Title: Prevalence and preferred niche of small eukaryotes with mixotrophic potentials in the global ocean

Running head: Prevalence of mixotrophs in global oceans

Kaiyi Dong[1], Ying Wang[1], Wenjing Zhang[1], Qian Li[1,2*]

[1]School of Oceanography, Shanghai Jiao Tong University, 1954 Huashan Road, Shanghai 200030, China.

[2] Key Laboratory of Polar Ecosystem and Climate Change (Shanghai Jiao Tong University), Ministry of Education, 1954 Huashan Road, Shanghai 200030, China.

*Correspondence: liqian00@sjtu.edu.cn

Keywords: mixotroph, heterotroph, autotroph, eukaryotes, niche, competition, *Tara Oceans*, 18S rRNA gene

Data Availability

All dataset used in this study were archived with Dryad (https://doi.org/10.5061/dryad.612jm649h). Original metabarcode and environmental context data were publicly deposited by the *Tara Oceans* Project, under Pangaea website (https://doi.pangaea.de/10.1594/PANGAEA.840721) and European Nucleotide Archive (project accession #PRJEB6610).

Funding

This study was supported by the Young Scientists Fund granted by the National Natural Science Foundation of China (Grant No. 42106097), and the Oceanic Interdisciplinary Program of Shanghai Jiao Tong University (project No. SL2021PT104).

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We thank the *Tara Oceans* project and their team for providing the original database of eukaryotic 18S rRNA metabarcodes and environmental parameters. We also acknowledge the funding resource of Young Scientists Fund of the National Natural Science Foundation of China, and the Oceanic Interdisciplinary Program of Shanghai Jiao Tong University.

Abstract

Unicellular eukaryotes that are capable of phago-mixotrophy in the ocean compete for inorganic nutrients and light with autotrophs, and for bacterial prey with heterotrophs. In this study, we ask what are the overall prevalence of eukaryotic mixotrophs in vast open oceans, and how the availability of inorganic nutrients, light, and prey affects their relative success. We utilized the *Tara Oceans* eukaryotic 18S rRNA gene and environmental context variables dataset to conduct large-scale field analysis. We also performed isolate-based culture experiments to verify growth and nutritional resource relationships for representative mixotrophic taxa. The field analysis suggested that the overall prevalence of mixotrophs was negatively correlated with nutrient concentrations and positively associated with light availability. Concentrations of bacterial prey also presented somewhat correlation with mixotrophs but to a less extent. In comparison, the culture experiments demonstrated a taxa-specific relationship between mixotrophic growth and nutrition resources, i.e., the growth of one group was significantly dependent on light availability while the other group was less affected by light when they received sufficient prey. Both groups were capable of growing efficiently with low inorganic nutrients when receiving sufficient prey and light. Therefore, our field analysis and culture experiments both suggest that phago-mixotrophy for ocean eukaryotes is seemingly to be an efficient strategy to compensate nutrient deficiency but unnecessarily light limitation. This study collectively revealed a close relationship between abiotic and biotic nutritional resources and the prevalence of trophic strategies, shedding light on the importance of light and nutrients for determining the competitive success of mixotrophs versus autotrophic and heterotrophic eukaryotes in the ocean.

Introduction

Pico- and nano-sized small eukaryotes (<3-20 μm) represent the dominant microbial communities in ocean food webs. Through metabolic activities such as photosynthesis and phagocytosis, they regulate the flux of elements and energy between primary producers and consumers. Traditionally, all protists with inherent chloroplasts were considered strict autotrophs (only photosynthesize), while those without inherited chloroplasts were regarded as strict heterotrophs (only phagocytose). It has become increasingly evident that many of those autotrophs/heterotrophs are in fact mixotrophs, phagotrophic mixotrophs particularly, they are capable of both photosynthesis and phagocytosis [1,2,3]. Depending on whether the organism possesses its own chloroplasts, mixotrophs can be further categorized into constitutive mixotrophs and non-constitutive mixotrophs. Constitutive mixotrophs, the focus of the current study, are normally small-size flagellated eukaryotes that primarily feeding on bacterial prey [4,5], whereas non-constitutive mixotrophs are larger-sized microplankton (20-200 μm) which graze on both eukaryotic and bacterial prey [6,7]. Due to the distinct metabolic preference and trophic activities, those two mixotrophic groups should be studied separately.

The relative abundance and prevalence for mixotrophs (comparing to other eukaryotes) in a given system can be partially interpreted as the competition results among mixotrophs, autotrophs, and heterotrophs. Specifically, mixotrophs and autotrophs compete for inorganic nutrients (hereinafter referred to as nutrients) and light for autotrophic growth, and for

bacterial prey (including heterotrophic and autotrophic bacteria) with bacterivorous heterotrophs for heterotrophic growth. Although results may vary, previous studies have suggested that the physiological performance and competitive advantage of mixotrophs can be related to these three nutritional resources of nutrients, light and prey [8,9,10,11]. Mixotrophy are most likely benefited from scenarios where nutrition is not sufficient to support a single trophic strategy, i.e., strict autotroph or heterotroph. For example, Ward (2019) has proposed a conceptual model arguing that mixotrophy can effectively alleviate dissolved inorganic nitrogen deficiency in oligotrophic environments via phagocytosis of high-nitrogen bacterial prey [12]. Meanwhile, it is necessary to acknowledge that the photosynthesis and/or phagocytosis performance alone is very likely penalized for mixotrophs [13], and these tradeoffs can differentiate significantly among different species. Li et al. (2022) have demonstrated a distinct variation for grazing/heterotrophic capability and trophic efficiency among different mixotrophic strains [5].

In regional areas of temperate sea [14,15,16] and oligotrophic ocean [11,17,18] or lakes [19,20,21], field observations have also shown that mixotrophy can be advantageous for phytoplankton community growth when nutrition are not ideal for autotrophy, which are mostly due to limiting nutrients and/or irradiance. In other words, conditions that are unfavorable for strict autotrophy or heterotrophy could possibly select for mixotrophy on the entire community level. Yet for larger spatial scales that encompass multiple ocean regions, there has been few studies to look into the broad-scale relationship between mixotrophic prevalence and environmental conditions related to nutrition availability. While acknowledging these relationships can be complex due to multi-factorial interactions among biotic and abiotic components in the food web, it is reasonable to speculate nutrients, light and prey can play an important role determining the relative success of mixotrophy. With climate change, evidence has shown that mixotrophs may play a more important role due to their metabolic plasticity, i.e., easier to adapt to changing environments [22,23,24,25]. Motivated by these findings, we are particularly interested to delineate the overall prevalence and realized niche favoring mixotrophs in comparison to strict autotrophs and heterotrophs in vast open oceans across the globe.

The broad goal of this study is to gain insights into the interrelationship between environmental conditions and trophic strategies on both community and species level, as well as the underlying mechanism behind it. We aim to answer the following scientific questions, 1) What are the overall prevalence of mixotrophic eukaryotes in open oceans? 2) Are there specific environments under which mixotrophs are favored over autotrophs and heterotrophs; 3) If so, what role does nutrients, light and bacterial prey play for determining the relative success and fitness of mixotrophy? We use the eukaryotic 18S rRNA gene dataset from *Tara Oceans Expedition* to assign trophic mode to each lineage/species and assess the relationship between trophic mode composition and contextual variables. Subsequently, we run culture experiments to confirm relationships between representative mixotrophic isolates and key nutritional resource variables. The *Tara Oceans* analysis revealed statistically important variables that are affiliated with the trophic mode prevalence, and the experimental results provided detailed and physiological evidence for it.

Materials and Methods

Study regions

The data generated by *Tara Oceans* expedition between 2009 and 2013 are particularly suited to study the genetic and functional diversity of plankton. In this study, a total of 104 stations and 165 samples were investigated, among which all 104 sites had surface samples and 61 sites had both surface and chlorophyll maximum layer (CML) samples. A surface dataset (SUR) consisting of 104 samples was used, while the surface and chlorophyll maximum depth dataset (SUR&CML) comprising 122 (the 61 stations have both SUR&CML samples) and 165 samples (all 104 SUR samples and 61 CML sample), was generated for different statistical analysis. These samples were collected during four different seasons (Spring, Summer, Autumn, Winter) between 2009 and 2013 from eight oceanic regions (Northern Atlantic Ocean, Mediterranean Sea, Red Sea, Indian Ocean, Southern Atlantic Ocean, Southern Ocean, Southern Pacific Ocean and Northern Pacific Ocean) with broad latitudinal (64.4ºS-43.7ºN) and longitudinal ranges (159.0ºW-73.9ºE). Bathymetry depths varied from 13 to 5964 km and distance to coast was between 0.9-1404 km.

Contextual environmental variables

We built an environmental context dataset for further statistical analysis following these three steps. First, representative environmental variables were obtained from both the Tara Oceans PANGAEA data repository (http://www.pangaea.de) and a variety of publications, including Pesant et al. (2015), Picheral et al. (2017), Ardyna et al. (2017), Ibarbalz et al. (2019), Faure et al. (2019), Karlusich et al. (2022) etc. (**Supplementary Dataset S1**) [26,27,28,29,30,31]. The major approaches for retrieving these variables included in situ sensors, laboratory analysis of water samples, climatological instruments and model simulations, following standard protocols and analytical methods (details are described in those aforementioned literatures). Second, these ~100 variables were manually filtered down to 77 by eliminating similar variables and less relevant ones, to reduce collinearity and inflation of dimensionality. Examples include salinity was chosen over conductivity, and excluding factors such as moon phase durations and current residual time etc. Lastly, when variables have values from multiple measurements and/or calibration/calculation methods, only the most accurate one or suitable one (for the purpose of this study) was selected. For example, median values for photosynthetically available radiation (PAR), averaged on an 8-day interval and calculated based on attenuation coefficient and surface irradiance observed by AMODIS were selected over a 1-day interval to avoid fluctuations caused by short-term weather conditions. Total chlorophyll *a* concentrations, derived from sensor measurements and calibrated with non-photosynthetic quenching and discrete water sample measurements, were used among other available measurements. Additional comparisons were also conducted to confirm there were no overall significant differences between the selected one and other measurements for the same variable, such as 8-day averaged PAR vs 30-day averaged PAR, total Chl *a* derived from sensor vs water samples vs MODIS, and nutrient concentrations derived from water samples and model simulation, etc.

The final contextual environmental dataset comprised 29 variables (**Supplementary Dataset S2)**, representing physical conditions such as temperature, salinity, density, attenuation coefficient of PAR (Kd PAR), chlorophyll maximum layer (CML), mixed layer depth (MLD), nitracline depth and euphotic depth; climatology such as PAR and sunshine duration; biochemical conditions such as total Chl *a*, oxygen, nutrients (NO_3^-&NO_2^-, NO_2^-, PO_4^{3-}, SiO_4^-), pH, total alkalinity, total carbon, fluorescence, particulate organic carbon

166 (POC), particulate inorganic carbon (PIC), and bacterial abundances including autotrophic
167 (*Prochlorococcus* and *Synechococcus*) and heterotrophic bacteria. Note that longitudes,
168 latitudes, distance to coasts, sampling depths and bathymetry depths were not included in the
169 analysis to better focus on the aforementioned climatological, biochemical and physical
170 variables.

171 **18S rRNA gene abundances and corrected cell abundances**

172 Detailed genomic sample processing methods were described in de Vargas et al. (2015)
173 [32]. Briefly, community genomic DNA was extracted from cells filtered through membrane
174 filters with different pore sizes (0.8µm, 5µm, 20µm) after pre-filtration with nylon sieves (to
175 retrieve different plankton communities). In this study, we focused on the 0.8-5µm size range,
176 which contributed the majority of eukaryotic community in sampled regions [32]. The
177 eukaryotic V9-18S rRNA gene region was sequenced through Illumina HiSeq, assembled,
178 filtered and classified with SWARM for operational taxonomic units (OTUs). The sequence
179 files were downloaded from the European Nucleotide Archive under project number
180 PRJEB6610. Subsequently, we re-assigned OTUs to the lowest taxonomic rank possible,
181 referring to the Protist Ribosomal Reference (PR^2) database [33]. Ciliophora, Radiolaria, and
182 Foraminifera that were likely derived from large-sized populations (contributing a small
183 portion of the total species), were excluded from further analysis to focus on small-sized
184 constitutive mixotrophs. Unkown Eukaryota and a small portion of parasites (e.g., Fungi,
185 Amoebozoa, Syndiniales, Dinoflagellata) were excluded from further trophic annotation, and
186 thus further analysis. This final gene dataset has generated 40,126,831 sequences and 845
187 lineages/species for surface samples, and 17,874,753 sequences and 803 lineages/species for
188 the CML samples, respectively (**Supplementary dataset S3**).

