bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **Monitoring fish communities through DNA metabarcoding in the fish pass system of the**
2 **second largest hydropower plant in the world**

3 Giorgi Dal Pont[1,2£], Camila Duarte Ritter[1,2,3*£], Andre Olivotto Agostinis[1], Paula Valeska
4 Stika[1,2], Aline Horodesky[2], Nathieli Cozer[1], Eduardo Balsanelli[4], Otto Samuel Mader Netto[2],
5 Caroline Henn[5], Antonio Ostrensky[1,2], Marcio Roberto Pie[1,2]

6 [1] Grupo Integrado de Aquicultura e Estudos Ambientais, Departamento de Zootecnia,
7 Universidade Federal do Paraná, Rua dos Funcionários, 1540, Juvevê, 80035-050 Curitiba, PR,
8 Brazil.

9 [2] ATGC Genética Ambiental LTDA. Rua dos Funcionários 1540, Juvevê, Curitiba, PR Brazil,
10 80035-050

11 [3] Eukaryotic Microbiology, Faculty of Biology, University of Duisburg-Essen,
12 Universitätsstrasse 5, D-45141 Essen, Germany.

13 [4] Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Paraná, Rua dos
14 Funcionários, 1540, Juvevê, 80035-050 Curitiba, PR, Brazil.

15 [5] Itaipu Binacional. Divisão de Reservatório - MARR.CD, Avenida Tancredo Neves, 6731, Foz
16 do Iguaçu, Paraná, CEP 85866-900, Brazil

17

18 *Corresponding author: Camila D. Ritter, kmicaduarte@gmail.com. Phone: +55 48991434597.
19 Postal address: Eukaryotic Microbiology, Faculty of Biology, University of Duisburg-Essen,
20 Universitätsstrasse 5, S05 R04 H83, D-45141 Essen, Germany.

21 £ Joint first authorship

22

23

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

24 **Abstract**: The Itaipu Hydroelectric Power Plant is the second largest in the world in power
25 generation. The artificial barrier created by its dam imposes an obstacle for fish migration. Thus,
26 in 2002, a fish pass system, named Piracema Channel, was built to allow fish to access areas
27 upstream of the reservoir. We tested the potential of metabarcoding to monitor the impact of both
28 the dam and associated fish pass system in the Paraná River fish communities and to compare it
29 with traditional monitoring methods. Using a fragment of the 12S gene, we characterized
30 richness and community composition based on amplicon sequence variants, operational
31 taxonomic units, and zero-radius OTUs. We combined GenBank and in-house data for
32 taxonomic assignment. We found that different bioinformatics approaches showed similar
33 results. Also, we found a decrease in fish diversity from 2019 to 2020 probably due to the recent
34 extreme drought experienced in southeastern Brazil. The highest alpha diversity was recorded in
35 the mouth of the fish pass system, located in a protected valley with the highest environmental
36 heterogeneity. Despite the clear indication that the reference databases need to be continuously
37 improved, our results demonstrate the analytical efficiency of the metabarcoding to monitor fish
38 species.

39 **Key-words**: Environmental monitoring, Dam, Fish passage, Neotropical ichthyofauna.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

40 **Background:**

41 The Itaipu Hydroelectric Power Plant, built at the border between Brazil and Paraguay, is the
42 second largest in the world in power generation[1], second only to the Three Gorges Power Plant
43 in China. With the formation and filling of its reservoir, in 1982[2], the natural barrier to the
44 migration of fishes of the middle section of the Paraná River (Sete Quedas falls) was replaced by
45 the artificial barrier of the Itaipu dam, located 170 km downstream. This artificial barrier (196 m
46 high) caused impacts on the adjacent fish assemblages, such as the reduction in reproductive
47 activity in the first kilometers downstream of the dam[3]. To allow for fish migration and mitigate
48 the environmental impact of the dam, a fish passage system known as the Piracema Channel was
49 created in 2002, linking the Paraná River to Itaipu's Reservoir[4]. However, the real contribution to
50 the reproductive success of the long-distance migratory species is still under investigation, and
51 this channel also allowed for the dispersal of species originally restricted to the lower Paraná
52 River upstream and species originally restricted to the upper Paraná River downstream[5]. These
53 potential impacts are continuous and can interact with natural disturbance, such as several
54 droughts as which happened in 2020. In this context, monitoring the impact of both dam and fish
55 pass system in the Paraná River fish communities is essential.

56 Fish diversity estimates in Brazilian freshwater are still imprecise due to the scarcity of complete
57 inventories[5–7]. Many species are described every year and several groups are in need of
58 taxonomic revision[5,8]. Furthermore, traditional assessment methods for fish diversity surveys are
59 costly and time consuming, given that they depend on capture (e.g. netting, trawling) or
60 observation[9,10] and expertise for taxonomic identification[11]. In this sense, designing methods for
61 cost-effective monitoring fish diversity and community composition is an urgent task. Most
62 sampling efforts in Brazil have historically been primarily funded by the hydroelectric sector,
63 focusing particularly on rivers where power dams were built[12]. The areas of the dam construction
64 have some of the most comprehensive knowledge of fish assemblage composition in comparison
65 with other Brazilian regions and therefore offer an ideal opportunity to compare taxonomic
66 surveys with molecular approaches.

67 A promising alternative to traditional taxonomic surveys and biomonitoring methods is the use
68 of environmental DNA (eDNA), combined with a high-throughput sequencing approach, as in
69 the case of metabarcoding[13]. This technique has the advantage of obtaining DNA from

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

70 environmental samples, such as water, without first isolating the target organism and therefore
71 can sample entire communities[14]. Metabarcoding is a powerful tool for biodiversity assessment
72 that has been widely used for several purposes and different taxonomic groups[15–17], and is
73 considered a transformative technology for the entire field[18]. However, some limitations, such as
74 the relative scarcity of DNA sequences for several species, which is even more problematic in
75 highly diverse regions such as the Neotropics[19], may create constraints that hamper its full
76 application[20,21].

77 The absence of a comprehensive DNA reference database may lead to a misidentification of
78 several species. Therefore, putting together a curated and complete DNA reference database is
79 fundamental for species identification through a metabarcoding approach[7]. But, even with an
80 incomplete DNA reference database, the use of molecular units, such taxonomic units clustered
81 by similarity (operational taxonomic units - OTUs[22]) or unique sequences (e.g. amplicon
82 sequence variants - ASVs[23], or zero-radius OTUs - ZOTUs[24]) allows for diversity monitoring in
83 the context of biodiversity assessment in megadiverse biomes. Such estimates without
84 comprehensive species identification limit ecological conclusions but allowed for monitoring of
85 natural and artificial impacts[16,25]. The metabarcoding approach has been successfully used for
86 molecular identification of several vertebrate groups in temperate regions[26,27], monitoring of
87 endangered species such as freshwater fish in Australia[28] and turtles in the United States[29], and to
88 describe biodiversity even with limited taxonomic identification[30].

89 In this context, our goal here is to describe an effective survey protocol for detecting fish
90 assemblages through eDNA metabarcoding in an ecologically complex and highly diverse
91 freshwater system, the Piracema channel, that connects the Paraná River with Itaipu Reservoir.
92 For this, we used an in-house molecular database of fishes occurring in the channel
93 complemented with GenBank sequences. We also describe fish alpha diversity and community
94 structure in the Piracema channel system. Additionally, we compare our metabarcoding results
95 with the traditional sampling campaigns made between 2017- 2021.

96

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

97 **Material and Methods:**

98 *Study area* – Our study was conducted at the Piracema channel (Fig. 1), a fish pass system
99 connecting the Paraná River with the Upper Paraná River floodplain (main reservoir). For the
100 traditional taxonomic survey, we sampled three points at mouth of channel at Paraná River (Fig.
101 1, blue square), the main lake at the Piracema channel (Fig. 1, red circle), and the reservoir near
102 the water intake to the Channel (Fig. 1, green triangle) between 2017 to 2021. Fish were
103 collected monthly during the fish reproductive period (October to March), and once during the
104 winter (July or August), employing active and passive methods (Table 1).

