



Swansea University
Prifysgol Abertawe



Cronfa - Swansea University Open Access Repository

This is an author produced version of a paper published in:

Global Ecology and Biogeography

Cronfa URL for this paper:

<http://cronfa.swan.ac.uk/Record/cronfa48817>

Paper:

Leles, S., Mitra, A., Flynn, K., Tillmann, U., Stoecker, D., Jeong, H., Burkholder, J., Hansen, P., Caron, D., et. al. (2019). Sampling bias misrepresents the biogeographical significance of constitutive mixotrophs across global oceans. *Global Ecology and Biogeography*

<http://dx.doi.org/10.1111/geb.12853>

This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder.

Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

<http://www.swansea.ac.uk/library/researchsupport/ris-support/>

1 **Title:**

2 **Sampling bias misrepresents the biogeographic significance of constitutive mixotrophs**
3 **across global oceans**

4 **Suzana Gonçalves Leles¹, Aditee Mitra^{1*}, Kevin John Flynn¹, Urban Tillmann², Diane**
5 **Stoecker³, Hae Jin Jeong⁴, JoAnn Burkholder⁵, Per Juel Hansen⁶, David A. Caron⁷,**
6 **Patricia M. Glibert³, Gustaaf Hallegraeff⁸, John Raven^{9,10}, Robert W. Sanders¹¹, Mikhail**
7 **Zubkov¹²**

8 **Affiliations**

9 ¹Swansea University, Singleton Park, Swansea SA2 8PP, UK

10 ²Alfred-Wegener-Institut, Helmholtz-Centre for Polar and Marine Research,
11 Am Handelshafen 12, 27570 Bremerhaven, Germany

12 ³University of Maryland Center for Environmental Science, Horn Point Laboratory, 2020
13 Horns Point Rd., PO Box 775, Cambridge MD 21613, USA

14 ⁴School of Earth and Environmental Sciences, College of Natural Sciences, Seoul National
15 University, Seoul 08826, Republic of Korea

16 ⁵Center for Applied Aquatic Ecology, Department of Applied Ecology, North Carolina State
17 University, Raleigh, NC 27606 USA

18 ⁶Marine Biological Section, Department of Biology, University of Copenhagen,
19 Strandpromenaden 5, DK-3000 Helsingør, Denmark

20 ⁷Department of Biological Sciences, University of Southern California, 3616 Trousdale
21 Parkway, Los Angeles, 90089-0371 USA

22 ⁹Division of Plant Science, University of Dundee at the James Hutton Institute, Invergowrie,
23 Dundee DD2 5DA, UK.

24 ¹⁰Climate Change Cluster, University of Technology Sydney, Ultimo, NSW 2007, Australia.

25 ¹¹Department of Biology, Temple University, Philadelphia, PA 19122, USA

26

27 *corresponding author: A.Mitra@swansea.ac.uk

28

29 **Running title:** Sampling bias in mixotroph biogeography

30

31 **ABSTRACT**

32 **Aim:** Most protist plankton are mixotrophic, with potential to engage in photoautotrophy and
33 phagotrophy; however, the ecology of these organisms has been misdiagnosed for over a
34 century. A large proportion of these organisms are constitutive mixotrophs (CMs), with an
35 innate ability to photosynthesize. Here, for the first time, an analysis is presented of the
36 biogeography of CMs across the oceans.

37 **Location:** Global marine ecosystems

38 **Time period:** 1970 to 2018

39 **Major taxa studied:** Marine planktonic protists

40 **Methods:** Records for CM species, primarily from the Ocean Biogeographic Information
41 System (OBIS), were grouped by taxonomy and size to evaluate sampling efforts across
42 Longhurst's oceanic provinces. Biases were evaluated through non-parametric tests and
43 multivariate analysis. Biogeographies of CMs from OBIS data were compared with data from
44 studies that specifically targeted these organisms.

45 **Results:** CMs of different taxonomic groups, across all size ranges, are ubiquitous. However,
46 strong database biases were detected with respect to organism size, taxonomic groups, and
47 region. A strong bias was seen towards dinophytes. Species < 20 µm, especially non-
48 dinophytes, were least represented, their recorded distribution limited to coastal regions and
49 to temperate and polar seas. Studies specifically targeting these organisms revealed their
50 distribution to be much wider. Such biases likely have occurred due to a failure to capture
51 and correctly identify these organisms in routine sampling protocols.

52 **Main conclusions:** CMs are dominant members of organisms traditionally termed
53 "phytoplankton". However, lack of routine protocols for measuring phagotrophy in
54 "phytoplankton" protists has led to widespread misrepresentation of the fundamental nature
55 of marine planktonic primary producers; most express 'animal-like' as well as 'plant-like'
56 nutrition. Our results have implications for studies of the global biogeography of plankton, of
57 food web dynamics (including models), and of biogeochemical cycling in the oceans.

58 **KEYWORDS**

59 allometry, biogeography, global, Longhurst, mixotrophy, oceans, phytoplankton, taxonomy

60

61 **Introduction**

62 The term biogeography was coined in the late nineteenth-century stemming from studies of
63 terrestrial plants and animals (Ebach, 2015). Studies of biogeography have typically focussed
64 on terrestrial ecosystems. Since the early twentieth century, however, there has been growing
65 recognition of the importance of the biogeography of marine microbes across the Earth's
66 oceans, the single largest continuous ecosystem (Smayda, 1958; Dolan, 2006; Cermeño, de
67 Vargas, Abrantes, & Falkowski, 2010). Over the last decade, various studies have highlighted
68 technical advancements, such as the development of molecular and statistical techniques and
69 the availability of online databases, as contributing factors to the rise in marine
70 biogeography-facing studies (Dolan, 2005; Fuhrman et al., 2008; de Vargas et al., 2015;
71 Biard et al., 2016; Leles et al., 2017).

72 Studies of biogeography of marine microbes (prokaryotes and eukaryotes) have
73 concentrated on functional groups, such as cyanobacteria vs phytoplankton vs
74 microzooplankton, or taxonomic groups, such as diatoms and tintinnids (Pierce & Turner,
75 1993; Cermeño & Falkowski, 2009; Malviya et al., 2016). Most functional or taxonomic
76 groupings have followed the traditional designations of either phototrophs or heterotrophs,
77 akin to the plant-animal dichotomy in terrestrial systems. Flynn et al. (2013) criticised this
78 dichotomy, identifying that the vast bulk of the protist plankton formally labelled as
79 “phytoplankton” or “microzooplankton” are potentially mixotrophic, merging phototrophy
80 and phagotrophy. Subsequently, Mitra et al. (2016) proposed a new functional classification
81 for marine protists to aid the exploration of the proposed new mixotroph-centric paradigm in
82 marine ecology (Zubkov & Tarran, 2008; Unrein, Gasol, Not, Forn, & Massana, 2014).
83 According to this functional classification, marine protists are broadly divided between six
84 functional groups; two of these align with the traditional non-phagotrophic phytoplankton
85 (notably diatoms) and non-phototrophic microzooplankton, while the other four represent
86 contrasting mixotroph functional groups. The mixotroph groups are divided between those
87 with an innate (constitutive) ability to photosynthesize (constitutive mixotrophs - CMs), and
88 three non-constitutive mixotroph (NCM) groups. The NCMs do not have the innate ability to
89 photosynthesize; they derive their photosynthetic capabilities by incorporating plastids from a
90 range of different prey (generalist NCMs, e.g., *Laboea*), or from very specific phototrophic
91 prey (specialist NCMs, e.g., *Mesodinium*), or by enslaving phototrophic prey as symbionts
92 (endosymbiotic NCMs, e.g., Rhizarians).

93 It has been shown that incorporating mixotrophic organisms within *in silico* food web
94 studies alters the dominance of different plankton functional groups in freshwater as well as
95 marine systems (Mitra & Flynn, 2010; Wilken, Huisman, Naus-Wiezer, & Van Donk, 2013;
96 Mitra et al., 2014). An understanding of the true trophic status of plankton, including
97 different mixotrophic types, across biogeographic areas is thus important. Leles et al. (2017)
98 undertook a biogeographic analysis of the non-constitutive mixotroph (NCM) functional
99 groups. These NCM organisms are undeniably mixotrophic, as they must feed to acquire
100 phototrophic potential. However, the real trophic status of constitutive mixotrophs is
101 problematic because these organisms may not need to eat frequently if de facto at all (Flynn
102 et al., 2013; Mitra et al., 2016). The default expectation has been to retain the traditional
103 “phytoplankton” label for these organisms. Yet, if they are indeed functional mixotrophs, this
104 traditional label would misrepresent biogeographic distributions and allied trophic dynamics.