189 Eukaryotic communities are known to exhibit substantially different 18S rRNA gene
190 copy numbers, varying from as low as one copy in species belonging to Ochrophyta to
191 thousands of copies per cell in dinoflagellates [5,34]. Overall, cellular biovolumes have been
192 shown to be positively correlated with 18S rRNA gene copy numbers but certain taxa, such
193 as dinoflagellates, tended to have particularly high copy numbers [35,36]. To correct for this
194 bias and obtain gene-converted-cell abundances, we first collected empirical data from
195 plankton, focusing on small cells with equivalent diameter of 5-10 µm and biovolume of
196 65-523 μm^3 (**Supplementary Table S1**). Then, we fit a linear correlation equation between
197 empirical 18S rRNA gene copy numbers and cellular biovolume for individual group of
198 dinoflagellates, diatoms, and other eukaryotes (including chlorophytes, haptophytes,
199 dictyochophytes, etc.), respectively. Finally, a correction factor (C.F.) for each group was
200 calculated using upper limit biovolume of 65 μm^3, that gives 59.3, 10.5, and 4.7 for
201 dinoflagellates, diatoms, and other eukaryotes, respectively. For comparison, Martin et al.
202 (2022) used 27.1, 4.4, and 0.9 C.F. for dinoflagellates, diatoms, and all other eukaryotes [36].
203 Gene abundances were converted into cell abundances by dividing gene numbers with C.F.
204 for each group. The corrected cell abundance was the focus for further analysis, but 18S
205 rRNA gene abundances were also assessed occasionally for comparison purposes.

206 **Assignment of trophic mode**

207 In this study, we focus on the trophic strategy potentials among different eukaryotic
208 lineages, rather than the in situ active trophic activities (are partially associated with
209 instantaneous conditions). Therefore, we assigned each lineage/species (when possible) into

210 three trophic groups by searching published database and references [5,37,38], including
211 (potential) mixotrophs (pigmented eukaryotes with phagocytosis capability), heterotrophs
212 (eukaryotes with no inherent chloroplasts) and autotrophs (pigmented eukaryotes with no
213 available evidence for phagocytosis). When trophic annotation was impossible at lineage or
214 species level, their affiliated genus or family was referred to for annotation. The filtration has
215 generated a final composition of 184-186 mixotrophic, 283-313 heterotrophic, and 303-377
216 autotrophic lineages/species for the SUR and SUR&CML samples, respectively
217 (**Supplementary dataset S3**). We acknowledge the limitation of this trophic annotation
218 methodology, and address it further in the Discussion.

219 **Trophic index and trophic composition**

220 The relative prevalence or success of each trophic group at a given site was expressed as
221 the trophic index of each group (TI_g), where *g* represents three trophic groups of mixotroph
222 (M), autotroph (A) and heterotroph (H). Therefore, TI_g of TI_M, TI_A, and TI_H can be quantified
223 as their relative abundance against total eukaryotic community (T), either in cell or gene
224 abundances (expressed as M:T, A:T and H:T). The combined three index of TI_M, TI_A and TI_H
225 can represent the complete results for trophic mode compositions at a given site.

226 **Statistical analysis**

227 The surface dataset was used for redundancy analysis (RDA) to elucidate the overall
228 distribution pattern of trophic groups among 104 stations, and potential relationships with
229 environmental variables. All 29 filtered contextual variables were considered for the analysis,
230 with missing values for each variable replaced with means and then log-transformed for
231 standardization prior to analysis. Hellinger transformation was applied to species abundances
232 and escoufier selection with a threshold of 0.9 was applied to the trophic species.
233 Environmental variables that passed significance test and collinearity test were presented on
234 the RDA plot. The RDA was carried out in R with the vegan package, following the script
235 outlined by Faure et al. (2019) [30]. Principal components analysis (PCA) was applied to
236 extract the major components that explained the environmental gradients from both surface
237 (SUR) and SUR&CML datasets with R package pcaMethods [39]. After extracting the main
238 variable components, TI_g indices were plotted over two axes of PC1 and PC2 with
239 generalized additive model. Stepwise regression between all filtered variables and TI_M were
240 conducted to retrieve important variables explaining mixotrophic prevalence, and linear
241 regression between single variables and TI_g were carried out to provide additional correlation
242 results. ANOVA was conducted for the culture experiments to reveal factorial/treatment
243 impact on growth rates and grazing rates among various mixotrophic isolates.

244 **Culture experiments**

245 Methods for isolating and cultivating mixotrophs from the North Pacific Subtropical
246 Gyre were described in Li et al. (2021; 2022) [5,40]. Briefly, mixotrophic eukaryotes were
247 isolated from the euphotic zone at Station ALOHA (22° 45' N, 158° 00' W) by enriching
248 seawater samples in nitrogen-depleted Keller medium (K-N medium) that supplemented with
249 bacterial prey. Six identified isolates belonging to Dictyochophyceae, Prymnesiophyceae, and
250 Chrysophyceae were selected for conducting growth experiments. These strains were
251 phylogenetically close to the *Tara Oceans* lineages that were among the top 15 abundant
252 lineages. Prior to experiments, all isolates were acclimated for 1-2 months (duplicated over
253 10 generations) under corresponding treatment conditions, i.e., low nutrients (K-N medium)

254 and high light, low nutrients and low light, high nutrients (K medium) and high light, and
255 high nutrients and low light. The high and low light treatments were receiving ~100 and ~10
256 μM photons m^{-2} s^{-1} irradiance on a light/dark cycle (12/12 h). All treatments were
257 supplemented with autotrophic bacterial prey of *Prochlorococcus* (strain MIT9301) at final
258 concentrations of ~2×10^6 cells mL^{-1}, which represents the second abundant prokaryotic
259 population at station ALOHA. Prey cells were centrifuged with a rotation speed of 2000-3000
260 × g for 3–5 min and gently resuspended for the enrichment. Controls without added prey or
261 grazers were also set up at each condition for comparison purposes. Cell concentrations of
262 prey and grazers were measured every 12–24 hours by flow cytometry, and ingestion
263 evidences were occasionally obtained via microscopic observation.
264 Based on our previous study on the grazing functional responses, most isolates have
265 ingestion rates saturated when prey concentration are higher than 10^6 cells mL^{-1}. Meanwhile,
266 clearance rates (ingestion rates divided by prey concentration) can roughly present the
267 maximal clearance rates with the final prey concentration used in these experiments.
268 Therefore, we use clearance rates instead of ingestion rates [1][41] to present grazing
269 performance. Ingestion rates (prey $grazer^{-1}$ h^{-1}) for each grazer were calculated as the amount
270 of prey removed between sampling time t+1 and t, divided by averaged grazer concentrations
271 over the same time period. Clearance rates were calculated by dividing the ingestion rate by
272 averaged prey concentration over the same interval. Body volume specific clearance rates
273 (body volume $grazer^{-1}$ h^{-1}) were calculated by dividing the clearance rate with cellular
274 biovolume (μm^3) of each grazer (measured under microscopy) [42]. Clearance rates reported
275 in the results were averaged over all sampling intervals prior to grazers reaching stationary
276 phase. Growth rates presented in the results were maximal values during exponential growth
277 phase (slope of Ln transformed cell abundance versus time).

278 **Results**

279 **Oceanographic characteristics across the *Tara Oceans* sampling regions**

280 The 104 surface samples encompassed eight ocean habitats on large spatial scale (**Fig.**
281 **1a**). Geographically they span roughly four latitudinal regions, from low (0º-20º) to
282 low-medium (20º-30º), to medium (30º-40º), and medium-high latitude (40º-60º). Moving
283 from low to high latitudinal regions, overall PAR decreased from 25.4 to 14.7 mol photons
284 m^{-2} d^{-1}, nutrients, e.g., NO_3^-&NO_2^- , increased from 0.03 to 17.3 μmol L^{-1}, and heterotrophic
285 bacterial abundances were relatively stable, ranging between 3.9-5.3 $\times10^5$ cells mL^{-1} (**Table 1;**
286 **Fig. 1b**). The globally averaged means were 21.3 mol photons m^{-2} d^{-1} of PAR, 0.14 μmol L^{-1}
287 of NO_3^-&NO_2^- and 4.5 $\times10^5$ cells mL^{-1} of heterotrophic bacteria, demonstrating most
288 sampling sites were oligotrophic environments. Smaller regional variations were also
289 observed. The higher nutrient concentrations in low latitude regions were possibly due to the
290 Peruvian coastal and eastern equatorial upwelling, while lower PAR in medium latitude areas
291 was because Mediterranean Seas were sampled during winter.

292 **Relative abundances and taxonomic compositions of three trophic modes**

293 The trophic index of TI_M, TI_A and TI_H varied from 0.23 to 0.34, 0.10 to 0.32 and 0.43 to
294 0.51 in cell abundance, and from 0.29 to 0.45, 0.08 to 0.24 and 0.38 to 0.46 in gene
295 abundance, respectively (**Table 1; Fig. 1c**). TI_M in cell abundance were overall somewhat
296 lower than gene abundance. Nevertheless, the highest TI_M median values were both retrieved
297 from low-medium latitude, and the lowest values were from low latitude (cell abundances)

298 and medium-high latitudes (gene abundances), respectively. In comparison, high and low TI_A
299 median values were from low and medium-high latitude and low-medium latitude (both cell
300 and gene abundances). TI_H did not change as much across latitudinal gradients, but highest
301 median values in cell and gene abundances were shown in low-medium and medium-high
302 latitudes, respectively.
303 Before the gene-cell abundances correction, mixotrophs were mostly composed of
304 lineages from Dinoflagellata, Haptophyta and Ochrophyta. After the correction for cell
305 abundances, the dominant mixotrophs were retrieved from Haptophyta and Ochrophyta (**Fig.**
306 **1d**). Proportions of dinoflagellates decreased for heterotrophic community after gene-cell
307 conversion, and the rest three most abundant heterotrophs were belonging to phyla of
308 Opalozoa, Discoba and Sagenista, both in cell and gene abundances. Differences for
309 autotrophic compositions on phyla level in cell and gene abundances were insignificant,
310 which were both dominated by Chlorophyta and Ochrophyta. Redundancy analysis suggested
311 stations from the same ocean or latitudinal regions tended to cluster together (**Fig. 1e;**
312 **Supplementary Fig. S1**). Certain species were more abundant in lower latitudes, such as
313 Chrysophyceae_Clade G_sp. of Ochrophyta, *Chrysochromulina* sp. and
314 Haptophyta_HAP2_XXX_sp. of Haptophyta, and *Karlodinium veneficum* of Dinoflagellata
315 for mixotrophs, MOCH.2_XXX sp. and Chlorophyta_XXXX_sp. of Chlorophyta for
316 autotrophs, and MAST.4C_XX sp. and Telonemia Goup2_X sp. for heterotrophs. Those taxa
317 were positively affiliated with PAR and CML, and negatively correlated with nutrients
318 (NO_3^-&NO_2^- and NO_2^-) and other variables located on the right side of the RDA plot. Other
319 species, including *Geminigera cryophila* of Cryptophyta and *Triparma mediterranean* of
320 Ochrophyta for mixotrophs and Telonemia Group1 for heterotrophs, were abundant at higher
321 latitudes and positively correlated with oxygen and sunshine duration. Additionally,
322 *Micromonas* A2 and Bacillariophyta_XXX sp. for autotrophs were associated with high
323 MLD, POC and NO_2^-, while Cryptomonadales_XX_sp. for mixotrophs and NPK2
324 lineage_X_sp. for heterotrophs were affiliated with high abundances of *Prochlorococcus*.

