

# Qualitative and quantitative DNA metabarcoding assessment of Neotropical ichthyoplankton communities is marker-dependent

Daniel Teixeira<sup>1</sup>, Heron Hilário<sup>2</sup>, Gilmar Santos<sup>3</sup>, and Daniel Carvalho<sup>2</sup>

<sup>1</sup>UFMG

<sup>2</sup>Pontifícia Universidade Católica de Minas Gerais

<sup>3</sup>Pontifícia Universidade Católica de Minas Gerais

June 20, 2023

## Abstract

The study of ichthyoplankton is paramount to understanding fish assemblages' reproductive dynamics. DNA metabarcoding has been applied as a rapid, cost-effective, and accurate taxonomy tool, allowing the identification of multiple individuals simultaneously. However, there remain significant challenges when using DNA metabarcoding, such as molecular marker choice according to the taxonomic resolution and length of the fragment to be sequenced, primer bias, incomplete reference databases, and qualitative inferences incongruences. Here, 30 ichthyoplankton pools collected from a Neotropical River were identified at a molecular level using DNA metabarcoding to compare the resolution, sensibility, specificity and relative read abundance (RRA) recovery of three molecular markers: the standard COI fragment (650pb, with each strand analyzed individually) and two short 12S rRNA genes markers (? 200bp - NeoFish and MiFish markers). The combined use of the three markers increased the genera detection rates by 25% to 87.5%, allowing an increased taxonomic coverage and robust taxonomic identification of complex neotropical ichthyoplankton communities. RRA is marker-dependent, indicating caution is still needed whilst inferring species abundance based on DNA metabarcoding data when using PCR-dependent protocols.

## Qualitative and quantitative DNA metabarcoding assessment of Neotropical ichthyoplankton communities is marker-dependent

Running Title: *Qualitative and quantitative DNA metabarcode*

Daniel Fonseca Teixeira<sup>1, 2</sup>, Heron Oliveira Hilário<sup>1</sup>, Gilmar Bastos Santos<sup>1</sup>, Daniel Cardoso Carvalho<sup>1\*</sup>

1- Post-Graduate Program in Vertebrate Biology, Pontifical Catholic University of Minas Gerais, PUC Minas, Belo Horizonte 30535-610, Brazil; danielcarvalho@pucminas.br (D.C.C.); heronoh@gmail.com (H.O.H); gilmarsantos4@hotmail.com (G.B.S.)

2- Post-Graduate Program in Genetics, Federal University of Minas Gerais, Belo Horizonte 31270-901, Brazil; danielfonsecat@gmail.com (D.F.T.);

\*Corresponding author: Daniel Cardoso Carvalho (danielcarvalho@pucminas.br)

## Abstract

The study of ichthyoplankton is paramount to understanding fish assemblages' reproductive dynamics. DNA metabarcoding has been applied as a rapid, cost-effective, and accurate taxonomy tool, allowing the identification of multiple individuals simultaneously. However, there remain significant challenges when using DNA metabarcoding, such as molecular marker choice according to the taxonomic resolution and length of the fragment to be sequenced, primer bias, incomplete reference databases, and qualitative inferences

incongruences. Here, 30 ichthyoplankton pools collected from a Neotropical River were identified at a molecular level using DNA metabarcoding to compare the resolution, sensibility, specificity and relative read abundance (RRA) recovery of three molecular markers: the standard COI fragment (650pb, with each strand analyzed individually) and two short 12S rRNA genes markers ([?] 200bp - NeoFish and MiFish markers). The combined use of the three markers increased the genera detection rates by 25% to 87.5%, allowing an increased taxonomic coverage and robust taxonomic identification of complex neotropical ichthyoplankton communities. RRA is marker-dependent, indicating caution is still needed whilst inferring species abundance based on DNA metabarcoding data when using PCR-dependent protocols.

**Keywords: Molecular taxonomy, COI, 12S rRNA, High-throughput DNA sequencing 1. Introduction**

The study of ichthyoplankton composition, abundance and distribution is pivotal for understanding the reproductive dynamics of local fish assemblages (Mariac et al., 2018). The analysis of these parameters allows the identification of spawning sites, nursery areas where recruitment occurs, migration routes, temporal and spatial pattern variations and differences in the reproduction patterns of migratory and nonmigratory fish (Baumgartner et al., 2004; Bialecki et al., 2005; Reynalte-Tataje et al., 2012). This information is instrumental in elucidating the influence of anthropogenic environmental alterations on fish reproduction and in the definition of effective management actions for species conservation and, consequently, fishing stock maintenance (da Silva et al., 2015; Silva et al., 2017).

Traditionally, ichthyoplankton taxonomy has applied the regressive development sequence technique, based on the morphological comparison of younger larvae with previously identified juveniles (Ahlstrom and Moser, 1976; Nakatani et al., 2001). However, due to the absence of morphological diagnostic characters during the egg stage, some authors exclude them from the studies and resort to identifying exclusively larvae, which in the initial stages is also a difficult task (Baumgartner et al., 2008; Reynalte-Tataje et al., 2012). Moreover, the accuracy of the traditional morphological identification can diverge between taxonomists and laboratories, according to their experience and specialty (Ko et al., 2013). These limitations can compromise surveying essential information to conserve the areas of interest (Nobile et al., 2019).

