Environmental Research 243 (2024) 117708

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

Environmental DNA metabarcoding reveals the effect of environmental selection on phytoplankton community structure along a subtropical river

Luping Zeng [a], Jing Wen [a], Bangjie Huang [a], Yang Yang [b], Zhiwei Huang [a], Fantang Zeng [a], Huaiyang Fang [a,**], Hongwei Du [a,*]

[a] The Key Laboratory of Water and Air Pollution Control of Guangdong Province, South China Institute of Environmental Sciences, Ministry of Ecology and Environment, No. 18 Ruihe Road, Guangzhou, 510530, China
[b] Research Center of Hydrobiology, Department of Ecology, Jinan University, Guangzhou, 510632, China

ARTICLE INFO

Keywords:
Phytoplankton diversity
eDNA metabarcoding
Environmental heterogeneity
Spatial pattern
Community assembly

ABSTRACT

The Dongjiang River, a major tributary of the Pearl River system that supplies water to more than 40 million people in Guangdong Province and neighboring regions of China, harbors rich biodiversity, including many endemic and endangered species. However, human activities such as urbanization, agriculture, and industrialization have posed serious threats to its water quality and biodiversity. To assess the status and drivers of phytoplankton diversity, which is a key indicator of aquatic ecosystem health, this study used Environmental DNA (eDNA) metabarcoding combined with machine learning methods to explore spatial variations in the composition and structure of phytoplankton communities along the Dongjiang River, including its estuary. The results showed that phytoplankton diversity exhibited spatial distribution patterns, with higher community structure similarity and lower network complexity in the upstream than in the downstream regions. Environmental selection was the main mechanism shaping phytoplankton community composition, with natural factors driving the dominance of Pyrrophyta, Ochrophyta, and Cryptophyta in the upstream regions and estuaries. In contrast, the downstream regions was influenced by high concentrations of pollutants, resulting in increased abundance of Cryptophyta. The random forest model identified temperature, dissolved oxygen, chlorophyll *a*, NO_2^-, and NH_4^+ as the main factors influencing the primary phytoplankton communities and could be used to predict changes during wet periods. This study provides valuable insights into the factors influencing phytoplankton diversity and community composition in the Dongjiang River, and demonstrates the application value of eDNA metabarcoding technique in large-scale, long-distance river biodiversity monitoring.

1. Introduction

Phytoplankton are important primary producers and the base of the food web in aquatic ecosystems. Their diversity and community composition directly affect water quality and biological resources (Zhang et al., 2022, 2023b). Traditional phytoplankton monitoring methods rely mainly on morphological identification, which requires high levels of technical expertise, is time-consuming and labor-intensive, and prone to sample loss and identification errors (Pawlowski et al., 2018). In recent years, environmental DNA (eDNA) metabarcoding has emerged as a new method for biodiversity monitoring, which extracts and detects phytoplankton DNA directly from environmental samples (Djurhuus et al., 2020; Ji et al., 2021; Zhang, 2019). This method overcomes the limitations of traditional methods, and can achieve comprehensive, fast, and non-invasive assessment of phytoplankton diversity in aquatic ecosystems (Cordier et al., 2021; Jarman et al., 2018; Van Driessche et al., 2023). eDNA metabarcoding has been rapidly developed and widely applied in recent years, especially in aquatic ecology, showing great potential and advantages (Barnes et al., 2014; Clark et al., 2020). Previous research has been conducted on phytoplankton diversity assessment using the eDNA metabarcoding technique (Diao et al., 2022; Lambert et al., 2019; Xue et al., 2018), mainly focusing on lake reservoirs, portions of rivers, or specific tributaries, with limited research at the spatial scale of the entire river. Additionally, current research still faces challenges discerning species distributions and biodiversity patterns in complex aquatic

---

* Corresponding author. South China Institute of Environmental Sciences, Ministry of Ecology and Environment, China.
** Corresponding author.South China Institute of Environmental Sciences, Ministry of Ecology and Environment, China.
E-mail addresses: fanghuaiyang@scies.org (H. Fang), duhongwei@scies.org (H. Du).

https://doi.org/10.1016/j.envres.2023.117708
Received 13 August 2023; Received in revised form 8 November 2023; Accepted 15 November 2023

environments (Datry et al., 2016; Zou et al., 2020).

In recent years, river ecosystems have been affected by human activities and global climate change, resulting in habitat degradation and biodiversity loss (Sun et al., 2023). Correspondingly, river aquatic organisms exhibit specific assembly patterns in response to environmental changes. Two important and complementary mechanisms are generally considered for community assembly, namely niche-based and neutral theories (Chase, 2010; Rosindell et al., 2011). Niche theory assumes that the formation of community structure is determined by deterministic processes, including biotic and abiotic factors. Species rely on their niches to acquire resources, avoid predators, and survive in various adverse environmental conditions (Chase and Myers, 2011). Neutral theory assumes that community structure is formed by random processes such as birth, death, immigration, speciation, and limited dispersal (Chen et al., 2019b; Qian et al., 2023). Environmental selection and passive dispersal are two major ecological processes that determine the community assembly of aquatic ecosystems (Hanson et al., 2012). Understanding the potential mechanisms underlying the phytoplankton community assembly can help improve the management of aquatic ecosystem functions, but this is not well understood in large-scale river ecosystems.

The Dongjiang River supplies water to Guangdong Province, China, but its water quality and biodiversity are under threat from human activities (Lv et al., 2022). Additionally, the Dongjiang River has complex geographical and climatic features, resulting in obvious heterogeneity of its aquatic environment. This environmental heterogeneity may have important impacts on phytoplankton community assembly, but there is a lack of systematic and in-depth research on this topic. Therefore, this study used eDNA metabarcoding combined with machine learning methods to investigate the spatial variation in phytoplankton community composition and structure along the Dongjiang River, including its estuary. Our research objectives are: (1) to reveal the spatial distribution patterns and trends of phytoplankton diversity in the Dongjiang River; (2) to explore the main environmental factors and mechanisms affecting phytoplankton community composition and structure; (3) to assess the degree and direction of environmental selection on phytoplankton community assembly; (4) to establish a random forest model to predict the effects of different environmental factors on phytoplankton dominant groups. This study has important implications for understanding and protecting the aquatic ecology of the Dongjiang River system, as well as providing new methods and perspectives for studying phytoplankton community assembly and environmental selection in other aquatic environments, and improving management strategies in these systems.

2. Materials and methods

2.1. Study area, sample collection and analysis

The Dongjiang River is a critical water source for the Guangdong-Hong Kong-Macao Greater Bay Area, one of the most densely populated and economically developed regions in China. The river's land use changes from natural vegetation in the upstream to urban areas in the downstream, where various industrial activities exert pressure on river ecosystems and planktonic diversity (Lv et al., 2022; Zhang et al., 2023a, 2023b; Zou et al., 2021). Dongjiang River is approximately 435 km long in Guangdong Province, and we sampled 19 sites along its length, with distances ranging from 8 to 27 km between locations. These locations were chosen to be representative of the different sections of the river, including upstream (UR) (S1–S4), midstream (MR) (S5–S11), and downstream regions (DR) (S12–S17), as well as the Pearl River Estuary (PRE) (S18–S19) (Fig. 1). We sampled two locations in the PRE, one in the northern part of the estuary and one in the southern part. They have different environmental conditions and may influence the composition and structure of phytoplankton communities. It should be noted that our findings may not be fully representative of the entire PRE, and further research is needed to fully understand the phytoplankton communities in this region. Detailed information about each sampling site is shown in