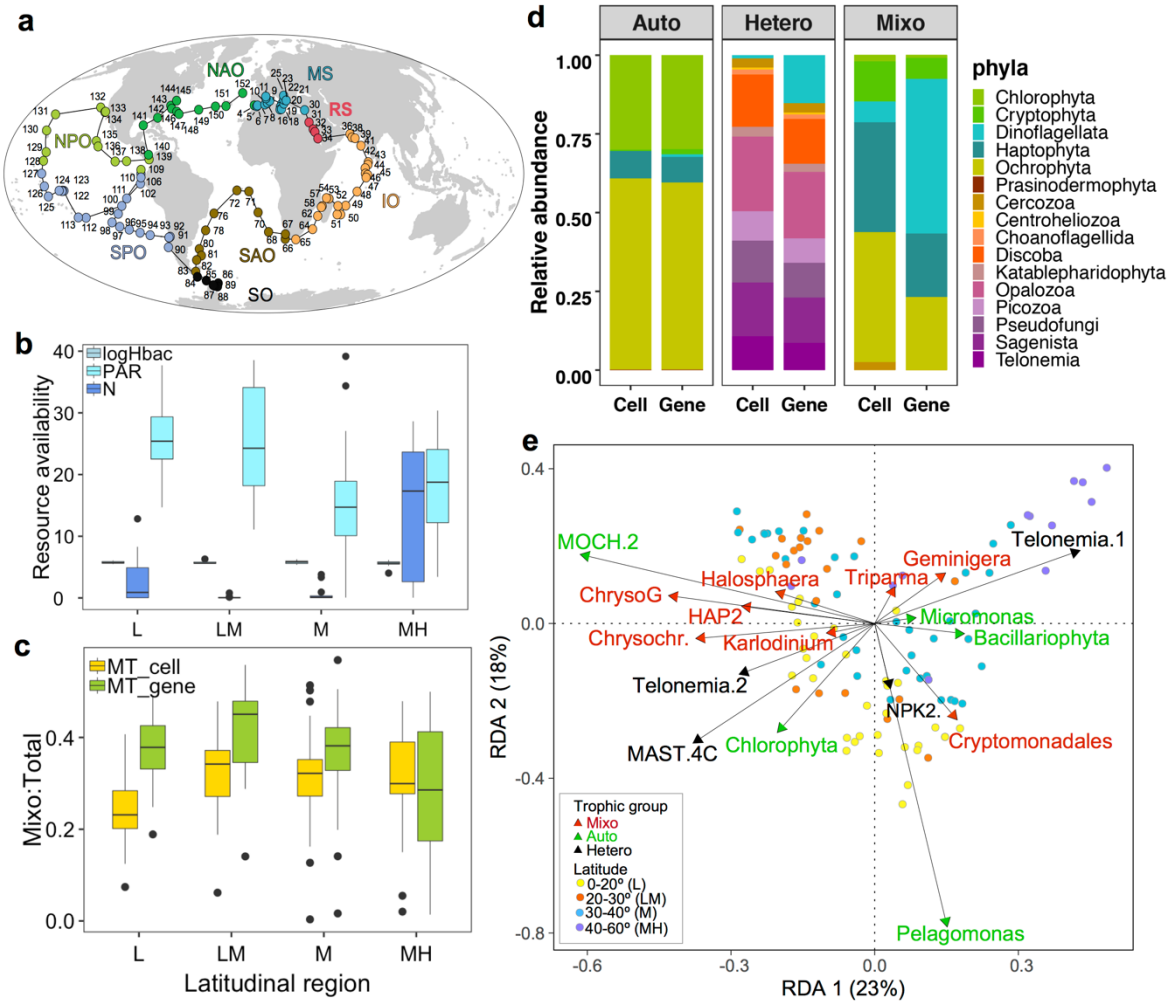


Fig. 1. Map of study sites (**a**), averaged median values of inorganic nutrients (NO_3^-&NO_2^-; N for abbreviation), light (PAR) and heterotrophic bacteria (**b**), and mixotrophic prevalence index TI_M (Mixo:Total) (**c**) at different latitudinal regions. Panel **d** presents composition of autotrophic (Auto), mixotrophic (Mixo) and Heterotrophic (Hetero) eukaryotes on phyla level derived from 18S rRNA gene and corrected cell abundances. Panel **e** is the redundancy analysis plot with distribution of trophic groups on species level. Selected RDA variables and *Tara Oceans* lineage full names can be found in Supplementary Fig. S1. Latitudinal region name and definition were shown in the legend of panel **e**.

Nutritional variables and the prevalence of three trophic modes

Among the 29 selected predictors, availability of nutrients (including NO_3^-&NO_2^- and silicate), PAR and temperature were the top variables revealed by principal component analysis for explaining the environmental conditions of surface (SUR) and surface and chlorophyll maxima layer (SUR&CML) samples (**Table 2; Figs. 2, 3**). Specifically, both principal component axes (PC1 and PC2) for the SUR&CML samples were mostly explained by PAR and nutrients (loading coefficients varied from -0.42 to 0.17; R^2 of 0.30). When only considering SUR samples, PC1 axis was mostly contributed by nutrients and temperature (R^2 of 0.35), while PC2 axis (R^2 of 0.24) were mostly explained by MLD, CML and PAR.

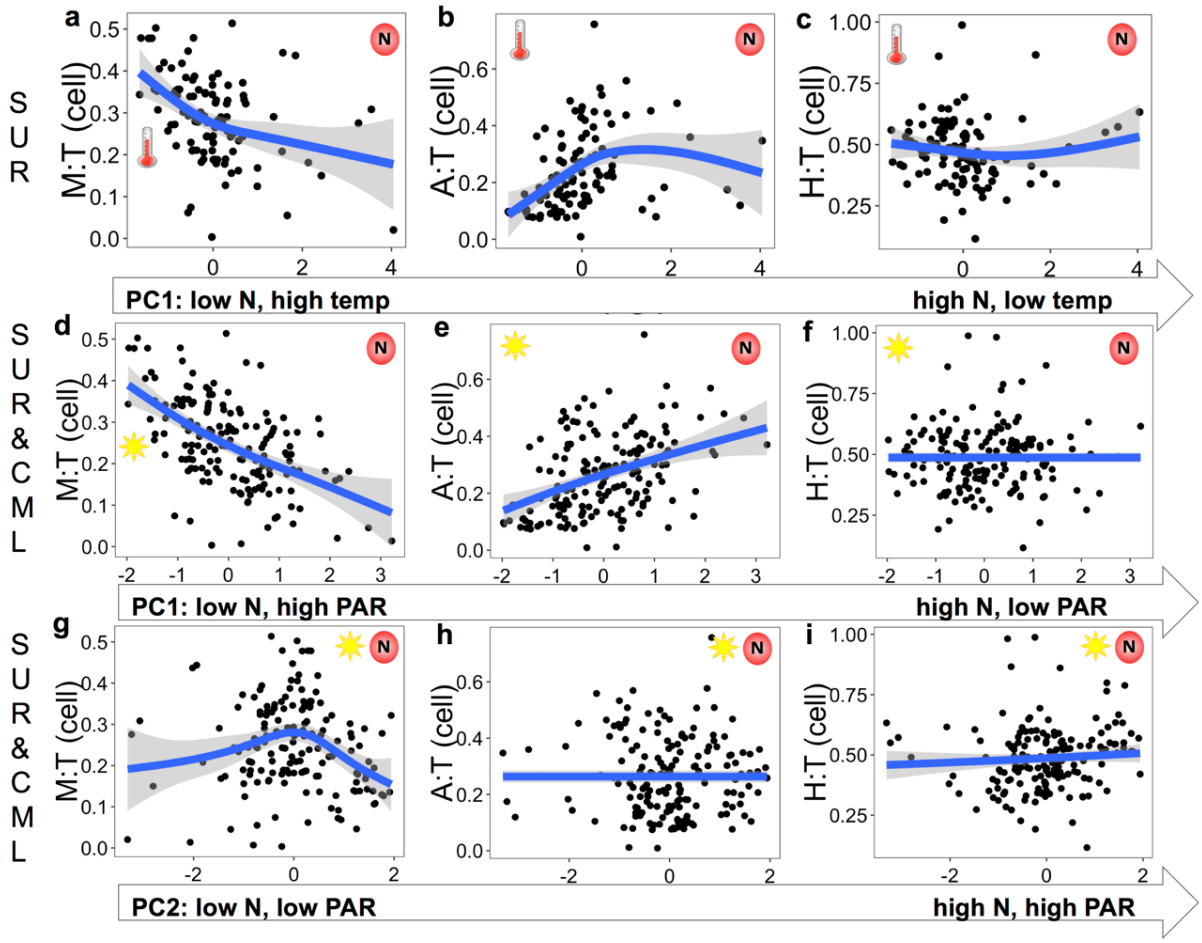


Fig. 2. Variation patterns of $TI_{A/M/H}$ in cell abundances along PC1 axis for surface (SUR) dataset (**a-c**), and PC1 (**d-f**) and PC2 (**g-i**) axes for surface and chlorophyll maximal layer (SUR&CML) dataset. Principal components of nutrients (NO_3^-&NO_2^-), PAR, and temperature were marked on each panel, and the location of symbols indicates higher values ends. Labels embedded on each PC-axis indicate how variables change from left to right. Blue lines and shaded bands were fitted generalized additive model smoother and 95% confidence interval.

The fitted smoothers of TI_g demonstrating the most substantial responses were from mixotrophs to changing principal components. Specifically, there was a decreasing trend of TI_M along PC1 for both datasets, indicating decreasing prevalence of mixotroph with increasing nutrients for SUR (**Fig. 2a**), as well as decreasing PAR and increasing nutrients for SUR&CML (**Fig. 2d**). TI_M did not demonstrate clear changes with the PC2 axis from SUR samples so results were not shown. It is noteworthy that the slope for decreasing TI_M when nutrients are increasing and PAR were decreasing was the steepest among all plots (**Fig. 2d**). For PC2 axes of SUR&CML dataset, TI_M presented a unimodal shape with changing nutrients and PAR conditions, with maximal values obtained around zero PC2 scores (**Fig. 2g**). These results may suggest a more complex non-linear relationship between TI_M and co-varying nutrients and light conditions. TI_A overall increased with increasing nutrients for the SUR dataset (**Fig. 2b**), and increasing nutrients and decreasing PAR for the SUR&CML

dataset (**Fig. 2e**). When nutrients and PAR changed in the same direction (PC2 for SUR&CML dataset), TI_A did not show clear variations (**Fig. 2h**). TI_H overall varied much less with changing conditions compared to TI_M and TI_A (**Figs. 2c, f, i**), among which a weak unimodal shape was observed with PC1 from SUR dataset (**Fig. 2c**). For the gene abundances denoted results, similar trends were found for TI_g variations over the principal components but with different slope shapes (**Fig. 3**). Particularly, a steeper decreasing slope was presented for TI_M on PC1of SUR samples and a steeper increasing slope (before the peak) was seen on PC2 for the SUR&CML samples (**Fig. 3a, g**). The weak increasing trend of TI_H over PC1 axis of SUR samples disappeared (**Fig. 3c**), and a somewhat stronger increase of TI_H were demonstrated along both axes of PC1 and PC2 for the SUR&CML dataset (**Fig. 3f, i).**

