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[12*]

[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;

Email of coauthors: Kaiyi Dong dongkaiyi@sjtu.edu.cn; Ying Wang wyhzau2016@163.com; Wenjing Zhang 022440910025@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 (Forthcoming on Dryad). 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 eukaryotic18S 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

Single-celled eukaryotes that are capable of phago-mixotrophy in the ocean compete inorganic nutrients and light with autotrophs and bacterial prey with heterotrophs, yet the role these factors play in shaping the success of mixotrophs remain ambiguous. In this study, we ask what environmental variables could impact the prevalence of small eukaryotic mixotrophs in vast open ocean, and what role does nutritional resource play in these interactions. We used the *Tara Oceans* eukaryotic 18S rRNA gene dataset and environmental variables to conduct large-scale field based analysis, and isolates based culture experiments to verify growth and nutritional resource relationships for representative mixotorphs. The field observation showed that comparing to autotrophic and heterotrophic eukaryotes, mixotrophs were favored by environments with lower nutrients and higher light availability. Bacterial prey and abiotic conditions including temperature and depths of nitracline, maximal chlorophyll and mixed layer also presented somewhat correlations with mixotrophic prevalence but to less significant extent. The culture experiments demonstrated that most mixotrophs could maintain similarly high growth rates when receiving sufficient light and nutrients, either as inorganic form or organic bacterial prey, but not necessarily with low light and high nutrients conditions. This study revealed close relationship between abiotic and biotic nutritional resource and prevalence of mixotrophy, shedding light on the importance of light and nutrients for determining 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, and through metabolic activities such as photosynthesis and phagocytosis, they control the flux of elements and energy between primary producers and consumers. All protists with inherent chloroplasts were traditionally treated as strict autotrophs (only photosynthesize), and those without inherited chloroplasts were treated as strict heterotrophs (only phagocytose). It has become now increasingly evident that many of those autotrophs/heterotrophs are in fact mixotrophs, phagotrophic mixotrophs in particularly, who are capable of both photosynthesis and phagocytosis (Sher and Sher, 2002; Stoecker et al., 2017; Millette et al., 2023). 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 current study) are normally small-size flagellated eukaryotes feeding on mostly bacterial prey (Bock et al., 2021; Li et al., 2022), whereas non-constitutive mixotrophs are larger-sized microplankton (20-200 μm) which graze on both eukaryotic and bacterial prey (Flynn et al., 2013; Mitra et al., 2016). Due to their distinct metabolic activities and trophic efficiencies, trophic (group) compositions can have substantial impact on biogeochemistry and function of the ecosystem.

In a given ecosystem, mixotrophs and autotrophs compete (inorganic) nutrients and light for autotrophic growth, and bacterial prey (including autotrophic bacteria) with bacterivorous heterotrophs for heterotrophic growth. Although results may vary, previous studies have suggested the important role these nutritional resource (nutrients, light and prey) play in terms of physiological performance and competitive advantages for mixotrophs (Rothhaupt,

1996; Thingstad et al., 1996; Anderson et al., 2018; Edwards, 2019). The principle is that mixotrophy could be highly beneficial when single trophic mode is not sufficiently sustained. Ward (2019) has proposed a conceptual model to claim 'mixotrophy is more than the sum of its parts' due to its highly flexible metabolism. Meanwhile, it is necessary to acknowledge the tradeoff behind this universal mixotrophic trait, that is, photosynthesis and/or phagocytosis performance alone is very likely penalized (Raven, 1997). Alongside these metabolic tradeoffs, complex biological interactions including the microbial loop (bacteria-DOC-nutrients) add on additional uncertainty for competitive outcomes among those three trophic groups. Nevertheless, it is logical to speculate that nutrients, light and bacterial prey are among the most crucial drivers affecting growth and fitness of these trophic groups.

Field observations have shown mixotrophy can be advantageous on community level when conditions are not ideal for autotrophy, such as mature state (summer season) of temperate sea (Mitra et al., 2014; Hansen et al., 2019; Leles et al., 2021) and oligotrophic ocean (Zubkov and Tarran 2008; Hartmann et al., 2012; Edwards, 2019) or lakes (Tittel et al. 2003; Hansson et al., 2019; LeNoac'h et al., 2024). In another word, conditions that are unfavorable for strict autotrophy or heterotrophy could possibly select for mixotrophy. With climate change, evidence have shown that mixotrophs may play a more important role due to their metabolic plasticity (Wilken et al., 2013; Berge et al., 2017; Lepori-Bui et al., 2022; Chu et al., 2023). All these factors make it challenging yet critical to understand the relationship between environment and trophic composition of protists in the ocean. Motivated by these findings, we are particularly interested to delineate the prevalence and favored niche of mixotrophs versus strict autotrophs and heterotrophs in vast open oceans across the globe.

The goal of this study is to gain insights on ocean environments that could possibly affect the relative success of mixotrophs and underlying mechanisms behind it. We aim to answer the following scientific questions, 1) What are the overall composition and distribution pattern of mixotrophic protists 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 in those selecting processes? We use the eukaryotic 18S rRNA gene dataset from *Tara Oceans Expedition* to assign trophic mode to each lineage/specie and assess relationship between trophic composition and environmental variables. We then run culture experiments to confirm relationships between representative mixotrophic isolates and key variables identified from the field (*Tara Oceans*). The *Tara Oceans* analysis revealed factors that are potentially important for shaping the trophic composition, and the experimental results provided detailed and metabolic evidence for it.

Materials and Methods

Study regions

Tara Oceans Data generated by *Tara Oceans expedition* between 2009 and 2013 are particularly suited to study the genetic and functional diversity of plankton (http://www.pangaea.de). In this study, a total of 104 stations and 165 samples were investigated, among which all 104 sites have surface samples and 61 sites had both surface and chlorophyll maximal depth. A surface dataset (SUR) that had 104 samples and surface and DCM dataset that include 122-165 samples (SUR&DCM) were generated for different statistical analysis. These samples were collected during four different seasons (Spring,

Summer, Autumn, Winter) between 2009-2013 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). Bathy depths varied from 13 to 5964 km and distance to coast are between 0.9-1404 km.

Contextual environmental variables

We built an environmental context datasets for further statistical analysis following these three steps. First, representative environmental variables were downloaded 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. (2014), Ardyna et al., (2017); Ibarbalz et al., (2019); Faure et al., (2019); Karlusich et al. (2022) etc. (**Supplementary Dataset S1**). Major approaches of retrieving these variables were in situ sensors, lab analysis of water samples, climatological instruments and model simulations, following standard protocols and analytical methods (details in those previously listed literatures). Second, these over 100 variables were manually filtered down to 77 (by eliminating similar and less relevant ones) to reduce colinearity and inflation of dimensionality. Examples include salinity was chosen over conductivity, density was chosen over Sigma-theta, and 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 or suitable one (for explaining community compositions) 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 measurement and calibrated with non-photosynthetic quenching and discrete water sample measurements were used among other available measurements. Additional comparisons were also done to confirm there were no overall significant differences between the chosen one and other measurements for the same variable, e.g. 8-day averaged PAR vs 30-day averaged PAR, total Chl *a* derived from sensor vs water samples vs MODIS, and nutrients concentrations derived from water samples and model simulation etc.

The final contextual environmental dataset generated 29 variables (**Supplementary Dataset S2)**, representing physical conditions such as temperature, salinity, density, depth of chlorophyll maximum (DCM), 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 (NO3+NO2, NO2, PO4, SiO4), PH, total alkalinity, total carbon, fluorescence, particulate organic carbon (POC), particulate inorganic carbon (PIC), attenuation coefficient of PAR (Kd PAR), and bacterial abundances including autotrophic (*Prochlorococcus* and *Synechococcus*) and heterotrophic bacteria. Note that longitudes, latitudes, distance to coasts, sampling depths and bathy depths were not included for the analysis to better focus on the aforementioned climatological, biochemical and physical variables.

18S rRNA gene abundances and corrected cell abundances

Detailed genomic sample process methods were described in de Vargas et al. (2015). Briefly, community genomic DNA were extracted from cells filtered through membrane

filters with different pore sizes (0.8μm, 5μm, 20μm) after pre-filtration with nylon sieves (to retrieve different plankton communities). In this study, we focused on the 0.8-5μm size range which contributed the majority of eukaryotic community in sampled regions (de Vargas et al., 2015). Eukaryotic V9-18S rRNA gene region were sequenced through Illumina HiSeq, assembled, filtered and classified with SWARM for OTUs. The sequence files were downloaded from the European Nucleotide Archive under project number PRJEB6610. Then we re-assigned OTUs to the lowest taxonomic rank possible referring to the Protist Ribosomal Reference (PR2) database (Guillou et al., 2013). Multicellular organisms from metazoan and larger plankton of Ciliophora, Radiolaria and Foraminifera that were unlikely the dominant populations for small eukaryotes below 5 μm were removed from further analysis. This final gene dataset has generated 64,010,687 sequences and 881 lineages/species for further analysis.

Eukaryotic community are known to have substantially different 18S rRNA gene copy numbers，varying from as low as one copy in flagellates belonging to Ochrophyta (Li et al., 2022) to thousands of copies per cell in dinoflagellates (Zhu et al. 2005). Overall cellular biovolume have been shown to be positively correlated with 18S rRNA gene copy numbers but certain taxa tended to have particularly high (skewed) copy numbers, such as dinoflagellates (Gong et al., 2019; Martin et al., 2022). To correct for this bias and obtain gene-converted-cell abundances, we first collected empirical data from broad eukaryotic populations, focusing on small cells with equivalent diameter <5-10 μm and biovolume <65-523 μm^3 (**Supplementary Table S1**), then fit a linear correlation equation between empirical 18S rRNA gene copy numbers and cellular biovolume for individual group of dinoflagellates, diatom and other small flagellates (haptophytes, dictyochophytes, chlorophytes etc.). Finally, a correction factor (C.F.) for each group was calculated using upper limit biovolume of 65 μm^3, that gives 59.3, 10.5 and 4.7 for dinoflagellates, diatom and all other small flagellates. For comparison, Martin et al. (2022) used 27.1, 4.4 and 0.9 C.F. for dinoflagellates, diatom and all other flagellates. Gene abundances were converted into cell abundances by diving gene numbers with C.F. for each group. The corrected cell abundance was the focus for further analysis but 18S rRNA gene abundances were also assessed occasionally for comparison purpose.

