

RESEARCH ARTICLE

Insights into the representativeness of biodiversity assessment in large reservoir through eDNA metabarcoding

Thainá Cortez¹*, André Torres²*, Murilo Guimarães³, Henrique Pinheiro¹, Marcelo Cabral⁴, Gabriel Zielinsky⁴, Camila Pereira⁵, Giovanni de Castro⁵, Luana Guerreiro⁵, Juliana Americo¹, Danielle do Amaral¹‡, Mauro Rebelo¹‡

1 Bio Bureau Biotechnology, Rio de Janeiro, Rio de Janeiro, Brazil, **2** Laboratório de Biologia, Controle e Vigilância de Insetos Vetores, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Rio de Janeiro, Brazil, **3** Biology Institute, Federal University of Piauí, Teresina, Piauí, Brazil, **4** Prooceano, Rio de Janeiro, Rio de Janeiro, Brazil, **5** Carlos Chagas Filho Institute of Biophysics, Federal University of Rio de Janeiro, Rio de Janeiro, Rio de Janeiro, Brazil

* These authors contributed equally to this work.

‡ DA and MR also contributed equally to this work.

* thainacortez@gmail.com



OPEN ACCESS

Citation: Cortez T, Torres A, Guimarães M, Pinheiro H, Cabral M, Zielinsky G, et al. (2025) Insights into the representativeness of biodiversity assessment in large reservoir through eDNA metabarcoding. *PLoS ONE* 20(1): e0314210. <https://doi.org/10.1371/journal.pone.0314210>

Editor: Petr Heneberg, Charles University, CZECHIA

Received: May 17, 2024

Accepted: November 7, 2024

Published: January 24, 2025

Copyright: © 2025 Cortez et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The sequences data underlying the results presented in the study are available from NCBI SRA (BioProject PRJNA1105563).

Funding: This work was financed by Tijoá Energia through the Agência Nacional de Energia Elétrica (ANEEL) Research and Development Program (Grant PD-09151-2001/2020). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Monitoring biodiversity on a large scale, such as in hydropower reservoirs, poses scientific challenges. Conventional methods such as passive fishing gear are prone to various biases, while the utilization of environmental DNA (eDNA) metabarcoding has been restricted. Most eDNA studies have primarily focused on replicating results from traditional methods, which themselves have limitations regarding representativeness and bias. In our study, we employed eDNA metabarcoding with three markers (12SrRNA, COI, and 16SrRNA) to evaluate the biodiversity of an 800 km² reservoir. We utilized hydrodynamic modeling to determine water flow velocity and the water renewal ratio throughout the study area. Additionally, we conducted statistical comparisons—rarefaction curves and multivariate methods—among samples as an alternative approach to assess biodiversity representation. The eDNA identified taxa previously documented in the reservoir by traditional monitoring methods, as well as revealed 29 –nine fishes and 20 non-fish—previously unreported species. These results highlight the robustness of eDNA as a biodiversity monitoring technique. Our findings also indicated that by randomly sampling 30% of the original number of samples, we could effectively capture the same biodiversity. This approach enabled us to comprehend the reservoir's biodiversity profile and propose a straightforward, cost-effective monitoring protocol for the future based on eDNA.

Introduction

Increasing the accuracy of biodiversity estimation has been a challenge in the face of lacking detailed information about population dynamics, habitat-based models, and either the occurrence or abundance of a species, especially for rare or elusive species [1–3]. Advancements in

Competing interests: The authors have declared that no competing interests exist.

technology and methods have evolved over time from identifying individual organisms based on specific DNA barcodes to the current capability to analyze hundreds of thousands of DNA barcodes obtained from environmental samples (e.g. water, soil, air), allowing the description of entire communities in terms of species composition and dynamics [4]. This high-throughput approach, known as 'DNA metabarcoding,' is especially potent at uncovering biodiversity [e.g. 5–8].

Monitoring biodiversity through environmental DNA (eDNA) poses enormous challenges, especially in aquatic systems, due to the vast scales involved, and the combination of environmental conditions such as water current speed and density, wind, and tides, which influence particle transport [9]. Hydrodynamic modeling is an alternative approach for investigating the environmental influence on the transport of molecules. Such modeling approaches offer physical insights into the water body by considering outflow volume, water renewal rate, current velocity, and flow direction. This information together can enhance our understanding of the fundamental processes affecting the spatial and temporal (including seasonal) distribution of eDNA [10–12].

In recent years, eDNA has gained acceptance due to its non-intrusive nature and cost-effectiveness [13, 14]. It has proven highly effective because eDNA metabarcoding studies involve multiple sampling sites with several replicates at each location, and, based on the assumption that eDNA in aquatic systems is dispersed by water flow and begins to degrade over time, increasing the number of sampling sites enhances the detection of a broader range of species [15, 16]. This approach addresses many concerns scientists and regulators have regarding the effectiveness of traditional environmental programs, which may struggle with issues related to longevity, continuity, scale, or statistical significance [17, 18]. Several approaches can help assess the optimal experimental design for a given study area. For instance, rarefaction curves can reveal the correlation between sampling size and the number of sequences per taxon (or Molecular Operational Taxonomic Unit, MOTUs), demonstrating the effectiveness of both the number of samples collected and the molecular approach [19]. Although this analysis is often employed in eDNA studies to determine whether the number of sampling sites or the molecular protocol sufficed to capture diversity [20–22], it has rarely been used to assess the optimal number of samples in large reservoirs with hundreds of square kilometers [23, 24]. In this context, multivariate analyses, such as Principal Component Analysis (PCA) and Discriminant Analysis of Principal Components (DAPC), can offer insights into the distribution of variation across all sites, indicating which samples contribute most to the overall variability [19]. Finally, diversity analyses, including richness, Shannon's H index, and species frequency computation, can complement these findings by providing visualizations that enable the recognition of community patterns and differences. These combined methods can inform how comprehensive the experimental design and the applied molecular approach were.

The aim of this study was to build a large-scale biodiversity inventory of a HPP reservoir utilizing eDNA and compare its findings to the previous reports either published in the scientific literature or filed over decades with regulators. The robustness of the inventory, as well as the most cost-effective experimental design and optimal number of samples, were investigated. Toward this goal, hydrodynamic modeling was initially conducted at an 800 km² HPP reservoir, Três Irmãos (TI, São Paulo, Brazil), coupled with pilot sampling at a single site to gain insight into local diversity using eDNA metabarcoding and three distinct markers (COI, 16SrRNA and 12SrRNA). Subsequently, multiple stations were sampled across the reservoir to evaluate large-scale biodiversity. Multivariate analyses, rarefaction curves, and community comparisons were employed to assess the effectiveness of the sampling design in compiling a comprehensive biodiversity inventory of the reservoir.

Material and methods

Study area

The Três Irmãos (TI) hydroelectric power plant reservoir is located at the lower Tietê River basin, in São Paulo State in the Southeastern region of Brazil (Fig 1A). A 9,600 m long artificial channel called Pereira Barreto—the second-largest freshwater canal in the world—connects the TI HPP reservoir to the nearby Ilha Solteira HPP reservoir on the São José dos Dourados River (Fig 1A). This enables coordinated water management of the two reservoirs and integrated energy operation of the two hydroelectric plants, both of which drain into the Paraná River.

Três Irmãos is the largest power plant on the Tietê River. Its operation began in November 1993 (CESP 2013). The reservoir spans 133 km in length, with a main central water channel width of about 5 km. Its largest arm stretches 23 km in one direction and extends another 20 km in the opposite direction. It covers a surface area of 815 square km and holds 13.4 billion cubic meters of water. The reservoir has a flooded area of 817 km², a volume of 13,800 x 10⁶ m³, and an average depth of 17 meters. Inflow into the TI reservoir is limited primarily by the

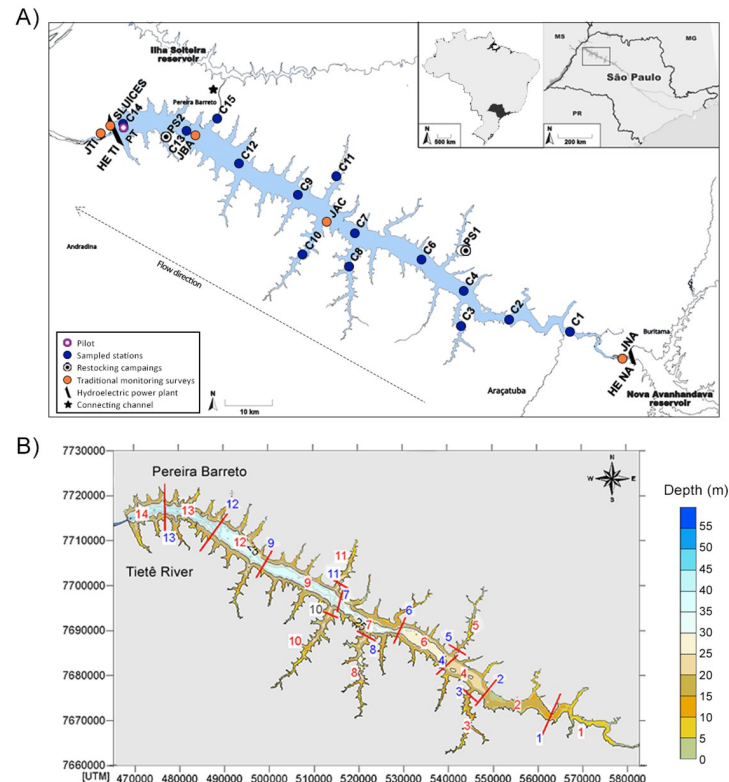


Fig 1. Trés Irmãos (TI) reservoir and hydrodynamic modeling compartments. A) The TI and all sampling sites are shown. Purple circle shows the pilot sampling location. Blue circles labeled C1-C4 and C6-C15 designate the positions of the 14 sampling stations along the TI reservoir. Stations abbreviations as in S1 Table. White circles (PS1 and PS2) indicate where fish restocking programs are conducted. Orange circles (JNA, JAC, JBA, SLUICES and JT) show the locations where the traditional taxonomic surveys are conducted by the hydroelectric company. Black trapezoid bars represent the positions of the Três Irmãos (UHE TI) and Nova Avanhandava (UHE NA) hydroelectric dams. Black star shows the Pereira Barreto service channel that connects the TI and IS reservoirs. B) For water renewal computations by hydrodynamic modeling, the reservoir was divided into 14 compartments. Red bars delineate the boundaries of the compartments. Compartment numbers are in red; the blue numbers are the control sections in which time series flows were calculated with the hydrodynamic model.

