJ3F_1_2109_7000_7275   |
| ASV_LGC_00008 | <i>Micromonas bravo</i> I         | 5459  | HWI-M02024_112_000000000-ACJ3F_1_1101_14140_2642  |
| ASV_LGC_00009 | <i>Micromonas</i> B warm I        | 5251  | HWI-M02024_112_000000000-ACJ3F_1_1106_22514_11146 |
| ASV_LGC_00010 | <i>Mantoniella squamata</i>       | 4500  | HWI-M02024_112_000000000-ACJ3F_1_1103_8489_15072  |
| ASV_LGC_00011 | <i>Mantoniella</i> clade B        | 4070  | HWI-M02024_112_000000000-ACJ3F_1_1101_16376_19283 |
| ASV_LGC_00012 | <i>Ostreococcus tauri</i> _3189   | 3558  | HWI-M02024_112_000000000-ACJ3F_1_1105_25414_3189  |
| ASV_LGC_00013 | <i>Ostreococcus tauri</i> _8393   | 3288  | HWI-M02024_112_000000000-ACJ3F_1_1116_6656_8393   |
| ASV_LGC_00014 | <i>Micromonas polaris</i>         | 2200  | HWI-M02024_112_000000000-ACJ3F_1_1104_18898_13160 |
| ASV_LGC_00015 | <i>Micromonas</i> Clade B..4      | 1616  | HWI-M02024_112_000000000-ACJ3F_1_1115_22402_20688 |
| ASV_LGC_00016 | <i>Micromonas pusilla</i>         | 1259  | HWI-M02024_112_000000000-ACJ3F_1_1102_17618_11009 |
| ASV_LGC_00017 | <i>Micromonas</i> B subarctic     | 1096  | HWI-M02024_112_000000000-ACJ3F_1_1106_10729_9219  |
| ASV_LGC_00018 | <i>Ostreococcus tauri</i> _10972  | 1061  | HWI-M02024_112_000000000-ACJ3F_1_1101_9670_10972  |
| ASV_LGC_00019 | <i>Micromonas</i> B warm II       | 1045  | HWI-M02024_112_000000000-ACJ3F_1_1103_25501_7171  |
| ASV_LGC_00022 | <i>Ostreococcus mediterraneus</i> | 369   | HWI-M02024_112_000000000-ACJ3F_1_1106_29619_14577 |
| ASV_LGC_00023 | <i>Ostreococcus</i> clade E       | 277   | HWI-M02024_112_000000000-ACJ3F_1_1116_18277_10405 |
| ASV_LGC_00024 | <i>Ostreococcus mediterraneus</i> | 248   | HWI-M02024_112_000000000-ACJ3F_1_1103_20917_24371 |
| ASV_LGC_00025 | <i>Ostreococcus lucimarinus</i>   | 234   | HWI-M02024_112_000000000-ACJ3F_1_1102_16960_7276  |
| ASV_LW_00001  | <i>Ostreococcus mediterraneus</i> | 67761 | M00390_102_000000000-ACEAN_1_1106_11177_14040     |
| ASV_LW_00002  | <i>Ostreococcus</i> clade E       | 37318 | M00390_102_000000000-ACEAN_1_1107_11816_19819     |
| ASV_LW_00003  | <i>Micromonas polaris</i>         | 35194 | M00390_102_000000000-ACEAN_1_2105_19731_21730     |
| ASV_LW_00004  | <i>Mantoniella squamata</i>       | 25692 | M00390_102_000000000-ACEAN_1_1107_20899_6104      |
| ASV_LW_00005  | <i>Bathycoccus prasinos</i>       | 20779 | M00390_102_000000000-ACEAN_1_1101_8267_5862       |
| ASV_LW_00006  | <i>Micromonas bravo</i> II        | 20153 | M00390_102_000000000-ACEAN_1_1101_13226_2947      |
| ASV_LW_00007  | <i>Micromonas commoda</i> C       | 14869 | M00390_102_000000000-ACEAN_1_1101_19723_8016      |
| ASV_LW_00008  | <i>Ostreococcus tauri</i> _3189   | 13511 | M00390_102_000000000-ACEAN_1_1101_10242_22531     |
| ASV_LW_00009  | <i>Ostreococcus lucimarinus</i>   | 7310  | M00390_102_000000000-ACEAN_1_1101_22616_1270      |
| ASV_LW_00010  | <i>Ostreococcus tauri</i> _10972  | 5640  | M00390_102_000000000-ACEAN_1_1101_13071_3223      |
| ASV_LW_00011  | <i>Micromonas pusilla</i>         | 2806  | M00390_102_000000000-ACEAN_1_1101_17876_4103      |
| ASV_LW_00012  | <i>Micromonas commoda</i> AB      | 2610  | M00390_102_000000000-ACEAN_1_1101_13887_6615      |
| ASV_LW_00013  | <i>Micromonas bravo</i> I         | 2357  | M00390_102_000000000-ACEAN_1_1108_21831_3480      |
| ASV_LW_00014  | <i>Micromonas</i> B subarctic     | 1862  | M00390_102_000000000-ACEAN_1_1112_28497_17492     |
| ASV_LW_00015  | <i>Micromonas</i> Clade B..4      | 1078  | M00390_102_000000000-ACEAN_1_1101_13697_13215     |
| ASV_LW_00016  | <i>Mantoniella</i> clade B        | 935   | M00390_102_000000000-ACEAN_1_1101_10638_9605      |
| ASV_LW_00017  | <i>Ostreococcus tauri</i> _8393   | 908   | M00390_102_000000000-ACEAN_1_1101_19812_1055      |
| ASV_LW_00018  | <i>Mantoniella</i> clade A        | 841   | M00390_102_000000000-ACEAN_1_1115_21286_22319     |
| ASV_LW_00020  | <i>Micromonas polaris</i>         | 668   | M00390_102_000000000-ACEAN_1_1101_14320_1948      |
| ASV_LW_00021  | <i>Micromonas</i> B warm II       | 661   | M00390_102_000000000-ACEAN_1_1101_13671_3127      |
| ASV_LW_00022  | <i>Micromonas commoda</i> C       | 462   | M00390_102_000000000-ACEAN_1_1104_4249_10680      |
| ASV_LW_00024  | <i>Ostreococcus</i> clade B       | 366   | M00390_102_000000000-ACEAN_1_1113_12543_2229      |
| ASV_LW_00027  | <i>Micromonas bravo</i> I         | 205   | M00390_102_000000000-ACEAN_1_1103_16789_17810     |

**Table 2.** Summary of the coastal distribution of Mamiellophyceae species and clades. The column indicate the number of samples where the species/clade represented more than 1% of Mamiellophyceae