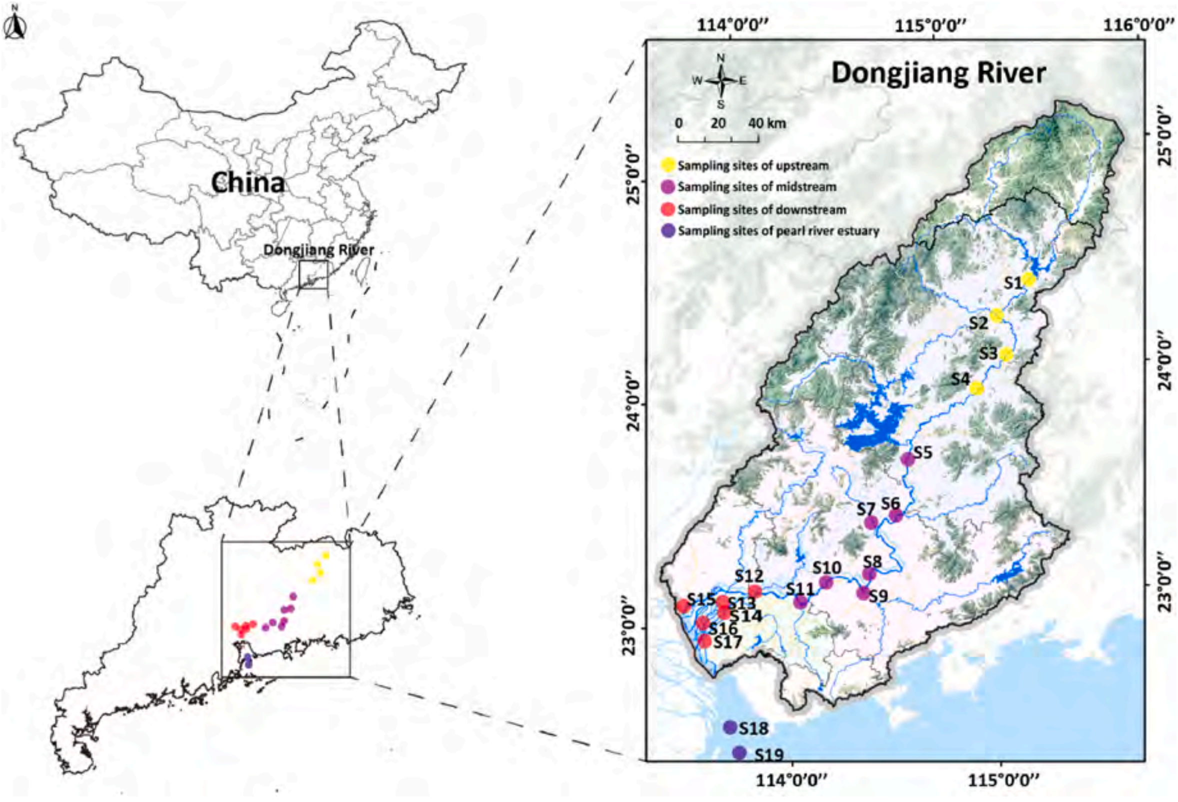

Fig. 1. Distribution of sampling sites.

Table S1, including the site name, location coordinates, location city and river section. Water samples were collected from 19 sites in August 2022.

Water temperature (Temp), pH, and dissolved oxygen (DO) were measured onsite using a YSI multiparameter water quality analyzer (YSI Incorporated, USA). A 2-L water sample was collected at each site and the pH of each water sample was adjusted to 3.0 with 4 M H_2SO_4 before being stored at 4 °C. The concentrations of total nitrogen (TN), total phosphorus (TP), ammonia nitrogen (NH_4^+-N), nitrate nitrogen (NO_3^-), nitrite nitrogen (NO_2^-), chlorophyll-a (Chl-a), total organic carbon (TOC), chemical oxygen demand (COD), biochemical oxygen demand (BOD5), chloride ion (Cl^-), and sulfate ion (SO_4^{2-}) were measured within 48 h according to the Chinese national standard method (SEPA, 2002).

2.2. eDNA collection and analysis

Three duplicate 1-L surface water samples were collected at each site from a depth of <1.5 m using sterile bottles. After collection, the water samples were kept in a low-temperature cooler and transported to the laboratory within 24 h. The samples were vacuum filtered through a 47-mm, 0.22-μm sterile microporous mixed cellulose membrane within 6 h and packed into 2-mL centrifuge tubes, which were then immediately stored at −80 °C until DNA extraction. The DNA samples were then sent to Shanghai BIOZERON Co., Ltd (Shanghai, China) for high-throughput sequencing on the Illumina MiSeq platform, and the sequencing depth for this study was 500,000 reads per sample.

2.3. DNA extraction, polymerase chain reaction (PCR) amplification, sequencing and bioinformatics

eDNA was extracted from the samples using the E. Z.N.A.® Soil DNA Kit (Omega Bio-tek, Norcross, GA, U.S.) according to the manufacturer's protocols. The V4 region of the 18S ribosomal RNA gene was amplified by PCR (95 °C for 4 min, followed by 35 cycles at 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 60 s, with a final extension at 72 °C for 10 min) using the primers TAReuk454FWD1-TAReukREV3 (TAReuk454FWD1: 5′-CCAGCASCYGCGGTAATTCC-3′, TAReukREV3: 5′-ACTTTCGTTGA-TYRA-3′). PCR reactions were performed in triplicate using a 20-μL mixture containing 4 μL of 5 × FastPfu Buffer, 2 μL of 2.5 mM deoxyribonucleotide triphosphate, 0.8 μL of each primer (5 μM), 0.4 μL of FastPfu Polymerase, and 10 μL of template DNA. The paired-end reads obtained from sequencing were assembled based on their overlap relationships, while sequence quality control and filtering were performed. DADA2 was used to generate feature sequences for QIIME2 and the Amplicon Sequence Variants (ASVs) concept was used to construct the feature table. Taxonomic annotations were performed using the SILVA database (Salmaso et al., 2022).

2.4. Statistical analysis

R software (v.4.0.2) was employed for statistical analysis and visualization. The vegan package (v.2.6.2) in R was utilized to calculate the alpha diversity index (Dixon, 2003), which includes various indices such as observed species, Chao1, ACE, Shannon, Simpson, and Pielou's evenness. Notably, Chao1 and ACE were both used to predict the number of operational taxonomic units in the sample, with higher values indicating greater species richness. The ACE calculation formula is as follows:

$$\mathrm{ACE} = S_{\mathrm{abund}} + \frac{S_{\mathrm{rare}}}{C_{\mathrm{ace}}} + \frac{F_1}{C_{\mathrm{ace}}}\gamma_{\mathrm{ace}}^2$$

where S_{abund} refers to the number of abundant species (with an abundance threshold $> n$), S_{rare} refers to the number of rare species (with an abundance threshold $\leq n$), F_1 refers to the number of species with just one individual, and γ_{ace}^2 is the estimated variance coefficient of the rare species.

The Shannon and Simpson indices were both employed to estimate microbial community diversity, with higher Shannon values indicating greater community diversity and higher Simpson values indicating lower community diversity. The Pielou's evenness describes the actual Shannon index value for a community relative to the maximum Shannon index value that can be attained in a community with the same species richness. The Pielou's evenness calculation formula is as follows:

$$\mathrm{Pielou}'\mathrm{sevenness} = \frac{H}{H_{\max}} = \frac{H}{\log_x S}$$

where H refers to the Shannon index, $H_{\max}$ refers to the maximum Shannon index that can be achieved for a given community richness, S refers to the community richness index, and x is the base number, such as 2 or e.

The calculation methods for the other indices are provided by Zeng et al. (2021). Additionally, the VennDiagram package (v.1.7.1) in R enabled Venn analysis. The multcomp package (v.1.4–18) was used to perform Tukey's HSD test for significant differences, while the vegan package was utilized to generate Bray-Curtis distance and carry out the analysis of Non-metric Multidimensional Scaling (NMDS), Anosim, and redundancy analysis (RDA). To visualize community composition differences among sampling areas, a principal coordinates analysis (PCoA) was performed (Paradis et al., 2004) using the ape package (v.5.6–1) in R. Furthermore, the gbm package (v.2.1.8) was used to carry out Aggregated Boosted Tree (ABT) analysis, which evaluates the contributions of environmental factors to community diversity (Ge et al., 2008), while the randomForest package (v.4.6–14) in R was employed to determine the relative importance of predictive factors in the model using Random Forest (RF) analysis. Distance decay analysis conducts linear regression of sample geographic distance or environmental similarity against community composition similarity, and was carried out using the lm function in R. Spearman's correlation was analyzed using the linkET package (v.0.0.2.2) in R, while the figures were visualized using ggplot2 (v.3.3.5) in R. To further explore microbial interactions, a co-occurrence network was designed for the ASVs with top 30% abundance and appearing in samples above 75%. Spearman's coefficient was used for calculation of correlations, with a threshold value of 0.60 set for retention as a network node, and edges with p-values greater than 0.05 were removed. Visualization of the network and the modular analysis was performed using Cytoscape (v.3.8.0) (Deng et al., 2016). Due to the insufficient number of PRE samples (less than 8), it was not suitable to construct a robust co-occurrence network. Therefore, this study conducted network analysis on UR, MR and DR. A neutral community model was used to predict the relationship between ASV detection frequency and their relative abundance across the wider metacommunity. Calculation of 95% confidence intervals around all fitting statistics was done by bootstrapping with 1000 bootstrap replicates (Chen et al., 2019b).

