



HAL
open science

Diversity, community structure, and quantity of eukaryotic phytoplankton revealed using 18S rRNA and plastid 16S rRNA genes and pigment markers: A case study of the Pearl River Estuary

Shumin Xu, Guihao Li, Cui He, Yi Huang, Dan Yu, Huiwen Deng, Zhuyin Tong, Yichong Wang, Christine Dupuy, Bangqin Huang, et al.

► To cite this version:

Shumin Xu, Guihao Li, Cui He, Yi Huang, Dan Yu, et al.. Diversity, community structure, and quantity of eukaryotic phytoplankton revealed using 18S rRNA and plastid 16S rRNA genes and pigment markers: A case study of the Pearl River Estuary. *Marine Life Science & Technology*, 2023, 5 (3), pp.415-430. 10.1007/s42995-023-00186-x . hal-04205394

HAL Id: hal-04205394

<https://hal.science/hal-04205394v1>

Submitted on 15 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Running title: Characterizing phytoplankton using multiple biomarkers

Diversity, community structure, and quantity of eukaryotic phytoplankton revealed using 18S rRNA and plastid 16S rRNA genes and pigment markers: A case study of the Pearl River Estuary

Shumin Xu^{1,2}, Guihao Li¹, Cui He¹, Yi Huang¹, Dan Yu¹, Huiwen Deng¹, Zhuyin Tong³, Yichong Wang³, Christine Dupuy⁴, Bangqin Huang³, Zhuo Shen¹, Jie Xu⁵, Jun Gong^{1,2*}

¹ School of Marine Sciences, Sun Yat-Sen University (Zhuhai Campus), and Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519000, China

² Guangdong Provincial Key Laboratory of Marine Resources and Coastal Engineering, Guangzhou 510006, China

³ State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen 361102, China

⁴ BIOFEEL, UMRi LIENSs, La Rochelle Université / CNRS, France

⁵ Centre for Regional Oceans, Department of Civil and Environmental Engineering, Faculty of Science and Technology, University of Macau, Macau, China

***Correspondence:**

Jun Gong, gongj27@mail.sysu.edu.cn

Abstract

Understanding consistencies and discrepancies in characterizing diversity and quantity of phytoplankton is essential for better modeling ecosystem change. In this study, eukaryotic phytoplankton in the Pearl River Estuary, South China Sea were investigated using nuclear 18S rRNA and plastid 16S or 23S rRNA genes and pigment analysis. It was found that 18S abundance poorly explained the variations in total chlorophyll *a* (Chl-*a*). However, the ratios of log-transformed 18S abundance to Chl-*a* in the major phytoplankton groups were generally environment-dependent, suggesting that the ratio has potential as an indicator of the physiological state of phytoplankton. The richness of 18S-based operational taxonomic units was positively correlated with the richness of 16S-based amplicon sequence variants of the whole phytoplankton community, but insignificant or weak for individual phytoplankton groups. Overall, the 18S-based, rather than the 16S-based, community structure had a greater similarity to pigment-based estimations. Relative to the pigment data, the proportion of haptophytes in the 18S dataset, and diatoms and cryptophytes in the 16S dataset, were underestimated. This study highlights that 18S metabarcoding tends to reflect biomass-based community organization of eukaryotic phytoplankton. Because there were lower copy numbers of plastid 16S than 18S per genome, metabarcoding of 16S probably approximates abundance-based community organization. Changes in biomass organization of the pigment-based community was sensitive to environmental changes. Taken together, it is recommended multiple methodologies be applied to more accurately profile the diversity and community composition of phytoplankton in natural ecosystems.

Key words: Accessory pigments · Chloroplast · Marker gene · Metabarcoding · Phytoplankton

Introduction

Phytoplankton are photosynthetic microbes comprising taxonomically diverse species of cyanobacteria and eukaryotic microalgae in aquatic ecosystems. Marine phytoplankton are responsible for about half of our planet's annual net primary production (Falkowski et al. 2004; Field et al. 1998). The tremendous importance of phytoplankton in ecology and biogeochemical processes (e.g., cycling of nitrogen, phosphorus, silicate and iron, driving carbon export to deep waters) has triggered the development of chemical and molecular methodologies to characterize their diversity and community composition across time and space (Abaychi and Riley 1979; Chen et al. 2022; Gao et al. 2020; Maki et al. 2017; Treusch et al. 2012; Xie et al. 2022). Relative to classical methodologies based on morphological identification and enumeration (Huang et al. 2020), chemotaxonomy of accessory pigments by high-performance liquid chromatography (HPLC) enables the rapid quantification of major groups of phytoplankton (e.g., diatoms, dinophytes, cyanobacteria, chlorophytes, haplotypes, cryptophytes; (Everitt et al. 1990; Mackey et al. 1996; Wright et al. 1991). The total chlorophyll *a* (Chl-*a*) content has been widely used as a proxy of phytoplankton biomass and the biomass of each phytoplankton group can also be quantified (Everitt et al. 1990; Yang et al. 2019). The analysis of pigments provides empirical data for calibrating remote sensing of functional types of phytoplankton in local and global oceans (Claustre et al. 2004; Sathyendranath et al. 2001). However, the ratio of cellular biomass in terms of carbon content to Chl-*a* content (C: Chl-*a*), an indicator of the physiological state of microalgae, varies greatly with environmental factors and among phytoplankton groups (Geider 1987; Sathyendranath et al. 2009). Furthermore, the taxonomic resolution of phytoplankton by chemotaxonomy is relatively low (class level, at best) (Eker-Develi et al. 2012).

Molecular approaches targeting nuclear 18S ribosomal RNA genes have been applied to characterize the diversity and community structure of unicellular eukaryotes at lower taxonomic ranks (e.g., family, genus, and even species levels) (Belevich and Milyutina 2022; Chen et al. 2019; Guo et al. 2016; Liu et al. 2021; Rii et al. 2018). With increasing sequencing depth and improvements in publicly available reference databases, rare eukaryotic taxa can be routinely sampled and classified (Tragin et al. 2018; Wang et al. 2022; Ye et al. 2022). Although metabarcoding of 18S rRNA genes covers both pigmented (phytoplankton) and non-pigmented eukaryotes (protozoa and fungi), the operational taxonomic units (OTUs) of eukaryotic phytoplankton taxa can be retrieved after classification to reassemble the eukaryotic phytoplankton communities (Kirkham et al. 2011; Trefault et al. 2021). For molecular quantification of phytoplankton

eukaryotes, the 18S rRNA gene abundance of a taxon can be calculated by multiplying the total 18S rRNA gene copy number of the whole community (determined using quantitative real-time PCR assay, qPCR) with the sequence proportion of that taxon (Gong et al. 2020a; Zhu et al. 2005). There was a good correlation between the relative abundance of 18S rRNA genes and the proportion of pigment content of major phytoplankton groups in the Neuse River Estuary (Gong et al. 2020b); nevertheless, the relative abundance of 18S rRNA genes might not always reflect variations in pigment content, such as on the Western Antarctica Peninsula (Lin et al. 2018). It thus remains important to investigate the relationship between pigment content and rDNA abundance in different phytoplankton groups, which may have varying genome sizes across taxa (Lin 2011; Veldhuis et al. 1997). Furthermore, microalgal pigment content in phytoplankton species is well known to be cell size-dependent, and regulated by multiple environmental factors (e.g., light intensity, nutrient supply, and temperature) (Geider et al. 1986; Kana et al. 1997). Given that the relationship between 18S rRNA gene copy number and biovolume of a protistan cell follows a power law function, and the biovolume (a proxy of biomass) also varies with temperature (Fu and Gong 2017; Godhe et al. 2008), it is interesting to explore the relative importance of environmental conditions in modulating the rRNA gene abundance - pigment content relationship in communities of eukaryotic phytoplankton.

The 16S rRNA genes in chloroplast genomes of eukaryotes and Cyanobacteria can be targeted and amplified using plastid-specific PCR primers (Decelle et al. 2015). In contrast to the high levels and large variability of copy numbers of 18S rRNA genes in eukaryotic genomes (Gong et al. 2013; Salmaso et al. 2020), the copy numbers of plastid 16S rRNA genes per cell are relatively low and constant (Bennke et al. 2018; Shi et al. 2011), which raises the question of how different the diversity and community structure are when determined by these two molecular markers. Recently, comparative studies targeting both plastid 16S and 18S rRNA genes have found similar temporal patterns of phytoplankton community structure (Needham and Fuhrman 2016), and good correlations between plastid 16S and 18S rRNA gene abundances of cryptophytes and three diatom species (Lin et al. 2018), and between 16S- and 18S-based ratios of relative abundance of many shared phytoplankton classes (Yeh and Fuhrman 2022). However, the quantitative relationships between plastid rRNA gene abundances and pigment contents in specific phytoplankton groups, whether the 16S- and 18S-based alpha diversity estimators of eukaryotic phytoplankton are consistent in spatiotemporal patterns, and which taxa are strongly biased in the 16S- or 18S-based phytoplankton communities, against those characterized using pigments, still need to be investigated.