Assignment of trophic mode

We focus on the trophic strategy potentials which reflect long-term eco-evolutionary results among different lineage, rather than the active mixotrophic activities which are associated with in situ physiological and environmental conditions. Therefore, we assigned each lineage/specie (when possible) into three trophic groups, including (potential) mixotrophs (pigmented eukaryotes with indicated phagocytosis capability), heterotrophs (flagellated protists and absence of inherent chloroplasts) and autotrophs (pigmented protists has no available evidence of phagocytosis) by searching evidence among published database and references (e.g. Leles et al., 2019; Li et al., 2022; 2023; Martin et al., 2022; Mitra et al., 2023). When trophic annotation is impossible at lineage or specie level, their affiliated genus or family was referred to for annotation. A small portion of parasites (e.g., Fungi, Amoebozoa, Syndiniales) and symbionts (e.g. Dinoflagellata) were assigned as 'others' and not included for further analysis. Some of the uncharacterized or uncultivated taxa were assigned based on trophic mode from phylogenetic relatives. The filtration has generated a

final trophic composition dataset with 58,001,584 sequences and 842 species (**Supplementary Dataset S3**). We acknowledge the limitation of this trophic annotation methodology, and address them further in the Discussion.

Trophic index and trophic composition

The relative prevalence or success of each trophic group at a given site was expressed as the trophic index of each group (TI_g), where g represents three trophic groups of mixotroph (M), autotroph (A) and heterotroph (H). Therefore, TI_g of TI_M, TI_A, and TI_H can be quantified as their relative abundance against total eukaryotic community (T), either in cell or gene abundances (expressed as M:T, A:T and H:T). The combined three index of TI_M, TI_A and TI_H can represent the overall picture of trophic composition at a given site.

Statistical analysis

Surface dataset (SUR) were used for redundancy analysis (RDA) to retrieve an overall distribution pattern of trophic groups among 104 stations, and potential relationship against environmental variables. The 29 filter contextual variables were all considered for the analysis, with missing values for each variable replaced with means and then log-transformed for standardization prior to analysis. Hellinger transformation was done to species abundances and escoufier selection with a 0.9 threshold was applied to the trophic species. Environmental variables passed significance test and colinearity test were presented on the RDA plot. The RDA was carried out in R with the vegan package, following Faure et al.'s script (2019). Principal components analysis (PCA) was applied to extract the major components that explaining the environmental gradients from both SUR and SUR&DCM datasets with R package pcaMethods (Stacklies et al., 2007). PCA was carried out for both SUR (104 samples) and SUR&DCM dataset (165 samples) separately to achieve the most comprehensive results. After main variable components were extracted, TI_g index were plotted over two axes of PC1 and PC2 with generalized additive model. Stepwise regression between all variables and TI_M were conducted to retrieve important variables explaining mixotrophic prevalence, and linear regression between single variables and TI_g were carried out to provide additional correlation results. ANOVA was conducted for the culture experiments to reveal factorial impact on growth rates and grazing rates among various mixotrophic isolates.

Culture experiments

Methods for isolating and cultivating mixotrophs from North Pacific Subtropical gyre was decribed in Li et al. (2021; 2022). Briefly, mixotrophs were isolated from euphotic zone samples at Station ALOHA (22° 45' N, 158° 00' W) by enriching in nitrogen-depleted Keller medium (K-N medium) that supplemented with bacterial prey. Six identified isolates belonging to different classes and phylogenetically close to the abundant *Tara Oceans* lineages were selected for grazing and growth experiments. Prior to experiments, all isolates were acclimated for 1-2 months (duplicated over 10 generations) under corresponding treatment conditions, i.e., low nutrients (K-N medium) and high light, low nutrients and low light, high nutrients (K medium) and high light, and high nutrients and low light. The high and low light treatments were receiving ~100 and ~10 μM photons m^{-2} s^{-1} irradiance on a 12 h:12 h light:dark cycle. All treatments were supplemented with gently centrifuged and washed bacterial prey of *Prochlorococcus* (str. MIT9301) at final concentrations of 10^6 cells mL^{-1} to estimate grazing capabilities of examined mixotrophs. Controls without added prey

255 or grazers were also set up at each condition for comparison purposes. Cell concentrations of
256 prey and grazers were measured every 12–24 hours by flow cytometry and ingestion
257 evidences were occasionally obtained via microscopic observation.
258 Ingestion rates, *I* (prey grazer^{-1} h^{-1}) for each grazer were calculated as the amount of prey
259 removed between sampling time t+1 and t, divided by averaged grazer concentrations over
260 the same time period. Clearance rates were calculated by dividing the ingestion rate by
261 average prey concentration over the same interval, and specific clearance rate (body volume
262 grazer^{-1} h^{-1}) was calculated by dividing the clearance rate by cellular biovolume (μm^3) of
263 each grazer (measured under microscopy). Clearance rates reported in the results were
264 averaged over all sampling intervals prior to grazers reaching stationary density. Growth rates
265 presented in the results were averaged mean rates during exponential growth phase (slope of
266 Ln transformed cell abundance vs time).
267 **Results**
268 **Trophic composition of eukaryotes across the *Tara Oceans* sampling sites**
269 The 104 surface samples encompassed eight ocean habitats on large spatial scale (**Fig.**
270 **1a**). Geographically they span roughly four latitudinal regions, from low (0º-20º) to
271 low-medium (20º-30º), medium (30º-40º), and medium-high latitude (40º-60º), with overall
272 decreasing PAR (from 27.3 to 14.7 mol m^{-2} d^{-1}), increasing nutrients (from 0.05 to 17.3 μM
273 NO3+NO2), and over stable heterotrophic bacterial abundances (4.3-5.3 ×10^5 cells mL^{-1})
274 (**Table 1; Fig. 1b**). The globally averaged means were 21.3 mol m^{-2} d^{-1}, 0.14 μM and 4.5
275 ×10^5 cells mL^{-1} for PAR, NO3+NO2 and heterotrophic bacteria, demonstrating typical open
276 ocean environments. Smaller-regional variations were also present. The higher nutrients
277 concentrations in low latitude regions were possibly due to Peruvian coast and eastern
278 equatorial upwelling, and lower PAR in medium latitudes was because Mediterranean Sea
279 were sampled in winter. Three trophic index of TI_M, TI_A, and TI_H varied from 0.23-0.34,
280 0.10-0.32 and 0.43-0.51 in cell abundance, and 0.29-0.45, 0.08-0.24 and 0.38-0.46 in gene
281 abundance, respectively. TI_M in cell abundance were overall somewhat lower than gene
282 abundance. Nevertheless, the highest TI_M median values were both retrieved from
283 low-medium latitude, and the lowest values were from low latitude (cell abundance) and
284 medium-high latitudes (gene abundance), respectively (**Table 1; Fig. 1c**). In comparison,
285 high and low TI_A median values were from low and medium-high latitude and low-medium
286 latitude (both cell and gene abundance). TI_H did not change as much across latitudinal
287 gradients, but highest median values in cell and gene abundance were shown in low-medium
288 and medium-high latitudes, respectively.
289 Before the gene-cell correction, mixotrophs were mostly composed of Dinoflagellata,
290 Haptophyta and Ochrophyta, and with corrected cell abundances, dominant mixotrophs were
291 retrieved from Haptophyta and Ochrophyta (**Fig. 1d**). Proportions of dinoflagellates
292 decreased for heterotrophic community after gene-cell conversion, and the rest abundant
293 heterotrophs were belonging to phyla of Telonemia, Opalozoa and Discoba, both in cell and
294 gene abundances. Differences for autotrophic compositions on phyla level in cell and gene
295 abundances were insignificant, which were both dominated by Chlorophyta and Ochrophyta.
296 Redundancy analysis (**Fig. 1e; Supplementary Fig. S1**) suggested stations from the same
297 ocean or latitudinal regions tended to cluster together. Certain species were more abundant in
298 lower latitudes, such as Chrysophyceae_Clade G_sp. of Ochrophyta, *Chrysochromulina* sp.

and Haptophyta_HAP2_XXX_sp. of Haptophyta, and *Karlodinium veneficum* of Dinoflagellata for mixotrophs, MOCH.2_XXX sp. and Chlorophyta_XXXX_sp. of Chlorophyta, and MAST.4C_XX sp. and Telonemia Goup2_X sp. for heterotrophs. Those taxa were positively affiliated with PAR and DCM, and negatively related with nutrients (NO3+NO2 and NO2) and other variables located on the right side of the RDA plot. Other species, including *Geminigera Cryophila* of Cryptophyta and *Triparma Mediterranean* of Ochrophyta for mixotrophs and Telonemia Group1 for heterotrophs, were abundant at higher latitude and positively correlated with oxygen and sunshine duration. *Micromonas* A2 and Bacillariophyta_XXX sp. for autotrophs were associated with high MLD, POC and NO2; and Cryptomonadales_XX_sp. for mixotrophs and NPK2 lineage_X_sp. for heterotrophs were affiliated with high abundances of *Prochlorococcus*.

Variables contributing to trophic composition

Among the 29 selected predictors, availability of nutrients (including NO3+NO2 and silicate), PAR and temperature were the top variables revealed by principle component analysis explaining environmental variations for both surface and surface & DCM datasets (**Table 2; Figs. 2&3**). Specifically, both PC axes (R^2 of 0.30) for the surface and DCM samples were mostly explained by PAR and nutrients (loading coefficients varied from -0.42 to 0.17). When only considering surface samples, PC1 axis (R^2 of 0.35) was mostly contributed by nutrients and temperature, while PC2 axis (R^2 of 0.24) were mostly explained by MLD, DCM and PAR.

Generalized additive model fitted smoothers have demonstrated the most substantial responses by mixotorphs to changing principle components. Specifically, there was a decreasing trend of TI_M along PC1 for both datasets, indicating decreasing fitness of mixotroph with increasing nutrients (surface) (**Fig. 2a**), as well as decreasing PAR and increasing nutrients (surface & DCM) (**Fig. 2d**). TI_M did not demonstrate changes with the PC2 axis from surface 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 most steep among all plots. For PC2 axes of surface &DCM 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 surface dataset (**Fig. 2b**), and increasing nutrients and decreasing PAR for the surface & DCM dataset (**Fig. 2e**). When nutrients and PAR changed in the same direction (PC2 for surface & DCM dataset), TI_A did not show clear variation (**Fig. 2h**). TI_H overall varied much less with changing PAR and nutrients availability comparing to TI_M and TI_A (**Figs. 2c, f, i**), among which a weak unimodal shape was observed with PC1 from surface dataset (**Figs. 2c**). For the gene abundances, overall similar trends were found (**Supplementary Fig. 3)**. Somewhat steeper decreasing slope for TI_M on PC1 for surface samples and increasing slope (before the peak) on PC2 for surface & DCM samples (**Supplementary Fig. 3a, g**). In comparison, the weak increasing trend of TI_H over surface sample PC1 disappeared (**Supplementary Fig. 3c)**, and a somewhat stronger increase of TI_H were demonstrated along both axes of surface & DCM dataset (**Supplementary Fig. 3f, i).**

Consistent with the PCA results, clear differences were observed for TI_g between surface and DCM samples, accompanied with distinct resources availability of nutrients, light, and

343 bacterial prey (**Fig. 4**). Significantly higher TI_M were seen in surface where PAR and
344 heterotrophic bacteria were higher and N (NO3+NO2) were lower (median 0.29 vs 0.18,
345 $P<0.001$, df=60). In comparison, TI_A were slightly higher in DCM comparing to surface but
346 not statistical significant (median 0.30 vs 0.29, $P>0.05$, df=60), and TI_H were significantly
347 higher in DCM than surface (median 0.52 vs 0.47, $P<0.05$, df=60) (**Fig. 4a, b**). Linear
348 regression between TI_M and these single variables have suggested a consistently positive
349 relationship with PAR and negative correlation with nutrients (NO3+NO2), from both
350 surface and surface & DCM dataset (**Fig. 4c, d**). TI_A demonstrated significantly positive and
351 negative correlations with nutrients and PAR, whereas neither was significantly associated
352 with TI_H (**Supplementary Figs. S2a-h**). For heterotrophic bacteria, a positive but weaker
353 relationship ($P=0.02$) were found between TI_M and the surface & DCM dataset but not the
354 surface dataset (**Fig. 4c, d**). In contrast, both datasets have demonstrated strong negative
355 associations between heterotrophic bacteria and TI_H, and a weaker but positive correlation
356 with TI_A in the surface ($P<0.05$) (**Supplementary Fig. S2i-l**). Other single variables showed
357 significant regressions with TI_M were all shown in **Supplementary Fig. S3**.