105 **Table 1.** Fish sampling methods at the Piracema Channel

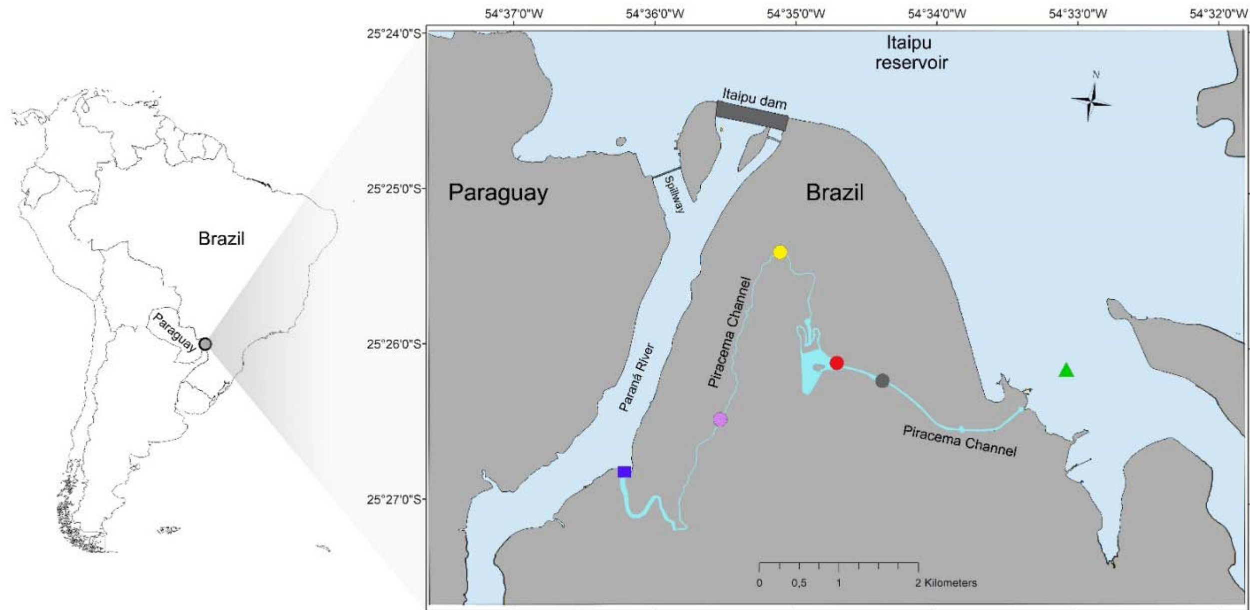
Fish sampling method	Quantitative and qualitative aspects
Gill nets	Mesh sizes: 1 a 10 cm (adjacent knots), each one 10m long and 1 to 1,5m high
Longlines	30 hooks, 10 of each size: /10, /8 and /6, fish pieces as bait
Cast nets	Mesh sizes: 3 e 6 cm (adjacent knots)
Electrofishing	Smith-Root, backpack electrofisher, 600V, 30Hz DC

106

107

108 For each point, gill nets and longlines were set out in the afternoon (16:00 h) and inspected every
109 4 hours during a 24 h cycle; cast nets were operated 3 times each mesh, after every gear
110 inspection. Boarded electrofishing was operated two times in each point, at dawn and at dusk,
111 covering 100m of the environment margin each time. Fish were euthanized by immersion in
112 benzocaine solution, following current legislation[31], and identified accordingly Britski et al.[32],
113 Ota et al.[33] and Neris et al.[34]. Fragments of muscle were collected with a scalpel, placed into 2 ml
114 tubes filled with 99.8% ethanol and stored at 4ºC until processing. Voucher specimens are
115 housed in the Nupelia-UEM fish collection.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



116

117 **Figure 1. Sampling location.** The map shows the sampling location of each collection point. We
118 sampled one point at mouth of channel at Paraná River (blue square), four points along the
119 Piracema Channel (circles; Bela Vista River 1 = purple, Bela Vista 2 = yellow, Brasilia stream =
120 gray, and lake = red), and one point at Itaipu's reservoir (green triangle). Up at figure is possible
121 to visualize the Itaipus' dam that created the reservoir. Inset panel shows the location of Itaipu's
122 dam in relation to South America. Map was created in QGIS v.3.6.2 software [88].

123

124 For metabarcoding, we sampled one site at the mouth of channel at Paraná River (Fig. 1, blue
125 square), four sites along the Piracema Channel (Fig. 1, circles), and one site at Itaipu Reservoir
126 (Fig. 1, green triangle). Each sampling point was collected in sextuplicate. All six sites were
127 sampled in 2019 and three sites (mouth of channel at Paraná River, lake at Piracema channel [red
128 circle], and the reservoir) were sampled again in 2020, totaling 54 samples. All sampling sites
129 were provided with GPS coordinates.

130 *Sampling design for molecular analysis* – We collected water by partially submerging a one litter
131 polypropylene bottle. The objective was to sample water at the air/water interface. After water
132 collection, bottles were closed and cleaned with a 10% sodium hypochlorite solution, following
133 by rinsing with distilled water. We used gloves which were changed in between each new
134 sampling replicate to reduce the risk of cross-sample contamination.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

135 After the collection and cleaning steps, the bottles were stored in polystyrene boxes containing
136 artificial ice to maintain the temperature of the samples at 4 to 10 °C. The samples were filtered,
137 on the same day of collection, using nitrocellulose membranes (0.45 μm pore) with the aid of a
138 vacuum pump. Filters were kept in 100% ethanol under refrigerated conditions until molecular
139 analysis was performed. All dry samples were processed at the ATGC laboratory at the
140 Universidade Federal do Paraná (UFPR).

141 *DNA extraction* – For total DNA extraction, we kept the collected filters at room temperature to
142 allow the residual ethanol to dry completely. After dried we extract the DNA using magnetic
143 beads (microspheres surrounded by magnetite and carboxyl), which bind to DNA (carboxyl bond
144 - DNA) by the process of Solid Phase Reversible Immobilization (SPRI). The DNA extract was
145 stored at –20 °C until the amplification. The extraction and quantification processes were carried
146 out in separate rooms, as suggested by Pie et al.[35]. We checked the DNA concentration using
147 both a spectroscope (Nanodrop, Thermo, USA) and a fluorimeter (Qubit, Invitrogen, USA).

148 *PCR amplification* – We targeted the 12S rRNA gene using the MiFish forward (5'-
149 GTCGGTAAAACTCGTGCCAGC-3') and reverse (5'-
150 CATAGTGGGGTATCTAATCCCAGTTTG-3') primers designed by Miya et al.[36] to yield 163–
151 185 bases long fragments. Amplification was performed in a total volume of 20 μl in GoTaqG2
152 system (Promega, USA), 500 nM of forward and reverse primers, and 20 ng of DNA template.
153 The PCR conditions consisted of an initial denaturation step of 2 min at 95 °C and then 25 cycles
154 of denaturation at 94 °C for 30 s, hybridization at 55 °C for 45 s, and elongation at 72 °C for 30
155 s, followed by a final elongation at 72 °C for 5 min and finishing at 4 °C. To avoid PCR
156 inhibition BSA (0.5 μg/μl) was added to the reaction as suggested by Boeger et al[37]. The quality
157 of amplification was verified on a 1.5% agarose gel in TBE buffer (9 mM TRIS, 9 mM boric
158 acid, 1 mM EDTA), stained with SYBR Safe DNA Gel Stain (ThermoFisher Scientific,
159 Country). All replicates from each sampling point were amplified to increase the chance of
160 detecting rare species. The PCR product was then diluted (20x) and used as a template for the
161 addition of adapters in the second PCR. Indexing was performed for Illumina MiSeq sequencing
162 (Illumina, USA), using the above PCR system with Nextera indexes (Illumina) in a total volume
163 of 10 μl. PCR conditions were an initial step of 95 °C for 3 min, following by 12 cycles of
164 denaturation at 94 °C for 30 s, hybridization at 55 °C for 45 s, and elongation at 72 °C for 30 s,

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

165 followed by a final elongation at 72 °C for 5 min and finishing at 4 °C. We checked the DNA
166 concentration in a Qubit® fluorimeter (Invitrogen, USA), normalized and pooled the PCR
167 products following the Illumina protocol. The samples were sequenced at GoGenetic (Curitiba,
168 Brazil) using Illumina MiSeq Reagent 600V3 (PE 300b). Three DNA extraction negative
169 controls were included for sequencing. We included three negative controls from the DNA
170 extraction to sequencing. The raw sequences are deposited in GenBank under Bioproject
171 PRJNA750895 (biosamples SAMN20500524 – SAMN20500577).