Studies have employed molecular techniques to strengthen the precision and reliability of ichthyoplankton taxonomy. Comparative investigations have demonstrated that molecular taxonomy using DNA barcoding is more efficient than traditional morphological taxonomy, identifying the eggs and larvae to lower taxonomic levels and correcting erroneous morphological identifications (Becker et al., 2015; Ko et al., 2013). Using DNA barcoding, (Frantine-Silva et al., 2015) identified over 99% of 536 ichthyoplankton samples at species levels, including eggs, which accounted for 30% of the observed species richness. Morphologically, (Becker et al., 2015) identified eggs only as migratory or nonmigratory, when possible, while DNA barcoding allowed the identification of eggs (plus damaged larvae) to species level, and highlighted imprecisions in the morphological taxonomy even with such broader analysis. Nonetheless, despite its great taxonomic precision, DNA barcoding relies on individual processing and sequencing of each organism, and can become expensive and laborious for large scale inventories (Taberlet et al., 2012; Yu et al., 2012), such as ichthyoplankton studies (Mariac et al., 2018; Nobile et al., 2019).

The DNA metabarcoding approach, using High-Throughput Sequencing (HTS), has gained prominence for its ability to allow massive biodiversity access and transform ecology (Yu et al., 2012). The method combines DNA barcode-based taxonomy with HTS to simultaneously identify hundreds to thousands of organisms. DNA metabarcoding analyses are economical, quick, broad, minimally dependent on taxonomic expertise, and its data remain available for further verification (Taberlet et al., 2012; Yu et al., 2012). This approach has allowed the reconstruction of ancestral communities (Jorgensen et al., 2012), biodiversity monitoring (Andersen et al., 2012), and detection of larger operational taxonomic units in a fraction of the time spent in conventional studies based on morphology and DNA barcoding (Fonseca et al., 2010). This approach has also shown high efficiency in ecological ichthyoplankton studies, allowing precise and reliable identification of fish egg and larva bulk samples (Kimmerling et al., 2018; Mariac et al., 2018).

Different from environmental samples (for example, soil and water), in which genetic material is often degraded, bulk samples usually provide genomic DNA of better quality, allowing the amplification of markers with more extensive sequences (Taberlet et al., 2012). However, the HTS platforms accessible to most research laboratories have limited sequencing lengths of up to 600 base pairs (bp). This hampers the usage of markers previously standardized for DNA barcoding, such as the 650 bp fragment of the mitochondrial cytochrome c oxidase subunit I (COI) gene commonly used for fish (Ward, 2009). Additionally, the variability in COI sequences hinders the design of internal minibarcode primers, taking some researchers to pass this gene over in favor of more conserved ones for metabarcoding (Deagle et al., 2014). Among these conserved genes, mitochondrial 12S rRNA has been highlighted as a good alternative for fish metabarcoding (Milan et al., 2020; Miya et al., 2020; Sales et al., 2021).

Besides marker selection, another challenge in DNA metabarcoding is quantitative analysis. Some factors can bias the number of read copies obtained for each individual or species, such as the number of mitochondria per cell, different-sized individuals in the same sample, and amplification bias (Carvalho, 2022; Fonseca, 2018). Nonetheless, some studies have shown a positive correlation between the number of eggs or larvae in mock samples and the number of reads obtained for each taxon using DNA metabarcoding with an amplification step (Duke and Burton, 2021; Nobile et al., 2019).

This study used DNA metabarcoding to analyze the composition of ichthyoplankton sampled at the Neotropical megadiverse Sao Francisco River Basin, in Brazil. Additionally, the sensibility, specificity, and taxonomic resolution of two 12S markers were tested and compared with the traditional COI fragment used for DNA barcoding. The results obtained here will contribute to an improved method for ecological studies focusing on the ichthyofauna reproductive dynamics, and to design management and conservation strategies for the maintenance of fish reproduction locally.

## 2. Materials and Methods

### *Sample collection*

The Sao Francisco River Basin harbors at least 205 fish species (Alves et al., 2007), making it an excellent challenge for testing the sensibility and resolution of different markers within such complex Neotropical ichthyofauna. Thirty ichthyoplankton samples were used, denoted SF01 to SF30, 22 composed of eggs and eight composed of larvae. The larvae samples were assembled with similar-sized fragments of all specimens to minimize bias due to larvae size and relative read abundance estimative of input DNA from each pool.

### *Genomic DNA extraction*

Genomic DNA was extracted from samples containing a pool of fish eggs and larvae fixed in ethanol. To ensure the complete evaporation of the alcohol, initially the excess was removed through pipetting, and then the microtubes were kept open for three hours at 55degC. We then added 600µl of TNES buffer to each sample and ground the bulk with a plastic pestle until only minuscule tissue fragments were left. Next, 20µl of proteinase K (20mg/ml) was added to each microtube. The samples were kept at 55°C until complete tissue digestion. Finally, the genomic DNA was extracted using a low-cost saline protocol adapted from (Aljanabi and Martinez, 1997).

The samples were quantified using a Qubit 4 Fluorometer (Thermo Fisher) with a 1x dsDNA HS Assay Kit (Thermo Fisher) to verify the success of the DNA extraction. The samples were then diluted to 100ng/µl.