<https://doi.org/10.1371/journal.pone.0314210.g001>

Nova Avanhandava (NA) HPP upstream of the reservoir, and secondarily by the flows across the Pereira Barreto channel which vary according to the operational requirements of the two hydroelectric power plants. The TI reservoir area has multiple uses, such as fish farming, professional and recreational fishery, tourism associated with boat trips, and cattle breeding, making it an anthropic area [25, 26].

Hydrodynamic modeling

The modeling domain, used for hydrodynamic modeling, covered the entire water mirror of the reservoir. For the water renewal ratio and stream speed computations, the hydrodynamic simulations utilized the Environmental Hydrodynamics Base System (SisBAHIA[®], <http://www.sisbahia.coppe.ufrj.br/>), a model used for natural water bodies under different scenarios [27]. The two-dimensional module of the FIST3D (Filtered in Space and Time 3D) hydrodynamic model was used to simulate water flow velocity and rate of water renewal during rainy and dry seasons. The FIST optimized spatial discretization allows for an optimal presentation of jagged contours and complex bathymetries. Bathymetric data was interpolated in the Surface Mapping System software (Surfer) using the Kriging regression method, with a spacing of 100 meters to obtain the Digital Terrain Model (DTM) for hydrodynamic modeling (S1 Fig). For the reservoir contour and spatial delimitation, the spatial discretization of the model was performed via biquadratic quadrangular finite elements. The water velocity was measured for both rainy (April 1 to June 30, 2017) and dry (October 23, 2020 to January 21, 2021) seasons for the entire area of the reservoir. The mesh has a grid resolution ranging from around 300 meters in the tributary rivers to approximately 2 km in the Três Irmãos Reservoir main body (S1 Fig).

Daily inflows from the NA reservoir and the arms where the Mato Grosso brook and Baixote stream are, and daily outflows of the TI dam and Pereira Barreto channel, in addition to the fluvial inflow of the tributary and the winds, were used as variables or constraints in the model. Due to the reservoir's large size (~132 km in length), a compartmentalized analysis of the rate of water renewal during a simulated three-months period was applied. The region was divided into 14 compartments (Fig 1B). The conservation of mass principle was applied to a generic substance diluted in the water mass with concentration C .

Pilot sampling

Pilot sampling was conducted in September 2020 at one site in the reservoir (Fig 1A). Water sampling was conducted in accordance with ethical guidelines and Brazilian legal regulations. No permits were required for the described study, which complied with all relevant regulations. To evaluate the best depth for sampling, three 1L water samples were collected at the surface, middle, and bottom of the water column (1-, 13-, and 25-meter depths, referred to as P01, P13 and P25, respectively) (S1 Table). Three independent samples (R1, R2 and R3) were collected at each depth, resulting in nine samples. For each sample, one liter of water was collected utilizing a Van Dorn bottle, then stored in sterilized polypropylene bottles (Nalgon, Cat. No. 2330) at 8° C until filtered using 0.22 µm Sterivex filters (Merck-Millipore, Cat. no. SVGP01050) in 60 mL increments using disposable syringes. After filtration, the remaining liquid was expelled by injecting air, and the filters were stored at 4° C in sterile Longmire's buffer (Tris 100 mM, EDTA 100 mM, NaCl 10 mM, SDS 0.5%) until the DNA extraction. Non-disposable equipment was cleaned with sodium hypochlorite solution (10% v/v) and rinsed with autoclaved distilled water to avoid cross-sampling contamination.

Sampling stations in Três Irmãos Reservoir

After analyzing the data from both pilot sampling and the hydrodynamic modeling (see [Results](#) section), 14 new stations distributed along the entire reservoir were sampled in October 2021 to establish a broader baseline, hereafter called stations C1–C4 and C6–C15 ([Fig 1A](#) and [S1 Table](#)). Sampling stations covered each compartment used to evaluate water renovation ratios across the river. Stations C1, C2, C4, C6, C7, C9, C12, C13, and C14 were located in the main water body. Four stations were placed in areas that comprise the arms with (C3, C8, C10) and without tributaries (C11). One station (C15) was located at the entrance to the Pereira Barreto channel that connects the TI reservoir to the neighboring IS reservoir ([Fig 1A](#)).

Three replicates (R1, R2 and R3) were collected within each station, all 1 meter above the benthic bottom, resulting in 42 samples at different depths across among the 14 stations ([S1 Table](#)). The replicates were collected within approximately 10-minutes intervals. All collection, processing, and storage procedures were the same as those described above for the pilot sampling.

DNA extraction and sequencing

DNA extraction from the Sterivex filters was performed using the DNeasy[®] Blood & Tissue kit (Qiagen) with adaptations [28]. The DNA concentration was determined using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). DNA samples are registered in the Brazilian National System for the Management of Genetic Heritage and Associated Traditional Knowledge (SISGEN), accession number A7E42BC.

The amplicons from mitochondrial genes targeting the minibarcodes of cytochrome c oxidase I (COI), 16SrRNA (16S), and 12SrRNA (12S) were amplified and sequenced using the pair of primers described in [S2 Table](#). Amplification was performed in 25 µl using Platinum[™] Multiplex PCR Master Mix (Applied Biosystems), 1 µM of each primer, and 5 µl of DNA. The thermocycling conditions included an initial denaturation step at 95°C for 2 minutes and then 25 denaturation cycles at 95°C for 30 seconds each, annealing at different temperatures for each primer pair (48°C for COI; 57°C for 12S and 53°C for 16S) for 90 seconds, and extension at 72°C for 30 seconds, followed by a final extension at 72°C for 10 minutes and finishing at 4°C. For the library preparation, the amplicons were indexed using Nextera XT indexes (IDT for Illumina). Library quality control was performed using TapeStation (Agilent). DNA concentration was normalized, and all amplicons were pooled in one library. The library was sequenced on Illumina NovaSeq 6000 S4 Flow Cell 2x150.

Bioinformatic analysis

After sequencing, raw sequence reads were evaluated using FASTQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) and MultiQC (<https://multiqc.info/>). All generated fastq files were imported into the QIIME2 pipeline [29] for further quality filtering and taxonomic classification. Primer trimming, read denoising, merging the remaining paired reads into amplicons, and chimera removal steps were carried out with DADA2 [30]. Here, all forward and reverse reads with less than 145 nucleotides overlapping were removed, as well as all reads with quality < 30 in Phred scale. The resulting Amplicon Sequence Variants (ASVs) were clustered into MOTUs (97% of identity among sequences) with DADA2. The taxonomy identification was performed using two different methods: 1) Naive Bayesian classifier [31] trained with the sequences' dataset of the 12S Mitohelper [32], COI classifier version 4.0 [33] and the MIDORI 16S dataset [34]; and 2) BLASTn searches [35] using the curated

mitochondrial reference database available in NCBI (<https://ftp.ncbi.nlm.nih.gov/refseq/release/mitochondrion/>), and the nucleotide non-redundant (nr) database [36]. Only hits with an e-value lower than $1e^{-50}$ and a query cover higher than 90% were allowed. ASVs and MOTUs that matched off-target species were discarded. Off-target species included non-eukaryotic organisms, human, domesticated animals (*Bos taurus*, *Canis familiaris*, *Felis catus*, *F. silvestris*, *Gallus gallus*, *Meleagris gallopavo*, *Mus musculus*, *Ovis vignei*) and marine animals (*Chimaera sp.*, *Coryphaena equiselis*, *C. hippurus*, *Scomber sp.*, *Homarus americanus*, *Rhinosomus triqueter*).

Ichthyofauna communities

For the pilot sampling, a Binomial Generalized Linear Model (GLM) was employed to estimate the probability of encountering fish species at each depth level, allowing for the assessment of the optimal sampling design. Heatmaps of the log₁₀ frequencies of MOTUs were constructed based on MOTUs for fish species. MOTUs frequencies were computed considering all three markers (CO1, 16S, and 12S) combined.

Based on fish MOTUs identified only up to the species level, ecological analyses of alpha- and beta-diversity were conducted in R. Two indices of alpha diversity were calculated with the vegan R [37] package within each locality: the Shannon Index (H') [38], and richness (number of MOTUs). To compare the diversity indices across the 14 locations (beta-diversity), we constructed bar plot charts to visualize the species diversity in terms of fish orders. All graphs were constructed in R scripts using the *ggplot2* package [39].

Evaluation of the eDNA-based biodiversity inventory completeness

To evaluate the completeness of our data, eDNA results were compared to the biodiversity data of TI built from Global Biodiversity Information Facility data (GBIF, <https://www.gbif.org/>) and a non-systematic literature review. In addition, both PCA and DAPC multivariate methods and rarefaction curve computation were employed. Using this integrative approach, we were able to assess the most informative samples and identify potentially redundant information. Only fish MOTUs identified up to the species level were considered, avoiding potential bias in biodiversity assessments.

The PCA was conducted to capture the variation among samples within the first two principal components (PCs), considering both MOTUs abundance and richness per locality. Individual and scree plots were generated using the *factoextra* package [40] from R v. 3.3.2 [41], respectively. The DAPC was implemented to cluster samples based on compositional similarity. Because DAPC maximizes differences among groups, it enables the exploration of sample similarities and the identification of potentially redundant samples in the experimental design. The optimal number of PCs was determined by the *optim.a.score* function from the *ade4* R package [42], and genetic clusters were identified using the *find.cluster* function. Individual cluster assignments from DAPC were visualized using the *assignplot* function from *ade4*. The contribution of each sample to the overall variability was assessed through computed eigenvalues within each PC.

The impact of reducing sampling effort on capturing local variation was tested through the creation of four sub-datasets, three of them lacking one of the three most contributing samples and one containing only the three most contributing samples. Rarefaction curves were then computed using all complete and sub-datasets using the function *specaccum* from the *vegan* R package [43].

Results

Hydrodynamic modeling shows a lentic environment and minimal water exchange in the TI Reservoir

Interpolation of bathymetry from the Três Irmãos (TI) Reservoir showed that the average depth of the region was 18.3 meters. The TI Reservoir is a predominantly lentic environment in both rainy and dry seasons, with more intense currents upstream, next to the Nova Avanhadava (NA) dam, and downstream near the TI dam. Hydrodynamic modeling results show that the water flows from the NA to the TI dam. The water velocity is higher near the main inflow into the reservoir at the NA dam (upstream in the reservoir) in both rainy and dry seasons (S2 Fig). Upstream, currents are more intense due to the topography of the channel, reaching up to 1.33 m/s and 0.56 m/s in the rainy and dry season, respectively. In the central region, the flow is very slow, with very low velocities (< 0.05 m/s) in rainy and dry seasons during all the modeled periods in the sections closest to the TI dam (S2 Fig). In the regions just upstream from the TI dam, currents are more intense than in the central region, but approximately 10% weaker than in the upstream region, presenting averages of 0.04 m/s in the rainy season and 0.02 m/s in the dry season.