172 *Sequence analyses and taxonomic assessment* – For the amplicon sequence variants (ASVs)
173 approach, we used the Cutadapt package[38] in Python v.3.3[39] to remove primers. We then used the
174 DADA2 package[23] in R v. 4.0.2[40] to quality filter reads, merge sequences, remove chimeras, and
175 to infer ASVs. We excluded reads with ambiguous bases (maxN=0). Based on the quality scores
176 of the forward and reverse sequences, each read was required to have <3 or <5 errors,
177 respectively (maxEE=c (3,5), truncQ=2). Therefore, ASVs were inferred for forward and reverse
178 reads for each sample using the run-specific error rates. To assemble paired-end reads, we
179 considered a minimum of 12 base pairs of overlap and excluded reads with mismatches in the
180 overlapping region. Chimeras were removed using the consensus method of
181 "removeBimeraDenovo" implemented in DADA2.

182 For operational taxonomic units (OTUs) and zero-radios OTU (ZOTUs) analyses, we used the
183 USEARCH/UPARSE v.11.0.667 Illumina paired reads pipeline[41] to primer remove, quality
184 filtering, dereplicate and sort reads by abundance, to infer OTUs and ZOTUs, and to remove
185 singletons. We filtered the sequences to discard chimeras and clustered sequences into OTUs at a
186 minimum similarity of 97% using a 'greedy' algorithm that performs chimera filtering and OTU
187 clustering simultaneously and the UNOISE algorithm to denoised sequence as zero-radios OTUs
188 to create or ZOTUs table[41,42].

189 We build a reference dataset of DNA sequences for the 205 fish taxa that have been historically
190 recorded in the Itaipu system using the following steps. First, we looked for 12S sequences of
191 these species in GenBank by searching for their corresponding names. We were able to find
192 sequences for 126 species in our reference database (Table S1). Additionally, we created an in-
193 house database which included sequences for 42 additional species to the 79 species previously
194 identified as present on Itaipu system but not available on GenBank. Sequences for the in-house

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

195 database were obtained via Sanger sequencing of tissue samples and were uploaded to GenBank
196 (accession numbers MZ778813- MZ778856). We manually blasted all sequences against the
197 NCBI GenBank database to verify misidentification or problematic sequences (e.g. blasted in the
198 different family). In total, our reference database included 168 (82%) sequences from the 205
199 taxa recorded in the Itaipu system. Finally, we blasted the ASVs, OTUs, and ZOTUs sequences
200 with our reference database to verify the taxonomic composition using the "Blastn" function of
201 the program Blast+[43] with an e-value of 0.001. We kept ASVs, OTUs, and ZOTUs that have
202 matched with a fish species at minimum level of 75% similarity (as these sequences are probably
203 fishes species not present in our reference database), and considered identified species just
204 ASVs, OTUs, and ZOTUs that matched in a minimum level of 97% similarity. ASVs, ZOTUs
205 and OTUs present with a proportion > 0.01% of reads in the sum of three negative controls were
206 discarded (13 ASVs, 1 ZOTU, and 5 OTUs).

207 *Statistical analysis* – We conducted all analyses in R using RStudio[44]. We used the tidyverse
208 package v. 1.3.0[45] for data curation and ggplot2 v. 3.3.2[46], ggfortify v. 0.4.11[47], gridExtra v.
209 2.3[48], and ggpubr v. 0.4.0[49] for data visualization (scripts in Appendix 1).

210 For analysis of alpha and beta diversity with metabarcoding data, we made the analysis at ASVs,
211 OTUs, and ZOTUs level. Since the number of observed ASVs, ZOTUs, and OTUs is dependent
212 on the number of reads, we rarefied all samples to the lowest number of reads obtained from any
213 one plot (157 for ASVs, 147 for ZOTUs, and 219 for OTUs; Fig. S1) using the "rarefy" function
214 with the vegan v.2.5.7[50] R package. Because in the ZOTUs table the minimum reads of a plot
215 was nine, we used the second lower value to avoid having to downsize the other samples to such
216 a low number of reads[51]. Because rarefying of counts is considered inappropriate for detection of
217 differentially abundant species[51], even more with so different sampling depth as in our case, we
218 also calculated true effective number of ASVs, ZOTUs, and OTUs of order q=1, which is
219 equivalent to the exponential of the Shannon entropy[52], using the function "AlphaDiversity" of
220 the Entropart v.1.6.7[53] R package. The effective number is more robust against biases arising
221 from uneven sampling depth than the simple counts of ASVs, ZOTUs, and OTUs[51,54].
222 Additionally, for alpha diversity, we also calculated the ASV, OTU, and ZOTU richness (the
223 number of ASV, OTU, and ZOTU per point), Chao1, and Fisher's alpha diversity (i.e., the

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

224 relationship between the number of ASV, OTU, and ZOTU in any given point and the number of
225 reads of each ASV, OTU, and ZOTU) using the phyloseq v.1.34.0[55] R package.

226 For beta diversity, we also used rarefaction (with "rrarefy" function of vegan package) and hill
227 number (with "varianceStabilizingTransformation" function in DESeq2 v.1.28.1[56] R package) to
228 normalize our data. While rarefaction normalizes data by random subsampling without
229 replacement, the hill number transformation normalizes the count data with respect to sample
230 size (number of reads in each sample) and variances, based on fitted dispersion-mean
231 relationships[56]. We then constructed two-dimensional Principal Coordination Analysis (PCoA)
232 ordinations of the abundance (reads) and presence/absence data for both rarefied and hill
233 numbered data. We used the 'cmdscale' function and Bray-Curtis distances in the vegan package
234 to assess community dissimilarity among all samples in the PCoA. We used the "envfit" method
235 in vegan to fit sampling localities and sample year onto the PCoA ordination as a measure of the
236 correlation among the sampling localities with the PCoA axes.

237 For traditional survey data, we calculated the alpha diversity using the observed richness, Chao1,
238 and ACE with the function "estimate", and Shannon index with the function "diversity" both
239 with vegan package. We also constructed two-dimensional Principal Coordination Analysis
240 (PCoA) ordinations of the abundance (reads) and presence/absence data, and used the "envfit" to
241 fit sampling localities and sample year onto the PCoA ordination.

242

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

243 **Results:**

244 For the traditional surveys, 4,430 fishes were collected, for a total 138 species. Most specimens
245 were collected at the mouth of channel at Paraná River, with 2,240 (51%) fishes belonging to
246 105 species. The reservoir showed the lowest number of collected specimens: 1,034 (23%) and a
247 total of 64 species.

248 For metabarcoding data, we obtained a total of 25,292,218 reads. After all cleaning steps, we
249 kept a total of 4,100,729 sequences belonging to 7,219 ASVs. Of these, 2,683,436 (65% of the
250 total) sequences belonging to 211 ASVs were classified into species corresponding to our
251 reference database in the level of 75% similarity, which included other samples from additional
252 projects sequenced together. From our samples we kept 190 ASVs (900,865 reads), of which 121
253 (64%) were classified in 35 species matches at a level > 97% of similarity (Table 2), which is
254 certainly an underestimation of the real number of species, since the other 69 ASVs should
255 belong to species do not present in our database.

256 **Table 2.** Species identified at level of > 97% similarity, the number of ASVs, ZOTUs, OTUs
257 identified per species (possible intra-specific variation) and the number of reads per species after
258 correction. In bold species that was not registered in one of the pipelines, five in total. It is
259 possible to observe that the reads are more similar for the pipelines of unique sequences (ASVs
260 and ZOTUs), while the cluster pipeline (OTUs) keep more reads.