### *DNA amplification*

The DNA was amplified using via polymerase chain reaction (PCR). For the 12S rRNA gene, NeoFish (Milan et al., 2020) and MiFish (Miya et al., 2015) markers were amplified, and the fragment of the COI gene was amplified using a cocktail of primers targeting the standard COI fragment (Ward et al., 2005). Each sample was amplified in triplicate using the primer with a specific barcode tag for demultiplexing. The PCR reaction solution had a final volume of 20µl, containing: 8.34µl of ultrapure water, 0.16µl of BSA (100µg/ml), 10µl of AmpliTaq Gold 360 Master Mix (Thermo Fisher), 0.25µl of each primer, and 1,0µl of DNA template. An

additional 1,0 $\mu$ l of ultrapure water was added for the negative control samples instead of the DNA template. For the positive control sample, 1.0 $\mu$ l of template DNA from a saltwater fish species, *Prionace glauca*, was added.

PCR conditions consisted of initial denaturation for 10 min at 95°C, followed by 35 cycles of denaturation for 1 min at 95°C, primers annealing for 30 sec at 56°C (COI), 60°C (MiFish) or 63°C (NeoFish) and extension for 30 sec at 72°C, with a final extension for 7 min at 72°C.

The PCR results were checked using 1.8% agarose gel electrophoresis. Both 12S markers presented bands of the expected size for all samples, including the positive control, whereas for COI, the sample SF08 failed to produce any bands but was also included in the sequencing step. The negative control samples did not produce any bands but were also sequenced with the other samples.

#### *Library preparation and DNA sequencing*

According to the manufacturer's protocol, one library for each marker was prepared using the Collibri PCR-free Kit (Thermo Fisher). The libraries were quantified at the start of the preparation, after each major step, and at the end, by fluorometry. The 12S libraries were sequenced in a MiniSeq Sequencing System (Illumina) using a MiniSeq Mid Output Kit (300-cycle), and the COI library was sequenced in a MiSeq Sequencing System (Illumina) using a MiSeq Reagent Kit v3 (600-cycle).

#### *Bioinformatic analyses*

The bioinformatics analyses were carried out using the metabarcoding analysis package DADA2 (Callahan et al., 2016) and the Phyloseq package (McMurdie and Holmes, 2013). A pipeline in R v4.4.0 (R Core Team, 2021) was used for read quality control, removal of adaptors (Cutadapt, Martin, 2011), removal of sequencing errors and chimeric reads, reads merge and for obtaining the Amplicon Sequencing Variants (ASVs) distribution visualization and for taxonomic assignment. The taxonomic assignment was conducted in two rounds. First, using the DADA2 RPD classifier against a custom 12S database based on the database developed by Milan et al. (2020) for both 12S markers, containing 252 DNA sequences, with 181 specifically from Sao Francisco Basin. Secondly, using local BLASTn (Camacho et al., 2008) against the NCBI nucleotide database (Sayers, 2022; NCBIInt). Both 98% and 99% percentual identity thresholds were applied for identifications at the species level for COI and 12S, respectively. The RRA (relative read abundance) was determined by dividing the absolute counts of each ASV by the sum of the absolute counts of all ASVs in a sample.

To compare species identifications between markers, Venn Diagrams were built using the web application Lucidchart (<https://lucid.app/>). To examine the potential difference between marker choice on sample composition, a Permutational Multivariate Analysis of Variance (PERMANOVA) and principal coordinate analysis (PCoA) were performed with 1000 permutations, applying the Jaccard and Bray-Curtis dissimilarity indexes using the function 'adonis' (vegan 2.5-2 R package).

Due to the maximum 600 bp length limitation of the sequencing technology available, the forward R1 and reverse R2 COI reads could not be merged by overlap to reconstruct the barcoding amplicon, as each strand covers a different region of the COI gene with possible distinct variations for each taxon. Therefore, reads R1 and R2 were analyzed separately, and each sample's taxonomic assignment results were combined.

The ASVs found in the negative controls were removed from all other samples. Additionally, considering that the high throughput could amplify contaminations not detected by negative controls, and the risk of false positives, but also aiming not to exclude underrepresented taxa, only ASVs with more than 0.01% of relative read abundance (RRA) in each sample were considered.

### **3. Results**

All primer sets produced successful sequencing results for most samples. However, one sample (SF08) for COI did not produce any amplification, even after further DNA purification, quantification, and a new PCR adjustment. Nevertheless, SF08 was successfully amplified and sequenced using the 12S markers Mifish and

Neofish. A low number of reads were observed for SF04 and SF08, resulting in only one and three reads, respectively, despite the latter not presenting any problems in the amplification process, resulting in 93.33% (two failed samples out of 30) sequencing success. Notably, both 12S markers resulted in 100% amplification and sequencing success, with at least 38,697 reads in a sample (SF27) for MiFish, and 37,228 reads (SF10) for NeoFish.

