Application of environmental DNA metabarcoding to identify fish community characteristics in subtropical river systems

Sai Wang[1], Dong-Hai Wu[1], En-Ni Wu[1], Nan-Lin Chen[1], Shi-Di Fan[1], Tuan-Tuan Wang[1], Zhong-Bing Chen[2], Yong-Duo Song[1], Yang Zhang[3], Zhuo-Luo a[4], Lei Xiang[5], Yu-Xi Yang[1], and Rong-Mei Zi[1]

[1]Hainan University
[2]Czech University of Life Sciences Prague
[3]Shenzhen Guanghuiyuan Environment Water Co., Ltd.
[4]China Water Resources Pearl River Planning Surveying & Designing Co., Ltd.
[5]Jinan University

August 4, 2023

Abstract

Fish play vital roles in river ecosystems; however, traditional investigations of fish usually cause certain ecological damage. Extracting DNA from aquatic environments and identifying DNA sequences offers an alternative, non-invasive approach for detecting fish species. In this study, environmental DNA (eDNA), coupled with PCR and next-generation sequencing, and electrofishing were used to compare their effects in identifying fish community characteristics. In three subtropical rivers of southern China, fish specimens and eDNA in water were collected from headwaters to estuaries. Both eDNA OTU richness and individual abundance (including number and biomass) could group 38 sampling sites into eight spatial zones with significant differences in local fish community composition. Compared with the order-/family-level grouping, the genus-/species-level grouping could more accurately recognize the differences between upstream zones I - III, midstream zones IV - V, and downstream zones VI – VIII. From headwaters to estuary, two environmental gradients significantly influenced the longitudinal distribution of fish species, including the first gradient composed of habitat and physical water parameters and the second gradient composed of chemical water parameters. The high regression coefficient of alpha diversity between eDNA and electrofishing methods as well as the accurate recognition of dominant, alien, and biomarker species at each spatial zone indicated that eDNA could characterize fish community attributes at a similar level of traditional approach. Generally, our results demonstrated that eDNA metabarcoding can be used as an effective tool in revealing fish composition and diversity, which is important for using the eDNA technique in aquatic field monitoring.

Hosted file

`image1.emf` available at https://authorea.com/users/646453/articles/658361-application-of-environmental-dna-metabarcoding-to-identify-fish-community-characteristics-in-subtropical-river-systems

1 Introduction

Large river systems are essential for providing critical foraging, breeding ground, and nursery habitats for a variety of fauna and are considered among the most productive ecosystems worldwide (Wang et al., 2020a; Wang et al., 2021b). Fish, as consumers at high trophic levels in river food webs, represent the sum of a wide range of complex trophic interactions (Wang et al., 2018b). Linking the ecological indicators of fish communities to human interference remains an important goal for river managers (Zou et al., 2020).

The distribution, composition and diversity of fish communities are commonly used as proxies to assess an ecosystem's health and integrity. To appropriately manage and protect aquatic ecosystems, it is essential to develop effective and eco-friendly monitoring approaches to collect field data and obtain biological parameters (Kumar et al., 2022; Shu et al., 2021).

Traditional monitoring of fish diversity has depended largely on census methods such as electrofishing, gill/hoop/seine netting, and dredging/trawling (Wang et al., 2020a; Wang et al., 2019a). However, those methods have always been limited by their low sampling efficiencies, destructiveness to organisms, and strict reliance on taxonomic expertise (Sakata et al., 2020; Zhang et al., 2020). The application of environmental DNA (eDNA) metabarcoding for fish diversity analysis has emerged and offers a new avenue for characterizing fish diversity in river ecosystems (Pont et al., 2018). It provides cost-effective, dependable, rapid and continuous investigations for monitoring and assessing fish diversity, which is crucial for the timely and effective management and conservation of river and estuary ecosystems (Garlapati et al., 2019).

The metabarcoding approach coupled with the use of eDNA is a potentially powerful tool for surveying and assessing aquatic diversity. Numerous studies have demonstrated the utility of eDNA metabarcoding for assessing fish diversity (Rourke et al., 2022). Researchers have successfully applied eDNA metabarcoding to monitor fish diversity in freshwater and seawater samples from different habitats, especially in streams, reservoirs, estuaries and oceans (Civade et al., 2016; Stoeckle et al., 2017; Yao et al., 2022; Zou et al., 2020). The results from these studies have shown that eDNA metabarcoding is a sound biomonitoring tool for use in the conservation and management of aquatic ecosystems (Nguyen et al., 2020). Currently, the application of eDNA metabarcoding to monitor and assess biodiversity is at the forefront of the available methods used by ecologists and conservation scientists (Beng and Corlett, 2020; Bernos et al., 2023).

Many studies have compared eDNA results to those from traditional methods and have shown a correlation between the results from the two approaches (Lacoursière-Roussel et al., 2016; Port et al., 2016). In some studies, eDNA analysis was superior for characterizing fish biodiversity compared to traditional techniques such as electrofishing and hoop netting (Nguyen et al., 2020; Pont et al., 2018). In other studies, the results obtained from eDNA have been comparable to those of traditional methods in which fishes are caught through visual dive surveys and trawling (Port et al., 2016; Zou et al., 2020). Previous studies have shown that eDNA metabarcoding retains a higher diversity of taxa than the traditional method; however, the practical application of eDNA in evaluating the composition and structure of fish communities has been less explored, and whether eDNA could replace traditional monitoring is still unknown.