Species	**Number of ASVs**	**Reads of ASVs**	**Number of ZOTUs**	**Reads of ZOTUs**	**Number of OTUs**	**Reads of OTUs**
Acestrorhynchus lacustris MG755503.1	1	9151	1	2201	1	2201
Apareiodon affinis NC_015834.1	5	44738	1	44852	1	44852
Astyanax altiparanae	12	109217	2	108603	5	108551
Characidae3	16	129645	7	128540	8	128576
Characidium sp. LC036706.1	3	6357	2	6389	1	6389
Cichla kelberi	3	2738	1	799	1	799
***Cichla ocellaris* LC069581.1**	**0**	**0**	**1**	**2019**	**1**	**2019**
Crenicichla britiskii	**2**	**14565**	**0**	**0**	**0**	**0**
Crenicichla sp.	4	28044	2	27982	4	28004
***Crenicichla sp.* LC069598.1**	**0**	**0**	**1**	**14350**	**1**	**14350**
Eigenmannia limbata MH263669.1	3	2431	2	97	1	97
Geophagus brasiliensis C_031181.1	4	10888	2	10351	2	10351
Gymnotus carapo AP011979.1	3	10408	3	10057	2	13415
Gymnotus sylvius MN583179.1	1	3517	1	3412	3	45
***Hemigrammus erythrozonus* MT484070.1**	**3**	**1153**	**0**	**0**	**0**	**0**
Hemigrammus marginatus MG755550.1	3	7335	1	7352	1	7352

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hemiodus orthonops	4	2695	1	2478	1	2478
Hemisorubim platyrhynchos JF898664.1	2	1727	1	1687	1	1687
Hoplias intermedius KU523584.1	2	11902	1	11274	1	11274
Hoplias sp1	8	46648	5	46823	3	46869
Hoplias sp2	3	4876	2	4915	2	4915
***Hyphessobrycon amandae* MT484069.1**	**1**	**1010**	**0**	**0**	**0**	**0**
Hypostomus affinis KT239013.1	4	14468	2	15871	2	338
Hypostomus albopunctatus	11	89735	2	93752	7	94997
Hypostomus ancistroides	3	5240	3	5239	2	19725
Hypostomus commersoni	2	10371	2	10395	2	10396
Hypostomus gymnorhynchus JN855752.1	3	4110	1	3196	1	3195
Leporellus vittatus LC104399.1	1	1728	1	1654	1	1654
Leporinus elongatus NC_034281.1	1	1375	1	1319	1	1319
Leporinus lacustris	1	3010	1	2973	1	2973
Leporinus octofasciatus	7	112049	3	113810	12	113954
***Loricaria sp Mato Grosso* KR478070.1**	**1**	**78**	**1**	**71**	**0**	**0**
***Loricaria sp Orinoco* KR478071.1**	**0**	**0**	**1**	**17**	**1**	**17**
Megaleporinus obtusidens NC_034945.1	2	8293	1	8228	1	8228
Myloplus tiete	1	915	1	865	1	865
Oligosarcus sp LC145855.1	6	17868	2	18059	1	18034
Oreochromis niloticus MN255618.1	8	9673	4	9799	5	10110
***Piaractus mesopotamicus* NC_024940.1**	**0**	**0**	**1**	**291**	**3**	**325**
Pimelodella cristata MH286807.1	3	5080	2	4922	3	4924
Pimelodus albicans JF898707.1	2	3586	2	3427	1	3424
***Pimelodus ornatus* JF898680.1**	**0**	**0**	**0**	**0**	**1**	**3**
Pirinampus pirinampu	1	453	1	428	1	428
Plagioscion squamosissimus MT080739.1	2	3431	1	3348	1	3348
Platydoras armatulus NC_025585.1	2	3264	2	3151	2	3151
Poecilia reticulata NC_024238.1	3	7459	1	5957	1	5957
Prochilodus lineatus NC_024939.1	8	79673	1	80216	5	80504
Pseudoplatystoma corruscans NC_026846.1	5	18627	3	16746	5	16736
***Pseudoplatystoma reticulatum* NC_033859.1**	**19**	**8235**	**0**	**0**	**3**	**104040**
***Pterodoras granulosus* AY264087.1**	**1**	**16**	**0**	**0**	**1**	**15**
Rhamdia quelen EU179824.1	2	28623	2	26811	8	29105
Salminus brasiliensis NC_024941.1	1	140	1	138	1	138
Salminus hilarii	1	1153	1	1229	1	1229
***Satanoperca jurupari* LC069577.1**	**0**	**0**	**1**	**82**	**1**	**82**
Schizodon borellii	3	10511	2	10611	5	10628
Trachelyopterus galeatus JX899742.1	1	288	1	269	1	340
Zungaro jahu EU179830.1	2	2368	2	2361	1	2361
Total	190	900865	87	879416	123	986767

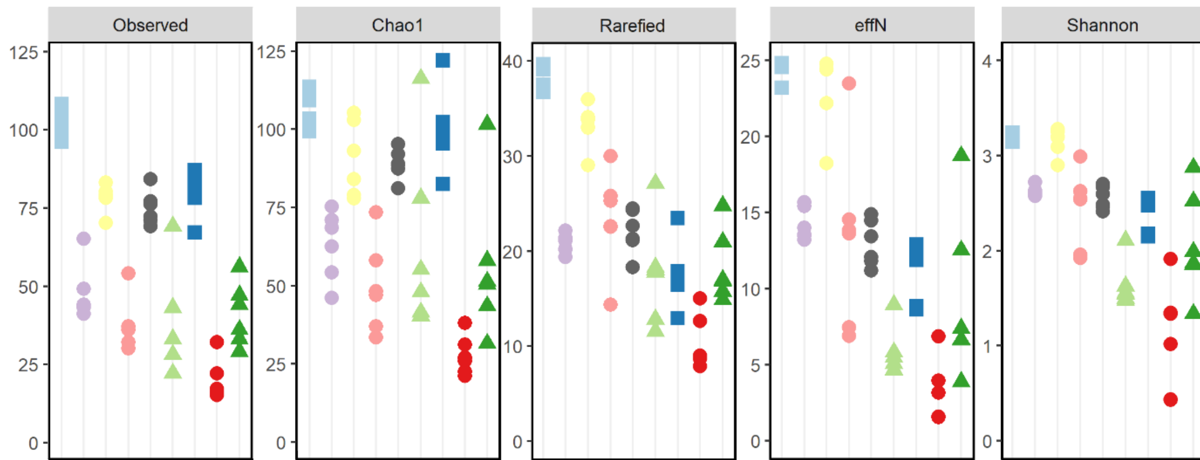
261

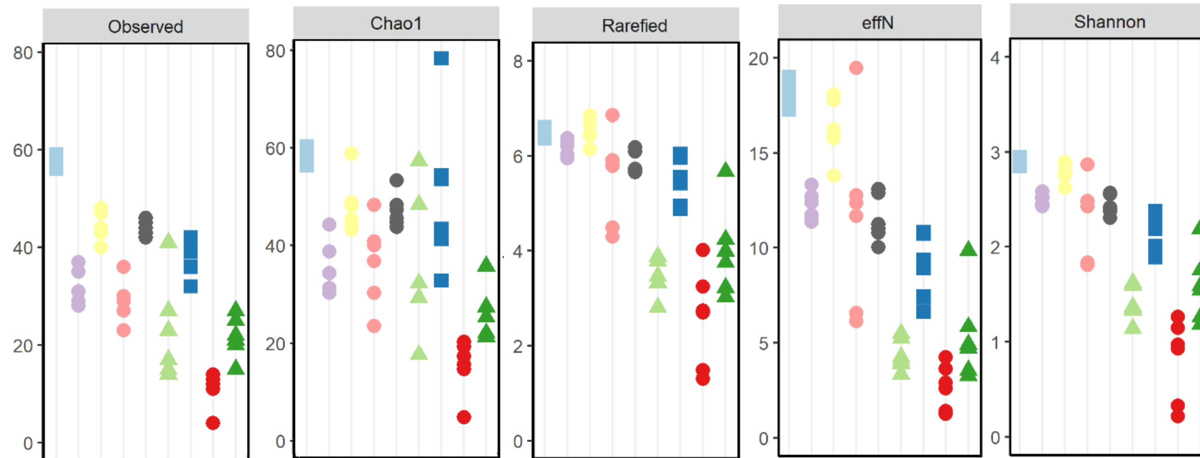
bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