The sequencing effort resulted in 4,505,309 reads for all markers after quality filtering. The number of reads showed considerable differences among markers and samples. COI produced 584,309 reads, averaging 19,477 reads per sample, ranging from one (SF04) to 29,485 (SF11). MiFish presented 1,919,545 total reads, averaging 63,985, with a minimum of 38,697 (SF27) and maximum of 83,734 (SF23). Sequencing with NeoFish resulted in 2,001,455 reads, with an average number of 66,715 per sample, varying from 37,228 (SF10) to 96,285 (SF23). After BLASTn searches, 1,699 COI reads remained without taxonomic assignment, and 69 were assigned to Bacteria. On the other hand, all MiFish and NeoFish reads were assigned to fish taxa.

ASVs were assigned to 26 fish taxa, from which 22 were identified at the species level, two at the genus level and three at the subfamily level. The 12S marker NeoFish was able to detect the highest number of orders, families, genera, and the same number of species as MiFish. In contrast, COI detected fewer species, genera, and families than the other markers and the same number of orders as MiFish (Table S1, Figs. 1, S1).

The combined use of the three markers increased the genera detection rates by 25% to 87.5% when considering an initial analysis with only NeoFish or COI, respectively (Fig. 2). The improvement in species recovery rates with the use of all three markers combined ranged from 31.25% to 61.54% when considering an initial analysis with either 12S gene markers or COI, respectively (Fig. 2).

The COI marker detected 16 taxa belonging to 13 species, eight genera, six families, three orders and one class (Fig. S1). Besides the 13 taxa identified at the species level, one was identified at the genus level, one at the subfamily level and another at the family level (Table S1). Among the 13 species, three were detected exclusively by the COI gene (*Bergiaria westermanni*, *Leporinus friderici* and *Prochilodus lineatus*). The species *B. westermanni* was present in 12 samples, with an average RRA of 2.29%, ranging from 0.13% to 6.16%. The anostomid *L. friderici* was detected in eight samples, with the RRA ranging from 0.13% to 32.64%, and an average of 7.81%. Lastly, *P. lineatus* was present in three samples and had an average RRA of 3.60%, ranging between 1.20% to 7.12% (Fig. 3).

With the 12S gene markers, 21 taxa were detected, including 18 species, 14 genera, 11 families, four orders and one class (Table S1). One of these taxa was identified at the genus level (*Characidium* sp.) and two at the subfamily level (Doradinae and Stevardiinae). Of the 18 identifications at species level, eight were exclusively detected with the 12S gene, all indigenous to the Sao Francisco Basin: *Cetopsorhamdia iheringi*, *Megaleporinus elongatus*, *M. reinhardti*, *Pachyurus squamipennis*, *Planaltina myersi*, *Pseudoplatystoma corruscans*, *Steindachnerina elegans* and *Sternopygus macrurus* (Fig. 1, Table S1).

Four of the eight species detected exclusively with the 12S markers were detected by both markers, but with some variation in samples and abundance. For example, the species *C. iheringi* was detected by MiFish and NeoFish in sample SF15, with 0.07% and 0.02% RRA, respectively (Fig. 3).

Overall, RRA and taxon detection was not consistent between each marker (Fig. 3). For instance, the most abundant taxon *P. pohli* had a total of 1,348,589 (70.79% of the total) for MiFish and 424,499 (73.02%) reads recovered for NeoFish. Notably, the highest RRA detected for NeoFish was *M. elongatus*, with 236,332 (20.90% of the total) reads. Additionally, in some samples (e.g., SF05, SF13, SF14 and SF15) where both MiFish and COI detected multiple taxa for Pimelodidae, NeoFish was not able to identify any taxa for this family (Fig. 3).

The PERMANOVA evidenced significant differentiation of fish communities using distinct molecular markers (Table 1). The influence of primer choice on taxa recovered in each sample was significant for both presence-absence (Jaccard) and RRA (Bray-Curtis) analyses (Table 1). While there is considerable overlap between

both 12S markers considering only presence-absence of each taxon in each sample, as can be seen in the PcoA plot (Fig. 4a), the analysis taking taxa abundance into account revealed a slight overlap between MiFish and COI (Fig. 4b).

#### 4. Discussion

DNA metabarcoding has become an essential tool for species inventory and monitoring. However, its use in identifying ichthyoplankton is still incipient in the Neotropics, with several methodological challenges and biases still needing to be tackled (Carvalho, 2023). Considering the crescent demand for innovative techniques to unravel the complex reproductive dynamics of fish communities for both research and practical applications, there is an underlying need for the continuous refinement of this methodology. The COI gene has been commonly used as the marker of choice (Mariat et al., 2018; Nobile et al., 2019) because of its well-established primers and complete reference sequences libraries encouraged by the global initiative Fish Barcode of Life. Here, each molecular marker recovers a distinct community structure when considering both quantitative and qualitative analysis.

Using a marker of choice still raises concerns since using several markers is still expensive when using HTS and because each marker has distinct amplification biases and taxonomic resolution (Deagle et al., 2014). The high interspecific variability of the COI gene, when compared to other mitochondrial genes (Hebert et al., 2003), can help differentiate closely related species. However, the same high variability creates the need to use universal degenerate primers with lower specificity than those designed for more conserved genes. Also, it hinders the design of internal minibarcodes for COI (Deagle et al., 2014). In the present study, while COI presented 93.33% sequencing success and some of its sequences remained unassigned or were assigned to Bacteria, both 12S markers were successfully sequenced for all samples, and all their sequences were assigned to fish taxa. Additionally, the technology used for sequencing limits the total fragment size to 600bp, precluding the merging of both COI strands from forming the full-sized barcode, therefore, each strand was analyzed independently. The loss of resolution power caused by this could explain why COI detected fewer species, genera, and families than the 12S markers and why two of the three exclusive species level identifications were assigned to nonnative fishes closely related to species from Sao Francisco.

