



## Article

# Prevalence and Preferred Niche of Small Eukaryotes with Mixotrophic Potentials in the Global Ocean

Kaiyi Dong <sup>1</sup>, Ying Wang <sup>1</sup>, Wenjing Zhang <sup>1</sup> and Qian Li <sup>1,2,\*</sup><sup>1</sup> School of Oceanography, Shanghai Jiao Tong University, 1954 Huashan Road, Shanghai 200030, China<sup>2</sup> 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

**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 the overall prevalence of eukaryotic mixotrophs in the vast open ocean is, 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 a 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 were negatively correlated with nutrient concentrations and positively associated with light availability. Concentrations of heterotrophic bacteria as a single variable also presented a positive correlation with mixotrophic prevalence, but to a lesser extent. On the other hand, 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 an efficient strategy to compensate for nutrient deficiency but unnecessary to compensate for light scarcity. 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.

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



**Citation:** Dong, K.; Wang, Y.; Zhang, W.; Li, Q. Prevalence and Preferred Niche of Small Eukaryotes with Mixotrophic Potentials in the Global Ocean. *Microorganisms* **2024**, *12*, 750. <https://doi.org/10.3390/microorganisms12040750>

Academic Editors: Darrell Jay Grimes and Shuo Shen

Received: 18 February 2024

Revised: 23 March 2024

Accepted: 24 March 2024

Published: 8 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. 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 (which only photosynthesize), while those without inherited chloroplasts were regarded as strict heterotrophs (which only phagocytose). It has become increasingly evident that many of those autotrophs/heterotrophs are in fact mixotrophs, phagotrophic mixotrophs particularly, as they are capable of both photosynthesis and phagocytosis [1–3]. 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-sized flagellated eukaryotes that primarily feed 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 their distinct metabolic preferences and trophic activities, those two mixotrophic groups should be studied separately.

The relative abundance and prevalence of mixotrophs (compared 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 [8–11]. Mixotrophs are most likely to benefit from scenarios where nutrition is not sufficient to support a single trophic strategy, i.e., strict autotrophy or heterotrophy. For example, Ward (2019) 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) demonstrated distinct grazing rates and biovolume conversion rates among different mixotrophic strains [5], suggesting varying capabilities of phagotrophy and gross growth efficiencies.

In regional areas of temperate sea [14–16] and oligotrophic ocean [11,17,18] or lakes [19–21], field observations have also shown that mixotrophy can be advantageous for phytoplankton community growth, when nutrition is not ideal for autotrophy, which is mostly due to limited nutrients and/or irradiance. In other words, conditions that are unfavorable for strict autotrophy or heterotrophy could possibly select for mixotrophy on an entire community level. However, for larger spatial scales that encompass multiple ocean regions, there have been few studies looking 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., they find it easier to adapt to changing environments [22–25]. Motivated by these findings, we are particularly interested in delineating 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 was to gain insights into the interrelationship between environmental conditions and trophic strategies on both a community and a species level, as well as the underlying mechanism behind it. We aimed to answer the following scientific questions: (1) What is 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 roles do nutrients, light, and bacterial prey play in determining the relative success and fitness of mixotrophy? We used the eukaryotic 18S rRNA gene dataset from the *Tara* Oceans Expedition to assign trophic modes to each lineage/species and to assess the relationship between trophic mode composition and contextual variables. Subsequently, we ran 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.

## 2. Materials and Methods

### 2.1. Study Regions

The data generated by the *Tara* Oceans expedition between 2009 and 2013 are particularly suited to studying 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 had both SUR and CML samples) and 165 samples (all 104 SUR samples and 61 CML samples), was generated for a different statistical analysis. These samples were collected during four different seasons (spring, summer, autumn, and 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 ( $64.4^{\circ}$  S– $43.7^{\circ}$  N) and longitudinal ranges ( $159.0^{\circ}$  W– $73.9^{\circ}$  E). Bathymetry depths varied from 13 to 5964 km and distance to coast was between 0.9 and 1404 km.

### 2.2. 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) [26–31]. 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 the aforementioned literature). Second, these 100 variables were manually filtered down to 77 by eliminating similar variables and less relevant ones, to reduce collinearity and the inflation of dimensionality. Examples include the fact that salinity was chosen over conductivity, and factors such as moon phase durations and current residual time, etc., were excluded. Lastly, when variables had 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 the attenuation coefficient and surface irradiance observed by Moderate Resolution Imaging Spectroradiometer (MODIS), 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 measurements 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 ( $K_d$  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 ( $\text{NO}_3^-$  &  $\text{NO}_2^-$ ,  $\text{NO}_2^-$ ,  $\text{PO}_4^{3-}$ ,  $\text{SiO}_4^-$ ), pH, total alkalinity, total carbon, fluorescence, particulate organic carbon (POC), particulate inorganic carbon (PIC), and bacterial abundances, including autotrophic (*Prochlorococcus* and *Synechococcus*) and heterotrophic bacteria. Note that longitudes, latitudes, distance to coasts, sampling depths, and bathymetry depths were not

included in the analysis, in order to better focus on the aforementioned climatological, biochemical, and physical variables.

### 2.3. 18S rRNA Gene Abundances and Corrected Cell Abundances

Detailed genomic sample processing methods were described in de Vargas et al. (2015) [32]. Briefly, community genomic DNA was extracted from cells filtered through membrane filters with different pore sizes (0.8  $\mu\text{m}$ , 5  $\mu\text{m}$ , and 20  $\mu\text{m}$ ) after pre-filtration with nylon sieves (to retrieve different plankton communities). In this study, we focused on the 0.8–5  $\mu\text{m}$  size range, which constituted the majority of eukaryotic communities in the sampled regions [32]. The eukaryotic V9-18S rRNA gene region was sequenced through Illumina HiSeq, assembled, filtered, and classified with SWARM for operational taxonomic units (OTUs). The sequence files were downloaded from the European Nucleotide Archive under project number PRJEB6610. Subsequently, we re-assigned OTUs to the lowest taxonomic rank possible, referring to the Protist Ribosomal Reference (PR<sup>2</sup>) database [33]. Ciliophora, Radiolaria, and Foraminifera, which were likely derived from large-sized populations (contributing a small portion of the total species), were excluded from further analysis to focus on small-sized constitutive mixotrophs. Unknown Eukaryota and a small portion of parasites (e.g., Fungi, Amoebozoa, Syndiniales, and Dinoflagellata) were excluded from further trophic annotation, and, thus, further analysis. This final gene dataset generated 37,181,755 sequences (gene abundances) and 725 lineages/species for surface samples, and 17,874,753 sequences (gene abundances) and 802 lineages/species for the CML samples, respectively (Supplementary dataset S3).

Eukaryotic communities are known to exhibit substantially different 18S rRNA gene copy numbers, varying from as low as one copy, in species belonging to Ochrophyta, to thousands of copies per cell in dinoflagellates [5,34]. Overall, cellular biovolumes have been shown to be positively correlated with 18S rRNA gene copy numbers, but certain taxa, such as dinoflagellates, tended to have particularly high copy numbers [35,36]. To correct for this bias and obtain gene-converted cell abundances, we first collected empirical data from plankton, focusing on small cells with equivalent diameters of 5–10  $\mu\text{m}$  and biovolumes of 65–523  $\mu\text{m}^3$  (Supplementary Table S1). Then, we fit a linear correlation equation between empirical 18S rRNA gene copy numbers and cellular biovolume for individual group of dinoflagellates, diatoms, and other eukaryotes (including chlorophytes, haptophytes, dictyochophytes, etc.), respectively. Finally, a correction factor (C.F.) for each group was calculated using upper limit biovolume of 65  $\mu\text{m}^3$ ; that gave 59.3, 10.5, and 4.7 for dinoflagellates, diatoms, and other eukaryotes, respectively. For comparison, Martin et al. (2022) used 27.1, 4.4, and 0.9 C.F. for dinoflagellates, diatoms, and all other eukaryotes [36]. Gene abundances were converted into cell abundances by dividing gene numbers with the 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 purposes.

### 2.4. Assignment of Trophic Modes

In this study, we focused on the trophic strategy potentials among different eukaryotic lineages, rather than the in situ active trophic activities (which are partially associated with instantaneous conditions). Therefore, we assigned each lineage/species (when possible) into three trophic groups by searching published database and references [5,37,38], including (potential) mixotrophs (pigmented eukaryotes with phagocytosis capability), heterotrophs (eukaryotes with no inherent chloroplasts), and autotrophs (pigmented eukaryotes with no available evidence for phagocytosis). When trophic annotation was impossible at lineage or species level, their affiliated genus or family was referred to for annotation. The filtration generated a final composition of 184–186 mixotrophic, 283–313 heterotrophic, and 303–377 autotrophic lineages/species for the SUR and SUR&CML samples, respectively (Supplementary dataset S3). We acknowledge the limitation of this trophic annotation methodology and address it further in the Discussion section.