The Pearl River Estuary (PRE) is a tropical estuary off the South China Sea, where phytoplankton blooms in the middle part of the estuary during the dry season (Jia et al. 2019; Lu and Gan 2015; Niu et al. 2020; Qiu et al. 2019). Phytoplankton diversity, abundance, and community structure have been studied based on either morphological identification (Chen et al. 2019; Jiang et al. 2015), 18S rRNA gene sequencing (e.g., Wu and Liu 2018), or pigment analysis (Chai et al. 2016). In this study, we investigate the spatial and seasonal patterns of eukaryotic phytoplankton in the PRE using a combination of three methodologies: pigment analysis and molecular approaches targeting both nuclear and plastid rRNA genes. The main aims were: 1) to assess the quantitative linkages between nuclear and plastid rRNA gene abundances and pigment contents of eukaryotic phytoplankton; 2) to quantify the relative importance of environmental factors in explaining the variation in the ratio of 18S rRNA gene copy number (log-transformed) to the content of Chl-*a* (a proxy of C: Chl-*a*); and 3) to explore potential advantages and limitations of the two molecular approaches in covering alpha and beta diversity of phytoplankton groups.

Results

Hydrological parameters

The profiles of temperature, salinity and Chl-*a* in the PRE showed significant spatial and seasonal variability (Fig. 2A-L). The water column was distinctly stratified in July, with high temperatures (28 - 30.2 °C) in the 10-m surface layer and lower temperatures (22 - 24 °C) below a depth of 15 m (Fig. 2A-C). The thermal stratification disappeared in November, with uniform temperature ranging from 23.6 to 25.0 °C (Fig. 2D). A saltwater wedge was distinct in all transects and both seasons (Fig. 2E-H). Along Transect 1, the freshwater tongue extended further seawards in July than in November. The highest total Chl-*a* concentration (determined by HPLC) in surface waters occurred along an arc linking the sites S10 (9.13 $\mu\text{g L}^{-1}$), S43 (9.18 $\mu\text{g L}^{-1}$), and S63 (6.93 $\mu\text{g L}^{-1}$; Fig. 2I-K) in July, which corresponded to a diatom bloom (Fig. 3I-K). However, the high levels of surface Chl-*a* (up to 3.2 ~ 5.4 $\mu\text{g L}^{-1}$) occurred at the inner estuary in November (Fig. 2L), indicating phytoplankton blooms in that area.

The concentration of dissolved oxygen (DO) ranged from 0.55 to 6.11 mg L^{-1} in the summer with much lower values (< 3 mg L^{-1}) in the bottom waters (depth of 10 to 20 m) at sites S8, S10, S41, S43, and S63 (Fig. S1). The bottom hypoxic zones usually co-occurred when there were phytoplankton blooms in the surface waters. In general, the concentrations of NO_3^- (0.21-138.64 μM), NO_2^- (0.01-16.61 μM), soluble reactive

phosphorus (SRP, 0.02-1.11 μM), and dissolved silicate (DSi, 3.95-107.96 μM) varied markedly in space and between seasons but generally decreased from freshwater to more saline sites. However, NH_4^+ concentration (0.97-11.86 μM) was generally higher at the more saline sites and in the deeper layers (Fig. S1).

Nuclear 18S and plastid 23S rRNA gene abundances

The distribution of both nuclear 18S and plastid 16S rRNA gene copy numbers of eukaryotic phytoplankton was uncoupled with the concentration of Chl-*a* along the PRE transects (Fig. 2M - T). Unlike Chl-*a* that peaked at the surface, the 18S rRNA genes were most abundant at the bottom in Transects 1 and 2 (up to 7.5×10^8 and 3.2×10^8 copies/L). Nevertheless, for the samples collected at Transect 3 in July and Transect 1 in November, the higher abundances of 18S rRNA genes (4.2×10^9 copies/L) were coincident with higher levels of Chl-*a* (Fig. 2M - P). Unlike Chl-*a* and 18S rRNA genes, the plastid 23S rRNA gene abundance was not that variable in space ($0 \sim 6.52 \times 10^8$ copies/L), with peaks in a few surface samples and shallow sites in Transect 1 in both seasons (Fig. 2Q-T).

Distribution of major phytoplankton groups based on pigment content

The Chl-*a* contents of dinophytes, chlorophytes, diatoms, cryptophytes, and haptophytes exhibited high variability across the transects and the two seasons (Fig. 3A-T). Spatially, all groups were abundant in surface waters (depth <10 m), except for Haptophyta, which had the highest biomass (0.14 to 0.16 μg Chl-*a*/L) in the bottom waters at two marine sites (S13 and S45 with depths about 30 m; Fig. 3Q, R). During the summer cruise, the eukaryotic phytoplankton pigment was dominated by diatoms (0.03 – 5.96 μg Chl-*a*/L), followed by dinophytes (0 – 1.75 μg Chl-*a*/L) and chlorophytes (0.03 – 1.80 μg Chl-*a*/L); cryptophytes (0.02 – 0.94 μg Chl-*a*/L), and haptophytes (0 – 0.16 μg Chl-*a*/L) were minor. In November, dinophytes bloomed at the less saline sites of Transect 1, contributing up to 4.3 μg Chl-*a*/L (Fig. 3D), and other groups contributed only small fractions to the surface Chl-*a* (Fig. 3H, L, P, T).

Environmental factors influencing the relationships among pigments, nuclear and plastid rRNA abundances

There was no significant correlation between total Chl-*a* and 18S rRNA gene abundance of eukaryotic phytoplankton ($P > 0.05$; Fig. 4A). The 18S abundance explained 41% of the variability of the pigmented

dinophyte Chl-*a* content ($P < 0.05$). For other groups, the explanatory power reduced greatly, being 20% and 12% for haptophytes and diatoms, respectively ($P < 0.05$), and none for chlorophytes and cryptophytes ($P > 0.05$; Fig. 4B). The correlations between 18S and plastid 23S rRNA gene abundances and pigment contents were usually weak or insignificant for diatoms, chlorophytes, cryptophytes and haptophytes, but stronger for dinoflagellates ($0.17 < R^2 < 0.42$, $P < 0.05$; Fig. S2).

The ratios of 18S rRNA gene abundance ($\log(x+1)$ transformed) to Chl-*a* content of the four microalgal groups varied widely across season and space, ranging from 2 to 5726 (on average 398) in the pigmented dinophytes, and from 1 to 88 in diatoms (average 22), chlorophytes (average 40), and cryptophytes (average 45) (Fig. S3). Regression analysis indicated that the ratio for pigmented dinophytes was significantly related to NO_3^- ($R^2 = 0.57$; $P < 0.01$) and salinity ($R^2 = 0.57$; $P < 0.01$); Fig. 4C, D). The 18S to Chl-*a* ratio in both diatoms and chlorophytes tended to decrease with increasing total phytoplankton biomass ($R^2 = 0.25$; $P < 0.01$; Fig. 4E, F). In contrast, the cryptophytes had a higher level of 18S rRNA gene abundance per μg Chl-*a* in a deeper layer ($R^2 = 0.46$; $P < 0.01$), or in waters with a lower concentration of DSi ($R^2 = 0.42$; $P < 0.01$; Fig. 4G, H).

Diversity and taxonomic composition based on 18S, 16S, and pigment analysis

The raw sequencing data (4,635,358 reads of 18S rRNA genes and 4,755,849 reads of 16S rRNA genes) were processed and analyzed using QIIME2. A total of 356,968 and 550,538 reads were retained for nuclear 18S and plastid 16S rRNA genes after quality control, respectively. A total of 712 OTUs and 5531 amplicon sequence variants (ASVs) were detected for 18S and 16S rRNA genes, respectively. By rarefying at 3789 reads per sample for 18S and 3000 reads per sample for 16S, the abundance of both molecular markers in surface waters was mapped, showing a decrease towards the sea in July (Fig. 5A-C, E-G), but the opposite trend in November (Fig. 5D, H).