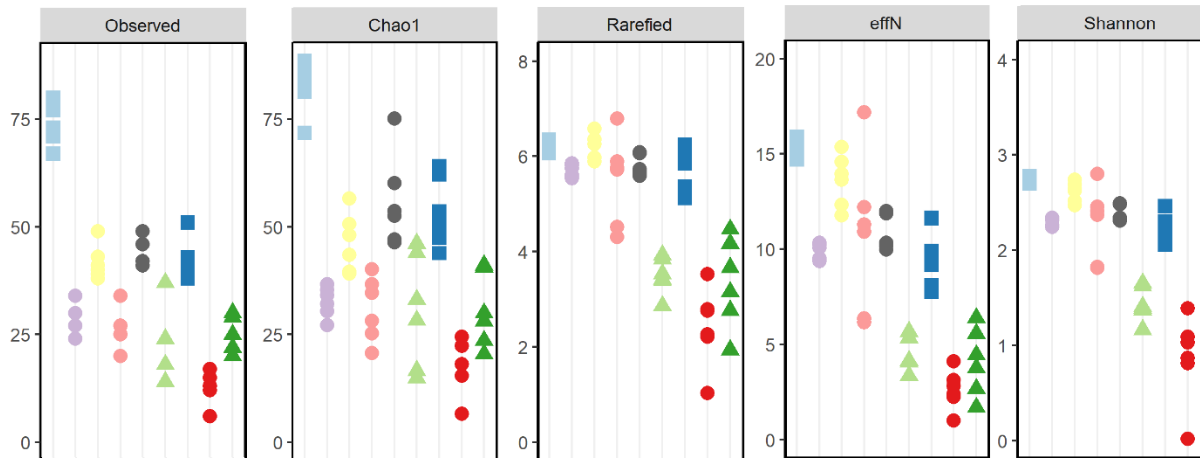
262 For the OTU and ZOTU analyses, after all cleaning steps, we obtained 2,835,679 and 2,819,524
263 reads belonging to 855 OTUs and 215 ZOTUs, respectively (considering samples from other
264 projects). Of these, 2,660,872 (94%) and 2,657,978 (94%) reads belonging to 136 OTUs and 94
265 ZOTUs, respectively, were classified into species corresponding to our reference database in the
266 level of 75% similarity. For our samples, we kept 123 OTUs (986,767 reads) and 87 ZOTUs
267 (879,416 reads). As the OTUs analysis already classified the sequences by 97% similarity, these
268 probably correspond to the number of species present in our samples (89% of species sampled in
269 five years with traditional surveys). Yet, only 37 species belonging to 42 OTUs and 34 species
270 belonging to 46 ZOTUs at > 97% similarity were assigned in both analyses. Eighty-one (66%)
271 OTUs and 41 (47%) ZOTUs were identified as a fish species with a similarity lower than 97%,
272 representing species not present in our reference database.

273 All the alpha diversity measures of ASVs, OTUs, and ZOTUs per sampling point varied, with
274 the point at mouth of the channel at Paraná River, in 2019, showing the highest diversity for all
275 molecular units and the lake at Piracema Channel in 2020 the lowest (Fig. 2). For the traditional
276 surveys, the variation was more random, but the mouth of channel at Paraná River also had the
277 highest diversity (Fig. S2).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.







278

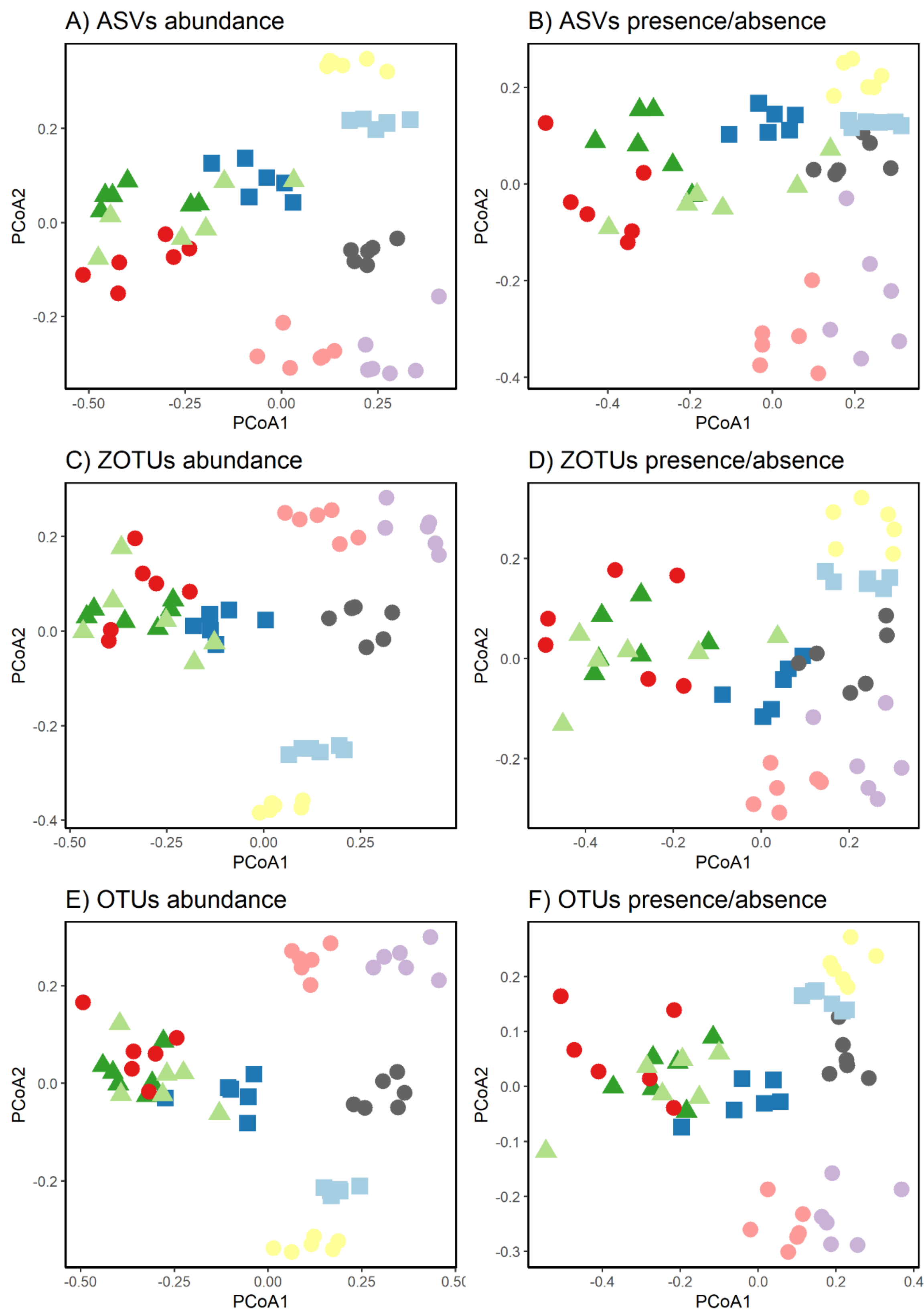
bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

279 **Figure 2.** Alpha diversity estimation for A) ASVs, B) ZOTUs, C) OTUS. Alpha diversity varied
280 by location and by sampling year. Each point is one of the replicates sampled. Colors and
281 symbols represent collection points (mouth of channel at Paraná River = blue square, Itaipu's
282 reservoir = green triangle, and Piracema Channel = circles [Bela Vista River 1 = purple, Bela
283 Vista 2 = yellow, Brasilia stream = gray, and lake = red]), and tone represent year of collection
284 (light = 2019, dark = 2020).