### 2.5. 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 three combined indices of  $TI_M$ ,  $TI_A$ , and  $TI_H$  represented the complete results for trophic mode compositions at a given site.

### 2.6. Statistical Analysis

The surface dataset was used for redundancy analysis (RDA) to elucidate the overall distribution pattern of trophic groups among 104 stations, and potential relationships with environmental variables. All 29 filtered contextual variables were 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 applied to species abundances and an Escoufier selection with a threshold of 0.9 was applied to the trophic species. Environmental variables that passed both a significance test and a collinearity test were presented on the RDA plot. The RDA was carried out in R using the *vegan* package, following the script outlined by Faure et al. (2019) [30]. Principal component analysis (PCA) was applied to extract the principal components (PC) that explained the environmental gradients from both surface (SUR) and SUR&CML datasets, using R package *pcaMethods* [39]. After extracting the main variable components,  $TI_g$  indices were plotted over two axes of PC1 and PC2 with a generalized additive model. Stepwise regressions between all filtered variables and  $TI_M$  were conducted to retrieve important variables explaining mixotrophic prevalence, and linear regressions between single variables and  $TI_g$  were carried out to provide additional correlation results. Analysis of variance (ANOVA) was conducted for the culture experiments to reveal factorial/treatment impacts on growth rates and grazing rates among various mixotrophic isolates.

### 2.7. Culture Experiments

Methods for isolating and cultivating mixotrophs from the North Pacific Subtropical Gyre were described in Li et al. (2021; 2022) [5,40]. Briefly, mixotrophic eukaryotes were isolated from the euphotic zone at Station ALOHA (22°45 'N, 158°00 'W), by enriching seawater samples in nitrogen-depleted Keller medium (K-N medium) that was supplemented with bacterial prey. Six identified isolates belonging to Dictyochophyceae, Prymnesiophyceae, and Chrysophyceae were selected for conducting growth experiments. These strains were phylogenetically close to the *Tara* Oceans lineages that were among the top 15 abundant lineages. 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  $\mu\text{M}$  photons  $\text{m}^{-2} \text{s}^{-1}$  irradiance on a light/dark cycle (12/12 h). All treatments were supplemented with autotrophic bacterial prey of *Prochlorococcus* (strain MIT9301) at final concentrations of  $\sim 2 \times 10^6$  cells  $\text{mL}^{-1}$ , which represents the second most abundant prokaryotic population at station ALOHA. Prey cells were centrifuged with a rotation speed of  $2000\text{--}3000 \times g$  for 3–5 min and gently resuspended for enrichment. Controls without added prey or grazers were also set up at each condition for comparison purposes. Cell concentrations of prey and grazers were measured every 12–24 h by flow cytometry, and ingestion evidences were occasionally obtained via microscopic observation.

Based on our previous study on grazing functional responses, most isolates have saturated ingestion rates when the prey concentration is higher than  $10^6$  cells  $\text{mL}^{-1}$ . Meanwhile, clearance rates (ingestion rates divided by prey concentration) can roughly present the maximal clearance rates with the final prey concentration used in these experiments. Therefore, we used clearance rates instead of ingestion rates [1,41] to represent grazing

performance. Ingestion rates (prey grazer<sup>-1</sup> h<sup>-1</sup>) for each grazer were calculated as the amount of prey removed between sampling time  $t + 1$  and  $t$ , divided by averaged grazer concentrations over the same time period. Clearance rates were calculated by dividing the ingestion rate by averaged prey concentration over the same interval. Body volume-specific clearance rates (body volume grazer<sup>-1</sup> h<sup>-1</sup>) were calculated by dividing the clearance rate by the cellular biovolume ( $\mu\text{m}^3$ ) of each grazer (measured under microscopy) [42]. Clearance rates reported in the results were averaged over all sampling intervals, prior to grazers reaching the stationary phase. The growth rates presented in the results were maximal values during exponential growth phase (slope of Ln transformed cell abundance versus time).

### 3. Results

#### 3.1. Oceanographic Characteristics across the Tara Oceans Sampling Regions

The 104 surface samples encompassed eight ocean habitats on a large spatial scale (Figure 1a). Geographically, they spanned roughly four latitudinal regions, from low (0–20°) to low–medium (20–30°), to medium (30–40°), and medium–high latitudes (40–60°). Moving from low to high latitudinal regions, overall PAR decreased from 25.4 to 14.7 mol photons m<sup>-2</sup> d<sup>-1</sup>, nutrients, e.g., NO<sub>3</sub><sup>-</sup> & NO<sub>2</sub><sup>-</sup>, increased from 0.03 to 17.3  $\mu\text{mol L}^{-1}$ , and heterotrophic bacterial abundances were relatively stable, ranging between 3.9 and 5.3 × 10<sup>5</sup> cells mL<sup>-1</sup> (Table 1; Figure 1b). The globally averaged means were 21.3 mol photons m<sup>-2</sup> d<sup>-1</sup> of PAR, 0.14  $\mu\text{mol L}^{-1}$  of NO<sub>3</sub><sup>-</sup> & NO<sub>2</sub><sup>-</sup>, and 4.5 × 10<sup>5</sup> cells mL<sup>-1</sup> of heterotrophic bacteria, demonstrating most sampling sites were oligotrophic environments. Smaller regional variations were also observed. The higher nutrient concentrations in low-latitude regions were possibly due to the Peruvian coastal and eastern equatorial upwelling, while lower PAR values in medium-latitude areas were because the Mediterranean Sea was sampled during winter.

**Table 1.** Oceanographic characters and trophic index of surface water samples. The abbreviations PAR, NO<sub>3</sub><sup>-</sup> & NO<sub>2</sub><sup>-</sup>, and Hbac stand for photosynthetically available radiation, dissolved nitrate and nitrite (NO<sub>3</sub><sup>-</sup> & NO<sub>2</sub><sup>-</sup>), and heterotrophic bacteria, respectively. M/T, M/A, and M/H denote ratios of either corrected cell abundance (Cell) or 18S rRNA gene abundance (Gene) among mixotrophs (M), autotrophs (A), heterotrophs (H), and total eukaryotes (T).

| Latitude | Oceanic Region <sup>†</sup>        | Station | Temp (°C) | Salinity | PAR (mol m <sup>-2</sup> d <sup>-1</sup> ) | NO <sub>3</sub> <sup>-</sup> & NO <sub>2</sub> <sup>-</sup> ( $\mu\text{mol L}^{-1}$ ) | Hbac (Cells mL <sup>-1</sup> ) | M/T (Cell/Gene) | A/T (Cell/Gene) | H/T (Cell/Gene) |
|----------|------------------------------------|---------|-----------|----------|--------------------------------------------|----------------------------------------------------------------------------------------|--------------------------------|-----------------|-----------------|-----------------|
| 0–20°    | SPO, IO, NPO, SAO, NAO             | 36      | 26.6      | 35.1     | 25.4                                       | 0.89                                                                                   | 4.9 × 10 <sup>5</sup>          | 0.23/<br>0.38   | 0.32/<br>0.24   | 0.43/<br>0.38   |
| 20–30°   | IO, NPO, SPO, SAO, RS              | 17      | 24.7      | 36.2     | 24.3                                       | 0.03                                                                                   | 3.9 × 10 <sup>5</sup>          | 0.34/<br>0.45   | 0.10/<br>0.08   | 0.51/<br>0.44   |
| 30–40°   | MS, SAO, NAO, SO                   | 37      | 19.5      | 36.4     | 14.7                                       | 0.13                                                                                   | 5.3 × 10 <sup>5</sup>          | 0.32/<br>0.38   | 0.19/<br>0.16   | 0.48/<br>0.43   |
| 40–60°   | SO, SAO, MS                        | 14      | 7.1       | 34.3     | 18.8                                       | 17.3                                                                                   | 4.5 × 10 <sup>5</sup>          | 0.30/<br>0.29   | 0.18/<br>0.22   | 0.47/<br>0.46   |
| Global   | SPO, IO, NPO, SAO, NAO, SO, RS, MS | 104     | 23.8      | 35.4     | 21.3                                       | 0.14                                                                                   | 4.5 × 10 <sup>5</sup>          | 0.29/<br>0.38   | 0.21/<br>0.19   | 0.47/<br>0.42   |

<sup>†</sup> Oceanic regions follow the Tara Oceans 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.