358 **Mixotrophic growth with varying nutritional resource**

359 The top abundant *Tara Oceans* mixotrophs were belonging to classes of
360 Prymnesiophyceae, Dictyochophyceae, Chrysophyceae, Dinophyceae and Cryptophyceae
361 (**Fig. 5a**). Mixotrophs we isolated from North Pacific Subtropical Gyre that have been shown
362 to be efficient bacteria (including *Prochlorococcus*) grazers are closely affiliated to those
363 abundant *Tara Oceans* lineages, i.e., Chrysochromulina sp. (Chrysochr.) and
364 Prymnesiophyceae_XXX_sp. (PrymXXX) of Prymnesiophyceae,
365 Chrysophyceae_Clade-H_X_sp. (ChrysoH) of Chrysophyceae, and *Florenciella Parvula* (F.
366 Parvula), Dictyochales_X_sp. (DictyX) and Florenciellales_X_sp. (FloX) of Chrysophyceae
367 (Li et al., 2021; 2022). The strain of FloX was not shown in Fig. 5a but ranked as the 26th
368 most abundant lineage across all stations. The relative abundances of those isolates retrieved
369 from surface were all higher than DCM except for DictyX (**Fig. 5b**), presenting similar
370 patterns as the entire mixotroph community (TI_M results in **Fig. 4a**). Considering the
371 consistent and substantial responses of TI_M to nutrients and PAR derived from statistical
372 analysis and field observations, we conducted experiments to assess impacts of different light
373 (high and low) and nutrients (inorganic and organic nutrients) availability on growth of those
374 six representative mixotrophs (**Fig. 5c**).

375 Distinct growth curves (increase slopes and final abundances) were presented among
376 different treatments for the six isolates (**Fig. 6a**), accompanied by different disappearance
377 rates of prey (**Fig. 6b**). Prey in controls (without mixotrophic grazers) all presented slower
378 disappearance rates comparing to cultures with grazers but to different levels among strains,
379 and grazers in control (without prey) grew to much lower concentrations, i.e, all $<1\times10^3$ cells
380 mL^{-1} in low nutrients (NO3) medium and similar concentrations to those with supplemented
381 prey in high nutrients medium (data not shown). Shown as exponential growth rates (bars in
382 **Fig. 6c**), the majority isolates presented the highest values when grown under high light (100
383 $\mu E\ m^{-2}\ s^{-1}$), either receiving high nutrients or sufficient prey. Lower growth rates were seen
384 with low light treatments for certain isolates, including Chrysochr., PrymXXX and FloX. For
385 grazing rates, estimated as clearance rates (biovolume $grazer^{-1}\ h^{-1}$), both low nutrients or
386 low-light conditions could stimulate the overall grazing rates, indicated by longer bars and

387 treatment cartons in **Fig. 6d**. In contrast, when mixotrophs were limited by both nutrients and
388 light, clearance rates could be significantly lower for some strains, such as DictyX and FloX.
389 ANOVA results have suggested that significant effects of light were observed among all
390 six isolates, i.e, to increase growth rates with increasing light, either alone, or together with
391 nutrients and species. Nutrients availability did not show significant impact on mixotrophic
392 growth rates as cultures were supplemented with prey. Less significant effects were seen for
393 clearance rates, as strains responded differently to changing light and nutrients conditions,
394 indicated by the high P (>0.05) values from both single and multi-factorial ANOVA (**Table**
395 **3**).
396 **Discussion**
397 In this study, we used large-scale field survey to retrieve prominent variables that
398 possibly affect the trophic composition of eukaryotic community in the field. We then
399 conducted finely controlled lab experiments using various mixotroph isolates to verify the
400 field observation. While acknowledging species-specific variations, field and experimental
401 results were overall consistent, proving insights into the close relationship between trophic
402 composition and nutritional resource (including nutrients, light and bacterial prey), as well as
403 well distinct niche preferences of mixotrophic eukaryotes versus autotrophs and heterotrophs
404 in the ocean.
405 **Key variables affecting mixotrophic prevalence**
406 Indicated by our field analysis, higher nutrients and/or lower light can promote the
407 overall success of autotrophs and oppositely, mixotrophs could possibly proliferate from low
408 nutrients and high light environments. Physical conditions of DCM, nitracline depths and
409 MLD also seemed to affect mixotrophic prevalence, though they were more or less associated
410 with nutrients and light conditions. For example, deeper nitraclines can potentially suppress
411 nutrients supply in the surface and therefore increase $TI_{M.}$ Similarly, deeper chlorophyll
412 maxima could be indicative of higher light availability and lower nutrients in photic depths
413 and can thus enhance TI_M as well. In contrast, deeper mixed layer depths (stronger mixing)
414 can reduce light availability and increase nutrients concentration, which both weakens TI_M.
415 The overall positive association between DCM and nitracline depths and TI_M, and negative
416 correlation between MLD and TI_M were in line with their overall relationships with nutrients
417 and light.
418 Heterotrophic bacteria maintained relatively consistent abundances in the sampled
419 oceans and was not selected as main explanatory factor for TI_g by RDA or PCA. Nevertheless,
420 they demonstrated as one of the predictors explaining TI_M based on stepwise regression
421 model (**Supplementary Table S2**). When treated as a single variable, a strong negative
422 correlation was found between TI_H and heterotrophic bacteria. Comparatively, the correlation
423 was positive for TI_A and less insignificant for TI_M. This interesting shifts likely suggested
424 distinct interactions between bacteria and three trophic groups. Heterotrophic bacteria benefit
425 from the dissolved organic carbon (DOC) excreted by autotrophs and recycle nutrients for
426 autotrophic growth which forms a positive relationship. Whereas for heterotrophs, this
427 negative association could be due to high grazing pressure exerted on heterotrophic bacteria
428 (Fuhrman and Noble, 1995; Pernthaler, 2005). The weaker but positive correlation between
429 heterotrophic bacteria and TI_M could be explained by the more complex relations among
430 bacteria (prey and decomposer) and mixotrophs (grazer and DOC provider), as well as the

lower grazing capability/pressure from mixotrophic grazers than heterotrophs (Raven 1997; Berge et al., 2017; Edwards et al., 2023; Mitra and Flynn, 2023).

Other field studies have suggested besides nutrients (Stukel et al., 2011; LeNoac'h et al., 2024) and light (Ptacnik et al. 2016; Millette et al., 2017; Anderson, et al., 2019), temperature (Lambert et al., 2021) and carbon composition (colored DOM and CO2) (Hansson et al., 2019) were also likely important for mixotrophic prevalence. Those variables were also included in our analysis and yielded varying relationships with TI_g (sometimes significant). These differences could be possibly due to the different environmental gradients covered among studies (global open oceans vs north Pacific vs boreal Canadian lakes). Stepwise regression model has suggested PAR, nutrients, density, PIC, DCM, heterotrophic bacteria and *Prochlorococcus* were the top predictors explaining TI_M with the *Tara Oceans* dataset (**Supplementary Table S2**). Future studies focusing on smaller regions with in situ experiments will provide additional insights into the trophic niche-environment relationships.

Mixotrophic physiology and fitness

Suggested by our field and experimental results, mixotrophy can likely function as an efficient strategy for compensating nutrients deficiency (through ingestion of bacterial prey), but not necessarily for supplementing light limitation. Other studies have also argued the important role light play for determining the competitive advantages of mixotrophs versus others. Derived from culture experiments and theoretical models, Rothhaupt (1996) and Edwards et al. (2019; 2022) have suggested that mixotorphs are proliferated under high light because they can outcompete heterotrophic grazers using extra energy gained from photosynthesis which suppress bacterial prey to a lower threshold. Both mechanisms could be true, and it is worthy to notice that variations exist among different strains and taxa. These globally dominant mixotrophs from small eukaryotes possibly align with the traditionally categorized Type II mixotrophs, i.e., primarily phototroph but phagocytose for nutrients (Stoecker, 1998). Based on resource allocation and plastid evolution theories, many of these small constitutive mixotrophs gain energy and carbon through photosynthesis and additional nutrients from prey (Jones, 1997; Raven, 1997; Mitra et al., 2014; Ward, 2019). Penalized photosynthetic performance, or higher requirements for light seems to be a significant trait tradeoff for them. Nevertheless, good mixotrophs (such as ChrysoH) performs better under both low light and low nutrients conditions due to higher grazing capabilities comparing to others (such as FloX), and are thus less likely penalized by light and nutrients deficiency. Instead, their success could be possibly depending on bacterial availability and other factors.

Methodology caveats

One potential caveat in our methodology is that mixotrophs were annotated based on published literature at species or higher taxonomic ranks. These trophic annotations were not necessarily aligning with in situ metabolic activities. Nevertheless, the potential or historically evolved capability of being mixotrophs are sufficient to delineate their ecological niche which was the focus of this study. We also recruited longer-term averaged values for some of the key variables (e.g. 8-day and 30-day PAR) as well as a variety of parameters (e.g. NO3, NO2, PO4 and SiO4 for nutrients) to reduce the false relationship. Lack of trophic evidence for some of the species/lineages could on the other hand, cause false annotation, and variations against other studies. For example, we annotated *Micromonas* belonging to Chlorophyta as autotrophs based on the best knowledge of ours (Jimenez et al., 2021;

Edwards et al., 2022) but others have argued possible phagotrophic evidence from the *Micromonas Polaris* (McKie-Krisberg and Sanders, 2014). We also acknowledge the gene-cell abundance correction factors could be inaccurate for certain groups, as we did not have first-hand cellular abundance data from in situ samples. Nevertheless, these calibrated abundance data could improve accuracy for some heavily over-estimated groups such as dinoflagellates and under-estimated small flagellates.