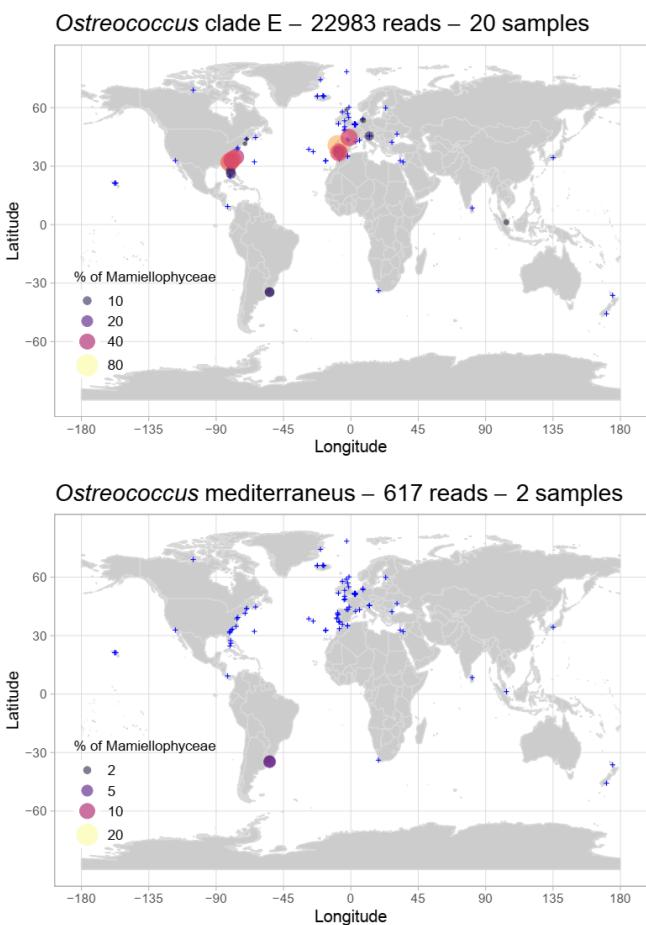
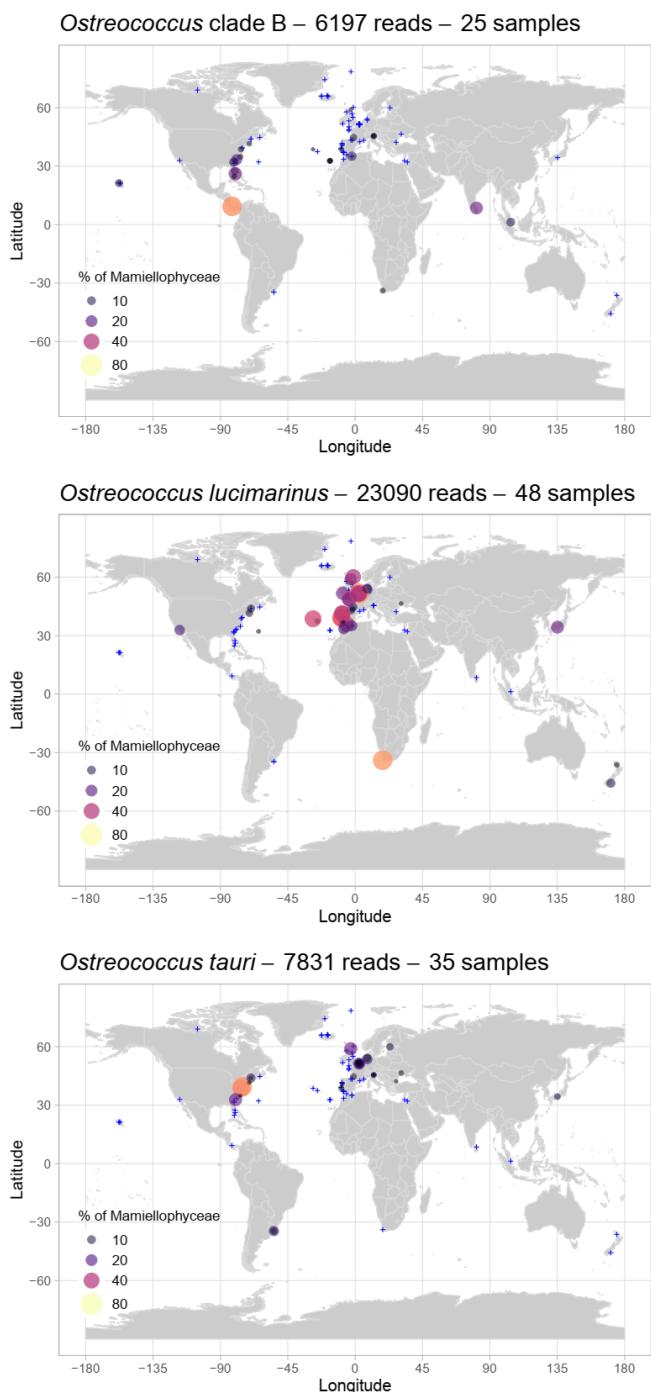
| Genus               | Species              | Clade       | Samples | Polar | Sub-polar | Temperate | Tropical | Mediterranean Sea | Remark                        |
|---------------------|----------------------|-------------|---------|-------|-----------|-----------|----------|-------------------|-------------------------------|
| <i>Bathycoccus</i>  | <i>prasinos</i>      |             | 72      |       |           |           |          |                   |                               |
| <i>Mantoniella</i>  | <i>squamata</i>      |             | 11      |       |           |           |          |                   | Hawaii                        |
|                     | <i>sp.</i>           | B           | 16      |       |           |           |          |                   | Hawaii                        |
| <i>Micromonas</i>   | <i>pusilla</i>       |             | 18      |       |           |           |          |                   | North Europe                  |
|                     | <i>bravo</i>         | I           | 38      |       |           |           |          |                   |                               |
|                     | <i>bravo</i>         | II          | 61      |       |           |           |          |                   |                               |
|                     | <i>sp.</i>           | B warm I    | 9       |       |           |           |          |                   |                               |
|                     | <i>sp.</i>           | B warm II   | 5       |       |           |           |          |                   |                               |
|                     | <i>sp.</i>           | B subarctic | 11      |       |           |           |          |                   |                               |
|                     | <i>sp.</i>           | B_.4        | 15      |       |           |           |          |                   | Below 45 °                    |
|                     | <i>commoda</i>       | AB          | 16      |       |           |           |          |                   |                               |
|                     | <i>commoda</i>       | C           | 57      |       |           |           |          |                   |                               |
|                     | <i>polaris</i>       |             | 4       |       |           |           |          |                   | Baltic Sea                    |
| <i>Ostreococcus</i> | <i>tauri</i>         |             | 35      |       |           |           |          |                   | Adriatic, Black Sea           |
|                     | <i>mediterraneus</i> |             | 2       |       |           |           |          |                   | Only in lagoon                |
|                     | <i>lucimarinus</i>   |             | 48      |       |           |           |          |                   |                               |
|                     | <i>sp.</i>           | E           | 20      |       |           |           |          |                   |                               |
|                     | <i>sp.</i>           | B           | 25      |       |           |           |          |                   | South temperate, Adriatic Sea |



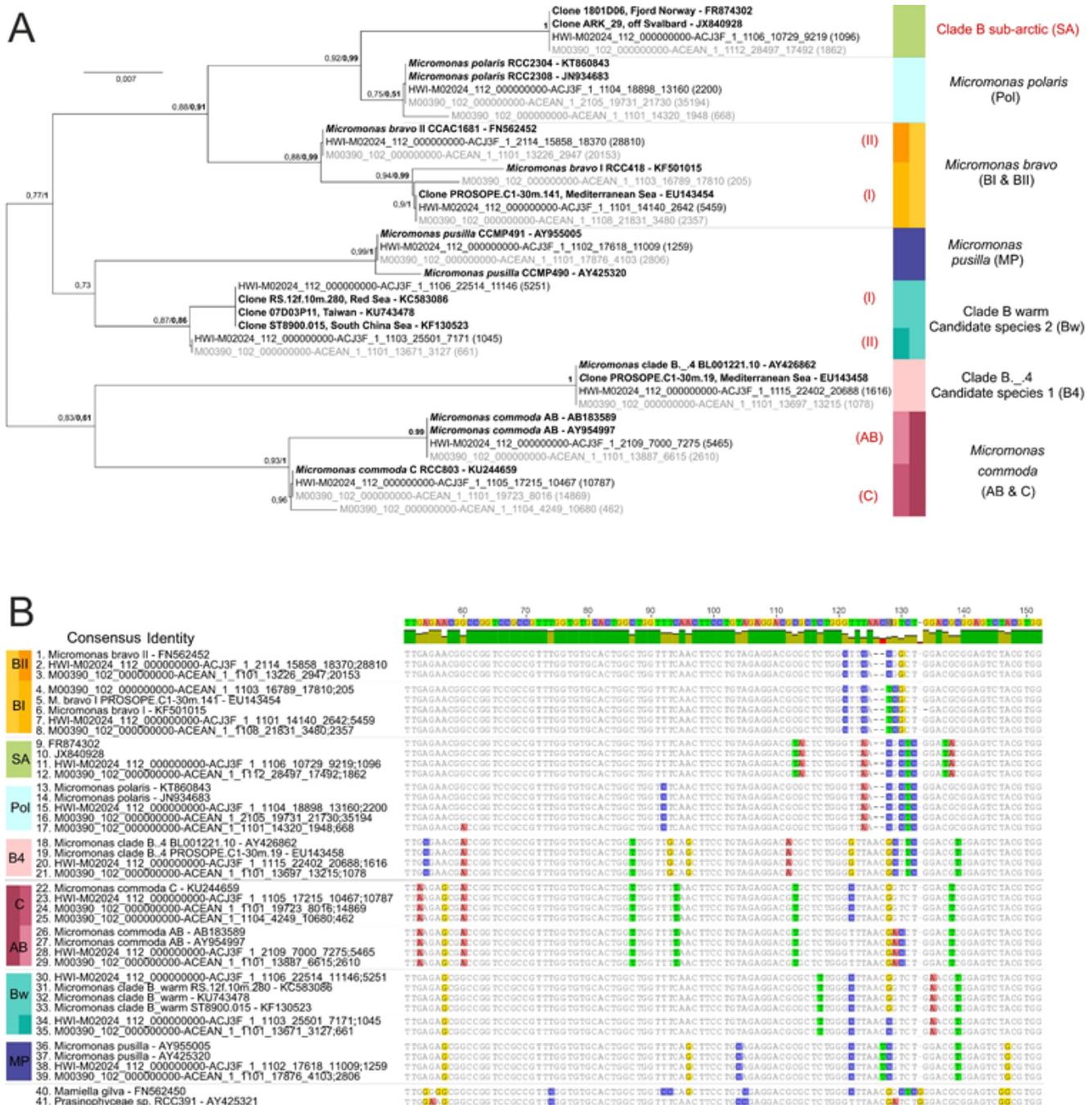
**Figure 1.** Treemap of the Mamiellophyceae genera and clade contribution for the OSD2014 LGC dataset. Only selected ASVs with more than 200 reads and stations with more than 100 reads were taken into account.



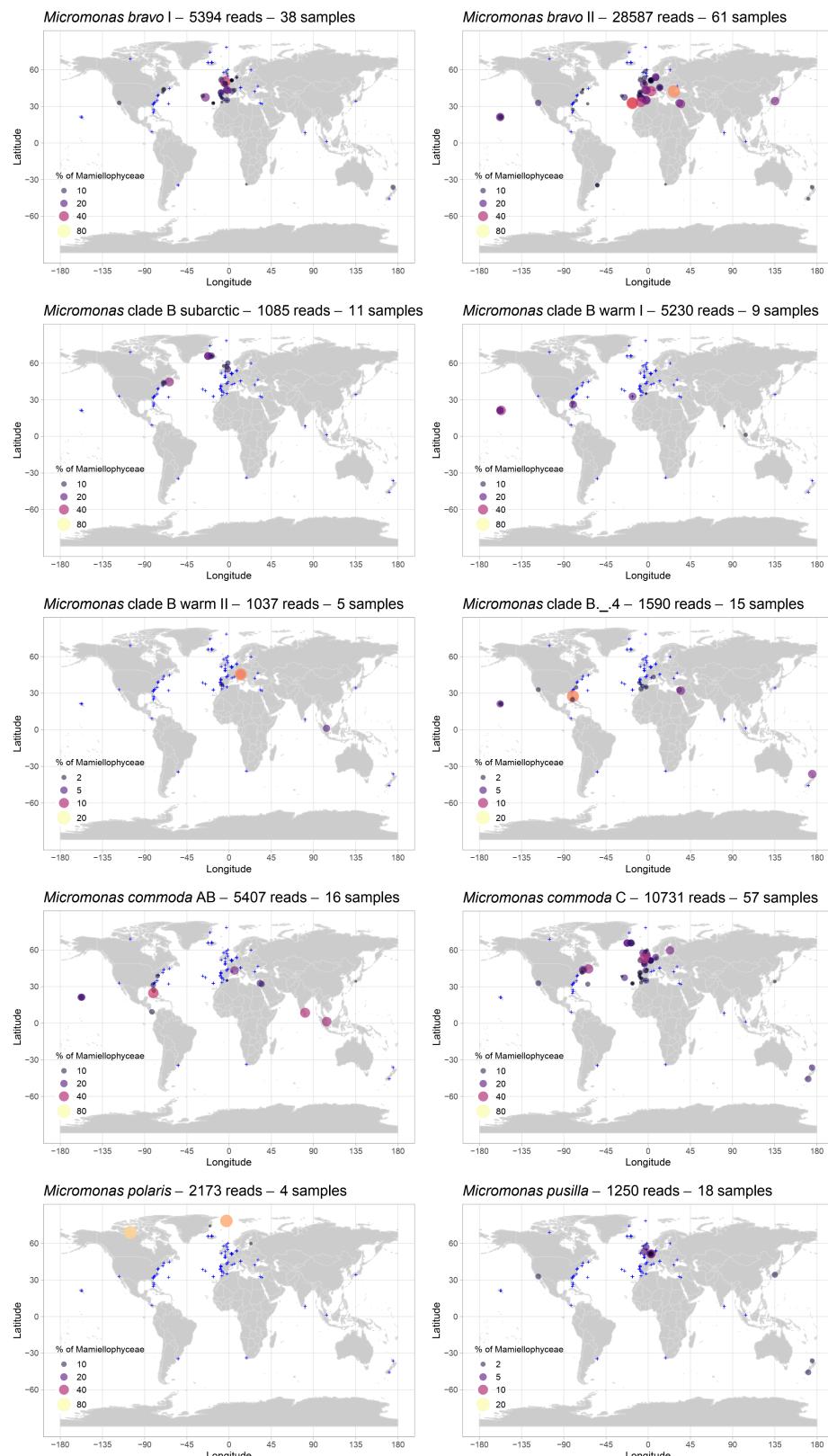
**Figure 2.** Phylogenetic diversity within the genus *Ostreococcus*. A. Phylogenetic tree of 26 *Ostreococcus* V4 regions of the 18S rRNA gene. The tree was rooted with *B. prasinos* and only ML bootstrap values higher than 70 % are represented and Bayesian posterior probabilities are in bold. Reference sequences from GenBank are in bold, ASVs from the LW dataset (starting with M) are in grey and ASVs from LGC in black (starting with H). Numbers in brackets correspond to the number of reads for each ASV. Only ASVs represented by more than 200 reads were taken into account. B. Alignment of 26 *Ostreococcus* V4 regions, the alignment is 341 bp long, but only the main signatures are shown (between positions 20 and 140 of the original alignment).