3. Results

3.1. Spatial patterns of phytoplankton community diversity

The eDNA samples of phytoplankton communities in the Dongjiang River were sequenced. The species accumulation curves show the relationship between the number of samples and the number of detected species. The curves tend to flatten when the sampling reaches saturation, suggesting that the sampling was sufficient to reflect the phytoplankton diversity in the Dongjiang River in this study (Fig. S1(a)). After clustering using the ASVs analysis pipeline, the eDNA sequencing data formed 4061 ASVs, annotated to 5 phyla, 15 families, 38 genera, and 89 species. The annotation rate of phytoplankton genera accounted for 82.91% of the total detected genera and 36.73% of the ASVs were annotated to the phytoplankton species level (Fig. S1(b)). The Venn diagram indicated that the number of unique phytoplankton species increased from the upstream to the downstream regions, although the estuary had the fewest unique species (Fig. 2(a)). The upstream regions had a significantly lower relative abundance of shared species than that of the midstream and downstream regions ($p < 0.05$, Fig. 2(b)). In the alpha diversity analysis, both Chao1 and ACE indices for the phytoplankton community in the Dongjiang River were significantly higher in the downstream regions than in the upstream and midstream regions ($p < 0.05$) (Fig. 2(c)). In contrast, the Pielou's evenness exhibited the opposite pattern ($p < 0.05$), with a significantly lower value in the downstream than in the upstream regions. However, no spatial differences in the Shannon or Simpson indices were observed within the basin ($p > 0.05$, Fig. S2).

3.2. *Spatial heterogeneity of the phytoplankton community structure*

The PCoA analysis revealed that the phytoplankton community structure was similar among the upstream, midstream, and downstream regions of the Dongjiang River, but distinctly different from the estuary regions ($p < 0.05$, Fig. 3(a)). Likewise, the NMDS analysis also showed that the phytoplankton composition of S19 in the estuary area and S12 in the downstream region of the Dongjiang River differed significantly from other sites in the basin (stress <0.2, Fig. 3(b)). The Bray-Curtis distance analysis indicated that there was a significant difference in the composition structure of phytoplankton along the river ($p < 0.05$, Fig. 3(c)), with small distances in the phytoplankton community within each site of the midstream and upstream regions, but significantly higher than those in the downstream regions. The species composition distance in the estuary area was also significantly different from those in the upstream regions.

Pyrrophyta, Cryptophyta, and Ochrophyta were dominant phytoplankton in the Dongjiang River (Fig. 4(a)). The upstream and downstream regions of the Dongjiang River were dominated by Pyrrophyta and Cryptophyta, respectively, with their relative abundance accounting for more than 50% of the study area, while the dominant phytoplankton in the midstream regions and estuary area were more evenly distributed, indicating a steady state structure. The Wilcoxon rank sum test was applied to identify species with significant abundance differences among groups (Fig. 4(b)). Pyrrophyta was the phylum in the upstream region that differed most significantly from the midstream and downstream regions ($p < 0.01$), while Bacillariophyta also had significant abundance differences between the upstream and downstream regions ($p < 0.05$).

Using the random forest algorithm to further explore phytoplankton genera (Fig. 4(c)), six phytoplankton genera were identified as predictive of the phytoplankton community, including *Colpodella*, *Chromera*, and *Perkinsidae* of Pyrrophyta, *Chroomonas*, *Hemiselmis*, *Kathablepharidae* of Cryptophyta. These genera were consistent with the above stacked bar plot results, i.e., Pyrrophyta and Cryptophyta were the key dominant phytoplankton in specific river regions. The random forest model based on phytoplankton genus predicted a sample classification accuracy for the Dongjiang River and estuary of 69.09% (Fig. S3), confirming the reliability of the random forest model.

The co-occurrence network diagram (Fig. 4(d)) shows that the network complexity of the phytoplankton community in the Dongjiang River increased from upstream to downstream regions, with the upstream regions having the lowest network modularity index of 6 and the downstream regions having the highest network modularity index of 14.

3.3. *Response of keystone species of phytoplankton to anthropogenic disturbance*

RDA analysis revealed that temperature, Chl-a, and DO were

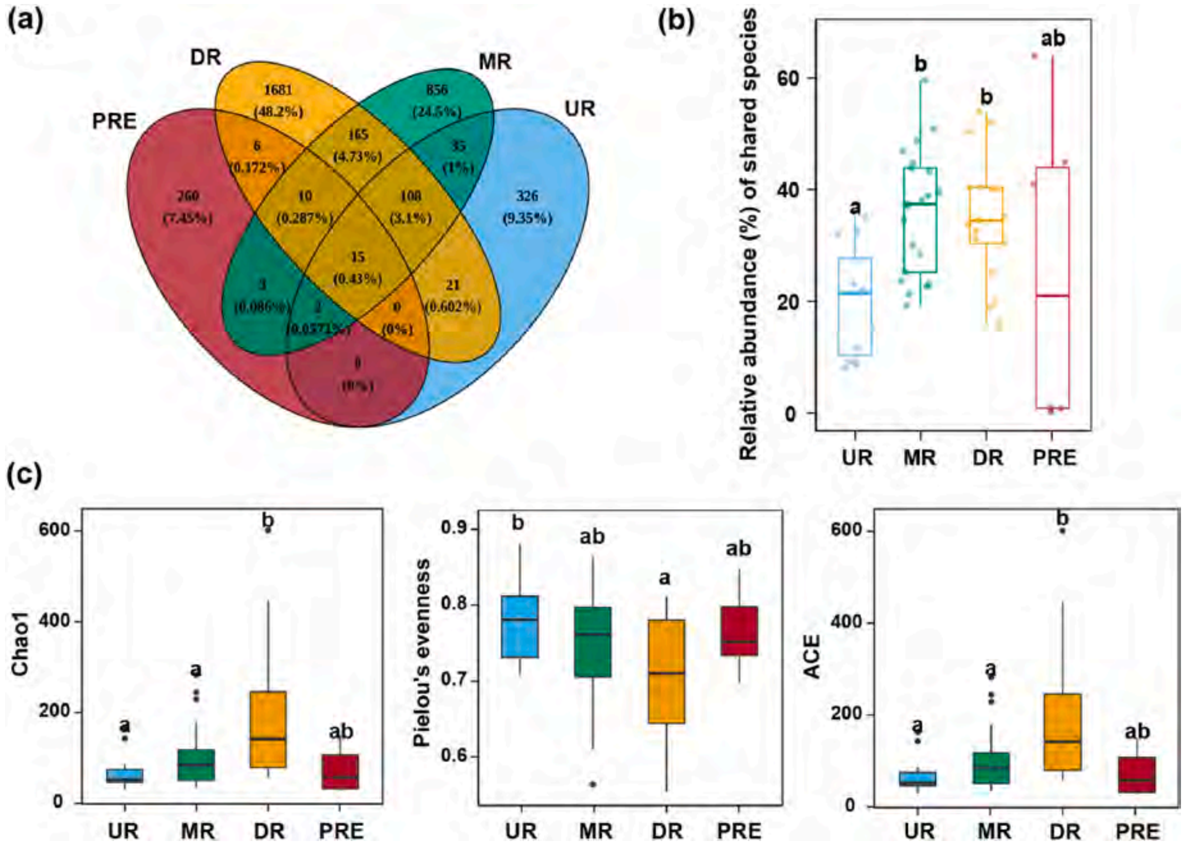

Fig. 2. Alpha diversity and composition of phytoplankton in the Dongjiang River. (a) Venn diagram of unique and shared ASVs in different regions. (b) Relative abundance of shared species. (c) Alpha diversity indices among regions. Different letters indicate significant differences at $p < 0.05$. UR: upstream regions, MR: midstream regions, DR: downstream regions, PRE: Pearl River Estuary.