285

286 Fish communities varied among sampled sites. For the abundance based in hill numbers, the first
287 axis of the PCoA separated the samples by year (envfit: R^2 = 0.30 [ZOTUs], 0.34 [ASVs], and
288 0.36 [OTUs], $p < 0.001$), except for the Itaipu's reservoir, with the positive values associated
289 with 2019 and the negative values associated with 2020 (Fig. 3). The second axis separated the
290 samples by locality with some overlap (envfit: R^2 = 0.95 [ZOTUs] - 0.96 [ASVs and OTUs], p <
291 0.001; Fig. 3). For the presence/absence data also based in hill numbers, the overlap was higher
292 but yet the separation by year (envfit: R^2 = 0.30[ZOTUs], 0.32 [OTUs], and 0.41 [ASVs], p <
293 0.001) and locality (envfit: R^2 = 0.91 [ZOTUs] - 0.92 [ASVs and OTUs], $p < 0.001$) was similar
294 to the abundance data (Fig.3). The results of rarefied data were similar with more overlap among
295 sampling points (Appendix 1, Fig. S3). For traditional surveys, the points are clustered by
296 localities (envfit: R^2 = 0.60 for abundance and 0.87 for presence/absence data, $p < 0.001$), but not
297 by year ($p > 0.05$, Fig. S4).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



298

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

299 **Figure 3.** Principal Coordinates Analysis (PCoA) of fishes' communities from Itaipu based in
300 hill numbers for A) ASVs abundance, B) ASVs presence/absence, C) ZOTUs abundance, D)
301 ZOTUs presence/absence, E) OTUs abundance, and F) OTUs presence/absence. The axis 1
302 separated mainly the samples by year, while the axis 2 separated samples mainly by locality.
303 Each point is one of the replicates sampled. Colors and symbols represent collection points
304 (mouth of channel at Paraná River = blue square, Itaipu's reservoir = green triangle, and
305 Piracema Channel = circles [Bela Vista River 1 = purple, Bela Vista 2 = yellow, Brasilia stream
306 = gray, and lake = red]), and tone represent year of collection (light = 2019, dark = 2020).
307

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

308 **Discussion:**

309 Our results support mounting evidence that eDNA analysis provides a cost-effective alternative
310 to characterize fish biodiversity. We also demonstrate that different bioinformatic approaches
311 show similar results in terms of alpha and beta diversity, supporting the use of molecular
312 approaches to monitor biodiversity even with incomplete taxonomic identification. However, a
313 serious caveat for using these molecular methods for biodiversity assessments is the scarcity of
314 comprehensive taxonomic reference databases, especially for the tropical regions of the globe.
315 Here, we also highlight these caveats for the Neotropical fish database, which are taxonomically
316 limited, limiting the identification of several species. With a complete reference database, eDNA
317 could detect mostly fish community and also fish species that are poorly or non-represented by
318 conventional methods, as suggested by our results.

319 We identified 35 species with ASVs, 37 with OTUs, and 34 with ZOTUs approaches at >97%
320 similarity. However, many other ASVs, OTUs, and ZOTUs were identified at <97% similarity,
321 representing species not present in our database. We created our database based on the historical
322 taxonomic survey of Piracema Channel that may prevent identification of species that had not
323 been recorded by conventional fish survey methods. However, the use of a database without
324 curatorship can include spurious species identifications, such as species unlikely to be physically
325 present at sampling sites[10]. That occurs because when the database does not contain the sequence
326 of a certain species, the sequences will match with the closest species in that database, which can
327 occur in a completely different environment (e.g. marine), beyond other factors that also
328 contribute to registering spurious species, such as misannotated sequences[57] or low variability in
329 the target sequenced region[10] that will sign any species with such similar sequence. For instance,
330 our sequence for *Prochilodus lineatus* is identical to other *Prochilodus* species, such as *P. harttii*
331 and *P. costatus*. Furthermore, there are many species undescribed, making it impossible to
332 identify them. A recent compilation to list Paraná state fish species included 42 undescribed
333 species[5], and this number may be underestimated due to the presence of crypt species and
334 sampling biases.

335 Even with the previously mentioned limitations, the use of molecular units such as amplicon
336 sequence variants (ASVs)[58], operational taxonomic units (OTUs)[22], and zero-radius OTUs

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

337 (ZOTUs)[24] allows for assessing of genetic diversity and enables comparison among multiple
338 sites[59], space-time dynamics[16] and evaluate natural and anthropogenic impacts[60]. For instance,
339 vertebrate populations from freshwater ecosystems are declining at alarming rates (83% decline
340 since 1970)[61], and their conservation and management are a priority for global biodiversity[62].
341 The Neotropical region harbors one of the largest freshwater biodiversity, with an estimated
342 9,000 described fish species (around 30% of total freshwater species)[11]. The increasing
343 construction of dams is threatening fish populations over the entire planet [63–65] but specially in
344 Neotropical countries such as Brazil[5,66,67], and effective ways to monitor fish biodiversity to
345 understand its impact is essential.

346 As observed with the use of conventional ichthyofauna monitoring methods[68], the number of
347 species, ASVs, OTUs, ZOTUs, or 12S gene sequence readouts identified in our study showed a
348 variation between the two sampling occasions (2019 and 2020). Such variations in fish
349 assemblages can be related to a series of factors, both biotic (ecological characteristics of the
350 species, for example) and abiotic (variations in water quality, and other environmental factors).
351 In addition, physical characteristics of the environment such as total water volume and
352 hydrological characteristics can also play a key role in the ecology and occurrence of fish
353 species[68]. For instance, the recent extreme drought experienced in southeastern Brazil[69] may
354 have impacted fish assemblages. Our results showed a decrease of alpha diversity in 2020 in both
355 mouth of channel at Paraná River (blue squares) and the lake (red circles; Fig. 2). In addition to
356 the direct effects caused by this type of climatic phenomenon, such as the reduction in the
357 volume of water, indirect effects such as reduced oxygen concentration in the water and food
358 availability can cause severe impacts on fish's communities[68,70,71]. Such effects were more
359 evident at the mouth of channel at Paraná River, where the water level dropped 7 m from 2019 to
360 2020. At the reservoir, alpha diversity did not vary as water level fluctuation was less evident as
361 a result of a stable environment due to the large size of this water body (green triangles; Fig. 2).
362 However, the traditional survey in Piracema Channel was unable to significantly detect the
363 diversity variation throughout the period of the study (Fig. S2), highlighting the high sensibility
364 of eDNA metabarcoding for monitoring.

365 Among sampling points, the highest alpha diversity was recorded in those collected in mouth of
366 channel at Paraná River, while the lowest alpha diversity was registered in the lake (Fig. 2).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

367 Habitat heterogeneity is recognized as a main factor supporting functional and phylogenetic
368 diversity, which is often reflected in the taxonomic richness of the fish communities[72]. Mouth of
369 channel at Paraná River, the entrance of the Piracema Channel, is in a protected valley, where the
370 riparian vegetation is conserved, allowing the colonization by a diversified flora and fauna.
371 Besides this, the confluence with the Paraná River produces adjacent lotic and lentic
372 microhabitats, supporting a higher alpha diversity when compared to the main lake or the water
373 intake of the Channel, which are lentic and uniform environments. Such pattern of fish diversity /
374 limnologic gradients meets the patterns previously assessed for the reservoir tributaries[73].

375 The beta diversity showed that in 2020, with the event of the extreme drought, a homogenization
376 of fish assemblage happened (Fig. 3 & Fig. 4). Both samples from the mouth of channel at
377 Paraná River (blue squares) and the lake (red circles) cluster together with the reservoir in both
378 years. The Itaipu's Reservoir was filled in 1982 and the Piracema Channel (a fish pass),
379 connecting the region just downstream from Itaipu Dam to the Itaipu Reservoir, was opened 38
380 years later. Both events allowed the dispersion of species (including non-native species) in both
381 directions promoting the homogenization of communities from upper and lower Paraná
382 River[5,74,75]. Our results show the importance of the closest rivers and streams for system diversity
383 and resilience, as the mostly community variation was found in the Boa Vista River and Brasilia
384 Stream (Fig. 3).