Regression analyses showed that the importance of environmental factors in driving the alpha diversity of eukaryotic phytoplankton was different when different molecular markers were considered (Fig. 6A-D, Fig. S4). Salinity ($R^2 = 0.67$, $P < 0.001$) and NO_3^- ($R^2 = 0.45$, $P < 0.001$) were the most significant environmental factors in explaining the variation in 18S-based OTU richness and Chao1 index (Fig. 6A, B, Fig. S4B), whereas SRP ($R^2 = 0.26$, $P < 0.001$) and salinity ($R^2 = 0.17$, $P < 0.001$) were most significant for 16S-based ASV richness (Fig. 6C, D).

Linear regression analysis showed that 18S OTU richness of eukaryotic phytoplankton was positively related to plastid 16S ASV richness (slope = 0.576), and explained 20% of the variation across all samples examined ($P < 0.01$; Fig. 6E). However, for an individual group of phytoplankton, the regressions were either insignificant (diatoms and dinophytes, $P > 0.05$; Fig. 6F, G) or much weaker (cryptophytes and haplotypes, Fig. 6I, J). There was even a negative relationship between the richness of these two markers in chlorophytes (Fig. 6H). Correlations between other 16S- and 18S-based alpha diversity indices of eukaryotic phytoplankton or an individual group were weak or insignificant too (Figs S5, S6).

As far as the five major microalgal groups were concerned, the community composition of eukaryotic phytoplankton characterized using the plastid 16S rRNA gene was quite different from those documented by metabarcoding 18S rRNA genes and using pigment analysis, whereas the latter two methods appeared to yield highly concordant results (Fig. 7). Most plastid 16S rRNA gene reads were affiliated with Chlorophyta (on average 55%) and pigmented Dinophyta (32%). Bacillariophyta (< 5%) and Cryptophyta (< 1.2%), which had much lower proportions in 16S-based communities, however, appeared to be much more abundant in both 18S (average 36% and 3%) and pigment (average 43% and 18%) datasets. The haptophyte pigment was frequently abundant (0 – 34%) in the HPLC samples, but this group was rarely identified in the plastid 16S (< 1.3%), and hardly detected (< 0.7%) in the 18S dataset. Other Ochrophyta (i.e., the ochrophytes not including diatoms) frequently occurred with low to moderate proportions of reads (0.2% – 21%) in the 18S dataset, however, was detected in only a few samples by 16S sequencing, and was never identified by pigment analysis (Fig. 7). Blooms of dinophytes in seven samples (Chl-*a* percentages > 50%, and total Chl-*a* > 1 µg/L) were well reflected by their high sequence proportions in the 18S dataset (asterisks in Fig. 7). However, these signals of dinophyte blooms were not distinguishable in the 16S dataset. In hypoxic waters, pigment-based biomass of chlorophytes appeared to be lower than those in normoxic samples, which, however, was not pronounced in the 18S and 16S datasets (arrows in Fig. 7).

The community composition of eukaryotic phytoplankton resolved at the order level showed that some orders were identified only in one of the 18S and 16S datasets (Fig. S7). For example, orders Gymnodiniales and Suessiales, Prorocentrales, and Gonyaulacales were not resolved in the 16S (Fig. S7A-C). Two diatom species *Cyclotella* sp. (Order Thalassiosirales) and *Guinardia delicatula* (Order Rhizosoleniales) appeared to be dominant in the 18S dataset but were not identified from the 16S (Fig. S7D-F). The most dominant taxa within cryptophytes were assigned to two different orders in 18S (Order Cryptomonadales) and 16S

datasets (Order Pyrenomonadales) (Fig. S7J-L).

Regression analysis showed that the ratio of 18S rRNA gene abundances between two of the four microalgal groups generally related well to their ratio of Chl-*a* contents ($0.3 < R^2 < 0.66$, $P < 0.05$; Fig. 8). However, the correlations between chlorophytes and cryptophytes ($R^2 = 0.12$, $P < 0.01$), and between haptophytes versus all other groups were weak ($0.05 < R^2 < 0.12$; Fig. 8). There were no significant relationships between the ratio of 18S and plastid 16S rRNA gene abundance of any two of the five major microalgal groups (Fig. S8A). The relations between the ratio of plastid 16S and the ratio of pigment content were weak or not significant (Fig. S8B).

The community structure of eukaryotic phytoplankton based on the five major microalgal groups revealed using plastid 16S rRNA genes were much more different from those assessed by pigment analysis than by metabarcoding 18S rRNA genes (Fig. 9A). Compared with the community structure assessed by pigment contents, chlorophytes, and pigmented dinoflagellates tended to be overestimated in both 18S and 16S based communities, whereas cryptophytes might be underestimated in 18S-based communities. The metabarcoding of 18S rRNA genes, however, suffered from bias against haptophytes, which might be underestimated in 16S-based phytoplankton community as well (Fig. 9B). Pairwise correlations of community distance obtained from different methods also supported that 18S rRNA gene could reflect more variability in pigment-based community structure of eukaryotic phytoplankton in the PRE ($r = 0.39$, $P < 0.001$) than plastid 16S rRNA genes ($r = 0.13$, $P = 0.064$) (Fig. 9C - E).

Environmental factors affecting the community structure of eukaryotic phytoplankton

The CCA plots showed that the measured environmental factors could explain a small portion of variations in both 18S-OTU and plastid 16S-ASV-based community structure of eukaryotic phytoplankton, for which NO_2^- , temperature, salinity, and nutrients were the most significant driving factors (Fig. 10A - B). However, NO_2^- , depth and SRP were found to be significant factors driving the changes in the eukaryotic phytoplankton community based on the pigment proportions of the five major groups (Fig. 10C). The higher concentration of NO_2^- , the higher proportion of 18S and pigment-based biomass of dinophytes (Fig. 10D, F). Relative quantities of the other four phytoplankton groups were mainly influenced by Chl-*a* and NO_3^- in 18S-based communities (Fig. 10D), and by SRP and depth in pigment-based communities (Fig. 10F). In contrast, totally different environmental variables (salinity and DSi) were selected as the major driving

factors for phytoplankton community in the application of plastid 16S approach (Fig. 10E).

Discussion

The spatial and seasonal variations in the community composition of eukaryotic phytoplankton in the tropical PRE provided a good experimental field to examine the consistency and discrepancy of different approaches in characterizing quantity and diversity. Basically, the 18S and pigment-based data of 2020 were consistent with existing surveys of phytoplankton in the PRE using morphological observations. For example, all the studies of Jia et al. (2019), Xu et al. (2022), and Zhong et al. (2020a) demonstrated that diatoms were the dominant phytoplankton group in July, followed by chlorophytes. Dong et al. (2020) observed a bloom of the dinoflagellate *Cochlodinium geminatum* in November 2018. Similarly, a dinoflagellate bloom (likely *Polykrikos geminatus*) was observed in November 2020, which led to significant high 18S sequence proportions of dinoflagellates (> 50%) at all the stations (Fig. 6).

It was found that the 18S-based community structure of eukaryotic phytoplankton was highly concordant with that based on pigment content. This judgment is based on at least the following three facts: (1) the rRNA gene abundance ratio between any two groups of dinophytes, chlorophytes, diatoms, and cryptophytes generally well reflected the ratio of their pigment contents (Fig. 6D-H); (2) the dinophyte blooms identified in the pigment analysis were also well captured in the 18S dataset (Fig. 6A-C); and (3) 18S- and pigment-based community distances were well correlated with each other (Fig. 7C). This result is consistent with Gong et al. (2020b), who demonstrated similar relationships between 18S rRNA gene abundance and pigment content in a study of the community structure of eukaryotic phytoplankton in a temperate estuary of USA. Lin et al. (2019) even found that the absolute abundance of 18S rRNA genes was an even better correlate of pigment contents of cryptophytes, diatoms, and *Phaeocystis* in a coastal region of the Antarctic. Since cellular Chl-*a* content has long been known to be a good indicator of microalgal biomass, these results indicate that the 18S rRNA gene-based community structure approximates a biomass-based community structure of eukaryotic plankton. This notion is supported by the finding of power-law relationships between cellular 18S rRNA gene copy number and cell biovolume in dinoflagellates (Godhe et al. 2008), and several heterotrophic protistan species (Fu and Gong 2017; Zou et al. 2021).