Combing various approaches including in situ grazing experiments, field observation, model simulation and single-cell tools can be powerful to retrieve the most reliable data for future mixotrophy research (Beisner et al. 2019; Millette et al., 2023).

Conclusion

This study provides the first *Tara Oceans* analysis for niche partitioning among mixotrophs, autotrophs and heterotrophs of the eukaryotic community. Our results suggested that in vast open oceans, phagotrophic mixotrophy is an efficient trophic strategy that small eukaryotes have evolved to survive nutrients impoverish, deepening our understanding about their prevalence and distinct niche in comparison to autotrophs and heterotrophs.

Table 1. Oceanographic characters and trophic index of surface water samples. Abbreviations for PAR, N+N, Hbac stands for photosynthetically available radiation, dissolved nitrate and nitrite (NO3+NO2), and heterotrophic bacteria, respectively. M:T, M:A and M:H denotes rations of either cell abundance (cell) or 18S rRNA gene abundance (gene) among mixotroph (M), autotroph (A), heterotroph (H) and total eukaryotes (T).

Latitude	Oceanic Region†	Station	Temp (℃)	Salinity	PAR (mol m^{-2} d^{-1})	N+N (μM)	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

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

Table 2. Major environmental variables retrieved by principle component analysis for surface and surface & DCM datasets. Top variables from each principle 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.

	Surface		**Surface & DCM**		
	PC1	**PC2**	**PC1**	**PC2**	**sum**[†]
PAR	-0.08	**-0.14**	**-0.42**	**-0.35**	0.99
$NO3+NO_2$	**0.31**	**-0.14**	**0.23**	**-0.25**	0.93
Silicate	**0.23**	-0.10	**0.17**	**-0.17**	0.67
Temp	**-0.24**	0.08	-0.14	0.13	0.59
Nitracline	-0.20	-0.11	-0.15	0.12	0.58
DCM	0.06	**-0.17**	-0.10	0.07	0.40
MLD	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

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

	Growth rate					**Clearance rate**				
	single factor			**multi-factors**		**single factor**			**multi-factors**	
	L	**N**	**sp.**	**L&N**	**sp.&N&L**	**L**	**N**	**sp.**	**L&N**	**sp.&N&L**
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[¶]	+	+	/	/	/	+	–	/	/	/

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

References

Anderson, R., Charvet, S. and Hansen, P. J. (2018). Mixotrophy in chlorophytes and haptophytes—effect of irradiance, macronutrient, micronutrient and vitamin limitation. Front. microbiol. 9: 1704. doi:10.3389/fmicb.2018.01704

Anderson, R. and P. J. Hansen. (2019). Meteorological conditions induce strong shifts in mixotrophic and heterotrophic flagellate bacterivory over small spatio-temporal scales. Limnol. Oceanogr. 65: 1189–1199. doi:10.1002/lno.11381

Ardyna, M. (2017) Tara Oceans Consortium, Coordinators; Tara Oceans Expedition, Participants (2017): 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. PANGAEA, https://doi.org/10.1594/PANGAEA.883614

Beisner, B. E., Grossart, H. and Gasol, J. M. (2019). A guide to methods for estimating phago-mixotrophy in nanophytoplankton. J. Plankton Res. 41: 77–89. doi:10.1093/plankt/fbz008

Berge, T., et al. (2017). Modeling succession of key resource–harvesting traits of mixotrophic plankton. The ISME Journal 11(1): 212–223.

Bock, N., et al. (2021). Experimental identification and in silico prediction of bacterivory in green algae. The ISME Journal. doi.org/10.1038/s41396-021-00899-w.

Chu, T.Y., Moeller, H.V. and Archibald, K.M. (2023). Competition between phytoplankton and mixotrophs leads to metabolic character displacement. Ecological Modelling 481. 110331; doi.org/10.1016/j.ecolmodel.2023.110331

de Vargas, C., et al. (2015). Eukaryotic plankton diversity in the sunlit ocean. Science 348(6237): 1261605.

Edwards, K. F. (2019). Mixotrophy in nanoflagellates across environmental gradients in the ocean. Proc. Natl. Acad. Sci. USA 116(13): 6211–6220.

Edwards, K. F., et al. (2022). Trophic strategies explain the ocean niches of small eukaryotic phytoplankton. Proc. Royal Soc. B: Biological Sciences 290(1991): 20222021.

Edwards, K. F., et al. (2023). Ingestion kinetics of mixotrophic and heterotrophic flagellates. Limnol. Oceanogr. 68(4): 917–927.

Faure, E., et al. (2019). Mixotrophic protists display contrasted biogeographies in the global ocean. The ISME Journal 13(4): 1072–1083.

Flynn, K.J., et al. (2019). Mixotrophic protists and a new paradigm for marine ecology: where does plankton research go now? J. Plankton Res. 41(4): 375–391. doi:10.1093/plankt/fbz026

Fuhrman, J.A. and Noble, R.T. (1995). Viruses and protists cause similar bacterial mortality in coastal seawater. Limnol. Oceanogr. 40(7): 1236-1242; doi.org/10.4319/lo.1995.40.7.1236

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

Guillou, L., et al. 2013. The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Res. 41: D597–D604. doi:10.1093/nar/gks1160

Hansson, T.H., Grossart, H. P., Giorgio, P.A. Gelais, N. F. S. and Beisner, B. E. (2019). Environmental drivers of mixotrophs in boreal lakes. Limnol. Oceanogr. 64, 1688–1705; doi: 10.1002/lno.11144

Hartmann, M., Grob, C., Tarran, G. A., Martin, A. P., Burkill, P. H., Scanlan, D. J. and Zubkov. M. V. (2012). Mixotrophic basis of Atlantic oligotrophic ecosystems. Proc. Natl. Acad. Sci. USA 109: 5756–5760. doi:10.1073/pnas.1118179109

Hansen, P.J. et al. (2019). Mixotrophy Among Freshwater and Marine Protists. In book: Encyclopedia of Microbiology, 4 edition. Elsevier. 10.1016/B978-0-12-809633-8.20685-7

Ibarbalz, F.M., et al. (2019). Global Trends in Marine Plankton Diversity across Kingdoms of Life. Cell 179, 1084–1097. doi:10.1016/j.cell.2019.10.008

Jimenez, V., Burns, J. A., Le Gall, F., Not, F. and Vaulot, D. (2021). No evidence of phago-mixotropy in Micromonas polaris (Mamiellophyceae), the dominant picophytoplankton species in the arctic. J. Phycol. 57: 435–446. doi:10.1111/jpy.13125-20-132

Jones, H.L.J. (1997). A classification of mixotrophic protists based on their behavior. Freshw. Biol. 37 (1): 35-43. doi:10.1046/j.1365-2427.1997.00138.x

Karlusich, J. J. P., et al. (2022). A robust approach to estimate relative phytoplankton cell abundances from metagenomes. Mol. Ecol. Resour. 23: 16–40. doi:10.1111/1755–0998.13592

Lambert, B. S., et al. (2022). The dynamic trophic architecture of open–ocean protist communities revealed through machine–guided metatranscriptomics. Proceedings of the National Academy of Sciences 119(7): e2100916119.

Leles, S. G., Bruggeman J., Polimene L., Blackford J., Flynn K. J. and Mitra A. (2021). Differences in physiology explain succession of mixoplankton functional types and affect carbon fluxes in temperate seas. Prog. Oceanogr. 190. doi:10.1016/j.pocean.2020.102481

LeNoac'h, P., et al. (2024) Nutrient availability is the main driver of nanophytoplankton phago-mixotrophy in North American lake surface waters. J. Plankton Res. 00, 1–16. doi:10.1093/plankt/fbad054

Lepori-Bui, M., Paight, C., Eberhard, E., Mertz, C.M. and Moeller, H.V. (2022). Evidence for evolutionary adaptation of mixotrophic nanoflagellates to warmer temperatures. Glob. Chang. Biol. 28: 7094-7107.

Li, Q., Edwards, K. F., Schvarcz, C. R., Selph, K. E. and Steward, G. F. (2021). Plasticity in the grazing ecophysiology of Florenciella (Dictyochophyceae), a mixotrophic nanoflagellate that consumes Prochlorococcus and other bacteria. Limnol. Oceanogr. 66: 47–60. doi:10.1002/lno.11585

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

Li, Q., Dong, K.Y., Wang, Y. and Edwards, K. F. (2023). Relative importance of bacterivorous mixotrophs in an estuary-coast environment. Limnol. Oceanogr. Lett. 9(1): 81–91. doi:10.1002/lol2.10362

Martin, J.L., Santi, L., Pitta, P., John, U. and Gypens, N. (2022). Towards quantitative metabarcoding of eukaryotic plankton: an approach to improve 18S rRNA gene copy number bias. Metabarcoding and Metagenomics. doi: 10.3897/mbmg.6.85794

McKie–Krisberg, Z. and Sanders, R.W. (2014). Phagotrophy by the picoeukaryotic green alga Micromonas: implications for Arctic Oceans. The ISME Journal. 8: 1953–1961.

Millette, N. C., Pierson, J. J., Aceves, A. and Stoecker, D. K. (2017). Mixotrophy in Heterocapsa rotundata: A mechanism for dominating the winter phytoplankton. Limnol. Oceanogr. 62: 836–845. doi:10.1002/lno.10470

Millette, N. C., et al. (2023). Mixoplankton and mixotrophy: future research priorities. J.Plankton Res. 45 (4): 576–596. doi:10.1093/plankt/fbad020

Mitra, A., et al. (2014). The role of mixotrophic protists in the biological carbon pump. Biogeosciences 11:995–1005.

Mitra, A., et al. (2016). Defining Planktonic Protist Functional Groups on Mechanisms for Energy and Nutrient Acquisition: Incorporation of Diverse Mixotrophic Strategies. Protist 167(2): 106–120.

Mitra, A., et al. (2023). The Mixoplankton Database (MDB): Diversity of photo-phago-trophic plankton in form, function, and distribution across the global ocean. J. Eukaryot. Microbiol. 70: e12972. doi:10.1111/jeu.12972

Pernthaler, J. (2005). Predation on prokaryotes in the water column and its ecological implications. Nat. Rev. Microbiol. 3: 537–546. doi:10.1038/nrmicro1180

Pesant, S., et al. (2015). Open science resources for the discovery and analysis of Tara Oceans data. Scientific Data 2(1): 150023.