105 In this study we have undertaken the first global biogeographic analysis of marine
106 protists with a constitutive ability to photosynthesize and that are documented as being
107 significantly mixotrophic (i.e., organisms that are undeniably CMs). The CM group includes
108 various ecologically important taxonomic groups within a wide range of sizes, such as
109 species of picoplankton and nanoplankton in oligotrophic oceans and polar regions (Stoecker
110 & Lavrentyev, 2018), as well as various harmful algal bloom species in coastal waters (e.g.,
111 *Karlodinium* spp., *Alexandrium* spp.) which result in fish kills and closure of aquaculture
112 facilities (Mitra et al., 2016; Stoecker, Hansen, Caron, & Mitra, 2017; Shumway, Burkholder,
113 & Morton, 2018). Biogeographic analysis of this important group of planktonic protists will
114 aid our understanding of impacts of environmental drivers on community composition and
115 ecosystem functioning especially in the face of climate change events. In our analyses, we
116 specifically identify how biases in the scientific community’s perception of these organisms,
117 as well as sampling methods, impact on our understanding of their biogeography.

118

119 **Methods**

120 We conducted a global analysis of field data for different groups of constitutive mixotrophs
121 (CMs). CMs are defined according to Mitra et al. (2016); these are planktonic protists with an
122 inherent capability to photosynthesize and also a demonstrable potential to engage in
123 phagotrophy for their nutritional needs. Traditionally mixotrophy within many planktonic
124 phototrophs has included phototrophy *plus* osmotrophy (i.e., uptake of dissolved organic
125 substances; e.g., Glibert & Legrand, 2006; Burkholder, Glibert, & Skelton, 2008). Here,
126 however, we focus on photoautotrophic protists which engage in phagotrophy because
127 osmotrophy appears to be ubiquitous in protists; thus, it is assumed here that all species are
128 capable of osmotrophy (see Flynn et al. 2013; Mitra et al. 2016).

129 *Data compilation*

130 Data were compiled according to species name. AlgaeBase (<http://www.algaebase.org/>) was
131 used to resolve synonyms and basionyms of various species. Using expert knowledge within
132 our team (authors), we first assembled a database of all protist species known to be CMs
133 (Appendix S1, Table S1.1); the definitions of Mitra et al. (2016) were used to differentiate
134 between CMs and strict autotrophs (i.e., phytoplankton). We did not consider
135 coccolithophorids; to date, there has been only one published study reporting phagotrophic
136 activity by the cosmopolitan species *Emiliana huxleyi* (Rokitta et al., 2011).

137 Data for the global distribution of the CMs within our list were acquired through
138 interrogation of the Ocean Biogeographic Information System database (OBIS;
139 <http://www.iobis.org/>) on 20th January 2018. Geographic coordinates corresponding to the
140 locations where the CMs were recorded were obtained. Records with possible spatial errors,
141 such as data points located inland, were excluded from the analysis. Georeferenced
142 occurrence data were retrieved from OBIS using the ‘devtools’ and ‘robis’ packages in R (R
143 Core Team, 2017). Other packages in R used for data compilation and visualization were
144 ‘rgdal’, ‘plyr’, ‘ggplot2’, ‘ggalt’, and ‘gridExtra’.

145 In order to ensure that we captured distribution data of CMs across different size
146 classes, we conducted a survey of published literature in the electronic databases ISI Web of
147 Science and Elsevier on 20th January 2018. All data sources are listed within Appendix 1.
148 Smaller planktonic protists are rarely, if at all, identified down to species level in field
149 surveys. The aim of these surveys was therefore to obtain records from studies which
150 specifically targeted known CMs in the nanoplankton spectrum (2–20 µm length). These

151 studies, albeit not being species-specific, reported *in situ* measurements of the abundance of
152 actively feeding mixotrophic nanoflagellates (i.e., CMs of < 20 μm length) or the relative
153 contribution of pico- and nano- CMs to total bacterivory by flagellates. Geographic
154 coordinates were retrieved from each of these studies in order to compare these data with
155 species-specific data obtained from the OBIS database.

156 *Spatial analysis*

157 Records compiled from online databases were aligned with the biogeographic classification
158 of the ocean according to Longhurst (2007). As in Leles et al. (2017), the 54 biogeographic
159 provinces proposed by Longhurst (2007) were grouped into seven principal biomes according
160 to primary production and physical forcing: Mediterranean Sea, Coastal Seas, Polar Seas,
161 Temperate Seas, Oligotrophic Gyres, Coastal Upwelling, and Equator (Appendix S1, Table
162 S1.2). However, due to a lack of data we did not include the Coastal Upwelling biome in our
163 analysis. Grids corresponding to Longhurst provinces used in the maps were obtained from
164 <http://www.marineregions.org/>. Geographic coordinates corresponding to the exact location
165 of where the CM species were found were then aligned with biogeographic provinces. The
166 records obtained for each species within each biogeographic province were exploited to
167 produce global distribution maps.

168 *Data analysis*

169 The biogeography of CMs was investigated according to taxonomy and size classes across
170 different oceanic provinces. Species size was obtained from the literature (Tomas, 1997;
171 Hoppenrath & Leander, 2007; Berge, Hansen, & Moestrup, 2008; Nézan & Chomérat, 2009;
172 Yoo et al., 2010; Kang et al., 2011; Lim et al., 2015; Jang, Jeong, Kwon, & Lee, 2017; Ok,
173 Jeong, Lim, & Lee, 2017) and online repositories (<http://nordicmicroalgae.org>;
174 <http://www.sccap.dk>; <http://www.marinespecies.org>). Global distribution maps were
175 generated to visualize the presence of each taxonomic group within each size class according
176 to the biogeographic provinces. The sampling locations from all 178 cruises included in the
177 analysis were also mapped in order to compare the biogeographic patterns with the total
178 sampling effort. It is noteworthy that we did not assume *a priori* that all taxonomic groups
179 should contain individuals within all size-classes; indeed, that is not the case. We used expert
180 knowledge and the literature to identify size class boundaries within each taxonomic group.
181 For example, there are no known species within prasinophytes, chrysophytes, and
182 haptophytes which are > 15 μm in size (measured across the major cell axis). Likewise, all

183 known raphidophyte species are $> 10 \mu\text{m}$. While there are some suggestions of presence of
184 so-called picoplanktonic marine dinoflagellates ($< 5 \mu\text{m}$), this information is derived from
185 initial sequence data only (Moon-van der Staay, De Wachter, & Vaulot, 2001; Lin, Zhang,
186 Hou, Miranda, & Bhattacharya, 2006); thus, we could not include such organisms in this
187 analysis. Dinophytes were, therefore, not assumed to occur within the $< 5 \mu\text{m}$ size-class.

188 Potential biases related to size and location, when tracing the global distribution of
189 CMs, were analysed quantitatively. For this, we used the mean number of records obtained
190 for each size class across different oceanic biomes. Our dataset did not follow a normal
191 distribution; therefore, Kruskal-Wallis rank tests were used to evaluate the effect of size and
192 of biome on CM distribution. Species were grouped according to size within the different
193 oceanic biomes to test for differences amongst size classes and across the biomes. Post-hoc
194 non-parametric tests (Dunn's test) were performed to identify pairs of size classes or oceanic
195 biomes that were significantly different from each other. The same procedure was performed
196 to evaluate the potential bias related to CM taxonomy and location, grouping species
197 according to taxonomy instead of size to test for any differences between taxonomic groups
198 and across the biomes. These analyses were conducted using the 'dunn.test' package in R.

199 Dissimilarities between occurrence patterns of CM species across the different
200 biogeographic provinces were explored using the non-metric multidimensional scaling
201 (NMDS) technique; this is akin to the analysis undertaken by Leles et al. (2017) on NCMs.
202 The georeferenced data compiled from OBIS (Appendix S2, Table S2.1) were used to build a
203 matrix with the number of occurrences for each species within each of the 54 Longhurst's
204 biogeographic provinces. CM species were then grouped according to size and taxonomy.
205 This analysis thus enables positioning of species and biomes within a 2D space; the former
206 are placed with respect to their distribution across the biogeographic provinces and the latter
207 according to species occurrence. Merging the positions of both species and biomes allowed
208 us to explore the placement of species with respect to each other as well as in relation to the
209 different biomes. The distance matrix used in the NMDS analysis was calculated after square
210 root transformation using the Bray-Curtis distance. NMDS was performed using the
211 'metaMDS' function ('vegan' package in R). Progressively lower number of dimensions ($k =$
212 5 to 2) were used in order to attain the value for k which provided the lowest stress value
213 (Legendre & Legendre, 1998).