Water renewal in the reservoir compartments varies between rainy and dry seasons. During floods, water renews faster, peaking at 50 days and then increases exponentially. Compartment 1 (closest to NA dam) renews 50% of its water in less than five days during the rainy season but takes 10 days in dry periods. In compartments 2, 4, and 6, a similar trend is observed for the first 50 days. After 70 days, sections 1, 2, and 4 have more than 95% of their water renewed during floods. Sections 1 and 2 in dry periods reach over 95% renewal in 90 days, while compartment 4 reaches about 80%. The furthest compartments (12, 13, and 14) only receive "new" water during the rainy season, renewing up to 10% after 50 days (S3 Fig).

The results indicated that the water flow in the arms tends to stagnate, regardless of the time of year, meaning there is a low exchange of water between the main flow and the arms. Taken together, the reservoir depth, water velocity, and water renewal rate supported the distribution of stations for water sampling, where they are allocated to capture eDNA in the different hydrological profiles of the reservoir and the arms.

Bottom sampling and three-marker analyses improve the chances of fish eDNA capture

The pilot sampling obtained an average of 1.5 million of 150 bp paired-end reads per depth. The raw sequence data are publicly available at the SRA database under the BioProject accession code PRJNA1105563. After the read cleaning, amplicon merging, chimera removal, and clustering of identical amplicons, 615, 2,993, and 4,404 ASVs remained for 12S, 16S, and COI datasets, respectively (S3 Table). After applying the Bayesian classifier from QIIME2 and removing all off-target species, COI exclusively identified 24 taxa, while 16S found 3 and 12S, none. These datasets merged resulted in 27 non-fish taxa, which belonged to Rotifera, Copepoda, Annelida, Algae and Viridiplantae clades (S4 Table).

Regarding ichthyofauna, the GLM revealed that sampling at the benthic bottom (25 m) increases the chance of capturing fish eDNA (S4 Fig). COI, 12S and 16S detected five, four and one taxa, respectively, resulting in 10 taxa. Four species were identified by both 16S and COI, one by 12S and COI, and one by 12S and 16S. Only 4 species were detected by all three mini-barcodes (S4 Table).

eDNA metabarcoding uncovers 171 MOTUs in the TI reservoir, including 29 previously unrecorded species

As described in the Methods section, following the pilot assessments, a sampling campaign covering the entire reservoir through 14 stations was performed. After sequencing the 42 samples from 14 stations, 67.9 million paired reads for 12S, 31.1 million for 16S, and 74.9 million for COI were obtained (S3 Table). The raw sequence data are also publicly available at the SRA database under the BioProject accession code PRJNA1105563. Following trimming and clustering steps, 2.9 million amplicons were identified for 12S, 22 million for 16S, and 14.8 million for COI. A total of 1187, 3687, and 7437 ASVs remained for 12S, 16S, and COI datasets, respectively. Subsequently, the classification using the Bayesian classifier from QIIME2 and removal of all off-target species, the ASVs clusters that share 97% of identity presented 34, 41, and 190 MOTUs, respectively.

The biodiversity inventory of TI Reservoir, combining historical and eDNA-based data, encompasses 577 taxa (S5 and S6 Tables). The dataset was divided into non-fish (S5 Table) and fish (S6 Table) to permit a broader view of the fishes detected, the main group of interest. The final dataset of non-fish metazoan resulted in 422 taxa, where 36 (~8%) were also detected in this study (S5 Table). Of these, 20 taxa (~5% of the 422) were exclusively detected using eDNA, with species belonging to Annelida (Clitellata and Polychaeta), Arthropoda (Copepoda, Ostracoda, Insecta, and Branchiopoda), Chordata (Avians) and Rotifera. The remaining taxa, previously reported in the literature, include Annelida, Arthropoda, Chordata, and Mollusca species (Fig 2 and S5 Table).

Arthropods, especially copepods, were highly abundant across all stations, including the pilot station. We also found four non-fish taxa never previously reported in the TI reservoir, even though they had been reported in upstream reservoirs and other rivers of the Paraná Basin: an Olygochaeta (*Chaetogaster diastrophus*), a copepod (*Acanthocyclops vernalis*) and two rotifers (*Polyarthra remata* and *Synchaeta* sp.). SAR (Stramenopila, Alveolata, and Rhizaria) MOTUs belonging to the Stramenopila clade were not identified beyond this level because of the lack of reference sequences.

Metabarcoding eDNA matched 12 species previously identified in TI through traditional taxonomy: two annelids (*Branchiura sowerbyi* and *Limnodrilus* sp.), one arthropoda (*Diaphanosoma spinulosum*), three birds (*Nycticorax nycticorax*, *Mycteria americana* and *Phalacrocorax brasilianus*), one mammal (*Hydrochoerus hydrochaeris*), one mollusc (*Limnoperna fortunei*), and four rotifers (*Asplanchna* sp., *Euchlanis dilatata*, *Polyarthra* sp. and *Trichocerca* sp.). Although some taxa remained unclassified, most belong to groups previously found in TI and other rivers of the Paraná Basin (S6 Table). Nonetheless, a few species present in previous reports in the literature were not detected through DNA sequences obtained in this work, such as Cyanophyta, predominant during 2011 [43, 44]. No macro-crustacean was detected, including the exotic *Macrobrachium amazonicum* and *Macrobrachium jelskii*, both previously found in TI [45]. Insects [46], Molluscs [46], except for *L. fortunei*, and Platyhelminthes already reported in this area were not detected either.

For the ichthyofauna, the resulting inventory comprised 135 species (S6 Table), with 35 species being identified through both literature review and eDNA surveys. In comparison, 81 species (~60%) were only documented in the literature over the past 50 years. Nine species detected with eDNA metabarcoding had not previously been recorded in TI: one cypriniform (*Gobio gobio*), seven characiforms (*Hemiodus orthonops*, *Hoplias aimara*, *Hyphessobrycon erythrostigma*, *Leporinus piau*, *Schizodon fasciatus*, *Schizodon knerii*, *Metynniss melanogrammus*), and one cichliform (*Crenicichla cyanotus*), here found in multiple sites (Fig 3). Approximately 60% of all species were captured through both eDNA and traditional monitoring in

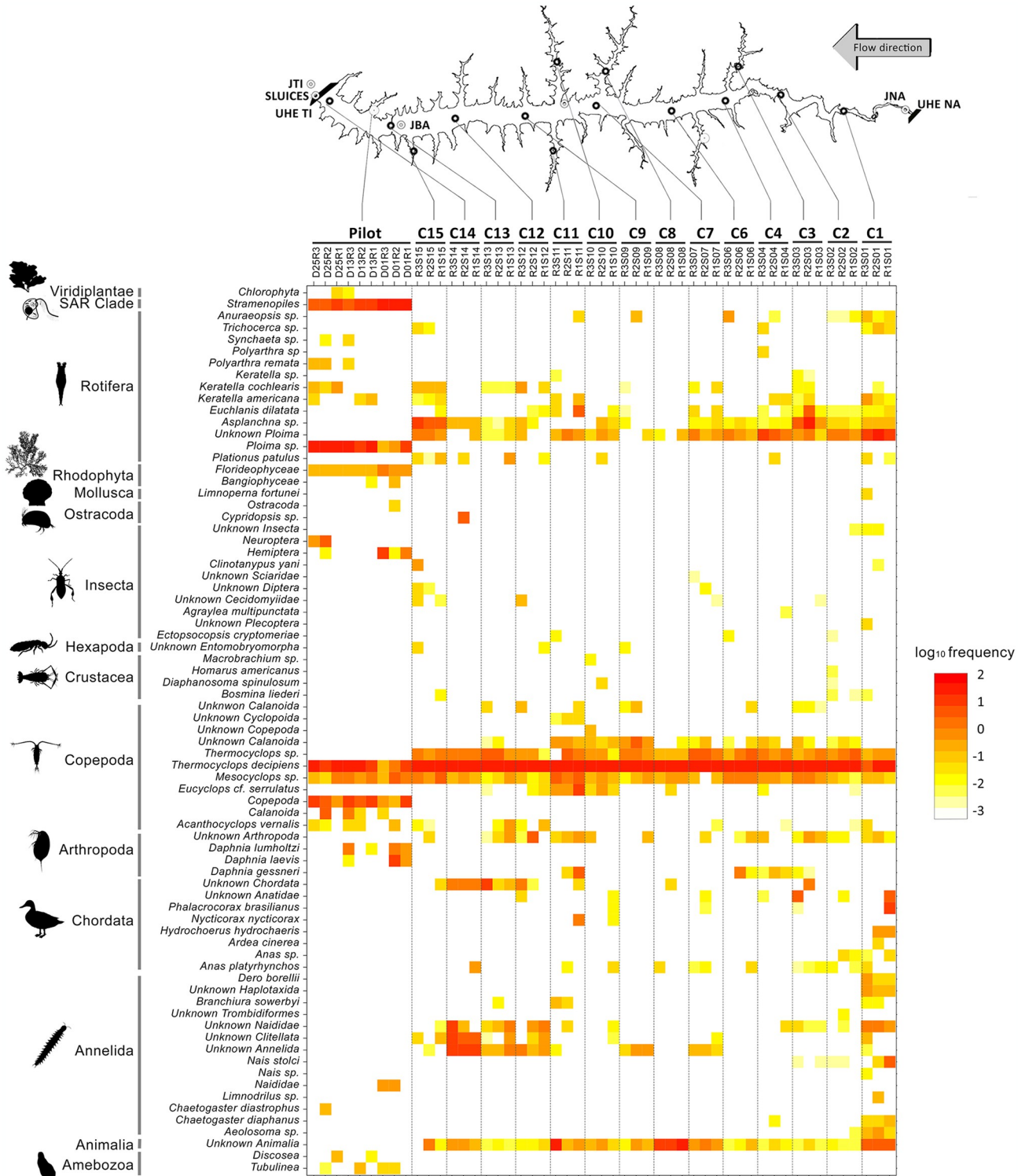


Fig 2. Non-fish taxa distribution and frequency across the Três Irmãos (TI) reservoir. Heatmap illustrating the distribution of non-fish Molecular Operational Taxonomic Units (MOTUs) identified by eDNA across sampling stations in the Três Irmãos (TI) reservoir. The TI reservoir is shown at the top with all sampled stations C1-C4 and C6-C15. Stations abbreviations as in S1 Table. Hydroelectric power plants Três Irmãos (TI) and Nova Avanhandava (NA) are shown at the extremes of the reservoir. The yellow to red color spectrum represents the log₁₀ transformation of the average frequency. Solid black lines on the left delineate the observed phyla.