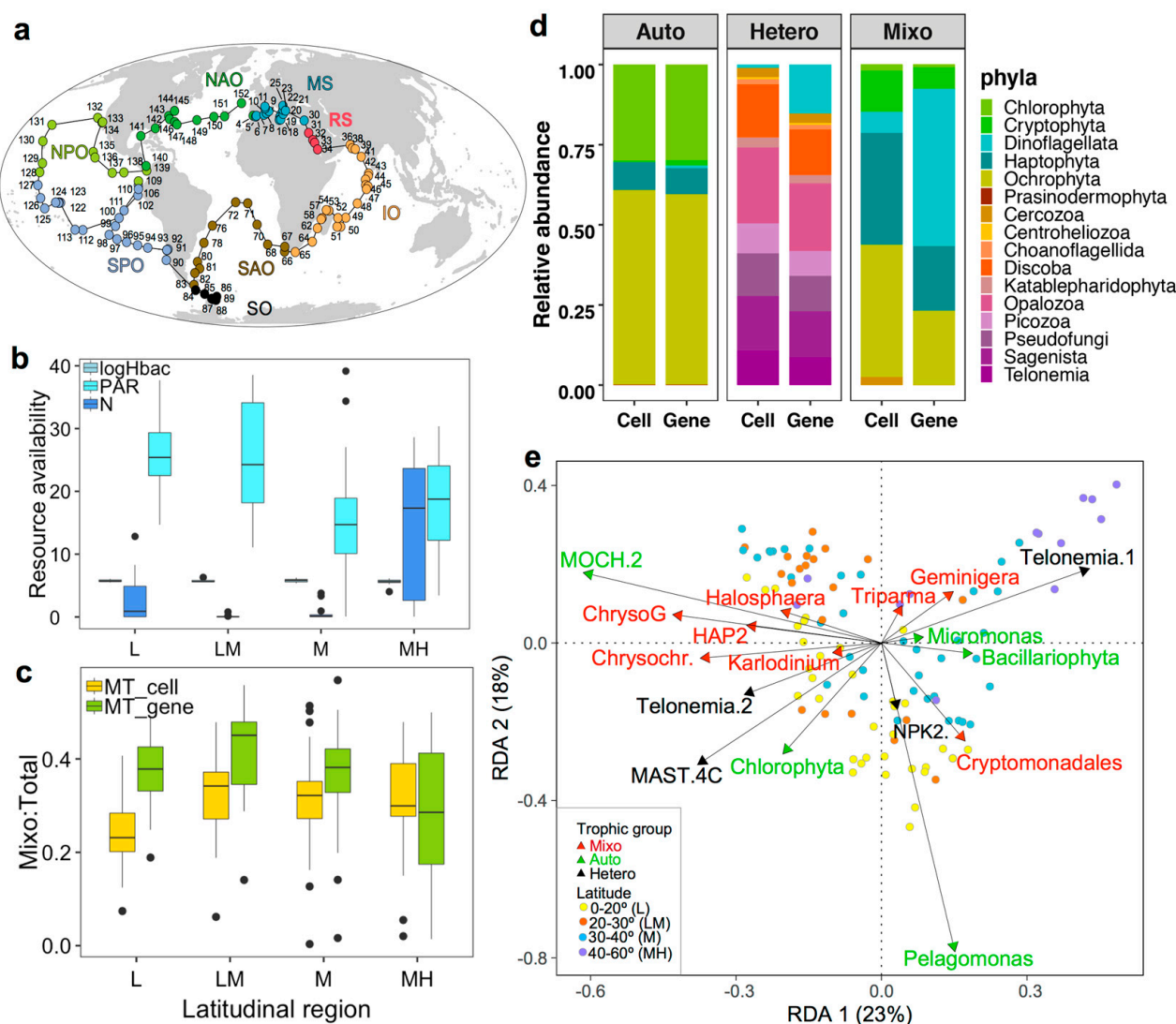


**Figure 1.** Map of study sites (a), averaged median values of inorganic nutrients ( $\text{NO}_3^-$  &  $\text{NO}_2^-$ ; N for abbreviation), light (PAR) and heterotrophic bacteria (Hbac) (b), and mixotrophic prevalence index  $\text{TI}_M$  (Mixo/Total) derived from 18S rRNA gene (MT\_gene) and corrected cell abundances (MT\_cell) (c) at different latitudinal regions. (d) presents the composition of autotrophic (Auto), mixotrophic (Mixo), and heterotrophic (Hetero) eukaryotes on a phyla level, derived from 18S rRNA gene and corrected cell abundances. (e) is a redundancy analysis plot showing the distribution of trophic groups on a species level. Oceanic region abbreviations in panel (a) are, 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. Latitudinal region names and definitions in (b,c) are shown in the legend of (e), and black dots were potential outliers from individual samples. *Tara* Oceans lineages in (e) were shown with genera name, and full names of species can be found in Supplementary Figure S1 and Supplementary dataset S3.

### 3.2. Relative Abundances and Taxonomic Compositions of Three Trophic Modes

The trophic index of  $\text{TI}_M$ ,  $\text{TI}_A$ , and  $\text{TI}_H$  varied from 0.23 to 0.34, from 0.10 to 0.32, and from 0.43 to 0.51 in cell abundance, and from 0.29 to 0.45, from 0.08 to 0.24, and from 0.38 to 0.46 in gene abundance, respectively (Table 1; Figure 1c).  $\text{TI}_M$  values in cell abundance were overall somewhat lower than for gene abundance. Nevertheless, the highest  $\text{TI}_M$  median values were both retrieved from low-medium latitudes, and the lowest values were from low latitudes (cell abundances) and medium-high latitudes (gene abundances), respectively. In comparison, high and low  $\text{TI}_A$  median values were from low and medium-

high latitudes and low–medium latitudes (both cell and gene abundances).  $TI_H$  did not change as much across latitudinal gradients, but the highest median values in cell and gene abundances were shown in low–medium and medium–high latitudes, respectively.

Before the gene and cell abundances correction, mixotrophs were mostly composed of lineages from Dinoflagellata, Haptophyta, and Ochrophyta. After the correction for cell abundances, the dominant mixotrophs were retrieved from Haptophyta and Ochrophyta (Figure 1d). Proportions of dinoflagellates decreased in heterotrophic communities after gene–cell conversion, and the other three most abundant heterotrophs belonged to the phyla Opalozoa, Discoba, and Sagenista, both in cell and gene abundances. Differences in autotrophic compositions on the phyla level in cell and gene abundances were insignificant, which were both dominated by Chlorophyta and Ochrophyta. Redundancy analysis suggested that stations from the same ocean or latitudinal regions tended to cluster together (Figure 1e; Supplementary Figure S1). Certain species were more abundant in 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 for autotrophs, and *MAST.4C\_XX\_sp.* and *Telonemia Goup2\_X\_sp.* for heterotrophs. Those taxa were positively affiliated with PAR and CML, and negatively correlated with nutrients ( $NO_3^-$  &  $NO_2^-$  and  $NO_2^-$ ) 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 latitudes and positively correlated with oxygen and sunshine duration. Additionally, *Micromonas A2* and *Bacillariophyta\_XXX\_sp.* for autotrophs were associated with high MLD, POC, and  $NO_2^-$ , while *Cryptomonadales\_XX\_sp.* for mixotrophs and *NPK2 lineage\_X\_sp.* for heterotrophs were affiliated with high abundances of *Prochlorococcus*.

### 3.3. Nutritional Variables and the Prevalence of Three Trophic Modes

Among the 29 selected predictors, the 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; Figures 2 and 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 mostly consisted of nutrients and temperature ( $R^2$  of  $0.35$ ), while the PC2 axis ( $R^2$  of  $0.24$ ) was mostly explained by MLD, CML, and PAR.

**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 ‘ $-$ ’ explain the changing variable values decrease from left to right on the axes.

|                     | SUR                       |                           | SUR&CML                   |                           | Sum <sup>†</sup> |
|---------------------|---------------------------|---------------------------|---------------------------|---------------------------|------------------|
|                     | PC1                       | PC2                       | PC1                       | PC2                       |                  |
| PAR                 | $-0.08$                   | <b><math>-0.14</math></b> | <b><math>-0.42</math></b> | <b><math>-0.35</math></b> | 0.99             |
| $NO_3^-$ & $NO_2^-$ | <b>0.31</b>               | <b><math>-0.14</math></b> | <b>0.23</b>               | <b><math>-0.25</math></b> | 0.93             |
| Silicate            | <b>0.23</b>               | $-0.10$                   | <b>0.17</b>               | <b><math>-0.17</math></b> | 0.67             |
| Temperature         | <b><math>-0.24</math></b> | 0.08                      | $-0.14$                   | 0.13                      | 0.59             |
| Nitracline depth    | $-0.20$                   | $-0.11$                   | $-0.15$                   | 0.12                      | 0.58             |
| CML                 | 0.06                      | <b><math>-0.17</math></b> | $-0.10$                   | 0.07                      | 0.40             |
| Mixed layer depth   | 0.06                      | <b><math>-0.22</math></b> | $-0.04$                   | $-0.06$                   | 0.38             |
| $R^2$               | 0.35                      | 0.24                      | 0.30                      | 0.25                      |                  |

<sup>†</sup> summed absolute values from PC1 and PC2 for both datasets.