214 The volume of seawater analysed will inevitably influence the identification of rare
215 species (i.e., those present in low abundances). However, we could not account for this factor

216 since most publications do not report the volume sampled and analysed and, therefore, this
217 information could not be retrieved from OBIS. Even if this information were available, the
218 volume of water that ideally should be analysed can be highly variable depending on the
219 trophic status of the system (e.g., larger volumes of water should be inspected in oligotrophic
220 waters compared to eutrophic systems, and also depending on bloom seasonality); additional
221 knowledge on the different marine ecosystems investigated would be required to interpret
222 such data. Moreover, we would expect a minimal effect in our biogeographic analysis since
223 we grouped constitutive mixotrophs into different taxonomic groups and size-classes.

224

225 **Results**

226 The compiled list of CMs includes 80 species across the different taxonomic groups of
227 primarily cryptophytes, chrysophytes, haptophytes, prasinophytes, raphidophytes, dinophytes,
228 chlorarachniophytes, and synchronophytes (Appendix S1, Table S1.1). Nearly 250,000
229 records were obtained from OBIS for 51 out of the 80 species that comprised the initial
230 species list (Appendix S2, Table S2.1); these species were placed into taxonomic groups and
231 five size classes (Appendix S2, Table S2.1). The global distributions of CMs across the
232 different biogeographic provinces were mapped according to taxonomic groups and
233 allometrics (Figure 1). At least one record was necessary to assume the presence of a species
234 within a province; thus, only five taxonomic groups (Appendix S2, Table S2.1) were mapped
235 (Figure 1).

236 These maps demonstrate a strong bias in the available data towards larger species, and
237 specifically towards dinophytes. Nearly 50% of the constitutive mixotrophic species were
238 larger than 20 μm ; except for two raphidophytes, all species $> 20 \mu\text{m}$ were dinophytes. The
239 dearth of data for other CM groups across the different size classes is indicated by the “non-
240 coloured” (white) biogeographic maps in Figure 1. In order to compare these biogeographic
241 patterns with the total sampling effort, we also mapped the sampling locations from all 178
242 cruises included in the analysis (Appendix S3, Figure S3.1). From this it can be seen which
243 biogeographic provinces were visited at different points in time and space but were not
244 associated with the presence of a CM group (white biogeographic provinces in Figures 1 and
245 2 vs Appendix S3, Figure S3.1).

246 The potential for biases relating to species size, taxonomic grouping, and location
247 were tested using quantitative data (Appendix S3, Figure S3.2). The number of records
248 differed significantly among size classes with a clear bias towards larger size classes ($H =$

249 11.8, $p = 0.02$; Appendix S3, Figure S3.2a). Post-hoc comparisons confirmed that sampling
250 effort was highest for species with maximum cell dimension $> 20 \mu\text{m}$ (Figure 2a; Appendix
251 S4, Table S4.1). Differences were also found when comparing taxonomic groups ($H = 12.2$, p
252 $= 0.02$; Appendix S3, Figure S3.2b). The number of records obtained for dinophytes was
253 significantly higher than those retrieved for any other taxonomic group and no significant
254 difference was found among the other groups (Figure 2b; Appendix S4, Table S4.2). We also
255 tested whether the number of records varied across biomes for each dataset (i.e., grouped by
256 size or by taxonomy, Appendix S3, Figure S3.1), and there were significant differences in
257 both analyses ($H = 13.6$, $p = 0.02$; and $H = 11.2$, $p = 0.05$, respectively). Fewer records were
258 available from oligotrophic gyres and equatorial regions compared to other biomes
259 (Appendix S3, Figure S3.1; Appendix S4, Tables S4.3 and S4.4); indeed, data within
260 oligotrophic gyres and equatorial regions were available primarily for dinophytes except for
261 one record of a raphidophyte species (Figure 2b). The analysis revealed no clear difference
262 among the other biomes (Appendix S4, Tables S4.3 and S4.4).

263 The relationships between size classes and taxonomy in the global distribution of
264 CMs were explored through the NMDS analysis (Figure 3). Our analysis revealed dinophyte
265 species within the larger size class ($> 20 \mu\text{m}$) closer to the different biomes thus displaying a
266 broader geographic distribution compared to other species (Figure 3). The different colours in
267 this figure allow the reader to discern that different taxonomic groups have different
268 distribution patterns. Through reference to the different symbols (Figure 3), it is clear that
269 size and taxonomy are not independent and most species of dinophytes are $> 20 \mu\text{m}$ while
270 most species within other groups are $< 20 \mu\text{m}$. However, even among dinophytes there were
271 various outliers indicating a bias towards focussing on sampling species $> 20 \mu\text{m}$ in field
272 studies; *Alexandrium andersonii* and *Fragilidium subglobosum* are examples of these
273 “outlier” large dinophytes which ordinated closer to species belonging to other taxonomic
274 groups, such as the haptophyte *Prymnesium polylepis* and the prasinophyte *Cymbomonas*
275 *tetramitiformis*.

276 Non-dinophyte species $< 20 \mu\text{m}$ were the least represented group of CMs. To further
277 illustrate the under-representation of these groups within global databases, and thus the
278 incomplete status of global distribution data, we plotted the data available for these groups
279 from generic studies (Figure 4) and from studies that specifically targeted these groups
280 (Appendix S2). A total of 48 records from 21 studies were located when interrogating studies
281 which specifically targeted known CMs in the nanoplankton spectrum ($2\text{--}20 \mu\text{m}$; Appendix

282 S2, Table S2.2). It should be noted that these studies do not provide species-specific
283 information; unless a sampling study specifically targets smaller CM species and uses
284 specialist identification methods and skills, it is highly unlikely, and often impossible, for
285 these taxa to be identified to species level using standard light microscopy within generic
286 sampling protocols. Therefore, the maps derived using data from the generic versus the
287 specific studies present very different outputs (Figure 4). The map obtained from the generic
288 data depicts a limited distribution pattern for CMs < 20 μm with occurrences recorded in
289 coastal regions, and in temperate and polar seas (Figure 4a). In contrast, studies which have
290 targeted these species show their distribution to largely encompass open oceans, including
291 regions closer to the Equator and within oligotrophic gyres in the Atlantic and the Pacific
292 Oceans (Figure 4b).

293

294 **Discussion**

295 Our results indicate a clear bias in the data for the biogeographic distribution of constitutive
296 mixotrophs (CMs) across and within taxonomic groups, different size spectra, and
297 Longhurst's oceanic regions. This appears primarily a reflection of the difficulties associated
298 with sampling and identification of diagnostic features for small cells in natural water
299 samples. Nevertheless, the wide-scale distributions of CMs of various sizes across the
300 different provinces in the global oceans indicate the importance of considering this mixotroph
301 functional group in studies of protist biogeography and thence in ecology.

302 Our current knowledge of CM species distribution remains incomplete primarily due
303 to various biases in the data within global repositories. Even though it is now recognised that
304 most eukaryote "phytoplankton" groups, with the important exception of diatoms, have the
305 potential to demonstrate mixotrophy (Flynn et al., 2013; Stoecker et al., 2017), clear evidence
306 of mixotrophy has been obtained for fewer than 150 species (Appendix S1, Table S1.1 and
307 Leles et al., 2017). To place this in context, the total number of "phytoplankton" species are
308 (as a guesstimate), often rounded to the nearest thousand within a major grouping (phylum).

309 Our work is the first attempt to document the biogeographies of organisms that are
310 undeniably CMs (Appendix S2, Table S2.1; Figures 1 and 2). During interrogation of the
311 online databases, we assumed that the species detected were originally correctly identified
312 and reported. An allied challenge is determining which species indeed express phagotrophy.
313 A recent study suggests that standard methodologies for detecting phagotrophy in field

314 plankton samples are inappropriate for use on these organisms and has most likely resulted in
315 non-identification of CM activity (Anderson et al., 2017). While Anderson et al. (2017)
316 focussed on nano-CMs, their findings suggest that similar methodological problems occur
317 widely, which could easily result in serious misinformation in the protist records of the
318 different databases.