<https://doi.org/10.1371/journal.pone.0314210.g002>

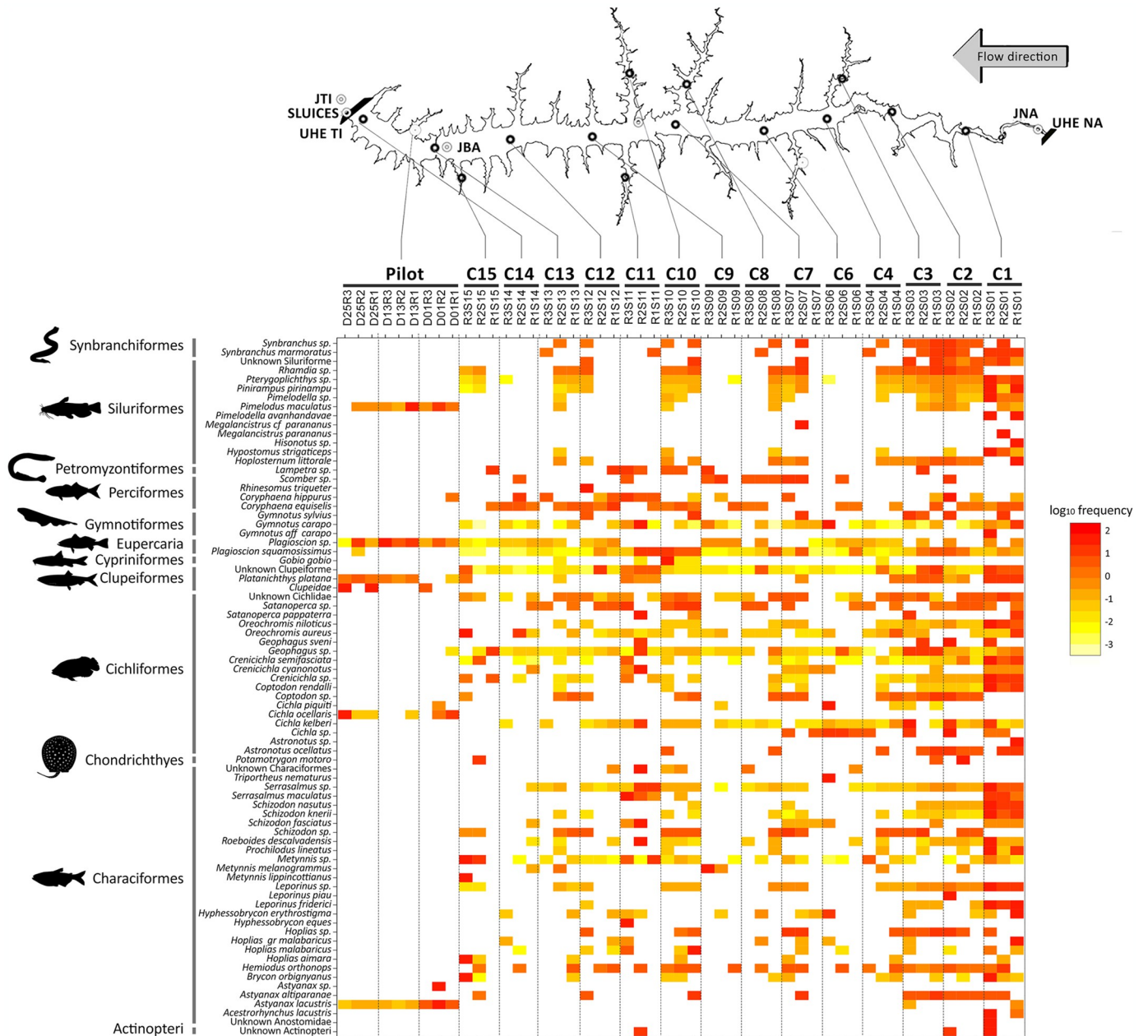


Fig 3. Ichthyofauna distribution and frequency across the Três Irmãos (TI) reservoir. Heatmap illustrating the distribution of fish Molecular Operational Taxonomic Units (MOTUs) identified by eDNA across sampling stations in the Três Irmãos (TI) reservoir. The TI reservoir is shown at the top with all sampled stations C1–C4 and C6–C15. Stations abbreviations as in S1 Table. Hydroelectric power plants Três Irmãos (TI) and Nova Avanhandava (NA) are shown at the beginning and at the end of the reservoir. The yellow to red color spectrum represents the log₁₀ transformation of the average frequency. Solid black lines on the left delineate the observed orders.

<https://doi.org/10.1371/journal.pone.0314210.g003>

2020–2021, including *Roebooides descalvadensis*, *Hoplias malabaricus*, *Crenicichla semifasciata*, *Cichla kelberi*, *C. piquiti*, *Satanoperca pappaterra*, *Serrasalmus maculatus*, *Oreochromis niloticus*, *O. aureus* and *Plagioscion squamosissimus*, for instance. Only 19 species (~14%) had GBIF records for the TI reservoir.

Seventy-seven fish species were identified through eDNA when considering all locations from both sampling and station samples (Fig 3). These species are distributed among nine

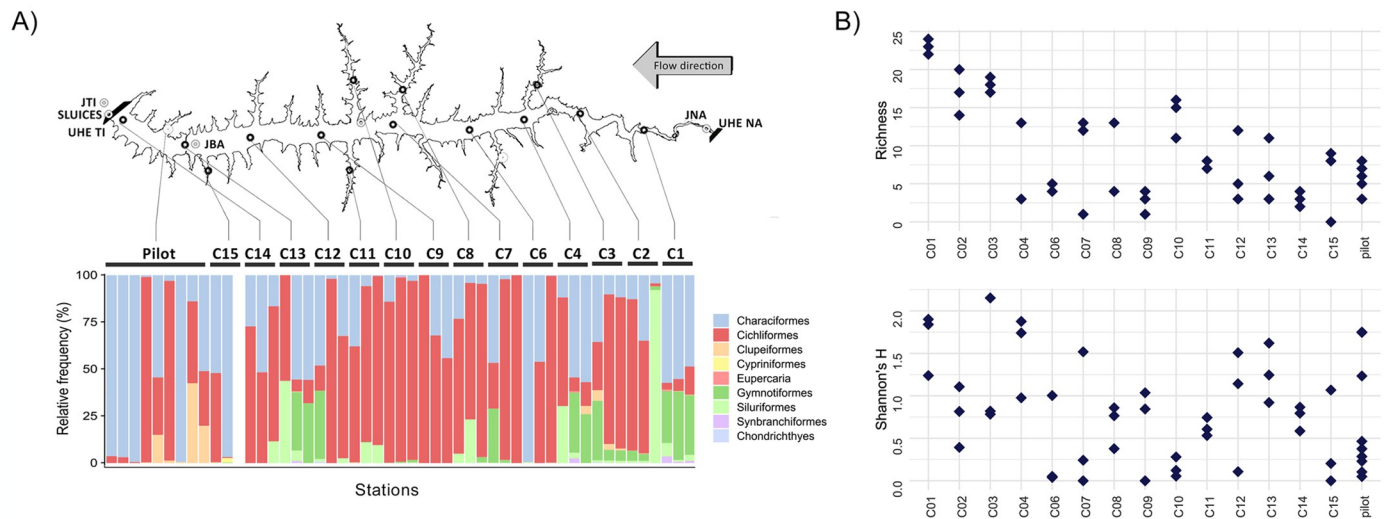


Fig 4. Alpha diversity results for ichthyofauna taxa detected in the Três Irmãos (TI) reservoir. Based on fish Molecular Operational Taxonomic Units (MOTUs) identified through eDNA per sampling station, A) the barplot displays all orders found within the Três Irmãos (TI) reservoir. Colors indicate the orders as shown in the legend. The TI reservoir is shown at the top with all sampled stations C1-C4 and C6-C15. Stations abbreviations as in S1 Table. B) Richness and Shannon's H index (y-axis) were computed for each station (x-axis). Each diamond represents one sample collected at the respective station.

<https://doi.org/10.1371/journal.pone.0314210.g004>

orders, namely Cichliformes, Characiformes, Chondrichthyes, Cypriniformes, Eupercaria, Gymnotiformes, Myliobatiformes, Siluriformes, and Synbranchiformes. The Cichliformes and Characiformes orders were the most abundant in almost all stations (Fig 4A), followed by Gymnotiformes, represented here by species of the genus *Gymnotus*, exhibiting significant presence, especially in station C01. The order Clupeiformes, represented by a single species (*Platanichthys platana*), was found in the pilot sampling, and identified in three other stations (C03, C04, and C15). Chondrichthyes species were detected in several stations, but in low frequencies, ranging between 3% and 8% per sample.

Station C01 exhibited higher variability in terms of the number of taxa and also had higher MOTU frequencies compared to all other stations, accounting for ~54% of all species found across all locations, including the pilot sampling. It also revealed some exclusive species, such as *Acestrorhynchus lacustris*, *Pimelodella avanhandavae*, and *Gymnotus carapo*. The only autochthonous species identified was the Siluriforme *Pimelodus maculatus*. The other six species are allochthonous (originally from other basins) such as the *Plagioscion squamosissimus* (Corvina) and the three cichliforms (*Cichla piquiti*, *C. kelberi*, and *C. ocellaris*) all native to the Amazon.

From the six native fish species known to be introduced annually into the TI reservoir as part of a local restocking program, *Brycon orbygnianus* (Piracanjuba) and *Prochilodus lineatus* (Curimbatá) were detected by eDNA metabarcoding at nine and seven sampling stations, respectively. The other four species that are part of the restocking program *Piaractus mesopotamicus* (Pacu-guaçu), *Pseudoplatystoma corruscans* (Pintado), *Salminus brasiliensis* (Dourado) and *Leporinus elongatus* (Piapara) were not detected with eDNA.