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 the decreasing prevalence of mixotrophs with

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

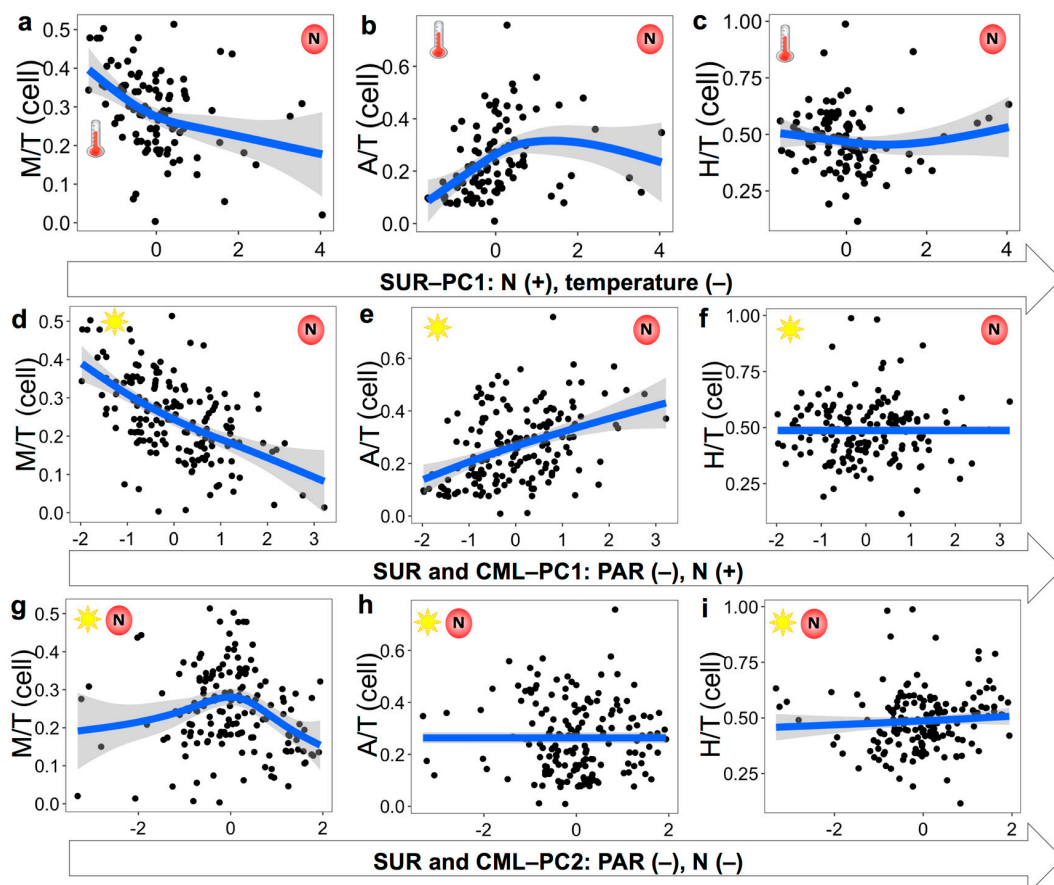


**Figure 2.** Variation patterns of  $TI_{A/M/H}$  in corrected cell abundances along PC1 axis for surface (SUR) dataset (a–c) and PC1 and PC2 axes for surface and chlorophyll maximal layer (SUR&CML) dataset (d–i). Cartoon symbols of nutrients ( $NO_3^-$  &  $NO_2^-$ ) (letter N with a red circle), PAR (the sun), and temperature (a thermometer) were marked on each panel, and the location of them indicates higher values ends. Each black dot represents observed value from two datasets. Blue lines and shaded bands were fitted to a generalized additive model smoother and with 95% confidence intervals.

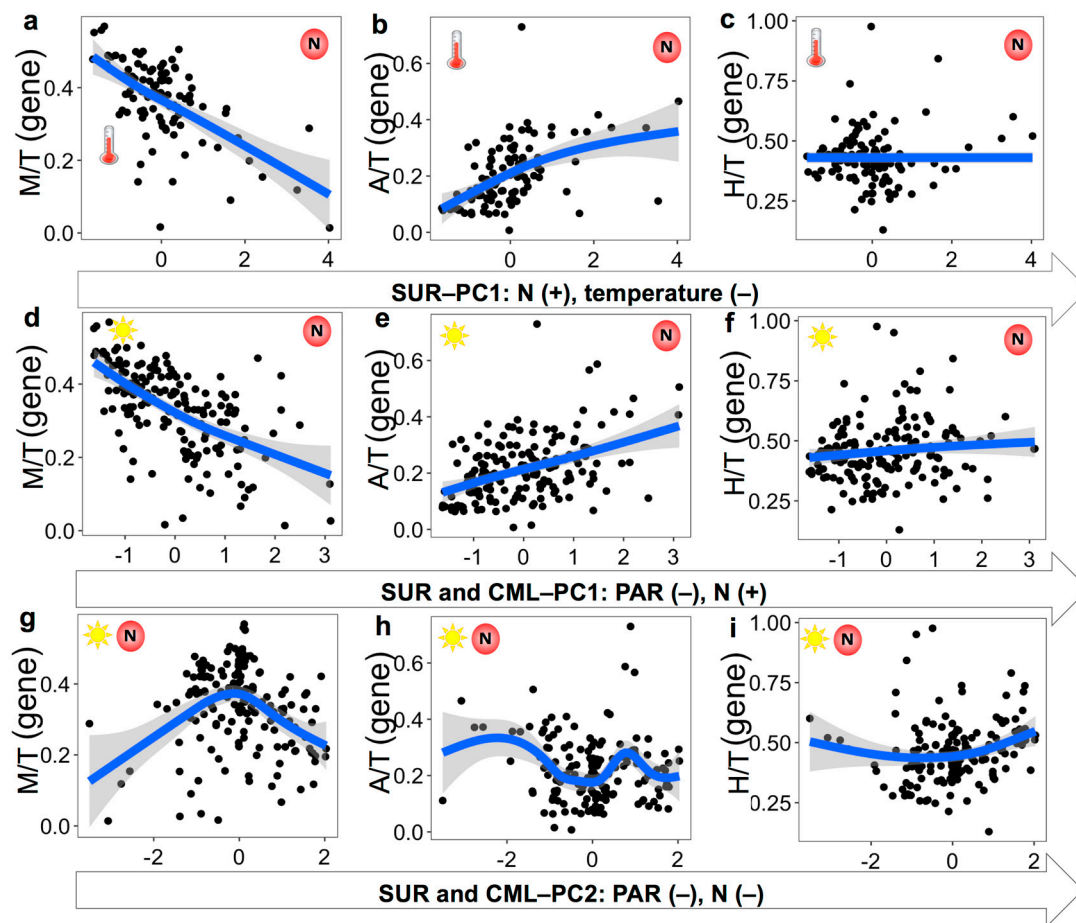


**Figure 3.** Variation patterns of  $TI_{A/M/H}$  in 18S rRNA gene abundances along the PC1 axis for the surface (SUR) dataset (a–c) and PC1 and PC2 axes for the surface and chlorophyll maximal layer (SUR&CML) dataset (d–i). Cartoon symbols of nutrients ( $NO_3^-$  &  $NO_2^-$ ) (a letter N with a red circle), PAR (the sun), and temperature (a thermometer) were marked on each panel, and the location of them indicates higher values ends. Each black dot represents observed value from two datasets. Blue lines and shaded bands were fitted to the generalized additive model smoother with 95% confidence intervals.

Consistent with the PCA results, clear differences were observed for  $TI_g$  between SUR and CML samples (from 61 stations, we have both SUR and CML samples), accompanied with distinct resource availability of nutrients, light, and bacterial prey (Figure 4). Significantly higher  $TI_M$  values were seen in surface samples where PAR and bacteria were higher and N ( $NO_3^-$  &  $NO_2^-$ ) was lower (median 0.29 vs. 0.18,  $p < 0.001$ ,  $df = 60$ ). In comparison,  $TI_A$  was slightly higher in CML compared to surface samples, but this was not statistically significant (median 0.30 vs. 0.29,  $p > 0.05$ ,  $df = 60$ ), and  $TI_H$  values were significantly higher in CML than surface samples (median 0.52 vs. 0.47,  $p < 0.05$ ,  $df = 60$ ) (Figure 4a,b). Linear regressions derived from all samples suggested a consistently positive relationship for  $TI_M$  with PAR and negative correlations with nutrients ( $NO_3^-$  &  $NO_2^-$ ), from both the SUR (122 samples; Figure 4c) and the SUR&CML datasets (165 samples; Figure 4d).  $TI_A$  demonstrated a significantly positive and negative correlation with nutrients and PAR, respectively, whereas neither was significantly associated with  $TI_H$  (Supplementary Figure S2a–h). For heterotrophic bacteria, a positive but weaker relationship ( $p = 0.02$ ) was found between  $TI_M$  and the SUR&CML dataset, but not with the SUR dataset (Figure 4c,d). In contrast, both datasets demonstrated strong negative associations between heterotrophic bacteria and  $TI_H$  values, and a weaker but positive correlation between heterotrophic bac-

teria and  $TI_A$  in the surface ( $p < 0.05$ ) (Supplementary Figure S2i–l). Other single variables showing significant regressions with  $TI_M$  were all shown in Supplementary Figure S3.