Minibarcodes for the 12S gene have been developed and applied to environmental DNA metabarcoding studies (Milan et al., 2020; Miya et al., 2020; Sales et al., 2021) and, more recently, to ichthyoplankton studies as well (Jiang et al., 2022; Van Nynatten et al., 2023). One of the main concerns when using these markers is the conserved nature of the gene, which can impact their ability to differentiate closely related species, especially in diverse regions. However, the current study shows that both MiFish and NeoFish were able to successfully identify and distinguish multiple congeneric species, such as *Leporinus piau* and *L. taeniatus*, *Megaleporinus elongatus* and *M. reinhardti*, *Pimelodus fur*, *P. maculatus* and *P. pohli*, and *Prochilodus argenteus* and *P. costatus*. Moreover, the 12S markers have a higher species detection sensibility than COI, considering that the exclusive fishes they retrieved were underrepresented, with low RRA. This could result from low-efficiency primer binding by COI, which can lead to a lack of amplification (Zhang et al., 2020).

Database completeness is another variable that directly impacts species detection, as a lack of reference sequences for a given species may hamper accurate taxonomic assignment (Collins et al., 2019). This aspect has affected both COI and 12S markers in this study. For instance, while *Pachyurus squamipennis* is not represented by any COI reference sequence in the public databases and was exclusively detected by 12S, the only native species retrieved solely by COI, *Bergiaria westermanni*, does not have any 12S representative sequence in neither the public nor our custom library. These limitations highlight the importance of continuous sequencing efforts to broaden reference sequence databases, especially for megadiverse regions.

Considering that each marker has advantages and limitations, some studies suggest combining multiple primer sets to increase taxonomic coverage (Liu and Zhang 2021; Zhang et al., 2020). In a metabarcoding study using multiplexed markers to identify zooplankton mock communities, Zhang et al. (2018) demonstrated that a multi-maker approach can improve species detection and allow the cross-validation of taxa

detected by each marker. Our results support this conclusion, as using the three markers combined increased the genera detection by up to 87.5% and species detection by up to 61.54%. Therefore, employing multiple markers, whenever feasible, reduces the likelihood of overlooking species or incorrectly classifying them due to the absence or mislabeling of sequences in the reference database (Locatelli et al., 2020).

Discrepancies between markers were observed in the quantitative analysis using the RRA estimative. Although some studies with mock samples of eggs (Duke and Burton, 2020) and larvae (Nobile et al., 2019) yielded a positive correlation between input organisms and output reads for each species, the results from this study support the idea that amplification bias is one of the main pitfalls for quantitative metabarcoding analyses, as already reported (Carvalho, 2023; Fonseca, 2018). While MiFish and COI presented RRA resemblance for samples in which both detected similar taxa, low-efficiency primer binding to Siluriformes and especially Pimelodidae sequences resulted in completely different abundance patterns for NeoFish.

In conclusion, using multiple markers from two distinct genes and lengths allowed an increased taxonomic coverage and robust taxonomic classification of complex neotropical ichthyoplankton communities. Finally, precaution is still needed when inferring species abundance based on DNA metabarcoding data when using PCR-dependent protocols since it is marker dependent. Nonetheless, ichthyoplankton metabarcoding offers superior resolution and feasible scalability compared to traditional techniques, and provides qualitative information, which is paramount for characterizing reproducing species and definition of conservation strategies.

### Acknowledgements

We are grateful to Amanda Jessica Silva Santos for bench work assistance, Guilherme Moreira Santos and Gustavo Ribeiro Rosa for helping with taxonomic assignment curation and Izabela Santos Mendes for contributing to statistical analysis. This work has been financially supported by FAPEMIG, CAPES, FIP-PUC Minas and ANEEL/CEMIG (P&D 0624). We thank CNPq for providing a research fellowship (312102/2022-4) to DCC.

### Data availability statement

Raw Illumina sequences and corresponding metadata are deposited in the Zenodo repository (DOI 10.5281/zenodo.8038835).

### References

- Ahlstrom, E. H., & Moser, H. G. (1976). Eggs and larvae of fishes and their role in systematic investigations and in fisheries. *Revue des Travaux de L'institut des Peches Maritimes* , 40 (3), 379-398.
- Aljanabi, S. M., & Martinez, I. (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic acids research* , 25 (22), 4692-4693.
- Alves, C. B. M., Vieira, F., Pompeu, P. S., & Ministerio do Meio Ambiente. (2011). Ictiofauna da bacia hidrografica do rio Sao Francisco. *Diagnostico do macrozoneamento ecologico-economico da bacia hidrografica do rio Sao Francisco. Ministerio do Meio Ambiente, Brasilia* , 226-241.
- Andersen, K., Bird, K. L., Rasmussen, M., Haile, J., Breuning-Madsen, H., Kjaer, K. H., . . . Willerslev, E. (2012). Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Molecular Ecology* , 21 (8), 1966-1979.
- Baumgartner, G., Nakatani, K., Gomes, L. C., Bialecki, A., & Sanches, P. V. (2004). Identification of spawning sites and natural nurseries of fishes in the upper Parana River, Brazil. *Environmental Biology of Fishes* , 71 (2), 115-125.
- Baumgartner, G., Nakatani, K., Gomes, L. C., Bialecki, A., Sanches, P. V., & Makrakis, M. C. (2008). Fish larvae from the upper Parana River: do abiotic factors affect larval density?. *Neotropical Ichthyology* , 6 , 551-558.