**Figure 3.** Distribution of the major *Ostreococcus* ASVs for OSD2014 (LGC). The three major *O. tauri* ASVs have been pooled together. Circle surface corresponds to the percentage of ASV reads relative to the total number of Mamiellophyceae reads. Samples for which the ASV contribution was lower than 1 % are represented by blue crosses. A zoom on European waters is provided in Fig. S3.

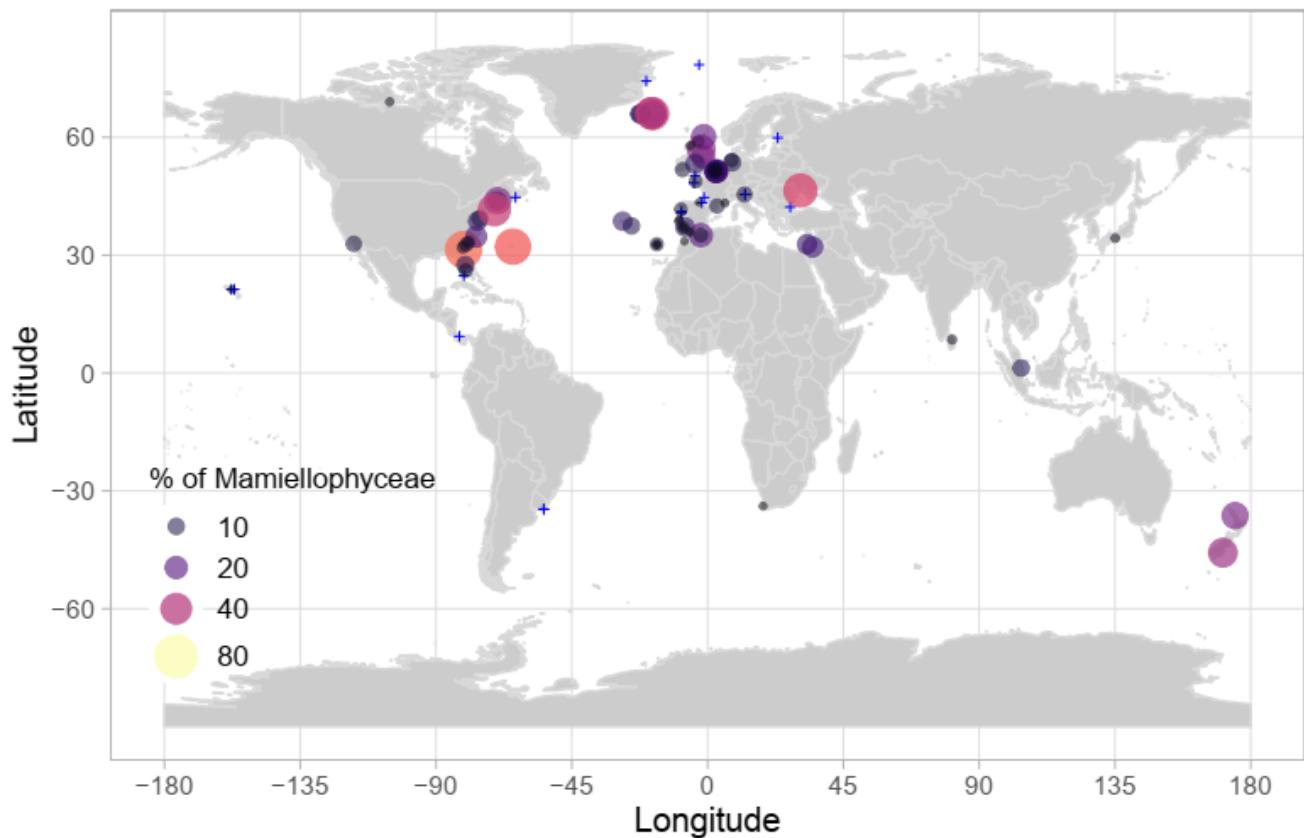


**Figure 4.** Phylogenetic diversity within the genus *Micromonas*. A. Phylogenetic tree of 39 *Micromonas* V4 regions of the 18S rRNA gene (Fastree). The tree was rooted with Mamiellales (RCC391, AY425321 and *Mamiella gilva*, FN562450). Legend as Fig. 2. B. Alignment of 39 *Micromonas* V4 regions, the alignment is 327 bp long, but only the main signatures are shown (between positions 50 and 150 of the original alignment).



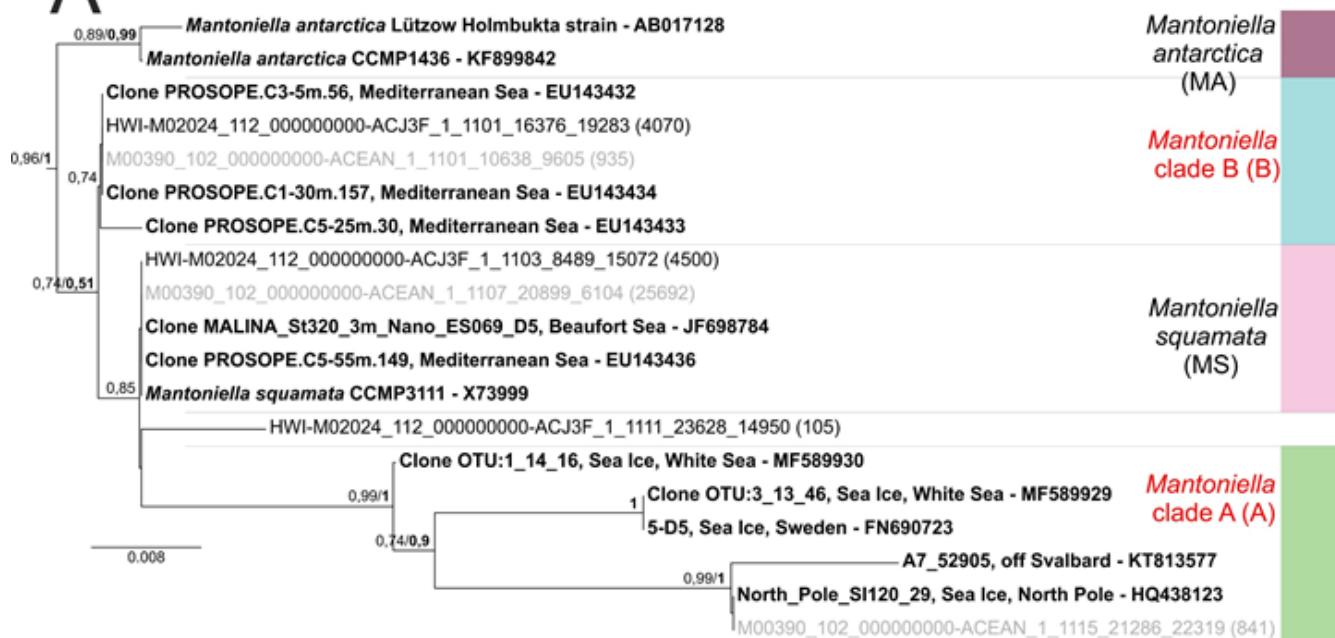
**Figure 5.** Distribution of the major *Micromonas* ASVs for OSD2014 (LGC). Legend as Fig. 3. A zoom on European waters is provided in Fig. S5.