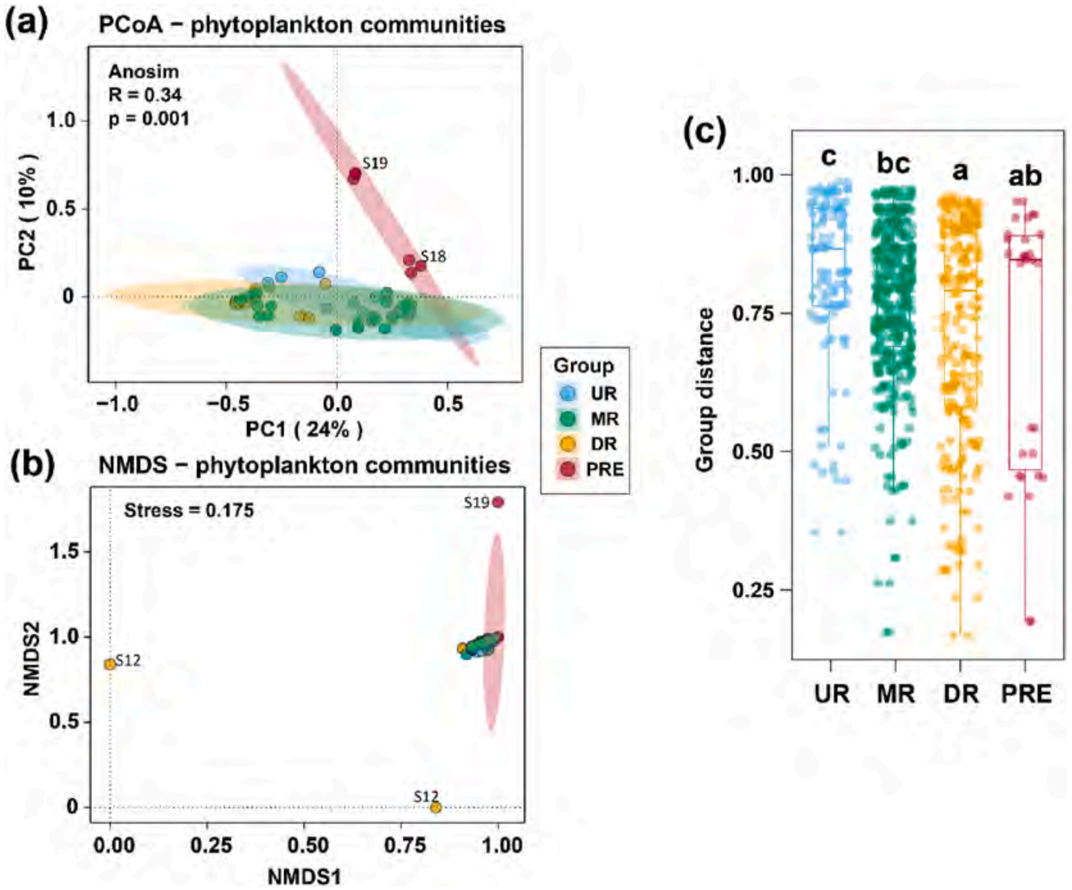

Fig. 3. Beta diversity analysis of the phytoplankton community in the Dongjiang River. (a) and (b) show principal coordinate analysis (PCoA) and non-metric multidimensional scaling (NMDS) ordination based on Bray-Curtis dissimilarity of phytoplankton, respectively. Ellipses represent 95% confidence intervals. (c) Group distance based on Bray-Curtis dissimilarity. Different letters indicate significant differences at $p < 0.05$. UR: upstream regions, MR: midstream regions, DR: downstream regions, PRE: Pearl River Estuary.

important environmental driving factors affecting the phytoplankton community structure in the upstream Dongjiang River, the downstream regions were mainly constrained by high concentrations of nutrients NH_4^+, TP, TN, and BOD_5, while Cl^- and SO_4^{2-} are important environmental factors affecting phytoplankton distribution in the estuary (Fig. 5 (a)). Among the species, *Chroomonas* of Cryptophyta was positively correlated with NH_4^+, *Syndiniales. Group.I* of Pyrrophyta and *Hemiselmis* of Cryptophyta were positively correlated with Cl^- and SO_4^{2-}, while *Colpodella* and *Gymnodiniphycidae* of Pyrrophyta were positively correlated with temperature, Chl-a, and DO. The ABT model was used to evaluated the contributions of environmental factors to community diversity (Fig. 5(b)). Cl^- was the primary factor affecting community composition, followed by NH_4^+ and SO_4^{2-}. Temp, Chl-a, DO, and COD together accounted for 72.03% of the explanation ratio for spatial differences in community diversity.

The explanatory proportion and importance of environmental factors on the variation of keystone species were analyzed by the random forest method (Fig. 5(c)). The results indicated that temperature had a significant negative correlation with *Kathablepharidae* of Cryptophyta, and accounted for more than 85% of its variation in abundance. For *Colpodella* of Pyrrophyta and *Chroomonas* of Cryptophyta, the potential key factors affecting their abundance were temperature, DO, Chl-a, and NO_2^-, and DO and NH_4^+, respectively. These environmental factors accounted for ≥75% of the variations in abundance of these phytoplankton. They are important environmental factors determining the key phytoplankton and were the characteristic parameters used by the random forest algorithm to predict and evaluate differences in the phytoplankton communities. DO exhibited a highly significant positive correlation with *Colpodella* and a highly significant negative correlation with *Chroomonas*.

3.4. Mechanisms of phytoplankton community assembly

Linear regression analysis of environmental heterogeneity and geographic distance among samples with phytoplankton community similarity (Fig. 6(a)) showed that similarities among phytoplankton communities decreased with increasing geographic distance ($R^2 = 0.112$, $p < 0.0001$) and increased with increasing environmental similarity ($R^2 = 0.336$, $p < 0.0001$). The explanatory percentages in the neutral community model for the studied regions ranged from 0 to 25.6% (Fig. 6(b)). Even from the perspective of the entire river basin, only 34.3% could be explained by the neutral model. There was no significant difference in the niche breadth of the phytoplankton communities among the different research regions ($p > 0.05$, Fig. 6(c)).

4. Discussion

4.1. The ecological significance of phytoplankton community structure along river gradient changes

Phytoplankton are highly sensitive to environmental changes (Chang et al., 2022; Huo et al., 2022; Zhang et al., 2022) and often serve as important indicators of aquatic ecological changes induced by external disturbances (Zhang et al., 2023b). This study systematically analyzed the phytoplankton community structure and environmental factors in different regions of the Dongjiang River, and revealed the spatial