The river systems of southern China, in a typical subtropical monsoon climate zone, serve as reserves for biodiversity conservation. Due to disproportionate use of coastal wetland resources and intense anthropogenic activities (i.e., drainage, reclamation, and pollution), subtropical river ecosystems (e.g., the Pearl River) have been badly damaged, and their biodiversity and bioresources have seriously declined. To investigate the distribution and composition of fish communities in this area, eDNA metabarcoding studies combined with electrofishing surveys were conducted. In addition, the diversity of fluvial fishes observed by electrofishing with the characterization of eDNA collected concurrently from rivers by metabarcoding was compared. The objectives of our study include 1) using a metabarcoding protocol to assess the eDNA-based composition and diversity of fish communities; 2) analysing the response between eDNA OTU richness and fish amounts (e.g., individual number and biomass); and 3) exploring the application of eDNA in assessing environmental influences on fish diversity.

2 Materials and methods

2.1 Study area and sampling sites

The Pearl River, the second largest river in China, is in the tropical and subtropical areas of Southeast Asia and serves as a critical water source for Guangdong Province and Hong Kong. The water sources of the Pearl River are essential for supplying drinking water, generating power, and performing irrigation in Guangdong Province. Thus, the ecological health of the Pearl River is important for the sustainable development of the Pearl River Delta. The Pearl River consists of multiple tributaries – the North, West, East, Liuxi, and

Zeng Rivers – which ultimately merge into the Pearl River Estuary in southern China. The current survey originated from the Liuxi River and Zeng River (Fig. 1), which are regarded as the water sources with high cultural significance for Guangzhou, the capital city of Guangdong Province in southern China. The Liuxi River and Zeng River are the upstream source waters for the Pearl River, including both upstream wadable and downstream nonwadable habitats. Thus, they are important for the comparison between the eDNA metabarcoding protocol and electrofishing method.

Thirty-eight sampling sites were chosen from the headwaters of the Liuxi (L1 – L14) and Zeng Rivers (Z1 – Z12) to the mainstream of the Pearl River (P1 – P12) and then to the Pearl River mouth connected with the South China Sea (Fig. 1). The landforms within the drainage basin are dominated by medium-low mountains and hills. The mean annual temperature and precipitation are 21.6 °C and 2188 mm, respectively. Approximately 83.3% of the runoff is discharged in the flood season (from April to September). The zonal soil and vegetation are Udic Ferralisols and southern subtropical evergreen broad-leaved forest, respectively. Forest coverage is approximately 50%. Arable land accounts for only 8.7% of the drainage area, with paddy fields accounting for most of this. Environmental data, including physicochemical parameters of water quality, habitat characteristics, and bacterial amounts (see Table S1 in the Supplementary Material), were provided by two nationally accredited (China Metrology Accreditation) third-party testing agencies (Hainan Qingxiao Environmental Testing Co., Ltd. and Hainan Qianchao Ecological Technology Co., Ltd.).

2.2 Traditional method (electrofishing) for fish investigation

Fish were collected from the headwaters to the estuary (Fig. 1) during the rainy season (from June to July) in 2022; each site was sampled three times over a rainy season following basic guidelines (Barbour et al., 1999; Hauer and Lamberti, 2007). Electrofishing equipment consisted of a 24-kW generator, a 12 V-160 A lithium battery, a silicon-controlled inverter, and two continuously adjustable voltage and frequency regulators. This equipment was used to effectively stun and collect fish (individual weight $<$ 10 kg) in a 2 m wide $\times$ 2 m long $\times$ 3.5 m deep water column. Due to varying water levels, two electrofishing operations were conducted as follows: 1) At wadeable sites, single-pass backpack electrofishing was performed by two operators moving in a zig-zag fashion. Electrofishing was adjusted at low voltage and mixed frequency, and the walking speed was controlled to ensure a sampling effort of approximately 8 m^2 min^{-1}over 30 minutes. 2) At nonwadeable sites, personnel and equipment were loaded with a welded hull boat, and a bamboo quant was used to propel the boat to eliminate noise disturbance to fish. Electrofishing was adjusted at a high voltage and main frequency, and the paddling speed was controlled to ensure a sampling effort of approximately 6 m^2 min^{-1}. Due to the high water depth, a large scoop net was used by a sternward auxiliary to collect the stunned benthic fish that floated slowly upwards. Boat-electrofishing was conducted over a distance of 500 m, spanning both riverbanks at a depth of 1-3 m (Flotemersch et al., 2006). The number and biomass (weight mass) of fish individuals were collected as traditional fishing data.

2.3 Environmental DNA extraction and metabarcoding

Within 24 hr of collection, all water samples were filtered using Whatman glass microfiber filter papers (47 mm diameter, 1.2 μm pore size). Prior to filtering each river sample, 500 ml of ddH_2O was filtered on a separate filter to act as laboratory controls, followed by filtration of the river sample on new filters using the same filtration apparatus. DNA from water was extracted using the Qiagen DNeasy Tissue and Blood DNA extraction kit following the manufacturer's protocol with minor modifications. Three membranes (500 mL of water per membrane) for each sample were cut into pieces, ground and mixed. Then, the sample was soaked in 600 μL of 2 $\times$ lysis buffer and 40 μL of proteinase K. Incubation with this mixture was performed at 56 °C for 2.5 hr. Finally, we washed the filters in the mixture and performed elution in 200 μL of AE buffer. Filtration blanks and negative controls were coextracted alongside the samples and were subjected to the same protocol as the samples. The DNA concentration was determined using the Qubit dsDNA HS Assay Kit and detected in a 1.0% agarose gel. No data or bands were observed for the filtration blanks or negative controls.