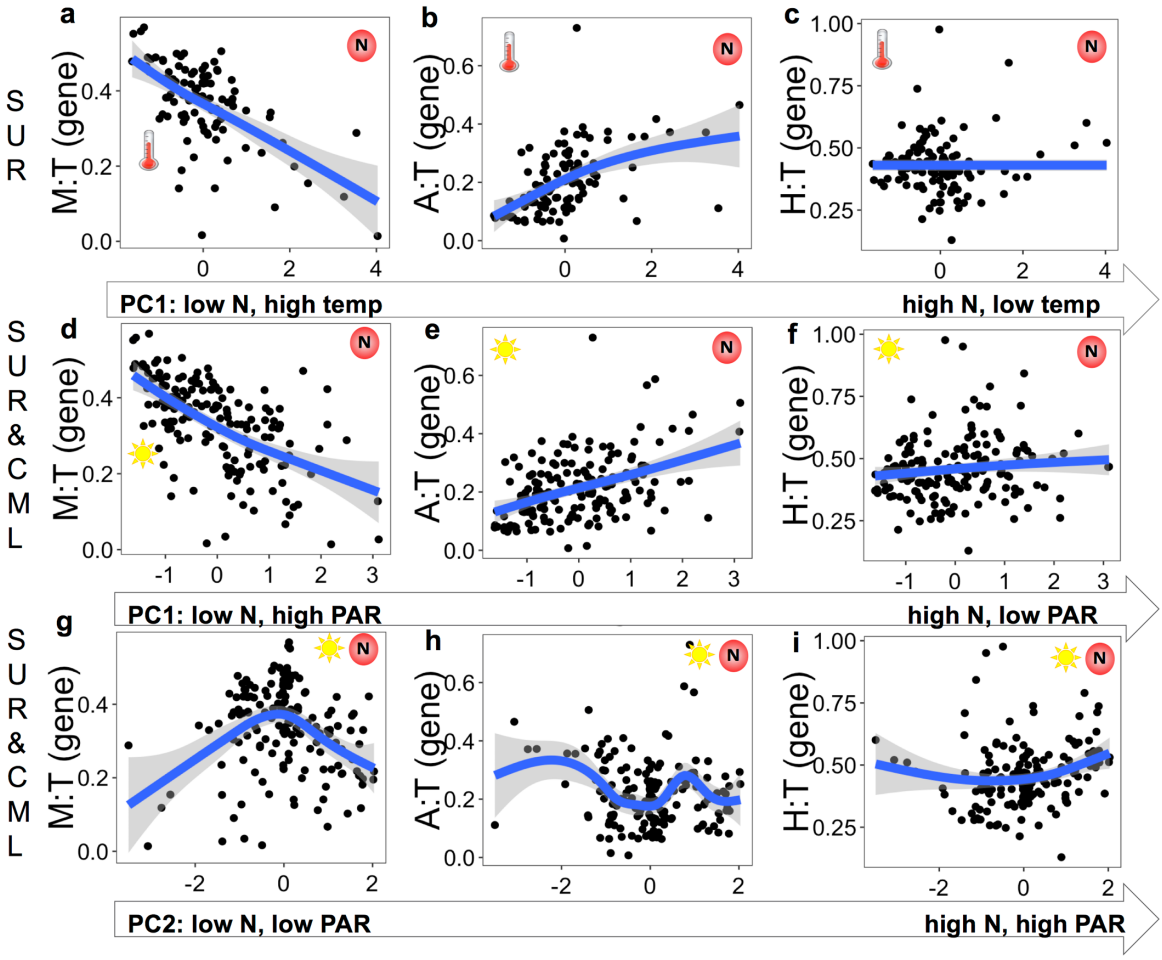
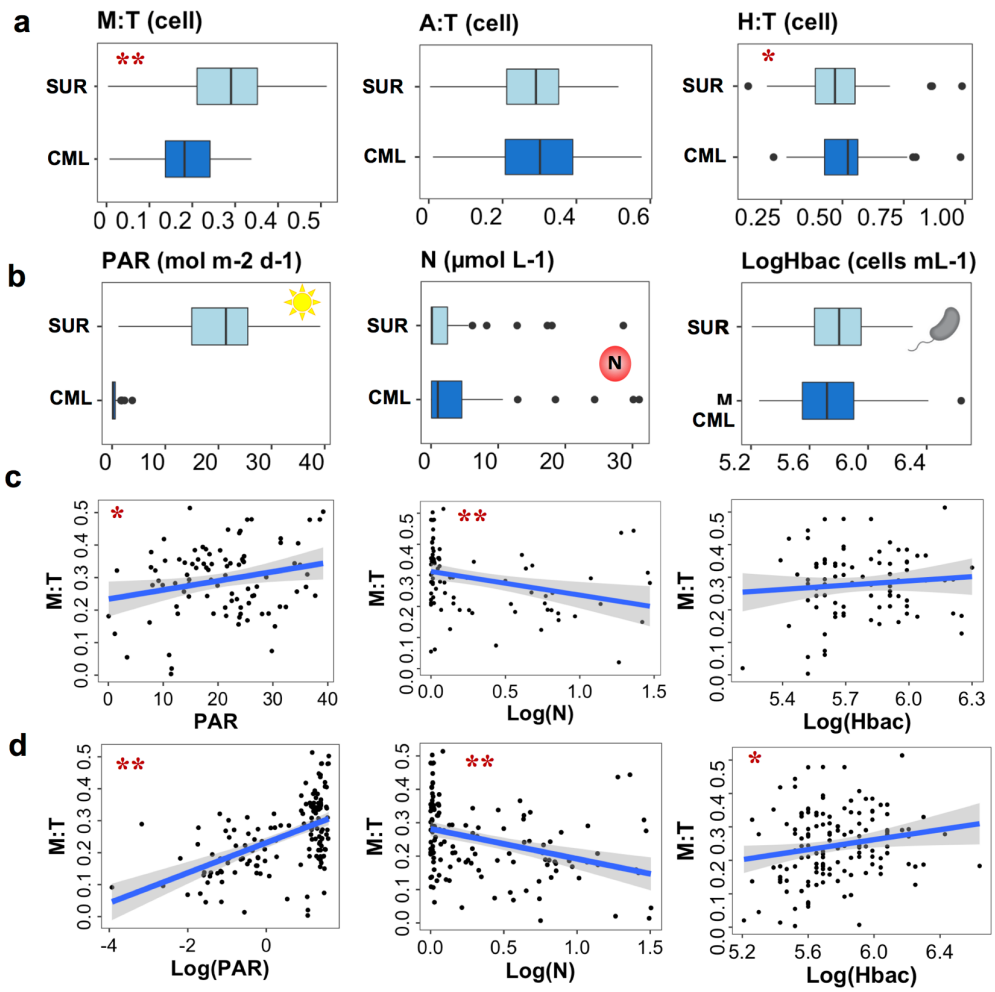


Fig. 3. Variation patterns of $TI_{A/M/H}$ in 18S rRNA gene abundances along PC1 axis for surface (SUR) dataset (**a-c**) and PC1 (**d-f**) and PC2 (**g-i**) axes for surface and chlorophyll maximal layer (SUR&CML) dataset. Principal components of nutrients (NO_3^-&NO_2^-), PAR, and temperature were marked on each panel, and the location of the symbols indicates higher values ends. Labels embedded on each PC-axis indicate how variables change from left to right. Blue lines and shaded bands were fitted generalized additive model smoother and 95% confidence interval.

Consistent with the PCA results, clear differences were observed for TI_g between SUR and CML samples (from 61 stations have both SUR and CML samples), accompanied with distinct resources availability of nutrients, light and bacterial prey (**Fig. 4**). Significantly

386 higher TI_M was seen in surface samples where PAR and bacteria were higher and N
387 (NO_3^-&NO_2^-) were lower (median 0.29 vs 0.18, $p<0.001$, df=60). In comparison, TI_A was
388 slightly higher in DCM compared to surface samples but not statistically significant (median
389 0.30 vs 0.29, $p>0.05$, df=60), and TI_H were significantly higher in CML than surface (median
390 0.52 vs 0.47, $p<0.05$, df=60) (**Fig. 4a, b**). Linear regression derived from all samples
391 suggested a consistently positive relationship for TI_M with PAR and negative correlation with
392 nutrients (NO_3^-&NO_2^-), from both SUR (122 samples; **Fig. 4c**) and SUR&CML dataset (165
393 samples; **Fig. 4d**). TI_A demonstrated a significantly positive and negative correlation with
394 nutrients and PAR, whereas neither was significantly associated with TI_H (**Supplementary**
395 **Figs. S2a-h**). For heterotrophic bacteria, a positive but weaker relationship ($p=0.02$) was
396 found between TI_M and the SUR&CML dataset but not the SUR dataset (**Fig. 4c, d**). In
397 contrast, both datasets have retrieved a strong negative association between heterotrophic
398 bacteria and TI_H, and a weaker but positive correlation between heterotrophic bacteria and
399 TI_A in the surface ($p<0.05$) (**Supplementary Fig. S2i-l**). Other single variables showing
400 significant regressions with TI_M were all shown in **Supplementary Fig. S3**.



401 **Fig. 4.** Comparison of TI_g (**a**) and nutritional resource availability (**b**) between surface (SUR)
402 and chlorophyll maximum layer (CML). Panels **c** and **d** demonstrate linear regressions
403 between TI_M and three resource variables of PAR, NO_3^-&NO_2^- (N), and heterotrophic bacteria
404 (Hbac) with SUR (**c**) and SUR&CML dataset (**d**), respectively. Variables were sometimes

Log-transformed for the best representation purpose. Significance codes were shown in red stars (one star if 0.001<P<0.05 and two stars if P<0.001), and shaded bands are pointwise 95% confidence interval on the fitted linear regression (blue lines).

Mixotrophic growth with varying nutritional resource

The most abundant mixotrophs identified from the *Tara Oceans* dataset belonged to classes of Prymnesiophyceae, Dictyochophyceae, Chrysophyceae, Dinophyceae and Cryptophyceae (**Fig. 5a**). Mixotrophic strains we isolated from the North Pacific had been shown to be efficient grazers of heterotrophic bacteria and cyanobacteria (including *Prochlorococcus* and *Synechococcus*), and they were closely affiliated to those abundant *Tara Oceans* lineages [5,40]. Six strains selected for the experimental study were belonging to *Chrysochromulina*, Prymnesiophyceae_XXX of Prymnesiophyceae, Clade-H_X of Chrysophyceae, as well as *Florenciella*, Dictyochales_X and Florenciellales_X of Dictyochophyceae. Although the lineage from Florenciellales_X was not shown in Fig. 5a, it ranked as the 26th most abundant lineage across all stations. The relative abundances of those isolates retrieved from surface were all higher than CML except for Dictyochales_X (**Fig. 5b**), presenting similar patterns as the entire mixotrophic community (TI_M results in **Fig. 4a**). Given the consistent and substantial responses of TI_M to nutrients and PAR derived from field observations, we designed multifactorial bioassay experiments to assess impacts of light and nutrients (inorganic and organic nutrients) availability on the growth of mixotrophs (**Fig. 5c**).

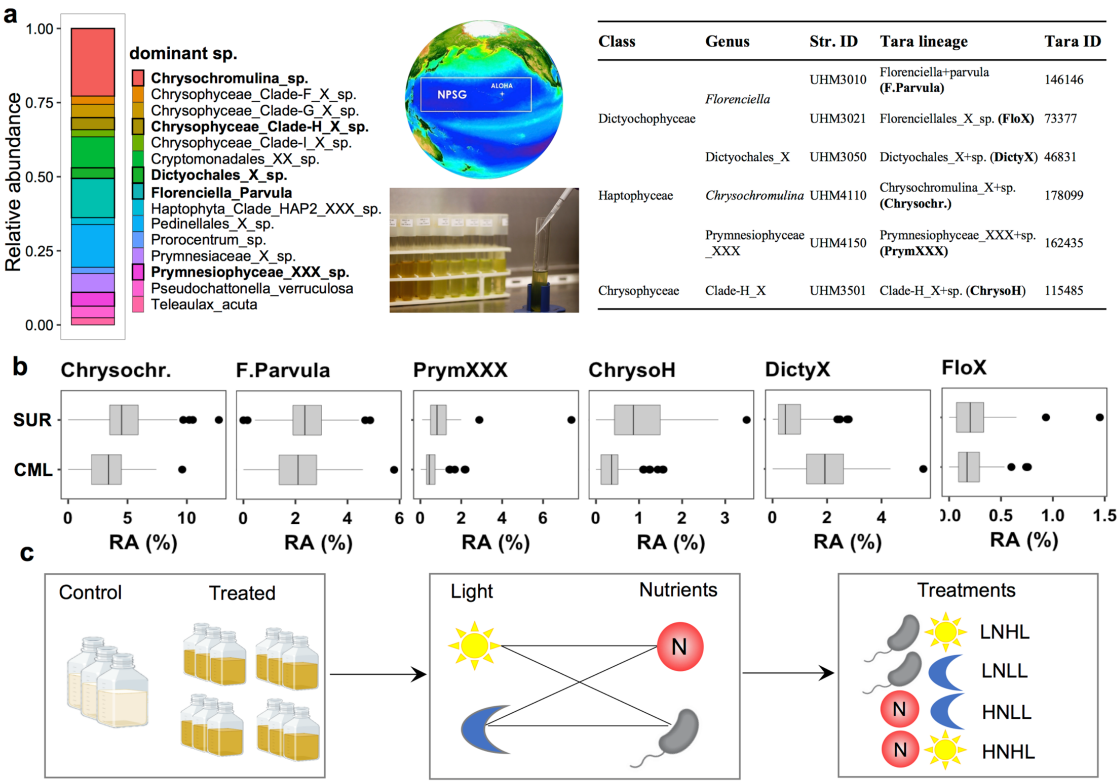


Fig. 5. Relative abundances of top 15 most abundant *Tara Oceans* mixotrophic lineages (in gene converted cell abundances), and phylogenetically related strains isolated from North Pacific Subtropical Gyre (**a**). Comparison of relative abundances in surface (SUR) and chlorophyll maximum layer (CML) (**b**), and an illustrative flow chart showing how different treatments were set up for the growth experiments (**c**).

430
431
432 Distinct growth curve shapes and final abundances were presented among different
433 treatments for the six isolates (**Fig. 6a**), accompanied by different disappearance rates of prey
434 (**Fig. 6b**). Prey in controls (without mixotrophic grazers) all presented slower disappearance
435 rates compared to cultures with grazers but to different levels among strains. Grazers in low
436 nutrients control (without prey) grew to much lower concentrations, i.e., all $<1\times10^3$ cells mL^{-1}
437 comparing to 10^4-10^5 cells mL^{-1} when added prey (data not shown). Shown as exponential
438 growth rates (bars in **Fig. 6c**), most isolates presented the highest values when grown under
439 high light, either when receiving high nutrients or sufficient prey (low nutrients). Lower
440 growth rates were observed with low light treatments (marked with black arrows in **Fig. 6c**)
441 for strain 4110 from *Chrysochromulina*, 4150 from Prymnesiophyceae_XXX and 3021 from
442 Florenciellales_X, i.e., categorized as functional type I. The rest three isolates did not show
443 significant impact by light, and thus were classified as functional type II. For grazing rates,
444 most strains showed the highest clearance rates when receiving low nutrients and high light
445 (marked with blue arrows in **Fig. 6d**). Only two strains belonging to Type II were capable of
446 grazing much faster when receiving low light and high nutrients (marked with red arrows in
447 **Fig. 6d**).
448 ANOVA results revealed a significant effect on growth by light, i.e., to increase growth
449 rates with increasing light, either alone, or together with nutrients and species. However,
450 (inorganic) nutrients availability did not show significant impact on growth rates, as cultures
451 supplemented with bacterial prey can sufficiently compensate for nutrients deficiency. Less
452 significant effects were observed for clearance rates, as strains responded differently to
453 changing light and nutrients conditions, indicated by the high p (>0.05) values from both
454 single and multi-factorial ANOVA (**Table 3**).

