

Research article

New insights into biologic interpretation of bioinformatic pipelines for fish eDNA metabarcoding: A case study in Pearl River estuary

Zhuoying Li ^a, Wencheng Zhao ^a, Yun Jiang ^a, Yongjing Wen ^a, Min Li ^{b, **}, Li Liu ^a, Keshu Zou ^{a,*}

^a Joint Laboratory of Guangdong Province and Hong Kong Region on Marine Bioresource Conservation and Exploitation, College of Marine Sciences, South China Agricultural University, 510642, Guangzhou, China

^b Key Laboratory for Sustainable Utilization of Open-sea Fishery, Ministry of Agriculture and Rural Affairs, South China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou, 510300, China

ARTICLE INFO

Handling editor: Jason Michael Evans

Keywords:

eDNA metabarcoding
Fish community
Bioinformatic pipelines
Distance matrix
Comparative evaluation

ABSTRACT

Environmental DNA (eDNA) metabarcoding is an emerging tool for monitoring biological communities in aquatic ecosystems. The selection of bioinformatic pipelines significantly impacts the results of biodiversity assessments. However, there is currently no consensus on the appropriate bioinformatic pipelines for fish community analysis in eDNA metabarcoding. In this study, we compared three bioinformatic pipelines (Uparse, DADA2, and UNOISE3) using real and mock (constructed with 15/30 known fish) communities to investigate the differences in biological interpretation during the data analysis process in eDNA metabarcoding. Performance evaluation and diversity analyses revealed that the choice of bioinformatic pipeline could impact the biological results of metabarcoding experiments. Among the three pipelines, the operational taxonomic units (OTU)-based pipeline (Uparse) showed the best performance (sensitivity: 0.6250 ± 0.0166 ; compositional similarity: 0.4000 ± 0.0571), the highest richness (25–102) and minimal inter-group differences in alpha diversity. It suggested the OTU-based pipeline possessed superior capability in fish diversity monitoring compared to ASV/ZOTU-based pipeline. Additionally, the Bray-Curtis distance matrix achieved the highest discriminative effect in the PCoA (43.3%–53.89%) and inter-group analysis ($P < 0.01$), indicating it was better at distinguishing compositional differences or specific genera of fish community at different sampling sites than other distance matrices. These findings provide new insights into fish community monitoring through eDNA metabarcoding in estuarine environments.

1. Introduction

Fish diversity is closely linked to food web stability and the overall functioning and balance of ecosystems, making it crucial for monitoring the health of aquatic ecosystems (Guzman et al., 2022; Sales et al., 2021; Zhou et al., 2022b). A comprehensive understanding of fish diversity is vital for effective fisheries and ecosystem management. Commonly, traditional ecological methods, such as direct underwater visual observation, trawling, and electric fishing, have been commonly employed to monitor fish diversity (Dugal et al., 2022; Goutte et al., 2020). However, limitations and drawbacks of traditional methods have been restricted their applications. For instance, the process of sampling is inherently complex, costly, biased, and often invasive (e.g., detrimental effects on ecosystems, fish and other organism), posing significant challenges in

capturing low-abundance and elusive species (Couton et al., 2023; Elliott et al., 2019; Keck et al., 2022). Additionally, species identification in traditional investigations requires specialized taxonomic expertise (Dugal et al., 2022; Pukk et al., 2021). Consequently, it is urgently need for new, non-invasive, efficient, and cost-effective approaches to replace traditional methods in fish diversity monitoring.

With the emergence of high-throughput sequencing, environmental DNA (eDNA) metabarcoding has become a popular and efficient tool for investigating biodiversity in various ecosystems, including estuaries, coastlines, wetlands, freshwater environments, and so on (Bylemans et al., 2018; Deiner et al., 2017; DiBattista et al., 2022). By providing information on fish community composition and diversity, eDNA metabarcoding enables the ecological assessment of aquatic environments (Handley et al., 2023). Moreover, it has the potential to serve as an

* Corresponding author.

** Corresponding author.

E-mail addresses: limin@scsfri.ac.cn (M. Li), zoukeshu@scau.edu.cn (K. Zou).

newly method for investigating relative species abundance and quantifying fish diversity by leveraging the positive correlation between species abundance and read abundance (García-Machado et al., 2022; Handley et al., 2019; Li et al., 2021; Zou et al., 2020). Compared to traditional ecological surveys, eDNA metabarcoding overcomes some limitations of traditional methods and provides a broader and higher-throughput monitoring approach with more comprehensive diversity information, as well as higher efficiency and greater detection accuracy, even it also requires a certain level of taxonomic expertise (Beng and Corlet, 2020; Keck et al., 2022; Ruppert et al., 2019; Zhou et al., 2022a). The eDNA technology allows for the collection of fish diversity data in aquatic ecosystems without invasive sampling, making it a promising and rapidly developing innovative technology (Couton et al., 2023; Takahashib et al., 2023).

Currently, the protocols used for fish diversity assessment through eDNA metabarcoding are complex, lack standardization and uniformity. The workflows for fish eDNA metabarcoding encompass various steps, including sampling, eDNA extraction, library preparation, high-throughput sequencing, and bioinformatic pipelines (Andres et al., 2023; Mathon et al., 2022; Ruan et al., 2022). Each step in these workflows can significantly influence biodiversity outcomes (Joos et al., 2020; Zaiko et al., 2022). One of the emerging challenges in eDNA metabarcoding is the analysis and interpretation of data, with a particular focus on the bioinformatic analysis of sequence data (DiBattista et al., 2022). Different bioinformatic pipelines can be utilized for specific bioinformatic analyses (Brandies and Hogg, 2021). For example, Operational Taxonomic Units (OTUs) are required to cluster the effective sequences of the samples in order to facilitate the analysis of species composition diversity information. And clean data are clustered into OTUs by using a default sequence similarity threshold (such as 97%), and each cluster is assigned a random identifier (OTU ID). Usually, the bioinformatic pipeline for OTU based on minimum sequence similarity to cluster sequences (Uparse algorithm) has been widely used in eDNA metabarcoding data processing. It helps mitigate overestimation of diversity or noise caused by sequencing errors and artifacts, limiting errors from experimental workflows or bias from intragenomic variations (Siddiqui et al., 2022; Sokal and Crovello, 1970; Xiong et al., 2022).

Sequencing errors in OTU-based pipeline can lead to misclassification, nested sequences, and expansion of taxonomic number (Kachroo et al., 2021; Poncheewin et al., 2020; Skelton et al., 2023). To address the limitations, denoising algorithms have been developed to mitigate the impact of sequencing errors on data analysis (Brandt et al., 2021; Kachroo et al., 2021; Moossavi et al., 2020; Zhou et al., 2021). Denoising algorithms preserve precise sequence variants based on sequence correction/denoising methods, resulting in improved results in terms of sensitivity and accuracy in correcting erroneous sequences (Liu et al., 2022; Needham et al., 2017). Unlike representative sequences among similar sequences, the denoising algorithm directly derives precise sequence variants from a biological entity (Joos et al., 2020; Modin et al., 2020; Siddiqui et al., 2022). DADA2 and UNOISE3 are also popularly used bioinformatic pipelines of denoising algorithms for assessing biodiversity in eDNA technology data analysis (Antich et al., 2021; Edgar, 2016a; Hakimzadeh et al., 2024; Tsuji et al., 2020). DADA2 algorithm achieves single-nucleotide resolution Amplicon Sequence Variant (ASVs) through sequence correction (Callahan et al., 2016). While the UNOISE3 algorithm utilizes the unoise3 command to denoise and output biologically meaningful sequences based on the Uparse algorithm, known as Zero-radius Operational Taxonomic Units (ZOTUs) (Edgar, 2016b). These denoising algorithms (DADA2 and UNOISE3) identify taxa with higher resolution through requiring high-quality sequencing information (García-García et al., 2019; Liu et al., 2022). Simultaneously, it will lead to a reduction in the number of detected taxa and then underestimation on the correlation between the community composition and its associated factors (e.g., environmental factors). Thus, the choice of bioinformatic pipelines has a significant impact on the biological interpretation and outcomes of eDNA metabarcoding

experiments (Joos et al., 2020; Moossavi et al., 2020).

Most studies on bioinformatic pipelines have focused on microbial communities, with few comparative studies conducted on fish communities or at the clustering or denoising level using mock and real sequencing data (Dal Pont et al., 2021; Kachroo et al., 2021; Mathon et al., 2021; Reitmeier et al., 2021; Xiong et al., 2022). Therefore, it is crucial to conduct comparative studies on bioinformatic pipelines in fish eDNA metabarcoding to enhance the accuracy and reproducibility of taxonomic assignments and the recovery of species abundance (Mathon et al., 2022; Miya, 2022; Pauvert et al., 2019). In this study, we collected fish eDNA metabarcoding data based on mock community and real sequences from samples of the Pearl River Estuary (PRE), an estuary disturbed by highly anthropogenic activities. In order to identify the most suitable bioinformatic pipeline for analyzing fish eDNA metabarcoding data, we employed three bioinformatic pipelines (Uparse, DADA2, and UNOISE3) to analyze the composition and diversity of the fish community. And the similarities and differences in the abilities of the pipelines to monitor taxonomy, alpha and beta diversity of the fish community were evaluated. Additionally, different bioinformatics analysis workflows are not entirely consistent in their suitability for distance matrices. Thus, the effectiveness in separating fish community compositions based on different bioinformatics pipelines using three distance matrices was assessed.

2. Material and methods

2.1. Sampling collection and DNA extraction

The Pearl River Estuary (PRE) is an estuarine ecosystem located in the southern region of the Tropic of Cancer in China. It is not only crucial in the northern South China Sea but also one of the most significantly human-impacted estuaries globally (Lin et al., 2020; Pan et al., 2023). In recent years, the PRE has experienced a decline in overall fish resources, including a reduction in species richness, density, and biomass (Zeng et al., 2022; Zhou et al., 2019). This study specifically focused on four entrances within the Pearl River: Humen (HU), Jiaomen (JM), Hongqimen (HQ), and Jitimén (JT). These entrances are situated between latitudes 22°00' to 22°48' N and longitudes 113°24' to 113°41' E. Environmental DNA (eDNA) samples were collected from August to November in 2019, specifically on non-rainy days. A total of 21 samples were obtained from the four estuaries, with each estuary comprising three sampling sites. To ensure accuracy, three duplicate samples were collected per individual site, with each sample consisting of 3 L of surface water collected along the coast. Additionally, prior to sampling at each location, the water carriers and sterile plastic bottles are soaked in 10% bleach solution (final concentration of 1% sodium hypochlorite) for at least 30 min. In each site, samples and sterile double steaming water (as blank control for avoiding contaminations in sampling process) were filtered in the field using 0.45 µm nitrocellulose filters (Xinya, Shanghai, China). The water filtration was done with dedicated clean consumables, and filters were stored and transported in liquid nitrogen until eDNA extraction. Then, TIANamp Marine Animals DNA Kit (Tiangen Biotech) was utilized to extract eDNA, with negative controls and blanks extracted alongside the samples. The following amplification and experimentations were followed the methodology outlined in Li et al. (2023) and Zou et al. (2020). All the laboratory experiments were done in a clean and specialized laboratory for eDNA research.

2.2. Mock community

We conducted simulations using four different fish assemblages using gillnetting, which were collected from the same sites as the eDNA sampling. These simulations, referred to as "mock communities," included fish species that had been previously identified in literature surveys conducted in the PRE (Jiang et al., 2022; Li et al., 2023; Zou et al., 2020). Each mock community sample consisted of either 15 fish

individuals (group: A/C) or 30 fish individuals (B/D), with detailed information provided in Table S1. The composition of mock community included randomly selected species from various orders, families, and genera. Genomic DNA was then extracted from each fish individual using the TIANamp Marine Animals DNA Kit (Tiangen Biotech, Beijing, China) in a laboratory that underwent thorough decontamination and sterilization procedures, including treatment with ultraviolet light and a 1% sodium hypochlorite for 30 min each.