Metabarcoding was performed in duplicate on each DNA extract with the primers MiFish-F

(5'GTCGGTAAAACTCGTGCCAGC-3') and MiFish-R (5'-CATAGTGGGGTATCTAATCCCAGTTTG-3') (Miya et al., 2015), which target the 12S rDNA region (amplifying an ˜180 bp region) of the mitochondrial genome, to identify fish species. DNA amplifications were performed in a two-step PCR protocol designed for the BGISEQ-500 platform. Three PCR replicates were performed for each sample. For each set of replications, environmental samples, filtration blanks and negative controls were included. The PCR assay volume was 50 μL, including 0.3 μL of Takara Ex Taq (5 U/μL), 5 μL of 10×Ex Taq buffer (20 mM Mg^{2+} plus), 4 μL of a dNTP mixture, 1 μL of the forwards and reverse primers (10 μM), 1 μL of the DNA template (environmental samples, filtration blanks and negative controls), and molecular biology-grade water added to 50 μL. For all samples, the first-step PCR was performed as follows: 94 °C for 5 min, followed by 10 cycles of 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 30 s, with a final extension at 72 °C for 10 min. The first-step PCR products were diluted 5 times with molecular biology grade water and used as the templates for the second-step PCR. For subsequent sequencing on the BGISEQ-500 platform, two or three random nucleotides were inserted into the MiFish primers (to increase sequence diversity during sequencing). The second-step PCR was carried out in a 50 μL reaction volume including 0.3 μL of Takara Ex Taq (5 U/μL), 5 μL of 10× Ex Taq buffer (20 mM Mg^{2+} plus), 4 μL of a dNTP Mixture, 1 μL of the forwards and reverse primers with the BGISEQ-500 adapter (10 μM), 2 μL of the template, and molecular biology-grade water added to 50 μL. The PCR procedure was as follows: 94 °C for 5 min, followed by 20 cycles of 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 30 s, with a final extension at 72 °C for 10 min. After the two-step PCR amplification, the PCR products were detected in a 1.5% agarose gel. None of the filtration blanks or negative controls showed amplification.

2.4 Library construction and sequencing

The PCR products showing the target bands were mixed in equal amounts, followed by electrophoresis in a 2% agarose gel and gel cutting. The PCR products were purified using a QIAquick Gel Extraction kit. Sixty nanograms of the purified PCR products was denatured at 95 °C and ligated with T4 DNA ligase at 37 °C to generate a single-stranded circular DNA library. Library replicates and library blanks were simultaneously included throughout the process. The concentrations and fragment size distributions of the libraries were checked on an Agilent 2100 Bioanalyzer. All libraries were subsequently pooled in equal amounts to generate DNA nanoballs (DNBs). Each DNB was pooled into 1 lane for sequencing. Sequencing was performed via 150 bp paired-end sequencing on the BGISEQ-500 high-throughput platform.

To obtain clean reads, the raw data were filtered to eliminate adapter contamination and reads of low quality. Then, paired-end reads were combined with tags based on overlaps using FLASH (Magoč and Salzberg, 2011). The tags were clustered into OTUs (operational taxonomic units) using USEARCH with a 97% threshold, and chimaera were filtered out using UCHIME (Rognes et al., 2016). All tags were mapped to each representative OTU sequence using USEARCH to obtain the OTU richness table. The taxonomic assignment of OTU sequences was mapped to the NT database downloaded from the National Center for Biotechnology Information (NCBI) GenBank database using the Blastn tool. Sequences were designated as belonging to a species if there was [?] 96% sequence identity to the NT and NCBI database barcode across the entire length of the amplicon, if a sequence from at least one other species within the same genus was available for comparison (and $< 96\%$ identical). An OTU was categorized into another OTU if a sequence could not be assigned to a species. If a sequence could be assigned to several species ([?] 99% matching rate) and the species belonged to the same genus, the taxonomic resolution collapsed to the genus level.

2.5 Data processing and statistical analysis

All statistical analyses were conducted in R 4.2.3 (R Core Team, 2021) with the primary packages *cluster* , *factoextra* ,*phyloseq* , *vegan* , and *ggplot2* . Nonmetric multidimensional scaling (NMDS) was used to group the 38 sampling sites into spatial zones with distinct differences in fish community composition according to the relative OTU richness, fish individual number and biomass. NMDS relies on the rank order of pairwise variable dissimilarities (Euclidean distance in this study) and does not make any underlying distributional assumptions of the data (Borcard et al., 2011). Sampling sites were plotted in ordination space with the distance between points positively related to the dissimilarity of output parameters (i.e., sites with similar

output parameters were plotted closer to one another). The analysis of similarity (ANOSIM) test was used to evaluate the dissimilarity matrix and test whether groups of objects had significantly ($P< 0.05$) different mean dissimilarities.

Based on the OTU richness, the pairwise taxonomic Bray-Curtis dissimilarity matrix between different samples was calculated using the*microeco* package (Liu et al., 2021). Environmental factors and fish OTU richness that showed significant variations in their values were used, and stepwise forwards selection was performed to linearly reduce the correlated variables along the axes. A permutation limit (with a P value of 0.05) was used to determine which variables to incorporate into the final model. The relationship between eDNA-based and number-/biomass-based alpha diversity was estimated by linear regression. Linear dependencies were explored by computing the variable variance inflation factors to ensure no confounding colinearity. The statistical significance of the axes derived from each analysis was tested with a Monte Carlo test (999 permutations).