319 Studies that specifically target CMs have various limitations. Quantitative estimates
320 of abundance are challenging because current methods can only account for mixotrophs that
321 were actively feeding at the time of the sampling/experiment (Safi & Hall, 1999; Gast et al.,
322 2014; Sato, Shiozaki, & Hashihama, 2017). Estimates of bacterivory rates are also
323 problematic because they are based on the assumptions that bacterivory remains constant
324 over the short time scales of incubation studies and that community ingestion rates can be
325 approximated to the average ingestion rate of all feeding individuals (Anderson et al., 2017).
326 In reality, feeding varies over the diel cycle (Tsai, Chin, & Chiang, 2009), and only a small
327 proportion of the total mixotroph assemblage is actively feeding at any time during an
328 experiment (Christaki, Wambeke, & Dolan, 1999; Anderson et al., 2017).

329 The CM species database is biased towards certain species and/or groups that have
330 been extensively investigated due to their perceived environmental importance. For example,
331 studies in the coastal regions covering low-salinity, estuarine, and marine waters have
332 focussed on harmful algal bloom (HAB) or toxic species within the dinophyte taxa
333 (Shumway et al., 2018). Some, if not most, of these HAB species are known mixotrophs
334 (Shumway et al., 2018). Some co-occurring protist species (including competitors, prey, and
335 predators) are also known mixotrophs as well, though many are not often recorded as they are
336 not themselves HAB species. Indeed, of the 80 CM species included within our analysis, 51
337 species were dinophytes, despite the importance of mixotrophy having been well recognized
338 in other taxonomic groups (Gast, McKie-Krisberg, Fay, Rose, & Sanders, 2014; Unrein et al.,
339 2014; Stoecker & Lavrentyev, 2018).

340 Biases among records of CMs may also be particularly high due to differential efforts
341 at sampling CMs of different sizes. It is not common practice to identify to species level in
342 plankton surveys, especially among the smaller flagellate cells which are coincidentally
343 mixotrophic (Sanders & Gast, 2012; Stoecker & Lavrentyev, 2018). While incubation
344 experiments have demonstrated ingestion of prey by these small CMs (< 20 μm), taxonomic
345 identifications using traditional diagnostic approaches has been difficult due to few

346 distinctive features for these species. The available evidence indicates that field studies
347 typically neglect most CMs < 20 μm . However, mixotrophic pico- and nano- flagellates for
348 example, have been shown to be abundant and to play a major role as bacterivores and
349 primary producers in oligotrophic oceans (Zubkov & Tarran, 2008; Hartmann, Zubkov,
350 Scanlan, & Lepère, 2013; Mitra et al., 2014). Yet, global plankton databases have failed to
351 indicate the ubiquity of these organisms across the global oceans, particularly in the open
352 oceans (Figure 4). On the other hand, while according to survey data dinophytes appear to be
353 important in oligotrophic oceans (Figure 1), actually they are relatively minor contributors to
354 the total abundance and activity of mixotrophic flagellates within these areas (Unrein et al.,
355 2014).

356 The strong bias towards larger cells and dinophytes is in part also attributed to
357 traditional sampling and preservation methods that are known to underestimate the
358 abundance of more fragile groups or smaller size fractions within the plankton (Gifford &
359 Caron, 2000; Edwards, Johns, Leterme, Svendsen, & Richardson, 2006). The resultant
360 distorted picture of plankton composition across the oceans (Biard et al., 2016) is typified by,
361 for example, survey data from the historic Continuous Plankton Recorder (CPR), a device
362 best suited (designed) for sampling mesozooplankton and larger (> 250 μm size) robust
363 phytoplankton (Richardson et al., 2006). The CPR cannot capture data for the majority of the
364 non-diatom protist species, an exception being very large dinoflagellates such as CMs within
365 the genus *Tripos* (formally *Ceratium*). Given the changes in our understanding of marine
366 ecology, with recognition of the increasing importance of the microbial loop and mixotrophic
367 protists, the CPR requires augmentation with a device specifically designed for sampling
368 small protists.

369 Recent advances in metagenomics and computational analysis allow identification of
370 the taxonomic and genomic content of marine communities, and investigation of their
371 functional potential (Sunagawa et al., 2015). Nevertheless, we could not exploit such datasets
372 in our analysis. While sequence datasets offer a powerful approach to provide molecular
373 taxonomy in the future, protistologists likely have decades of work ahead of them to match
374 sequence data against morphology/physiology. For the sequence datasets, one of the most
375 problematic issues is converting the data into species information (Bucklin, Lindeque,
376 Rodriguez-Ezpeleta, Albaina, & Lehtiniemi, 2016; Leray & Knowlton, 2016). Within the
377 arena of molecular analysis itself, there are different methodologies for obtaining sequence

378 information, so that forming operational taxonomic units (OTUs) from those data are far from
379 standardized (e.g. Callahan, McMurdie, & Holmes, 2017).

380 In addition, most mixotrophic protists fall into often closely-related groups that are
381 composed of a mixture of species with different nutritional modes, i.e., autotrophic,
382 heterotrophic, and mixotrophic. Although metagenomics may offer a means to identify active
383 mixotrophs, (Rokitta et al., 2011; Yelton et al., 2016), it remains to be fully resolved how to
384 use the potential of metagenomics to differentiate mixotrophs from their autotrophic or
385 heterotrophic counterparts (Santoferrara & McManus, 2017). Further, OTUs far outnumber
386 the number of morphologically-described or nutritionally-described species that have been
387 sequenced; there is presently no way of differentiating the mixotrophs among the many
388 OTUs typically recorded in molecular datasets. Meanwhile experimental studies continue to
389 provide increasing evidence that more and more species previously thought to be “strictly”
390 photoautotrophic phytoplankton are, in fact, CMs (e.g., Hoppenrath and Leander, 2007;
391 Berge et al., 2008; Nézan & Chomérat, 2009; Yoo et al., 2010; Kang et al., 2011; Lim et al.,
392 2015; Jang et al., 2017; Ok et al., 2017; Kamennaya, Kennaway, Fuchs, & Zubkov, 2018).

393 In conclusion, our analyses show clearly that CMs of different taxonomic groups and
394 size ranges are present and active throughout the oceans. The oceanic plankton community
395 contributes an estimated 50% of total planetary primary production. The CMs are a major
396 contributing component of this community. Therefore, it is important that we are aware of the
397 biogeography of these organisms with reference to their physiology, ecology and seasonality.
398 Our findings are highly relevant to plankton biogeography, with implications scaling up to
399 the functioning of food webs and biogeochemical cycles in the global oceans. The challenge
400 now facing marine scientists is to develop widely accepted routine protocols for determining
401 mixotrophic potential and improved survey methods for sampling and identifying the smaller
402 plankton members. It is important that developments in plankton protist ecophysiology,
403 traditional and molecular taxonomy, field experimentation and sampling all remain well
404 aligned during the coming decade to ensure that we adequately and speedily resolve the
405 global importance of oceanic mixotrophy. Without this alignment, our abilities to develop
406 and deploy models to explore the consequences of climate change on oceanic (and thence
407 planetary) processes will be impaired.

408

409 **References**

- 410 Anderson, R., Jürgens, K., & Hansen, P. J. (2017). Mixotrophic phytoflagellate bacterivory
411 field measurements strongly biased by standard approaches: a case study. *Frontiers in*
412 *Microbiology*, *8*, 1–12.
- 413 Berge, T., Hansen, P. J., & Moestrup, Ø. (2008). Prey size spectrum and bioenergetics of the
414 mixotrophic dinoflagellate *Karlodinium armiger*. *Aquatic Microbial Ecology*, *50*, 289–
415 299.
- 416 Biard, T., Stemmann, L., Picheral, M., Mayot, N., Vandromme, P., Hauss, H., ... Not, F.
417 (2016). In situ imaging reveals the biomass of giant protists in the global ocean. *Nature*,
418 *532*, 504–507.
- 419 Bucklin, A., Lindeque, P. K., Rodriguez-Ezpeleta, N., Albaina, A., & Lehtiniemi, M. (2016).
420 Metabarcoding of marine zooplankton: prospects, progress and pitfalls. *Journal of*
421 *Plankton Research*, *38*, 393–400.
- 422 Burkholder, J.M., Glibert, P.M., & Skelton, H.M. (2008). Mixotrophy, a major mode of
423 nutrition for harmful algal species in eutrophic waters. *Harmful Algae*, *8*, 77–93.
- 424 Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should
425 replace operational taxonomic units in marker-gene data analysis. *ISME Journal*, *11*,
426 2639–2643.
- 427 Cermeño, P., de Vargas, C., Abrantes, F., & Falkowski, P. G. (2010). Phytoplankton
428 biogeography and community stability in the ocean. *PLoS ONE*, *5*.
429 doi:10.1371/journal.pone.0010037
- 430 Cermeño, P., & Falkowski, P. (2009). Controls on Diatom Biogeography in the Ocean.
431 *Science*, *325*, 1539–1542.
- 432 Christaki, U., Van Wambeke, F., & Dolan, J. R. (1999). Nanoflagellates (mixotrophs,
433 heterotrophs and autotrophs) in the oligotrophic eastern Mediterranean: Standing stocks,
434 bacterivory and relationships with bacterial production. *Marine Ecology Progress*
435 *Series*, *181*, 297–307.
- 436 de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., ... Karsenti, E.
437 (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science*, *348*, 1–11.
- 438 Dolan, J. R. (2005). An introduction to the biogeography of aquatic microbes. *Aquatic*
439 *Microbial Ecology*, *41*, 39–48.
- 440 Dolan, J. R. (2006). Microbial biogeography? *Journal of Biogeography*, *33*, 199–200.
- 441 Ebach, M. C. (2015). *Origins of biogeography*. New York, NY, Springer.
- 442 Edwards, M., Johns, D., Leterme, S. C., Svendsen, E., & Richardson, A. J. (2006). Regional
443 climate change and harmful algal blooms in the northeast Atlantic. *Limnology and*
444 *Oceanography*, *51*, 820–829.
- 445 Flynn, K. J., Stoecker, D. K., Mitra, A., Raven, J. A., Glibert, P. M., Hansen, P. J., ...
446 Burkholder, J. M. (2013). Misuse of the phytoplankton-zooplankton dichotomy: The
447 need to assign organisms as mixotrophs within plankton functional types. *Journal of*
448 *Plankton Research*, *35*, 3–11.