## *Bathycoccus prasinos* – 16651 reads – 72 samples

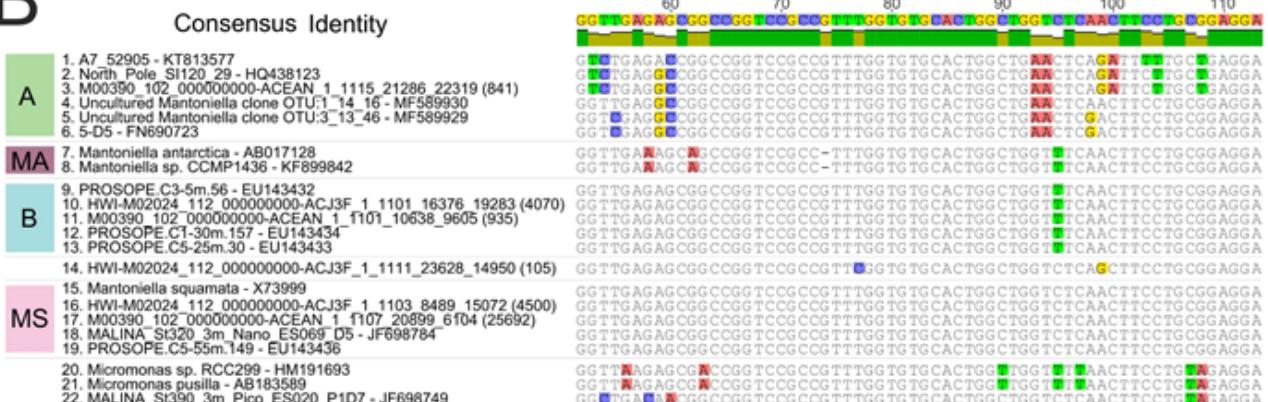


**Figure 6.** Distribution of the major *Bathycoccus* ASV for OSD2014 (LGC). Legend as Fig. 3. A zoom on European waters is provided in Fig. S6.

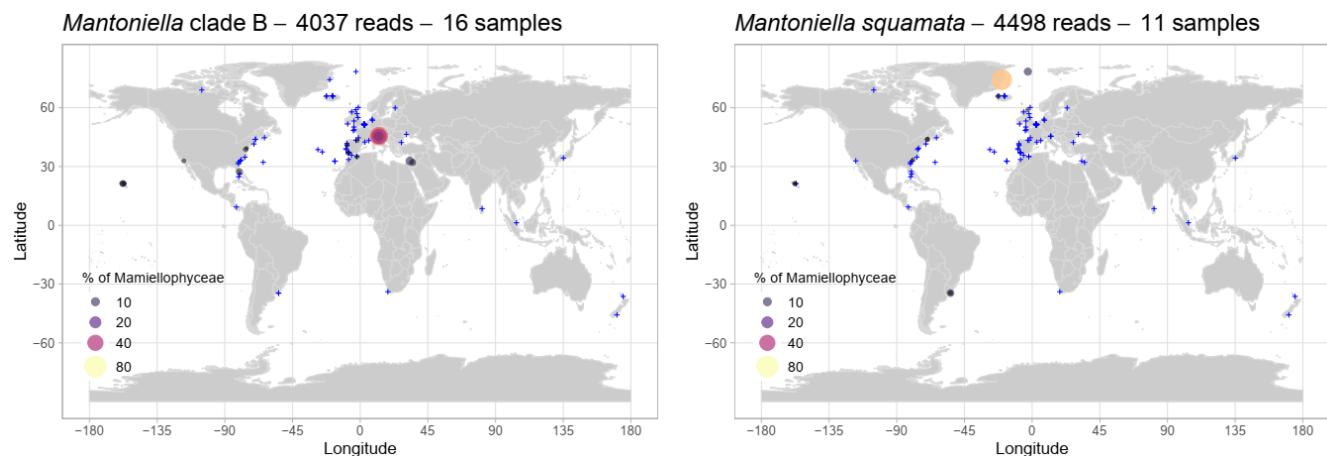
A



B



**Figure 7.** Phylogenetic diversity within the genus *Mantonella*. A. Phylogenetic tree of 19 *Mantonella* V4 regions of the 18S rRNA gene (Fasttree). The tree was rooted with the 3 *Micromonas* sequences (AB183589, HM191693, JF698749) and only bootstrap values higher than 70 % are represented. Legend as Fig. 2. B. Alignment of 19 *Mantonella* V4 regions, the alignment was 368 bp long, but only the main signatures are shown (around the 20th and 141th position of the original alignment).



**Figure 8.** Distribution of the major *Mantiella* ASVs for OSD2014 (LGC). Legend as Fig. 3. A zoom on European waters is provided in Fig. S7.

# **Novel diversity within marine Mamiellophyceae (Chlorophyta) unveiled by metabarcoding**

## **Supplementary material**

**Margot Tragin**

**Daniel Vaulot**

Sorbonne Université, CNRS, UMR 7144

Station Biologique, Place Georges Teissier, 29680 Roscoff, France

Date: October 21, 2018

All supplementary data have been deposited to

Figshare: <https://figshare.com/s/6e074ed24b01d3d1f8ea>

GitHub : [https://github.com/vaulot/Paper-2018\\_Tragin\\_Mamiellophyceae-R-scripts](https://github.com/vaulot/Paper-2018_Tragin_Mamiellophyceae-R-scripts)

1 **Supplementary data**

- 2 Supplementary Data S1: (OSD\_mamiello\_mothur.sh) Mothur script for sequence analysis.
- 3 Supplementary Data S2: (OSD\_Mamiello.html). R script for generating distribution maps and
- 4 heatmaps with output.
- 5 Supplementary Data S3: (Excel file OSD\_Mamiello\_Tables\_Supplementary.xlsx). Description of
- 6 the different sheets:
- 7 • "metadata": list of OSD stations with coordinates and metadata.
- 8 • "samples": list of OSD samples.
- 9 • "otus LGC Mamiello": Mamiellophyceae LGC ASV table with taxonomy and number of
- 10 reads assigned to each ASV per OSD 2014 samples.
- 11 • "otus LW Mamiello": Mamiellophyceae LW ASV table with taxonomy and number of reads
- 12 assigned to each ASV per OSD 2014 samples.
- 13 • "sequences GenBank": Genbank accession numbers for sequences used to build the
- 14 trees.
- 15 • "simil\_Ostreococcus": Matrix of the pairwise percent identity between *Ostreococcus* clades.
- 16 The calculation was done based on the alignment available as Supplementary data.
- 17 • "simil\_Micromonas": idem for *Micromonas*.
- 18 • "simil\_Mantoniella": idem for *Mantoniella*.
- 19 Supplementary Data S4: (OSD\_Mamiello\_ASV\_LGC.fasta) OSD 2014 Mamiellophyceae ASV
- 20 representative sequences (LGC sequencing).
- 21 Supplementary Data S5: (OSD\_Mamiello\_ASV\_LW.fasta) OSD 2014 Mamiellophyceae ASV
- 22 representative sequences (LW sequencing).
- 23 Supplementary Data S6: (OSD\_Mamiello\_ASV\_LGC\_Ostreococcus\_align.fasta) Alignment of
- 24 major LGC and LW *Ostreococcus* ASV with related GenBank sequences from clones and
- 25 cultures.

<sup>26</sup> Supplementary Data S7: (OSD\_Mamiello\_ASV\_LGC\_Micromonas\_align.fasta) Idem for  
<sup>27</sup> *Micromonas*.

<sup>28</sup> Supplementary Data S8: (OSD\_Mamiello\_ASV\_LGC\_Mantoniella\_align.fasta) Idem for  
<sup>29</sup> *Mantoniella*.

## <sup>30</sup> List of supplementary tables

- <sup>31</sup> Table S1 List of OSD stations.
- <sup>32</sup> Table S2 Matrix of the pairwise percent identity between *Ostreococcus* clades. The calcula-  
<sup>33</sup> tion was done based on the alignment available as Supplementary data S6.
- <sup>34</sup> Table S3 Matrix of the pairwise percent identity between *Micromonas* clades. The calculation  
<sup>35</sup> was done based on the alignment available as Supplementary data S7.
- <sup>36</sup> Table S4 Matrix of the pairwise percent identity between *Mantoniella* clades. The calculation  
<sup>37</sup> was done based on the alignment available as Supplementary data S8.

## <sup>38</sup> List of supplementary figures

- <sup>39</sup> Fig. S1 A. Percentage of OSD samples where a given ASV was detected (at least one read).  
<sup>40</sup> Numbers at right of bars correspond to number of samples. B. Idem but for samples  
<sup>41</sup> where the ASV contributed more than 1% of the Mamiellophyceae reads. Stations  
<sup>42</sup> with less than 100 Mamiellophyceae reads were not considered. The three major *O.*  
<sup>43</sup> *tauri* ASVs have been pooled together.
- <sup>44</sup> Fig. S2 Heatmap of the Mamiellophyceae communities in OSD samples. The number  
<sup>45</sup> of read corresponding to the major ASVs were normalized by the number of  
<sup>46</sup> Mamiellophyceae reads per station. Stations with less than 100 Mamiellophyceae  
<sup>47</sup> reads were not considered.

- 48 Fig. S3 Distribution of major *Ostreococcus* ASVs for OSD2014 (LGC) in European wa-  
49 ters. The three major *O. tauri* ASVs have been pooled together. Circle surface  
50 corresponds to the percentage of ASV reads relative to the total number of Mamiel-  
51 lophyceae reads. Samples for which the ASV contribution was lower than 1 % are  
52 represented by blue crosses.
- 53 Fig. S4 Alignments of *Ostreococcus* V4 regions and localization of the qPCR primers  
54 and probes (Demir-Hilton et al. 2011) used to quantify the OI (A) and OII (B)  
55 clades. Sequences with no mismatches to each probe set are in bold. Blue and  
56 green sequences correspond to LGC and LW datasets, respectively.
- 57 Fig. S5 Same as Fig. S3 for *Micromomas*.
- 58 Fig. S6 Same as Fig. S3 for *Bathycoccus*.
- 59 Fig. S7 Same as Fig. S3 for *Mantoniella*.