Two types of fish communities were constructed for further experiments: a uniform community with equal molar concentrations (group: A/B; final concentration of 16 ng/ μ L) and a staggered community with abundances spanning four orders of magnitude (C/D; final concentrations of 20, 2, and 0.2 ng/ μ L). To ensure the robustness and reliability of the results, three biological replicates were included for each mock community, and the detailed construction process of each biological replicate could be found in Appendix 1.

2.3. Bioinformatic

The PCR amplifications in this study utilized eDNA and DNA from

mock communities as templates. eDNA and DNA of mock community as templates, specific primers (MiFish-F: 5'- GTCGGTAAACTCGTGC-CAGC-3'; MiFish-R: 5'- CATACTGGGGGTATCTAATCCCAGTTG-3') were employed for amplification (Miya et al., 2015), and three PCR technical replicates were conducted in each sample. Each amplification set included samples, negative controls, filter blanks, and positive controls (DNA templates extracted from several fish species captured at the same location). No amplification fragments were present in the filtering of blank and negative controls in this study. Subsequently, libraries were constructed in equal amounts and sequenced separately on the Illumina Nova6000 platform (Guangdong Magigene Biotechnology Co., Ltd. Guangzhou, China) to generate 250 bp paired-end reads. The specific details of the PCR amplifications and sequencing procedures were cited from a previous study (Li et al., 2023). Subsequently, the sequencing data (raw data) was analyzed using three bioinformatic pipelines. These pipelines differed primarily in four aspects: (1) paired-end sequence merging, (2) primer trimming and quality control, (3) treatment of chimeric sequences, and (4) construction of a sequence table. The detailed explanation of each step involved in the bioinformatic analysis can be found in Fig. 1.

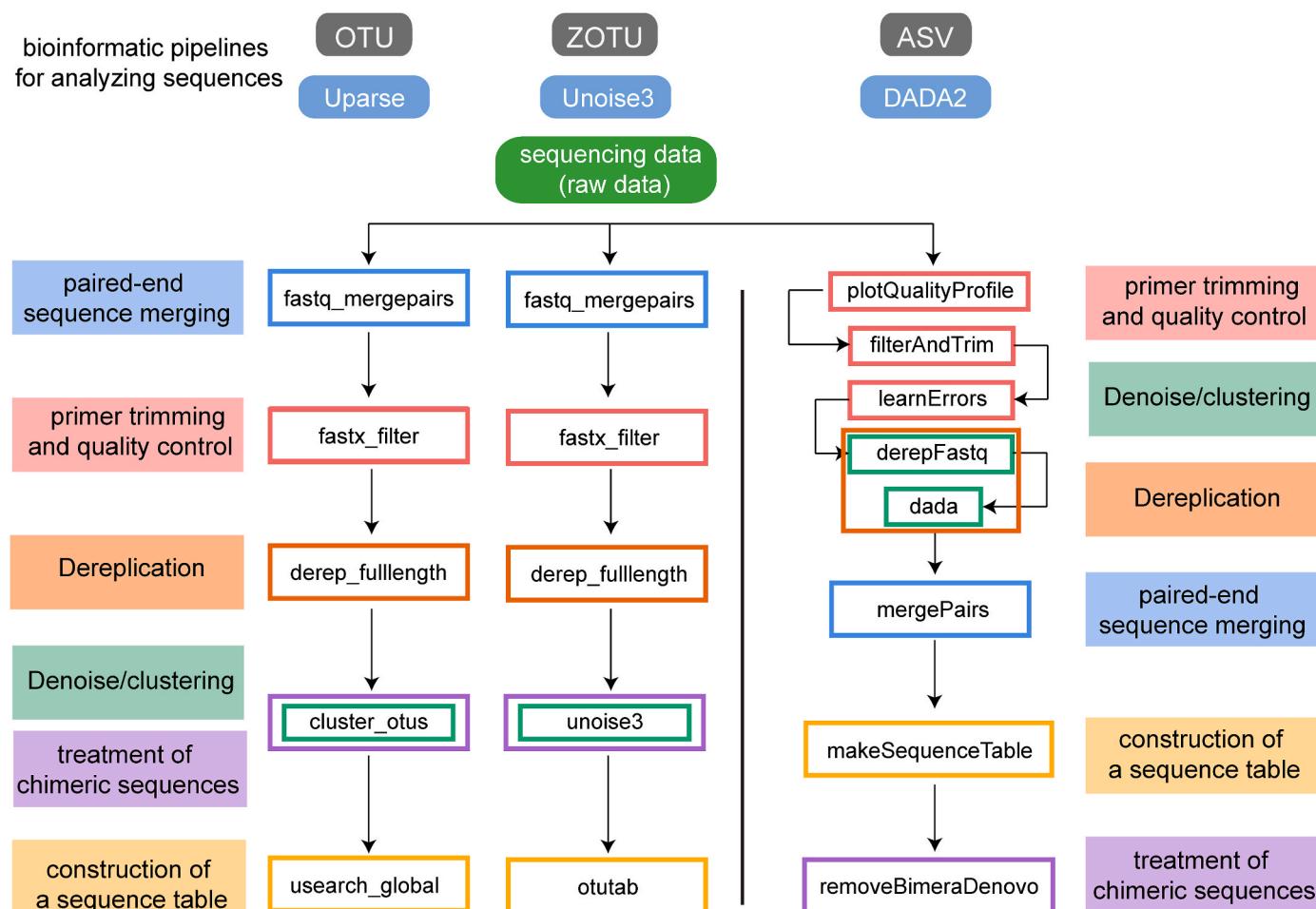


Fig. 1. Overview of the bioinformatic pipelines compared in the study, with each of the four steps represented by different colors. All bioinformatic pipelines were executed using R V3.6.3.

*paired-end sequence merging: refers to combining forward and reverse sequence files to obtain the complete sequence.

*primer trimming and quality control: refers to trimming the primers and adapters of the forward/reverse/merged sequences and conducting quality control based on default data.

*Dereplication: primarily involves removing duplicates in the sequences, identifying unique read sequences and their abundance.

*Denoise/clustering: Denoising focuses on removing sequencing errors and PCR errors, with DADA2 simultaneously de-replicating. Clustering mainly involves grouping sequences into OTUs based on their similarity to each other (97%).

*treatment of chimeric sequences: primarily involves removing sequences composed of different species obtained from sequencing.

*construction of a sequence table: generates a matrix of reads counts corresponding to samples and OTUs/ZOTUs/ASVs.

2.3.1. OTU (Uparse)/ZOTU (UNOISE3)

Uparse algorithm clustered sequences with a similarity greater than 97% (with the principle of clustering low-abundance sequences onto high-abundance sequences) to obtain OTUs. On the other hand, UNOISE3 was an optimized version of the Uparse algorithm that utilized a non-clustering denoising approach to obtain ZOTUs. In this study, sequencing data (raw data) were analyzed using OTU/ZOTU-based pipelines, both employing usearch with R v3.6.3 but utilizing different algorithms. The OTU-based pipeline utilized the Uparse for clustering algorithm, while the ZOTU-based pipeline employed the UNOISE3 for denoising algorithm. The initial steps of the two pipelines were identical and performed only once, which included sequence merging, primer removal, quality control, and sequence de-redundancy (Dereplication). Initially, the "fastq_mergelpairs" command merged the two-terminal sequences. Subsequently, primer removal and quality control were conducted using the "fastx_filter" command (fastq_maxee_rate was set to 0.01), in which the forward and reverse sequences were trimmed to exclude the regions containing primers and barcodes, specifically 21 and 27 bp, respectively. Finally, the "derep_fulllength" command (minuniquesize was set to 10) eliminated redundant sequences. The subsequent stage involved branching the pipeline into OTU clustering or ZOTU denoising. For OTU clustering, the sequences were clustered at a 97% similarity threshold using the "cluster_otsu" command. The resulting OTU table was generated using the "usearch_global" command. In contrast, for ZOTU denoising, the "unoise3" command directly performed denoising without clustering. The final step involved using the "otutab" command to generate the ZOTU table.

2.3.2. ASV (DADA2)

ASV was an operational unit obtained through sequence correction and denoising algorithm using the DADA2 (Callahan et al., 2016). Sequencing data (raw data) was analyzed using an ASV-based pipeline implemented in DADA2. The analysis followed the guidelines provided by the instructor, which can be found at <https://benjneb.github.io/dada2/tutorial.html>. Firstly, the sequence quality was checked using the "plotQualityProfile()" function. Secondly, the primer and barcode regions, which had a total length of 21 or 27 base pairs for the forward and reverse primers respectively, were trimmed using the "filterAndTrim()" command (parameter: maxN = 0, maxEE = c(2,2), truncQ = 2, rm.phix = TRUE, compress = TRUE, multithread = T). Thirdly, the error rate was estimated iteratively, and the error model was learned from the reference sample sequence until the learning model converged to the true error rate. Fourthly, the "derepFastqO" function was employed to remove duplicate sequences, and the "dadaO" function was used to further control the sequence quality based on the error model. Fifthly, the "mergePairsO" command was applied to merge the forward and reverse sequences, resulting in the generation of complete sequences. The merged sequence was considered valid only if the forward and reverse sequences overlapped by at least 20 bases and were identical in the overlap region. Sixthly, a sequence list was created using the "makeSequenceTableO" function. Finally, the "removeBimeraDenovoO" function (parameter: method = "consensus") was used to eliminate chimeric sequences. The resulting ASV table was then exported.

2.3.3. Allocation by category

To enhance the reproducibility of the experiment and reduce false positive rate, we verified the consistency of results through technical replicates and excluded species that appeared in only one technical replicate. Each unique OTU, ZOTU, and ASV was locally blasted against the non-redundant nucleotide database of the National Center for Biotechnology Information (NCBI). The initial screening of the blast results prioritized hits with the highest scores, and only OTUs, ZOTUs, and ASVs with a sequence identity ranging from 97% to 100% and sequence coverage ranging from 99% to 100% to fish species were retained. Any OTUs, ZOTUs, or ASVs that did not meet these criteria were excluded from subsequent analysis. During this step, 23 OTUs

(consisting of 908,148 reads), 23 ZOTUs (1,040,784 reads), and 30 ASVs (64,627 reads) could not be annotated to the species level. Following that, taxonomic information specific to fish species was annotated, and any fish species not found in the Pearl River, PRE, or South China Sea were removed from the dataset (Kuang et al., 2021; Ruan et al., 2022; Zou et al., 2020).

2.4. Performance evaluation for mock community

Mock communities were consolidated based on fish species and replicates after taxonomic assignment. To evaluate the performance of each bioinformatic pipeline in recovering mock communities, three criteria were defined: sensitivity, F-measure, and compositional similarity. Sensitivity and the F-measure are commonly used metrics for predicting the "accuracy" of measurements, which could assess the pipelines' ability to generate accurate species lists (Gardner et al., 2019; Lever et al., 2016). These two criteria offer complementary advantages, sensitivity focused on identifying missing rare taxa, while the F-measure emphasized the detection of false positives (Gardner et al., 2019). The number of false positives (FP), true positives (TP), and false negatives (FN) were calculated for each bioinformatics pipeline, and then derived the sensitivity (Equation (1)) and F-measure (Equation (2)) for each sample of mock community and biological replicate. FP referred to species present in the output data but absent in the sample of mock community, TP represented species present in both the sample of mock community and the output data, and FN denoted species present in the sample of mock community but not in the output data.

$$\text{sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (1)$$

$$F - \text{measure} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (2)$$

$$\text{compositional similarity} = 1 - \text{BC} \quad (3)$$

Compositional similarity (Equation (3)) was defined as the Bray-Curtis similarity between the recovered community and the mock community (Odum, 1950). It was computed as 1 minus the Bray-Curtis dissimilarity (Pauvert et al., 2019). This criterion assessed the level of difference between two abundance lists: a higher compositional similarity indicated a greater resemblance in relative abundance between the pipeline output and the mock community.