- 449 Fuhrman, J. A., Steele, J. A., Hewson, I., Schwalbach, M. S., Brown, M. V., Green, J. L., &
450 Brown, J. H. (2008). A latitudinal diversity gradient in planktonic marine bacteria.
451 *Proceedings of the National Academy of Sciences*, *105*, 7774–8.
- 452 Gast, R. J., McKie-Krisberg, Z. M., Fay, S. A., Rose, J. M., & Sanders, R. W. (2014).
453 Antarctic mixotrophic protist abundances by microscopy and molecular methods. *FEMS*
454 *Microbiology Ecology*, *89*, 388–401.
- 455 Gifford, D. J., & Caron, D. A. (2000). Sampling, preservation, enumeration and biomass of
456 marine protozooplankton. In R. Harris, P. Wiebe, J. Lenz, H. R. Skjoldal, & M. Huntley
457 (Eds.), *ICES Zooplankton Methodology Manual Book* (pp. 193–221). Elsevier Ltd.
- 458 Glibert, P.M., & Legendre, C. (2006). The diverse nutritional strategies of harmful algae:
459 focus on osmotrophy. In E. Granéli & J. Turner (Eds.), *The Ecology of Harmful Algae*
460 (pp. 163-175). Springer-Verlag.
- 461 Hartmann, M., Zubkov, M. V., Scanlan, D. J., & Lepère, C. (2013). In situ interactions
462 between photosynthetic picoeukaryotes and bacterioplankton in the Atlantic Ocean:
463 Evidence for mixotrophy. *Environmental Microbiology Reports*, *5*, 835–840.
- 464 Hoppenrath, M., & Leander, B. S. (2007). Character evolution in polykrikoid dinoflagellates.
465 *Journal of Phycology*, *43*, 366–377.
- 466 Jang, S. H., Jeong, H. J., Kwon, J. E., & Lee, K. H. (2017). Mixotrophy in the newly
467 described dinoflagellate *Yihiella yeosuensis*: A small, fast dinoflagellate predator that
468 grows mixotrophically, but not autotrophically. *Harmful Algae*, *62*, 94–103.
- 469 Kamennaya, N. A., Kennaway, G., Fuchs, B. M., & Zubkov, M. V. (2018). “Pomacytosis”—
470 Semi-extracellular phagocytosis of cyanobacteria by the smallest marine algae. *PLoS*
471 *Biology*, *16*, 1–13.
- 472 Kang, N. S., Jeong, H. J., Yoo, Y. Du, Yoon, E. Y., Lee, K. H., Lee, K., & Kim, G. (2011).
473 Mixotrophy in the newly described phototrophic dinoflagellate *Woloszynskia cincta*
474 from western Korean waters: Feeding mechanism, prey species and Effect of prey
475 concentration. *Journal of Eukaryotic Microbiology*, *58*, 152–170.
- 476 Legendre, P., & Legendre, L. (1998). *Numerical Ecology* (Second). Amsterdam: Elsevier.
477 Retrieved from <http://www.onmyphd.com/?p=legendre.fenchel.transform>
- 478 Leles, S. G., Mitra, A., Flynn, K. J., Stoecker, D. K., Hansen, P. J., Calbet, A., ... Våge, S.
479 (2017). Oceanic protists with different forms of acquired phototrophy display
480 contrasting biogeographies and abundance. *Proceedings of The Royal Society B*, *284*,
481 20170664.
- 482 Leray, M., & Knowlton, N. (2016). Censusing marine eukaryotic diversity in the twenty-first
483 century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*.
484 doi:10.1098/rstb.2015.0331
- 485 Lim, A. S., Jeong, H. J., Kim, J. H., Jang, S. H., Lee, M. J., & Lee, K. (2015). Mixotrophy in
486 the newly described dinoflagellate *Alexandrium pohangense*: A specialist for feeding on
487 the fast-swimming ichthyotoxic dinoflagellate *Cochlodinium polykrikoides*. *Harmful*
488 *Algae*, *49*, 10–18.
- 489 Lin, S., Zhang, H., Hou, Y., Miranda, L., & Bhattacharya, D. (2006). Development of a
490 dinoflagellate-oriented PCR primer set leads to detection of picoplanktonic
491 dinoflagellates from Long Island Sound. *Applied and Environmental Microbiology*, *72*,

492 5626–5630.

493

494 Longhurst, A. (2007). *Ecological Geography of the sea* (Second Edi). MA: Academic Press.

495 Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., ... Bowler, C.
 496 (2016). Insights into global diatom distribution and diversity in the world's ocean.
 497 *Proceedings of the National Academy of Sciences*, *113*, E1516–E1525.

498 Mitra, A., & Flynn, K. J. (2010). Modelling mixotrophy in harmful algal blooms: More or
 499 less the sum of the parts? *Journal of Marine Systems*, *83*, 158–169.

500 Mitra, A., Flynn, K. J., Burkholder, J. M., Berge, T., Calbet, A., Raven, J. A., ... Zubkov, M.
 501 V. (2014). The role of mixotrophic protists in the biological carbon pump.
 502 *Biogeosciences*, *11*, 995–1005.

503 Mitra, A., Flynn, K. J., Tillmann, U., Raven, J. A., Caron, D., Stoecker, D. K., ... Lundgren,
 504 V. (2016). Defining Planktonic Protist Functional Groups on Mechanisms for Energy
 505 and Nutrient Acquisition: Incorporation of Diverse Mixotrophic Strategies. *Protist*, *167*,
 506 106–120.

507 Moon-Van Der Staay, S. Y., De Wachter, R., & Vault, D. (2001). Oceanic 18S rDNA
 508 sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature*, *409*, 607–
 509 610.

510 Nézan, E., & Chomérat, N. (2009). *Fragilidium duplocampanaeforme* sp. nov.
 511 (Dinophyceae): A new phagotrophic dinoflagellate from the French Atlantic coast.
 512 *European Journal of Protistology*, *45*, 2–12.

513 Ok, J. H., Jeong, H. J., Lim, A. S., & Lee, K. H. (2017). Interactions between the mixotrophic
 514 dinoflagellate *Takayama helix* and common heterotrophic protists. *Harmful Algae*, *68*,
 515 178–191.

516 Pierce, R. W., & Turner, J. T. (1993). Global biogeography of marine tintinnids. *Marine*
 517 *Ecology Progress Series*, *94*, 11–26.

518 R Core Team (2017). R: A language and environment for statistical computing. Vienna,
 519 Austria: R Foundation for Statistical Computing.

520 Richardson, A. J., Walne, A. W., John, A. W. G., Jonas, T. D., Lindley, J. A., Sims, D. W.,
 521 ... Witt, M. (2006). Using continuous plankton recorder data. *Progress in*
 522 *Oceanography*, *68*, 27–74.

523 Rokitta, S. D., de Nooijer, L. J., Trimborn, S., de Vargas, C., Rost, B., & John, U. (2011).
 524 Transcriptome analyses reveal differential gene expression patterns between the life-
 525 cycle stages of *Emiliania huxleyi* (haptophyta) and reflect specialization to different
 526 ecological niches. *Journal of Phycology*, *47*, 829–838.