**Table S1.** List of OSD stations.

| Station | Name                     | Ocean             | Country                             |
|---------|--------------------------|-------------------|-------------------------------------|
| OSD1    | Plymouth - L4            | Atlantic Ocean    | UK                                  |
| OSD2    | Roscoff - SOMLIT         | Atlantic Ocean    | France                              |
| OSD3    | Helgoland                | Atlantic Ocean    | Germany                             |
| OSD4    | LTER-MC                  | Mediterranean Sea | Italy                               |
| OSD5    | Crete                    | Mediterranean Sea | Greece                              |
| OSD6    | Blanes                   | Mediterranean Sea | Spain                               |
| OSD7    | Moorea - Tiahura         | Pacific Ocean     | Tahiti                              |
| OSD8    | BATS                     | Atlantic Ocean    | Bermuda                             |
| OSD10   | Lake Erie W4             |                   | USA                                 |
| OSD13   | Varna Bay                | Black Sea         | Bulgaria                            |
| OSD14   | Banyuls                  | Mediterranean Sea | France                              |
| OSD15   | Villefranche - SOMLIT    | Mediterranean Sea | France                              |
| OSD17   | VLIZ                     | Atlantic Ocean    | Belgium                             |
| OSD18   | Kyrenia                  | Mediterranean Sea | Turkish Republic of Northern Cyprus |
| OSD19   | Famagusta                | Mediterranean Sea | Turkish Republic of Northern Cyprus |
| OSD20   | Faxaflooi                | Atlantic Ocean    | Iceland                             |
| OSD21   | Croatia                  | Mediterranean Sea | Croatia                             |
| OSD22   | Marseille Solemio SOMLIT | Mediterranean Sea | France                              |
| OSD24   | Marchica                 | Mediterranean Sea | Morocco                             |
| OSD25   | Saidia Rocher            | Mediterranean Sea | Morocco                             |
| OSD26   | Tangier                  | Atlantic Ocean    | Morocco                             |
| OSD28   | Belize                   | Atlantic Ocean    | Belize                              |
| OSD29   | Florida                  | Atlantic Ocean    | USA                                 |
| OSD30   | Tvärminne                | Atlantic Ocean    | Finland                             |
| OSD34   | Alexandria               | Mediterranean Sea | Egypt                               |

|       |                                |                   |             |
|-------|--------------------------------|-------------------|-------------|
| OSD35 | Cheasapeake Bay                | Atlantic Ocean    | USA         |
| OSD36 | Delaware                       | Atlantic Ocean    | USA         |
| OSD37 | Port Everglades                | Atlantic Ocean    | USA         |
| OSD38 | Long Key                       | Atlantic Ocean    | USA         |
| OSD39 | Charleston Harbor              | Atlantic Ocean    | USA         |
| OSD41 | Sequim Bay Park                | Pacific Ocean     | Alaska      |
| OSD42 | Faro Lake                      | Mediterranean Sea | Italy       |
| OSD43 | SIO Pier                       | Pacific Ocean     | USA         |
| OSD45 | Tampa Bay                      | Atlantic Ocean    | USA         |
| OSD46 | Horn Island                    | Atlantic Ocean    | USA         |
| OSD47 | Venice Lagoon                  | Mediterranean Sea | Italy       |
| OSD48 | Venice Gulf                    | Mediterranean Sea | Italy       |
| OSD49 | Vida                           | Mediterranean Sea | Slovenia    |
| OSD50 | Pasaia                         | Atlantic Ocean    | Spain       |
| OSD51 | Bocas del Toro                 | Atlantic Ocean    | USA         |
| OSD52 | Abu Hashish                    | Mediterranean Sea | Egypt       |
| OSD53 | Ras Disha                      | Mediterranean Sea | Egypt       |
| OSD54 | Maine Booth Bay                | Atlantic Ocean    | USA         |
| OSD55 | Maine Damariscotta River       | Atlantic Ocean    | USA         |
| OSD56 | Hawaii Kakaako                 | Pacific Ocean     | Hawaii      |
| OSD57 | Hawaii Oahu                    | Pacific Ocean     | Hawaii      |
| OSD58 | PICO                           | Atlantic Ocean    | USA         |
| OSD60 | South Carolina 2 - North Inlet | Atlantic Ocean    | USA         |
| OSD61 | Vineyard Sound                 | Atlantic Ocean    | USA         |
| OSD62 | Manai Straits                  | Atlantic Ocean    | UK          |
| OSD63 | Venice Acqua Alta              | Mediterranean Sea | Italy       |
| OSD64 | Odessa                         | Black Sea         | Ukraine     |
| OSD65 | Leigh Marine Laboratory        | Pacific Ocean     | New Zealand |

|        |                        |                   |             |
|--------|------------------------|-------------------|-------------|
| OSD69  | Marghera               | Mediterranean Sea | Italy       |
| OSD70  | Lido                   | Mediterranean Sea | Italy       |
| OSD71  | Otago                  | Pacific Ocean     | New Zealand |
| OSD72  | Boknis Eck             | Atlantic Ocean    | Germany     |
| OSD73  | Lima Estuary           | Atlantic Ocean    | Portugal    |
| OSD74  | Douro Estuary          | Atlantic Ocean    | Portugal    |
| OSD76  | Foglia                 | Mediterranean Sea | Italy       |
| OSD77  | Metauro                | Mediterranean Sea | Italy       |
| OSD78  | CONISMA                | Mediterranean Sea | Italy       |
| OSD80  | Young Sound            | Arctic Ocean      | Greenland   |
| OSD81  | Ria Formosa Lagoon     | Atlantic Ocean    | Portugal    |
| OSD90  | Etoliko Lagoon         | Mediterranean Sea | Greece      |
| OSD91  | Oualidiya              | Atlantic Ocean    | Morocco     |
| OSD92  | Casablanca             | Atlantic Ocean    | Morocco     |
| OSD93  | Eljadida               | Atlantic Ocean    | Morocco     |
| OSD94  | Saidia Marina          | Mediterranean Sea | Morocco     |
| OSD95  | Singapore Indigo_V     | Pacific Ocean     | Singapore   |
| OSD96  | Sao Miguel Azores I    | Atlantic Ocean    | Portugal    |
| OSD97  | Faial Azores           | Atlantic Ocean    | Portugal    |
| OSD98  | Sao Jorge Azores       | Atlantic Ocean    | Portugal    |
| OSD99  | C1                     | Mediterranean Sea | Italy       |
| OSD100 | Crete - GOS            | Mediterranean Sea | Greece      |
| OSD101 | Quinta do Lorde        | Atlantic Ocean    | Portugal    |
| OSD102 | Marina do Funchal      | Atlantic Ocean    | Portugal    |
| OSD103 | Porto da Cruz          | Atlantic Ocean    | Portugal    |
| OSD105 | Cambridge Bay, Nunavut | Arctic Ocean      | USA         |
| OSD106 | REYKIS                 | Atlantic Ocean    | Iceland     |
| OSD107 | Lisboa                 | Atlantic Ocean    | Portugal    |

|        |                                       |                      |              |
|--------|---------------------------------------|----------------------|--------------|
| OSD108 | Alcochete                             | Atlantic Ocean       | Portugal     |
| OSD109 | Rosario                               | Atlantic Ocean       | Portugal     |
| OSD110 | Figueira da Foz                       | Atlantic Ocean       | Portugal     |
| OSD111 | Ria de Aveiro_1                       | Atlantic Ocean       | Portugal     |
| OSD113 | CascaisWatch                          | North Atlantic Ocean | Portugal     |
| OSD114 | BerlengasWatch                        | Atlantic Ocean       | Portugal     |
| OSD115 | Santa Cruz                            | Atlantic Ocean       | Portugal     |
| OSD116 | Lagoa de Obidos                       | Atlantic Ocean       | Portugal     |
| OSD117 | Tavira Beach                          | Atlantic Ocean       | Portugal     |
| OSD118 | Lough Hyne                            | Atlantic Ocean       | Ireland      |
| OSD122 | Station A Gulf Of Eilat               | Red Sea              | Israel       |
| OSD123 | Shikmona                              | Mediterranean Sea    | Israel       |
| OSD124 | Osaka Bay                             | Pacific Ocean        | Japan        |
| OSD125 | Cullercoats Beach                     | Arctic Ocean         | UK           |
| OSD126 | Eyafjordur_1                          | Arctic Ocean         | Iceland      |
| OSD128 | Eyafjordur_3                          | Arctic Ocean         | Iceland      |
| OSD129 | Eyafjordur_4                          | Arctic Ocean         | Iceland      |
| OSD130 | Eyafjordur_5                          | Arctic Ocean         | Iceland      |
| OSD131 | Zlatna ribka                          | Black Sea            | Bulgaria     |
| OSD132 | Sdot YAM                              | Mediterranean Sea    | Israel       |
| OSD133 | Robben Island                         | Atlantic Ocean       | South Africa |
| OSD141 | Raunefjorden                          | Atlantic Ocean       | Norway       |
| OSD142 | Gray's Reef National Marine Sanctuary | Atlantic Ocean       | USA          |
| OSD143 | Skidaway Institute of Oceanography    | Atlantic Ocean       | USA          |
| OSD144 | Maunalua Bay O'ahu                    | Pacific Ocean        | USA          |
| OSD145 | Blankenberge                          | Atlantic Ocean       | Belgium      |
| OSD146 | Fram Strait                           | Arctic Ocean         | Germany      |
| OSD147 | Rajarata                              | Indian Ocean         | Sri Lanka    |

|        |                                      |                   |          |
|--------|--------------------------------------|-------------------|----------|
| OSD148 | Wadden Sea                           | Atlantic Ocean    | Germany  |
| OSD149 | Laguna Rocha Norte                   | Atlantic Ocean    | Uruguay  |
| OSD150 | Laguna Rocha Sur                     | Atlantic Ocean    | Uruguay  |
| OSD151 | South Atlantic Microbial Observatory | Atlantic Ocean    | Uruguay  |
| OSD152 | Compass Buoy Station                 | Atlantic Ocean    | Canada   |
| OSD153 | Faro Island                          | Atlantic Ocean    | Portugal |
| OSD154 | Arcachon-SOMLIT                      | Atlantic Ocean    | France   |
| OSD155 | Steilene Oslofjord                   | Atlantic Ocean    | Norway   |
| OSD156 | Hvaler Tisler Site                   | Atlantic Ocean    | Norway   |
| OSD157 | ELLElm2                              | Atlantic Ocean    | Norway   |
| OSD158 | Sao Miguel Azores II                 | Atlantic Ocean    | Portugal |
| OSD159 | Brest-SOMLIT                         | Atlantic Ocean    | France   |
| OSD162 | Stonehaven                           | Atlantic Ocean    | UK       |
| OSD163 | Scapa                                | Atlantic Ocean    | UK       |
| OSD164 | Scalloway                            | Atlantic Ocean    | UK       |
| OSD165 | Loch Ewe                             | Atlantic Ocean    | UK       |
| OSD166 | Armintza                             | Atlantic Ocean    | Spain    |
| OSD167 | Eyafjordur_6                         | Arctic Ocean      | Iceland  |
| OSD168 | IMST_izmir                           | Mediterranean Sea | Turkey   |
| OSD169 | Brightlingsea Creek, Essex           | Atlantic Ocean    | UK       |
| OSD170 | Belgium - 130                        | Atlantic Ocean    | Belgium  |
| OSD171 | Belgium - 230                        | Atlantic Ocean    | Belgium  |
| OSD173 | Belgium - 710                        | Atlantic Ocean    | Belgium  |
| OSD174 | Belgium - 780                        | Atlantic Ocean    | Belgium  |
| OSD175 | ZG02                                 | Atlantic Ocean    | Belgium  |
| OSD176 | Belgium - 215                        | Atlantic Ocean    | Belgium  |
| OSD177 | Belgium - 120                        | Atlantic Ocean    | Belgium  |
| OSD178 | Belgium - 435                        | Atlantic Ocean    | Belgium  |