- Becker, R. A., Sales, N. G., Santos, G. M., Santos, G. B., & Carvalho, D. C. (2015). DNA barcoding and morphological identification of neotropical ichthyoplankton from the Upper Parana and Sao Francisco. *Journal of Fish Biology* , 87 (1), 159-168.
- Bialetzki, A., Nakatani, K., Sanches, P. V., Baumgartner, G., & Gomes, L. C. (2005). Larval fish assemblage in the Baia River (Mato Grosso do Sul State, Brazil): temporal and spatial patterns. *Environmental Biology of Fishes* , 73 , 37-47.
- 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. *Nature methods* , 13 (7), 581-583.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC bioinformatics* , 10 , 1-9.
- Carvalho, D. C. (2022). Ichthyoplankton DNA metabarcoding: Challenges and perspectives. *Molecular Ecology* , 31 (6), 1612-1614.
- Collins, R. A., Bakker, J., Wangensteen, O. S., Soto, A. Z., Corrigan, L., Sims, D. W., ... & Mariani, S. (2019). Non-specific amplification compromises environmental DNA metabarcoding with COI. *Methods in Ecology and Evolution* , 10 (11), 1985-2001.
- Da Silva, P. S., Makrakis, M. C., Miranda, L. E., Makrakis, S., Assumpcao, L., Paula, S., ... & Marques, H. (2015). Importance of reservoir tributaries to spawning of migratory fish in the upper Parana River. *River Research and Applications* , 31 (3), 313-322.
- Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., & Taberlet, P. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology letters* , 10 (9), 20140562.
- Duke, E. M., & Burton, R. S. (2020). Efficacy of metabarcoding for identification of fish eggs evaluated with mock communities. *Ecology and evolution* , 10 (7), 3463-3476.
- Fonseca, V. G. (2018). Pitfalls in relative abundance estimation using eDNA metabarcoding. *Molecular Ecology Resources* . 18 (5), 923-926.
- Fonseca, V. G., Carvalho, G. R., Sung, W., Johnson, H. F., Power, D. M., Neill, S. P., ... & Creer, S. (2010). Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature communications* , 1 (1), 98.
- Frantine-Silva, W., Sofia, S. H., Orsi, M. L., & Almeida, F. S. (2015). DNA barcoding of freshwater ichthyoplankton in the Neotropics as a tool for ecological monitoring. *Molecular Ecology Resources* , 15 (5), 1226-1237.
- Hebert, P. D., Ratnasingham, S., & De Waard, J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences* , 270 (suppl\_1), S96-S99.
- Jiang, R., Lusana, J. L., & Chen, Y. (2022). High-Throughput DNA Metabarcoding as an Approach for Ichthyoplankton Survey in Oujiang River Estuary, China. *Diversity* , 14 (12), 1111.
- Jorgensen, T., Haile, J., Moller, P. E. R., Andreev, A., Boessenkool, S., Rasmussen, M., ... Willerslev, E. (2012). A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. *Molecular Ecology* , 21 (8), 1989-2003.
- Kimmerling, N., Zuerqert, O., Amitai, G., Gurevich, T., Armoza-Zvuloni, R., Kolesnikov, I., ... Sorek, R. (2018). Quantitative species-level ecology of reef fish larvae via metabarcoding. *Nature ecology & evolution* , 2 (2), 306-316.
- Ko, H. L., Wang, Y. T., Chiu, T. S., Lee, M. A., Leu, M. Y., Chang, K. Z., ... & Shao, K. T. (2013). Evaluating the accuracy of morphological identification of larval fishes by applying DNA barcoding. *PLoS*



*One* , 8 (1), e53451.

Liu, J., & Zhang, H. (2021). Combining multiple markers in environmental DNA metabarcoding to assess deep-sea benthic biodiversity. *Frontiers in Marine Science* , 8 , 684955.

Locatelli, N. S., McIntyre, P. B., Therkildsen, N. O., & Baetscher, D. S. (2020). GenBank's reliability is uncertain for biodiversity researchers seeking species-level assignment for eDNA. *Proceedings of the National Academy of Sciences* , 117 (51), 32211-32212.

Mariac, C., Vigouroux, Y., Duponchelle, F., Garcia-Davila, C., Nunez, J., Desmarais, E., & Renno, J. F. (2018). Metabarcoding by capture using a single COI probe (MCSP) to identify and quantify fish species in ichthyoplankton swarms. *PLoS One* , 13 (9), e0202976.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* , 17 (1), 10-12.

McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one* , 8 (4), e61217.