For each sample of mock community, the mean and standard error of the three criteria were calculated for each biological replicate. The mean and standard error between replicates represented the within-sample variability of a specific bioinformatic pipeline, whereas the mean and standard error across sample of mock community represented the between-sample variability of the results of each bioinformatics pipeline.

2.5. Statistical analysis

All data analyses were performed using R v3.6.3 (R Core Team, 2019). The alpha diversity, including the Shannon index, richness, and Pielou evenness index, was calculated using the diversity and spec-number function from the package "vegan" v2.5-2 (Oksanen et al., 2018). To determine the sequencing depth of the sample, the rarefaction curve was performed using rarecurve function from the package "vegan" v2.5-2. To mitigate the impact of outliers, it is necessary to standardize the data before conducting subsequent analyses (St-Pierre et al., 2018). Prior to the follow-up statistical analysis, the Hellinger transformation (using the adonis2 function from the package "vegan" v2.5-2) was applied to address the issue of double-zero values and minimize residual effects (Laporte et al., 2021; Oksanen et al., 2018; Zhang et al., 2021).

Histograms (package "ggplot2" v3.2.0) and upset maps (package "upsetR" v1.3.3) based on Hellinger-transformed abundance data were

utilized to visualize the fish community composition and species richness (Conway et al., 2017; Wickham, 2016). We performed Procrustes analysis using the procrustes function from the package "vegan" v2.5-2 to assess the consistency between two bioinformatics pipeline data. It compared the alignment of two datasets by minimizing the sum of squared differences between corresponding points (Gower, 1975; Zhao et al., 2019). To compare the distribution patterns of fish communities across different sites or samples, Principal Coordinates Analysis (PCoA) was conducted using the cmdscale function from the package "vegan" v2.5-2 (Oksanen et al., 2018). To assess any significant differences between all simulated instances and bioinformatics pipelines, Analysis of Similarities (ANOSIM) was performed using the "anosim" function from the package "vegan" v2.5-2 (Warton et al., 2012). A permutation variance multivariate analysis (PERMANOVA, 999 permutation, "adonis2" function from the package "vegan" v2.5-2) that explained the unbalanced, and multivariate design was used to test the significance of differences in the abundant genera (genera commonly detected with relatively high abundance) among different distance matrices the three bioinformatic pipelines (Anderson, 2001; Cholet et al., 2022). Three distance matrices were considered: Bray-Curtis distance (BC), calculated based on species abundance information; unweighted UniFrac (UU), incorporated taxonomic information but only considerd the presence or absence of species in the community; and Weighted UniFrac (WU), which not only incorporated taxonomic information but also took into account the presence or absence of species and their relative abundances in the community (Bray and Curtis, 1957; Lozupone et al., 2007; Maki et al., 2021). Note, we performed permutation multivariate dispersion (PERMDISP) analysis using the betadisper and permute functions from the package "vegan" v2.5-2 prior to conducting the PERMANOVA analysis. It allowed us to assess the multivariate homogeneity of dispersion within the analysis groups (Xu et al., 2019). Additionally, Mantel tests (parameters: correlation coefficient: Spearman, and the number of permutations: 1000) for similarity, implemented separately using the "linkET" package (implemented using the "mantel_test" function form the package "vegan" v2.5-2), were used to evaluate the correlation between fish community compositions obtained from the three bioinformatic pipelines and the BC, UU, and WU distance matrices (Huang, 2021).

3. Results

3.1. Performance evalution of mock and real communities

More than two million and over one million raw sequences were obtained from the real community of fish eDNA metabarcoding and mock communities, respectively, using three different pipelines (OTU, ZOTU, and ASV) (Tables S2, S3, and S4). For the real communities, the sequence counts were reduced by 6.55% in OTU, 14.09% in ZOTU, and 25.33% in ASV after quality filtering, denoising, merging, and chimera removal (Tables S2, S3, and S4). After processing, an average of 43,288 effective OTUs, 49,561 effective ZOTUs, and 3077 effective ASVs were generated per sample. Annotation of taxonomic information at the species level (excluding non-fish reads) revealed that the number of annotated sequences for the OTU, ZOTU, and ASV pipelines were 908,148, 1,040,784, and 64,627, respectively (Tables S2, S3, and S4). Finally, a total of 140 OTUs, 111 ZOTUs, and 109 ASVs were retained for further analysis of the real communities (Table 1). In contrast, the percentage of sequence reduction (after quality filtering, denoising, merging, and chimera removal) was lower in the mock communities compared to the real communities. OTU, ZOTU, and ASV sequence counts decreased by 4.97%, 15.07%, and 21.01%, respectively (Tables S2, S3, S4). A total of 31 OTUs (1,025,182 reads), 30 ZOTUs (1,376,625 reads), and 26 ASVs (1,235,798 reads) were retained for subsequent analysis (Table 1).

In the real communities, a total of 19 orders, 67 families, 113 genera, and 156 species were identified using the three pipelines (OTU, ZOTU,

Table 1

The numbers of taxa assigned at different levels were recorded for each pipeline.

Community	Level	OTU	ZOTU	ASV	Total
Real	Order	18	18	15	19
	Family	67	60	59	67
	Genus	108	91	90	113
	Species	140	111	109	156
Mock	Order	9	9	9	9
	Family	18	18	18	20
	Genus	24	26	24	31
	Species	31	30	26	41

The "Total" represents the taxonomic data obtained from all three bioinformatics pipelines. The "Mock" refers to the data from the mock community, while the "Real" refers to the data from the real community.

and ASV) (Table 1). Conversely, in the mock communities, 9 orders, 20 families, 31 genera, and 41 species were identified. Irrespective of the taxonomic level, the OTU-based pipeline yielded the highest number of identifications, while the ASV-based pipeline yielded the lowest. This pattern was consistent in both the real and mock communities. At the order, family and genus levels, there were no differences in the quantities obtained through the three pipelines ($P > 0.05$). However, significant differences were observed at the species levels (real: $P = 0.030$; mock: $P = 0.019$; Table 1).

No significant differences in compositional similarity were observed among the three pipelines when analyzing the mock community ($P = 0.100$, Table S5). In order to mitigate the potential impact of insufficient sequencing depth, we conducted rarefaction curve analysis on all replicate samples in the mock community. The results revealed that the curves for all samples eventually plateaued, indicating sufficient sampling and sequencing depth for the samples (Fig. S1). However, significant differences were found in sensitivity and F-measure between almost all pairwise comparisons in the pipelines, except for the F-measure between OTU and ASV (Table S5). The OTU-based pipeline exhibited the highest sensitivity (0.625 ± 0.017), F-measure (0.628 ± 0.014), and compositional similarity (0.400 ± 0.057). These results indicate that the OTU-based pipeline had the fewest missing species, the lowest false positives, and the strongest similarity in composition with the mock community (Table 2). On the other hand, the ZOTU-based pipeline demonstrated the lowest sensitivity (0.500 ± 0.039) and F-measure (0.422 ± 0.128) in classification assignment, suggesting it had the most missing species and the highest false positives. Lastly, the ASV-based pipeline showed the lowest compositional similarity (0.057 ± 0.000), indicating a weak similarity in composition with the mock community.

3.2. Species composition in the real and mock fish community

The real fish communities collectively shared a total of 81 genera as identified by the three bioinformatic pipelines (Fig. 2a and b). Among these, only 14 genera were detected in any two pipelines, while a single pipeline exclusively detected 18 genera in the real community. The shared genera accounted for 99.36%, 99.79%, and 97.03% of the OTU, ZOTU, and ASV sequences, respectively (Fig. 2c and d). A similar pattern was observed in the mock communities, where 12 species were shared among the three pipelines, accounting for 79.12% (OTU), 99.40% (ZOTU), and 77.80% (ASV) of the sequences, respectively. Generally,

Table 2

Comparative analysis of performance evalution indices for three bioinformatic pipelines utilizing mock communities.

Pipeline	sensitivity	F-measure	compositional similarity
OTU	0.6250 ± 0.0166	0.6276 ± 0.0144	0.4000 ± 0.0571
ZOTU	0.5000 ± 0.0385	0.4222 ± 0.1283	0.1714 ± 0.0989
ASV	0.5862 ± 0.0000	0.6182 ± 0.0000	0.0571 ± 0.0000

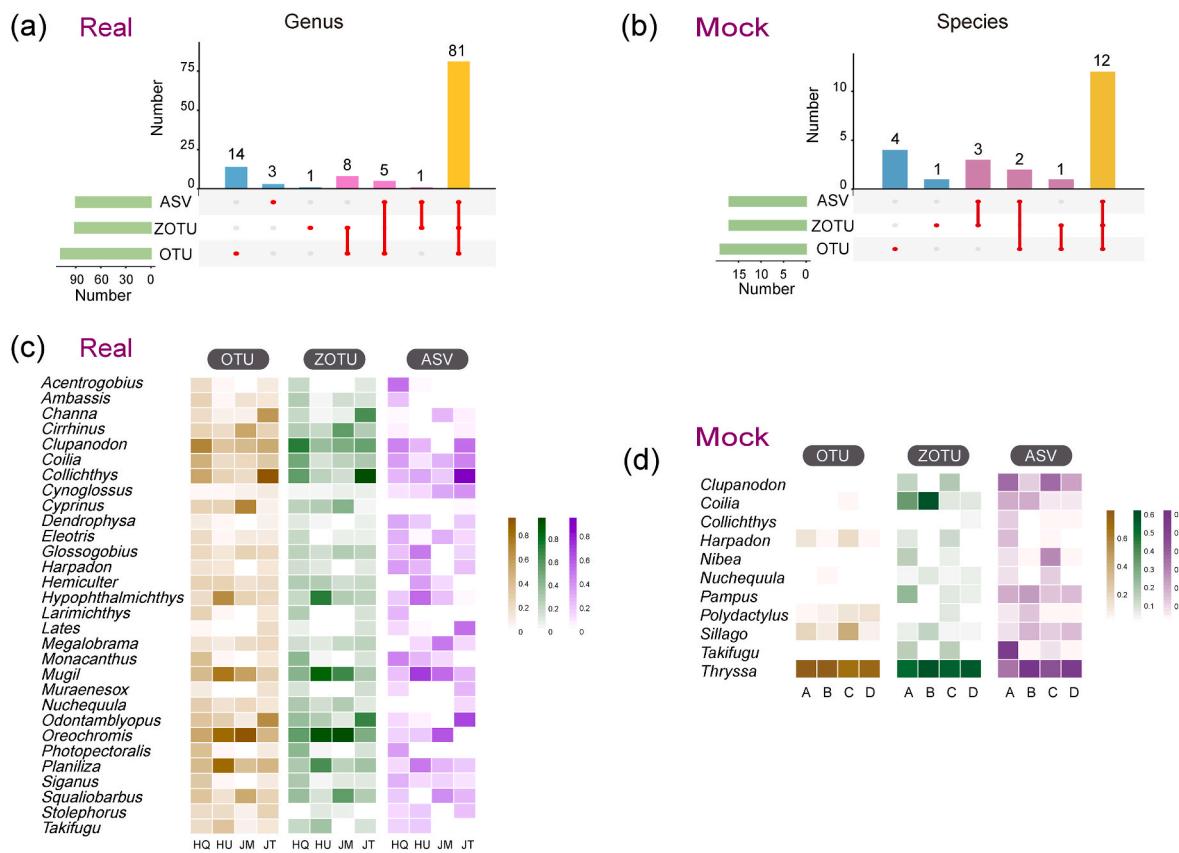


Fig. 2. Number and abundance information of genera/species information across different bioinformatic pipelines: shared and unique genera/species in the real (a)/mock (b) communities, relative abundance of the top 30/11 genera commonly detected in the real (c)/mock (d) communities. The X-axis represented the count of genera/species obtained from each bioinformatics pipeline. The blue bars indicated the count of genera/species obtained from a single bioinformatic pipeline, while the pink bars represented the count of genera/species obtained from any two bioinformatic pipelines. The yellow bars represented the count of genera/species shared by all three bioinformatic pipelines. The red dots or lines below the blue/pink/yellow bars indicated the bioinformatic pipeline responsible for detecting the data represented by that bar, while the detected count was displayed in black font above the blue/pink/yellow bars.

the exclusively detected genera/species in one pipeline were characterized by low-abundance sequences, which constituted a small proportion of the total sequences obtained from their respective pipelines. In the real communities, these exclusive sequences accounted for an average of 0.10% (range: $9.61 \times 10^{-5}\%$ –0.503%) of the total sequences, while in the mock communities, this percentage was even lower, at 0.04% (maximum: 0.14%; minimum: $2.03 \times 10^{-3}\%$).