It was found that plastid 16S ASV richness was positively related to 18S OTU richness across all samples, suggesting co-evolution between plastid and nuclear genes in eukaryotic phytoplankton, as

previously demonstrated for plants (Forsythe et al. 2020). Nevertheless, despite a looser definition of “OTU” (i.e., ASV) being applied for 16S than 18S, the 16S richness varied to a lesser extent (with a slope of regression < 1) than the 18S OTU richness, an explanation for which is that plastid 16S rRNA genes are generally more conservative than nuclear 18S rRNA genes in reflecting phytoplankton diversity. Similar results on lower evolutionary rates of plastid genes relative to nuclear ones have been noted in plants (Drouin et al. 2008; Wolfe et al. 1987). Therefore, the use of plastid markers for assessing phytoplankton diversity may be compensatory to the approach of nuclear 18S rRNA genes of eukaryotic phytoplankton (particularly for the lineages which were commonly detected from both 16S and 18S datasets), which frequently leads to an overestimation of species diversity due to intragenomic polymorphisms (Zou et al. 2021).

It has long been known that the ratio cellular C: Chl-*a* is a sensitive indicator of the physiological state of phytoplankton, usually ranging from 10 to 130, and tending to be higher in larger cell size, at higher levels of irradiance, and nitrate availability and growth rate, and at lower temperatures (Geider et al. 1986; Taylor et al. 1997). The ratio also differs between microalgal groups, with the value increasing in the following order: chlorophytes $<$ diatoms $<$ dinophytes (Geider et al. 1986). The 18S copy number per unit Chl-*a* was also found to be highest in dinophytes. Moreover, higher levels of 18S: Chl-*a* of dinophytes occurred in the waters with higher nitrate, and higher 18S: Chl-*a* ratios of diatoms and chlorophytes were detected in the samples with lower total Chl-*a* in bottom waters (Fig. 4C-F), which was consistent with the negative relationships between C: Chl-*a* and total Chl-*a* in the North Atlantic (Taylor et al. 1997) and with the environmental effects on C: Chl-*a* discussed above. The reason for this consistency may be that the cellular 18S rRNA gene copy number of protists scales with cell biovolume (volume-based biomass), as shown for protist species (Fu and Gong 2017; Godhe et al. 2008; Zou et al. 2021). However, what was not expected was that the ratio of cryptophyte 18S: Chl-*a* was higher at the bottom waters, where light availability was low, which is contradictory to the usual increase in pigment content by photoadaptation in low light (Geider et al. 1986; Kana et al. 1997). A possible explanation for this observation is that cryptophytes have accessory pigments such as phycoerythrin and phycocyanin, which enable high photo-absorption and growth rates in red- and blue-light dominated environments, such as at depth in estuaries (Heidenreich and Richardson 2020). Furthermore, the mixotrophic cryptophytes can adapt by shifting to a heterotrophic lifestyle as bacterial grazers (Hansen et al. 2019), when the irradiance became limiting for active photosynthesis. In short, this study suggests that the ratio 18S: Chl-*a* could be a potential alternative to C: Chl-*a* in correcting and

modeling pigment-based biomass of phytoplankton.

The proportion of haptophytes was less than 0.7% in the 18S dataset, although this could be underestimated, considering the pigment-based biomass of haptophytes in PRE was up to 0.2 $\mu\text{g Chl-}a/L$, and 34% of the eukaryotic phytoplankton community in the present study and in a previous investigation (Chai et al. 2016). This underestimation of haptophytes in the 18S dataset could be due to the eukaryote-universal 18S rRNA gene primers being strongly biased against Prymnesiales (Haptophyta) (Yeh et al. 2021; Liu et al. 2009), and their higher GC content (57%) in 18S than many other groups, which undermines the efficiency of PCR amplification (Liu et al. 2009). This bias could also explain the observations that the 18S ratio of haptophytes and other microalgal groups poorly reflects their pigment-based biomass proportions (Fig. 8). Similarly, the cryptophyte biomass (average 18%) might also be underestimated in the 18S dataset (average 3%). Since the 18S ratio of cryptophytes to other groups was well correlated with their pigment ratio (Fig. 8), indicating other cellular characteristics (e.g., having fewer rRNA gene copies per cell than other microalgae) may underlie the disproportional rRNA gene of this group. Given the importance of haptophyte and cryptophyte biomass in the phytoplankton of many PRE samples, these two groups could be investigated using group-specific primers targeting 18S rRNA genes; cellular rRNA gene copies have yet to be quantified for a better understanding of their diversity and quantity in coastal systems.

The plastid 16S-based community structure of eukaryotic phytoplankton was much different from those of both the pigment- and 18S-based structures. It was also demonstrated that 515F and 926R have good coverage for the plastid 16S (Mcnicol et al. 2021). Similar to a previous study that found contrasting relative abundances of diatoms and cryptophytes in 18S and 16S datasets across some Antarctic samples (Hamilton et al. 2021), the proportions of diatoms, cryptophytes, haptophytes in the 16S databases were also found to be much lower than those in the 18S-based datasets. There are many 18S rRNA gene copies (thousands to hundreds of thousands) in a microalgal cell, whereas the plastid 16S rRNA gene copy numbers have been reported to be much lower and less variable, often one to few dozen copies per genome in some phytoplankton species (Bennke et al. 2018, Needham and Fuhrman 2016). This suggests that 16S-datasets likely approximate cell abundance-based structures rather than a biomass-oriented organization of eukaryotic communities, to which pigment and 18S datasets are more relevant. Nevertheless, the copy number of plastid 16S of more phytoplankton species, and the relationship between cell abundance and total plastid 16S copy

number has yet to be further experimentally explored. Furthermore, there were discrepancies in identifying lower-ranked taxa using these two markers in this study. Only a few phytoplankton genera were commonly detected (Trefault et al. 2021). The differences in recovering lower-ranked taxa could contribute to not only their proportional differences in the communities but also the diversity estimators by using 18S and 16S rRNA gene sequencing. In addition, it should be noted that, although the pigmented dinophyte 16S appeared to be abundant, underestimation of their abundance is still possible, since the plastid rRNA genes of many dinoflagellates are difficult to amplify and there are not many sequences curated in the PhytoREF database (Decelle et al. 2015).

Concluding remarks

The molecular and pigment data collected from a tropical estuary were analyzed to explore whether similar variational patterns in diversity, quantity, and community structure of eukaryotic phytoplankton could be obtained using different methodologies. In general, it was found that there were insignificant or poor correlations among 18S rRNA, plastid 23S rRNA gene abundance, and Chl-*a* content, and between richness of 18S OTUs and plastid 16S ASVs of eukaryotic phytoplankton. The 18S- and the pigment-based community structure were more similar to each other than to the 16S-based structure. Not surprisingly, these inconsistencies resulted in different sets of major environmental drivers being identified for the datasets obtained using 18S, 16S, and pigment approaches. The discrepancies between the two molecular approaches might be caused by primer bias, different genome sizes and gene copy numbers among phytoplankton groups, and insufficient reference sequences with high taxon coverage in the database. The Chl-*a*-proxied biomass of phytoplankton has also been known to be environment-dependent. Moreover, the predictive accuracy of CHEMTAX is determined by the pigment ratios utilized in the reference matrix, and it has been suggested that CHEMTAX should be calibrated to the assemblages from which samples will be taken (Mackey et al. 1996). To summarize, this study highlights both the advantages and limitations of interpreting molecular and pigment data and suggests that multiple methods be applied to accurately characterize the spatial and temporal variations in the diversity and community structure of phytoplankton.

Materials and methods

Sampling

Two cruises were carried out in the PRE in July and November 2020 (Fig. 1). Water samples were collected from 18 sites. At most sites, water samples at three depth layers, i.e., the surface (5 m), the middle (half of the water depths, ranging from 3 m to 15 m), and the bottom (5 m above the seafloor), were collected using Niskin bottles mounted on a rosette sampler, which was equipped with a conductivity-temperature-depth (CTD) sensor (Sea-Bird SBE 11plus) and an SBE43 dissolved oxygen (DO) sensor (SeaBird, Bellevue, WA, USA). Subsamples were filtered through 0.45 μm cellulose acetate fiber membranes and the filtrate was stored at -20°C until measurements of nutrients.

Determination of physicochemical variables

Water temperature, salinity, pH, and DO concentration were determined *in situ* using the CTD and YSI sensors. The frozen subsamples were thawed at room temperature in laboratory and the concentrations of NO_2^- , NO_3^- , NH_4^+ , soluble reactive phosphorus (SRP) and dissolved silicate (DSi) were measured using a continuous flow analyzer (Seal AA3, Norderstedt, Germany), with analytical precisions at 0.03 μM , 0.03 μM , 0.03 μM , 0.02 μM , and 0.1 μM , respectively.