527 Safi, K. A., & Hall, J. A. (1999). Mixotrophic and heterotrophic nanoflagellate grazing in the
 528 convergence zone east of New Zealand. *Aquatic Microbial Ecology*, *20*, 83–93.

529 Santoferrara, L. F., & McManus, G. B. (2017). Integrating dimensions of biodiversity in
 530 choreotrichs and oligotrichs of marine plankton. *European Journal of Protistology*, *61*,
 531 323–330.

532 Sato, M., Shiozaki, T., & Hashihama, F. (2017). Distribution of mixotrophic nanoflagellates

- 533 along the latitudinal transect of the central North Pacific. *Journal of Oceanography*, 73,
534 159–168.
- 535 Shumway, S. E., Burkholder, J. A. M., Morton, S. L. (2018). *Harmful algal blooms: a*
536 *compendium desk reference*. Hoboken, NJ, John Wiley & Sons.
- 537 Smayda, T. J. (1958). Biogeographical studies of marine phytoplankton. *Oikos*, 9, 158–191.
- 538 Stoecker, D. K., Hansen, P. J., Caron, D. A., & Mitra, A. (2017). Mixotrophy in the Marine
539 Plankton. *Annual Review of Marine Science*, 9, annurev-marine-010816-060617.
- 540 Stoecker, D.K., Lavrentyev, P.J., 2018. Mixotrophic Plankton in the Polar Seas: A Pan-Arctic
541 Review. *Front. Mar. Sci.* 5
- 542 Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., ... Bork,
543 P. (2015). Structure and function of the global ocean microbiome. *Science*, 348.
544 doi:10.1126/science.1261359
- 545 Tomas, C. R. (1997). *Identifying marine phytoplankton*. Academic Press, San Diego,
546 California.
- 547 Tsai, A.-Y., Chin, W.-M., & Chiang, K.-P. (2009). Diel patterns of grazing by pigmented
548 nanoflagellates on *Synechococcus* spp. In the coastal ecosystem of subtropical western
549 pacific. *Hydrobiologia*, 636, 249–256.
- 550 Unrein, F., Gasol, J. M., Not, F., Forn, I., & Massana, R. (2014). Mixotrophic haptophytes
551 are key bacterial grazers in oligotrophic coastal waters. *The ISME Journal*, 8, 164–176.
- 552 Wilken, S., Huisman, J., Naus-Wiezer, S., & Van Donk, E. (2013). Mixotrophic organisms
553 become more heterotrophic with rising temperature. *Ecology Letters*, 16, 225–233.
- 554 Yelton, A. P., Acinas, S. G., Sunagawa, S., Bork, P., Pedrós-Alió, C., & Chisholm, S. W.
555 (2016). Global genetic capacity for mixotrophy in marine picocyanobacteria. *ISME*
556 *Journal*, 10, 2946–2957.
- 557 Yoo, Y. Du, Jeong, H. J., Kang, N. S., Kim, J. S., Kim, T. H., & Yoon, E. Y. (2010). Ecology
558 of *Gymnodinium aureolum*. II. Predation by common heterotrophic dinoflagellates and a
559 ciliate. *Aquatic Microbial Ecology*, 59, 257–272.
- 560 Zubkov, M. V., & Tarran, G. A. (2008). High bacterivory by the smallest phytoplankton in the
561 North Atlantic Ocean. *Nature*, 455, 224–226.

562

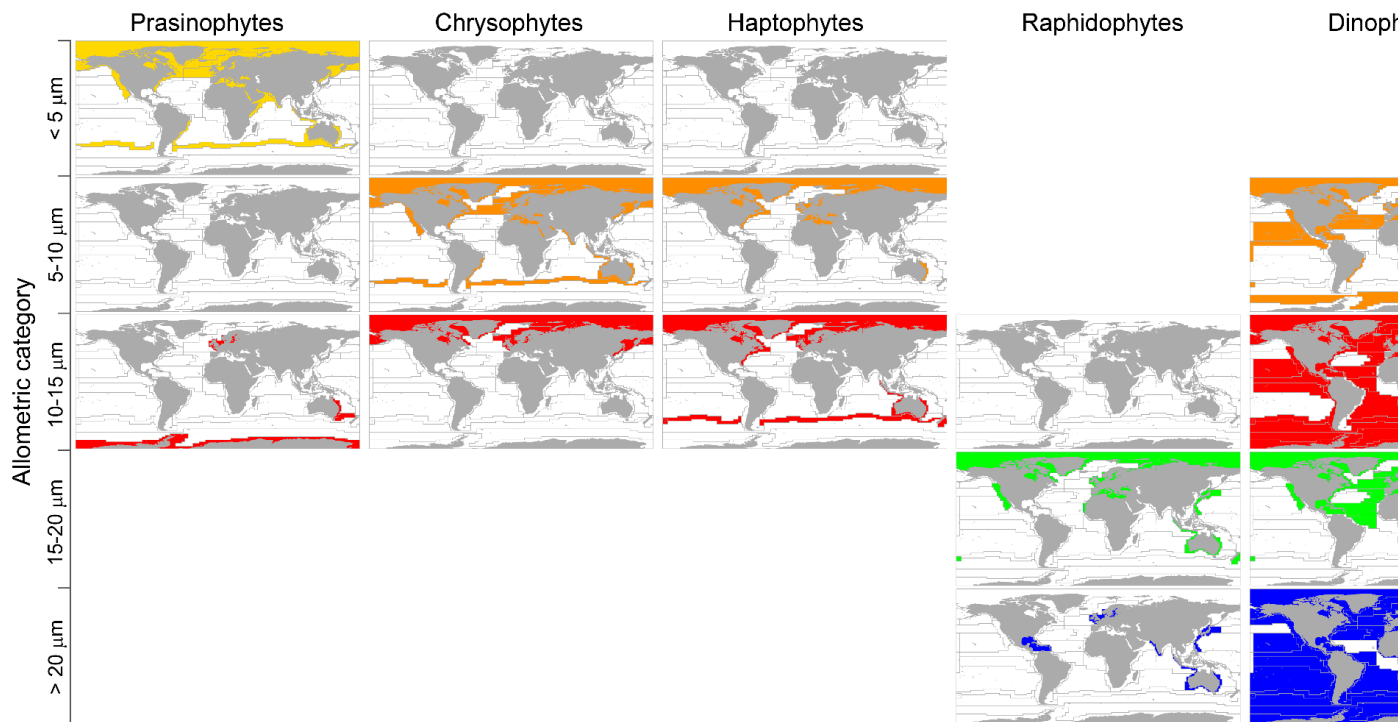
563 **Data Accessibility**

564 CM species and the geographic coordinates of studies which specifically targeted CMs are
565 collated within the Supplementary Information (Appendix S1, Table S1.1 and Appendix S2,
566 Table S2.2, respectively). Records for different CM species are available as open access
567 within the OBIS database (<http://www.iobis.org/>); we also provide this information in

568 Appendix S2, Table S2.1. Grids of Longhurst's biogeographic provinces (Appendix S1,
569 Table S1.2) can be downloaded from <http://www.marineregions.org/>.