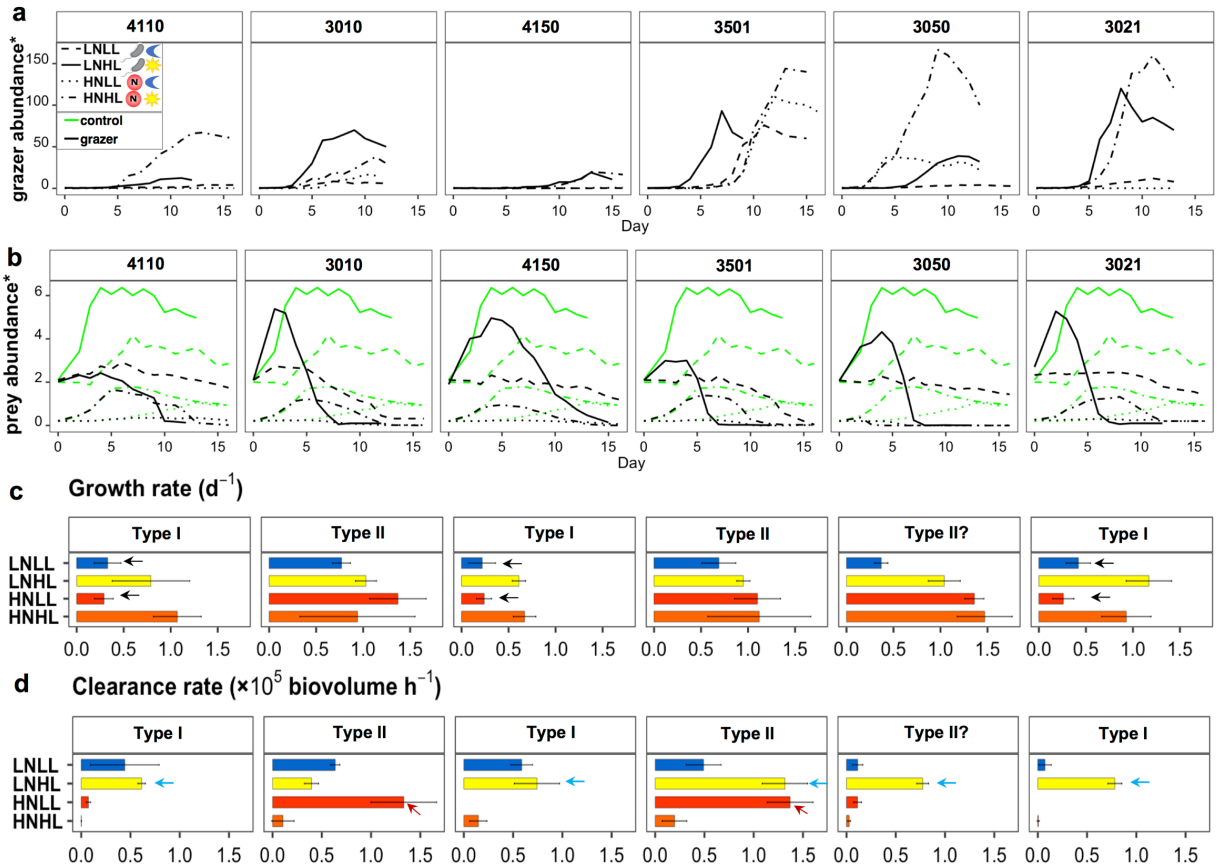


Fig. 6. Panel **a** presents growth curves of six mixotrophic isolates grown under high light (HL) and low light (LL), as well as high inorganic nitrogen (HN) and low inorganic nitrogen (LN) conditions. All treatments were supplemented with *Procholorococcus* prey. Panel **b** presents removal of prey in all cultures (ID table in Fig. 5a); green lines indicate prey control without grazers and black lines were prey with grazers. Panels **c** and **d** demonstrate estimated growth rates during exponential phase and biovolume clearance rates for each strain. Errors bars in c and d denote standard deviation among replicates for each treatment but were not shown in panel a and b for more readability. *For grazer concentrations in HNHL, they were divided by 50000 cells mL^{-1} for #4110 and 5000 cells mL^{-1} for other strains, and 1000 cells mL^{-1} for all the rest treatments. *Prey abundances in all HN treatments were divided by 10^7 cells mL^{-1}, and all LN were divided by 10^6 cells mL^{-1}. Arrows in **c** and **d** indicate treatments that had caused significantly lower growth rates (c), and significantly higher grazing rates (d).

Discussion

In this study, we used large-scale field surveys to retrieve prominent variables that possibly affect the trophic mode structure of eukaryotes on community levels, with a particular focus on three nutritional resources. We then conducted finely controlled lab experiments using isolates to investigate how mixotrophic growth respond to changing nutritional resources, in terms of both forms (inorganic nutrients and organic nutrients) and availability. While acknowledging species-specific variations, both field and experimental results retrieved a close relationship between mixotrophic prevalence and nutritional resource

among inorganic nutrients, light and prey, which have possibly introduced distinct niche preferences for mixotrophs versus autotrophic and heterotrophic eukaryotes in the ocean.

Mixotrophic growth responding to changing nutritional resources

Suggested by our experimental results, mixotrophy can function as an efficient strategy for compensating inorganic nutrients deficiency for most strains through ingestion of bacterial prey, but not necessarily for supplementing light limitation (Fig. **6c**). Among the six studied cultures, three of them were efficient with compensating light/energy deficiency via heterotrophic nutrition (grazing), but not the other three. These results indicated that mixotrophic controlling mechanisms for metabolic partitioning between autotrophy and heterotrophy were species-specific. The Type I isolates that were light-dependent possibly align with the 'primarily photosynthesis and phagocytosis for nutrients' functional group proposed by Jones (1997) and Stoecker (1998) [51] [52]. The other three strains of Type II were light-independent, and likely belong to the 'phagocytose when light is limiting' or 'perfectly balanced mixotrophy' functional group. Based on proposed resource allocation theories, those light-dependent mixotrophs probably rely on photosynthesis for gaining most energy and organic carbon [12,13,14]. The light-independent group performed better with low light availability, suggesting a higher flexibility governing autotrophic and heterotrophic growth in response to the changing conditions.

Relationship between mixotrophy and nutritional resources on large spatial scale

Our field analysis also indicated that availability of three nutritional resources could affect the relative success among three trophic strategies on community level, and across large spatial scales. Specifically, higher-nutrients and lower-light environments could possibly promote the relative abundances of autotrophs, whereas mixotrophic community proliferated from low nutrients and high light environments. This positive light-mixotrophy relationship in the field can be partially explained by the possibility that the light-depend Type I mixotrophs were dominant across the investigated ocean regions. On the other hands, Rothhaupt (1996) and Edwards et al. (2019; 2022) have suggested that during a resource competition process, mixotrophs could outcompete heterotrophs, because they can use extra energy gained from photosynthesis which enable them to suppress bacterial prey to very low level (below the threshold of heterotrophic grazers) [8,11,53]. Both mechanisms could be true, and it is worthy to conduct further studies to investigate the fundamental mechanisms related to metabolic partitioning and resource allocation for mixotrophic strategy.

Heterotrophic bacteria maintained relatively consistent abundances in the sampled oceans and was not selected as main explanatory factor for TI_g by RDA or PCA. Nevertheless, they demonstrated as one of the predictors explaining TI_M based on stepwise regression model (**Supplementary Table S2**). When treated as a single variable, a strong negative correlation was found between TI_H and heterotrophic bacteria. Comparatively, the correlation was positive for TI_A and less significant for TI_M. This interesting shifts likely suggested distinct interactions between bacteria and three trophic groups. Heterotrophic bacteria benefit from the dissolved organic carbon (DOC) excreted by autotrophs and recycle nutrients for autotrophic growth, forming a positive relationship. Whereas for heterotrophic eukaryotes, this negative association could be attributed to high grazing pressure they exerted on heterotrophic bacteria [43,44]. The weaker but positive correlation between heterotrophic bacteria and TI_M could be explained by the more complex relations among bacteria (prey and

521 decomposer) and mixotrophs (grazer and DOC provider), as well as the lower grazing
522 capability/pressure from mixotrophic grazers than heterotrophs [13,23,38,45].
523 **Environmental variables could affect the prevalence of trophic strategies**
524 Other oceanographic conditions such as CML, nitracline depths and MLD also seemed to
525 affect mixotrophic prevalence, though they are physical processes that are more or less
526 associated with nutrients and light conditions. For example, deeper nitraclines can potentially
527 suppress nutrients supply in the surface and therefore increase $TI_{M.}$ Similarly, deeper
528 chlorophyll maxima could be indicative of higher light availability and lower nutrients in
529 photic depths, thereby enhancing TI_M. In contrast, deeper mixed layer depths (stronger
530 mixing) could reduce light availability and increase nutrients concentration, which both
531 weaken TI_M. The overall positive association between CML and nitracline depths and TI_M,
532 and negative correlation between MLD and TI_M were in line with their overall relationships
533 with nutrients and light.
534 Previous field studies suggested that besides nutrients [21,46] and light [47,48,49],
535 variables such as temperature [50] and carbon composition (colored DOM and CO_2) [20]
536 were also likely important for affecting mixotrophic prevalence. These variables were also
537 included in our analysis and yielded varying relationships with TI_g. These differences could
538 possibly be due to the different environmental gradients covered among our studies (various
539 open oceans) versus others (North Pacific, and boreal Canadian lakes). Stepwise regression
540 model suggested that PAR, nutrients, density, PIC, DCM, heterotrophic bacteria and
541 *Prochlorococcus* were the top predictors explaining TI_M with the *Tara Oceans* dataset
542 (**Supplementary Table S2**).
543 Indeed, it is not always straightforward to delineate relationships between a single
544 environmental variable and trophic mode prevalence. Results from our statistical analysis
545 rather provide insights into the possible correlations between them. Future studies focusing
546 on smaller regions with in situ experiments targeting active trophic behaviors will provide
547 additional knowledge for interpreting the trophic mode and environment relationships.
548 **Methodology caveats**
549 One potential caveat in our methodology is that mixotrophs were annotated based on
550 published literature at species or higher taxonomic ranks. These trophic annotations were not
551 necessarily aligning with in situ metabolic activities. Nevertheless, the potential or
552 historically evolved capability of phago-mixotrophy are sufficient to delineate their
553 ecological niche which was the focus of this study. We also recruited longer-term averaged
554 values for some of the key variables (e.g., 8-day and 30-day PAR) as well as a variety of
555 parameters (e.g., NO_3^-, NO_2^-, PO_4^{3-} and SiO_4^- for nutrients) to reduce the possibility of
556 retrieving false relationships. Lack of trophic evidence for some of the species/lineages could
557 on the other hand, cause false annotation, and variations against other studies. For example,
558 we annotated *Micromonas*, a member of Chlorophyta, as autotrophs based on the best
559 knowledge of ours [51,54]. Howerver, others have argued possible phagotrophic evidence in
560 the *Micromonas polaris* [55]. We also recognize the gene-cell abundance correction factors
561 could be inaccurate for certain groups, as we did not have first-hand cellular abundance data
562 from in situ samples. Nevertheless, these calibrated abundance data could improve accuracy
563 for some heavily over-estimated groups such as dinoflagellates and under-estimated small
564 flagellates.

565 Combining various approaches including experimental study, field observation, model
566 simulation and bioinformatics tools can be powerful to retrieve the most reliable data for
567 future mixotrophy research [3,56].

568 **Conclusion**

569 This study provides the first *Tara Oceans* analysis for niche partitioning among three
570 trophic populations of mixotrophs, autotrophs and heterotrophs of the eukaryotic community.
571 Our results suggested that either on community level or species level, phagotrophic
572 mixotrophy are a trophic strategy marine eukaryotes had evolved to survive nutrients
573 impoverishment and light limitation, deepening our understanding about mechanisms
574 controlling their prevalence in the ocean.

575

576 **Table 1.** Oceanographic characters and trophic index of surface water samples.
577 Abbreviations for PAR, NO_3^-&NO_2^-, Hbac stands for photosynthetically available radiation,
578 dissolved nitrate and nitrite (NO_3^-&NO_2^-), and heterotrophic bacteria, respectively. M:T, M:A
579 and M:H denotes rations of either cell abundance (cell) or 18S rRNA gene abundance (gene)
580 among mixotroph (M), autotroph (A), heterotroph (H) and total eukaryotes (T).