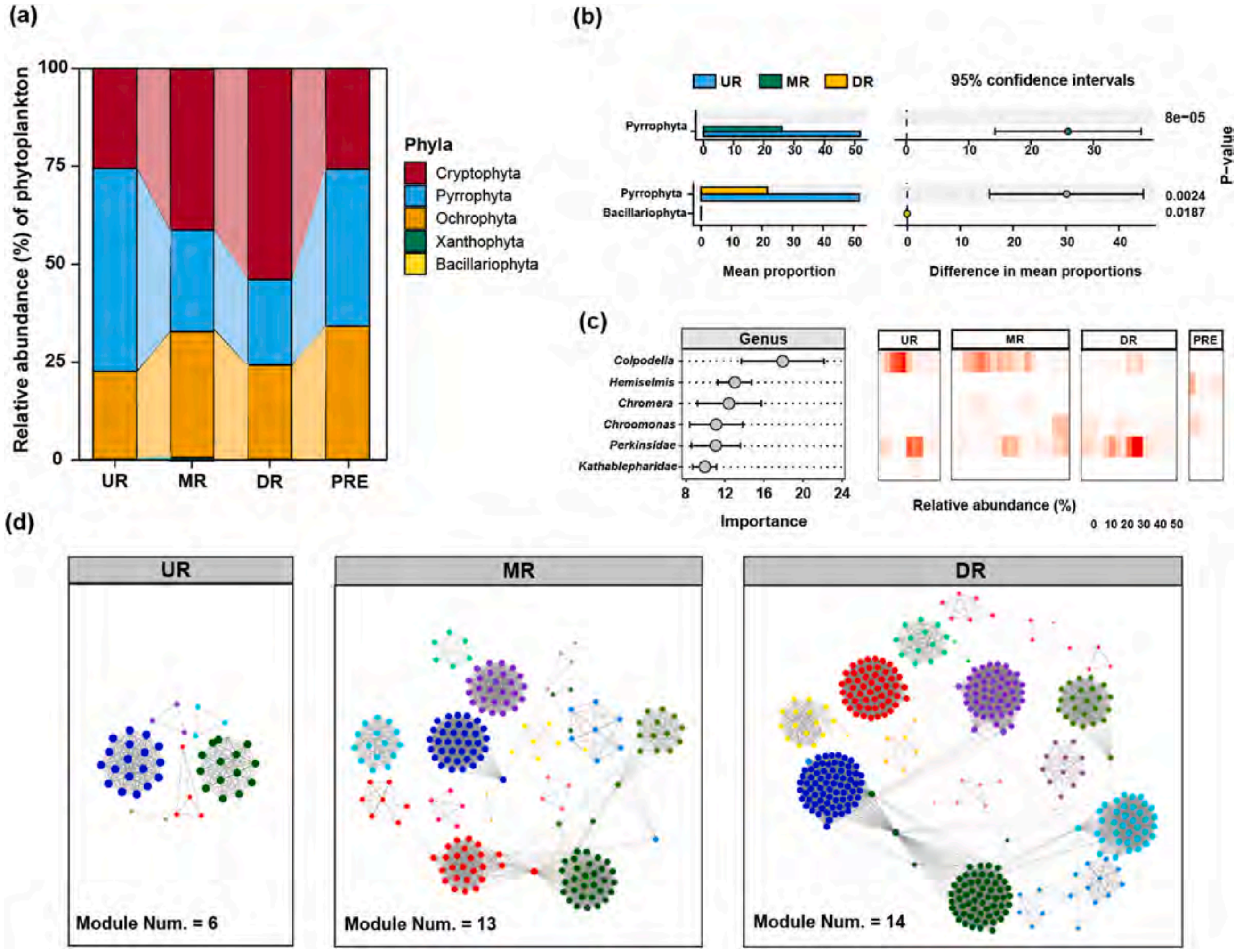

Fig. 4. Dominant and keystone phytoplankton and co-occurrence network of ASVs in the Dongjiang River. (a) Relative abundance of phytoplankton at the phylum level; (b) Phytoplankton with significant differences in abundance among regions at the phylum level. Left: average relative abundance; right: variation in abundance. Rightmost values are *p* values for Wilcoxon rank-sum test. (c) Importance of key phytoplankton genera based on random forest model. Left: importance for grouping and prediction; right: relative abundance. (d) Co-occurrence network of ASVs. Connections represent strong (Spearman's coefficient $|r| > 0.6$) and significant ($p < 0.05$) correlations. Node size is proportional to number of connections. Nodes are colored by modularity. UR: upstream regions, MR: midstream regions, DR: downstream regions, PRE: Pearl River Estuary.

heterogeneity of the phytoplankton community, as well as its relationship with environmental gradients. The phytoplankton community structure showed a trend of increasing abundance, diversity, and complexity from upstream to downstream, reflecting the adaptation and response of phytoplankton to environmental conditions. The phytoplankton community structure was mainly influenced by different environmental factors in different regions. The upstream region was mainly constrained by natural factors (such as temperature, Chl-a, DO), while the downstream region was mainly driven by anthropogenic factors (such as nutrient input and pollution discharge) (Liu et al., 2021; Tao et al., 2022). Excessive nutrients have previously been identified as important variables for phytoplankton mass reproduction (Baek et al., 2022). The estuary area is characterized by complex water flow exchange caused by salt tide backflow (Li et al., 2023; Monteiro et al., 2022). Thus, Cl^- and SO_4^{2-} are important environmental factors affecting phytoplankton distribution in this area. In fact, the terrain in the upstream region of the Dongjiang River is mountainous and hilly, with rapid river flow, which is not conducive to mass reproduction of phytoplankton. Similarly, the water flow in the estuary confluence region is turbulent (Li et al., 2023), and a large number of suspended solids and sediments reduces the transparency and also affects the distribution of phytoplankton (Golubkov and Golubkov, 2022). Therefore, there are few unique phytoplankton species in these regions. The downstream region is also subject to high intensity of human activity disturbance, which results in large inputs of nutrient elements entrained by the river runoff (Liu et al., 2021; Lv et al., 2022). The water flow speed in this region is relatively low, compared to the upstream and estuary areas. Consequently, multiple factors favor the proliferation of phytoplankton in the downstream segment, generating more phytoplankton species that are endemic to this area.

The co-occurrence network complexity of the phytoplankton community in the Dongjiang River increased from upstream to downstream regions, indicating a higher niche differentiation degree within the community and stronger network structure compactness and community connectivity (Shuwang et al., 2023). Studies have shown that in polluted river sections, the phytoplankton symbiotic network will present a more closely associated state (Geng et al., 2022), as higher diversity is conducive to community adaptation to a frequently disturbed environment and faster recovery (Chen et al., 2019a). Higher environmental heterogeneity also produces greater network complexity and

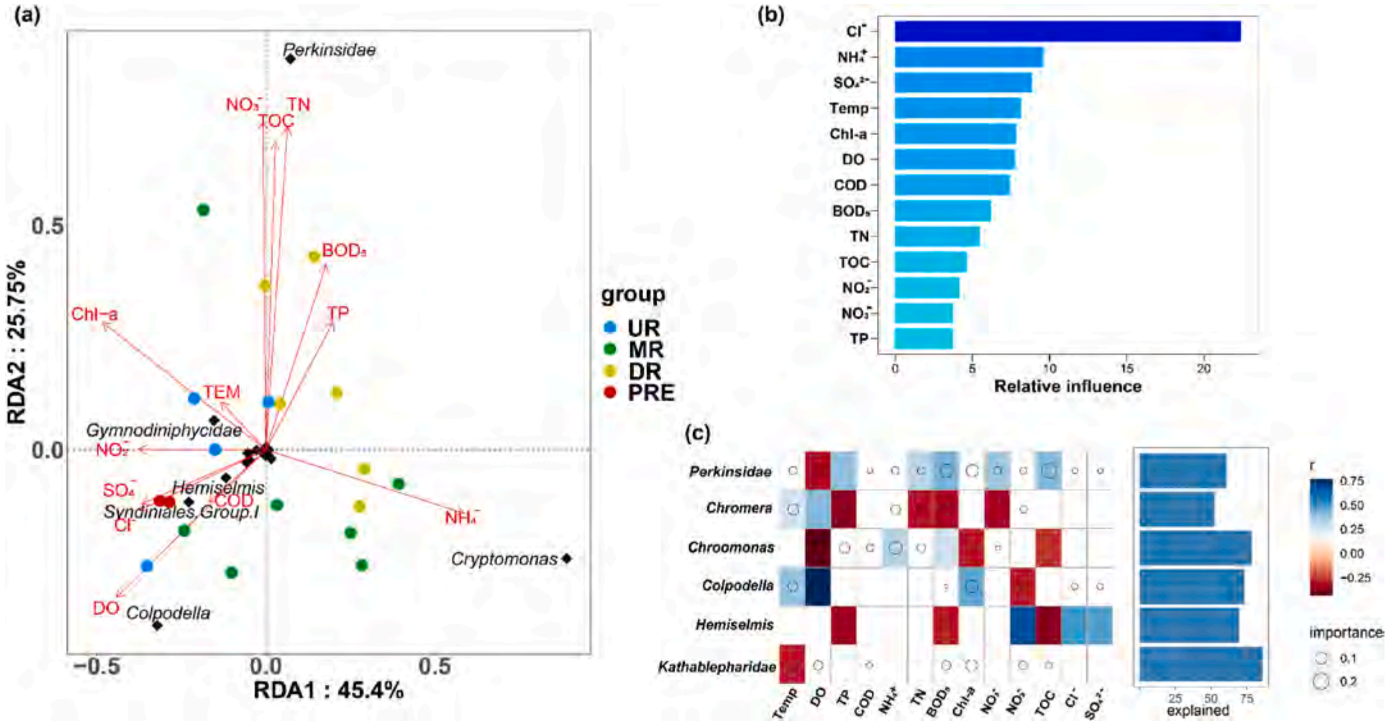

Fig. 5. Correlation analysis between phytoplankton and environmental factors. (a) Redundancy analysis (RDA) of phytoplankton genera, sampling regions, and environmental variables. (b) Aggregated Boosted Tree (ABT) bar chart of environmental factor importance. (c) Correlation coefficients between phytoplankton genera and environmental factors. Results with $p < 0.05$ are shown as color blocks. Circles indicate potential key factors; circle size represents factor importance. The bar chart on the right shows the explanatory ratio of environmental factors to changes in the abundance of key genera. UR: upstream regions, MR: midstream regions, DR: downstream regions, PRE: Pearl River Estuary.

connectivity (Mougi and Kondoh, 2012).

Machine learning has great potential for uncovering nonlinear interactions between multiple variables, it can be used to investigate the impact of complex environmental factors on community distributions, as well as to identify dominant species (Ekundayo et al., 2023; Fan et al., 2020). Using the random forest algorithm to further explore phytoplankton genera, there were some key species with important ecological functions and positions in the phytoplankton community, such as *Colpodella*, *Chromera*, *Perkinsidae*, *Chroomonas*, *Hemiselmis*, *Kathablepharidae*. They showed significant correlations with specific environmental factors and were mainly controlled by these environmental factors.