|        |                           |                |            |
|--------|---------------------------|----------------|------------|
| OSD182 | W08                       | Atlantic Ocean | Belgium    |
| OSD183 | W09                       | Atlantic Ocean | Belgium    |
| OSD184 | W10                       | Atlantic Ocean | Belgium    |
| OSD185 | Belgium - 421             | Atlantic Ocean | Belgium    |
| OSD186 | SERC Rhode River Maryland | Atlantic Ocean | USA        |
| OSD187 | Palmer station            | Southern Ocean | Antarctica |

**Table S2.** Matrix of the pairwise percent identity between *Ostreococcus* clades. The calculation was done based on the alignment available as Supplementary data S6.

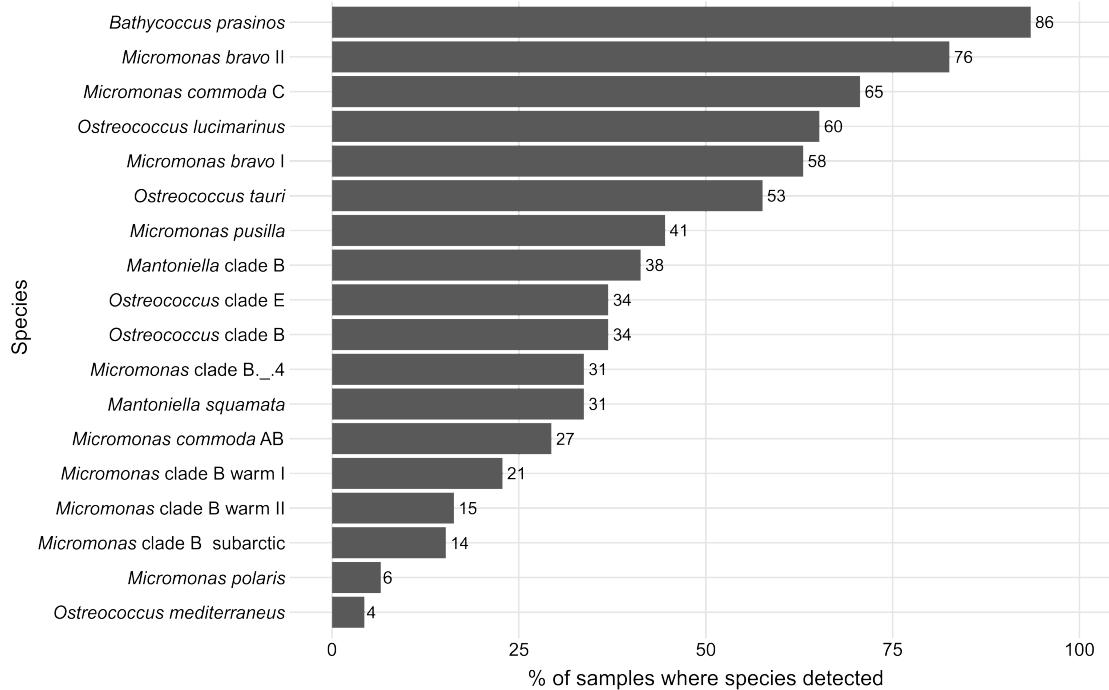
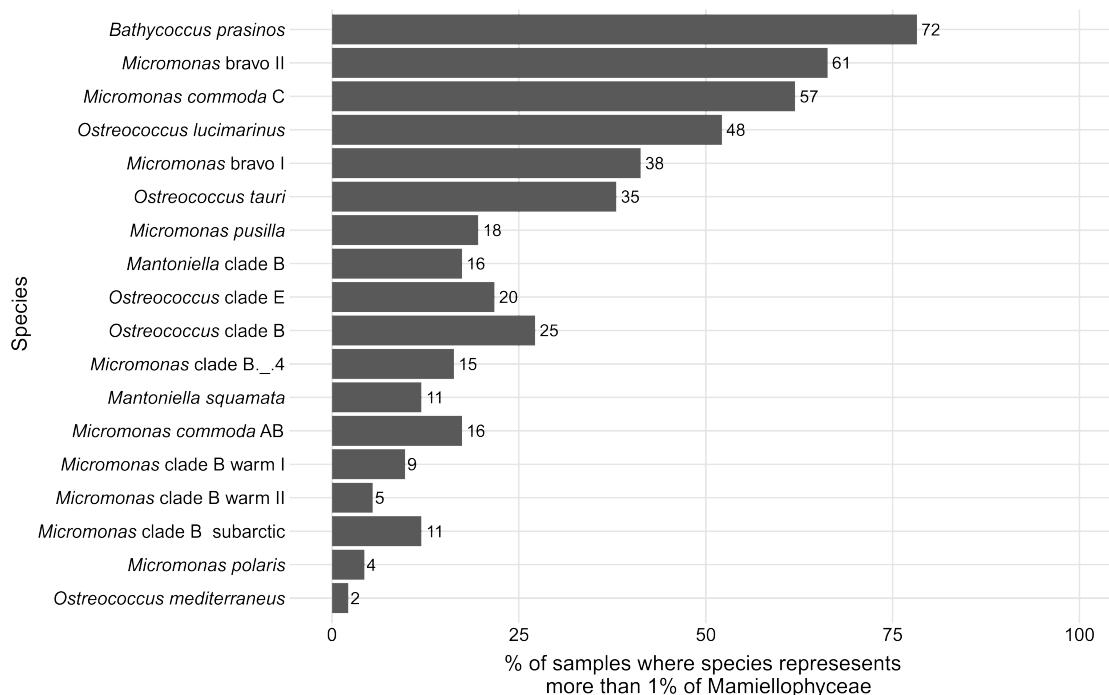
| Ostreococcus     | "O. lucimarinus" | O. clade B | O. tauri | O.mediterraneus | clade E |
|------------------|------------------|------------|----------|-----------------|---------|
| "O. lucimarinus" | 99.70            | 98.80      | 99.40    | 98.00           | 99.20   |
| O. clade B       | 98.80            | 99.70      | 99.10    | 97.70           | 99.40   |
| O. tauri         | 99.40            | 99.10      | 99.70    | 98.30           | 99.50   |
| O.mediterraneus  | 98.00            | 97.70      | 98.30    | 99.80           | 98.00   |
| clade E          | 99.20            | 99.40      | 99.50    | 98.00           | 99.80   |

**Table S3.** Matrix of the pairwise percent identity between *Micromonas* clades. The calculation was done based on the alignment available as Supplementary data S7.

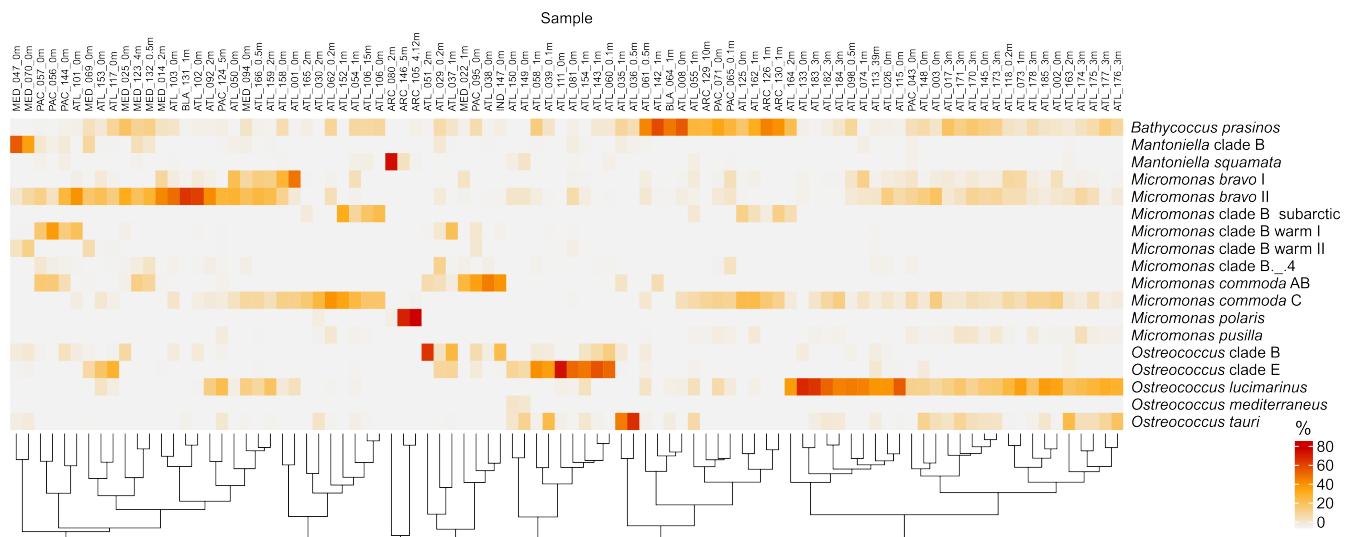
| Micromonas     | M. polaris | B sub-arctic | M. bravo II | M. bravo I | B._.4  | M. pusilla | B warm I | B warm II | M. commoda(C) | M. commoda(AB) |
|----------------|------------|--------------|-------------|------------|--------|------------|----------|-----------|---------------|----------------|
| M. polaris     | 99.90      | 99.10        | 98.90       | 98.30      | 97.40  | 97.30      | 97.70    | 98.20     | 97.20         | 97.40          |
| B sub-arctic   | 99.10      | 100.00       | 98.40       | 97.80      | 96.80  | 96.80      | 97.20    | 97.50     | 97.10         | 97.40          |
| M. bravo II    | 98.90      | 98.40        | 100.00      | 99.40      | 97.00  | 97.60      | 98.10    | 98.20     | 97.50         | 97.50          |
| M. bravo I     | 98.30      | 97.80        | 99.40       | 99.50      | 97.00  | 97.10      | 97.70    | 97.90     | 97.10         | 97.30          |
| B._.4          | 97.40      | 96.80        | 97.00       | 97.00      | 100.00 | 97.10      | 97.70    | 97.70     | 97.50         | 97.40          |
| M. pusilla     | 97.30      | 96.80        | 97.60       | 97.10      | 97.10  | 99.80      | 98.30    | 98.60     | 97.40         | 96.90          |
| B warm I       | 97.70      | 97.20        | 98.10       | 97.70      | 97.70  | 98.30      | 100.00   | 99.80     | 98.30         | 97.90          |
| B warm II      | 98.20      | 97.50        | 98.20       | 97.90      | 97.70  | 98.60      | 99.80    | 100.00    | 98.30         | 97.20          |
| M. commoda(C)  | 97.20      | 97.10        | 97.50       | 97.10      | 97.50  | 97.40      | 98.30    | 98.30     | 99.80         | 99.20          |
| M. commoda(AB) | 97.40      | 97.40        | 97.50       | 97.30      | 97.40  | 96.90      | 97.90    | 97.20     | 99.20         | 100.00         |