Latitude	Oceanic Region†	Station	Temp (℃)	Salinity	PAR (mol m^{-2} d^{-1})	NO_3^-&NO_2^- (μmol L^{-1})	Hbac (cells mL^{-1})	M:T (Cell /Gene)	A:T (Cell /Gene)	H:T (Cell /Gene)
0-20º	SPO, IO, NPO, SAO, NAO	36	26.6	35.1	25.4	0.89	4.9e5	0.23/ 0.38	0.32/ 0.24	0.43/ 0.38
20-30º	IO, NPO, SPO, SAO, RS	17	24.7	36.2	24.3	0.03	3.9e5	0.34/ 0.45	0.10/ 0.08	0.51/ 0.44
30-40º	MS, SAO, NAO, SO	37	19.5	36.4	14.7	0.13	5.3e5	0.32/ 0.38	0.19/ 0.16	0.48/ 0.43
40-60º	SO, SAO, MS	14	7.1	34.3	18.8	17.3	4.5e5	0.30/ 0.29	0.18/ 0.22	0.47/ 0.46
Global	SPO, IO, NPO, SAO, NAO, SO, RS, MS	104	23.8	35.4	21.3	0.14	4.5e5	0.29/ 0.38	0.21/ 0.19	0.47/ 0.42

581 † Oceanic regions follow the Tara Oceans project standards. NAO: North Atlantic Ocean; MS:
582 Mediterranean Sea; RS: Red Sea; IO: Indian Ocean; SAO: South Atlantic Ocean; SO:
583 Southern Ocean; SPO: South Pacific Ocean; NPO: North Pacific Ocean.

Table 2. Major environmental variables retrieved by principal component analysis for surface (SUR) and surface and chlorophyll maximal layer (SUR&CML) datasets. Top variables from each principal component axis (PC1 and PC2) for both datasets were shown with their loading coefficients (marked in bold). Symbols of '+' and '−' explain the direction of changing variable values from left to right on the axes.

	SUR		**SUR& DCM**		
	PC1	**PC2**	**PC1**	**PC2**	**sum**[†]
PAR	-0.08	**-0.14**	**-0.42**	**-0.35**	0.99
NO_3^-&NO_2^-	**0.31**	**-0.14**	**0.23**	**-0.25**	0.93
Silicate	**0.23**	-0.10	**0.17**	**-0.17**	0.67
Temperature	**-0.24**	0.08	-0.14	0.13	0.59
Nitracline depth	-0.20	-0.11	-0.15	0.12	0.58
CML	0.06	**-0.17**	-0.10	0.07	0.40
Mixed layer depth	0.06	**-0.22**	-0.04	-0.06	0.38
R^2	0.35	0.24	0.30	0.25	

[†] summed absolute values from PC1 and PC2 for both datasets

591 **Table 3.** ANOVA of different growth rates and clearance rates among six species cultivated
592 under different light (PAR) and nutrients (N) conditions (results shown in Fig. 6). Individual
593 and interactive effects of light, nutrients and species were analyzed via single factor and
594 multi-factors ANOVA, respectively.

	Growth rates					Clearance rates				
	single factor			multi-factors		single factor			multi-factors	
	PAR	N	Sp.	PA&N	PAR&N&Sp.	PAR	N	Sp.	PAR&N	PAR&N&Sp.
P[†]	*	>0.05	>0.05	*	*	>0.05	>0.05	>0.05	*	0.13
F[‡]	6.5	1.7	2.1	5.1	3.6	0.00	2.98	1.35	3.96	2.02
Dfn[§]	1	1	5	3	5	1	1	5	3	5
Dfd[‖]	20	20	17	18	18	20	20	17	18	18
Effect[¶]	+	+	/	/	/	+	–	/	/	/

595 [†] Significance level of ANOVA, '*' denotes P<0.05.
596 [‡] Variation between sample means / variation within the samples.
597 [§] Degrees of freedom in the numerator, equals to the groups (of factors) – 1.
598 [‖] Degrees of freedom in the denominator, equals to the total number of subjects in the
599 experiment – the groups (of factors).
600 [¶]Symbols of + and – indicate a positive and negative impacts from increasing light and
601 nutrients availability, and / denotes analysis unavailable. Only those marked with * P values
602 in the first row had significant impact.
603

604 **References**

605 1. Sherr, E.B.; Sherr, B.F. Significance of predation by protists in aquatic microbial food
606 webs. *Antonie van Leeuwenhoek* **2002**, *81*, 293-308, doi:10.1023/A:1020591307260.
607 2. Stoecker, D.K.; Hansen, P.J.; Caron, D.A.; Mitra, A. Mixotrophy in the Marine
608 Plankton. *Annual Review of Marine Science* **2017**, *9*, 311-335,
609 doi:https://doi.org/10.1146/annurev-marine-010816-060617.
610 3. Millette, N.e.a. Mixoplankton and mixotrophy: future research priorities. *Journal of*
611 *Plankton Research* **2023**, *45*, 576-596, doi:10.1093/plankt/fbad020.
612 4. Bock, N.A.; Charvet, S.; Burns, J.; Gyaltshen, Y.; Rozenberg, A.; Duhamel, S.; Kim,
613 E. Experimental identification and in silico prediction of bacterivory in green algae.
614 *The ISME Journal* **2021**, *15*, 1987-2000, doi:10.1038/s41396-021-00899-w.
615 5. Li, Q.; Edwards, K.F.; Schvarcz, C.R.; Steward, G.F. Broad phylogenetic and
616 functional diversity among mixotrophic consumers of Prochlorococcus. *The ISME*
617 *Journal* **2022**, *16*, 1557-1569, doi:10.1038/s41396-022-01204-z.
618 6. Flynn, K.J.; Mitra, A.; Anestis, K.; Anschütz, A.A.; Calbet, A.; Ferreira, G.D.;
619 Gypens, N.; Hansen, P.J.; John, U.; Martin, J.L.; et al. Mixotrophic protists and a new
620 paradigm for marine ecology: where does plankton research go now? *Journal of*
621 *Plankton Research* **2019**, *41*, 375-391, doi:10.1093/plankt/fbz026.
622 7. Mitra, A.; Flynn, K.J.; Tillmann, U.; Raven, J.A.; Caron, D.; Stoecker, D.K.; Not, F.;
623 Hansen, P.J.; Hallegraeff, G.; Sanders, R.; et al. Defining Planktonic Protist

Functional Groups on Mechanisms for Energy and Nutrient Acquisition: Incorporation of Diverse Mixotrophic Strategies. *Protist* **2016**, *167*, 106-120, doi:https://doi.org/10.1016/j.protis.2016.01.003.

8. Rothhaupt, K.O. Laboratorary Experiments with a Mixotrophic Chrysophyte and Obligately Phagotrophic and Photographic Competitors. *Ecology* **1996**, *77*, 716-724, doi:https://doi.org/10.2307/2265496.
9. Thingstad, T.F.; Havskum, H.; Garde, K.; Riemann, B. On the Strategy of 'Eating Your Competitor': A Mathematical Analysis of Algal Mixotrophy. *Ecology* **1996**, *77*, 2108-2118, doi:https://doi.org/10.2307/2265705.
10. Anderson, R.; Charvet, S.; Hansen, P.J. Mixotrophy in Chlorophytes and Haptophytes—Effect of Irradiance, Macronutrient, Micronutrient and Vitamin Limitation. *Frontiers in Microbiology* **2018**, *9*, doi:10.3389/fmicb.2018.01704.
11. Edwards, K.F. Mixotrophy in nanoflagellates across environmental gradients in the ocean. *Proceedings of the National Academy of Sciences* **2019**, *116*, 6211-6220, doi:doi:10.1073/pnas.1814860116.
12. Ward, B.A. Mixotroph ecology: More than the sum of its parts. *Proceedings of the National Academy of Sciences* **2019**, *116*, 5846-5848, doi:doi:10.1073/pnas.1902106116.
13. Raven, J.A. Phagotrophy in phototrophs. *Limnology and Oceanography* **1997**, *42*, 198-205, doi:https://doi.org/10.4319/lo.1997.42.1.0198.
14. Mitra, A.; Flynn, K.J.; Burkholder, J.M.; Berge, T.; Calbet, A.; Raven, J.A.; Granéli, E.; Glibert, P.M.; Hansen, P.J.; Stoecker, D.K.; et al. The role of mixotrophic protists in the biological carbon pump. *Biogeosciences* **2014**, *11*, 995-1005, doi:10.5194/bg-11-995-2014.
15. Hansen, P.; Anderson, R.; Stoecker, D.; Decelle, J.; Altenburger, A.; Blossom, H.; Drumm, K.; Mitra, A.; Flynn, K. Mixotrophy Among Freshwater and Marine Protists. 2019.
16. Leles, S.G.; Bruggeman, J.; Polimene, L.; Blackford, J.; Flynn, K.J.; Mitra, A. Differences in physiology explain succession of mixoplankton functional types and affect carbon fluxes in temperate seas. *Progress in Oceanography* **2021**, *190*, 102481, doi:https://doi.org/10.1016/j.pocean.2020.102481.
17. Zubkov, M.V.; Tarran, G.A. High bacterivory by the smallest phytoplankton in the North Atlantic Ocean. *Nature* **2008**, *455*, 224-226, doi:10.1038/nature07236.
18. Hartmann, M.; Grob, C.; Tarran, G.A.; Martin, A.P.; Burkill, P.H.; Scanlan, D.J.; Zubkov, M.V. Mixotrophic basis of Atlantic oligotrophic ecosystems. *Proceedings of the National Academy of Sciences* **2012**, *109*, 5756-5760, doi:doi:10.1073/pnas.1118179109.
19. Tittel, J.; Bissinger, V.; Zippel, B.; Gaedke, U.; Bell, E.; Lorke, A.; Kamjunke, N. Mixotrophs combine resource use to outcompete specialists: Implications for aquatic food webs. *Proceedings of the National Academy of Sciences* **2003**, *100*, 12776-12781, doi:doi:10.1073/pnas.2130696100.
20. Hansson, T.H.; Grossart, H.-P.; del Giorgio, P.A.; St-Gelais, N.F.; Beisner, B.E. Environmental drivers of mixotrophs in boreal lakes. *Limnology and Oceanography* **2019**, *64*, 1688-1705, doi:https://doi.org/10.1002/lno.11144.

21. Le Noac'h, P.; Cremella, B.; Kim, J.; Soria-Píriz, S.; del Giorgio, P.A.; Pollard, A.I.; Huot, Y.; Beisner, B.E. Nutrient availability is the main driver of nanophytoplankton phago-mixotrophy in North American lake surface waters. *Journal of Plankton Research* **2024**, *46*, 9-24, doi:10.1093/plankt/fbad054.
22. Wilken, S.; Huisman, J.; Naus-Wiezer, S.; Van Donk, E. Mixotrophic organisms become more heterotrophic with rising temperature. *Ecology Letters* **2013**, *16*, 225-233, doi:https://doi.org/10.1111/ele.12033.
23. Berge, T.; Chakraborty, S.; Hansen, P.J.; Andersen, K.H. Modeling succession of key resource-harvesting traits of mixotrophic plankton. *The ISME Journal* **2017**, *11*, 212-223, doi:10.1038/ismej.2016.92.
24. Lepori-Bui, M.; Paight, C.; Eberhard, E.; Mertz, C.M.; Moeller, H.V. Evidence for evolutionary adaptation of mixotrophic nanoflagellates to warmer temperatures. *Global Change Biology* **2022**, *28*, 7094-7107, doi:https://doi.org/10.1111/gcb.16431.
25. Chu, T.; Moeller, H.V.; Archibald, K.M. Competition between phytoplankton and mixotrophs leads to metabolic character displacement. *Ecological Modelling* **2023**, *481*, 110331, doi:https://doi.org/10.1016/j.ecolmodel.2023.110331.
26. Pesant, S.; Not, F.; Picheral, M.; Kandels-Lewis, S.; Le Bescot, N.; Gorsky, G.; Iudicone, D.; Karsenti, E.; Speich, S.; Troublé, R.; et al. Open science resources for the discovery and analysis of Tara Oceans data. *Scientific Data* **2015**, *2*, 150023, doi:10.1038/sdata.2015.23.
27. Picheral, M.; Searson, S.; Taillandier, V.; Bricaud, A.; Boss, E.; Stemmann, L.; Gorsky, G.; Tara Oceans Consortium, C.; Tara Oceans Expedition, P. Vertical profiles of environmental parameters measured from physical, optical and imaging sensors during station TARA_181 of the Tara Oceans expedition 2009-2013. **2014**, doi:10.1594/PANGAEA.836578.
28. Ardyna, M.; Tara Oceans Consortium, C.; Tara Oceans Expedition, P. Environmental context of all stations from the Tara Oceans Expedition (2009-2013), about the annual cycle of key parameters estimated daily from remote sensing products at a spatial resolution of 100km. *In: Tara Oceans Consortium, Coordinators; Tara Oceans Expedition, Participants (2017): Registry of all samples from the Tara Oceans Expedition (2009-2013). PANGAEA, https://doi.org/10.1594/PANGAEA.875582* **2017**, doi:10.1594/PANGAEA.883614.
29. Ibarbalz, F.M.; Henry, N.; Brandão, M.C.; Martini, S.; Busseni, G.; Byrne, H.; Coelho, L.P.; Endo, H.; Gasol, J.M.; Gregory, A.C.; et al. Global Trends in Marine Plankton Diversity across Kingdoms of Life. *Cell* **2019**, *179*, 1084-1097.e1021, doi:https://doi.org/10.1016/j.cell.2019.10.008.
30. Faure, E.; Not, F.; Benoiston, A.-S.; Labadie, K.; Bittner, L.; Ayata, S.-D. Mixotrophic protists display contrasted biogeographies in the global ocean. *The ISME Journal* **2019**, *13*, 1072-1083, doi:10.1038/s41396-018-0340-5.
31. Pierella Karlusich, J.J.; Pelletier, E.; Zinger, L.; Lombard, F.; Zingone, A.; Colin, S.; Gasol, J.M.; Dorrell, R.G.; Henry, N.; Scalco, E.; et al. A robust approach to estimate relative phytoplankton cell abundances from metagenomes. *Molecular Ecology Resources* **2023**, *23*, 16-40, doi:https://doi.org/10.1111/1755-0998.13592.