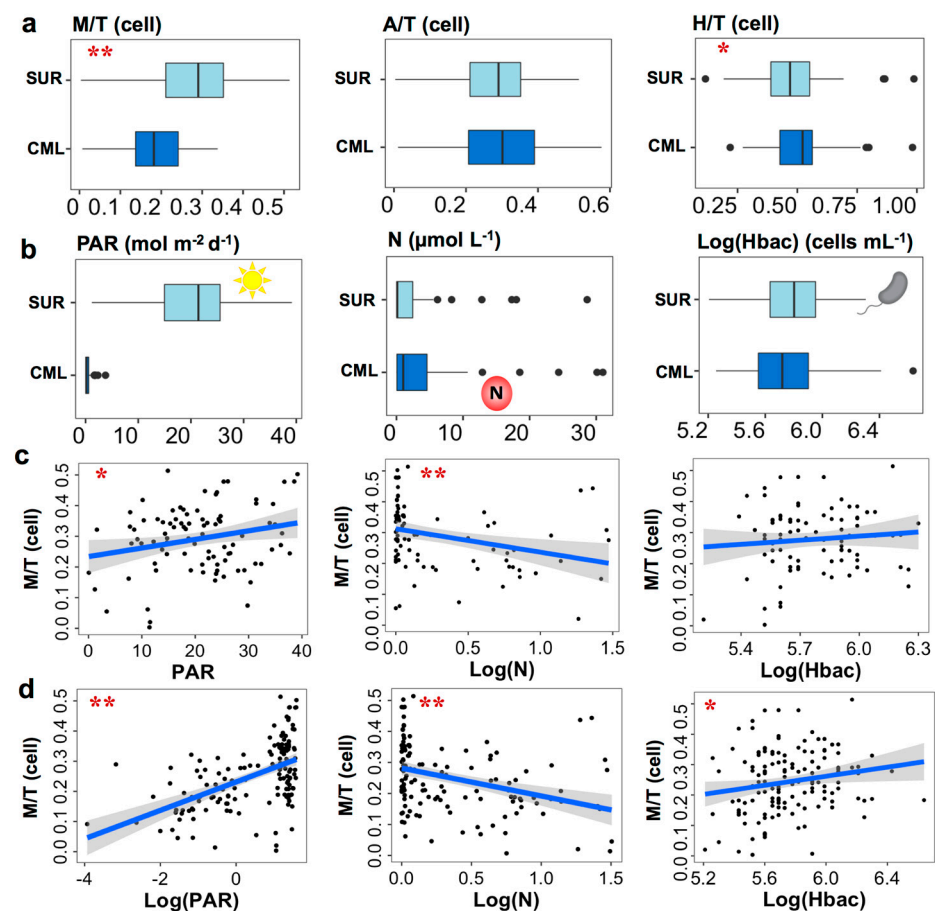


**Figure 4.** Comparison of  $TI_g$  (a) and nutritional resource availability (b) between surface (SUR) and chlorophyll maximum layer (CML). (c,d) demonstrate linear regressions between  $TI_M$  and three resource variables of PAR,  $NO_3^-$  &  $NO_2^-$  (N), and heterotrophic bacteria (Hbac) with the SUR dataset (c) and the SUR&CML dataset (d), respectively. Variables in (c,d) were sometimes Log-transformed for the best representation purpose, and each black dot represents observed value from two datasets. Significance codes were shown using red stars (one star if  $0.001 < p < 0.05$  and two stars if  $p < 0.001$ ), and shaded bands are pointwise 95% confidence intervals on the fitted linear regression (blue lines). The cartoon symbols of the sun, letter N with red circles, and bacterial cell shape represents PAR,  $NO_3^-$  &  $NO_2^-$ , and Hbac concentrations, respectively.

### 3.4. Mixotrophic Growth with Varying Nutritional Resource

The most abundant mixotrophs identified from the Tara Oceans dataset belonged to the classes of Prymnesiophyceae, Dictyochophyceae, Chrysophyceae, Dinophyceae, and Cryptophyceae (Figure 5a). Mixotrophic strains that we isolated from the North Pacific had been previously 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 belonged 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 Figure 5a, it ranked as the 26th most abundant lineage across all stations. The relative abundances of those isolates retrieved from surface samples were all higher than from CML, except for Dictyochales\_X (Figure 5b), presenting similar patterns as the entire mixotrophic community ( $TI_M$  results in Figure 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 the impacts of light and nutrient (inorganic and organic nutrients) availability on the growth of mixotrophs (Figure 5c).

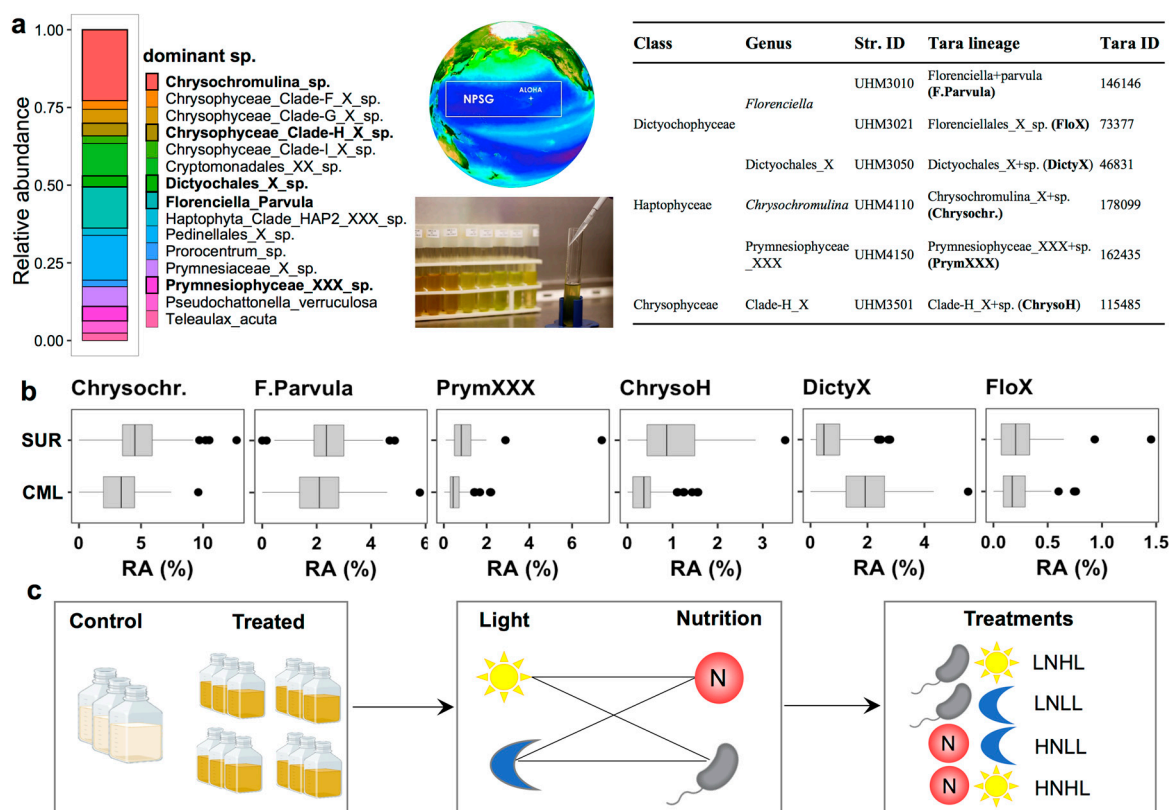


**Figure 5.** Relative abundances of the top 15 most abundant *Tara* Oceans mixotrophic lineages (in gene converted cell abundances), and phylogenetically related strains isolated from station ALOHA in the North Pacific Subtropical Gyre (a). Comparison of relative abundances in surface (SUR) and chlorophyll maximum layer (CML) (b), and an illustrative flow chart showing how different treatments were set up for the growth experiments (c). Cartoon symbols of the sun, the moon, bacterial shape and letter N with circles in (c) stand for treatment of high light, bacterial prey as the only primary nutrition (low inorganic nutrients), low light and high inorganic nitrogen levels, respectively. Images resources of two middle pictures in (a) are the Hawaiian Ocean Time Series crew and Dr. C.R. Schvarcz (was previously affiliated with the University of Hawaii).