Linear discriminant analysis effect size (LEfSe) is an algorithm for high-dimensional indicator discovery that identifies taxa by characterizing the differences between two or more biological conditions (Segata et al., 2011). LEfSe emphasizes both statistical significance and biological relevance, allowing researchers to identify discriminative features that are significantly different between biological classes. The nonparametric factorial Kruskal-Wallis sum-rank test was first used to detect features with significant differential abundance with respect to the class of interest. Second, LEfSe uses linear discriminant analysis to estimate the effect size of each differentially abundant feature and rank the feature accordingly (Liu et al., 2021).

Two-way hierarchical clustering analysis was performed using the*pheatmap* package. The package used clustering distances and methods implemented in the *dist* and *hclust* functions in R. The clustering analysis divided fish species with similar responses to the environmental factors into a group. Statistically significant cluster trees were identified using a bootstrap randomization technique in which the nonzero values were resampled and used to generate pseudovalues under the null hypothesis. The result was displayed as a heatmap.

3 Results

3.1 Spatial distribution pattern of fish communities along the river

There were 131 fish species belonging to 22 orders, 59 families, and 103 genera detected by the eDNA sampling protocol. There were 92 fish species belonging to 16 orders, 48 families, and 79 genera detected by electrofishing in situ. The fish species detected by eDNA across the 38 sampling sites contained all the fish species sampled by electrofishing, indicating that the capacity of eDNA to find species was stronger than that of electrofishing. NMDS results showed that, regardless of the data types (i.e., novel eDNA OTU richness and traditional biological counting), the relative percentage (%) of OTU richness (Fig. 2A), individual number (Fig. 2B), and biomass (Fig. 2C) could group the 38 sampling sites into eight spatial zones, including I (L1 - L2 and Z1 – Z3), II (L3 - L6), III (L7 - L11 and Z4), IV (L12 - L14 and P1 - P2), V (Z6 – Z7 and P3 - P4), VI (Z8 - Z11), VII (P5 - P8), and VIII (Z5 and P9 - P12). Along the river continuum, these zones were distributed in sections with heterogeneous habitats, including headwaters, upper stream, middle river, lower reaches, and river mouth. Our results suggested that eDNA could indicate the spatial distribution pattern of fish communities as well as the pattern reflected by electrofishing.

3.2 The composition of fish communities at different taxonomic levels

The relative percentage of eDNA OTU richness showed that, at the order level (Fig. 3A), Cypriniformes, Cichliformes, Siluriformes, and Gobiiformes made up $> 83\%$ of the fish communities in the three rivers. At the family level (Fig. 3B), Cichlidae, Cultrinae, Cyprinidae, and Labeoninae made up $> 56\%$ of the fish communities in Zones II - VIII, whereas Cyprinidae, Oxudercidae, and Heteropteridae made up $> 50\%$ of the fish communities in Zone I. At the genus (Fig. 3C) and species levels (Fig. 3D), the composition of fish communities was dispersed and determined by different genera and species. Notably, except for zones I and II, where native *Oryzias curvinotus*and *Hemiculter leucisculus* had the highest OTU richness, respectively,

exotic *Coptodon zillii* and *Oreochromis aureus* had the highest OTU richness in the other zones, indicating that the structure of fish communities in the middle and lower reaches was dominated by alien species. This result indicated that the eDNA metabarcoding protocol had advantages in discovering dominant species and exotic species.

The cluster analysis between the eight spatial zones showed that the order- and family-level grouping (Figs. 3A – 3B) was insufficient to distinguish the longitudinal differences in fish community composition from headwaters to the estuary. Compared with the order- and family-level grouping, the genus-level (Fig. 3C) grouping could more clearly distinguish the differences between upstream zones I - IV and downstream zones V - VIII. Moreover, the species-level grouping (Fig. 3D) was the most effective in distinguishing upstream zones I - III, midstream zones IV - V, and downstream zones VI – VIII. An interesting finding was that at the family level, upstream zone II and downstream zone VIII were clustered in the same group, which was caused by the widespread species of Cichlidae, Cultrinae, and Oxudercidae that could not be distinguished by family-level identification. In contrast, compared with the family level, the higher resolution at the genus and species levels could distinguish *Coptodon* and *Oreochromis* of Cichlidae as well as *Hemiculter* and *Parabramis* of Cultrinae, which were the key genera that distinguished fish distribution in the upper and lower reaches.

3.3 The diversity relationships between eDNA and electrofishing sampling

The linear regression results showed that the relationships between the alpha diversity calculated by eDNA OTU richness and fish individual number exhibited significant ($P < 0.05$) positive correlations (Fig. 4A). The regression equation is fitted as $Y = 0.9572[?]X + 0.5425$ ($R^2 = 0.8218$), where Y is the number-based diversity, X is the eDNA-based diversity, and$R$ -squared (R^2) measures how close the data points are to the fitted line. The linear regression between the fish diversity calculated by eDNA OTU richness and by fish biomass (i.e., wet weight) exhibited significant ($P < 0.05$) positive correlations. The regression equation is fitted as $Y = 0.4339[?]X + 2.3043$ ($R\ 2 = 0.4558$), where Y is the biomass-based diversity and X is the eDNA-based diversity (Fig. 4B). These results indicated that the eDNA metabarcoding protocol could reflect the alpha diversity of local fish communities as well as the diversity determined by traditional biological counting, and the eDNA-based diversity had a higher matching degree with number-based diversity than with biomass-based diversity. In addition,R^2 showed significant ($P < 0.05$) intergroup differences between the spatial zones (Fig. 4C). The highest R^2 values were observed in zones I - II and VII – VIII, which were significantly higher than those in zones IV - V. This result indicated that the eDNA-based diversity had higher linear regression with number-based diversity in headwater and the estuarial zones.