Pigment analysis

To measure the Chl-*a* content and phytoplankton structure, 0.5 to 1 L of subsamples were prefiltered on board using a 200- μm mesh, then filtered onto 0.7- μm GF/F filters (47-mm in diameter; Whatman, Little Chalfont, Buckinghamshire, UK) under low light and vacuum pressure (< 0.03 MPa) and immediately frozen in liquid nitrogen. Pigment analysis was performed using HPLC (Dionex UltiMate 3000 LC system, Thermo Scientific) (Zhong et al. 2020b). Thirteen diagnostic pigments were used to estimate the relative contributions of five eukaryotic phytoplankton groups [dinoflagellates, diatoms, chlorophytes (including prasinophytes), cryptophytes, haptophytes (a combination of haptophytes_8 and haptophytes_6)] and two prokaryotic groups (*Prochlorococcus* and *Synechococcus*) to the total Chl-*a* using the CHEMTAX program (ver. 1.95) (Mackey et al. 1996). The initial input matrix of ratios of diagnostic pigments to total Chl-*a* followed a previous study of the northern South China Sea (Wang et al. 2015), which included a number of samples from the Pearl River Estuary.

DNA extraction and high throughput sequencing

The prefiltered subsamples (0.3 - 0.5 L) for molecular analysis were filtered onto 0.2- μ m-pore-sized (47 mm in diameter) polycarbonate membrane (Millipore, Carrigtwohill, Cork, Ireland) under low vacuum. The filters with biomass were put into 2-mL cryovials and immediately stored in liquid nitrogen. DNA extraction was conducted using a FastDNA Spin Kit (MP Biomedical, Santa Anna, CA, USA) according to the manufacturer's instructions. Eventually, a total of 52 water (18 surface, 16 middle and 18 bottom) samples were subjected to high throughput sequencing of both nuclear 18S and plastid 16S rRNA genes.

The V4 region of 18S rRNA genes were PCR amplified using the primer set TAREuk454FWD1 (5'-CCAGCASCYGC GGTAATTCC -3') and TAREukREV3 (5'- ACTTTCGTTCTTGATYRA -3') (Stoeck et al. 2010). The highly variable V4-V5 regions of plastid 16S rRNA genes were amplified using the primers 515F (5'- GTGYCAGCMGCCGCGGTAA -3') and 926R (5'- CCGYCAATTYMTTTRAGTTT -3'), which were known also to target the 16S rRNA genes of bacteria including cyanobacteria (Mcnicol et al. 2021; Needham and Fuhrman 2016). The PCR reaction solution of each tube (30 μ L) contained 1 μ L of each primer (10 μ M), and 25 μ L $2\times$ Taq PCR Master Mix, and 2.5 μ L DNA template. The PCR program ran under the following conditions: 95 $^{\circ}$ C for 5 min, then 7 cycles of 95 $^{\circ}$ C for 45 s, 65 $^{\circ}$ C for 1 min (decreasing at 2 $^{\circ}$ C /cycle), and 72 $^{\circ}$ C for 90 s, followed by 30 cycles of 95 $^{\circ}$ C for 45 s, 50 $^{\circ}$ C for 30 s, and 72 $^{\circ}$ C for 95 s, with a final extension step for 10 min at 72 $^{\circ}$ C. The libraries of 16S and 18S rRNA amplicons were sequenced on the Illumina NovaSeq PE250 platform at a commercial company (Novogene, China).

Processing of sequencing data

The sequence data were processed using QIIME2 (Bolyen et al. 2019). ASVs were generated by trimming the raw amplicon sequences of primers using *cutadapt* (Martin 2011) and inputting them into the DADA2 pipeline (ver. 1.8) (Callahan et al. 2016). Reads were filtered with the following parameters: truncLen and trimLen = c(0, 0), truncQ = 2, maxEE = 2, and then the forward and reverse reads was merged using the default parameters (minOverlap = 12, maxMismatch = 0). Chimeras were removed using the *removeBimeraDenovo* command.

For plastid 16S rRNA genes, taxonomy was assigned using the classifier tool implemented in QIIME2 against the PhytoREF database (Decelle et al. 2015). Considering that evolutionary rates of plastid genes were significantly slower than nuclear ones (Wolfe et al. 1987), and the number of plastid 16S reads (on

average 10587 reads per sample) obtained was much lower than that of nuclear 18S (on average 30365 reads per sample), it was decided to define 16S sequence diversity at a finer resolution by clustering the reads into ASVs. The taxonomic identity of these ASVs was also examined using BLASTn searches against the plastid database in NCBI (<https://ftp.ncbi.nlm.nih.gov/refseq/release/plastid>).

For 18S rRNA gene sequences, the qualified reads were clustered into OTUs at 97% identity to minimize inflation of OTU richness due to intragenomic polymorphisms, as suggested by (Zou et al. 2021). The obtained OTUs were classified using the Protist Ribosomal Reference database (PR2, ver. 4.14) (Guillou et al. 2012) and SILVA (ver. 138) (Quast et al. 2012). For diversity and community structure of eukaryotic phytoplankton communities, the macroalgae (Rhodophyta, Streptophyta and Ulvophyceae) and non-pigmented dinoflagellates, Cercozoa, Ciliophora, Mesomycetozoa, Radiolaria and Fungi were discarded, and only the OTUs of the photosynthetic groups, i.e., pigmented dinophytes (including Gymnodiniales, Peridiniales, Gonyaulacales, Suessiales, Prorocentrale and Dinophysiales), Bacillariophyta, Chlorophyta, Cryptophyta, Haptophyta, and Ochrophyta were retained for subsequent analyses. The reads generated from MiSeq sequencing of 18S and plastid 16S rRNA genes have been deposited in the NCBI database (accession numbers: PRJNA904229 and PRJNA904929).

Quantitative real-time PCR (qPCR)

The qPCR assays were performed as previously described (Gong et al. 2013), with some modifications. Since the primers for plastid 16S rRNA genes also target heterotrophic bacteria, another set of primers was selected that specifically targeted the plastid 23S rRNA gene of eukaryotic algae (Kang et al. 2018). These two plastid genes are thought to have identical copy numbers in a chloroplastid genome. The plasmid standards (18S rRNA gene of *Thalassiosira* sp. and plastid 23S rRNA gene of *Synechococcus*) were serially diluted in eight 10-fold dilutions. The PCR reaction mixture (20 µL) contained 10 µL Master, 1 µL each primer, 2 µL DNA template and 6 µL double-distilled water. The primer sets 345F (5'-AAGGAAGGCAGCAGGCG-3') and 499R (5'- CACCAGACTTGCCCTCYAAT -3' (Zhu et al. 2005), and P23MISQF1 (5'- GGACARWAAGACCCTATGMAG -3') and P23MISR1 (5'- AGATYAGCCTGTTATCCCT -3' (Kang et al. 2018) were used for amplifying 18S and 23S rRNA gene abundances of phytoplankton, respectively. The qPCR assay was based on the fluorescence intensity of the SYBR Green dye, and reactions for each sample were carried out in a Roche LightCycler 96 System (Roche

Diagnostics, Mannheim, Germany). The cycling conditions were programmed for 18S rRNA genes as follows: an initial denaturation step at 95°C for 7 min, followed by 40 cycles at 95°C for 15 s, 60°C for 60 s and 77°C for 25 s. A dissociation curve was examined at 95°C for 15 s, 60°C for 60 s and 97°C for 1 s, to ensure that the target sequences being specifically amplified. The thermal cycling for amplifying plastid 23S rRNA genes consisted of 6 min at 94°C, followed by 40 cycles at 94°C for 10 s, 55°C for 20 s, with a final step at 72°C for 20 s. The dissociation curve was verified at 95°C for 10 s, 65°C for 60 s and 97°C for 1 s. The ‘absolute abundance’ (i.e., rRNA gene copy number per liter of seawater) of an individual taxon of microbial eukaryotes was calculated as follows:

$$\text{Absolute gene abundance of a taxon} = CN_{total} * RP$$

where CN_{total} was the total copy number of 18S (or plastid 16S) rRNA genes determined using qPCR for the whole community; and RP represented the relative proportion of the taxon revealed by high throughput sequencing of 18S (or plastid 16S) rRNA genes.