Distinct growth curve shapes and final abundances were presented among different treatments for the six isolates (Figure 6a), accompanied by different disappearance rates of prey (Figure 6b). Prey in controls (without mixotrophic grazers) all presented slower disappearance rates compared to cultures with grazers, but to different levels among strains. Grazers in the low nutrient control (without prey) grew to much lower concentrations, i.e., all  $< 1 \times 10^3$  cells  $\text{mL}^{-1}$ , compared to  $10^4$ – $10^5$  cells  $\text{mL}^{-1}$  when prey was added. Shown as exponential growth rates (bars in Figure 6c), most isolates presented the highest values when grown under high amounts of light, either when receiving high nutrients or sufficient prey (low nutrients). Lower growth rates were observed with low-light treatments (marked with black arrows in Figure 6c) for strain 4110 from *Chrysochromulina*, 4150 from Prymnesiophyceae\_XXX, and 3021 from Florenciellales\_X, i.e., those categorized as trophic type 1. Growth of one isolate from *Florenciella* (strain 3010) and one isolate from Chrysophyceae Clade H\_X (strain 3501) did not show significant impacts by light and, thus, were classified as trophic type 2. One additional strain 3050 from Dictyochales\_X represented slightly different characters from both Type 1 and Type 2 and were thus categorized as Type 1/2. Specifically, they performed better under low light than Type 1

strains when receiving high nutrients, but not as good as Type 2 strains with low light and low nutrients. More experiments need to be done to verify its trophic traits. For grazing rates, most strains showed the highest clearance rates when receiving low levels of nutrients and high amounts of light (marked with blue arrows in Figure 6d). Only two strains belonging to Type 2 were capable of grazing much faster when receiving low light levels and high nutrient levels (marked with red arrows in Figure 6d).

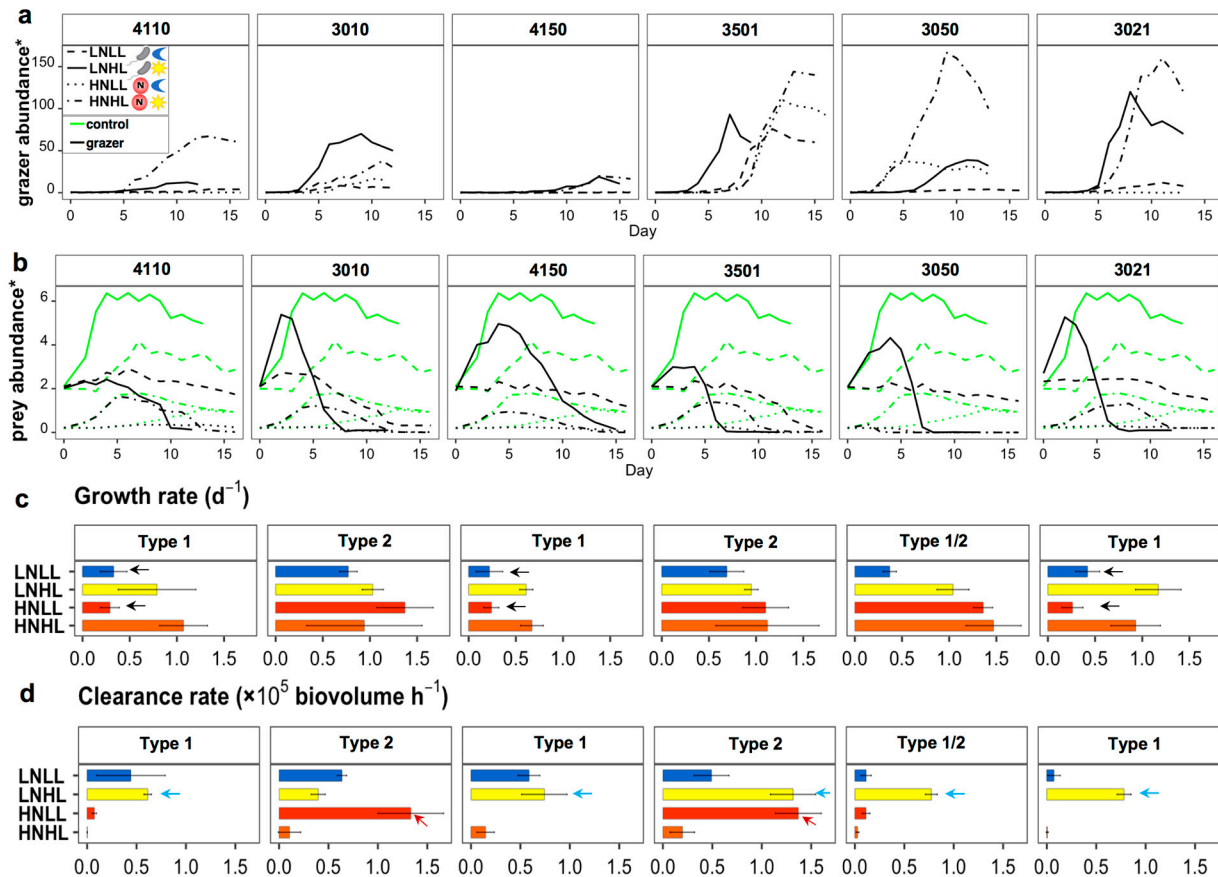


**Figure 6.** (a) presents growth curves of six mixotrophic isolates grown under high light (HL) and low light (LL), as well as high inorganic nitrogen (HN) and low inorganic nitrogen (LN) conditions. All treatments were supplemented with *Prochlorococcus* prey. (b) presents removal of prey in all cultures (ID table in Figure 5a); green lines indicate prey control without grazers and black lines were prey with grazers. (c,d) demonstrate estimated growth rates during exponential phase and biovolume clearance rates for each strain. Errors bars in (c,d) denote standard deviation among replicates for each treatment, which were not shown in (a,b) for higher readability. \* For grazer concentrations in HNHL, they were divided by 50,000 cells mL<sup>-1</sup> for strain 4110 and 5000 cells mL<sup>-1</sup> for other strains, and 1000 cells mL<sup>-1</sup> for all the other treatments. \* Prey abundances in all HN treatments were divided by 10<sup>7</sup> cells mL<sup>-1</sup>, and all LN were divided by 10<sup>6</sup> cells mL<sup>-1</sup>. Arrows in (c,d) indicate treatments that caused significantly lower growth rates (c), and significantly higher grazing rates (d).

The ANOVA results revealed a significant effect on growth by light, i.e., to increase growth rates with increasing light, either alone, or together with nutrients and species. However, (inorganic) nutrient availability did not show significant impact on growth rates, as cultures supplemented with bacterial prey can sufficiently compensate for nutrient deficiency. Less significant effects were observed for clearance rates, as strains responded differently to changing light and nutrient conditions, indicated by the high  $p$  ( $>0.05$ ) values from both single and multi-factorial ANOVAs (Table 3).

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

|                     | Growth Rates  |       |       |               |                   | Clearance Rates |       |       |               |                   |
|---------------------|---------------|-------|-------|---------------|-------------------|-----------------|-------|-------|---------------|-------------------|
|                     | Single Factor |       |       | Multi-Factors |                   | Single Factor   |       |       | Multi-Factors |                   |
|                     | PAR           | N     | Sp.   | PA and N      | PAR and N and Sp. | PAR             | N     | Sp.   | PAR and N     | PAR and N and Sp. |
| $p$ <sup>†</sup>    | *             | >0.05 | >0.05 | *             | *                 | >0.05           | >0.05 | >0.05 | *             | 0.13              |
| F <sup>‡</sup>      | 6.5           | 1.7   | 2.1   | 5.1           | 3.6               | 0.00            | 2.98  | 1.35  | 3.96          | 2.02              |
| Dfn <sup>§</sup>    | 1             | 1     | 5     | 3             | 5                 | 1               | 1     | 5     | 3             | 5                 |
| Dfd <sup>  </sup>   | 20            | 20    | 17    | 18            | 18                | 20              | 20    | 17    | 18            | 18                |
| Effect <sup>¶</sup> | +             | +     | /     | /             | /                 | +               | –     | /     | /             | /                 |

<sup>†</sup> Significance level of ANOVA, “\*” denotes  $p < 0.05$ . <sup>‡</sup> Variation between sample means/variation within the samples. <sup>§</sup> Degrees of freedom in the numerator (Dfn), equals to the groups (of factors). <sup>||</sup> Degrees of freedom in the denominator (Dfd), equals to the total number of subjects in the experiment—the groups (of factors). <sup>¶</sup> Symbols of + and – indicate positive and negative impacts from increasing light and nutrient availability, and / denotes “analysis unavailable”. Only factors with  $p$  values marked with \* in the first row had a significant impact.

#### 4. 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, and across large spatial scales, with a particular focus on three nutritional resources. We then conducted finely controlled lab experiments using isolates to investigate how mixotrophic growth responds 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 demonstrated a close relationship between mixotrophic prevalence and nutritional resource among inorganic nutrients, light, and prey, which possibly introduced distinct niche preferences for mixotrophs versus autotrophic and heterotrophic eukaryotes in the ocean.