Milan, D. T., Mendes, I. S., Damasceno, J. S., Teixeira, D. F., Sales, N. G., & Carvalho, D. C. (2020). New 12S metabarcoding primers for enhanced Neotropical freshwater fish biodiversity assessment. *Scientific Reports* , 10 (1), 1-12.

Miya, M., Gotoh, R. O., & Sado, T. (2020). MiFish metabarcoding: a high-throughput approach for simultaneous detection of multiple fish species from environmental DNA and other samples. *Fisheries Science* , 86 (6), 939-970.

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., ... Iwasaki, W. (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 (7), 150088.

Nakatani, K., Agostinho, A.A., Baumgartner, G., Bialetzki, A., Sanches, P.V., Makrakis, M.C., Pavanelli, C.S. (2001). Ovos e larvas de peixes de agua doce: desenvolvimento e manual de identificacao, first ed. EDUEM, Maringa.

Nobile, A. B., Freitas-Souza, D., Ruiz-Ruano, F. J., Nobile, M. L. M., Costa, G. O., De Lima, F. P., ... & Oliveira, C. (2019). DNA metabarcoding of Neotropical ichthyoplankton: Enabling high accuracy with lower cost. *Metabarcoding and Metagenomics* , 3 , e35060.

R Core Team (2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing* . Vienna, Austria. <https://www.R-project.org/>.

Reynalte-Tataje, D. A., Agostinho, A. A., Bialetzki, A., Hermes-Silva, S., Fernandes, R., & Zaniboni-Filho, E. (2012). Spatial and temporal variation of the ichthyoplankton in a subtropical river in Brazil. *Environmental Biology of Fishes* , 94 , 403-419.

Sales, N. G., Wangensteen, O. S., Carvalho, D. C., Deiner, K., Praebel, K., Coscia, I., ... Mariani, S. (2021). Space-time dynamics in monitoring neotropical fish communities using eDNA metabarcoding. *Science of the Total Environment* , 754 , 142096.

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., ... Sherry, S. T. (2023). Database resources of the National Center for Biotechnology Information in 2023. *Nucleic Acids Research* , 51 (D1), D29.

Silva, C. B., Dias, J. D., & Bialetzki, A. (2017). Fish larvae diversity in a conservation area of a neotropical floodplain: influence of temporal and spatial scales. *Hydrobiologia* , 787 , 141-152.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology* , 21 (8), 2045-2050.

Van Nynatten, A., Gallage, K. S., Lujan, N. K., Mandrak, N. E., & Lovejoy, N. R. (2023). Ichthyoplankton metabarcoding: An efficient tool for early detection of invasive species establishment. *Molecular Ecology Resources* .

Ward, R. D. (2009). DNA barcode divergence among species and genera of birds and fishes. *Molecular Ecology Resources* , 9 (4), 1077-1085.

Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., & Hebert, P. D. (2005). DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B: Biological Sciences* , 360 (1462), 1847-1857.

Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution* , 3 (4), 613-623.

Zhang, G. K., Chain, F. J., Abbott, C. L., & Cristescu, M. E. (2018). Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. *Evolutionary applications* , 11 (10), 1901-1914.

Zhang, S., Zhao, J., & Yao, M. (2020). A comprehensive and comparative evaluation of primers for metabarcoding eDNA from fish. *Methods in Ecology and Evolution* , 11 (12), 1609-1625.

### Figure caption

Figure 1. Venn diagrams recovered by each molecular marker MiFish, NeoFish and COI considering distinct taxonomic levels: (a) order, (b) family/(subfamily), and (c) genus/species. Species marked with an asterisk (\*) have not been reported for the Sao Francisco River Basin.

Figure 2. Improvement in detected taxa when using a single marker for the COI gene, combining COI and Mifish and combining COI, Mifish and Neofish.