385 Although eDNA metabarcoding is a powerful tool for biodiversity, as it has been widely used for
386 different purposes and different taxonomic groups, including identification and quantification of
387 Neotropical ichthyofauna[16,76,77], many issues can hamper the metabarcoding results[7,10,78,79]. Shaw
388 et al.[10] drew attention to methodological considerations related to the eDNA sampling process
389 for freshwater fishes. According to them, the number of replicates is extremely important to
390 obtain accurate data. Specifically, they demonstrated that the collection of two eDNA replicates
391 per point were insufficient to detect less abundant taxa; however, adopting five replicates must
392 have a 100% detection rate. In addition, the eDNA sampling water column obtained is more
393 effective in detecting fish communities than sediment eDNA[10]. Here, we collected six replicates
394 per sampling point on the water surface. Furthermore, our rarefaction analysis clearly shows that
395 many individual samples have a very low sequencing depth, but considering the replicates all our
396 sampled localities reach the asymptote (Fig. S1).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

397 The bioinformatic methodological choices can also affect the metabarcoding results. Here, we
398 used three pipelines that showed the best results compared with other approaches[80]. We used
399 both OTU-level clustering at 97% level, with UPARSE[41], and the unique sequences with zero-
400 radius ZOTU-level denoising, with UNOISE3[24], and ASV-level Divisive Amplicon Denoising
401 Algorithm 2, with DADA2[23]. Both the OTUs and the ZOTUs are created using in USEARCH[81].
402 The initial steps as merging, filtering, and deduplicating are the same for both approaches, with
403 just the last step been different. The third approach generated ASVs through a parametric model,
404 based in Q-scores to calculate a substitution model, estimating a probability for each possible
405 base substitution, to infer true biological sequences from reads as implemented in DADA2[23].
406 Although we recorded some variation in the number of reads and "species" registered in each
407 pipeline, the results are very similar, highlighting their robustness.

408 Another potential bias in the results is data treatment. Here we used several data normalizations
409 for both alpha and beta diversity. Although historically more used, rarefied data is biased to
410 detect differentially abundant species[51] and the hill numbers are considered the best approach for
411 metabarcoding data[54]. Also, due to PCR biases, variation in the copy number of 12S genes per
412 cell/genome, as well as differences in size and biomass across the targeted organisms can
413 compromise a straightforward interpretation of OTU reads as an abundance measure[82–84].
414 However, rare (low abundant) ASVs, ZOTUs and OTUs are more likely to be an artefact (both
415 erroneous sequence or because of cross-talk[85]) and the true sequences are more stochastically
416 distributed due to the intrinsic low occurrence and detection probability[86,87]. Therefore, analyses
417 that weight more the most abundant molecular units could be preferable. As each method has its
418 own biases, we present here both approaches.

419 Finally, it is important to highlight that, in general, molecular data derived from “environmental
420 sequencing” should be seen as complementary to, rather than as competing with, traditional
421 taxonomic studies. Indeed, a confluence of both lines of evidence is highly warranted, as it will
422 be necessary to overcome their respective shortcomings. For instance, we have shown here that
423 many species occurring in the Itaipu fish pass system have no genetic data to be identified. To
424 improve the species detection with metabarcoding it is crucial to enhance the genetic reference
425 database through traditional inventories. Indeed, the metabarcoding approach is an intricate web
426 of feedback loops with the species taxonomy and ecology.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

427 **Conclusion:**

428 Despite the clear indication that the reference databases need to be continuously fed with
429 additional information on species that occur in the region, our results demonstrate the analytical
430 efficiency of the metabarcoding approach for monitoring fish species in the Itaipu´s fish pass
431 system. In addition, the methodology allowed, even when the specific identity of the ASVs,
432 OTUs, and ZOTUs were below 97% similarity with the species in our database, to carry out
433 estimates of species alpha and beta diversity. The use of such a methodology enables the
434 monitoring of the fish community with sufficient sensitivity to detect changes due to some
435 natural or anthropogenic event.

436 **Acknowledgements:**

437 Funding was provided by Itaipu Binacional to AON (grant #4500049847). We thank Itaipu
438 Binacional for providing postdoctoral fellow grants to GDP and NC and the Alexander van
439 Humboldt foundation for providing a postdoctoral fellow grant to CDR. We also thank Conselho
440 Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for awarding AO and MRP with
441 research fellowship grants (grants #304633/2017-8 and #302904/2020-4, respectively).

442 **References:**

443 1. de Souza Dias, V., Pereira da Luz, M., Medero, G. M. & Tarley Ferreira Nascimento, D.
444 An overview of hydropower reservoirs in Brazil: Current situation, future perspectives and
445 impacts of climate change. *Water* **10**, 592 (2018).

446 2. Patias, J., Zuquette, L. V. & Rodrigues-Carvalho, J. A. Piezometric variations in the
447 basaltic massif beneath the Itaipu hydroelectric plant (Brazil/Paraguay border): Right
448 Buttress Dam. *Bull. Eng. Geol. Environ.* **74**, 207–231 (2015).

449 3. Agostinho, A. A. Pesquisas, monitoramento e manejo da fauna aquática em
450 empreendimentos hidrelétricos. in *SEMINÁRIO SOBRE FAUNA AQUÁTICA E O SETOR*
451 *ELÉTRICO BRASILEIRO* 38–59 (Brasil, 1994).

452 4. Makrakis, S., Gomes, L. C., Makrakis, M. C., Fernandez, D. R. & Pavanelli, C. S. The
453 Canal da Piracema at Itaipu Dam as a fish pass system. *Neotrop. Ichthyol.* **5**, 185–195

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

454 (2007).

455 5. Dos Reis, R. B., Frota, A., Depra, G. D. C., Ota, R. Rú. & Da Graca, W. J. Freshwater
456 fishes from Paraná State, Brazil: an annotated list, with comments on biogeographic
457 patterns, threats, and future perspectives. *Zootaxa* **4868**, 451–494 (2020).

458 6. Becker, R. A., Sales, N. G., Santos, G. M., Santos, G. B. & Carvalho, D. C. DNA
459 barcoding and morphological identification of neotropical ichthyoplankton from the
460 Upper Paraná and São Francisco. *J. Fish Biol.* **87**, 159–168 (2015).

461 7. Milan, D. T. *et al.* New 12S metabarcoding primers for enhanced Neotropical freshwater
462 fish biodiversity assessment. *Sci. Rep.* **10**, 1–12 (2020).

463 8. Agostinho, A. A., Pelicice, F. M. & Gomes, L. C. Dams and the fish fauna of the
464 Neotropical region: impacts and management related to diversity and fisheries. *Brazilian*
465 *J. Biol.* **68**, 1119–1132 (2008).

466 9. Bonar, S. A., Hubert, W. A. & Willis, D. W. Standard methods for sampling North
467 American freshwater fishes. (2009).

468 10. Shaw, J. L. A. *et al.* Comparison of environmental DNA metabarcoding and conventional
469 fish survey methods in a river system. *Biol. Conserv.* **197**, 131–138 (2016).

470 11. Reis, R. E. *et al.* Fish biodiversity and conservation in South America. *J. Fish Biol.* **89**,
471 12–47 (2016).

472 12. Baumgartner, G. *et al. Peixes do baixo rio Iguaçu*. (Eduem, 2012).

473 13. Taberlet, P., Bonin, A., Coissac, E. & Zinger, L. *Environmental DNA: For biodiversity*
474 *research and monitoring*. (Oxford University Press, 2018).

475 14. Taberlet, P., Coissac, E., Pompanon, F., Christian, B. & Willerslev., E. Towards next-
476 generation biodiversity assessment using DNA metabarcoding. **33**, 2045–2050 (2012).

477 15. Ritter, C. D. *et al.* The pitfalls of biodiversity proxies: Differences in richness patterns of
478 birds, trees and understudied diversity across Amazonia. *Sci. Rep.* **9**, (2019).