Statistical analysis

The vertical profiles of environmental factors, rRNA gene abundances, concentrations of nutrients, major phytoplankton groups in pigments, and alpha diversity estimators (OTU or ASV richness, Shannon, Simpson and Chao1) were generated using Ocean Data View (Brown 1998). Alpha diversity indices were computed using *vegan* in R (ver. 4.1.3) (Dixon 2003).

Regression analysis was conducted to explore the correlations between the diversity indices of pigmented eukaryotes and environmental variables using the function *stats* in R. Pearson correlations were performed to test the hypothesis that the relative abundance of a major microalgal group in the phytoplankton community showed a similar variational pattern in applications of all three methods (nuclear 18S rRNA gene, plastid 16S rRNA gene and pigment analysis). Non-metric multidimensional scaling (NMDS) was based on Bray-Curtis dissimilarities of the phytoplankton community structure and performed using the package *vegan* in R. Analysis of similarity (ANOSIM) was conducted to statistically test the hypothesis that the community structure of phytoplankton based on variations in the five major eukaryotic groups (diatoms, dinoflagellates, chlorophytes, haptophytes and cryptophytes) was independent of the three methodologies. Principal components analysis (PCA) was performed to visualize the differences in phytoplankton community structure generated using the three methods. The relationships between

community structure and environmental factors were explored using canonical correspondence analysis (CCA). All statistical analyses were carried out using the packages in R (ver. 4.1.3).

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Nos. 41976128 and 31970486), the Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (No. 311021004).

Authors' Contributions

JG designed the experiment. SX and YH conducted sampling. SX, CH, DY, HD, ZT, YW and BH performed laboratory work. SX, ZS and GL analyzed the data. SX and JG wrote the manuscript. All authors approved the final version of manuscript.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflicts of interest. Author Jun Gong is one of the Editorial Board Members, but he was not involved in the journal's review of, or decision related to, this manuscript.

Animal and human rights statement No animal and human rights are involved in this article.

References

- Abaychi J, Riley J (1979) The determination of phytoplankton pigments by high-performance liquid chromatography. *Anal Chim Acta* 107:1-11
- Belevich T, Milyutina I (2022) Species diversity of phototrophic picoplankton in the Kara and Laptev Seas. *Microbiology* 91:67-76
- Bennke CM, Pollehne F, Müller A, Hansen R, Kreikemeyer B, Labrenz M (2018) The distribution of phytoplankton in the Baltic Sea assessed by a prokaryotic 16S rRNA gene primer system. *J Plankton Res* 40:244-254
- Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodriguez AM, Chase J, Cope EK et al. (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 37:852-857
- Brown M (1998) Ocean data view 4.0. *Oceanography* 30:663-671
- Callahan BJ, Mcmurdie PJ, Rosen MJ, Han AW, Johnson A, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581-583
- Chai C, Jiang T, Cen J, Ge W, Lu S (2016) Phytoplankton pigments and functional community structure in relation to environmental factors in the Pearl River Estuary. *Oceanologia* 58:201-211
- Chen H, Li DH, Jiang AJ, Li XG, Wu SJ, Chen JW, Qu MJ, Qi XQ, Dai J, Zhao R, Zhang WJ, Liu SS, Wu LF (2022) Metagenomic analysis reveals wide distribution of phototrophic bacteria in hydrothermal vents on the ultraslow-spreading Southwest Indian Ridge. *Mar Life Sci Technol* 4:255-267
- Chen ZF, Zhang QC, Kong FZ, Liu Y, Zhao Y, Zhou ZX, Geng HX, Dai L, Zhou MJ, Yu RC (2019) Resolving phytoplankton taxa based on high-throughput sequencing during brown tides in the Bohai Sea, China. *Harmful Algae* 84:127-138
- Claustre H, Hooker SB, Van Heukelem L, Berthon J-F, Barlow R, Ras J, Sessions H, Targa C, Thomas CS, van der Linde D (2004) An intercomparison of HPLC phytoplankton pigment methods using in situ samples: Application to remote sensing and database activities. *Mar Chem* 85:41-61
- Decelle J, Romac S, Stern RF, Bendif el M, Zingone A, Audic S, Guiry MD, Guillou L, Tessier D, Le Gall F, Gourvil P, Dos Santos AL, Probert I, Vaulot D, de Vargas C, Christen R (2015) PhytoREF: a reference database of the plastidial 16S rRNA gene of photosynthetic eukaryotes with curated taxonomy. *Mol Ecol Resour* 15:1435-1445
- Dixon P (2003) VEGAN, a package of R functions for community ecology. *J Veg Sci* 14:927-930
- Drouin G, Daoud H, Xia J (2008) Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol Phylogeny Evol* 49:827-831
- Eker-Develi E, Berthon J-F, Canuti E, Slabakova N, Moncheva S, Shtereva G, Dzhirova B (2012) Phytoplankton taxonomy based on CHEMTAX and microscopy in the northwestern Black Sea. *J Mar Syst* 94:18-32
- Everitt D, Wright S, Volkman J, Thomas D, Lindstrom E (1990) Phytoplankton community compositions in the western equatorial Pacific determined from chlorophyll and carotenoid pigment distributions. *Deep-sea Res Pt A* 37:975-997
- Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O (2004) The evolution of modern eukaryotic phytoplankton. *Science* 305:354-360
- Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281:237-240
- Forsythe ES, Williams AM, Sloan DB (2020) Genome-wide signatures of plastid-nuclear coevolution point to repeated perturbations of plastid proteostasis systems across angiosperms. *Plant Cell* 33:980-997
- Fu R, Gong J (2017) Single cell analysis linking ribosomal (r)DNA and rRNA copy numbers to cell size and growth rate provides insights into molecular protistan ecology. *J Eukaryot Microbiol* 64:885-896

- Gao K, Gao G, Wang Y, Dupont S (2020) Impacts of ocean acidification under multiple stressors on typical organisms and ecological processes. *Mar Life Sci Technol* 2:279-291
- Geider RJ (1987) Light and temperature dependence of the carbon to chlorophyll a ratio in microalgae and cyanobacteria: implications for physiology and growth of phytoplankton. *New Phytol* 106:1-34
- Geider RJ, Platt T, Raven JA (1986) Size dependence of growth and photosynthesis in diatoms: a synthesis. *Mar Ecol Prog Ser* 30:93-104
- Godhe A, Asplund ME, Härnström K, Saravanan V, Tyagi A, Karunasagar I (2008) Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR. *Appl Environ Microbiol* 74:7174-7182
- Gong F, Li G, Wang Y, Liu Q, Huang F, Yin K, Gong J, Dolan J (2020a) Spatial shifts in size structure, phylogenetic diversity, community composition and abundance of small eukaryotic plankton in a coastal upwelling area of the northern South China Sea. *J Plankton Res* 42:650-667
- Gong J, Dong J, Liu X, Massana R (2013) Extremely high copy numbers and polymorphisms of the rDNA operon estimated from single cell analysis of oligotrich and peritrich ciliates. *Protist* 164:369-379
- Gong W, Hall N, Paerl H, Marchetti A (2020b) Phytoplankton composition in a eutrophic estuary: Comparison of multiple taxonomic approaches and influence of environmental factors. *Environ Microbiol* 22:4718-4731
- Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, de Vargas C, Decelle J, del Campo J, Dolan JR, Dunthorn M, Edvardsen B, Holzmann M, Kooistra WHCF, Lara E, Le Bescot N, Logares R, Mahé F et al. (2012) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res* 41:D597-604
- Guo L, Sui Z, Liu Y (2016) Quantitative analysis of dinoflagellates and diatoms community via Miseq sequencing of actin gene and v9 region of 18S rDNA. *Sci Rep* 6:1-9
- Hansen PJ, Anderson R, Stoecker DK, Decelle J, Altenburger A, Blossom HE, Drumm K, Mitra A, Flynn KJ (2019) Mixotrophy among freshwater and marine protists. *Encyclopedia of Microbiology (Fourth Edition)*:Elsevier, pp. 199-210
- Heidenreich KM, Richardson TL (2020) Photopigment, absorption, and growth responses of marine cryptophytes to varying spectral irradiance. *J Phycol* 56:507-520
- Huang H, Yang J, Huang S, Gu B, Xu D (2021) Spatial distribution of planktonic ciliates in the western Pacific Ocean: along the transect from Shenzhen (China) to Pohnpei (Micronesia). *Mar Life Sci Technol* 3:103-115
- Jia H, Chen M, Su W, Zhang S, Zhao K (2019) Structural characteristics and associated factors influencing phytoplankton abundance and species composition in Huangmaohai Bay, Pearl River Estuary. *J Coast Res* 35:72-81
- Jiang ZY, Wang YS, Cheng H, Sun CC, Wu ML (2015) Variation of phytoplankton community structure from the Pearl River estuary to South China Sea. *Ecotoxicology* 24:1442-1449
- Kana TM, Geider RJ, Critchley C (1997) Regulation of photosynthetic pigments in micro-algae by multiple environmental factors: a dynamic balance hypothesis. *New Phytol* 137:629-638
- Kang HE, Yoon TH, Yoon S, Kim HJ, Park H, Kang CK, Kim HW (2018) Genomic analysis of red-tide water bloomed with *Heterosigma akashiwo* in Geoje. *PeerJ* 6:e4854
- Kirkham AR, Jardillier LE, Tiganescu A, Pearman J, Zubkov MV, Scanlan DJ (2011) Basin-scale distribution patterns of photosynthetic picoeukaryotes along an Atlantic meridional transect. *Environ Microbiol* 13:975-990
- Lin S (2011) Genomic understanding of dinoflagellates. *Res Microbiol* 162:551-569
- Lin Y, Gifford S, Ducklow H, Schofield O, Cassar N (2018) Towards quantitative microbiome community profiling using internal standards. *Appl Environ Microbiol* 85:e02634-18
- Liu Q, Zhao Q, McMinn A, Yang EJ, Jiang Y (2021) Planktonic microbial eukaryotes in polar surface waters: recent