PERMDISP analysis was conducted for species with relatively high abundance (species that ranked within the top 5), and it showed no statistically significant differences in distribution between different bioinformatics pipelines, which was consistent with the assumptions of the PERMANOVA analysis (Table S6). Further PERMANOVA analysis indicated all three bioinformatic pipelines exhibited minimal differences in between-group differences that were not statistically significant ($P > 0.05$), indicating strong consistency. In the real community, all three bioinformatic pipelines demonstrated strong consistency for *Odontamblyopus* (PERMANOVA: $P = 1.0000$, $F = 0.0608$) and *Collichthys* ($P = 0.9333$, $F = 0.0963$, Fig. 2c) at the JT. Similarly, strong consistency was observed for *Thryssa* in the B ($P = 0.7333$, $F = 0.3818$) and D ($P = 0.6667$, $F = 0.4338$) of the mock community. Mantel analysis on shared genera revealed that ASV had relatively weaker consistency with the other two bioinformatic pipelines ($P < 0.004$, $R_{ASV/ZOTU} = 0.6344$, $R_{ASV/OTU} = 0.7083$) in the real community, while the OTU and ZOTU exhibited higher species similarity and the strongest consistency ($P < 0.001$, $R = 0.9464$) (Fig. 2c). In the mock community, OTU exhibited weaker species consistency compared to the other two bioinformatic pipelines ($P < 0.007$, $R_{OTU/ASV} = 0.5187$, $R_{OTU/ZOTU} = 0.5959$), while OTU and ZOTU showed higher species similarity and the strongest

consistency ($P < 0.001$, $R = 0.7237$).

Fish communities in the real and mock communities exhibited similar compositions at the family level. A total of 55 families were observable in the three pipelines in the real communities, collectively constituting over 97.74% of the total sequences in the OTU (97.74%), ZOTU (98.43%), and ASV (99.99%), respectively (Fig. 3). Meanwhile, in the mock communities, the 15 shared families accounted for 99.98%, 99.79%, and 97.48% of the OTU, ZOTU, and ASV sequences, respectively. The top 10 families with the highest relative abundance in the real/mock communities were detected by all three pipelines and each exhibited distinct relative abundance patterns (Fig. 3). Additionally, certain families such as Ariidae, Centrolophidae, Elopidae, Soleidae, Paralichthyidae, Loricariidae, and Lutjanidae were exclusively detected in the OTU, while the Serrasalmidae (real) and Bagridae (mock) family was only found in the ZOTU. Moreover, the families exclusively identified by a single pipeline consisted of low-abundance sequences, representing a small proportion of the dataset. In real communities, the average proportion of sequences identified at the family level using a single pipeline was 0.074%, ranging from $9.61 \times 10^{-5}\%$ –0.51%. On the other hand, in the mock communities, the proportion of sequences detected at the family level ranged from $5.5 \times 10^{-3}\%$ –1.122%, with an average of 0.208% (Fig. 3).

3.3. Comparison on composition and diversity of real and mock communities

When comparing the consistency of community composition at different taxonomic levels (order, family, genus, and species) among the

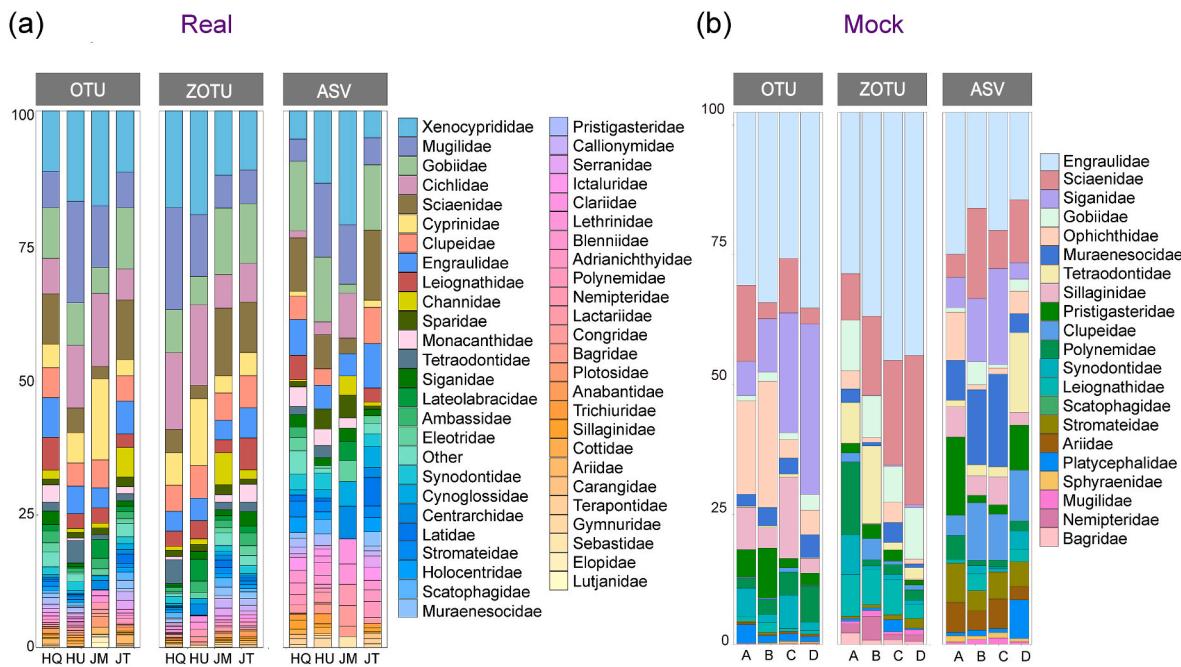


Fig. 3. Relative reads abundance of fish community at the family level was determined using different bioinformatic pipelines in the real (a) and mock community (b). The order was positively correlated with abundance, demonstrating a decreasing trend from top to bottom and from left to right. The category labeled as "Other" represents groups with a relative abundance of less than 0.01%.

three pipelines, the results indicated significant differences ($P < 0.007$) at the family and genus levels in the real communities (Fig. 4a). In the real communities, better consistency in community composition at the

family and genus levels was observed between the OTU and ZOTU, as these two pipelines being more similar to each other (family: $M^2 = 0.0439$; genus: $M^2 = 0.0318$). On the other hand, poorer consistency in



Fig. 4. Procrustes analysis of pairwise comparisons on real (a) and mock (b) communities between the bioinformatic pipelines at different taxonomic levels. P -value is significant ($P < 0.05$), it can be considered that the M^2 of the original observation value is not caused by chance. Smaller M^2 values indicate a higher degree of association between the two datasets.

community composition was found between the ASV and the other two pipelines ($M^2 = 0.6221\text{--}0.7084$). In the mock communities, significant differences ($P < 0.006$) were observed at the order, family, and genus levels (Fig. 4b). However, similar consistencies in community composition were found among the three bioinformatic pipelines in mock communities. The OTU and ZOTU exhibited better consistency in community composition at the family and genus levels (family: $M^2 = 0.0774$; genus: $M^2 = 0.0464$). It is worth noting that the consistency of community composition at the family level was higher compared to other taxonomic levels in both real and mock communities.

In both real and mock communities, the OTU had the highest species richness at all sampled sites, while the ASV showed the highest Pielou evenness (Fig. 5 and Fig. S2). Importantly, the Shannon indexes for real communities were similar across different pipelines. In the real communities, the Shannon index ranged from 2.4150 to 3.5401, with richness ranging from 25 to 102 and Pielou evenness ranging from 0.5316 to 0.9413 (Fig. 5). Significant or highly significant differences were observed between groups in all three alpha diversity indices obtained from the ASV. Further analysis of the pairwise correlations between ASV samples revealed that most of the biological replicates of ASV had low correlations ($-0.38 < r < 0.47$, Fig. S3), except for the correlation between HQ5 and HQ2 ($r = 0.6$). Additionally, the OTU and ZOTU

exhibited significant inter-group differences in the Shannon and Pielou evenness indexes. However, no significant differences were observed between groups in the three alpha diversity indices obtained from the three pipelines for the mock communities (Fig. S2). Group D had the lowest Shannon (1.3598 ± 0.3780) and Pielou evenness index (0.4069 ± 0.1200), while group A exhibited the highest richness (29 ± 2.6458).

Similarity analysis on alpha indices in the real and mock communities indicated that there were no statistically significant differences in the Shannon and Pielou evenness based on the three bioinformatic pipelines (Fig. 6). However, when considering the richness, the ASV exhibited significant differences ($P < 0.01$, Fig. 6a) compared to the other two pipelines in the real communities. In contrast, no differences were observed among the three pipelines in mock communities (Fig. 6b).

3.4. Distance matrix impact on community composition/diversity

Planiliza and *Thryssa* were found to be highly abundant in the real or mock communities when analyzed using the three bioinformatic pipelines (Fig. 2c and d). PERMDISP analysis was conducted for these two genera, and the results indicated no statistically significant differences in the distribution between most of the sites, which aligns with the

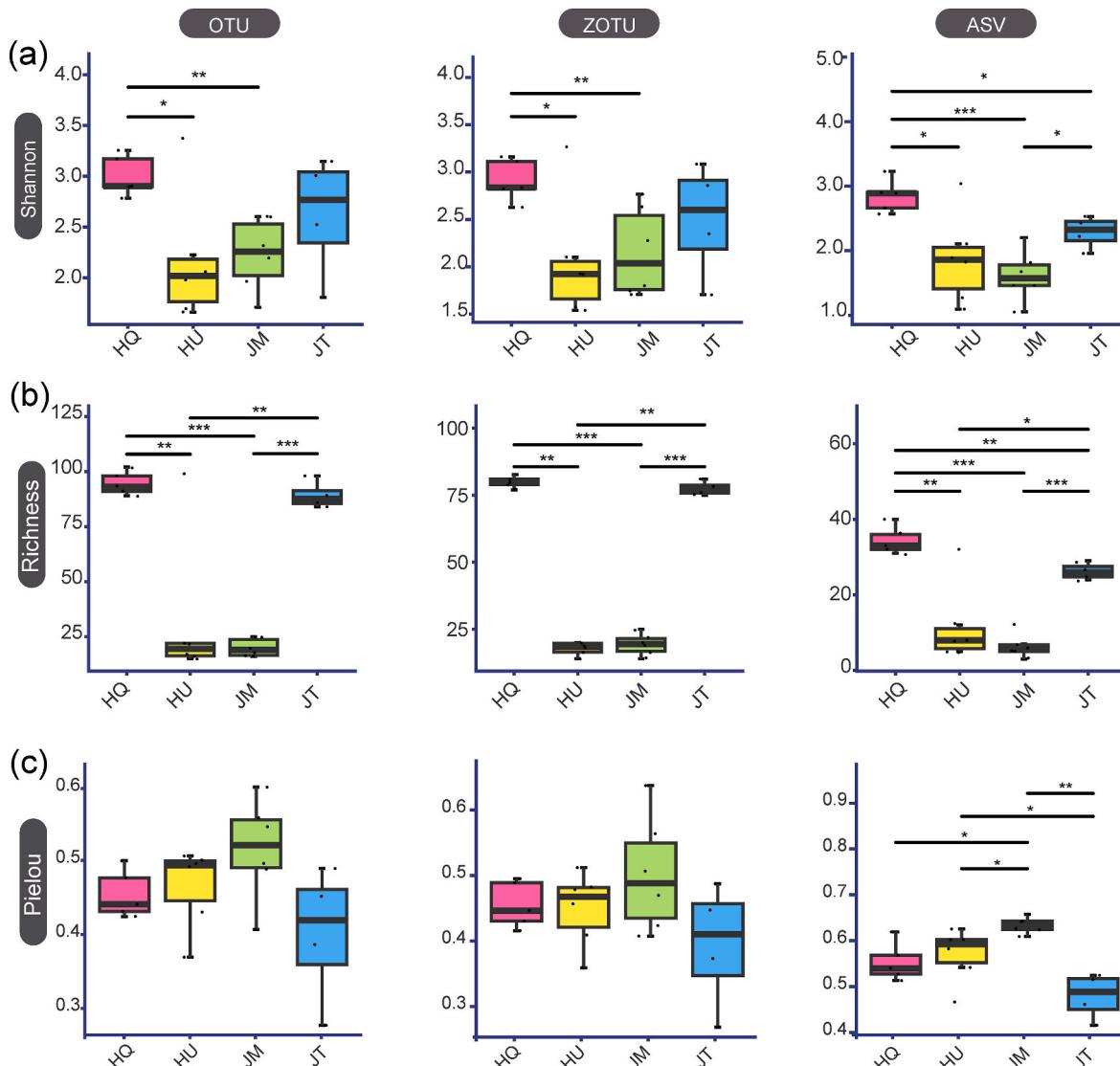


Fig. 5. Multiple comparisons of alpha diversity in the mock community with different bioinformatic pipelines, and the outcomes of inter-group similarity analysis were presented at the uppermost section of each figure. *, $0.05 > P > 0.01$; **, $0.01 > P > 0.001$; ***, $P < 0.001$.