Diversity variation among samples from the same station is greater than across stations

Despite being collected only minutes apart, diversity indexes, i.e. richness and Shannon's H, of the ichthyofauna communities showed greater variation among samples from the same station

than across stations (Fig 4B). For instance, the C12 station presented richness in one of the replicates that was almost two times greater than the other two replicates (12 compared to five and three). Stations C01, C02, and C03 exhibited the greatest richness, whereas the highest Shannon's H values were observed in stations C03, C01, and C04, respectively.

A quarter of the sampling effort was sufficient to reach the biodiversity plateau of the TI reservoir

A PCA was built with only the fish MOTUs identified to the species level. The individual position in the graph indicated a high similarity in composition among the samples, except for those from station C01 (C01R1, C01R2, and C01R3) and a few from stations C02 (C02R3), C03 (C03R1) and C07 (C07R2) (Fig 5A). These samples were collected at stations considerably distant from all others, potentially suggesting the presence of species exclusive to these sites.

In contrast, no distinct clustering pattern was observed for all other sites, indicating that despite the pilot sampling being conducted one year earlier, the composition remained very similar, preventing the segregation of these samples in the ordination space. However, the maximization of differences between these groups reinforces the pattern observed in the PCA,

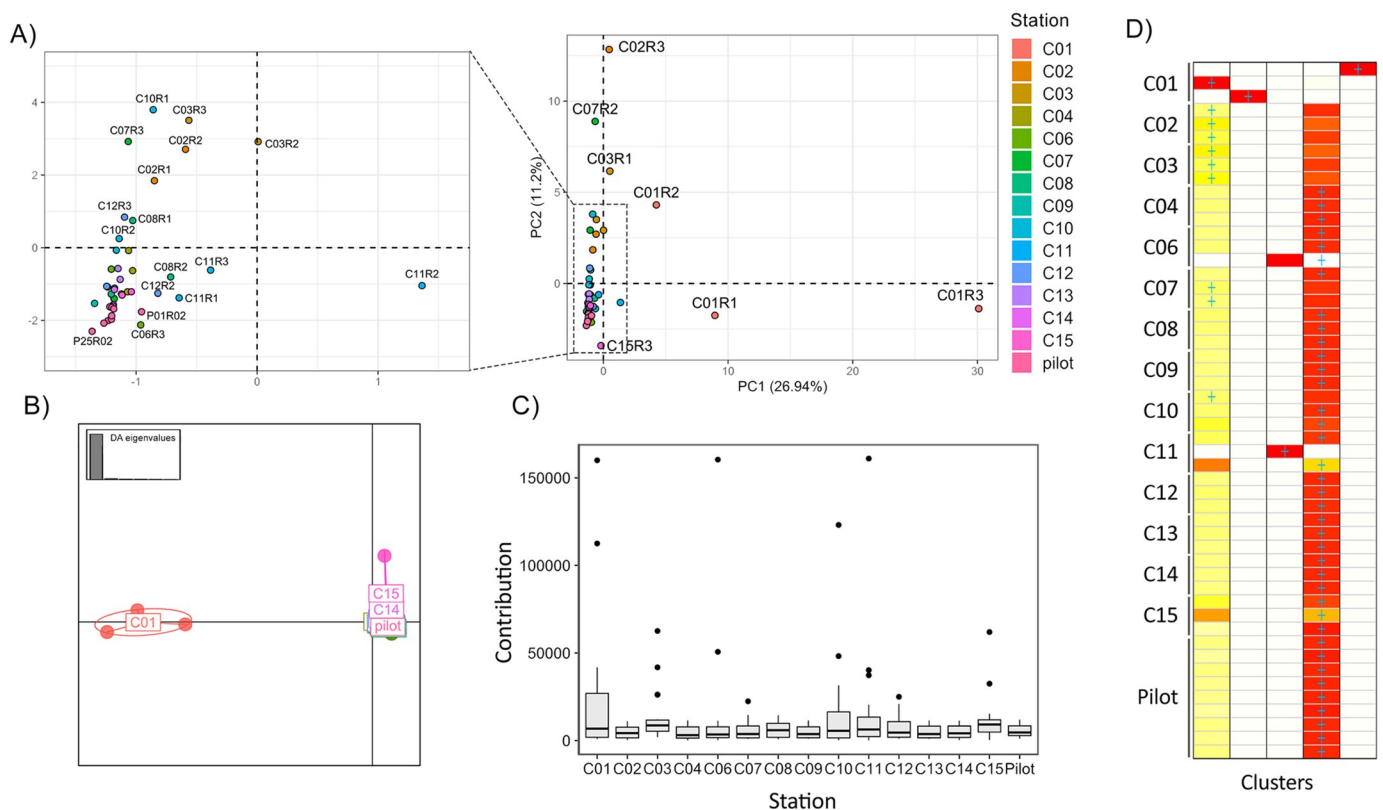


Fig 5. Beta diversity results for ichthyofauna taxa detected in the Três Irmãos (TI) reservoir. Based on fish Molecular Operational Taxonomic Units (MOTUs) identified through eDNA per sampling station, A) Principal Component Analysis (PCA) illustrates the variance in community composition. Each dot represents a sample from a station, with colors designating stations, as in the legend. Stations abbreviations as in S1 Table. Samples plotted close to each other are zoomed in the minor graph on the left side. B) Discriminant Analysis of Principal Components (DAPC) exhibits the maximization of differences among groups (6 PCs, $K = 5$ according to the BIC method). The colors correspond to those in PCA. C) Boxplot shows the contribution of each station and their samples to the segregation into five clusters. D) Assign plot illustrates each individual sample's assignment, considering $K = 5$. Darker shades of red indicate a higher chance of the sample belonging to that cluster.

<https://doi.org/10.1371/journal.pone.0314210.g005>

where only C01 appears different from all others, and the segregation among the remaining locations becomes indistinguishable in the ordination space (Fig 5B). When analyzing the sums of contributions from each station to the six principal components, stations C01, C10 and C11 exhibited the highest indices (Fig 5C and S5 Fig). By using six PCs according to the alpha score results, which reached 98.7% of the total variation, the DAPC identified five distinct groups: three groups comprised of one sample each from station C01, one group comprised of one sample each from C06 and C11, and the fifth group comprised of all the remaining samples (Fig 5D). In addition, a single sample from station C15 presented equal chances of belonging to two distinct groups.

The composition graph of all five clusters revealed that samples from C01 belong to different clusters, suggesting that even being sampled within the same station, the three obtained samples were different enough to be considered distinct community groups (Fig 5D). After these findings, using the samples with the highest contribution for the six PCs, rarefaction curves were computed for the datasets: i) containing all samples; ii) without C01; iii) without C11; iv) without C10; and v) containing only C01, C11, and C10 samples. For the complete dataset, as well as for those lacking one sample, the curves reached the plateau between 10–12 samples. The graphs indicated minimal changes in the plateau area, with slightly more noticeable differences observed when removing C01, suggesting that additional samples would be needed to fully capture the observed diversity (Fig 6). When C10 and C11 were removed, the results were relatively similar and did not exhibit a clear alteration in the curve shape. The rarefaction curve built using only C01, C10, and C11 samples almost reached the plateau area after nine samples, the total number (S5 Fig).

Discussion

This study provides an important insight into the optimal sampling strategy when conducting eDNA surveys in hydroelectric power plant reservoirs. Here, through the combination of hydrodynamic modeling and pilot sampling, we could delineate an optimal experimental design for the TI reservoir. In this study, we managed to perform a biodiversity baseline of a large reservoir for the first time in South America.

Metabarcoding successfully detected a substantial number of species previously documented in the literature over the past 50 years, along with nine fish species never reported in the reservoir before. We found that larger sampling sizes might not be the best approach when assessing biodiversity in lentic reservoirs using eDNA. Instead, it appears preferable to sample fewer locations with more replicates. Moreover, a one-year interval between sample collections did not seem to significantly influence diversity measures or the detection of species. Lastly, our data provides a valuable contribution to future biodiversity monitoring studies in reservoirs, aiming to achieve optimal utilization of resources with lower costs and better cost-effectiveness through molecular data.

This study contributed 20 new previously unreported non-fish species for the TI area (~5%). Overall, there was a high diversity of non-fish taxa, with annelids and rotifers found across several samples (Fig 2). Copepods, on the other hand, were abundant but less diverse, mainly represented by the genera *Thermocyclops* and *Mesocyclops*. Chordates were predominantly represented by birds with marine habitats, such as herons (*Ardea cinerea*) and Anatidae species (ducks, geese, and swans), found just upstream from the dam (C01-C03) and in the middle of the reservoir (C04, C07, C10, and C12). The only non-avian chordate detected was the capybara (*Hydrochoerus hydrochaeris*) in C01, a mammal commonly found in riverside and lakeside areas [47]. Contrary to what would be expected, the invasive golden mussel species *Limnoperna fortunei*, highly abundant in the TI HPP intake [48], was detected only in