##### 4.1. Mixotrophic Growth Responding to Changing Nutritional Resources

Suggested by our experimental results, mixotrophy can function as an efficient strategy for compensating for inorganic nutrient deficiency for most strains through the ingestion of bacterial prey, but not necessarily for supplementing light limitation (Figure 6c). Among the six studied cultures, two of them were efficient with compensating light/energy deficiency via heterotrophic nutrition (grazing), but not the others. These results indicated that mixotrophic controlling mechanisms for metabolic partitioning between autotrophy and heterotrophy were species-specific. The Type 1 isolates that were light-dependent possibly align with the ‘primarily photosynthesis and phagocytosis for nutrients’ functional group proposed by Jones (1997) and Stoecker (1998) [43,44]. The growth rates of Type 2 isolates seemed to be light- or nutrients- independent, and likely belong to the ‘phagocytose when light is limiting’ or ‘perfectly balanced mixotrophy’ functional group. The Type 1/2 strain showed varying responses to light when nutrients were high or low, and are thus seemingly to be in between or different from the other two trophic types. It is also possible that the culture under high nutrient and low light were not well acclimated for grazing, and more experiments need to be conducted for further verification. Based on proposed resource allocation theories, those light-dependent mixotrophs (Type 1) probably rely on photosynthesis for gaining most energy and organic carbon [12–14]. In comparison, the light-independent group performed better with low light availability (Type 2), suggesting a higher flexibility governing autotrophic and heterotrophic growth in response to the changing conditions.

#### 4.2. 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-nutrient and lower-light environments could possibly promote the relative abundances of autotrophs, whereas mixotrophic community overall proliferated from low-nutrient and high-light environments. This positive light–mixotrophy relationship in the field can be partially explained by the possibility that the light-dependent Type 1 mixotrophs were dominant across the investigated ocean regions. On the other hands, Rothhaupt (1996) and Edwards et al. (2019; 2022) suggested that, during the resource competition process, mixotrophs could outcompete heterotrophs, because they can use extra energy gained from photosynthesis which enables them to suppress bacterial prey to a very low level (below the threshold of heterotrophic grazers) [8,11,45]. 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 one of the predictors explaining  $TI_M$ , based on a 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 shift 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 the high grazing pressure they exert on heterotrophic bacteria [46,47]. The weaker but positive correlation between heterotrophic bacteria and  $TI_M$  could be explained by the more complex relations among bacteria (prey and decomposer) and mixotrophs (grazer and DOC provider), as well as the lower grazing capability/pressure from mixotrophic grazers than heterotrophs [13,23,38,48].

#### 4.3. Environmental Variables Affecting the Prevalence of Trophic Strategies

Other oceanographic conditions such as CML, nitracline depths, and MLD also seemed to affect mixotrophic prevalence, though they are physical processes that are more or less associated with nutrient and light conditions. For example, deeper nitraclines can potentially suppress nutrient supply in the surface and, therefore, increase  $TI_M$  values. Similarly, deeper chlorophyll maxima could be indicative of a higher light availability and lower nutrient availability in the photic depths, thereby enhancing  $TI_M$ . In contrast, deeper mixed layer depths (stronger mixing) could reduce light availability and increase nutrient concentration, which both weaken  $TI_M$ . The overall positive association between CML and nitracline depths and  $TI_M$ , and the negative correlation between MLD and  $TI_M$ , are in line with their overall relationships with nutrients and light.

Previous field studies suggested that, besides nutrients [21,49] and light [50–52], variables such as temperature [53] and carbon composition (colored dissolved organic matters and  $CO_2$ ) [20] were also likely important for affecting mixotrophic prevalence. These variables were also included in our analysis and yielded varying relationships with  $TI_g$ . These differences could possibly be due to the different environmental gradients covered among our studies (various open oceans) versus others (North Pacific, and boreal Canadian lakes). The stepwise regression model suggested that PAR, nutrients, density, PIC, CML, heterotrophic bacteria, and *Prochlorococcus* were the top predictors explaining  $TI_M$  for the Tara Oceans dataset (Supplementary Table S2).

Indeed, it is not always straightforward to delineate relationships between a single environmental variable and trophic mode prevalence. Results from our statistical analysis instead provide insights into the possible correlations between them, and unnecessarily the causal relationships. Future studies focusing on smaller regions, with in situ

experiments targeting active trophic behaviors, will provide additional knowledge for interpreting the relationships between trophic mode and the environment.

#### 4.4. Methodology Caveats

One potential caveat in our methodology is that mixotrophs were annotated based on the published literature at a species or higher taxonomic rank. These trophic annotations did not necessarily align with in situ metabolic activities. Nevertheless, the potential or historically evolved capability of phago-mixotrophy was 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.,  $\text{NO}_3^-$ ,  $\text{NO}_2^-$ ,  $\text{PO}_4^{3-}$  and  $\text{SiO}_4^-$  for nutrients), to reduce the possibility of retrieving false relationships. A 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*, a member of Chlorophyta, as autotrophs, based on the best knowledge we have [43,54]. However, others have argued for possible phagotrophic evidence in *Micromonas polaris* [55]. We also recognize that the gene and 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 for under-estimated small flagellates.

Combining various approaches including experimental study, field observation, model simulation, and bioinformatics tools can be a powerful solution in terms of retrieving the most reliable data for future mixotrophy research [3,56].

## 5. Conclusions

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

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/microorganisms12040750/s1>, Reference [57] is cited in the Supplementary Materials.

**Author Contributions:** Conceptualization, Q.L.; methodology, Q.L. and K.D.; formal analysis, Q.L. and K.D.; writing—original draft preparation, K.D.; writing—review and editing, Q.L., K.D., Y.W. and W.Z.; supervision, Q.L.; funding acquisition, Q.L. All authors have read and agreed to the published version of the manuscript.

**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).

**Data Availability Statement:** All datasets used in this study can be downloaded from the Supplementary Material. 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).

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

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Sherr, E.B.; Sherr, B.F. Significance of predation by protists in aquatic microbial food webs. *Antonie Leeuwenhoek* **2002**, *81*, 293–308. [\[CrossRef\]](#) [\[PubMed\]](#)
- Stoecker, D.K.; Hansen, P.J.; Caron, D.A.; Mitra, A. Mixotrophy in the Marine Plankton. *Annu. Rev. Mar. Sci.* **2017**, *9*, 311–335. [\[CrossRef\]](#) [\[PubMed\]](#)
- Millette, N.C.; Gast, R.J.; Luo, J.Y.; Moeller, H.V.; Stamieszkin, K.; Andersen, K.H.; Brownlee, E.F.; Cohen, N.R.; Duhamel, S.; Dutkiewicz, S.; et al. Mixoplankton and mixotrophy: Future research priorities. *J. Plankton Res.* **2023**, *45*, 576–596. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bock, N.A.; Charvet, S.; Burns, J.; Gyaltsen, Y.; Rozenberg, A.; Duhamel, S.; Kim, E. Experimental identification and in silico prediction of bacterivory in green algae. *ISME J.* **2021**, *15*, 1987–2000. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, Q.; Edwards, K.F.; Schvarcz, C.R.; Steward, G.F. Broad phylogenetic and functional diversity among mixotrophic consumers of *Prochlorococcus*. *ISME J.* **2022**, *16*, 1557–1569. [\[CrossRef\]](#) [\[PubMed\]](#)
- Flynn, K.J.; Mitra, A.; Anestis, K.; Anschütz, A.A.; Calbet, A.; Ferreira, G.D.; Gypens, N.; Hansen, P.J.; John, U.; Martin, J.L.; et al. Mixotrophic protists and a new paradigm for marine ecology: Where does plankton research go now? *J. Plankton Res.* **2019**, *41*, 375–391. [\[CrossRef\]](#)
- Mitra, A.; Flynn, K.J.; Tillmann, U.; Raven, J.A.; Caron, D.; Stoecker, D.K.; Not, F.; 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. [\[CrossRef\]](#) [\[PubMed\]](#)
- Rothhaupt, K.O. Laboratory Experiments with a Mixotrophic Chrysophyte and Obligately Phagotrophic and Photographic Competitors. *Ecology* **1996**, *77*, 716–724. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Anderson, R.; Charvet, S.; Hansen, P.J. Mixotrophy in Chlorophytes and Haptophytes—Effect of Irradiance, Macronutrient, Micronutrient and Vitamin Limitation. *Front. Microbiol.* **2018**, *9*, 1704. [\[CrossRef\]](#)
- Edwards, K.F. Mixotrophy in nanoflagellates across environmental gradients in the ocean. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 6211–6220. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ward, B.A. Mixotroph ecology: More than the sum of its parts. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 5846–5848. [\[CrossRef\]](#) [\[PubMed\]](#)
- Raven, J.A. Phagotrophy in phototrophs. *Limnol. Oceanogr.* **1997**, *42*, 198–205. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Hansen, P.; Anderson, R.; Stoecker, D.; Decelle, J.; Altenburger, A.; Blossom, H.; Drumm, K.; Mitra, A.; Flynn, K. Mixotrophy among Freshwater and Marine Protists. In *Encyclopedia of Microbiology*; Elsevier: Amsterdam, The Netherlands, 2019.
- 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. *Prog. Oceanogr.* **2021**, *190*, 102481. [\[CrossRef\]](#)
- Zubkov, M.V.; Tarran, G.A. High bacterivory by the smallest phytoplankton in the North Atlantic Ocean. *Nature* **2008**, *455*, 224–226. [\[CrossRef\]](#) [\[PubMed\]](#)
- 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. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 5756–5760. [\[CrossRef\]](#) [\[PubMed\]](#)
- 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. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12776–12781. [\[CrossRef\]](#)
- Hansson, T.H.; Grossart, H.-P.; del Giorgio, P.A.; St-Gelais, N.F.; Beisner, B.E. Environmental drivers of mixotrophs in boreal lakes. *Limnol. Oceanogr.* **2019**, *64*, 1688–1705. [\[CrossRef\]](#)
- Le Noac’h, P.; Cremella, B.; Kim, J.; Soria-Piriz, 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. *J. Plankton Res.* **2024**, *46*, 9–24. [\[CrossRef\]](#)
- Wilken, S.; Huisman, J.; Naus-Wiezer, S.; Van Donk, E. Mixotrophic organisms become more heterotrophic with rising temperature. *Ecol. Lett.* **2013**, *16*, 225–233. [\[CrossRef\]](#) [\[PubMed\]](#)
- Berge, T.; Chakraborty, S.; Hansen, P.J.; Andersen, K.H. Modeling succession of key resource-harvesting traits of mixotrophic plankton. *ISME J.* **2017**, *11*, 212–223. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lepori-Bui, M.; Paight, C.; Eberhard, E.; Mertz, C.M.; Moeller, H.V. Evidence for evolutionary adaptation of mixotrophic nanoflagellates to warmer temperatures. *Glob. Change Biol.* **2022**, *28*, 7094–7107. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chu, T.; Moeller, H.V.; Archibald, K.M. Competition between phytoplankton and mixotrophs leads to metabolic character displacement. *Ecol. Model.* **2023**, *481*, 110331. [\[CrossRef\]](#)
- 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. *Sci. Data* **2015**, *2*, 150023. [\[CrossRef\]](#) [\[PubMed\]](#)
- Picheral, M.; Searson, S.; Taillandier, V.; Bricaud, A.; Boss, E.; Stemann, 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. *Sci. Data* **2014**, *10*, 347. [\[CrossRef\]](#)