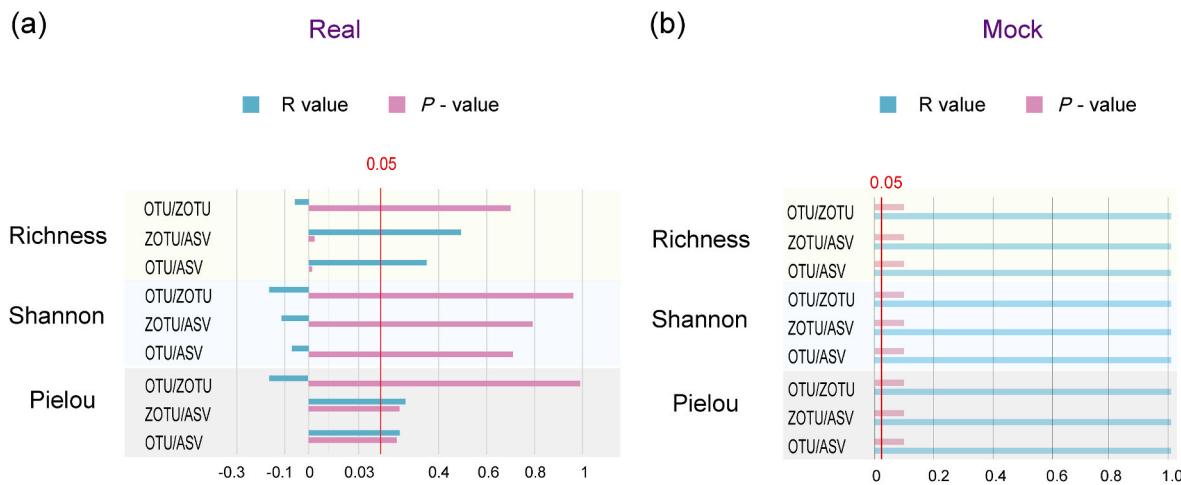


Fig. 6. Pairwise ANOSIM analysis of alpha diversity indices in real (a) and mock (b) communities. The significance threshold of the P -value (0.05) was indicated by the vertical red line.

assumptions of the PERMANOVA analysis (except for the ASV of *Plani*za with P -value of 0.013, Tables S7 and S8). PERMANOVA analysis based on these two genera revealed that the choice of distance matrix significantly influenced the dissimilarity in fish composition among groups (Fig. S4). When analyzing the similarity of community compositions among the three distance matrices, highly significant differences

($P < 0.001$, Fig. 7) were observed, indicating that the type of distance matrix used could greatly impact the assessment of fish community structure. In the real communities, the ASV-based pipeline exhibited relatively low similarity in community composition compared to the other two pipelines (Fig. 7a). The BC distance matrix showed the lowest similarity ($R_{ASV/OTU} = 0.1929$, $R_{ASV/ZOTU} = 0.1872$), while the UU

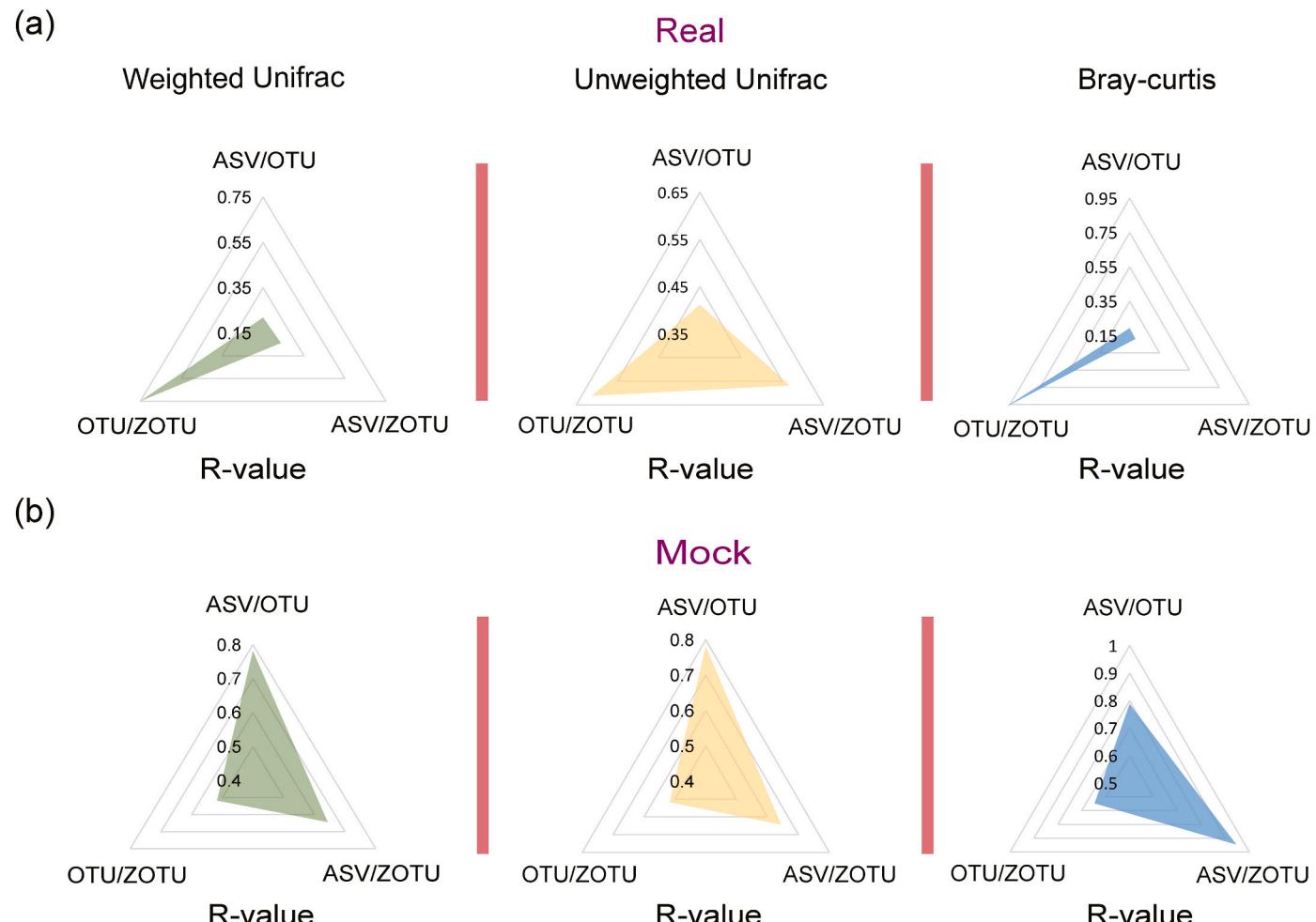


Fig. 7. Similarity of fish community among different bioinformatic pipelines in the real (a) and mock community (b). The Mantel test results reveal a remarkably significant correlation ($P < 0.001$) among the bioinformatic pipelines illustrated in the figure above.

distance matrix showed the highest similarity ($R_{ASV/OTU} = 0.4114$, $R_{ASV/ZOTU} = 0.5678$). Notably, the similarity of community composition between OTU and ZOTU was the strongest ($R_{WU} = 0.7457$; $R_{UU} = 0.6109$; $R_{BC} = 0.9671$). Based on these results, the UU distance matrix provided the highest similarity among the bioinformatic pipelines in real communities. Conversely, in the mock communities, the BC distance matrix yielded the highest similarity between the bioinformatic pipelines. The highest similarity between ASV and OTU was observed in the WU ($R_{WU} = 0.7804$) and UU ($R_{UU} = 0.7804$) distance matrices. In the BC distance matrix, the highest similarity was observed between ASV and ZOTU ($R_{BC} = 0.9449$) in the mock communities (Fig. 7b). Additionally, the lowest similarity in community composition was observed between OTU and ZOTU ($R_{WU} = 0.5164$, $R_{UU} = 0.5164$, $R_{BC} = 0.6463$).

The BC distance matrix performed the best in discriminated community differences between groups (Table 3). In the real communities, the BC distance matrix effectively discriminated community differences between groups based on both the OTU and ASV. The first (PCoA1) and second (PCoA2) axes explained 53.89% and 43.30% of the variation, respectively. Similarly, in the mock community, the BC distance matrix also outperformed the other types in distinguishing community differences between groups across all three pipelines. The first and second axes explained 53.06% (OTU), 51.10% (ZOTU), and 51.38% (ASV) of the variation, respectively. When examining inter-group community differences, the WU distance matrix based on ZOTU was found to be the most effective, explaining 57.91% of the variation in the real communities. Conversely, the BC distance matrix based on OTU was found to be the most effective in the mock communities.

4. Discussion

The interpretation of raw data in eDNA metabarcoding has traditionally relied on the selection of a bioinformatic analysis workflow during the data or information capturing processes (Yu et al., 2022). The choice of bioinformatic analysis workflow directly impacts specific biodiversity outcomes, such as community composition, relative species abundance, and differences in abundance between treatments/sample types, consequently influencing the results of biocenosis investigations utilizing eDNA metabarcoding (Deiner et al., 2017; Joos et al., 2020; Antich et al., 2021; Hakimzadeh et al., 2024). In this study, we compared different bioinformatic analysis pipelines for eDNA metabarcoding based on real and mock fish communities in PRE. Moreover, we expanded upon the conventional focus on species composition, abundance, and community diversity, and also evaluated the performance of bioinformatic pipelines and compared distance matrices for analyzing fish community (Jeske and Gallert, 2022; Pascoal et al., 2021;

Zhou et al., 2021).

4.1. Variation of ecological indicators among different bioinformatic pipelines

Benchmark studies commonly use sensitivity and F-measure to evaluate the quantity of recovered OTUs, ZOTUs, or ASVs (Weiss et al., 2016). On the other hand, compositional similarity is primarily employed to assess the relative abundances between species (Bakker, 2018). Here, we evaluated the performance of three bioinformatic pipelines for fish eDNA metabarcoding using mock communities, and found significant differences in sensitivity, F-measure, and compositional similarity. These results highlight the substantial impact of bioinformatic pipelines on outcomes of fish eDNA metabarcoding. Similar to the findings of Mathon et al. (2021) who assessed the performance of mock communities consisting of 29 different simulated species sets using various bioinformatic programs and pipelines, our results also revealed that the OTU-based pipeline had the highest compositional similarity to mock community with the highest sensitivity (Table 1, Fig. 2a and b). Additionally, the OTU-based pipeline had the lowest false positive rate, while the ZOTU-based pipeline had the highest, indicating that the performance discrepancies among different pipelines were contingent on the selection of steps in each pipeline (Mathon et al., 2021).