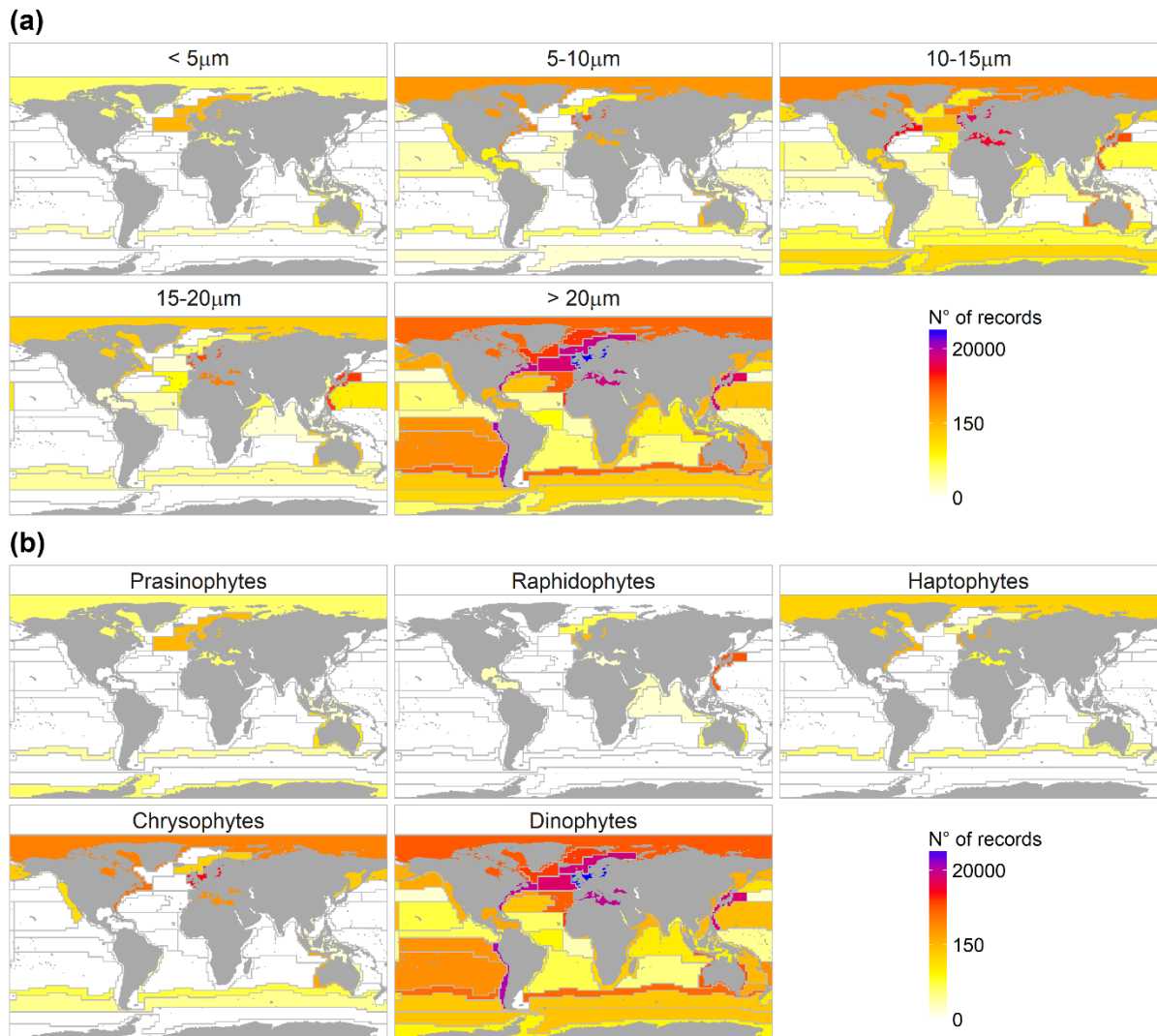
570

571 **Figures**
572
573



574
575
576 FIGURE 1 Global distribution of CMs across Longhurst's biogeographic provinces. Distribution maps are shown for five groups (prasinophytes, raphidophytes, haptophytes, chrysophytes, and dinophytes) across different size classes (maximum cell diameter: yellow, <math>< 5 \mu\text{m}</math>; orange, 5–10 $\mu\text{m}</math>; red, 10–15 $\mu\text{m}</math>; green, 15–20 $\mu\text{m}</math>; blue, > 20 $\mu\text{m}</math>). Colour-cast provinces indicate the presence of species in that province. White spaces indicate no data. The absence of maps (i.e., white spaces) indicate that there are no known species of that size class in that province.
579
580 See also Appendix S2, Table S2.1.$$$$

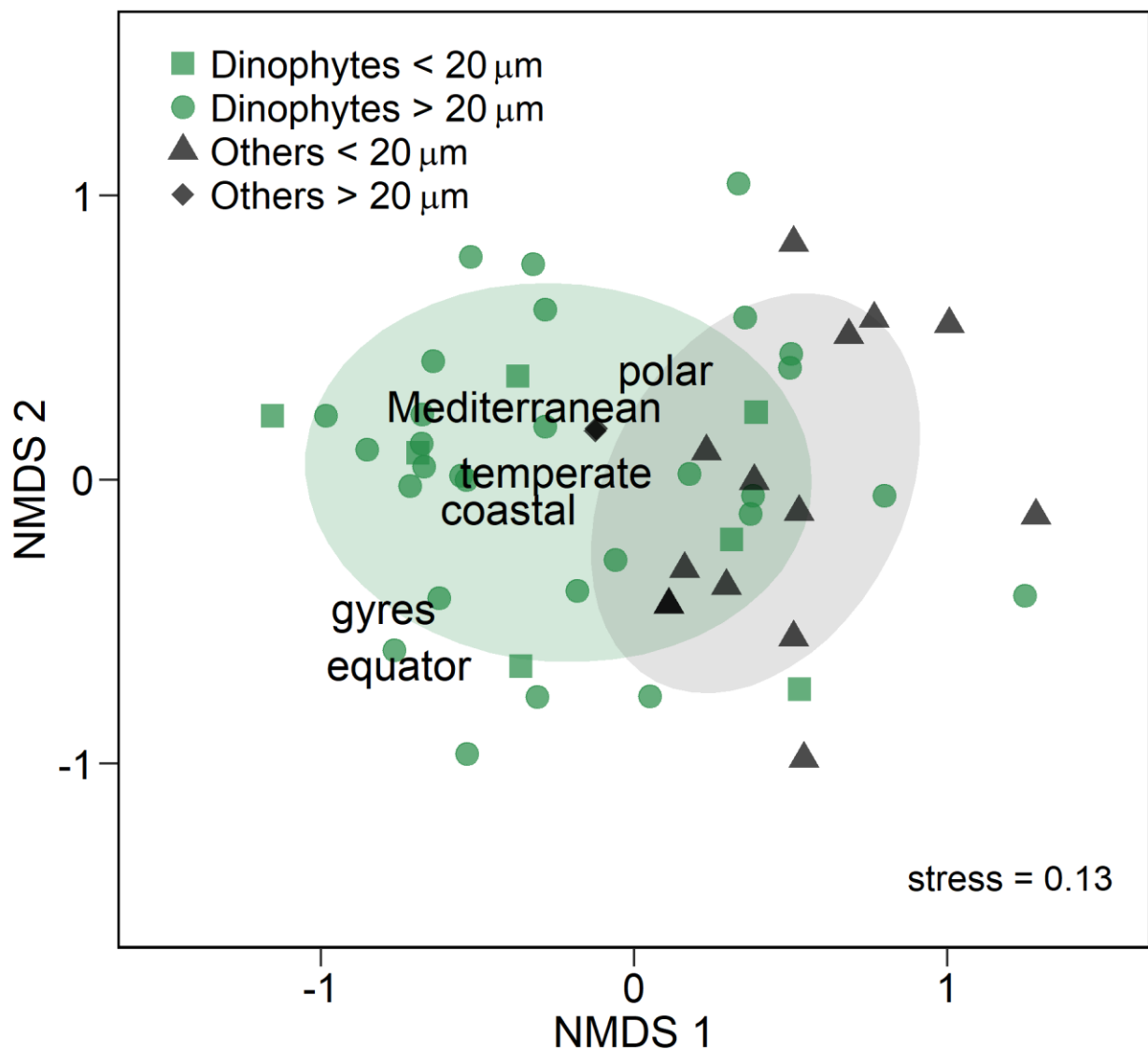
581



582
583

584 FIGURE 2 Number of records for CMs across different biogeographic provinces (Appendix
585 S1, Table S1.2 and Appendix S2, Table S2.1). Number of records provided for: (a) different
586 size-classes (μm , length), and, (b) different taxonomic groups. See also Appendix S3, Figure
587 S3.2.