3.4 Biomarkers that determined site-specific fish communities

The OTU-based LEfSe selected 21 families, 34 genera, and 37 species as biomarkers (Fig. 4A). The number- and biomass-based LEfSe selected 15 families, 24 genera, and 27 species (Fig. 4B) as well as 22 families, 34 genera, and 37 species (Fig. 4C). No order-level biomarker was selected, indicating that the spatial variation in fish communities at the order level was not significant. The common biomarkers selected by LEfSe from eDNA and electrofishing (see Table 1) were Barbinae|*Acrossocheilus* |*A. beijiangensis* , Nemacheilidae|*Schistura* |*S. fasciolata* , Cyprinidae|Nicholsicypris|*N. normalis* , Danioninae|*Zacco* |*Z. platypus* , Siluridae|*Silurus*|*S. asotus* , Oxudercidae|*Rhinogobius* |*R. giurinusi* and*R. cliffordpopei* , Acheilognathinae|*Rhodeus* in zone I; *Culter* , *Rhodeus* |*R. ocellatus* ,*Hemiculter* |*H. leucisculus* ,*Opsariichthys* |*O. bidens* ,*Paramisgurnus* |*P. dabryanus* , *Clarias fuscus*in zone II; Gobioninae|*Squalidus* |*S. argentatus* , *Coptodon* |*C. zillii* in zone III;*Carassius* |*C. auratus* ,*Pseudorasbora* |*P. parva* , Gastromyzontidae|*Vanmanenia* |*V. pingchowensis* in zone IV; *Osteochilus* |*O. salsburyi* in zone V; Hypophthalmichthyinae|*Hypophthalmichthys* |*H. nobilis* and *H. molitrix* , Mugilidae|*Liza* |*L. subviridis* , Cichlidae|*Oreochromis* |*O. aureus* , Scaridae|*Hipposcarus* |*H. harid* , Salmonidae|*Salmo* |*S . salar* ,*Squaliobarbus* |*S. curriculus* in zone VII; Gobiidae|*Glossogobius* ,*Konosirus* , *Mugil* |*M. cephalus* , *Liza carinata* in zone VIII. Notably, there were no common biomarkers observed in zone VI, indicating a great difference in the composition of fish communities detected by eDNA and electrofishing.

An interesting finding was that the eDNA-based biomarkers were similar to the number-based biomarkers but greatly differed in biomass-based biomarkers. Generally, the biomass-based biomarkers that were not detected by eDNA were bottom-dwelling species with large body sizes, such as Xenocyprinae|*Xenocypris* |*X. argentea* , Leuciscinae|*Squaliobarbus* |*S. curriculus* , Loricariidae|*Pterygoplichthys* |*P. anisitsi* in zone IV, and Ictaluridae|*Ictalurus* |*I. punctatus* , Sparidae|*Rhabdosargus* |*R. haffara* in zone VII. Another finding was that the order-level biomarkers appeared more in headwater zone I and downstream zones VII-VIII than in other zones, indicating a habitat-specific distribution of these order-level biomarkers. For example, in zone I, Barbinae|*Acrossocheilus* , Danioninae|*Zacco* , Siluridae|*Silurus* , and Oxudercidae|*Rhinogo bius*were typical rheophilic fish living in rapids and riffles with widespread cobble substrates. In zone VII, Hypophthalmichthyinae|*Hypophthalmichthys* , Mugilidae|*Liza* , Cichlidae|*Oreochromis* , Scaridae|*Hipposcarus* , and Salmonidae|*Salmo* preferred slow-flowing and open water bodies. In zone VIII, the appearance of Serrasalmidae|*Piaractus* , Clupeidae|*Konosirus* , Gobiidae|*Glossogobius* , and Mugilidae|*Mugil* indicated brackish waters near the estuary. Moreover, *Coptodon zillii* and *Labeo rohita*in zone III, *Pterygoplichthys pardalis* in zone V,*Oreochromis aureus* in zone VII, and *Piaractus brachypomus*in zone VIII were typical invasive species that could be selected as biomarkers, indicating the sensitivity of eDNA in detecting invasive species.

3.5 Relationships between environmental factors and fish community composition

The correlation analysis between environmental factors (see Table S1 in the Supplementary Material) and the composition of fish communities showed that 44 (Fig. 6A), 41 (Fig. 6B), and 40 (Fig. 6C) species had significant relationships with environmental factors according to the eDNA, individual number, and biomass data types, respectively. In addition, for eDNA-based fish community composition, at the order, family, genus, and species levels, there were 5, 18, 41, and 44 units, respectively, that significantly ($P < 0.05$) correlated with the environmental factors (see Fig. S1 in the Supplementary Material). This result indicated that the species-level identification of the fish community compositions had the strongest response to the environmental changes, followed by genus-level identification. Interestingly, regardless of input data types, there were two environmental gradients that significantly influenced the longitudinal distribution of fish communities. The first environmental gradient was mostly composed of physical water parameters and habitat factors, including elevation (m), dissolved oxygen (mg/L), flow velocity (m/s), riffle area (%), and vegetation cover (%). The high values of these factors indicated pristine habitats in the headwaters and upper reaches, where there was less human disturbance. The second environmental gradient was mostly composed of the chemical water parameters (e.g., high COD_{Mn} [mg/L] and BOD_5 [mg/L]), conductivity (μs/cm), bacteria levels (total number of bacteria and coliform), and heavy metals (e.g., Hg, As, and Cu [mg/L]). The high values of these factors indicated water pollution and extraneous interference that commonly occurred in urban and industrial areas of the lower reaches. These results indicated that eDNA-based analysis of fish community composition could reflect the longitudinal changes in environmental factors along the river, which was similar to the pattern reflected by individual number and biomass.