The utilization of different processing and algorithmic approaches in bioinformatic pipelines for analyzing raw sequence data can lead to varying output results (Jeske and Gallert, 2022; Özkurt et al., 2022). In line with the findings of Grützke (2019) and Tremblay and Yergeau (2019), our results demonstrated that the ASV-based pipeline produced lower sequence counts and species richness in both real and mock communities compared to the OTU-based pipeline. Furthermore, a higher proportion of sequences were discarded in the ASV-based pipeline (real: 25.33%; mock: 21.02%) compared to the OTU (real: 6.55%; mock: 4.97%). This difference can be attributed to the fact that the ASV-based pipeline removes and excludes more sequences to enhance the accuracy of the final sequences. Firstly, the DADA2 algorithm used in the ASV-based pipeline eliminates singletons and corrects potential sequencing errors in low-abundance sequences, resulting in fewer sequences (Callahan et al., 2016; Dahan et al., 2018; Grützke et al., 2019; Hakimzadeh et al., 2024). Secondly, during the clustering process in the OTU-based pipeline, many shorter sequences are merged into larger sequences, whereas all these sequences ultimately form distinct ASVs and subsequently get discarded if it does not meet the threshold in the ASV-based pipeline (Yu et al., 2022).

Different bioinformatic workflows have varying capabilities in recovering amplicon sequence counts, and the introduction of further taxonomic annotations may introduce biases in data interpretation (Straub et al., 2020). These factors can lead to variations in composition within the same community when different bioinformatic pipelines are employed. Our results showed that the fish community composition exhibited the highest similarity between OTU and ZOTU (significant differences with ASV), and higher number of taxa were identified based on OTU (Fig. 4, Table 2). The observed differences in community composition resulting from different bioinformatic pipelines can be primarily attributed to the handling of low-abundance sequences (Ans-lan et al., 2021; Joos et al., 2020). The ASV-based pipeline involves the removal of numerous low-abundance sequences, which may result in the loss of a few or rare species (Hakimzadeh et al., 2024). In contrast, the clustering algorithm (OTU) is more effective in distinguishing closely related taxa, as compared to the algorithms DADA2 (ASV) and UNOISE3 (ZOTU) (Dahan et al., 2018; García-García et al., 2019). Furthermore, our results showed that different proportions of sequences in ASV and ZOTU cannot be annotated at the species level (Table S4). Therefore, the OTU-based pipeline is better suited for detecting "true" species information in fish communities, as it aligns closer to mock community data with fewer false positives and a higher number of recovered taxa.

Table 3
PCoA analysis based on three distance matrices.

Community	Pipeline	Distance	PCoA1	PCoA2	P-value
Real	OTU	BC	40.63%	13.26%	<0.001
		WU	14.08%	11.10%	
		UU	15.07%	10.96%	
	ZOTU	BC	38.61%	12.20%	
		WU	42.93%	14.98%	
		UU	19.45%	8.83%	
	ASV	BC	28.81%	14.49%	
		WU	11.82%	10.83%	
		UU	18.28%	9.63%	
Mock	OTU	BC	34.50%	18.56%	<0.001
		WU	22.05%	15.98%	
		UU	34.32%	17.85%	
	ZOTU	BC	30.69%	20.41%	
		WU	25.31%	17.99%	
		UU	35.34%	20.27%	
	ASV	BC	30.63%	20.75%	
		WU	19.04%	16.40%	
		UU	29.65%	20.27%	

BC: Bray-Curtis; WU: Weighted UniFrac; UU: Unweighted UniFrac.

4.2. Monitoring capabilities of bioinformatic pipelines in fish diversity

In the study, we identified at least three high-abundance taxa (*Platiliza*, *Cynoglossus*, and *Glossogobius* in the real community; *Thryssa*, *Siganus*, and *Coilia* in the mock community) across all three pipelines. However, the variations in alpha diversity indices based on the three bioinformatic pipelines were observed, irrespective of whether in the real or mock community (Fig. 5, Fig. S2). It is widely acknowledged that the choice of bioinformatic pipelines can influence the results of biodiversity obtained from eDNA metabarcoding (Abellán-Schneyder et al., 2021). Specifically, in the real community, the three pipelines exhibited similar Shannon and Pielou evenness indices, with no significant differences observed. Moreover, the OTU-based pipeline exhibited the highest richness, with significantly different from ASV-based pipeline. Previous study has also reported no significant differences in Shannon and Pielou evenness indices among different bioinformatic pipelines, but did find significant differences in richness (Joos et al., 2020). The richness indices obtained from the bioinformatic pipelines are based on the number of OTUs/ZOTUs/ASVs (Caruso et al., 2019).

Variations in diversity between bioinformatic pipelines may be attributed to differences in output quantity, sequences, taxonomic classification, or the presence of high-abundance taxa (Straub et al., 2020). Here, the OTU-based pipeline yielded the largest taxa at the species level, while the ZOTU with the highest number of clean reads and ASV with the lowest number of clean reads and the fewest taxa, respectively (Table 1, Fig. S4). The clustering algorithm (OTU) is more effective in distinguishing closely related taxonomic groups, leading to the detection of a greater number of species and significant differences with ZOTU/ASV community (real: $P = 0.030$; mock: $P = 0.019$, Table 2; Dahan et al., 2018; García-García et al., 2019). Simultaneously, different alpha diversity showed in the three pipelines, suggesting that the variations in sequence and taxonomic numbers generated by the bioinformatic pipelines may contribute to the different alpha diversity observed in the fish community. Besides, significant differences were observed in abundance between the OTU and ASV-based pipelines in the real fish community ($P < 0.01$, Fig. 6). Similar phenomena have been reported in bacterial communities (Chiarello et al., 2022). Compared to ASV and ZOTU, the OTU-based pipeline resulted in lower Pielou evenness and a larger number of low-abundance sequences in the real community, which could affect the "true" richness within the community (Chiarello et al., 2022; Modin et al., 2020). Furthermore, the DADA2 algorithm (ASV-based pipeline) exhibited lower reproducibility between duplicate samples in the real community, leading to more pronounced inter-group differences in alpha diversity and larger inter-sample variances (Fig. 5, Fig. S1; Özkar et al., 2022). Consequently, the OTU-based pipeline yielded higher fish diversity (the highest richness) with less inter-group differences in alpha diversity, indicating its superior capability for monitoring fish diversity in PRE. The findings further highlight the critical role of selecting appropriate bioinformatic pipelines in assessing the biodiversity of fish communities.

4.3. Selection of distance matrix in bioinformatic pipelines

Distance matrices are frequently used to detect subtle variations in communities that are almost identical (Chao and Chiu, 2016; Chiarello et al., 2022; Parks and Beiko, 2013). Significantly different ($P < 0.001$) in correlation between bioinformatic pipelines and distance matrices indicate that the selection of a distance matrix significantly impact the information obtained for assessing fish community in eDNA metabarcoding. For instance, the similarity of fish community among different bioinformatic pipelines ranged from 0.1872 to 0.9671 in the real community (Fig. 7a), suggesting that different distance matrix can vary influences on the dataset and clustering results (Cholet et al., 2022; García-López et al., 2021; Pérez-Cobas et al., 2020).

Here, consistency in the changes of fish community similarity across

different bioinformatic pipelines was observed. And stronger correlations between the pipelines and BC/WU in the mock community suggested that the consistency of the pipelines for fish community analysis was influenced by the enriched taxa (Zhou et al., 2021). Consistent with a study by Martins et al. (2022) the abundance-weighted indices (BC and WU) in PCoA were found to be more effective in distinguishing and explaining changes in fish community composition compared to the presence-absence distance matrix (UU) (Table 3). Because the abundance-weighted indices incorporate more information about the community, while the presence-absence distance matrix only considers the presence or absence of taxa (Martino et al., 2019). Furthermore, the inter-group analysis based on *Thryssa* also revealed that the BC outperformed in distinguishing inter-group differences across all three bioinformatic pipelines (Fig. S4). The BC distance matrix is more suitable for differentiating fish communities or specific genera to reveal the greatest variances. Therefore, it is crucial to consider both the selection of bioinformatic pipeline and the distance matrix within the same one for the interpretation and analysis of fish eDNA metabarcoding data.

5. Conclusion

The selection of bioinformatic pipelines is crucial for assessing biodiversity and richness in fish community. Our study highlights the importance of bioinformatic pipelines selection and compares the differences among three bioinformatic pipelines in various aspects, including fish community composition, diversity, and the impact of distance matrices. Here, we have demonstrated that different bioinformatic analysis workflows can influence the obtained information regarding to fish community composition, with each workflow exhibiting its own biases. The OTU-based pipeline outperformed the others in terms of overall performance, including sensitivity, similarity to mock community composition, species richness, and diversity. It exhibited the fewest missing species, lowest false positives, and strongest similarity to community composition when analyzing fish eDNA data based on a mock community. Furthermore, different bioinformatic analysis workflows are not entirely consistent in terms of the suitable distance matrix. The Bray-Curtis distance matrix proved more effective in distinguishing fish communities or specific genera compared to the weighted UniFrac and unweighted UniFrac distance matrices. In summary, important implications for the data interpretation and analysis of fish eDNA metabarcoding (by selecting appropriate bioinformatic pipelines) in this study can promote to better understanding in the preferences of bioinformatic pipelines.

CRediT authorship contribution statement

Zhuoying Li: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Wencheng Zhao:** Methodology, Investigation, Formal analysis, Data curation. **Yun Jiang:** Validation, Methodology, Investigation, Formal analysis. **Yongjing Wen:** Visualization, Investigation, Formal analysis. **Min Li:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Formal analysis. **Li Liu:** Supervision, Methodology, Formal analysis. **Keshu Zou:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis.

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

The data that has been used is confidential.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (32102793), and the Central Public-interest Scientific Institution Basal Research Fund, South China Sea Fisheries Research Institute, CAFS (2021SD18).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jenvman.2024.122136>.

Appendix 1

For each biological replicate sample, we took 20 mg of fin clip and cut it into fragments, then used the TIANamp Marine Animals DNA Kit (Tiangen Biotech) to extract DNA according to the manufacturer's protocol. Negative controls and blanks were extracted alongside the samples and using the same protocol as the samples. The DNA quality and concentration were assessed using a 1.0% agarose gel and the Nanodrop 2000 system (Thermo Fisher Scientific, Waltham, MA, USA). No bands or data were detected in the filtered blank or negative controls. All samples were stored at -20 °C until PCR amplification testing was performed.