Picheral, M, et al. (2017). Tara Oceans Consortium, Coordinators; Tara Oceans Expedition, Participants (2014): Vertical profiles of environmental parameters measured from physical, optical and imaging sensors during station TARA_181 of the Tara Oceans expedition 2009-2013. PANGAEA, https://doi.org/10.1594/PANGAEA.836578

Ptacnik, R., et al. (2016). A light–induced shortcut in the planktonic microbial loop. Sci. Rep. 6(1): 29286.

Raven, J. A. (1997). Phagotrophy in phototrophs. Limnology and Oceanography 42: 198–205. doi:10.4319/lo.1997.42.1.0198

Rothhaupt, K. O. (1996). Laboratorary Experiments with a Mixotrophic Chrysophyte and Obligately Phagotrophic and Photographic Competitors. Ecology 77 (3). doi:10.2307/2265496

Sherr, E. B. and B. F. Sherr. (2002). Significance of predation by protists in aquatic microbial food webs. Antonie van Leeuwenhoek 81: 293–308. doi:10.1023/A:1020591307260

Stoecker, D. K. (1998). Conceptual models of mixotrophy in planktonic protists and some ecological and evolutionary implications. European Journal of Protistology 34: 281–290.

Stoecker, D. K., Hansen, P. J., Caron, D. A. and Mitra, A. (2017). Mixotrophy in the marine plankton. Ann Rev Mar Sci. 9: 311–335.

Stukel, M. R., Landry, M. R. and Selph. K. E. (2011). Nanoplankton mixotrophy in the eastern equatorial Pacific. Deep-Sea Res. II: Top. Stud. Oceanogr. 58: 378–386. doi:10.1016/j.dsr2.2010.08.016

Thingstad, T.F., Havskum, H., Garde, K., Riemann, B. (1996). On the strategy of "eating your competitor": A mathematical analysis of algal mixotrophy. Ecology 77, 2108–2118.

Tittel, J., et al. (2003). Mixotrophs combine resource use to outcompete specialists: Implications for aquatic food webs. Proc. Natl. Acad. Sci. USA 100(22): 12776–12781.

Ward, B. A. (2019). Mixotroph ecology: More than the sum of its parts. Proc. Natl. Acad. Sci. Comment. 116 (18): 5846–5848.

Wilken, S., et al. (2013). Mixotrophic organisms become more heterotrophic with rising temperature. Ecology Letters 16(2): 225–233.

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

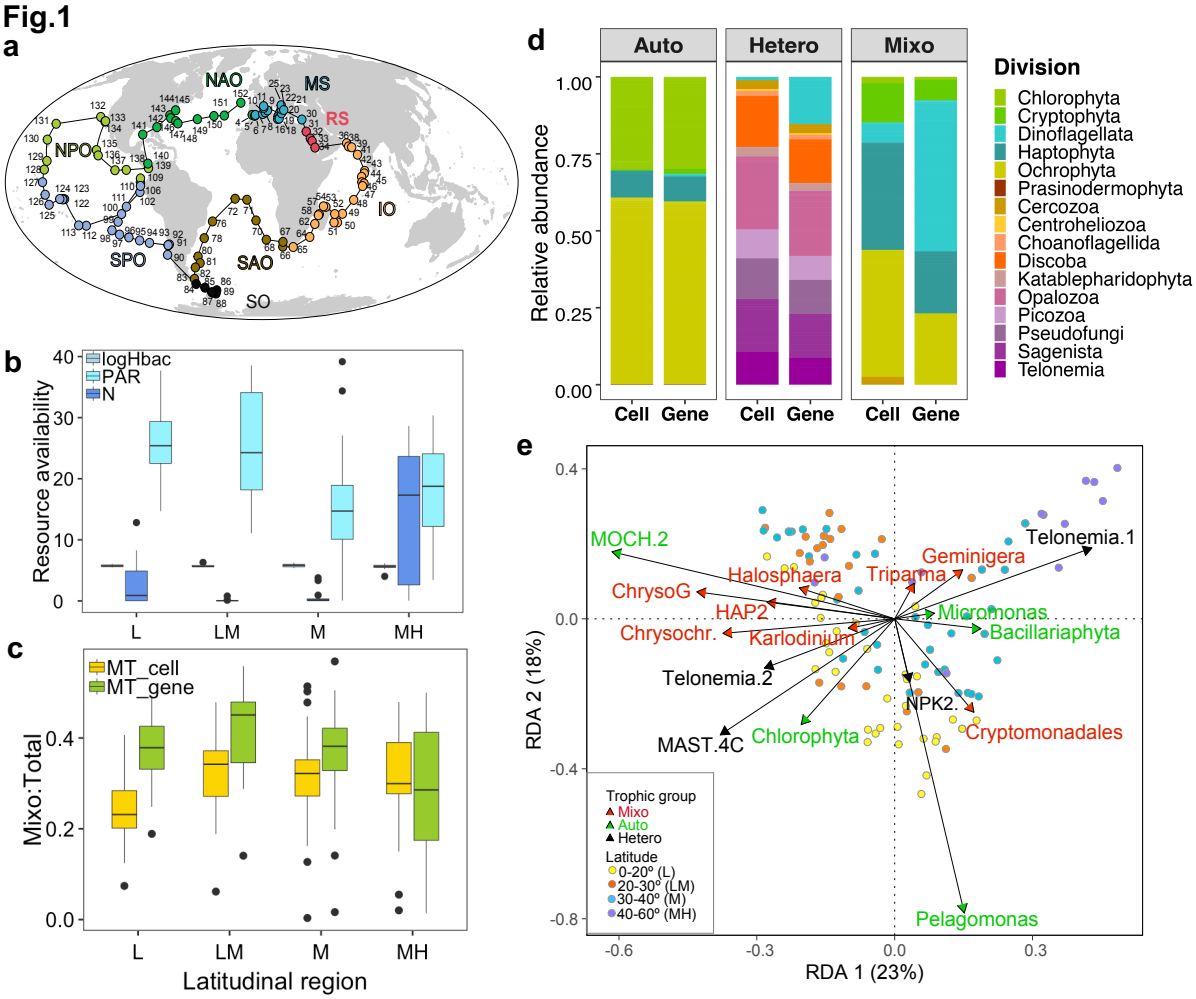
Zubkov, M. V. and Tarran G. A. (2008). High bacterivory by the smallest phytoplankton in the North Atlantic Ocean. Nature 455: 224–226.

Figure legends

Fig. 1. Map of studied sites (**a**), median resource availability of nutrients (NO3+NO2; N for abbreviation), light (PAR) and heterotrophic bacteria (**b**) and TI_M (Mixo:Total) (**c**) across different latitudinal regions. Panel **d** presents composition of autotrophic (Auto), mixotrophic (Mixo) and Heterotrophic (Hetero) protists on phyla level derived from 18S rRNA gene and corrected cell abundances. Panel **e** shows redundancy analysis between environmental variables and distribution of trophic groups on species level. Selected RDA variables and

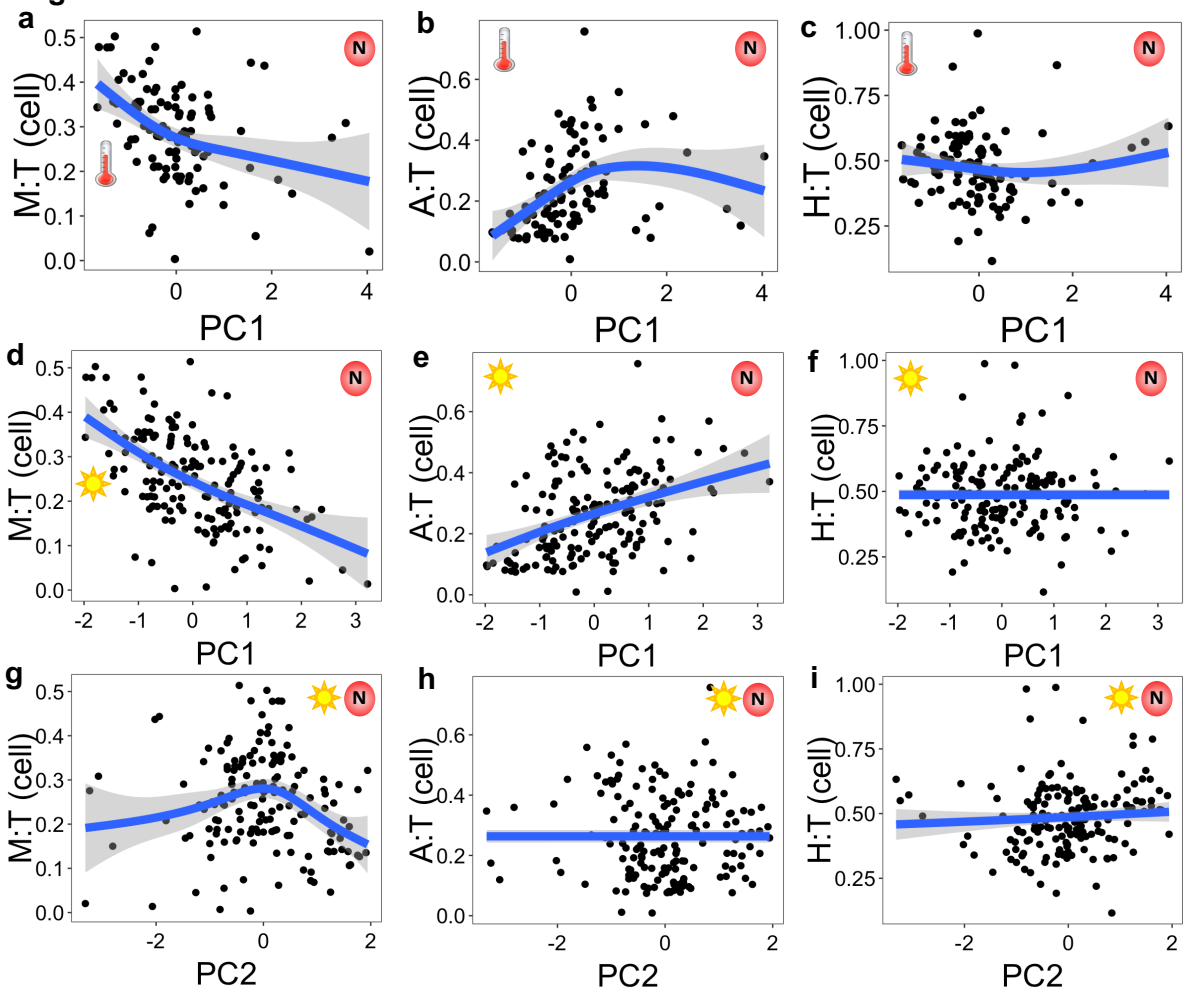
670 *Tara Oceans* lineage full names can be found in Supplementary Fig. S1. Latitudinal region
671 categories in b and c were shown in the legend of panel e.
672 **Fig. 2.** Variation patterns of $TI_{A/M/H}$ in cell abundances along PC1 axis for surface dataset
673 (**a-c**) and PC1 and PC2 axes for surface & DCM dataset (**d-i**). Cartoon symbols of nutrients
674 (N) of NO3+NO2, PAR and temperature marked on each panel denotes they are the main
675 components of each axis, and the location of each cartoon symbol indicates higher values.
676 Blue lines and shaded bands were fitted generalized additive model smoother and 95%
677 confidence interval.
678 **Fig. 3.** Variation patterns of $TI_{A/M/H}$ in 18S rRNA gene abundances along PC1 axis for
679 surface dataset (**a-c**) and PC1 and PC2 axes for surface & DCM dataset (**d-i**). Cartoon
680 symbols of nutrients (N) of NO3+NO2, PAR and temperature marked on each panel denotes
681 they are the main components of each axis, and the location of each cartoon symbol indicates
682 higher values. Blue lines and shaded bands were fitted generalized additive model smoother
683 and 95% confidence interval.
684 **Fig. 4.** Comparison of TI_g (**a**) and nutritional resource availability (**b**) between surface (SUR)
685 and chlorophyll maximum depths (DCM). Panel c and d demonstrate linear regressions
686 between TI_M and three resource variables of PAR, NO3+NO2 (N), and heterotrophic bacteria
687 (Hbac) with surface (**c**) and surface & DCM dataset (**d**), respectively. Variables were
688 sometimes Log-transformed for the best representation purpose. Significance codes were
689 shown in red stars (one star if $0.001<P<0.05$ and two stars if $P<0.001$), and shaded bands are
690 pointwise 95% confidence interval on the fitted linear regression (blue lines).
691 **Fig. 5.** Relative abundances of top 15 most abundant *Tara Oceans* mixotrophic lineages (cell
692 abundances), and closely related strains isolated from North Pacific Subtropical Gyre (**a**).
693 Comparison of relative abundances in surface (SUR) and maximal chlorophyll depths (DCM)
694 (**c**), and an illustrative flow chart of different treatments set up for the culture experiments (**c**).
695 **Fig. 6.** Growth curve of six mixotrophic isolates under high light (HL), low light (LL) and
696 high inorganic nitrogen (HN), low inorganic nitrogen (LN) conditions (**a**). Removal of
697 *Procholorococcus* prey in all cultures; green lines indicate prey control without grazers and
698 black lines were cultures with grazers (**b**). Panel (**c**) and (**d**) demonstrate estimated growth
699 rates during exponential phase and biovolume calibrated clearance rates, respectively. Errors
700 bars in c and d denote standard deviation among replicates for each treatment which were not
701 shown in panel a and b for more readability. *Grazer abundances in HNHL were divided by
702 50000 cells mL^{-1} for *Chrysochromulina* sp. and 5000 cells mL^{-1} for other sp., and 1000 cells
703 mL^{-1} for all the rest treatments. *Prey abundances in all HN were divided by 10^7 cells mL^{-1},
704 and all LN were divided by 10^6 cells mL^{-1}. Cartoon symbols in **c** and **d** indicate treatments
705 that showed significantly higher growth and grazing rates, respectively.
706