4 Discussion

The application of nondestructive techniques to monitor and characterize fish diversity has drawn increasing worldwide attention for assessing aquatic ecosystem health. Given that it is difficult and costly to investigate fish diversity in large rivers, finding a new monitoring method with applicable technology has drawn great attention. Negative impacts of increasing anthropogenic activities on aquatic organisms, especially fish, have been verified in the Pearl River and other rivers and reservoirs connected with it (Wang et al., 2020b; Wang et al., 2019a; Wang et al., 2018a; Wang et al., 2019b; Wang et al., 2021a). To protect and maintain the health of subtropical river systems, we investigated the fish diversity along a river continuum using a combination of electrofishing and eDNA metabarcoding. Our results presented evidence that eDNA can be used to monitor fish diversity in this kind of aquatic ecosystem.

4.1 The eDNA-based distribution and composition of fish communities in rivers

The fish species detected by eDNA across the 38 sampling sites contained all the fish species sampled by electrofishing, indicating that the capacity of eDNA to find species was stronger than that of electrofishing

(also see Balasingham et al. (2018; Beng and Corlett (2020). Generally, the extra species detected by eDNA were the species that have been reported in the Pearl River system but were not distributed in the designed sampling sites of this study. This indicated that on the one hand, eDNA could trace the DNA released by fish species that were difficult to sample by traditional methods; on the other hand, the DNA of these hard-catch or rare fish species might be carried by upstream or tributary waters and then collected at downstream or mainstream sites. Therefore, eDNA in water reflected more information on species distribution than the traditional method. However, eDNA contains more extraneous species information than traditional in situ field monitoring, which confuses site-specific assessments of ecological health or biological integrity. In fact, the similar spatial distribution pattern shown by NMDS and the high regression coefficient between eDNA- and number-based diversity indicated that eDNA could reveal the core information on the composition and diversity of fish communities as comprehensively as those reflected by individual number and biomass collected by traditional in situ investigation. Therefore, we suggest that researchers and managers use eDNA as an ecological tool to identify fish community structure and associated environmental factors.

4.2 Appropriate threshold of taxonomic classification for evaluating fish diversity

Compared with aquatic invertebrates, fish have more comprehensive DNA information in the gene bank, and thus, using eDNA metabarcoding to distinguish fish diversity has broad application prospects. Wang et al. (2023) suggested that in complex habitats, such as fresh and brackish converging lakes, analysing the invertebrate composition or diversity at the phylum or class level (at least to family) is effective enough to reflect the environmental properties. The great difference between eDNA usefulness for fish and invertebrates is the structure of primary taxonomic units, which is much simpler for fish than for invertebrates. At the class or order level, different invertebrate assemblages had totally different ecological properties (e.g., freshwater vs. brackish or littoral vs. limnetic); however, for fishes, order-level identification could not distinguish the habitat-specific ecological properties. Thus, according to our results, we recommend that an appropriate threshold for eDNA-based fish monitoring is at the genus or species level, which could be further chosen according to the monitoring or research targets. For example, if users focus on the influence of environmental changes on fish community structure, genus-level identification with eDNA is effective; if users focus on the fish distribution, species-level identification might be favourable.

4.3 Biomarkers that can distinguish the differences between fish communities

The LEfSe results showed that biomarkers selected by both eDNA and electrofishing exhibited similarity in headwaters and lower reaches but differences in middle reaches. eDNA could identify the fishes of Danioninae, Siluridae, Nemacheilidae, Oxudercidae, Acheilognathinae, and Barbinae, indicating that eDNA has a high degree of recognition for rheophilic fish living in headwater habitats. With widespread rapids and riffles, rushing waters in headwaters had a great perturbation effect on DNA remaining in the environment, especially for those deposited in the sediments. Thus, the eDNA samples collected in rapids (e.g., sites L1 – L2 in the Liuxi River and Z1 – Z3 in the Zeng River) contained more species OTU information than the eDNA samples collected in slow-flowing or limnetic waters of the middle and lower reaches.

An interesting finding was the difference between eDNA-based biomarkers and biomass-based biomarkers. The biomarkers selected by LEfSe through relative biomass preferred to screen out species with larger body sizes, such as Xenocypris, Squaliobarbus, and *Pterygoplichthys* in zone IV and *Ictalurus* and *Rhabdosargus* in zone VII. This is a problem that limits the promotion and application of eDNA in a wider research area. As suggested by Rourke et al. (2022), key influencing biotic factors on eDNA effects included the taxon examined as well as their body size, distribution, reproduction, and migration. Nevertheless, there is considerable evidence to support using eDNA as an ancillary tool for assessing fish population abundance and/or biomass across discrete spatiotemporal scales, following preliminary investigations to determine species and context-specific factors that influence the eDNA abundance/biomass relationship (Bylemans et al., 2019; Doi et al., 2017). Advantages of eDNA monitoring relative to other approaches include reduced costs, increased efficiencies, and nonlethal sampling (Stewart, 2019).