32. de Vargas, C.; Audic, S.; Henry, N.; Decelle, J.; Mahé, F.; Logares, R.; Lara, E.; Berney, C.; Le Bescot, N.; Probert, I.; et al. Eukaryotic plankton diversity in the sunlit ocean. *Science* **2015**, *348*, 1261605, doi:doi:10.1126/science.1261605.
33. Guillou, L.; Bachar, D.; Audic, S.; Bass, D.; Berney, C.; Bittner, L.; Boutte, C.; Burgaud, G.; de Vargas, C.; Decelle, J.; et al. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research* **2012**, *41*, D597-D604, doi:10.1093/nar/gks1160.
34. Zhu, F.; Massana, R.; Not, F.; Marie, D.; Vaulot, D. Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiology Ecology* **2005**, *52*, 79-92, doi:10.1016/j.femsec.2004.10.006.
35. Gong, W.; Marchetti, A. Estimation of 18S Gene Copy Number in Marine Eukaryotic Plankton Using a Next-Generation Sequencing Approach. *Frontiers in Marine Science* **2019**, *6*, doi:10.3389/fmars.2019.00219.
36. Martin, J.L.; Santi, I.; Pitta, P.; John, U.; Gypens, N. Towards quantitative metabarcoding of eukaryotic plankton: an approach to improve 18S rRNA gene copy number bias. *Metabarcoding and Metagenomics* **2022**, *6*, doi:10.3897/mbmg.6.85794.
37. Li, Q.; Dong, K.; Wang, Y.; Edwards, K.F. Relative importance of bacterivorous mixotrophs in an estuary-coast environment. *bioRxiv* **2023**, 2023.2009.2030.560290, doi:10.1101/2023.09.30.560290.
38. Mitra, A.; Caron, D.A.; Faure, E.; Flynn, K.J.; Leles, S.G.; Hansen, P.J.; McManus, G.B.; Not, F.; do Rosario Gomes, H.; Santoferrara, L.F.; et al. The Mixoplankton Database (MDB): Diversity of photo-phago-trophic plankton in form, function, and distribution across the global ocean. *Journal of Eukaryotic Microbiology* **2023**, *70*, e12972, doi:https://doi.org/10.1111/jeu.12972.
39. Stacklies, W.; Redestig, H.; Scholz, M.; Walther, D.; Selbig, J. pcaMethods—a bioconductor package providing PCA methods for incomplete data. *Bioinformatics* **2007**, *23*, 1164-1167, doi:10.1093/bioinformatics/btm069.
40. Li, Q.; Edwards, K.F.; Schvarcz, C.R.; Selph, K.E.; Steward, G.F. Plasticity in the grazing ecophysiology of Florenciella (Dichtyochophyceae), a mixotrophic nanoflagellate that consumes Prochlorococcus and other bacteria. *Limnology and Oceanography* **2021**, *66*, 47-60, doi:https://doi.org/10.1002/lno.11585.
41. Montagnes, D.J.S.; Barbosa, A.B.; Boenigk, J.; Davidson, K.; Jürgens, K.; Macek, M.; Parry, J.D.; Roberts, E.C.; imek, K. Selective feeding behaviour of key free-living protists: avenues for continued study. *Aquatic Microbial Ecology* **2008**, *53*, 83-98.
42. Fenchel, T. The Ecology of Heterotrophic Microflagellates. In *Advances in Microbial Ecology*, Marshall, K.C., Ed.; Springer US: Boston, MA, 1986; pp. 57-97.
43. Fuhrman, J.A.; Noble, R.T. Viruses and protists cause similar bacterial mortality in coastal seawater. *Limnology and Oceanography* **1995**, *40*, 1236-1242, doi:https://doi.org/10.4319/lo.1995.40.7.1236.
44. Pernthaler, J. Predation on prokaryotes in the water column and its ecological implications. *Nature Reviews Microbiology* **2005**, *3*, 537-546, doi:10.1038/nrmicro1180.

45. Edwards, K.F.; Li, Q.; Steward, G.F. Ingestion kinetics of mixotrophic and heterotrophic flagellates. *Limnology and Oceanography* **2023**, *68*, 917-927, doi:https://doi.org/10.1002/lno.12320.
46. Stukel, M.R.; Landry, M.R.; Selph, K.E. Nanoplankton mixotrophy in the eastern equatorial Pacific. *Deep Sea Research Part II: Topical Studies in Oceanography* **2011**, *58*, 378-386, doi:https://doi.org/10.1016/j.dsr2.2010.08.016.
47. Ptacnik, R.; Gomes, A.; Royer, S.-J.; Berger, S.A.; Calbet, A.; Nejstgaard, J.C.; Gasol, J.M.; Isari, S.; Moorthi, S.D.; Ptacnikova, R.; et al. A light-induced shortcut in the planktonic microbial loop. *Scientific Reports* **2016**, *6*, 29286, doi:10.1038/srep29286.
48. Millette, N.C.; Pierson, J.J.; Aceves, A.; Stoecker, D.K. Mixotrophy in Heterocapsa rotundata: A mechanism for dominating the winter phytoplankton. *Limnology and Oceanography* **2017**, *62*, 836-845, doi:https://doi.org/10.1002/lno.10470.
49. Anderson, R.; Hansen, P.J. Meteorological conditions induce strong shifts in mixotrophic and heterotrophic flagellate bacterivory over small spatio-temporal scales. *Limnology and Oceanography* **2020**, *65*, 1189-1199, doi:https://doi.org/10.1002/lno.11381.
50. Lambert, B.S.; Groussman, R.D.; Schatz, M.J.; Coesel, S.N.; Durham, B.P.; Alverson, A.J.; White, A.E.; Armbrust, E.V. The dynamic trophic architecture of open-ocean protist communities revealed through machine-guided metatranscriptomics. *Proceedings of the National Academy of Sciences* **2022**, *119*, e2100916119, doi:doi:10.1073/pnas.2100916119.
51. Edwards, K.F.; Li, Q.; McBeain, K.A.; Schvarcz, C.R.; Steward, G.F. Trophic strategies explain the ocean niches of small eukaryotic phytoplankton. *Proceedings of the Royal Society B: Biological Sciences* **2023**, *290*, 20222021, doi:doi:10.1098/rspb.2022.2021.
52. Stoecker, D.K. Conceptual models of mixotrophy in planktonic protists and some ecological and evolutionary implications. *European Journal of Protistology* **1998**, *34*, 281-290, doi:https://doi.org/10.1016/S0932-4739(98)80055-2.
53. Jones, H. A classification of mixotrophic protists based on their behaviour. *Freshwater Biology* **1997**, *37*, 35-43, doi:https://doi.org/10.1046/j.1365-2427.1997.00138.x.
54. Jimenez, V.; Burns, J.A.; Le Gall, F.; Not, F.; Vaulot, D. No evidence of Phago-mixotropy in Micromonas polaris (Mamiellophyceae), the Dominant Picophytoplankton Species in the Arctic. *Journal of Phycology* **2021**, *57*, 435-446, doi:https://doi.org/10.1111/jpy.13125.
55. McKie-Krisberg, Z.M.; Sanders, R.W. Phagotrophy by the picoeukaryotic green alga Micromonas: implications for Arctic Oceans. *The ISME Journal* **2014**, *8*, 1953-1961, doi:10.1038/ismej.2014.16.
56. Beisner, B.E.; Grossart, H.-P.; Gasol, J.M. A guide to methods for estimating phago-mixotrophy in nanophytoplankton. *Journal of Plankton Research* **2019**, *41*, 77-89, doi:10.1093/plankt/fbz008.

796 **Supplementary Table S1.** Species and cellular 18S rRNA gene copy numbers used in this
797 study to retrieve gene-cell abundance correction factors for small eukaryotes.

Taxa	Group	Reference	Method	Biovolume (μm^3)	Gene Copies per Cell
Ostreococcus tauri	other flagellates	Gong et al. 2019	bioinformatics	0.5	3.4
Trebouxia sp.		Gong et al. 2019	bioinformatics	1767.1	78.1
Bathycoccus prasinos		Zhu et al. 2005	qPCR	3.8	5.2
Emiliania huxleyi		Zhu et al. 2005	qPCR	64.3	2.5
Mesopedinella arctica		Zhu et al. 2005	qPCR	252.8	7.4
Micromonas pusilla		Zhu et al. 2005	qPCR	4.0	3.0
Nannochloropsis salina		Zhu et al. 2005	qPCR	4.0	1.0
Ostreococcus tauri		Zhu et al. 2005	qPCR	0.3	2.5
Pelagomonas calceolata		Zhu et al. 2005	qPCR	7.7	1.3
Rhodomonas salina		Zhu et al. 2005	qPCR	217.0	28.8
Tetraselmis sp.		Zhu et al. 2005	qPCR	208.7	7.6
Chrysochromulina sp.		Li et al., 2022	qPCR	20.6	2.0
Chrysochromulina sp.		Li et al., 2022	qPCR	41.6	2.0
ChrysoH		Li et al., 2022	qPCR	5.6	1.0
DictyX		Li et al., 2022	qPCR	78.0	4.5
Florenciella Parvula		Li et al., 2022	qPCR	26.5	1.2
Hap2		Li et al., 2022	qPCR	47.7	3.0
Rhizochromulina		Li et al., 2022	qPCR	14.1	1.2
Triparma		Li et al., 2022	qPCR	8.2	0.8
Florenciella sp.		Li et al., 2022	qPCR	31.1	2.0
Symbiodinium kawagutii	dinoflagellate	Gong et al. 2019	bioinformatics	288.7	160.5
Symbiodinium minutum		Gong et al. 2019	bioinformatics	220.9	116.0
Amphidinium carterae		Zhu et al. 2005	qPCR	432.0	83.7
Prorocentrum nux		Zhu et al. 2005	qPCR	79.9	22.1
Extubocellulus spinifer		Godhe et al. 2008	qPCR	150.0	19.0
Phaeodactylum triconutum	diatom	Gong et al. 2019	bioinformatics	101.6	4.2
Thalassiosira oceaniaca		Gong et al. 2019	bioinformatics	179.5	17.2
Nitzschia closterium		Zhu et al. 2005	qPCR	640.0	28.0
Thalassiosira weissflogii		Zhu et al. 2005	qPCR	471.0	14.8

798

Supplementary Table S2. Stepwise regression analysis revealed significant variables contributing to the prevalence of mixotrophs in both surface and chlorophyll maximal layer (SUR&CML) samples, and surface (SUR) samples alone. Significance of P values and coefficient for 8 selected predictors were given in each row.