Fig.1


a
NAO
MS
RS
NPO
IO
SPO
SAO
SO
b
logHbac
PAR
N
Resource availability
L
LM
M
MH
c
MT_cell
MT_gene
Mixo:Total
Latitudinal region
d
Auto
Hetero
Mixo
Relative abundance
Cell
Gene
Division
Chlorophyta
Cryptophyta
Dinoflagellata
Haptophyta
Ochrophyta
Prasinodermophyta
Cercozoa
Centroheliozoa
Choanoflagellida
Discoba
Katablepharidophyta
Opalozoa
Picozoa
Pseudofungi
Sagenista
Telonemia
e
MOCH.2
Telonemia.1
Geminigera
Halosphaera
Triparma
ChrysoG
HAP2
Microhmonas
Chrysochr.
Karlodinium
Bacillariaphyta
Telonemia.2
NPK2.
MAST.4C
Chlorophyta
Cryptomonadales
Pelagomonas
Trophic group
Mixo
Auto
Hetero
Latitude
0-20° (L)
20-30° (LM)
30-40° (M)
40-60° (MH)
RDA 2 (18%)
RDA 1 (23%)

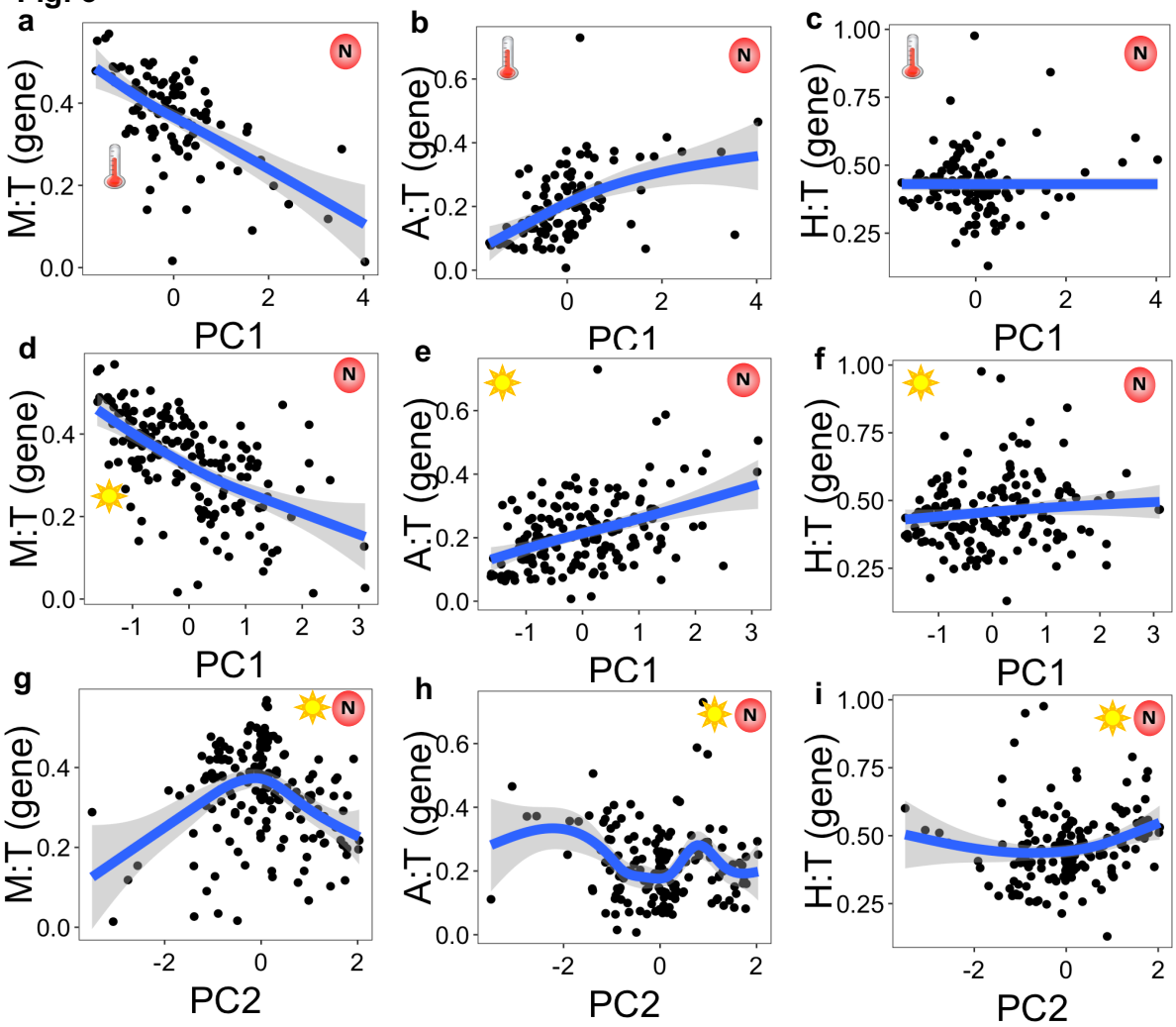
707

Fig. 2



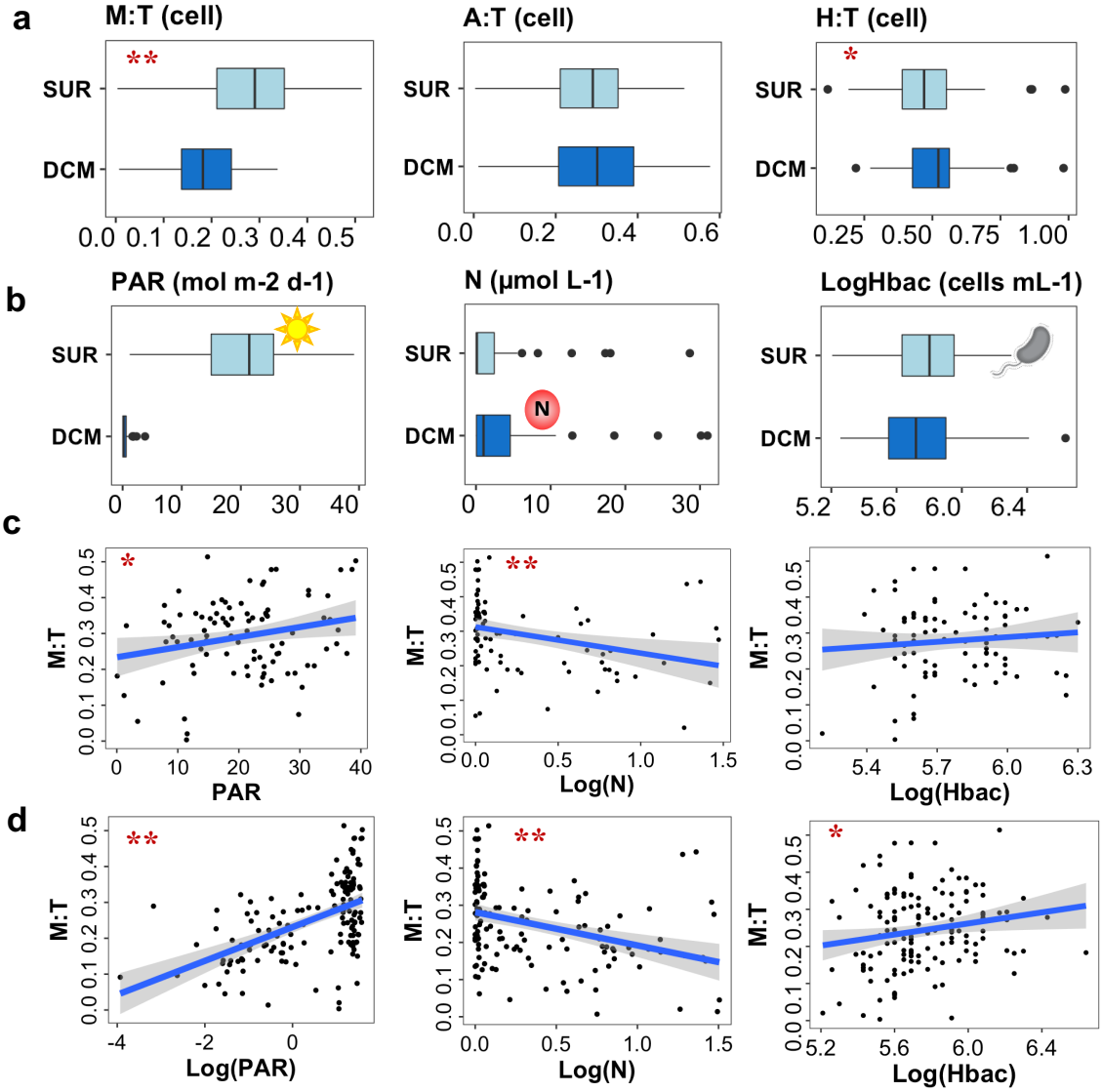
708

Fig. 3



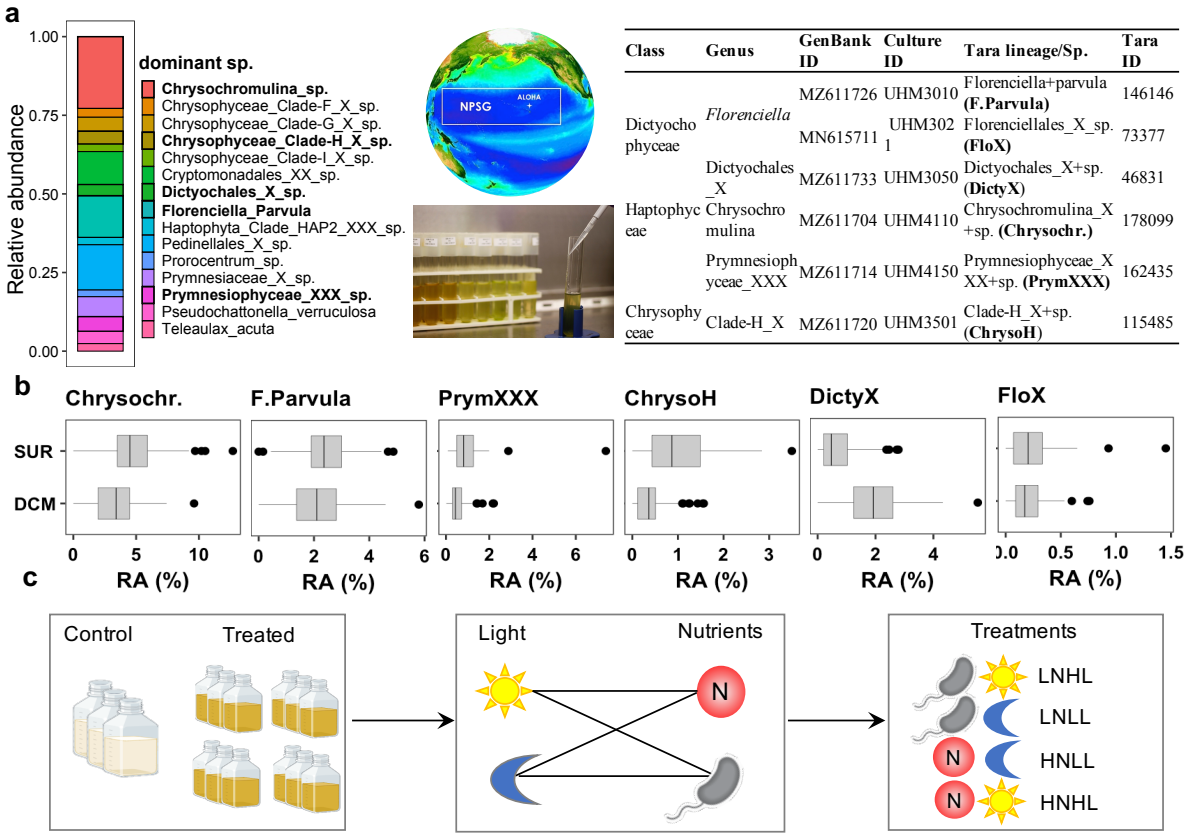
709
710

Fig. 4


a
M:T (cell)
A:T (cell)
H:T (cell)
SUR
DCM
**
*
0.0 0.1 0.2 0.3 0.4 0.5
0.0 0.2 0.4 0.6
0.25 0.50 0.75 1.00
b
PAR (mol m-2 d-1)
N (µmol L-1)
LogHbac (cells mL-1)
N
0 10 20 30 40
0 10 20 30
5.2 5.6 6.0 6.4
c
M:T
PAR
Log(N)
Log(Hbac)
0.0 0.5 1.0 1.5
5.4 5.7 6.0 6.3
d
Log(PAR)
-4 -2 0

711

Fig. 5

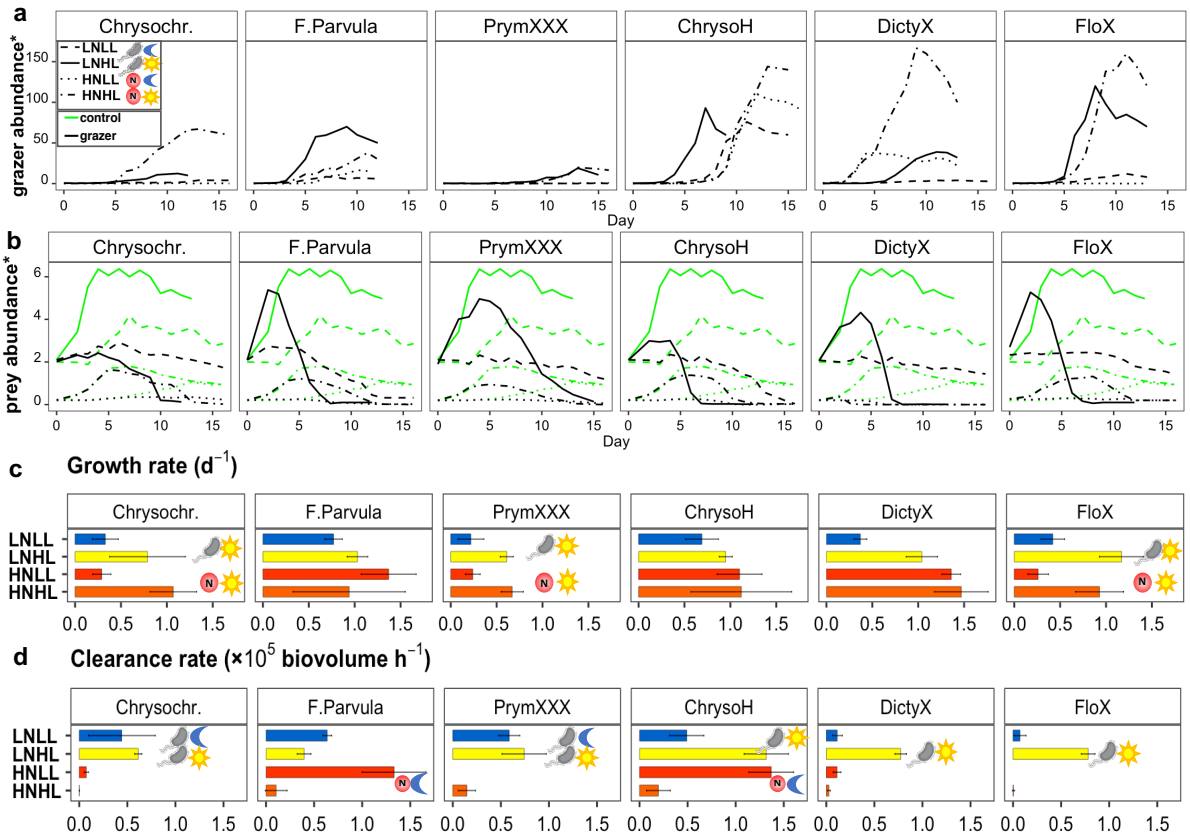


Class	Genus	GenBank ID	Culture ID	Tara lineage/Sp.	Tara ID
Dictyochophyceae	*Florenciella*	MZ611726	UHM3010	Florenciella+parvula **(F.Parvula)**	146146
		MN615711	UHM3021	Florenciellales_X_sp. **(FloX)**	73377
	Dictyochales_X	MZ611733	UHM3050	Dictyochales_X+sp. **(DictyX)**	46831
Haptophyceae	Chrysochromulina	MZ611704	UHM4110	Chrysochromulina_X+sp. **(Chrysochr.)**	178099
	Prymnesiophyceae_XXX	MZ611714	UHM4150	Prymnesiophyceae_XXX+sp. **(PrymXXX)**	162435
Chrysophyceae	Clade-H_X	MZ611720	UHM3501	Clade-H_X+sp. **(ChrysoH)**	115485

712

713

Fig. 6


a
Chrysochr.