**Table S4.** Matrix of the pairwise percent identity between *Mantoniella* clades. The calculation was done based on the alignment available as Supplementary data S8.

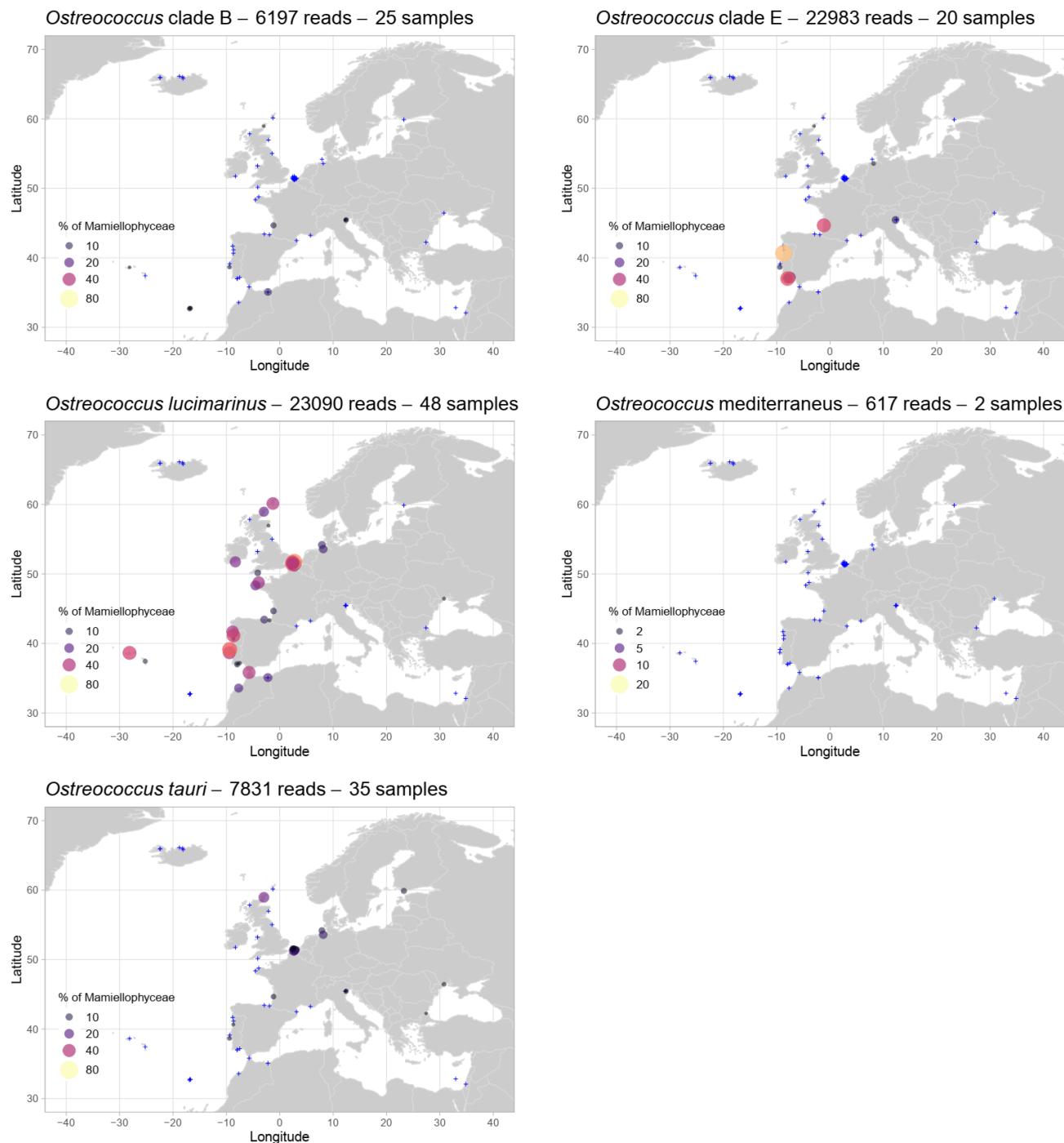
| Mantoniella   | M. antarctica | M. squamata | M. clade A | M. clade B |
|---------------|---------------|-------------|------------|------------|
| M. antarctica | 99.70         | 99.00       | 96.70      | 99.30      |
| M. squamata   | 99.00         | 100.00      | 97.50      | 99.80      |
| M. clade A    | 96.70         | 97.50       | 97.60      | 97.30      |
| M. clade B    | 99.30         | 99.80       | 97.30      | 99.90      |

**A****B**

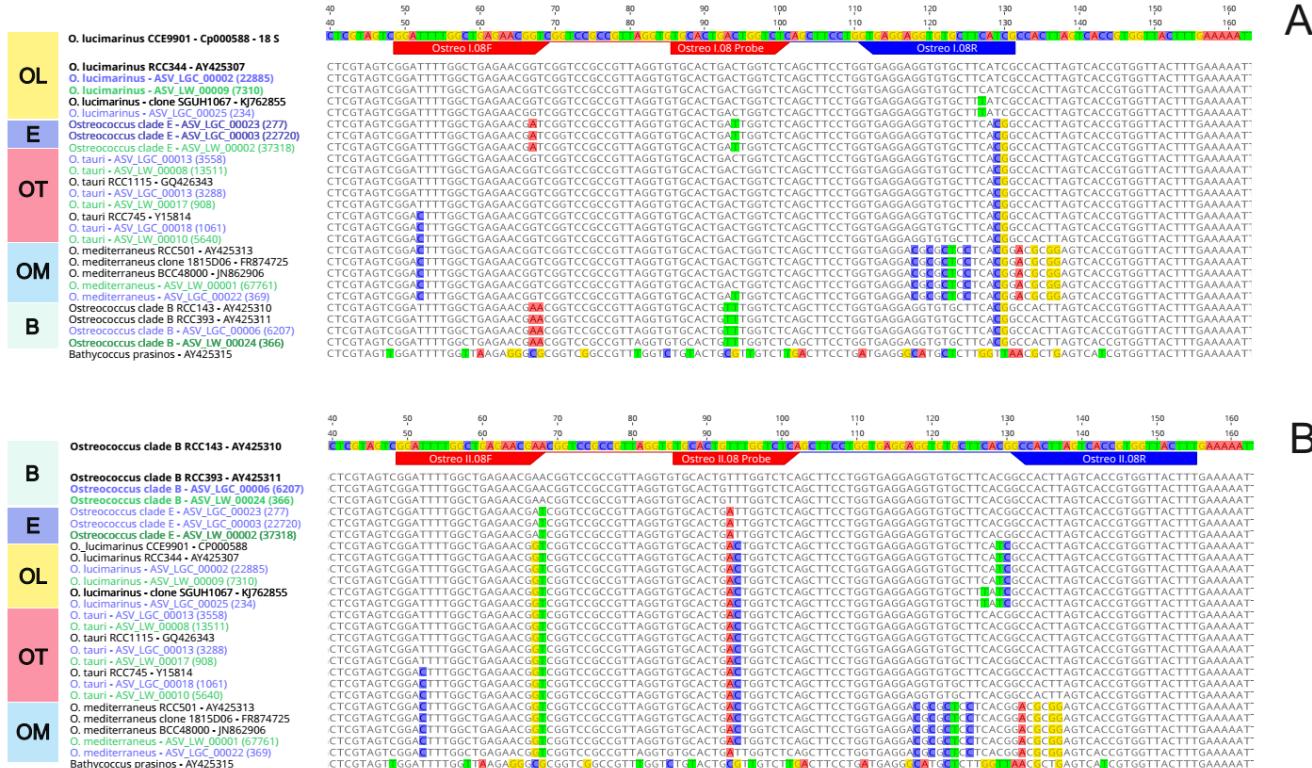
**Figure S1.** A. Percentage of OSD samples where a given ASV was detected (at least one read). Numbers at right of bars correspond to number of samples. B. Idem but for samples where the ASV contributed more than 1% of the Mamiellophyceae reads. Stations with less than 100 Mamiellophyceae reads were not considered. The three major *O. tauri* ASVs have been pooled together.