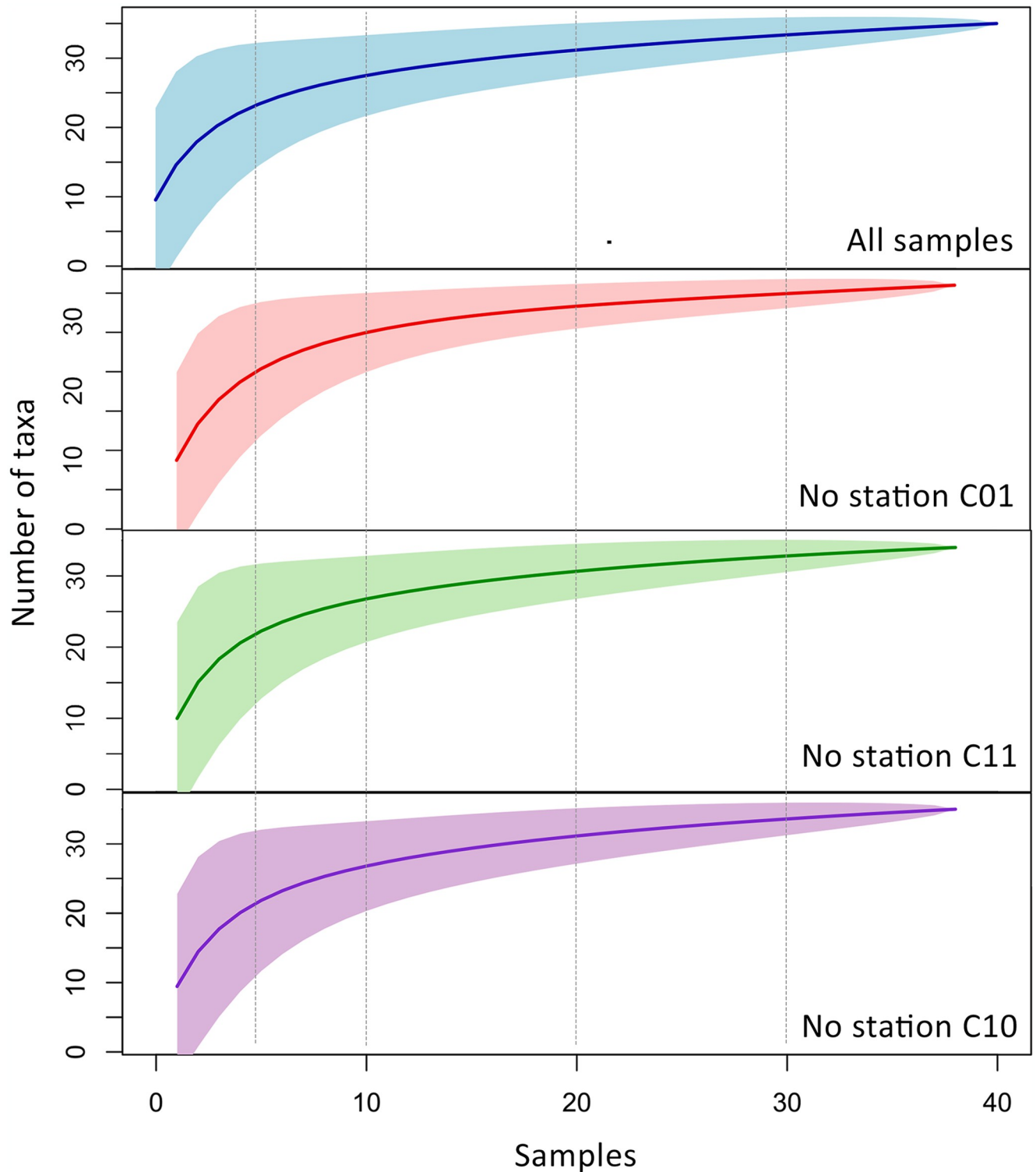


Fig 6. Rarefaction curves constructed based on fish richness using distinct sub-datasets. From top to bottom, the curves represented were built using all data, without station C01, without station C11 and without station C10, respectively. The number of taxa is indicated in the y-axis, and the number of samples (42 in total), in the x-axis.

<https://doi.org/10.1371/journal.pone.0314210.g006>

C01. The animal's body size and the primers chosen have direct influence on the eDNA survey; low DNA concentration [49] or the lack of amplification [16] or could explain its absence. Additionally, since most non-fish species are microscopic or from meiofauna, there is a potential for false negatives, i.e., the erroneous absence of a species. This bias can affect diversity estimates and, consequently, impact comparisons between traditional surveys and eDNA assessments [50]. Regarding the ichthyofauna, the eDNA survey identified nine species unreported to date, which corresponds to a ~6% increase in all species in the inventory of the study area. Nine major fish orders were detected across all samples, reinforcing previous reports for the TI. We identified abundant carnivorous species such as *O. aureus*, *O. niloticus*, *C. piquiti*, *C. kelberi*, *C. ocellaris* and *Plagioscion squamosissimus*. In fact, three of the most abundant species captured by traditional monitoring during 2021 [51] were also among the most frequently detected at multiple stations using eDNA (*P. squamosissimus* and *C. kelberi*), supporting our findings.

We also detected two native *Crenicichla* species, one of the richest groups in terms of species among the South American cichlids [52]. From the six taxa never previously reported in the TI reservoir, *Hemiodus orthonops* was detected at high frequencies at all sampling points. Some species were detected using traditional methods but not using eDNA, including *Crenicichla haroldoi*, *Cichla monoculus*, *Metynnis maculatus*, *Moenkhausia* sp., *Astyanax fasciatus* and *Leporinus*, for instance. Most of the species identified through traditional methods but not detected by eDNA have been reported only sporadically in local monitoring reports. This may suggest a seasonal reproductive or migratory behavior that makes them less consistently detectable [53]. Furthermore, the slow water flow, as those found for most of our sites according to hydrodynamic modeling, may hinder the transport and preservation of eDNA over extended periods, making species detection more challenging [54, 55]. This potential for seasonal presence could explain the absence of these species in the eDNA samples, despite their detection through traditional survey techniques. Even though it is tempting to look for an ecological explanation for the presence or lack of some species in certain locations, the absence of a species should not be solely attributed to its lack of occurrence since it can be a false negative. Several factors like species occupancy [56, 57], eDNA persistence and quantity [58, 59], primer bias [16] and DNA concentrations decrease [49] may influence the distribution pattern of eDNA. Following the same logic, the frequency of DNA fragments in the heatmap represents a rough estimation of the abundance of the species. Also factor such PCR and/or errors might produce false positives due errors can occur due to random mutations during amplification or misreading during sequencing [60]. For these reasons, the eDNA abundance and/or absence should be interpreted with caution as it may reflect body size of individual species [61] or drift [62], even for fishes [63].

When compared to the short-term traditional monitoring conducted during 2020–2021, eDNA successfully detected 60% of all fish taxa captured using traditional methods and more than 80% of all fish genera identified during the last 50 years (S6 Table). Several studies comparing eDNA metabarcoding and established monitoring methods have been conducted in rivers [19, 64, 65] or estuaries [52]. Consistent with our findings, most of these studies have shown that eDNA methods exhibit similar performance compared to traditional methods. For example, Goutte *et al.* (2020) found that eDNA was more efficient than traditional surveys lasting 3 to 6 years.

In traditional community ecology, which relies on specimen counting, the distribution of organisms is usually clustered under the influence of environmental and biological factors such as physical barriers, nutrient availability and competition. However, eDNA sampling is not bound by the same constraints. Instead, it is influenced by drift, sedimentation, and degradation rates [5, 58, 61, 63] and, once released, eDNA tends to homogenize in the water flow,

resulting in a gradient-like distribution that becomes more uniform over time. These processes effectively blur the boundaries of the originally distinct organismal clusters, illustrating the distinct dynamics between direct ecological observations and molecular ecological data. Here, eDNA could also detect species never reported for the study area, compensating for some of the limitations of the traditional monitoring methods, such as erroneous identification, dedicated time, and seasonal changes [66]. These findings support eDNA metabarcoding as an effective method when evaluating fish communities, as already affirmed in previous studies [67]. For this reason, the eDNA can be employed for rapid and large-scale biodiversity assessments for monitoring programs [17], reducing costs and minimizing interventions in aquatic environments.

The use of multivariate analysis combined with rarefaction curves allowed us to assess whether the eDNA-based diversity was robust and to identify potential redundant sampling sites while highlighting the most diverse locations. Our findings show that even though the water samples were collected at the exact same location within a few minutes' interval, the eDNA obtained was different enough to distinguish these water samples. Both alpha diversity analysis and PCA showed that many samples within the same stations are highly divergent, presenting distinct numbers of taxa and do not contain similar variance regarding community composition (Figs 4 and 5A). The PCA revealed that station C01 contributed the most to the observed variation (Fig 2A). This overrepresentation appears to result from species exclusive to this compartment, as supported by the frequency heatmaps (Fig 3). The C01 station alone accounted for 37 of 79 non-fish MOTUs (~46%) and 32 (80%) of all fish species. However, the three samples from C01 presented higher variation than expected from replicates, as observed in the DAPC clustering analysis. Due to potential differences in composition and abundance, two out of the three samples from C01 are distinct enough to be considered a separate ecological group (Fig 5D). For C01, a potential explanation for these results is the higher current speed (0.5 m/s compared to 0.05 m/s in other stations). Because eDNA behaves like fine particulate organic matter, the concentration of eDNA can be influenced by the hydrological profile, being transported according to the velocity of water movement (65). Consequently, the eDNA obtained during each sampling would substantially differ from each other. However, it is also possible that the ichthyofauna habitat preferences are shaping these communities' differences. As previously demonstrated in the literature, the ichthyofauna may exhibit distinct preferences regarding water velocity, temperature and depth, for example, in lotic systems [68, 69].

It is worth noting that in the PCA, all pilot sampling was very close to all stations except C01 (Fig 2A). This indicates that despite the one-year difference between the samples, this time gap did not result in significant differences in the composition of local communities. However, the different sample sizes from the two samplings, plus the short duration of the study, highlights the need for more comprehensive spatial analyses to confirm the time-scale effects on fish communities in the TI. Previous studies have indicated significant changes in community composition after a few weeks [70]. In some cases, temporal dynamics *per se* are not the most determinant factors; rather, environmental factors such as seasonality, salinity, or precipitation can have a greater influence on species distribution and occurrence [66].

Following station C01, stations C10 and C11 demonstrated significant contributions to the clustering analysis of groups from DAPC (Fig 2D). Interestingly, these samples are at the boundaries (C01), and at the center arm (C10 and C11) of the TI reservoir (Fig 1A). Hydrodynamic modeling revealed that water velocity at station C01 (0.05–0.1 m/s) is higher than at other stations, which show velocities of 0.0–0.05 m/s (S2 Fig). This elevated water flow in C01 likely contributes to the increased diversity observed among replicates, as faster flow enhances mixing and eDNA transport. Furthermore, the water renewal rate at C01 is substantially higher, with 50% of the water being renewed every 5–10 days. In contrast, stations C10 and

C11, representing the largest tributaries, experience much lower renewal rates. This should lead to the establishment of distinct communities, minimally influenced by eDNA from the reservoir's main body. To assess the impact of sampling design with and without these stations, we constructed rarefaction curves to determine the number of samples to achieve our observed number of MOTUs. When any of the C01, C11, or C10 samples were excluded, the plateau area of the curve did not change significantly, remaining around 10 to 12 samples (Fig 6). However, when using only these three stations, the rarefaction curve almost reached the plateau shape in a smaller sample size (~8 samples) (S5 Fig). These data show that our sampling effort with 42 samples and 14 stations was excessive, resulting in a considerable amount of redundant information. For this reason, we suggest that for the TI reservoir, a sampling size with only three strategically positioned stations with more samples within each one would be capable of capturing the same number of MOTUs as 42 samples within 14 stations. The location with higher diversity was the one with higher current speed (C01), supporting that the observed diversity does not follow a geographic pattern per se and, instead, it can be a result from either abiotic or biotic factor iterations, still unknown. Either way, the observed diversity could be expected at random sampling with at least 12 sites across the entire reservoir.