advances in high-throughput sequencing. *Mar Life Sci Technol* 3:94-102

- Lu Z, Gan J (2015) Controls of seasonal variability of phytoplankton blooms in the Pearl River Estuary. *Deep-sea Res Pt II* 117:86-96
- Mackey M, Mackey DJ, Higgins HW, Wright SW (1996) CHEMTAX - A program for estimating class abundances from chemical markers: Application to HPLC measurements of phytoplankton. *Mar Ecol Prog Ser* 144:265-283
- Maki A, Salmi P, Mikkonen A, Kremp A, Tirola M (2017) Sample preservation, DNA or RNA extraction and data analysis for high-throughput phytoplankton community sequencing. *Front Microbiol* 8:1848
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17:10-12
- McNichol J, Berube PM, Biller SJ, Fuhrman JA (2021) Evaluating and improving SSU rRNA PCR primer coverage for bacteria, archaea, and eukaryotes using metagenomes from global ocean surveys. *mSystems* 6:e0056521
- Needham DM, Fuhrman JA (2016) Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nat Microbiol* 1:16005
- Niu L, Luo X, Hu S, Liu F, Cai H, Ren L, Ou S, Zeng D, Yang Q (2020) Impact of anthropogenic forcing on the environmental controls of phytoplankton dynamics between 1974 and 2017 in the Pearl River estuary, China. *Ecol Indic* 116:106484
- Qiu D, Zhong Y, Chen Y, Tan Y, Song X, Huang L (2019) Short-term phytoplankton dynamics during typhoon season in and near the Pearl River Estuary, South China Sea. *J Geophys Res Biogeosci* 124:274-292
- Quast C, Pruesse E, Yilmaz P, Gerken J, Glckner FO (2012) The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res* 41:D590-596
- Rii YM, Bidigare RR, Church MJ (2018) Differential responses of eukaryotic phytoplankton to nitrogenous nutrients in the North Pacific Subtropical Gyre. *Front Mar Sci* 5:92
- Salmaso N, Boscaini A, Pindo M (2020) Unraveling the diversity of eukaryotic microplankton in a large and deep perialpine lake using a high throughput sequencing approach. *Front Microbiol* 11:789
- Sathyendranath S, Cota G, Stuart V, Maass H, Platt T (2001) Remote sensing of phytoplankton pigments: A comparison of empirical and theoretical approaches. *Int J Remote Sens* 22:249-273
- Sathyendranath S, Stuart V, Nair A, Oka K, Nakane T, Bouman H, Forget M-H, Maass H, Platt T (2009) Carbon-to-chlorophyll ratio and growth rate of phytoplankton in the sea. *Mar Ecol Prog Ser* 383:73-84
- Shi XL, Lepere C, Scanlan DJ, Vault D (2011) Plastid 16S rRNA gene diversity among eukaryotic picophytoplankton sorted by flow cytometry from the South Pacific Ocean. *PLoS ONE* 6:e18979
- Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner HW, Richards TA (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol Ecol* 19:21-31
- Taylor AH, Geider RJ, Gilbert FJ (1997) Seasonal and latitudinal dependencies of phytoplankton carbon-to-chlorophyll a ratios: results of a modelling study. *Mar Ecol Prog Ser* 152:51-66
- Tragin M, Zingone A, Vault D (2018) Comparison of coastal phytoplankton composition estimated from the V4 and V9 regions of the 18S rRNA gene with a focus on photosynthetic groups and especially Chlorophyta. *Environ Microbiol* 20:506-520
- Trefault N, De la Iglesia R, Moreno-Pino M, Lopes Dos Santos A, Gerikas Ribeiro C, Parada-Pozo G, Cristi A, Marie D, Vault D (2021) Annual phytoplankton dynamics in coastal waters from Fildes Bay, Western Antarctic Peninsula. *Sci Rep* 11:1368
- Treusch AH, Demir-Hilton E, Vergin KL, Worden AZ, Carlson CA, Donatz MG, Burton RM, Giovannoni SJ (2012) Phytoplankton distribution patterns in the northwestern Sargasso Sea revealed by small subunit rRNA genes from plastids. *ISME J* 6:481-492
- Veldhuis MJ, Cucci TL, Sieracki ME (1997) Cellular DNA content of marine phytoplankton using two new

fluorochromes: Taxonomic and ecological implications 1. *J Phycol* 33:527-541

- Wang L, Huang B, Liu X, Xiao W (2015) The modification and optimizing of the CHEMTAX running in the South China Sea. *Acta Oceanol Sin* 34:124-131
- Wang Z, Chi Y, Li T, Song W, Wang Y, Wu T, Zhang G, Liu Y, Ma H, Song W (2022) Biodiversity of freshwater ciliates (Protista, Ciliophora) in the Lake Weishan Wetland, China: the state of the art. *Mar Life Sci Technol* 4:429-451
- Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci USA* 84:9054-9058
- Wright S, Jeffrey S, Mantoura R, Llewellyn C, Bjørnland T, Repeta D, Welschmeyer N (1991) Improved HPLC method for the analysis of chlorophylls and carotenoids from marine phytoplankton. *Mar Ecol Prog Ser*:183-196
- Wu W, Liu H (2018) Disentangling protist communities identified from DNA and RNA surveys in the Pearl River - South China Sea continuum during the wet and dry seasons. *Mol Ecol* 27:4627-4640
- Xie ZX, Yan KQ, Kong LF, Gai YB, Jin T, He YB, Wang YY, Chen F, Lin L, Lin ZL, Xu HK, Shao ZZ, Liu SQ, Wang DZ (2022) Metabolic tuning of a stable microbial community in the surface oligotrophic Indian Ocean revealed by integrated meta-omics. *Mar Life Sci Technol* 4:277-290
- Xu S, Liu Y, Fan J, Xiao Y, Qi Z, Lakshmikandan M (2022) Impact of salinity variation and silicate distribution on phytoplankton community composition in Pearl River estuary, China. *Ecohydrol Hydrobiol* 22:466-475
- Yang S, Cui Z, Zhang Y, Jiang T, Yang Q, Sun Y (2019) Photosynthetic pigments in surface sediments in the northwest of the Bohai Sea, China: Potential implications for sediment deposition of brown tides of *Aureococcus anophagefferens* in coastal waters. *Ecol Indic* 102:145-153
- Ye T, Jiang Y, Chen S, Xu Y, Li L, Shin MK, Chen X (2022) The widely reported but poorly studied ciliate family Folliculinidae (Protozoa, Ciliophora, Heterotrichea): a revision with notes on its taxonomy, morphology and phylogenetic relationships. *Mar Life Sci Technol* 4:471-492
- Yeh Y-C, Fuhrman JA (2022) Contrasting diversity patterns of prokaryotes and protists over time and depth at the San-Pedro Ocean Time series. *ISME Commun* 2:1-12
- Zhong Q, Xue B, Noman MA, Wei Y, Liu H, Liu H, Zheng L, Jing H, Sun J (2020a) Effect of river plume on phytoplankton community structure in Zhujiang River estuary. *J Oceanol Limnol* 39:550-565
- Zhong Y, Liu X, Xiao W, Laws EA, Chen J, Wang L, Liu S, Zhang F, Huang B (2020b) Phytoplankton community patterns in the Taiwan Strait match the characteristics of their realized niches. *Prog Oceanogr* 186:102366
- Zhu F, Massana R, Not F, Marie D, Vaulot D (2005) Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol Ecol* 52:79-92
- Zou S, Fu R, Zhang Q, Gentekaki E, Gong J (2021) Coupling between ribotypic and phenotypic traits of protists across life-cycle stages and temperatures. *Microbiol Spectr* 9:e0173821
- Hamilton M, Mascioni M, Hehenberger E, Bachy C, Yung C, Vernet M, Worden AZ (2021) Spatiotemporal variations in Antarctic protistan communities highlight phytoplankton diversity and seasonal dominance by a novel cryptophyte lineage. *mBio* 12:e02973-21.
- Dong Y, Cui L, Cao R, Cen J, Zou J, Zhou X, Lu S (2020) Ecological characteristics and teratogenic retinal determination of *Cochlodinium geminatum* blooms in Pearl River Estuary, South China. *Ecotoxicol Environ Saf* 191:110226