4.4 Use of fish diversity to evaluate the water quality

Considering different monitoring targets, the use of eDNA-based fish monitoring could be summarized into three goals: 1) water pollution that may reduce fish diversity in water bodies, 2) evaluation of the effects of serious environmental events on local aquatic communities, and 3) tracing the target species such as indicator, protected, and endangered fish in the river system. Our results demonstrated that the environmental influence on fish distribution and composition analysed by the eDNA method showed a similar tendency to that analysed by the traditional method (e.g., individual number and biomass), suggesting that eDNA is a useful tool for monitoring biological communities in the field. In this study, DO, velocity, salinity, depth, COD_{Mn} (or BOD_5), and the bacteria amount in the water were the key environmental factors that influenced the fish communities. In fact, given that fishes are not only the crucial components of the local food web, but also the high-level predators that exhibit top-down control effects, key environmental factors that influence fish distribution and composition may also influence food web properties (e.g., ominivory, connectivity and stability; see Wang et al., 2018a; Wang et al., 2019b; Wang et al., 2021a). Our further research will focus on the eDNA-based relationships between the attributes of fish communities and those of invertebrate communities as well as on the structure and functioning of aquatic food webs.

5 Conclusions

Our study used eDNA macrobarcoding technology to investigate the distribution and composition of fish communities in subtropical river systems. eDNA-based fish investigation was as effective as electrofishing-based investigation in both composition and diversity analysis. The relative OTU richness (%) at the taxon-specific level could effectively distinguish the dominance of fish communities. The composition of fish communities in each spatial zone could be indicated by local biomarkers, which could identify the specificity of the local environment in each spatial zone. The response between environmental factors and genus- and species-level OTU richness was effective enough to indicate the ecological relationships between the environment and fish communities. Macrobarcoding-based eDNA technology has great potential to be used in field monitoring of fishes, and even extend to invertebrates and the food web.

ACKNOWLEDGEMENTS

This research was financially supported by Key R&D Program of Hainan Province (Nos. ZDYF2022SHFZ034, ZDYF2022SHFZ032, and ZDYF2021SHFZ064), National Key R&D Program of China (No. 2022YFD2401301), Natural Science Foundation of Hainan Province (Nos. 421QN196 and 421QN195), Open Project of State Key Laboratory of Marine Resource Utilization in South China Sea (Nos. MRUKF2023005 and MRUKF2023002), Hainan University Start-up Funding for Scientific Research (Nos. KYQD[ZR]-21015 and KYQD[ZR]-21033), the Program of Hainan Association for Science and Technology Plans to Youth R&D Innovation (No. QCXM201914), Guangdong Water Conservancy Science and Technology Innovation Project (2020-26), and Research Project of China Water Resources Pearl River Planning Surveying & Designing Company (201904).

COMPETING INTERESTS

The authors declare that they have no competing interests.

DATA AVAILABILITY STATEMENT

All data supporting this study are provided as supporting information accompanying this manuscript.

REFERENCES

Balasingham K. D., Walter R. P., Mandrak N. E., Heath D. D. J. M. E., 2018. Environmental DNA detection of rare and invasive fish species in two Great Lakes tributaries. 27, 112-127.

Barbour M. T., Gerritsen J., Snyder B., Stribling J. Rapid bioassessment protocols for use in streams and wadeable rivers. Washington, USA: USEPA, 1999.

Beng K. C., Corlett R. T., 2020. Applications of environmental DNA (eDNA) in ecology and conservation: opportunities, challenges and prospects. Biodiversity and Conservation, 29, 2089-2121.

Bernos T. A., Yates M. C., Docker M. F., Fitzgerald A., Hanner R., Heath D., et al., 2023. Environmental DNA (eDNA) applications in freshwater fisheries management and conservation in Canada: overview of current challenges and opportunities. Canadian Journal of Fisheries and Aquatic Sciences, 80, 1170-1186.

Borcard D., Gillet F., Legendre P. Numerical ecology with R. New York, USA: Springer, 2011.

Bylemans J., Gleeson D. M., Duncan R. P., Hardy C. M., Furlan E. M., 2019. A performance evaluation of targeted eDNA and eDNA metabarcoding analyses for freshwater fishes. Environmental DNA, 1, 402-414.

Civade R., Dejean T., Valentini A., Roset N., Raymond J.-C., Bonin A., et al., 2016. Spatial representativeness of environmental DNA metabarcoding signal for fish biodiversity assessment in a natural freshwater system. PLOS ONE, 11, e0157366.

Doi H., Inui R., Akamatsu Y., Kanno K., Yamanaka H., Takahara T., et al., 2017. Environmental DNA analysis for estimating the abundance and biomass of stream fish. Freshwater Biology, 62, 30-39.

Flotemersch J. E., Stribling J. B., Paul M. J. Concepts and Approaches for the Bioassessment of Non-wadeable Streams and Rivers. Cincinnati, Ohio, USA: Office of Research and Development, USEPA, 2006.

Garlapati D., Charankumar B., Ramu K., Madeswaran P., Ramana Murthy M. V., 2019. A review on the applications and recent advances in environmental DNA (eDNA) metagenomics. Reviews in Environmental Science and Bio/Technology, 18, 389-411.

Hauer F. R., Lamberti G. A. Methods in Stream Ecology. London, UK: Academic Press, 2007.

Kumar G., Reaume A. M., Farrell E., Gaither M. R., 2022. Comparing eDNA metabarcoding primers for assessing fish communities in a biodiverse estuary. PLoS One, 17, e0266720.

Lacoursière-Roussel A., Côté G., Leclerc V., Bernatchez L., 2016. Quantifying relative fish abundance with eDNA: a promising tool for fisheries management. Journal of Applied Ecology, 53, 1148-1157.

Liu C., Cui Y., Li X., Yao M., 2021. microeco: an R package for data mining in microbial community ecology. FEMS Microbiology Ecology, 97, fiaa255.

Magoč T., Salzberg S. L., 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics, 27, 2957-63.

Miya M., Sato Y., Fukunaga T., Sado T., Poulsen J. Y., Sato K., et al., 2015. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. Royal Society Open Science, 2.