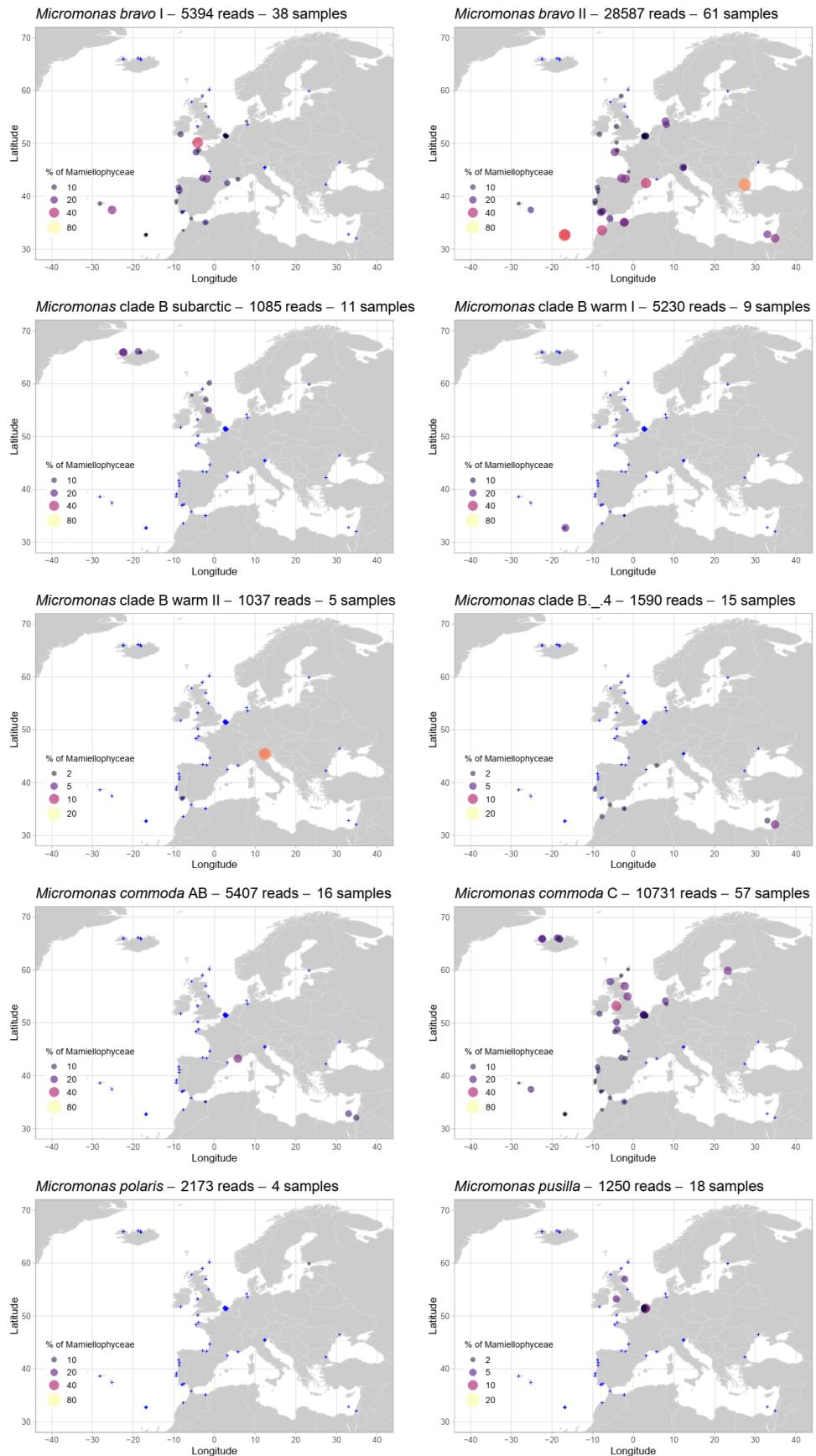
**Figure S2.** Heatmap of the Mamiellophyceae communities in OSD samples. The number of read corresponding to the major ASVs were normalized by the number of Mamiellophyceae reads per station. Stations with less than 100 Mamiellophyceae reads were not considered.



**Figure S3.** Distribution of major *Ostreococcus* ASVs for OSD2014 (LGC) in European waters. The three major *O. tauri* ASVs have been pooled together. Circle surface corresponds to the percentage of ASV reads relative to the total number of Mamiellophyceae reads. Samples for which the ASV contribution was lower than 1 % are represented by blue crosses.

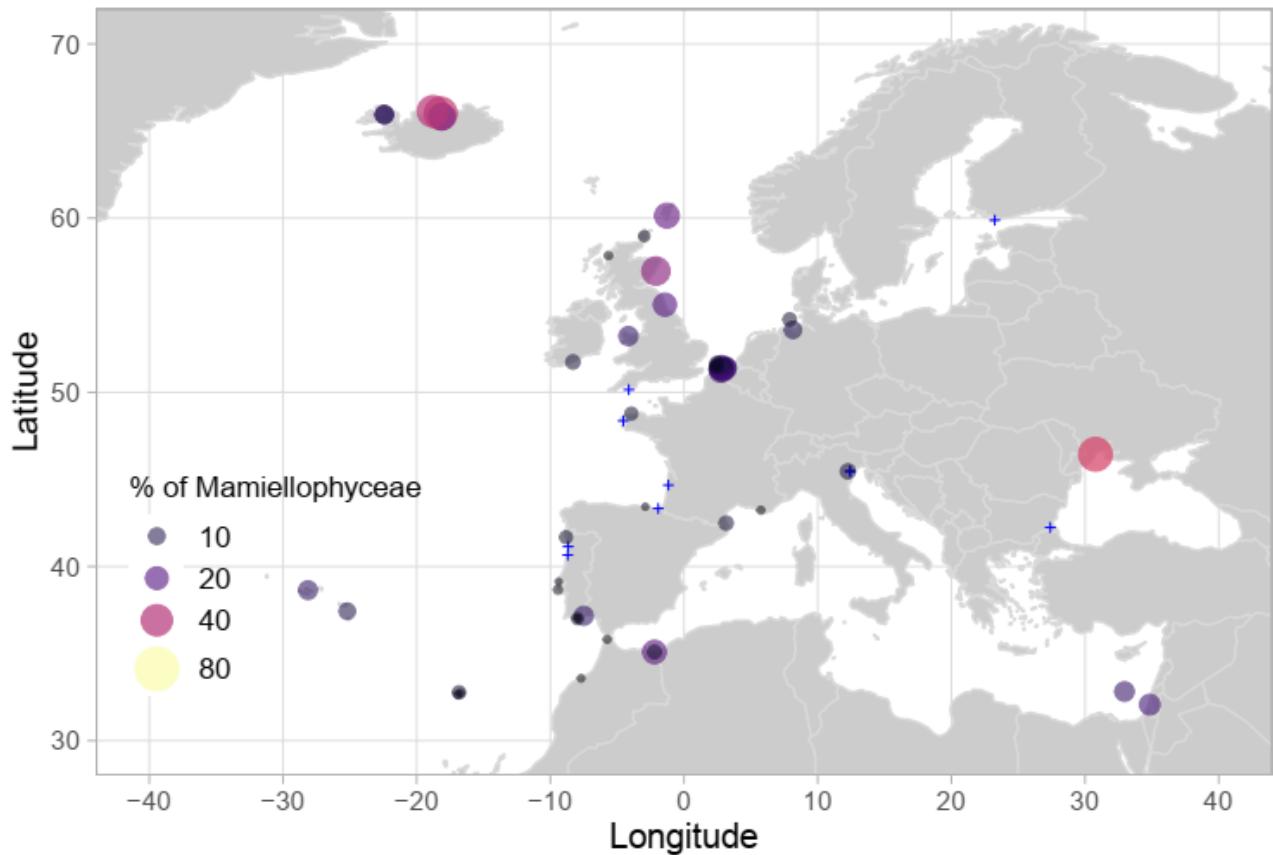


**Figure S4.** Alignments of *Ostreococcus* V4 regions and localization of the qPCR primers and probes (Demir-Hilton et al. 2011) used to quantify the OI (A) and OII (B) clades. Sequences with no mismatches to each probe set are in bold. Blue and green sequences correspond to LGC and LW datasets, respectively.

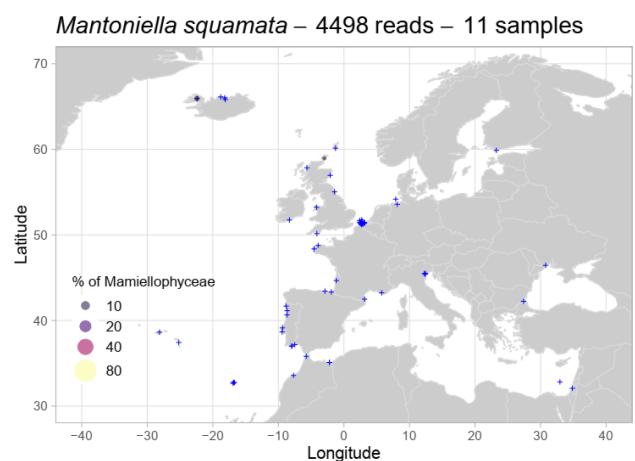
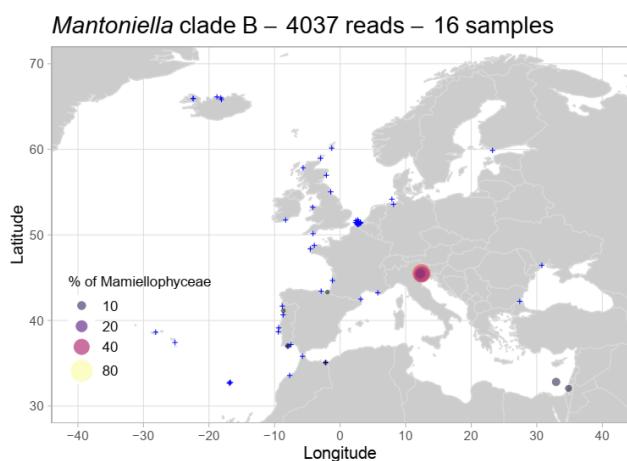


**Figure S5.** Same as Fig. S3 for *Micromonas*.

## *Bathycoccus prasinos* – 16651 reads – 72 samples



**Figure S6.** Same as Fig. S3 for *Bathycoccus*.



**Figure S7.** Same as Fig. S3 for *Mantoniella*.