Figure legends

Fig. 1 A map showing sampling sites in Transects 1 to 3 in the Pearl River Estuary, South China Sea. Twelve sites (depicted in blue) and 6 sites (in green) were visited in July and November, respectively. The color bar shows the water depths.

Fig. 2 Vertical profiles of temperature (**A - D**), salinity (**E - H**), total chlorophyll *a* (Chl-*a*, **I - L**), nuclear 18S rRNA gene (**M - P**), and plastid 23S rRNA gene abundances (**Q - T**) of eukaryotic phytoplankton along three summertime transects (including sites S1-S13, S41 - S45, and S65 - S61) and a wintertime transect (including sites A3 - A15) in the Pearl River Estuary.

Fig. 3 Vertical distribution of chlorophyll *a* content contributed by the five major phytoplankton groups, pigmented dinoflagellates (**A - D**), chlorophytes (**E - H**), diatoms (**I - L**), cryptophytes (**M - P**), and haptophytes (**Q - T**).

Fig. 4 Regression analysis showing significant relationships between ratio of 18S rRNA gene abundance ($\log(x+1)$ transformed) to Chl-*a* and other environmental factors. Abbreviations: Baci, diatoms; Chl-*a*, chlorophyll *a*; Chlo, chlorophytes; Cryp, cryptophytes; Dino, pigmented dinoflagellates; DSi, dissolved silicate; Hapt, haptophytes; N.A., not applicable.

Fig. 5 Spatial and seasonal variations in 18S-based OTU richness (**A - D**) and 16S-based ASV richness (**E - H**) of eukaryotic phytoplankton along the three transects. Abbreviations: ASV, amplicon sequence variant; OTU, operational taxonomic unit.

Fig. 6 The relationships between 18S-based OTU richness and the identified environmental factors that were most significant in the regression analyses (**A - D**). Also note that the weak correlations between 18S-based OTU and 16S-based ASV richness of the whole eukaryotic phytoplankton communities (**E**), and of individual phytoplankton groups (**F - J**). Abbreviations: ASV, amplicon sequence variant; OTU, operational taxonomic unit; SRP, soluble reactive phosphorus.

Fig. 7 Community composition of eukaryotic phytoplankton in surface (left panel), middle layer (middle panel) and bottom waters (right panel) in the Pearl River Estuary, as revealed by high throughput sequencing of nuclear 18S rRNA genes (**A**), plastid 16S rRNA genes (**B**), and pigment analysis (**C**). Arrows and asterisks indicate the samples where hypoxia and blooms of pigmented dinophytes occurred.

Fig. 8 Scatter plots showing linear regressions between ratio of 18S rRNA gene abundance of two major microalgal groups and ratio of pigment contents of that microalgal pair. Abbreviations: Dino, pigmented dinophytes; Baci, Bacillariophyta; Chlo, Chlorophyta; Cryp, Cryptophyta; Hapt, Haptophyta.

Fig. 9 Variations in phytoplankton community composition as characterized by using nuclear 18S, plastid 16S, and pigment-based methods. A non-metric multidimensional scaling (NMDS) plot showing the variations in phytoplankton community structure resolved at the phylum level (**A**). A plot of principal component analysis showing that Haptophyta was underestimated in the 18S dataset, whereas diatoms and cryptophytes underestimated in the plastid 16S dataset (**B**). Scatter plots showing Pearson correlations between phytoplankton community distances based on 18S, plastid 16S, and pigments (**C - E**).

Fig. 10 Plots of canonical correspondence analysis (CCA) showing the environmental factors that significantly varied with the community structure of eukaryotic phytoplankton (**A - C**), and with relative abundances of major phytoplankton groups. The community structure was based on 18S (**A** and **D**), 16S rRNA genes (**B** and **E**), and pigments (**C** and **F**), and resolved at the OTU (**A**) and ASV (**B**) levels, and at major group levels (**C - E**). Note that different sets of environmental factors were selected for the phytoplankton communities characterized by using these methods, nevertheless, more similar sets of environmental variables were selected in those characterized by 18S sequencing and pigment analysis. The water depths (surface, middle, and bottom layers) and months (July and November) of the collected samples are annotated in (**A - C**). Abbreviations: Baci, Bacillariophyta; Chl-*a*, chlorophyll *a*; Chlo, Chlorophyta; Cryp, Cryptophyta; Dino, pigmented dinophytes; DIN, dissolved inorganic nitrogen; DSi, dissolved silicate; SRP, soluble reactive phosphorus; Hapt, Haptophyta.

Supplementary Figures

Fig. S1 Vertical profiles of concentrations of dissolved oxygen (DO), nitrate (NO_3^-), nitrite (NO_2^-), ammonium (NH_4^+), soluble reactive phosphorus (SRP), and dissolved silicate (DSi) during July (three columns on the left) and November (the rightmost column)

Fig. S2 Correlations among 18S rRNA gene abundance (copy number/L), plastid 23S rRNA gene abundance (copy number/L), and pigment (chlorophyll *a*) content ($\mu\text{g/L}$) of the pigmented dinophytes, chlorophytes, diatoms, cryptophytes, and haptophytes. All gene abundances and pigment contents are $\log_{10}(x+1)$ transformed

Fig. S3 Distribution of ratio of $\log_{10}(18\text{S rRNA gene abundance} + 1)$ to chlorophyll *a* in the pigmented dinophytes, diatoms, chlorophytes, and cryptophytes. Units of gene abundance and pigment content are copy number/L and $\mu\text{g/L}$, respectively

Fig. S4 Regression analysis between alpha diversity estimators (Shannon, Chao1 and Simpson indices) of 18S OTUs and 16S ASVs and environmental factors. Only the two most significant environmental factors for each index are shown. Abbreviations: Chl-*a*, chlorophyll *a*; DO, dissolved oxygen; DSi, dissolved silicate; SRP, soluble reactive phosphorus

Fig. S5 Correlations between of alpha diversity estimators (Chao1, Simpson and Shannon) of 18S and 16S of the total eukaryotic phytoplankton community

Fig. S6 Correlations between of alpha diversity estimators (Chao1, Simpson and Shannon) of 18S and 16S of pigmented dinophytes, diatoms, chlorophytes, cryptophytes and haptophytes

Fig. S7 Variations in community composition of eukaryotic phytoplankton in the surface waters. Reads of taxa are binned at order levels. The classification of 18S (left and middle panels) by PR2 yielded almost identical community composition; however the classification of plastid 16S rRNA genes by PhytoRef (right

panel) showed different community composition, no matter what resolutions (OTU, defined at a 97% similarity cutoff, and ASV, defined at a 100% similarity cutoff) were applied. N.D., not detected.

Fig. S8 Correlations between ratio of 18S rRNA gene abundance of two microalgal groups and ratio of 16S rRNA gene abundance of those two groups (**A**), and between 16S rRNA gene abundance of two microalgal groups and pigment contents of the taxon pairs (**B**). Note that both the gene abundance and pigment content were $\log_{10}(x+1)$ transformed. Abbreviations: Dino, pigmented dinophytes; Baci, Bacillariophyta; Chlo, Chlorophyta; Cryp, Cryptophyta; Hapt, Haptophyta.