Nguyen B. N., Shen E. W., Seemann J., Correa A. M. S., O'Donnell J. L., Altieri A. H., et al., 2020. Environmental DNA survey captures patterns of fish and invertebrate diversity across a tropical seascape. Scientific Reports, 10, 6729.

Pont D., Rocle M., Valentini A., Civade R., Jean P., Maire A., et al., 2018. Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. Scientific Reports, 8, 10361.

Port J. A., O'Donnell J. L., Romero-Maraccini O. C., Leary P. R., Litvin S. Y., Nickols K. J., et al., 2016. Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. Mol Ecol, 25, 527-41.

Rognes T., Flouri T., Nichols B., Quince C., Mahé F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ, 4, e2584.

Rourke M. L., Fowler A. M., Hughes J. M., Broadhurst M. K., DiBattista J. D., Fielder S., et al., 2022. Environmental DNA (eDNA) as a tool for assessing fish biomass: A review of approaches and future considerations for resource surveys. Environmental DNA, 4, 9-33.

Sakata M. K., Watanabe T., Maki N., Ikeda K., Kosuge T., Okada H., et al., 2020. Determining an effective sampling method for eDNA metabarcoding: a case study for fish biodiversity monitoring in a small, natural river. Limnology.

Segata N., Izard J., Waldron L., Gevers D., Miropolsky L., Garrett W. S., et al., 2011. Metagenomic biomarker discovery and explanation. Genome Biology, 12, 1-18.

Shu L., Ludwig A., Peng Z., 2021. Environmental DNA metabarcoding primers for freshwater fish detection and quantification: In silico and in tanks. Ecology and Evolution, 11, 8281-8294.

Stewart K. A., 2019. Understanding the effects of biotic and abiotic factors on sources of aquatic environmental DNA. Biodiversity and Conservation, 28, 983-1001.

Stoeckle M. Y., Lyubov S., Zachary C. P., Hideyuki D., 2017. Aquatic environmental DNA detects seasonal fish abundance and habitat preference in an urban estuary. Plos One, 12, e0175186.

Wang S., Luo B.-K., Qin Y.-J., Zhao J.-G., Wang T.-T., Stewart S. D., et al., 2020a. Fish isotopic niches associated with environmental indicators and human disturbance along a disturbed large subtropical river in China. Science of The Total Environment, 750, 141667.

Wang S., Su L.-H., Luo B.-K., Qin Y.-J., Stewart S. D., Tang J.-P., et al., 2020b. Stable isotopes reveal effects of natural drivers and anthropogenic pressures on isotopic niches of invertebrate communities in a large subtropical river of China. Environmental Science and Pollution Research, 27, 36132-36146.

Wang S., Tang J.-P., Su L.-H., Fan J.-J., Chang H.-Y., Wang T.-T., et al., 2019a. Fish feeding groups, food selectivity, and diet shifts associated with environmental factors and prey availability along a large subtropical river, China. Aquatic Sciences, 81, 31.

Wang S., Wang L., Chang H.-Y., Li F., Tang J.-P., Zhou X.-A., et al., 2018a. Longitudinal variation in energy flow networks along a large subtropical river, China. Ecological Modelling, 387, 83-95.

Wang S., Wang L., Zheng Y., Chen Z.-B., Yang Y., Lin H.-J., et al., 2019b. Application of mass-balance modelling to assess the effects of ecological restoration on energy flows in a subtropical reservoir, China. Science of The Total Environment, 664, 780-792.

Wang S., Wang T.-T., Lin H.-J., Stewart S. D., Cheng G., Li W., et al., 2021a. Impacts of environmental factors on the food web structure, energy flows, and system attributes along a subtropical urban river in southern China. Science of The Total Environment, 794, 148673.

Wang S., Wang T.-T., Tang J.-P., Wang L., Yang Y., Lin H.-J., et al., 2018b. Longitudinal variation in fish prey utilization, trophic guilds, and indicator species along a large subtropical river, China. Ecology and Evolution, 8, 11467-11483.

Wang S., Wang T.-T., Xia W.-T., Chen Z.-B., Stewart S. D., Yang F.-J., et al., 2021b. Longitudinal pattern of resource utilization by aquatic consumers along a disturbed subtropical urban river: Estimating the relative contribution of resources with stable isotope analysis. Ecology and Evolution, 11, 16763-16775.

Wang T.-T., Wang X.-D., Wang D.-Y., Fan S.-D., Wang S., Chen Z.-B., et al., 2023. Aquatic invertebrate diversity profiling in heterogeneous wetland habitats by environmental DNA metabarcoding. Ecological Indicators, 150, 110126.

Yao M., Zhang S., Lu Q., Chen X., Zhang S. Y., Kong Y., et al., 2022. Fishing for fish environmental DNA: Ecological applications, methodological considerations, surveying designs, and ways forward. Mol Ecol, 31, 5132-5164.

Zhang S., Lu Q., Wang Y., Wang X., Zhao J., Yao M., 2020. Assessment of fish communities using environmental DNA: Effect of spatial sampling design in lentic systems of different sizes. Molecular Ecology Resources, 20, 242-255.

Zou K., Chen J., Ruan H., Li Z., Guo W., Li M., et al., 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. Science of the Total Environment, 702, 134704.

Hosted file

`Table.docx` available at https://authorea.com/users/646453/articles/658361-application-of-environmental-dna-metabarcoding-to-identify-fish-community-characteristics-in-subtropical-river-systems

Hosted file

`Figure.docx` available at https://authorea.com/users/646453/articles/658361-application-of-environmental-dna-metabarcoding-to-identify-fish-community-characteristics-in-subtropical-river-systems