The above results demonstrate how phytoplankton can adapt to different hydrological and geographical conditions, and how they can serve as indicators for water environmental quality and function. The study reveal the ecological significance of phytoplankton community structure along the river gradient in the Dongjiang River, which has not been reported before.

4.2. Environmental selection shaped the phytoplankton assemblage in the Dongjiang River

The ecological processes that shape phytoplankton communities are a long-standing question in aquatic ecology; environmental and spatial factors are two main drivers influencing community aggregation in natural environments (Chen et al., 2019a; Wu et al., 2023). The similarity of phytoplankton community was negatively correlated with geographic distance and positively correlated with environmental similarity, indicating that both spatial and environmental factors had an impact on the phytoplankton community composition. However, the environmental heterogeneity was higher in the downstream and estuary area, resulting in a significant distance decay pattern of phytoplankton community similarity and a more spatially heterogeneous community composition. Numerous studies have shown that coastal areas are influenced by both marine and terrestrial factors, have many complex sources, transformation processes, and contaminant profiles, and are the largest ecological transition zones in the world. Urban areas and population density can cause various degrees of disturbance that affect the aquatic ecosystem (Maure et al., 2021; Xie et al., 2021). Dongjiang River has similar environmental characteristics. Its upstream areas are freshwater, while its downstream and estuary areas are affected by tides and seawater backflow. These hydrodynamic conditions make its aquatic environment more complex in downstream regions than in its upstream freshwater regions (Zhang et al., 2023a). The correlation coefficients of the distances also suggest that the phytoplankton community assembly was more influenced by environmental variables alone than by spatial variables.

Neutral models can further distinguish the relative roles of stochastic dispersal and environmental selection (Alonso et al., 2006; Lu et al., 2023). In addition, niche breadth was used to explore metabolic adaptation and environmental selection for species in the studied communities (Pandit et al., 2009). The neutral model explained a low percentage of the variation in phytoplankton community structure, and there were differences among different regions, indicating that stochastic processes played only a minor or local role in the assembly of phytoplankton community. The niche width did not show significant differences among different regions, indicating that the species in the phytoplankton community had similar metabolic adaptability and environmental selectivity.

In summary, we can infer that deterministic processes, namely environmental filtering, were the dominant mechanisms for the assembly of phytoplankton community in the Dongjiang River, while stochastic processes, namely dispersal and drift, were secondary or auxiliary mechanisms. This conclusion is consistent with previous studies that found that environmental variables can explain the community variation in stable and isolated ecosystems or aquatic ecosystems with strong environmental gradients (Little et al., 2019; Wu et al., 2023;

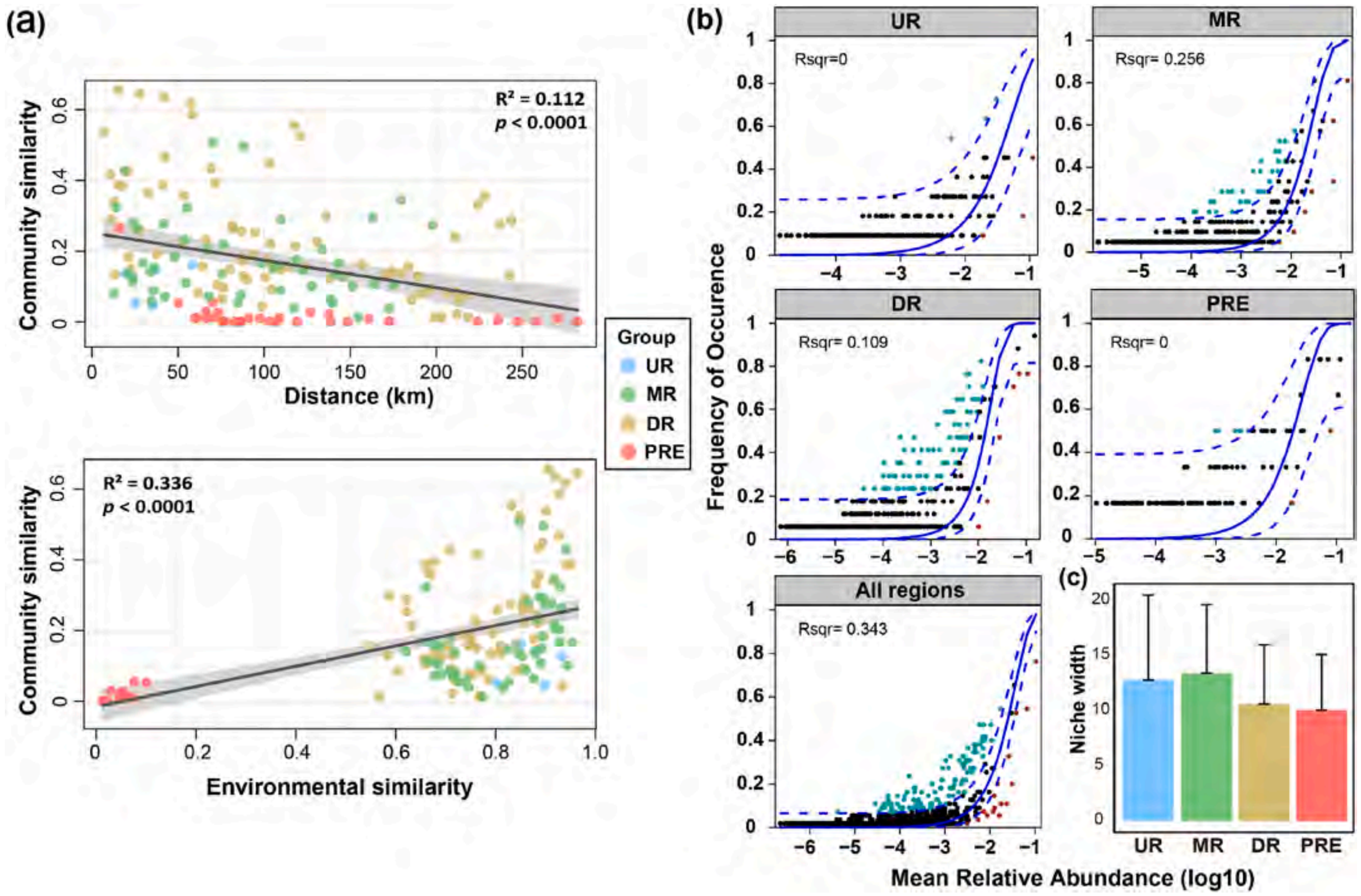

Fig. 6. Assembly mechanisms of the phytoplankton community in the Dongjiang River. (a) Distance dependence curve for phytoplankton community similarity and geographical and environmental similarity. Slope of the fitting curve indicates distance-dependent similarity. (b) Diagram of the phytoplankton neutral community model. Each dot represents an ASV; black, blue, and red dots indicate ASVs that conform, exceed, or fall below the neutral model, respectively. Blue solid line represents the optimal fit of the neutral model; blue dashed lines represent the 95% confidence interval. R^2 represents the goodness of fit of the neutral model. (c) The niche breadth of the phytoplankton community. Different letters indicate significant differences at $p < 0.05$. UR: upstream regions, MR: midstream regions, DR: downstream regions, PRE: Pearl River Estuary.

Zhang et al., 2023b). However, our study also differs from previous studies in several aspects. First, we used eDNA metabarcoding technique, which has higher sensitivity and resolution than traditional methods, to detect the phytoplankton diversity and composition. Second, we covered a large spatial scale and a fine spatial resolution, which enabled us to capture the spatial patterns and variations of phytoplankton communities along the river gradient. However, we only sampled during the dry season, which may not reflect the temporal changes and dynamics of phytoplankton communities. Therefore, our study provides a novel and comprehensive perspective on the phytoplankton diversity and assembly in the Dongjiang River, but also calls for further research on the temporal and seasonal effects.