When considering each replicate as a sample, the diversity is high and multiple species could be detected. However, the species distribution is very similar among them. Therefore, increasing the sampling effort by adding replicates in fewer locations could increase the number of taxa. It is not the first time that increasing the number of replicates was suggested for river systems. For instance, Shaw *et al.* [19] already emphasized the same discussion when evaluating a river area in South Australia, where the authors recommend more replicates to increase the chance of detecting DNA fragments from taxa that are patchily distributed, with low abundance, low biomass or rare occurrence.

Conclusions

We constructed a large-scale biodiversity inventory to evaluate the diversity of a reservoir's communities, focusing specifically on metazoans and ichthyofauna. Through the use of various descriptive and statistical metrics, we assessed the robustness of our biodiversity inventory and determined the optimal experimental design. We found that conducting consistent sampling at fewer sites with more replicates was more effective in capturing biodiversity. Metabarcoding proved to be a valuable tool, corroborating previous reports based on traditional monitoring surveys and identifying several previously unreported species.

Supporting information

S1 Fig. Metrics and spatial information regarding the Três Irmãos (TI) reservoir for hydrodynamic modeling. A) Bathymetry distribution across the reservoir. Darker shades show deeper sites. B) Biquadratic quadrangular finite elements for spatial discretization of the model.

(PNG)

S2 Fig. Hydrodynamic modeling for the entire Três Irmãos (TI) reservoir. Maximum current speed as measured during the A) rainy and B) dry seasons.

(PNG)

S3 Fig. Computed water renewal rate for the Três Irmãos (TI) reservoir. The water renewal is shown moving from Nova Avanhandava to Três Irmãos reservoir during A) rainy and B) dry scenarios.

(PNG)

S4 Fig. Probability of identifying fish by sampling for eDNA at different depths. Generalized Linear Model (GLM) showing the probability (y-axis) of sampling fish taxa in different depths (1 meter, 13 meters and 25 meters from the surface) after three samples (x-axis). Each line represents one depth according to the colors in the legend.
(PNG)

S5 Fig. Statistics from the Discriminant Analysis of Principal Components (DAPC). A) Sum of contribution of each sampled station for the DAPC clustering assignment. Given the differences in sampling size, the pilot sampling is not included. Station abbreviations as in [S1 Table](#). Each bar corresponds to one station, and the value shows the sum of contribution for the six first Principal Components (see [Methods](#)). The stations with the highest values are C01, C11 and C10, respectively. B) Rarefaction curve constructed using only samples from C01, C11 and C10, where the curve reaches the plateau curve around eight samples.
(PNG)

S1 Table. Sampling locations' information. Description, geographic coordinates, sampling date, water volume, number of replicates, depth and abbreviations.
(XLSX)

S2 Table. Pair of primers used to amplify the partial sequences of COI, 12SrRNA, and 16SrRNA genes. Sequences and references are shown for each marker.
(XLSX)

S3 Table. Summary of reads, Amplicon Sequence Variants (ASVs) and Molecular Operational Taxonomic Units (MOTUs) statistics. Based on eDNA analyses of both pilot and station sampling, values are shown after and before trimming steps.
(XLSX)

S4 Table. List of Amplicon Sequence Variants (ASVs) amplified using eDNA metabarcoding. For both pilot and sampling stations included, the molecular marker, frequency, number of samples, length and taxonomic classification information are displayed for each ASVs.
(XLSX)

S5 Table. Biodiversity inventory for non-fish taxa in the Três Irmãos (TI) reservoir. The inventory was built using previous literature, eDNA and GBIF information. For each taxon, the source of the record is indicated, as well as the reference, if needed.
(XLSX)

S6 Table. Biodiversity inventory for ichthyofauna in the Três Irmãos (TI) reservoir. The inventory was built using previous literature, eDNA and Global Biodiversity Information Facility (GBIF) information. For each taxon, the source of the record is indicated, as well as the reference, if needed.
(XLSX)

Acknowledgments

We thank Angélica Beccato, from Environment and Land Manager at Tijoá Energia, for her support and collaboration in developing this work.

Author Contributions

Conceptualization: Danielle do Amaral, Mauro Rebelo.

Data curation: Thainá Cortez, André Torres.

Formal analysis: Thainá Cortez, André Torres, Marcelo Cabral, Gabriel Zielinsky.

Funding acquisition: Danielle do Amaral, Mauro Rebelo.

Investigation: Thainá Cortez, André Torres, Murilo Guimarães, Henrique Pinheiro, Camila Pereira, Giovanni de Castro, Luana Guerreiro, Juliana Americo, Danielle do Amaral, Mauro Rebelo.

Methodology: Thainá Cortez, André Torres, Murilo Guimarães, Henrique Pinheiro, Marcelo Cabral, Gabriel Zielinsky, Juliana Americo, Danielle do Amaral.

Project administration: Juliana Americo, Danielle do Amaral, Mauro Rebelo.

Resources: Danielle do Amaral, Mauro Rebelo.

Supervision: Juliana Americo, Danielle do Amaral, Mauro Rebelo.

Validation: Henrique Pinheiro, Camila Pereira, Giovanni de Castro, Luana Guerreiro, Mauro Rebelo.

Visualization: Thainá Cortez, Murilo Guimarães, Henrique Pinheiro, Camila Pereira, Giovanni de Castro, Luana Guerreiro, Danielle do Amaral.

Writing – original draft: Thainá Cortez.

Writing – review & editing: Thainá Cortez, André Torres, Murilo Guimarães, Henrique Pinheiro, Marcelo Cabral, Gabriel Zielinsky, Camila Pereira, Giovanni de Castro, Luana Guerreiro, Juliana Americo, Danielle do Amaral, Mauro Rebelo.

References

1. Gu W, Swihart RK (2004) Absent or undetected? Effects of non-detection of species occurrence on wildlife—habitat models. *Biol Conserv* 116:195–203
2. Mächler E, Deiner K, Steinmann P, Altermatt F (2014) Utility of environmental DNA for monitoring rare and indicator macroinvertebrate species. *Freshwater Science* 33:1174–1183
3. Laramie MB, Pilliod DS, Goldberg CS (2015) Characterizing the distribution of an endangered salmonid using environmental DNA analysis. *Biol Conserv* 183:29–37
4. Handley LL (2015) How will the ‘molecular revolution’ contribute to biological recording? *Biological Journal of the Linnean Society* 115:750–766
5. de Faria LC, Domenico M Di, Andrade SCS, dos Santos MC, Fonseca G, Zanol J, Amaral ACZ (2018) The use of metabarcoding for meiofauna ecological patterns assessment. *Mar Environ Res* 140:160–168 <https://doi.org/10.1016/j.marenvres.2018.06.013> PMID: 29933903
6. Schwentner M, Zahiri R, Yamamoto S, Husemann M, Kullmann B, Thiel R (2021) eDNA as a tool for non-invasive monitoring of the fauna of a turbid, well-mixed system, the Elbe estuary in Germany. *PLoS One* 16:e0250452 <https://doi.org/10.1371/journal.pone.0250452> PMID: 33861810
7. Manandhar P, Manandhar S, Sherchan AM, et al (2023) Environmental DNA (eDNA) based fish biodiversity assessment of two Himalayan rivers of Nepal reveals diversity differences and highlights new species distribution records. *PLOS Water* 2:e0000099
8. Erős T, Funk A, Pont D, Hein T, Meulenbroek P, Preiszner B, et al (2024) eDNA metabarcoding reveals the role of habitat specialization and spatial and environmental variability in shaping diversity patterns of fish metacommunities. *PLoS One* 19:e0296310 <https://doi.org/10.1371/journal.pone.0296310> PMID: 38165893
9. Harrison JB, Sunday JM, Rogers SM (2019) Predicting the fate of eDNA in the environment and implications for studying biodiversity. *Proceedings of the Royal Society B: Biological Sciences* 286:20191409 <https://doi.org/10.1098/rspb.2019.1409> PMID: 31744434
10. Barnes MA, Turner CR (2016) The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics* 17:1–17

11. Fukaya K, Murakami H, Yoon S, et al (2021) Estimating fish population abundance by integrating quantitative data on environmental DNA and hydrodynamic modeling. *Mol Ecol* 30:3057–3067
12. Sinha S, Liu X, Garcia MH (2012) Three-dimensional hydrodynamic modeling of the Chicago River, Illinois. *Environmental Fluid Mechanics* 12:471–494
13. Bush A, Compson ZG, Monk WA, Porter TM, Steeves R, Emilson E, et al (2019) Studying Ecosystems With DNA Metabarcoding: Lessons From Biomonitoring of Aquatic Macroinvertebrates. *Front Ecol Evol*.
14. Shu L, Ludwig A, Peng Z (2020) Standards for Methods Utilizing Environmental DNA for Detection of Fish Species. *Genes (Basel)* 11:296 <https://doi.org/10.3390/genes11030296> PMID: 32168762
15. Cilleros K, Valentini A, Allard L, Dejean T, Etienne R, Grenouillet G, et al (2019) Unlocking biodiversity and conservation studies in high-diversity environments using environmental DNA (eDNA): A test with Guianese freshwater fishes. *Mol Ecol Resour* 19:27–46 <https://doi.org/10.1111/1755-0998.12900> PMID: 29768738
16. Holman LE, Chng Y, Rius M (2022) How does eDNA decay affect metabarcoding experiments? *Environmental DNA* 4:108–116
17. Biber E (2013) The Challenge of Collecting and Using Environmental Monitoring Data. *Ecology and Society* 18:art68
18. Haase P, Pauls SU, Schindehütte K, Sundermann A (2010) First audit of macroinvertebrate samples from an EU Water Framework Directive monitoring program: human error greatly lowers precision of assessment results. *J North Am Benthol Soc* 29:1279–1291
19. Shaw JLA, Weyrich L, Cooper A (2017) Using environmental (e)DNA sequencing for aquatic biodiversity surveys: a beginner's guide. *Mar Freshw Res* 68:20
20. Miya M (2022) Environmental DNA Metabarcoding: A Novel Method for Biodiversity Monitoring of Marine Fish Communities. *Ann Rev Mar Sci* 14:161–185 <https://doi.org/10.1146/annurev-marine-041421-082251> PMID: 34351788
21. Valdivia-Carrillo T, Rocha-Olivares A, Reyes-Bonilla H, Domínguez-Contreras JF, Munguia-Vega A (2021) Integrating eDNA metabarcoding and simultaneous underwater visual surveys to describe complex fish communities in a marine biodiversity hotspot. *Mol Ecol Resour* 21:1558–1574 <https://doi.org/10.1111/1755-0998.13375> PMID: 33683812
22. Pont D, Meulenbroek P, Bammer V, et al (2023) Quantitative monitoring of diverse fish communities on a large scale combining eDNA metabarcoding and qPCR. *Mol Ecol Resour* 23:396–409 <https://doi.org/10.1111/1755-0998.13715> PMID: 36151931
23. Boivin-Delisle D, Laporte M, Burton F, Dion R, Normandeau E, Bernatchez L (2021) Using environmental DNA for biomonitoring of freshwater fish communities: Comparison with established gillnet surveys in a boreal hydroelectric impoundment. *Environmental DNA* 3:105–120
24. Evans NT, Li Y, Renshaw MA, Olds BP, Deiner K, Turner CR, et al (2017) Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. *Canadian Journal of Fisheries and Aquatic Sciences* 74:1362–1374
25. Lucchetta BC, Watanabe FS, do Carmo NMRB (2023) A spatiotemporal classification approach to evaluate the impacts of land use and land cover changes before and after the Três Irmãos reservoir formation in the Tietê River, Brazil. *Model Earth Syst Environ* 9:4425–4435
26. de Miranda RB, Mauad FF (2015) Influence of Sedimentation on Hydroelectric Power Generation: Case Study of a Brazilian Reservoir. *Journal of Energy Engineering*.
27. Lamas BV, Rosman PCC, de Almeida Cruz DO (2018) Modelo Hidrodinâmico Computacional do Reservatório de Sobradinho.
28. Minamoto T, Miya M, Sado T, et al (2021) An illustrated manual for environmental DNA research: Water sampling guidelines and experimental protocols. *Environmental DNA* 3:8–13
29. Bolyen E, Rideout JR, Dillon MR, et al (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 37:852–857 <https://doi.org/10.1038/s41587-019-0209-9> PMID: 31341288
30. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581–583 <https://doi.org/10.1038/nmeth.3869> PMID: 27214047
31. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al (2018) Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 6:90 <https://doi.org/10.1186/s40168-018-0470-z> PMID: 29773078
32. Lim SJ, Thompson LR (2021) Mitohelper: A mitochondrial reference sequence analysis tool for fish eDNA studies. *Environmental DNA* 3:706–715