References

- Abellan-Schneyder, I., Matchado, M.S., Reitmeier, S., Sommer, A., Sewald, Z., Baumbach, J., List, M., Neuhaus, K., 2021. Primer, pipelines, parameters: issues in 16S rRNA gene sequencing. *mSphere* 6 (1), e01202, 20.
- Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 26 (1), 32–46.
- Andres, K.J., Lambert, T.D., Lodge, D.M., Andrés, J., Jackson, J.R., 2023. Combining sampling gear to optimally inventory species highlights the efficiency of eDNA metabarcoding. *Environmental DNA* 5 (1), 146–157.
- Anslan, S., Mikryukov, V., Armalaitis, K., Ankuda, J., Lazdina, D., Makovskis, K., Vesterdal, L., Schmidt, I.K., Tedersoo, L., 2021. Highly comparable metabarcoding results from MiG-Tech and Illumina sequencing platforms. *PeerJ* 9, e12254.
- Antich, A., Palacín, C., Wangensteen, O.S., Turon, X., 2021. To denoise or to cluster, that is not the question: optimizing pipelines for COI metabarcoding and metaphylogeny. *BMC Bioinf.* 22, 177.
- Bakker, M.G., 2018. A fungal mock community control for amplicon sequencing experiments. *Molecular Ecology Resources* 18, 541e556.
- Beng, K.C., Corlett, R.T., 2020. Applications of environmental DNA (eDNA) in ecology and conservation: opportunities, challenges and prospects. *Biodivers. Conserv.* 29 (7), 2089–2121.
- Brandies, P.A., Hogg, C.J., 2021. Ten simple rules for getting started with command-line bioinformatics. *PLoS Comput. Biol.* 17 (2), e1008645.
- Brandt, M.I., Trouche, B., Quintric, L., Günther, B., Wincker, P., Poulain, J., Arnaud-Haond, S., 2021. Bioinformatic pipelines combining denoising and clustering tools allow for more comprehensive prokaryotic and eukaryotic metabarcoding. *Molecular Ecology Resources* 21 (6), 1904–1921.
- Bray, J.R., Curtis, J.T., 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* 27 (4), 325–349.
- Bylemans, J., Gleeson, D.M., Linternmans, M., Hardy, C.M., Beitzel, M., Gilligan, D.M., Furlan, E.M., 2018. Monitoring riverine fish communities through eDNA metabarcoding: determining optimal sampling strategies along an altitudinal and biodiversity gradient. *Metabarcoding and Metagenomics* 2, e30457.
- Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13 (7), 581–583.
- Caruso, V., Song, X., Asquith, M., Karstens, L., 2019. Performance of microbiome sequence inference methods in environments with varying biomass. *mSphere* 4 (1), 10–1128.
- Chao, A., Chiu, C.H., 2016. Nonparametric estimation and comparison of species richness. *Encyclopedia of Life Sciences* 16, 1–11. John Wiley & Sons, May.
- Chiarello, M., McCauley, M., Villéger, S., Jackson, C.R., 2022. Ranking the biases: the choice of OTUs vs. ASVs in 16S rRNA amplicon data analysis has stronger effects on diversity measures than rarefaction and OTU identity threshold. *PLoS One* 17 (2), e0264443.
- Cholet, F., Lisik, A., Agogué, H., Ijaz, U.Z., Pineau, P., Lachaussée, N., Smith, C.J., 2022. Ecological observations based on functional gene sequencing are sensitive to the amplicon processing method. *mSphere* 7 (4), e00324, 22.
- Conway, J.R., Lex, A., Gehlenborg, N., 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33 (18), 2938–2940.
- Couton, M., Viard, F., Altermatt, F., 2023. Opportunities and inherent limits of using environmental DNA for population genetics. *Environmental DNA* 5 (5), 1048–1064.
- Dahan, D., Jude, B.A., Lamendella, R., Keesing, F., Perron, G.G., 2018. Exposure to arsenic alters the microbiome of larval zebrafish. *Front. Microbiol.* 9, 1323.
- Dal Pont, G., Duarte Ritter, C., Agostinis, A.O., Stica, P.V., Horodesky, A., Cozer, N., Balsanelli, E., Netto, O.S.M., Henn, C., Ostrensky, A., Pie, M.R., 2021. Monitoring fish communities through environmental DNA metabarcoding in the fish pass system of the second largest hydropower plant in the world. *Sci. Rep.* 11 (1), 23167.
- Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D.M., Vere, N.D., Pfrender, M.E., Bernatchez, L., 2017. Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *Mol. Ecol.* 26 (21), 5872–5895.
- DiBattista, J.D., Fowler, A.M., Riley, I.J., Reader, S., Hay, A., Parkinson, K., Hobbs, J.P. A., 2022. The use of environmental DNA to monitor impacted coastal estuaries. *Mar. Pollut. Bull.* 181, 113860.
- Dugal, L., Thomas, L., Wilkinson, S.P., Richards, Z.T., Alexander, J.B., Adam, A.A., Kennington, W.J., Jarman, S., Ryan, N.M., Bunce, M., Gilmour, J.P., 2022. Coral monitoring in northwest Australia with environmental DNA metabarcoding using a curated reference database for optimized detection. *Environmental DNA* 4 (1), 63–76.
- Edgar, R.C., 2016a. SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequences. *bioRxiv*, 074161.
- Edgar, R.C., 2016b. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv*, 081257.
- Elliott, V.L., Chheng, P., Uy, S., Holtgrieve, G.W., 2019. Monitoring of tropical freshwater fish resources for sustainable use. *J. Fish. Biol.* 94 (6), 1019–1025.
- García-García, N., Tamames, J., Linz, A.M., Pedrós-Alió, C., Puente-Sánchez, F., 2019. Microdiversity ensures the maintenance of functional microbial communities under changing environmental conditions. *ISME J.* 13 (12), 2969–2983.
- García-López, R., Cornejo-Granados, F., Lopez-Zavala, A.A., Cota-Húizar, A., Sotelo-Mundo, R.R., Gómez-Gil, B., Ochoa-Leyva, A., 2021. OTUs and ASVs produce comparable taxonomic and diversity from shrimp microbiota 16S profiles using tailored abundance filters. *Genes* 12 (4), 564.
- García-Machado, E., Laporte, M., Normandieu, E., Hernández, C., Côté, G., Paradis, Y., Mingelbier, M., Bernatchez, L., 2022. Fish community shifts along a strong fluvial environmental gradient revealed by eDNA metabarcoding. *Environmental DNA* 4 (1), 117–134.
- Gardner, P.P., Watson, R.J., Stott, M.B., Morales, S.E., Morgan, X.C., Finn, R.D., Draper, J.L., 2019. Identifying accurate metagenome and amplicon software via a meta-analysis of sequence to taxonomy benchmarking studies. *PeerJ* 7, 1–19.
- Goutte, A., Molbert, N., Guérin, S., Richoux, R., Rocher, V., 2020. Monitoring freshwater fish communities in large rivers using environmental DNA metabarcoding and a long-term electrofishing survey. *J. Fish. Biol.* 97 (2), 444–452.
- Gower, J.C., 1975. Generalized procrustes analysis. *Psychometrika* 40, 33–51.
- Grützke, J., Malorny, B., Hammerl, J.A., Busch, A., Tausch, S.H., Tomaso, H., Deneke, C., 2019. Fishing in the soup-pathogen detection in food safety using metabarcoding and metagenomic sequencing. *Front. Microbiol.* 10, 1805.
- Guzman, I.D., Altieri, P., Elosgé, A., Pérez-Calpe, A.V., Schiller, D.V., González, J.M., Brauns, M., Montoya, J.M., Larranaga, A., 2022. Water diversion and pollution interactively shape freshwater food webs through bottom-up mechanisms. *Global Change Biol.* 28 (3), 859–876.
- Handley, L., Read, D.S., Winfield, I.J., Kimbell, H., Johnson, H., Li, J., Hahn, C., Blackman, R., Wilcox, R., Donnelly, R., Szitenberg, A., 2019. Temporal and spatial variation in distribution of fish environmental DNA in England's largest lake. *Environmental DNA* 1, 26–39.
- Hakimzadeh, A., Abdala Asbun, A., Albanese, D., Bernard, M., Buchner, D., Callahan, B., Caporaso, J.G., Curd, E., Djemiel, C., Durling, M.B., Elbrecht, V., Gold, Z., Gweon, H. S., Hajibabaei, M., Hildebrand, F., Mikryukov, V., Normandieu, E., Özkurt, E., Palmer, J.M., Pascal, G., Porter, T.M., Straub, D., Vasar, M., Vetrovský, T., Zafeiropoulos, H., Anslan, S., 2024. A pile of pipelines: an overview of the bioinformatics software for metabarcoding data analyses. *Molecular Ecology Resources* 24 (5), e13847.
- Handley, L.L., Blackwell, T., Broadhurst, H.A., Clark, K., Davison, P.I., England, J., Stefano Mariani, S., McDevitt, A.D., Pillay, K., Read, D., Walsh, K., Nisbet, A., Creer, S., 2023. UK DNA working group eDNA week, January 2022. *Environmental DNA* 5 (1), 18–24.
- Huang, H.Y., 2021. linkET: Everything Is Linkable. R Package Version 0.0.3.
- Jeske, J.T., Gallert, C., 2022. Microbiome analysis via OTU and ASV-based pipelines—a comparative interpretation of ecological data in WWTP systems. *Bioengineering* 9 (4), 146.
- Jiang, P.W., Li, M., Zhang, S., Chen, Z.Z., Xu, S.N., 2022. Construction of DNA metabarcode database of fish in Pearl River Estuary based on mitochondrial cytochrome COI and 12S rDNA gene. *Southern Fisheries Science* 18 (3), 13–21 (in Chinese).
- Joos, L., Beirinckx, S., Haegeman, A., Debeke, J., Vandecasteele, B., Baeyen, S., Goormachtig, S., Clement, L., De Tender, C., 2020. Daring to be differential: metabarcoding analysis of soil and plant-related microbial communities using amplicon sequence variants and operational taxonomical units. *BMC Genom.* 21, 733.
- Kachroo, N., Lange, D., Penniston, K.L., Stern, J., Tasian, G., Bajic, P., Wolfe, A.J., Suryavanshi, M., Ticinesi, A., Meschi, Y., Monga, M., Miller, A.W., 2021. Meta-analysis of clinical microbiome studies in urolithiasis reveal age, stone composition, and study location as the predominant factors in urolithiasis-associated microbiome composition. *mBio* 12 (4), e02007–e02021.
- Keck, F., Blackman, R.C., Bossart, R., Brantschen, J., Couton, M., Hürlemann, S., Kirschner, D., Locher, N., Zhang, H., Altermatt, F., 2022. Meta-analysis shows both congruence and complementarity of DNA and eDNA metabarcoding to traditional methods for biological community assessment. *Mol. Ecol.* 31 (6), 1820–1835.