5. Conclusion

This research investigated the spatial heterogeneity, environmental driving factors, and community assembly mechanisms of the phytoplankton community distribution in the Dongjiang River and estuary area based on eDNA sequencing data. The results revealed that environmental selection largely shaped the structure of phytoplankton communities in the aquatic ecosystem of the Dongjiang River. The upstream regions and estuary area were mainly influenced by natural environmental factors, with temperature, Chl-a, DO, and Cl^-, SO_4^{2-} as the main environmental constraints, respectively. The downstream regions were largely influenced by human activities, with nutrients as the key environmental driving factors. Therefore, it is essential to take measures to reduce the input of anthropogenic pollutants in the downstream regions for maintaining the diversity, health, and stability of phytoplankton and the aquatic ecosystem. The results of this study provide scientific support for aquatic ecological environmental health and sustainable development of the Dongjiang River. However, the findings in this research may not fully represent the long-term dynamics of phytoplankton communities, further research is needed to fully understand the seasonal and interannual variations in these communities.

CRediT authorship contribution statement

Luping Zeng: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft. **Jing Wen:** Formal analysis, Investigation, Data curation, Writing - original draft. **Bangjie Huang:** Investigation, Software. **Yang Yang:** Writing - review & editing. **Zhiwei Huang:** Data curation. **Fantang Zeng:** Writing - review & editing. **Huaiyang Fang:** Supervision, Resources, Writing - review & editing. **Hongwei Du:** Supervision, Resources, Validation, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This research was financially supported by the National Key Research and Development Program of China (2022YFC3202202), the financial support from the Guangdong Basic and Applied Basic Research Foundation (2023A1515012179), the Fundamental Research Funds for the Central Public Welfare Research Institutes (PM-zx703-202305-227), the Key R & D program of Guangdong Province (2020B1111350001).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envres.2023.117708.

References

Alonso, D., Etienne, R.S., McKane, A.J., 2006. The merits of neutral theory. Trends Ecol. Evol. 21 (8), 451–457.

Baek, S.S., Jung, E.Y., Pyo, J., Pachepsky, Y., Son, H., Cho, K.H., 2022. Hierarchical deep learning model to simulate phytoplankton at phylum/class and genus levels and zooplankton at the genus level. Water Res. 218.

Barnes, M.A., Turner, C.R., Jerde, C.L., Renshaw, M.A., Chadderton, W.L., Lodge, D.M., 2014. Environmental conditions influence eDNA persistence in aquatic systems. Environ. Sci. Technol. 48 (3), 1819–1827.

Chang, C.W., Miki, T., Ye, H., Souissi, S., Adrian, R., Anneville, O., Agasild, H., Ban, S.H., Be`eri-Shlevin, Y., Chiang, Y.R., Feuchtmayr, H., Gal, G., Ichise, S., Kagami, M., Kumagai, M., Liu, X., Matsuzaki, S.I.S., Manca, M.M., Noges, P., Piscia, R., Rogora, M., Shiah, F.K., Thackeray, S.J., Widdicombe, C.E., Wu, J.T., Zohary, T., Hsieh, C.H., 2022. Causal networks of phytoplankton diversity and biomass are modulated by environmental context. vol 13 Nat. Commun. 13 (1), 1140, 2022.

Chase, J.M., 2010. Stochastic community assembly causes higher biodiversity in more productive environments. Science 328 (5984), 1388–1391.

Chase, J.M., Myers, J.A., 2011. Disentangling the importance of ecological niches from stochastic processes across scales. Phil. Trans. Biol. Sci. 366 (1576), 2351–2363.

Chen, B.Z., Smith, S.L., Wirtz, K.W., 2019a. Effect of phytoplankton size diversity on primary productivity in the North Pacific: trait distributions under environmental variability. Ecol. Lett. 22 (1), 56–66.

Chen, W.D., Ren, K.X., Isabwe, A., Chen, H.H., Liu, M., Yang, J., 2019b. Stochastic processes shape microeukaryotic community assembly in a subtropical river across wet and dry seasons. Microbiome 7 (1).

Clark, D.E., Pilditch, C.A., Pearman, J.K., Ellis, J.I., Zaiko, A., 2020. Environmental DNA metabarcoding reveals estuarine benthic community response to nutrient enrichment - evidence from an in-situ experiment. Environ. Pollut. 267.

Cordier, T., Alonso-Saez, L., Apotheloz-Perret-Gentil, L., Aylagas, E., Bohan, D.A., Bouchez, A., Chariton, A., Creer, S., Fruhe, L., Keck, F., Keeley, N., Laroche, O., Leese, F., Pochon, X., Stoeck, T., Pawlowski, J., Lanzen, A., 2021. Ecosystems monitoring powered by environmental genomics: a review of current strategies with an implementation roadmap. Mol. Ecol. 30 (13), 2937–2958.

Datry, T., Bonada, N., Heino, J., 2016. Towards understanding the organisation of metacommunities in highly dynamic ecological systems. Oikos 125 (2), 149–159.

Deng, Y., Zhang, P., Qin, Y.J., Tu, Q.C., Yang, Y.F., He, Z.L., Schadt, C.W., Zhou, J.Z., 2016. Network succession reveals the importance of competition in response to emulsified vegetable oil amendment for uranium bioremediation. Environ. Microbiol. 18 (1), 205–218.

Diao, C.Y., Jia, H., Guo, S.J., Hou, G., Xian, W.W., Zhang, H., 2022. Biodiversity exploration in autumn using environmental DNA in the South China sea. Environ. Res. 204.

Dixon, P., 2003. VEGAN, a package of R functions for community ecology. J. Veg. Sci. 14 (6), 927–930.

Djurhuus, A., Closek, C.J., Kelly, R.P., Pitz, K.J., Michisaki, R.P., Starks, H.A., Walz, K.R., Andruszkiewicz, E.A., Olesin, E., Hubbard, K., Montes, E., Otis, D., Muller-Karger, F. E., Chavez, F.P., Boehm, A.B., Breitbart, M., 2020. Environmental DNA reveals seasonal shifts and potential interactions in a marine community. Nat. Commun. 11 (1).

Ekundayo, T.C., Ijabadeniyi, O.A., Igbinosa, E.O., Okoh, A.I., 2023. Using machine learning models to predict the effects of seasonal fluxes on Plesiomonas shigelloides population density. Environ. Pollut. 317.

Fan, J.T., Wang, S.P., Li, H., Yan, Z.G., Zhang, Y.Z., Zheng, X., Wang, P.Y., 2020. Modeling the ecological status response of rivers to multiple stressors using machine learning: a comparison of environmental DNA metabarcoding and morphological data. Water Res. 183.

Ge, Y., He, J.Z., Zhu, Y.G., Zhang, J.B., Xu, Z.H., Zhang, L.M., Zheng, Y.M., 2008. Differences in soil bacterial diversity: driven by contemporary disturbances or historical contingencies? ISME J. 2 (3), 254–264.

Geng, Y.C., Peng, C.R., Zhou, W.C., Huang, S., Zhou, P.P., Wang, Z.C., Qin, H.J., Li, D.H., 2022. Gradient rise in seepage pollution levels in tailings ponds shapes closer linkages between phytoplankton and bacteria. J. Hazard Mater. 437.

Golubkov, M., Golubkov, S., 2022. Impact of the construction of new port facilities on primary production of plankton in the neva estuary (baltic sea). Front. Mar. Sci. 9.

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., Martiny, J.B.H., 2012. Beyond biogeographic patterns: processes shaping the microbial landscape. Nat. Rev. Microbiol. 10 (7), 497–506.

Huo, S.L., Zhang, H.X., Wang, J.F., Chen, J.A., Wu, F.C., 2022. Temperature and Precipitation Dominates Millennium Changes of Eukaryotic Algal Communities in Lake Yamzhog Yumco, Southern Tibetan Plateau, vol. 829. Science of the Total Environment.

Jarman, S.N., Berry, O., Bunce, M., 2018. The value of environmental DNA biobanking for long-term biomonitoring. Nat. Ecol. Evol. 2 (8), 1192–1193.

Ji, F.F., Yan, L., Yan, S.H., Qin, T.L., Shen, J.Z., Zha, J.M., 2021. Estimating aquatic plant diversity and distribution in rivers from Jingjinji region, China, using environmental DNA metabarcoding and a traditional survey method. Environ. Res. 199.

Lambert, S., Tragin, M., Lozano, J.C., Ghiglione, J.F., Vaulot, D., Bouget, F.Y., Galand, P. E., 2019. Rhythmicity of coastal marine picoeukaryotes, bacteria and archaea despite irregular environmental perturbations. ISME J. 13 (2), 388–401.

Li, Z.Y., Jiang, P.W., Wang, L.X., Liu, L., Li, M., Zou, K.S., 2023. A comparison of seasonal composition and structure of fish community between environmental DNA technology and gillnetting in the Pearl River Estuary, China. Ecol. Indicat. 147.