33. Porter TM, Hajibabaei M (2018) Automated high throughput animal CO1 metabarcode classification. *Sci Rep* 8:4226 <https://doi.org/10.1038/s41598-018-22505-4> PMID: 29523803
34. Machida RJ, Leray M, Ho S-L, Knowlton N (2017) Metazoan mitochondrial gene sequence reference datasets for taxonomic assignment of environmental samples. *Sci Data* 4:170027 <https://doi.org/10.1038/sdata.2017.27> PMID: 28291235
35. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410 [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: 2231712
36. Sayers EW, Beck J, Bolton EE, et al (2021) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 49:D10–D17 <https://doi.org/10.1093/nar/gkaa892> PMID: 33095870
37. Oksanen Jari, et al. (2013) Package 'vegan'. Community ecology package, version 2. 9: 1–295.
38. Smith EP, van Belle G (1984) Nonparametric Estimation of Species Richness. *Biometrics* 40:119
39. Wickham H, Chang W Package 'ggplot2'. R package version 1.0.0.
40. Kassambara A, Mundt F (2016) Package 'factoextra'.
41. R Core Team (2021) R: A language and environment for statistical computing. Foundation for Statistical Computing, Vienna, Austria.
42. Jombart T (2008) *ade4*: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405
43. Dixon P (2003) VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* 14:927–930
44. de Almeida FVR, N O Jr., Branco LHZ (2011) Flora de comunidades de macroalgas lólicas de fragmentos florestais remanescentes da região noroeste do Estado de São Paulo, Brasil. *Hoehnea* 38:553–568
45. Magalhães C, Bueno SLS, Bond-Buckup G, Valenti WC, da Silva HLM, Kiyohara F, et al (2005) Exotic species of freshwater decapod crustaceans in the state of São Paulo, Brazil: records and possible causes of their introduction. *Biodivers Conserv* 14:1929–1945
46. Stefani PM (2006) Ecologia trófica de espécies alóctones (*Cichla cf. ocellaris* e *Plagioscion squamosissimus*) e nativa (*Geophagus brasiliensis*) nos reservatórios do rio Tietê.
47. Ferraz KMPMDB, Ferraz SFDB, Moreira JR, Couto HTZ, Verdade LM (2007) *Capybara (Hydrochoerus hydrochaeris)* distribution in agroecosystems: a cross-scale habitat analysis. *J Biogeogr* 34:223–230
48. Pessotto MA, Nogueira MG (2018) More than two decades after the introduction of *Limnoperna fortunei* (Dunker 1857) in La Plata Basin. *Brazilian Journal of Biology* 78:773–784 <https://doi.org/10.1590/1519-6984.180789> PMID: 29340533
49. Goldberg CS, Turner CR, Deiner K, et al (2016) Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods Ecol Evol* 7:1299–1307
50. Carvalho C. S., De Oliveira M. E., Rodriguez-Castro K. G., Saranholi B. H., & Galetti P. M. Jr (2022). Efficiency of eDNA and iDNA in assessing vertebrate diversity and its abundance. *Molecular Ecology Resources*, 22(4), 1262–1273. <https://doi.org/10.1111/1755-0998.13543> PMID: 34724330
51. Campanha PMGDC, Matsumoto AA, Brazão ML, Basilio LM, Maruyama LS (2019) Length and weight relationships and biological aspects for 34 fish species from Três Irmãos reservoir, lower Tietê river basin, SP - Brazil. *Boletim do Instituto de Pesca*.
52. Kullander SO, de Lucena CAS (2006) A review of the species of *Crenicichla* (Teleostei: Cichlidae) from the Atlantic coastal rivers of southeastern Brazil from Bahia to Rio Grande do Sul States, with descriptions of three new species. *Neotropical Ichthyology* 4:127–146
53. Maruyama A., Sugatani K., Watanabe K., Yamanaka H., & Imamura A. (2018). Environmental DNA analysis as a non-invasive quantitative tool for reproductive migration of a threatened endemic fish in rivers. *Ecology and Evolution*, 8(23), 11964–11974. <https://doi.org/10.1002/ece3.4653> PMID: 30598791
54. Troth C. R., Sweet M. J., Nightingale J., & Burian A. (2021). Seasonality, DNA degradation and spatial heterogeneity as drivers of eDNA detection dynamics. *Science of the Total Environment*, 768, 144466. <https://doi.org/10.1016/j.scitotenv.2020.144466> PMID: 33736342
55. Stoeckle B. C., Kuehn R., & Geist J. (2016). Environmental DNA as a monitoring tool for the endangered freshwater pearl mussel (*Margaritifera margaritifera* L.): a substitute for classical monitoring approaches?. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 26(6), 1120–1129.
56. Hänfling B, Handley LL, Read DS, Hahn C, Li J, Nichols P, et al (2016) Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Mol Ecol* 25:3101–3119 <https://doi.org/10.1111/mec.13660> PMID: 27095076
57. Zhang S, Lu Q, Wang Y, Wang X, Zhao J, Yao M (2020) Assessment of fish communities using environmental DNA: Effect of spatial sampling design in lentic systems of different sizes. *Mol Ecol Resour* 20:242–255 <https://doi.org/10.1111/1755-0998.13105> PMID: 31625686

58. Turner CR, Uy KL, Everhart RC (2015) Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biol Conserv* 183:93–102
59. Lacoursière-Roussel A, Rosabal M, Bernatchez L (2016) Estimating fish abundance and biomass from eDNA concentrations: variability among capture methods and environmental conditions. *Mol Ecol Resour* 16:1401–1414 <https://doi.org/10.1111/1755-0998.12522> PMID: 26946353
60. Ficetola G. F., Pansu J., Bonin A., Coissac E., Giguët-Covex C., De Barba M., . . . et al. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular ecology resources*, 15(3), 543–556. <https://doi.org/10.1111/1755-0998.12338> PMID: 25327646
61. Maruyama A, Nakamura K, Yamanaka H, Kondoh M, Minamoto T (2014) The Release Rate of Environmental DNA from Juvenile and Adult Fish. *PLoS One* 9:e114639 <https://doi.org/10.1371/journal.pone.0114639> PMID: 25479160
62. Carraro L, Stauffer JB, Altermatt F (2021) How to design optimal eDNA sampling strategies for biomonitoring in river networks. *Environmental DNA* 3:157–172
63. Rourke ML, Fowler AM, Hughes JM, Broadhurst MK, DiBattista JD, Fielder S, et al (2022) Environmental DNA (eDNA) as a tool for assessing fish biomass: A review of approaches and future considerations for resource surveys. *Environmental DNA* 4:9–33
64. 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* 97:444–452 <https://doi.org/10.1111/jfb.14383> PMID: 32412670
65. Pont D, Rocle M, Valentini A, Civade R, Jean P, Maire A, et al (2018) Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. *Sci Rep* 8:10361 <https://doi.org/10.1038/s41598-018-28424-8> PMID: 29991759
66. Sales NG, Wangenstein OS, Carvalho DC, Deiner K, Præbel K, Coscia I, et al (2021) Space-time dynamics in monitoring neotropical fish communities using eDNA metabarcoding. *Science of The Total Environment* 754:142096 <https://doi.org/10.1016/j.scitotenv.2020.142096> PMID: 32898783
67. Hayami K, Sakata MK, Inagawa T, et al (2020) Effects of sampling seasons and locations on fish environmental DNA metabarcoding in dam reservoirs. *Ecol Evol* 10:5354–5367 <https://doi.org/10.1002/ece3.6279> PMID: 32607158
68. da Costa ID, de Carvalho Freitas CE (2013) Trophic ecology of the ichthyofauna of a stretch of the Urucu River (Coari, Amazonas, Brazil). *Acta Limnologica Brasiliensia* 25:54–67
69. Calazans AM, Martinez PA, Jacobina UP (2021) Lentic and lotic environments affect morphological diversity in characiformes fishes in the Neotropical São Francisco River Basin, Brazil. *Environ Biol Fishes* 104:977–987
70. Rivera SF, Vasselon V, Bouchez A, Rimet F (2023) eDNA metabarcoding from aquatic biofilms allows studying spatial and temporal fluctuations of fish communities from Lake Geneva. *Environmental DNA* 5:570–581