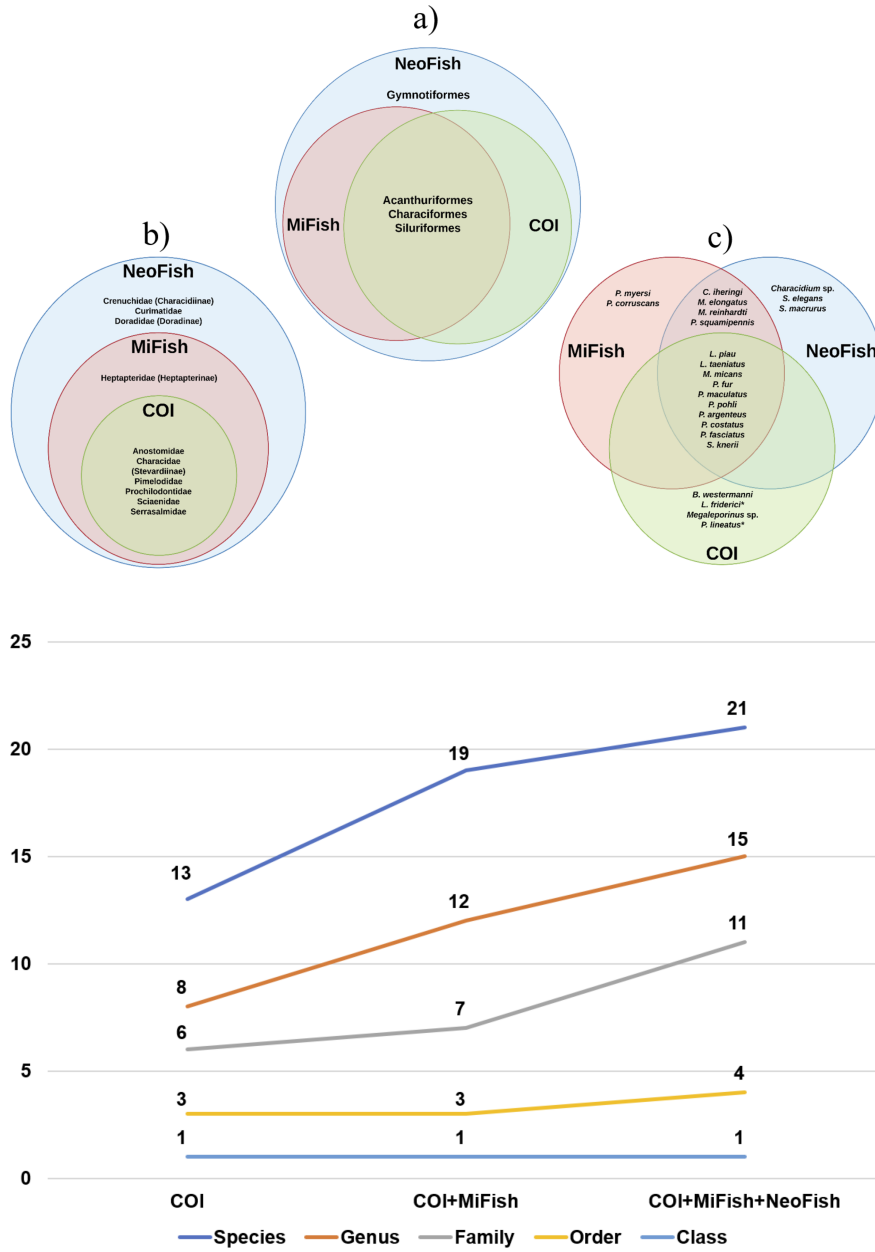
Figure 3. Relative read abundance (RRA) for each taxon, in each sample, with each marker.

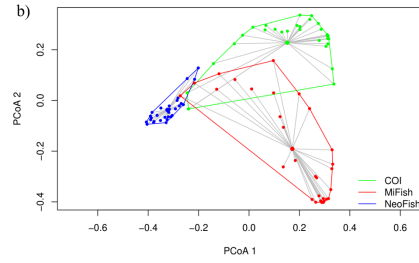
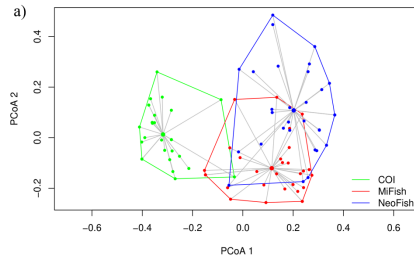
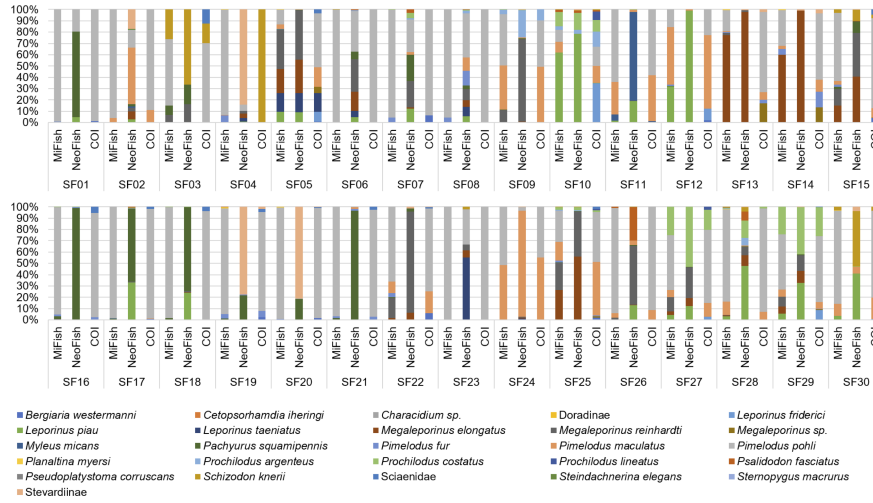
Figure 4. PCoA plots of 30 ichthyoplankton fish communities detected using three different molecular markers: COI, Mifish and Neofish. Analyses were conducted using (a) Jaccard coefficient considering the presence/absence of fish tax and (b) the Bray-Curtis coefficient, which uses a matrix of abundance based on each taxon.

### Table caption

Table 1. Summary of PERMANOVA results (R<sup>2</sup>-effect sizes and significance level)

showing the effect of marker choice on taxa recovered. Df= Degree of freedom; Sum of Squares, Value of Test F; and p-value associated with the F score.





## Hosted file

Table 1.docx available at <https://authorea.com/users/630890/articles/650481-qualitative-and-quantitative-dna-metabarcoding-assessment-of-neotropical-ichthyoplankton-communities-is-marker-dependent>