Little, C.J., Fronhofer, E.A., Altermatt, F., 2019. Dispersal syndromes can impact ecosystem functioning in spatially structured freshwater populations. Biol. Lett. 15 (3).

Liu, Q., Tian, Y.L., Liu, Y., Yu, M., Hou, Z.J., He, K.J., Xu, H., Cui, B.S., Jiang, Y., 2021. Relationship between dissolved organic matter and phytoplankton community dynamics in a human-impacted subtropical river. J. Clean. Prod. 289.

Lu, Q., Zhang, S.Y., Du, J.Q., Liu, Q., Dong, C.X., Zhao, J.D., Wang, Y.F., Yao, M., 2023. Multi-group biodiversity distributions and drivers of metacommunity organization along a glacial-fluvial-limnic pathway on the Tibetan plateau. Environ. Res. 220.

Lv, X.M., Chen, G.L., Wu, Y.C., Yu, L.L., Zhou, Y., Yu, Y.X., Lan, S.H., Hu, J.J., 2022. Ecological and AhR-mediated risk assessment of polycyclic aromatic hydrocarbons and polybrominated diphenyl ethers on multiple aquatic species in river water: a combined chemical analysis and in silico approach. Sci. Total Environ. 820.

Maure, E.D., Terauchi, G., Ishizaka, J., Clinton, N., DeWitt, M., 2021. Globally consistent assessment of coastal eutrophication. Nat. Commun. 12 (1).

Monteiro, I.B., Dantas, D.V., Makrakis, M.C., Lorenzi, L., Ribeiro, S.A., Pezzin, A.P.T., Silveira, V.F., Gentil, E., 2022. Composition and spatial distribution of floating plastic debris along the estuarine ecocline of a subtropical coastal lagoon in the Western Atlantic. Mar. Pollut. Bull. 179.

Mougi, A., Kondoh, M., 2012. Diversity of interaction types and ecological community stability. Science 337 (6092), 349–351.

Pandit, S.N., Kolasa, J., Cottenie, K., 2009. Contrasts between habitat generalists and specialists: an empirical extension to the basic metacommunity framework. Ecology 90 (8), 2253–2262.

Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20 (2), 289–290.

Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apotheloz-Perret-Gentil, L., Beja, P., Boggero, A., Borja, A., Bouchez, A., Cordier, T., Domaizon, I., Feio, M.J., Filipe, A.F., Fornaroli, R., Graf, W., Herder, J., van der Hoorn, B., Jones, J.I., Sagova-Mareckova, M., Moritz, C., Barquin, J., Piggott, J.J., Pinna, M., Rimet, F., Rinkevich, B., Sousa-Santos, C., Specchia, V., Trobajo, R., Vasselon, V., Vitecek, S., Zimmerman, J., Weigand, A., Leese, F., Kahlert, M., 2018. The future of biotic indices in the ecogenomic era: integrating (e) DNA metabarcoding in biological assessment of aquatic ecosystems. Sci. Total Environ. 637, 1295–1310.

Qian, X., Tang, X.H., Tian, W.W., Xiao, X.X., Wang, Y.L., Lv, Q.X., Li, H.Z., Feng, S., 2023. Climate factors and host chemical profiles jointly drives the bacterial community assembly in Mussaenda pubescens stems. Environ. Res. 235.

Rosindell, J., Hubbell, S.P., Etienne, R.S., 2011. The unified neutral theory of biodiversity and biogeography at age ten. Trends Ecol. Evol. 26 (7), 340–348.

Salmaso, N., Vasselon, V., Rimet, F., Vautier, M., Elersek, T., Boscaini, A., Donati, C., Moretto, M., Pindo, M., Riccioni, G., Stefani, E., Capelli, C., Lepori, F., Kurmayer, R., Mischke, U., Klemencic, A.K., Novak, K., Greco, C., Franzini, G., Fusato, G., Giacomazzi, F., Lea, A., Menegon, S., Zampieri, C., Macor, A., Virgilio, D., Zanut, E., Zorza, R., Buzzi, F., Domaizon, I., 2022. DNA sequence and taxonomic gap analyses to quantify the coverage of aquatic cyanobacteria and eukaryotic microalgae in reference databases: results of a survey in the Alpine region. Sci. Total Environ. 834.

SEPA, 2002. The State Environmental Protection Administration (SEPA), Water and Wastewater Monitoring and Analysis Method, fourth ed. China Environmental Science Press, Beijing.

Shuwang, X., Zhang, G.D., Li, D.Y., Wen, Y.J., Zhang, G.C., Sun, J., 2023. Spatial and temporal changes in the assembly mechanism and co-occurrence network of the chromophytic phytoplankton communities in coastal ecosystems under anthropogenic influences. Sci. Total Environ. 877.

Sun, Y., Li, H.J., Wang, X.C., Jin, Y., Nagai, S., Lin, S.J., 2023. Phytoplankton and microzooplankton community structure and assembly mechanisms in northwestern pacific ocean estuaries with environmental heterogeneity and geographic segregation. Microbiol. Spectr. 11 (2).

Tao, D.Y., Jin, Q.Q., Ruan, Y.F., Zhang, K., Jin, L.J., Zhan, Y.T., Su, G.Y., Wu, J.X., Leung, K.M.Y., Lam, P.K.S., He, Y.H., 2022. Widespread occurrence of emerging Ewaste contaminants-Liquid crystal monomers in sediments of the Pearl River Estuary, China. J. Hazard Mater. 437.

Van Driessche, C., Everts, T., Neyrinck, S., Halfmaerten, D., Haegeman, A., Ruttink, T., Bonte, D., Brys, R., 2023. Using environmental DNA metabarcoding to monitor fish communities in small rivers and large brooks: insights on the spatial scale of information. Environ. Res. 228.

Wu, S.H., Dong, Y.Z., Stoeck, T., Wang, S.J., Fan, H.A., Wang, Y.X., Zhuang, X.L., 2023. Geographic characteristics and environmental variables determine the diversities and assembly of the algal communities in interconnected river-lake system. Water Res. 233.

Xie, R.L., Zhao, G.F., Yang, J.H., Wang, Z.H., Xu, Y.P., Zhang, X.W., Wang, Z.J., 2021. eDNA metabarcoding revealed differential structures of aquatic communities in a dynamic freshwater ecosystem shaped by habitat heterogeneity. Environ. Res. 201.

Xue, Y.Y., Chen, H.H., Yang, J.R., Liu, M., Huang, B.Q., Yang, J., 2018. Distinct patterns and processes of abundant and rare eukaryotic plankton communities following a reservoir cyanobacterial bloom. ISME J. 12 (9), 2263–2277.

Zeng, L.P., Dai, Y.N., Zhang, X.M., Man, Y., Tai, Y.P., Yang, Y., Tao, R., 2021. Keystone species and niche differentiation promote microbial N, P, and COD removal in pilot scale constructed wetlands treating domestic sewage. Environ. Sci. Technol. 55 (18), 12652–12663.

Zhang, L.J., Yang, J.H., Zhang, Y., Shi, J.Z., Yu, H.X., Zhang, X.W., 2022. eDNA Biomonitoring Revealed the Ecological Effects of Water Diversion Projects between Yangtze River and Tai Lake, vol. 210. Water Research.

Zhang, W.Q., Han, S.J., Zhang, D.W., Shan, B.Q., Wei, D.Y., 2023a. Variations in dissolved oxygen and aquatic biological responses in China's coastal seas. Environ. Res. 223.

Zhang, X.W., 2019. Environmental DNA shaping a new era of ecotoxicological research. Environ. Sci. Technol. 53 (10), 5605–5612.

Zhang, Z., Li, J., Li, H.J., Wang, L.L., Zhou, Y.Q., Li, S.Z., Zhang, Z.J., Feng, K., Deng, Y., 2023b. Environmental DNA metabarcoding reveals the influence of human activities on microeukaryotic plankton along the Chinese coastline. Water Res. 233.

Zou, K.S., Chen, J.W., Ruan, H.T., Li, Z.H., Guo, W.J., Li, M., Liu, L., 2020. eDNA metabarcoding as a promising conservation tool for monitoring fish diversity in a coastal wetland of the Pearl River Estuary compared to bottom trawling. Sci. Total Environ. 702.

Zou, K.S., Wang, R.L., Xu, S.N., Li, Z.Y., Liu, L., Li, M., Zhou, L., 2021. Changes in protist communities in drainages across the Pearl River Delta under anthropogenic influence. Water Res. 200.