F.Parvula
PrymXXX
ChrysoH
DictyX
FloX
grazer abundance*
LNLL
LNHL
HNLL
HNHL
control
grazer
Day
b
prey abundance*
Day
c
Growth rate (d^{-1})
d
Clearance rate (×10^5 biovolume h^{-1})

714

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

Taxa	Group	Reference	Method	Biovolume (µm3)	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

717

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

Surface & DCM			**Surface**		
Predictor[†]	**P**[‡]	**Coefficient**	**Predictor**[†]	**P**[‡]	**Coefficient**
PAR	**	0.1	DCM	***	0.15
NO3+NO2	***	-0.08	*Prochlorococcus*	**	-0.82
Density	***	0.80	NO3+NO2	***	-0.15
PIC	**	207.5	PIC	*	218.3
DCM	***	0.10	PO4	*	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

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

Supplementary figures

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

Fig. S2. Linear regression between $TI_{A/H}$ (all in cell abundances) and resource variables retrieved from the surface (less sample dots) and surface & DCM dataset (more sample dots). Different significance levels were marked with different number of 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 values (the line). Abbreviations: PAR, photosynthetically active radiations; N: inorganic nutrients of nitrate (NO3) and nitrite (NO2); Hbac: heterotrophic bacteria.

Fig. S3. Linear regression between TI_g (all in cell abundances) and single environmental variables that have shown significant correlations, retrieved from both surface & DCM samples. Different significance levels were marked with different number of 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 values (the line). Abbreviations: DCM, depth of chlorophyll a max; Chla, total chlorophyll a, Pro, *Prochlorococcus*; sal, salinity.