588
589
590

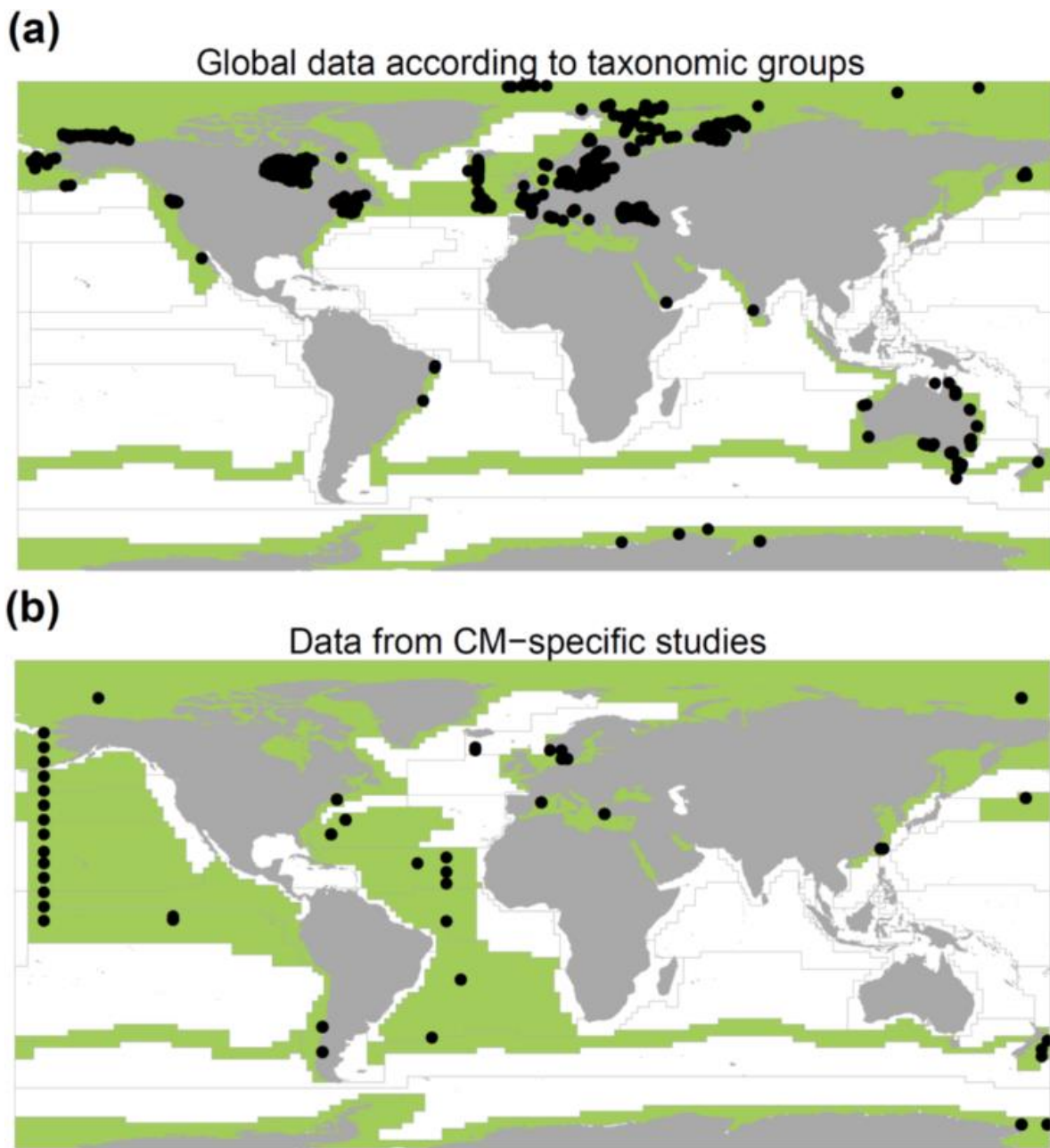


592
593

594 FIGURE 3 Spatial distribution of CMs from NMDS analysis. NMDS ordination was based
 595 on the number of records observed for each species within each biogeographic province;
 596 biomes were primarily derived from provinces (Appendix S3, Figure S3.3). Each symbol
 597 represents a species which was grouped by taxonomy (different colours) and by size and
 598 taxonomy (different symbols). Ellipses are shown at 75% confidence interval and were used
 599 to define dinophytes (green) and other groups (grey).

600
601

602
603



604
605

606 FIGURE 4 Global distribution of CMs with maximum cell dimension $< 20 \mu\text{m}$. (a) Records
607 obtained from OBIS (Appendix S2, Table S2.1); (b) Records obtained from studies of protist
608 nanoplankton bacterivory (Appendix S2, Table S2.2). Symbols correspond to the exact
609 location where CMs were found. Colour-cast provinces indicate the presence of CMs while
610 white provinces correspond to absence due to lack of data; the grid indicate Longhurst's
611 biogeographic provinces (Appendix S1, Table S1.2).

612
613

614 **Appendix 1:** Reference list of studies which specifically targeted CMs but not cited within
615 the main text (see also Appendix S2; Table S2.2)

616 Arenovski, A. L., Lim, E. L., & Caron, D. A. (1995). Mixotrophic nanoplankton in
617 oligotrophic surface waters of the Sargasso Sea may employ phagotrophy to obtain
618 major nutrients. *Journal of Plankton Research*, *17*, 801–820.

619 Czypionka, T., Vargas, C. A., Silva, N., Daneri, G., González, H. E., & Iriarte, J. L. (2011).
620 Importance of mixotrophic nanoplankton in Aysén Fjord (Southern Chile) during austral
621 winter. *Continental Shelf Research*, *31*, 216–224.

622 Hall, J. A., Barrett, D. P., & James, M. R. (1993). The importance of phytoflagellate,
623 heterotrophic flagellate and ciliate grazing on bacteria and picophytoplankton sized prey
624 in a coastal marine environment. *Journal of Plankton Research*, *15*, 1075–1086.

625 Hartmann, M., Grob, C., Tarran, G. Martin, A. P., Burkill, P. H., Scanlan, D. J. & Zubkov,
626 M. A. (2012). Mixotrophic basis of Atlantic oligotrophic ecosystems. *Proceedings of the*
627 *National Academy of Sciences*, *109*, 5756–5760.

628 Havskum, H., & Hansen, A. S. (1997). Importance of pigmented and colourless nano-sized
629 protists as grazers on nanoplankton in a phosphate-depleted Norwegian fjord and in
630 enclosures. *Aquatic Microbial Ecology*, *12*, 139–151.

631 Havskum, H., & Riemann, B. (1996). Ecological importance of bacterivorous, pigmented
632 flagellates (mixotrophs) in the Bay of Aarhus, Denmark. *Marine Ecology Progress*
633 *Series*, *137*, 251–263.

634 Moorthi, S., Caron, D. A., Gast, R. J., & Sanders, R. W. (2009). Mixotrophy: A widespread
635 and important ecological strategy for planktonic and sea-ice nanoflagellates in the Ross
636 Sea, Antarctica. *Aquatic Microbial Ecology*, *54*, 269–277.

637 Nygaard, K., & Tobiesen, A. (1993). Bacterivory in algae: A survival strategy during nutrient
638 limitation. *Limnology and Oceanography*, *38*, 273–279.

639 Sanders, R. W., Berninger, U. G., Lim, E. L., Kemp, P. F., & Caron, D. A. (2000).
640 Heterotrophic and mixotrophic nanoplankton predation on picoplankton in the Sargasso
641 Sea and on Georges Bank. *Marine Ecology Progress Series*, *192*, 103–118.

642 Sanders, R. W., & Gast, R. J. (2012). Bacterivory by phototrophic picoplankton and
643 nanoplankton in Arctic waters. *FEMS Microbiology Ecology*, *82*, 242–253.

- 644 Stukel, M. R., Landry, M. R., & Selph, K. E. (2011). Nanoplankton mixotrophy in the eastern
645 equatorial Pacific. *Deep-Sea Research Part II: Topical Studies in Oceanography*, 58,
646 378–386.
- 647 Tsai, A. Y., Gong, G. C., Sanders, R. W., Chen, W. H., Chao, C. F., & Chiang, K. P. (2011).
648 Importance of bacterivory by pigmented and heterotrophic nanoflagellates during the
649 warm season in a subtropical western Pacific coastal ecosystem. *Aquatic Microbial
650 Ecology*, 63, 9–18.
- 651 Unrein, F., Massana, R., Alonso-Sáez, L., & Gasol, J. M. (2007). Significant year-round
652 effect of small mixotrophic flagellates on bacterioplankton in an oligotrophic coastal
653 system. *Limnology and Oceanography*, 52, 456–469.
- 654 Vargas, C., Contreras, P., & Iriarte, J. (2012). Relative importance of phototrophic,
655 heterotrophic, and mixotrophic nanoflagellates in the microbial food web of a river-
656 influenced coastal upwelling area. *Aquatic Microbial Ecology*, 65, 233–248.

657

658 **Supporting Information**

659 **Appendix S1** List of CM species and biogeographic classification of the ocean.

660 **Appendix S2** Data compiled from OBIS and from studies which specifically targeted CMs.

661 **Appendix S3** Supplementary Figures.

662 **Appendix S4** Supplementary results for the post-hoc pairwise comparisons using Dunn's
663 test.

664

665 **Supporting information is provided in a separate Word document (Supporting
666 Information – Leles et al) and two Excel files (TableS1.1.xlsx and TableS2.2.xlsx).**

667

668