- Kuang, T.X., Chen, W.J., Huang, S.H., Liu, L., Zhou, L., 2021. Environmental drivers of the functional structure of fish communities in the Pearl River Estuary. *Estuar. Coast Shelf Sci.* 263, 107625.
- Laporte, M., Reny-Nolin, E., Chouinard, V., Hernandez, C., Normandeau, E., Bougas, B., Côté, C., Behmel, S., Bernatchez, L., 2021. Proper environmental DNA metabarcoding data transformation reveals temporal stability of fish communities in a dendritic river system. *Environmental DNA* 3 (5), 1007–1022.
- Lever, J., Krzywinski, M., Altman, N., 2016. Classification evaluation. *Nat. Methods* 13 (8), 603–604.
- Li, W., Zhai, D., Wang, C., Gao, X., Liu, H., Cao, W., 2021. Relationships among Trophic Niche Width, morphological variation, and genetic diversity of Hemiculter leuciscus in China. *Frontiers in Ecology and Evolution* 9, 691218.
- 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, 109915.
- Lin, D., Zhu, K., Qian, W., Putt, A., Chen, X.J., 2020. Fatty acid comparison of four sympatric loliginid squids in the northern South China Sea: indication for their similar feeding strategy. *PLoS One* 15 (6), e0234250.
- Liu, P.Y., Yang, S.H., Yang, S.Y., 2022. KTU: K-mer Taxonomic Units improve the biological relevance of amplicon sequence variant microbiota data. *Methods Ecol. Evol.* 13 (3), 560–568.
- Lozupone, C.A., Hamady, M., Kelley, S.T., Knight, R., 2007. Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. *Appl. Environ. Microbiol.* 73 (5), 1576–1585.
- Maki, K.A., Kazmi, N., Barb, J.J., Ames, N., 2021. The oral and gut bacterial microbiomes: similarities, differences, and connections. *Biol. Res. Nurs.* 23 (1), 7–20.
- Martino, C., Morton, J.T., Marotz, C.A., Thompson, L.R., Tripathi, A., Knight, R., Zengler, K., 2019. A novel sparse compositional technique reveals microbial perturbations. *mSystems* 4 (1), 10–1128.
- Martins, A.R., Bastos, D.A., Sousa, L.M., Giarrizzo, T., Vieira, T.B., Zuanon, J., 2022. Intense droughts affect temporal stability of Amazonian stream fish assemblages. *Freshw. Biol.* 67 (9), 1656–1667.
- Mathon, L., Marques, V., Mouillet, D., Albouy, C., Andrello, M., Baletaud, F., Baletaud, F., Borrero-Perez, H.G., Dejean, T.J., Edgar, G., Grondin, J., Guerin, P., Hocde, R., Juvel, J., Kadarusman, Maire, E., Mariani, G., McLean, M., Polanco, F.A., Pouyaud, L., Stuart-Smith, R.D., Sugeha, H.Y., Valentini, A., Vigliola, L., Vimono, I.B., Pellissier, L., Manel, S., 2022. Cross-ocean patterns and processes in fish biodiversity on coral reefs through the lens of eDNA metabarcoding. *Proceedings of the Royal Society B* 289, 20220162, 1973.
- Mathon, L., Valentini, A., Guérin, P.E., Normandeau, E., Noel, C., Lionnet, C., Boulange, E., Thuiller, W., Bernatchez, L., Mouillet, D., Dejean, T., Manel, S., 2021. Benchmarking bioinformatic tools for fast and accurate eDNA metabarcoding species identification. *Molecular Ecology Resources* 21 (7), 2565–2579.
- Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J.Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M., Iwasaki, W., 2015. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *R. Soc. Open Sci.* 2 (7), 150088.
- Miya, M., 2022. Environmental DNA Metabarcoding: a novel method for biodiversity monitoring of marine fish communities. *Ann. Rev. Mar. Sci.* 14 (1), 161–185.
- Modin, O., Liébana, R., Saheb-Alam, S., Wilén, B.M., Suarez, C., Hermansson, M., Persson, F., 2020. Hill-based dissimilarity indices and null models for analysis of microbial community assembly. *Microbiome* 8, 1–16.
- Moossavi, S., Atakora, F., Fehr, K., Khafipour, E., 2020. Biological observations in microbiota analysis are robust to the choice of 16S rRNA gene sequencing processing algorithm: case study on human milk microbiota. *BMC Microbiol.* 20 (1), 1–9.
- Needham, D.M., Sachdeva, R., Fuhrman, J.A., 2017. Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *ISME J.* 11 (7), 1614–1629.
- Odum, E.P., 1950. Bird populations of the highlands (North Carolina) plateau in relation to plant succession and avian invasion. *Ecology* 31, 587–605.
- Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Wagner, H.H., 2018. Vegan: community ecology package. R package version 2, 5-2.
- Özkurt, E., Fritscher, J., Soranzo, N., Ng, D.Y., Davey, R.P., Bahram, M., Hildebrand, F., 2022. LotuS2: an ultrafast and highly accurate tool for amplicon sequencing analysis. *Microbiome* 10 (1), 176.
- Pan, Z., Tang, C., Cao, Y., Xuan, Y., Zhou, Q., 2023. Distribution and source apportionment of phenolic EDCs in rivers in the Pearl River Delta, South China. *Environ. Sci. Pollut. Control Ser. A* 30 (16), 48248–48259.
- Parks, D.H., Beiko, R.G., 2013. Measures of phylogenetic differentiation provide robust and complementary insights into microbial communities. *ISME J.* 7 (1), 173–183.
- Pascoal, F., Costa, R., Assmy, P., Duarte, P., Magalhães, C., 2021. Exploration of the types of rarity in the arctic ocean from the perspective of multiple methodologies. *Microb. Ecol.* 84 (1), 59–72.
- Pauvert, C., Buée, M., Laval, V., Edel-Hermann, V., Fauchery, L., Gautier, A., Lesur, I., Vallance, J., Vacher, C., 2019. Bioinformatics matters: the accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline. *Fungal Ecology* 41, 23–33.
- Pérez-Cobas, A.E., Gomez-Valero, L., Buchrieser, C., 2020. Metagenomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses. *Microb. Genomics* 6 (8), 1–22.
- Poncheewin, W., Hermes, G.D., Van Dam, J.C., Koehorst, J.J., Smidt, H., Schaap, P.J., 2020. NG-Tax 2.0: a semantic framework for high-throughput amplicon analysis. *Front. Genet.* 10, 1366.
- Pukk, L., Kanefsky, J., Heathman, A.L., Weise, E.M., Nathan, L.R., Herbst, S.J., Sard, N.M., Scribner, K.T., Robinson, J.D., 2021. eDNA metabarcoding in lakes to quantify influences of landscape features and human activity on aquatic invasive species prevalence and fish community diversity. *Divers. Distrib.* 27 (10), 2016–2031.
- R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Reitmeier, S., Hitch, T.C., Treichel, N., Fikas, N., Hausmann, B., Ramer-Tait, A.E., Neuhaus, K., Berry, D., Haller, D., Lagkouvardos, I., Clavel, T., 2021. Handling of spurious sequences affects the outcome of high-throughput 16S rRNA gene amplicon profiling. *ISME Communications* 1 (1), 31.
- Ruan, H.T., Wang, R.L., Li, H.T., Liu, L., Kuang, T.X., Li, M., Zou, K.S., 2022. Effects of sampling strategies and DNA extraction methods on eDNA metabarcoding: a case study of estuarine fish diversity monitoring. *Zool. Res.* 43 (2), 192–204.
- Ruppert, K.M., Kline, R.J., Rahman, M.S., 2019. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: a systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation* 17, e00547.
- Sales, N.G., Wangensteen, O.S., Carvalho, D.C., Deiner, K., Præbel, K., Coscia, I., McDevitt, A.D., Mariani, S., 2021. Space-timedynamics in monitoring neotropical fish communities using eDNA metabarcoding. *Sci. Total Environ.* 754, 142096.
- Siddiqui, N.Y., Ma, L., Brubaker, L., Mao, J., Hoffman, C., Dahl, E.M., Wang, Z.Q., Karstens, L., 2022. Updating urinary microbiome analyses to enhance biologic interpretation. *Front. Cell. Infect. Microbiol.* 12, 789439.
- Skelton, J., Cauvin, A., Hunter, M.E., 2023. Environmental DNA metabarcoding read numbers and their variability predict species abundance, but weakly in non-dominant species. *Environmental DNA* 5 (5), 1092–1104.
- Sokal, R.R., Crovello, T.J., 1970. The Biological species concept: a critical evaluation. *Am. Nat.* 104 (936), 127–153.
- St-Pierre, A.P., Shikor, V., Schneider, D.C., 2018. Count data in biology-Data transformation or model reformation? *Ecol. Evol.* 8 (6), 3077–3085.
- Straub, D., Blackwell, N., Langarica-Fuentes, A., Peltzer, A., Nahnsen, S., Kleindienst, S., 2020. Interpretations of environmental microbial community studies are biased by the selected 16S rRNA (gene) amplicon sequencing pipeline. *Front. Microbiol.* 11, 550420.
- Takahashib, M., Saccò, M., Kestel, J.H., Nester, G., Campbell, M.A., Heyde, M.V.D., Heydenrych, M.J., Juszkiewicz, D.J., Nevill, P., Dawkins, K.L., Bessey, C., Fernandes, K., Miller, H., Power, M., Mousavi-Derazmahalleh, M., Newton, J.P., White, N.E., Richards, Z.T., Allentoft, M.E., 2023. Aquatic environmental DNA: a review of the macro-organismal biomonitoring revolution. *Sci. Total Environ.* 873 (15), 162332.
- Tremblay, J., Yergeau, E., 2019. Systematic processing of ribosomal RNA gene amplicon sequencing data. *GigaScience* 8 (12), giz146.
- Tsuji, S., Maruyama, A., Miya, M., Ushio, M., Sato, H., Minamoto, T., Yamanaka, H., 2020. Environmental DNA analysis shows high potential as a tool for estimating intraspecific genetic diversity in a wild fish population. *Molecular Ecology Resources* 20 (5), 1248–1258.
- Warton, D.I., Wright, S.T., Wang, Y., 2012. Distance-based multivariate analyses confound location and dispersion effects. *Methods Ecol. Evol.* 3 (1), 89–101.
- Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., Xia, L.C., Xu, Z.Z., Ursell, L., Alm, E.J., Birmingham, A., Cram, J.A., Fuhrman, J.A., Raes, J., Sun, F., Zhou, J., Knight, R., 2016. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J.* 10, 1669e1681.
- Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York City.
- Xiong, F., Shu, L., Gan, X., Zeng, H., He, S., Peng, Z., 2022. Methodology for fish biodiversity monitoring with environmental DNA metabarcoding: the primers, databases and bioinformatic pipelines. *Water Biol. Syst.* 1 (1), 100007.
- Xu, R., Yu, Z., Zhang, S., Meng, F., 2019. Bacterial assembly in the bio-cake of membrane bioreactors: stochastic vs. deterministic processes. *Water Res.* 157, 535–545.
- Yu, Z., Yu, X.F., Kerem, G., Ren, P.G., 2022. Perturbation on gut microbiota impedes the onset of obesity in high fat diet-induced mice. *Front. Endocrinol.* 13, 795371.
- Zaike, A., Greenfield, P., Abbott, C., von Ammon, U., Bleiwitch, J., Bunce, M., Crisostomi, M.E., Charlton, A., Dowle, E., Geller, J., Gutierrez, A.A., Hajibabaei, M., Haggard, E., Inglis, G.J., Lavery, S.D., Samoiloviene, A., Simpson, T., Stat, M., Stephenson, S., Sutherland, J., Thakur, V., Westfall, K., Wood, S.A., Wright, M., Zhang, G., Pochon, X., 2022. Towards reproducible metabarcoding data lessons from an international cross-laboratory experiment. *Molecular Ecology Resources* 22 (2), 519–538.
- Zeng, Z.Y., Cheung, W.W.L., Lai, H., Yi, H.D., Bi, S., Li, H.Y., Chen, X.L., Su, Y.Q., Liu, X.G., Chen, Q.X., Zhang, Z.L., Wei, X.C., Chen, J.H., Li, G.F., 2022. Species and functional dynamics of the demersal fish community and responses to disturbances in the pearl river estuary. *Front. Mar. Sci.* 9, 921595.
- Zhang, X., Xue, Y., Zhang, C., Ren, Y.P., Xu, B.D., Chen, Y., 2021. Sampling intensity influences the estimation of functional diversity indices of fish communities. *Ecol. Indicat.* 121, 107169.
- Zhao, R., Feng, J., Liu, J., Fu, W., Li, X., Li, B., 2019. Deciphering of microbial community and antibiotic resistance genes in activated sludge reactors under high selective pressure of different antibiotics. *Water Res.* 151, 388–402.
- Zhou, L., Wang, G.P., Kuang, T., Guo, D.L., Li, G.F., 2019. Fish assemblage in the Pearl River Estuary: spatial-seasonal variation, environmental influence and trends over the past three decades. *J. Appl. Ichthyol.* 35 (4), 884–895.
- Zhou, S., Fan, C.R., Xia, H.M., Zhang, J., Yang, W., Ji, D.J., Wang, L., Chen, L., Liu, N.N., 2022a. Combined use of eDNA metabarcoding and bottom trawling for the assessment of fish biodiversity in the Zhoushan Sea. *Front. Mar. Sci.* 8, 2056.
- Zhou, W.G., Xu, P., Ding, D.W., Suo, A.N., Xie, M.J., Zhang, L., Yue, W.Z., He, W.H., 2022b. Trophic structure of fishes and macroinvertebrates in relation to

environmental indicators in artificial reef ecosystems of Pearl River Estuary. *Ecol. Indicat.* 138, 108823.

Zhou, Y., Leung, M.H., Tong, X., Lee, J.Y., Lee, P.K., 2021. City-scale meta-analysis of indoor airborne microbiota reveals that taxonomic and functional compositions vary with building types. *Environ. Sci. Technol.* 55 (22), 15051–15062.

Zou, K.S., Chen, J.W., Ruan, H.T., Li, Z., Guo, W., 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, 134704.