Fig. S1

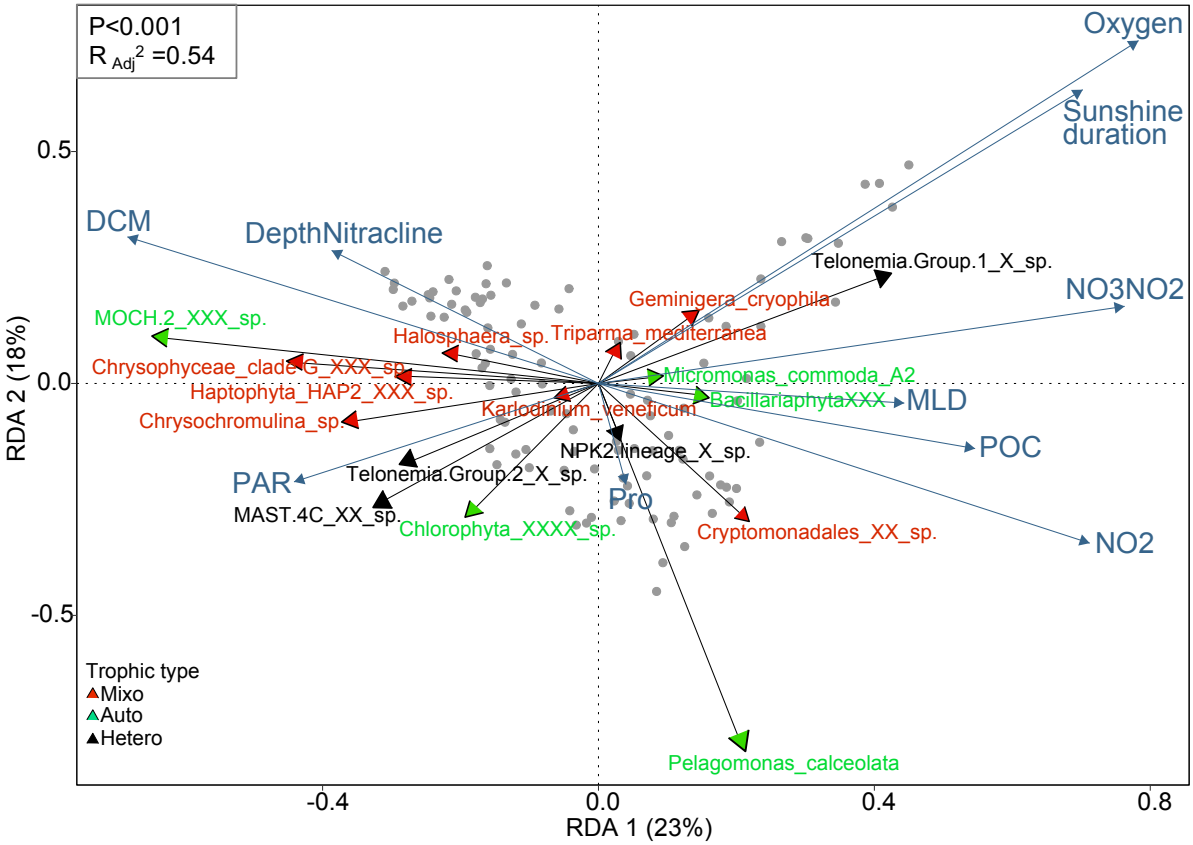
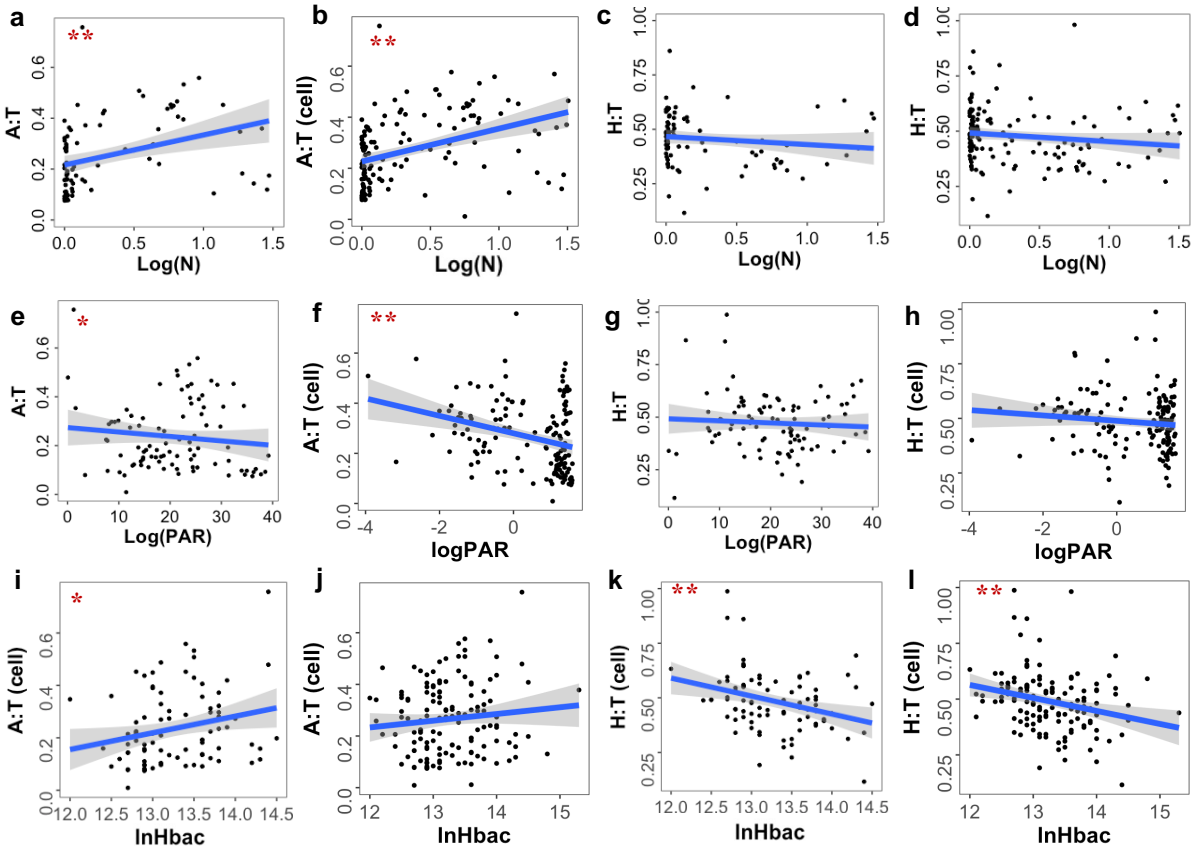
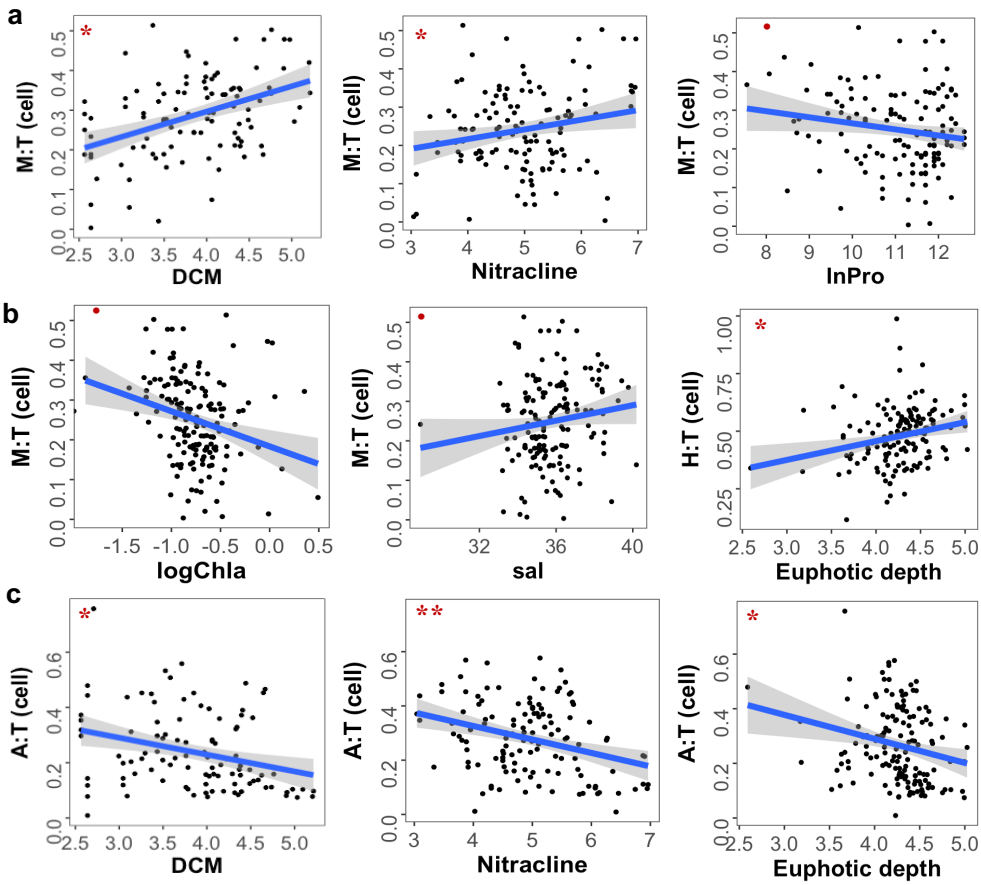

P<0.001
R Adj2 =0.54
Oxygen
Sunshine duration
DCM
DepthNitracline
Telonemia.Group.1_X_sp.
NO3NO2
Geminigera_cryophila
MOCH.2_XXX_sp.
Halosphaera_sp.
Triparma_mediterranea
Chrysophyceae_clade_G_XXX_sp.
Micromonas_commoda_A2
Haptophyta_HAP2_XXX_sp.
Karlodinium_veneficum
BacillariaphytaXXX
MLD
Chrysochromulina_sp.
NPK2.lineage_X_sp.
POC
PAR
Telonemia.Group.2_X_sp.
Pro
MAST.4C_XX_sp.
Chlorophyta_XXXX_sp.
Cryptomonadales_XX_sp.
NO2
Trophic type
Mixo
Auto
Hetero
Pelagomonas_calceolata
RDA 2 (18%)
RDA 1 (23%)
0.5
0.0
-0.5
-0.4
0.0
0.4
0.8

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
**
*

750

Fig. S3


a
M:T (cell)
DCM
Nitracline
lnPro
b
logChla
sal
H:T (cell)
Euphotic depth
c
A:T (cell)
DCM
Nitracline
Euphotic depth

751

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