SUR&CML			SUR		
Predictor[†]	**P**[‡]	**Coefficient**	**Predictor**[†]	**P**[‡]	**Coefficient**
PAR	**	0.1	CML	***	0.15
NO_3^-&NO_2^-	***	-0.08	*Prochlorococcus*	**	-0.82
Density	***	0.80	NO_3^-&NO_2^-	***	-0.15
PIC	**	207.5	PIC	*	218.3
CML	***	0.10	PO_4^{3-}	*	0.28
Hbac	**	1.25	Density	*	0.78
Prochlorococcus	*	-0.53	PAR	**	0.12
/	/	/	NPP	.	0.08

[†]Abbreviations: PAR, photosynthetically available radiation; PIC, particulate inorganic carbon; Hbac, heterotrophic bacteria; NPP, net primary production

[‡] Significant codes: '***' p<0.001, '**' p<0.01, '*' p<0.05, '.' p<0.1

807

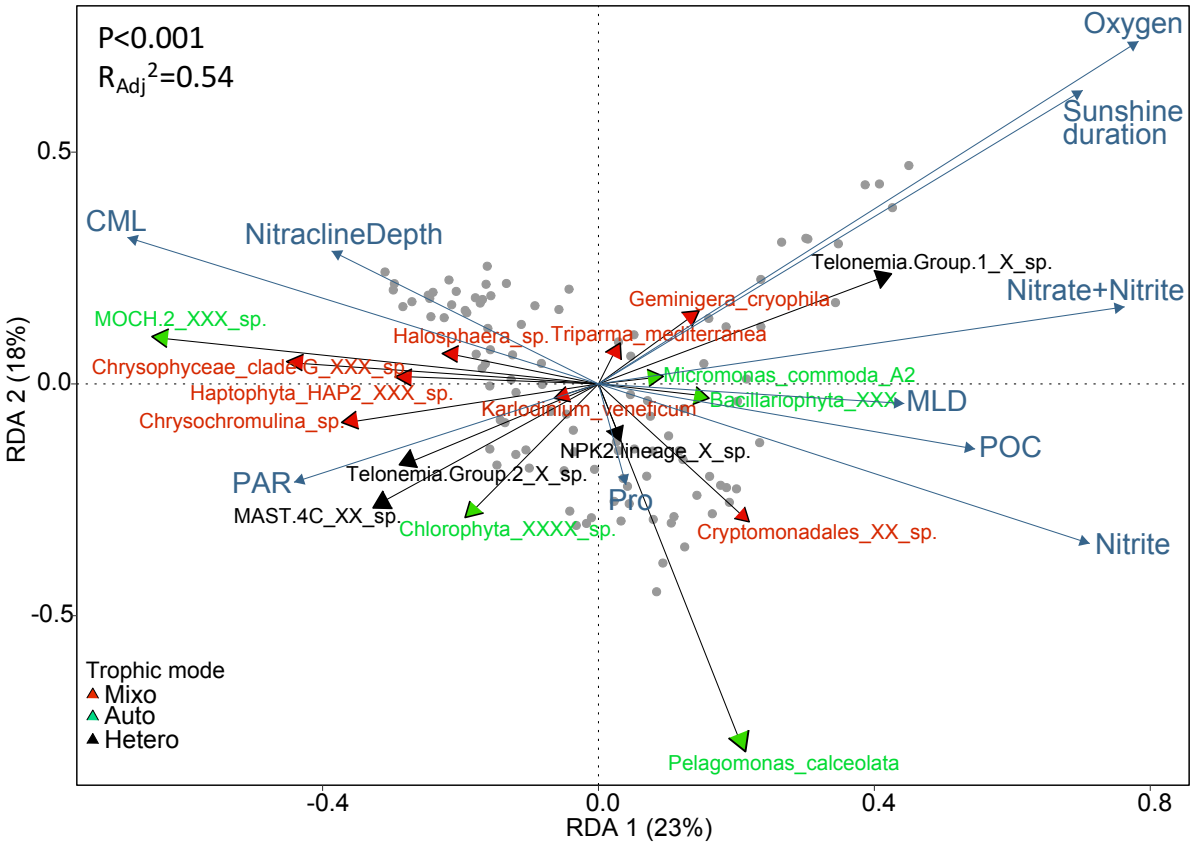
808 **Supplementary figures**

809 **Fig. S1.** Redundancy analysis for environmental impact on the distribution of trophic groups
810 in surface oceans, derived from 18S rRNA gene corrected cell abundances. Only 17 *Tara*
811 *Oceans* lineages passed Escoufier selectin (0.9 threshold) were shown, and adjusted R^2 of the
812 analysis is of 0.54. Environmental variables selected were denoted in cyan-blue arrows and
813 text. Abbreviations: PAR, photosynthetically active radiations; CML, chlorophyll maximum
814 layer; MLD, mixed layer depth; POC, particulate organic carbon; PIC, particulate inorganic
815 carbon; Pro, *Prochlorococcus*.

816 **Fig. S2.** Linear regression between $TI_{A/H}$ (all in cell abundances) and resource variables
817 retrieved from the SUR (104 samples) and SUR&CML dataset (165 samples). Different
818 significance levels were marked with different number of red stars (one star if $0.001<p<0.05$
819 and two stars if $p<0.001$), and shaded bands are pointwise 95% confidence interval on the
820 fitted values (the line). Abbreviations: PAR, photosynthetically active radiations; N:
821 inorganic nutrients of nitrate (NO_3^-) and nitrite (NO_2^-); Hbac: heterotrophic bacteria.

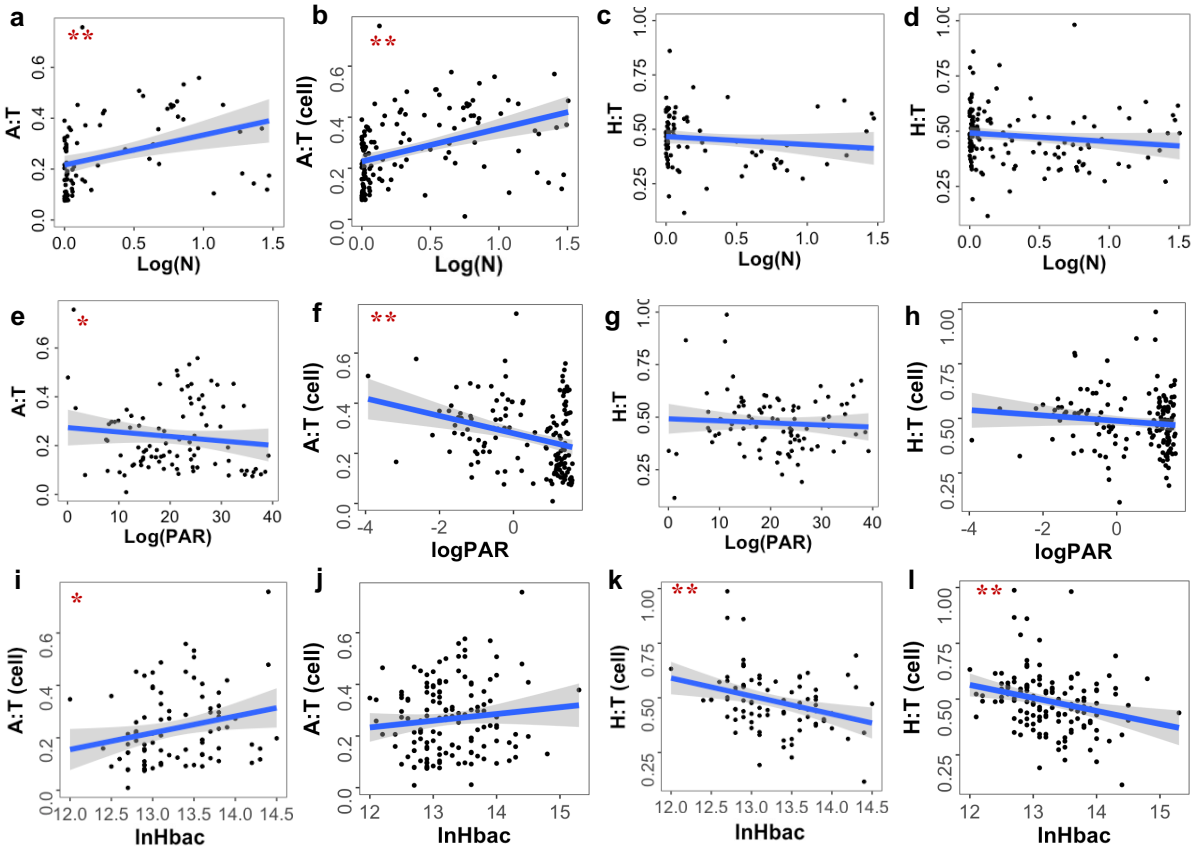
822 **Fig. S3.** Linear regression between TI_g (all in cell abundances) and single environmental
823 variables that have shown significant correlations, retrieved from all 165 SUR&CML
824 samples. Different significance levels were marked with different number of red stars (one
825 star if $0.001<p<0.05$ and two stars if $p<0.001$), and shaded bands are pointwise 95%
826 confidence interval on the fitted values (the line). Abbreviations: CML, chlorophyll maximal
827 layer; Chl a, total chlorophyll a, Pro, *Prochlorococcus*.

Fig. S1


P<0.001
R_{Adj}^2=0.54
Oxygen
Sunshine duration
CML
NitraclineDepth
Telonemia.Group.1_X_sp.
Nitrate+Nitrite
Geminigera_cryophila
MOCH.2_XXX_sp.
Halosphaera_sp.
Triparma_mediterranea
Chrysophyceae_clade_G_XXX_sp.
Micromonas_commoda_A2
Haptophyta_HAP2_XXX_sp.
Bacillariophyta_XXX
MLD
Chrysochromulina_sp.
Karlodinium_veneficum
POC
NPK2.lineage_X_sp.
PAR
Telonemia.Group.2_X_sp.
Pro
MAST.4C_XX_sp.
Chlorophyta_XXXX_sp.
Cryptomonadales_XX_sp.
Nitrite
Pelagomonas_calceolata
Trophic mode
Mixo
Auto
Hetero
RDA 2 (18%)
RDA 1 (23%)
0.5
0.0
-0.5
-0.4
0.0
0.4
0.8

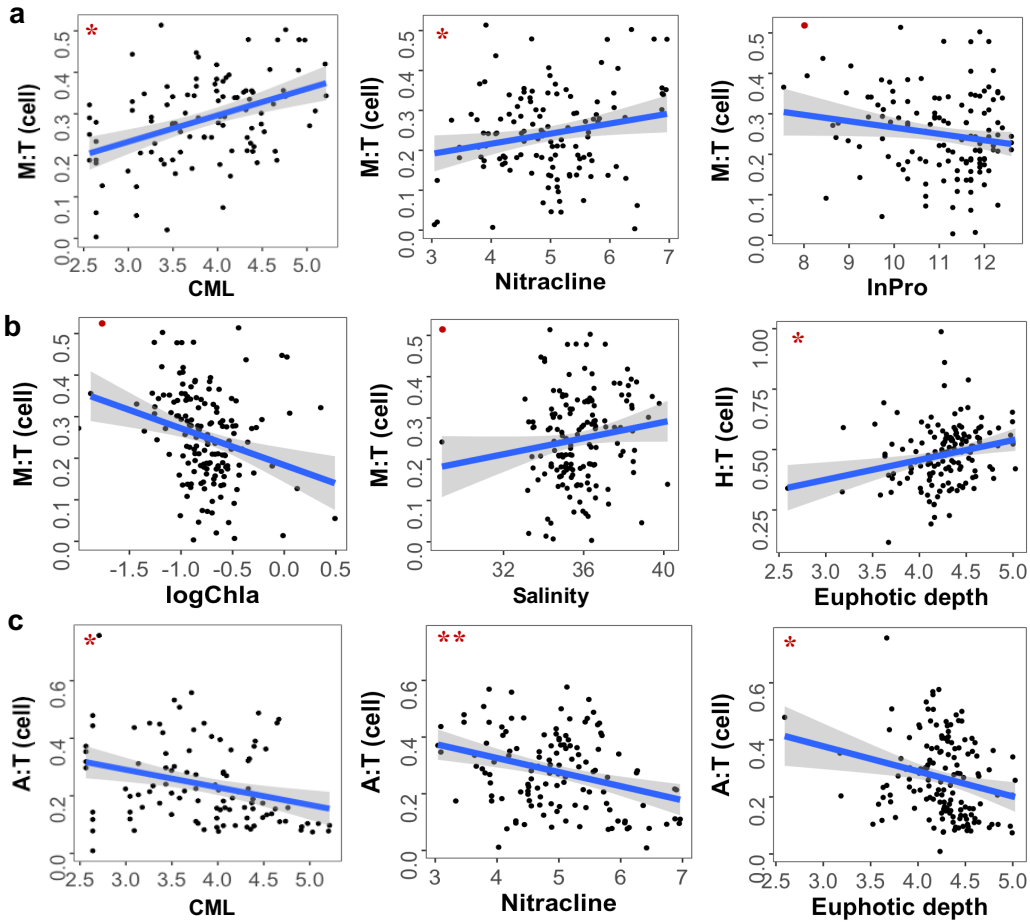
828

Fig. S2


a
b
c
d
e
f
g
h
i
j
k
l
A:T
A:T (cell)
H:T
H:T (cell)
Log(N)
Log(PAR)
logPAR
lnHbac
**
*

829

Fig. S3


a
M:T (cell)
CML
Nitracline
InPro
b
M:T (cell)
logChla
Salinity
H:T (cell)
Euphotic depth
c
A:T (cell)
CML
Nitracline
Euphotic depth

830

References

Godhe, A., Asplund, M. E., Härnström, K., Saravanan, V., Tyagi, A. and Karunasagar, I. (2008). Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR. Appl. Environ. Microbiol. 74: 7174–7182. doi:10.1128/AEM.01298-08

Gong, W.D. and Marchetti, A. (2019). Estimation of 18S Gene Copy Number in Marine Eukaryotic Plankton Using a Next-Generation Sequencing Approach. Front. Mar. Sci., Sec. Mar. Mol. Bio. Ecol. doi: 10.3389/fmars.2019.00219

Li, Q., Edwards, K. F., Schvarcz, C. R. and Steward., G. F. (2022). Broad phylogenetic and functional diversity among mixotrophic consumers of Prochlorococcus. The ISME Journal 16(6): 1557–1569. doi:10.1038/s41396–022–01204–z

Zhu, F., Massana, R., Not, F., Marie, D. and Vaulot, D. (2005). Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the18S rRNA gene. FEMS in Microbial Ecology 52(1): 79–92. doi:10.1016/j.femsec.2004.10.006