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 100 km. In Tara Oceans Consortium, Coordinators; Tara Oceans Expedition, Participants (2017): Registry of all Samples from the Tara Oceans Expedition (2009–2013). PANGAEA. 2017. Available online: <https://b2find9.cloud.dkrz.de/dataset/ac50a5b1-afdf-5ce5-bb99-78bb1b446c82> (accessed on 23 March 2024).
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. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Faure, E.; Not, F.; Benoiston, A.-S.; Labadie, K.; Bittner, L.; Ayata, S.-D. Mixotrophic protists display contrasted biogeographies in the global ocean. *ISME J.* **2019**, *13*, 1072–1083. [\[CrossRef\]](#) [\[PubMed\]](#)
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. *Mol. Ecol. Resour.* **2023**, *23*, 16–40. [\[CrossRef\]](#)
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. [\[CrossRef\]](#)
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 Res.* **2012**, *41*, D597–D604. [\[CrossRef\]](#) [\[PubMed\]](#)
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 Microbiol. Ecol.* **2005**, *52*, 79–92. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Gong, W.; Marchetti, A. Estimation of 18S Gene Copy Number in Marine Eukaryotic Plankton Using a Next-Generation Sequencing Approach. *Front. Mar. Sci.* **2019**, *6*, 219. [\[CrossRef\]](#)
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 Metagenom.* **2022**, *6*, e85794. [\[CrossRef\]](#)
37. Li, Q.; Dong, K.; Wang, Y.; Edwards, K.F. Relative importance of bacterivorous mixotrophs in an estuary-coast environment. *Limnol. Oceanogr. Lett.* **2023**, *9*, 81–91. [\[CrossRef\]](#)
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. *J. Eukaryot. Microbiol.* **2023**, *70*, e12972. [\[CrossRef\]](#) [\[PubMed\]](#)
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. [\[CrossRef\]](#)
40. Li, Q.; Edwards, K.F.; Schvarcz, C.R.; Selph, K.E.; Steward, G.F. Plasticity in the grazing ecophysiology of *Florentiella* (Dichtyochophyceae), a mixotrophic nanoflagellate that consumes *Prochlorococcus* and other bacteria. *Limnol. Oceanogr.* **2021**, *66*, 47–60. [\[CrossRef\]](#)
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. *Aquat. Microb. Ecol.* **2008**, *53*, 83–98. [\[CrossRef\]](#)
42. Fenchel, T. The Ecology of Heterotrophic Microflagellates. In *Advances in Microbial Ecology*; Marshall, K.C., Ed.; Springer: Boston, MA, USA, 1986; pp. 57–97.
43. Edwards, K.F.; Li, Q.; McBeain, K.A.; Schvarcz, C.R.; Steward, G.F. Trophic strategies explain the ocean niches of small eukaryotic phytoplankton. *Proc. R. Soc. B Biol. Sci. USA* **2023**, *290*, 20222021. [\[CrossRef\]](#)
44. Stoecker, D.K. Conceptual models of mixotrophy in planktonic protists and some ecological and evolutionary implications. *Eur. J. Protistol.* **1998**, *34*, 281–290. [\[CrossRef\]](#)
45. Jones, H. A classification of mixotrophic protists based on their behaviour. *Freshw. Biol.* **1997**, *37*, 35–43. [\[CrossRef\]](#)
46. Fuhrman, J.A.; Noble, R.T. Viruses and protists cause similar bacterial mortality in coastal seawater. *Limnol. Oceanogr.* **1995**, *40*, 1236–1242. [\[CrossRef\]](#)
47. Pernthaler, J. Predation on prokaryotes in the water column and its ecological implications. *Nat. Rev. Microbiol.* **2005**, *3*, 537–546. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Edwards, K.F.; Li, Q.; Steward, G.F. Ingestion kinetics of mixotrophic and heterotrophic flagellates. *Limnol. Oceanogr.* **2023**, *68*, 917–927. [\[CrossRef\]](#)
49. Stukel, M.R.; Landry, M.R.; Selph, K.E. Nanoplankton mixotrophy in the eastern equatorial Pacific. *Deep. Sea Res. Part II Top. Stud. Oceanogr.* **2011**, *58*, 378–386. [\[CrossRef\]](#)
50. 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. *Sci. Rep.* **2016**, *6*, 29286. [\[CrossRef\]](#)
51. Millette, N.C.; Pierson, J.J.; Aceves, A.; Stoecker, D.K. Mixotrophy in *Heterocapsa rotundata*: A mechanism for dominating the winter phytoplankton. *Limnol. Oceanogr.* **2017**, *62*, 836–845. [\[CrossRef\]](#)
52. Anderson, R.; Hansen, P.J. Meteorological conditions induce strong shifts in mixotrophic and heterotrophic flagellate bacterivory over small spatio-temporal scales. *Limnol. Oceanogr.* **2020**, *65*, 1189–1199. [\[CrossRef\]](#)
53. 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. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2100916119. [\[CrossRef\]](#)

54. Jimenez, V.; Burns, J.A.; Le Gall, F.; Not, F.; Vault, D. No evidence of Phago-mixotrophy in *Micromonas polaris* (Mamiellophyceae), the Dominant Picophytoplankton Species in the Arctic. *J. Phycol.* **2021**, *57*, 435–446. [[CrossRef](#)] [[PubMed](#)]
55. McKie-Krisberg, Z.M.; Sanders, R.W. Phagotrophy by the picoeukaryotic green alga *Micromonas*: Implications for Arctic Oceans. *ISME J.* **2014**, *8*, 1953–1961. [[CrossRef](#)] [[PubMed](#)]
56. Beisner, B.E.; Grossart, H.-P.; Gasol, J.M. A guide to methods for estimating phago-mixotrophy in nanophytoplankton. *J. Plankton Res.* **2019**, *41*, 77–89. [[CrossRef](#)]
57. Godhe, A.; Asplund, M.E.; Hårnström, K.; Saravanan, V.; Tyagi, A.; Karunasagar, I. Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR. *Appl. Environ. Microbiol.* **2008**, *74*, 7174–7182. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.