479 16. Sales, N. G. *et al.* Space-time dynamics in monitoring neotropical fish communities using
480 eDNA metabarcoding. *Sci. Total Environ.* **754**, 142096 (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

481 17. Zinger, L. *et al.* Body size determines soil community assembly in a tropical forest. *Mol.*
482 *Ecol.* **28**, 528–543 (2019).

483 18. Baird, D. J. & Hajibabaei, M. Biomonitoring 2.0: a new paradigm in ecosystem
484 assessment made possible by next□generation DNA sequencing. *Mol. Ecol.* **21**, 2039–
485 2044 (2012).

486 19. Zinger, L. *et al.* Advances and prospects of environmental DNA in neotropical rainforests.
487 *Adv. Ecol. Res.* **62**, 331–373 (2020).

488 20. Cilleros, K. *et al.* Unlocking biodiversity and conservation studies in high□diversity
489 environments using environmental DNA (eDNA): A test with Guianese freshwater fishes.
490 *Mol. Ecol. Resour.* **19**, 27–46 (2019).

491 21. Sales, N. G., Wangensteen, O. S., Carvalho, D. C. & Mariani, S. Influence of preservation
492 methods, sample medium and sampling time on eDNA recovery in a neotropical river.
493 *Environ. DNA* **1**, (2019).

494 22. Blaxter, M. *et al.* Defining operational taxonomic units using DNA barcode data. 1935–
495 1943 (2005). doi:10.1098/rstb.2005.1725

496 23. Callahan, B. J. *et al.* DADA2: high-resolution sample inference from Illumina amplicon
497 data. *Nat. Methods* **13**, 581–583 (2016).

498 24. Edgar, R. C. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon
499 sequencing. *BioRxiv* 81257 (2016).

500 25. Muha, T. P., Rodriguez-Barreto, D., O'Rorke, R., Garcia de Leaniz, C. & Consuegra, S.
501 Using eDNA metabarcoding to monitor changes in fish community composition after
502 barrier removal. *Front. Ecol. Evol.* **9**, 28 (2021).

503 26. Kitano, T., Umetsu, K., Tian, W. & Osawa, M. Two universal primer sets for species
504 identification among vertebrates. *Int. J. Legal Med.* **121**, 423–427 (2007).

505 27. Stoeckle, M. Y., Soboleva, L. & Charlop-Powers, Z. Aquatic environmental DNA detects
506 seasonal fish abundance and habitat preference in an urban estuary. *PLoS One* **12**,
507 e0175186 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

508 28. Bylemans, J. *et al.* An environmental DNA□based method for monitoring spawning
509 activity: A case study, using the endangered Macquarie perch (Macquaria australasica).
510 *Methods Ecol. Evol.* **8**, 646–655 (2017).

511 29. De Souza, L. S., Godwin, J. C., Renshaw, M. A. & Larson, E. Environmental DNA
512 (eDNA) detection probability is influenced by seasonal activity of organisms. *PLoS One*
513 **11**, e0165273 (2016).

514 30. Ritter, C. D. *et al.* Locality or habitat? Exploring predictors of biodiversity in Amazonia.
515 *Ecography (Cop.).* **42**, 321–333 (2019).

516 31. *CFMV- Resolução n° 1000 de 11 de maio de 2012 - Dispõe sobre procedimentos e*
517 *métodos de eutanásia em animais e dá outras providências.* (2012).

518 32. Britski, H. A., de Silimon, K. Z. de S. & Lopes, B. S. Peixes do Pantanal: manual de
519 identificação, ampl. *Brasília, DF, Embrapa Informação Tecnológica* (2007).

520 33. Ota, R. R., Deprá, G. de C., Graça, W. J. da & Pavanelli, C. S. Peixes da planície de
521 inundação do alto rio Paraná e áreas adjacentes: revised, annotated and updated. *Neotrop.*
522 *Ichthyol.* **16**, (2018).

523 34. Neris, N., Villalba, F., Kamada, D. & Viré, S. Guía de peces del Paraguay/Guide of fishes
524 of Paraguay. (2010).

525 35. Pie, M. R. *et al.* Development of a real-time PCR assay for the detection of the golden
526 mussel (Limnoperna fortunei, Mytilidae) in environmental samples. *An. Acad. Bras.*
527 *Cienc.* **89**, 1041–1045 (2017).

528 36. Miya, M. *et al.* MiFish, a set of universal PCR primers for metabarcoding environmental
529 DNA from fishes: detection of more than 230 subtropical marine species. *R. Soc. open Sci.*
530 **2**, 150088 (2015).

531 37. Boeger, W. A. *et al.* Testing a molecular protocol to monitor the presence of golden
532 mussel larvae (Limnoperna fortunei) in plankton samples. *J. Plankton Res.* **29**, 1015–1019
533 (2007).

534 38. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

535 *EMBnet. J.* **17**, 10–12 (2011).

536 39. Van Rossum, G. & Drake, F. L. Python 3 References Manual. Scotts Valley CA:
537 CreateSpace. (2009).

538 40. R Core Team. R: the R project for statistical computing. 2019. *URL: https://www. r-*
539 *project. org/[accessed 2020-03-30]* (2020).

540 41. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads.
541 *Nat. Methods* **10**, 996–998 (2013).

542 42. Edgar, R. C. & Flyvbjerg, H. Error filtering, pair assembly and error correction for next-
543 generation sequencing reads. *Bioinformatics* **31**, 3476–3482 (2015).

544 43. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421
545 (2009).

546 44. Team, Rs. RStudio: integrated development for R. *RStudio, Inc., Boston, MA URL*
547 *http//www. rstudio. com* **42**, 84 (2015).

548 45. Wickham, H. tidyverse: Easily Install and Load "Tidyverse" Packages (Version R
549 package version 1.1. 1). (2017).

550 46. Wickham, H. *ggplot2: elegant graphics for data analysis*. (Springer, 2016).

551 47. Tang, Y., Horikoshi, M. & Li, W. ggfortify: unified interface to visualize statistical results
552 of popular R packages. *R J.* **8**, 474 (2016).

553 48. Auguie, B. & Antonov, A. gridExtra: Miscellaneous functions for "grid" graphics
554 (Version 2.2. 1)[Computer software]. (2016).

555 49. Kassambara, A. & Kassambara, M. A. Package 'ggpubr'. (2020).

556 50. Oksanen, J. *et al.* Vegan: community ecology package. R package version 1.17-4.
557 *http//cran. r-project. org>. Acesso em* **23**, 2010 (2010).

558 51. McMurdie, P. J. & Holmes, S. Waste not, want not: why rarefying microbiome data is
559 inadmissible. *PLoS Comput. Biol.* **10**, e1003531 (2014).

560 52. Jost, L. Entropy and diversity. *Oikos* **113**, 363–375 (2006).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

561 53. Marcon, E., Herault, B. & Marcon, M. E. Package 'entropart'. (2021).

562 54. Mächler, E., Walser, J.-C. & Altermatt, F. Decision making and best practices for
563 taxonomy-free eDNA metabarcoding in biomonitoring using Hill numbers. *BioRxiv*
564 (2020).

565 55. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive
566 analysis and graphics of microbiome census data. *PLoS One* **8**, e61217 (2013).

567 56. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion
568 for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

569 57. León, A., Reyes, J., Burriel, V. & Valverde, F. Data quality problems when integrating
570 genomic information. in *International Conference on Conceptual Modeling* 173–182
571 (Springer, 2016).

572 58. Callahan, B. J., McMurdie, P. J. & Holmes, S. P. Exact sequence variants should replace
573 operational taxonomic units in marker-gene data analysis. *ISME J.* **11**, 2639–2643 (2017).

574 59. Stahlhut, J. K. *et al.* DNA barcoding reveals diversity of hymenoptera and the dominance
575 of parasitoids in a sub-arctic environment. *BMC Ecol.* **13**, (2013).

576 60. Gillet, B. *et al.* Direct fishing and eDNA metabarcoding for biomonitoring during a 3-year
577 survey significantly improves number of fish detected around a South East Asian
578 reservoir. *PLoS One* **13**, e0208592 (2018).

579 61. Barrett, M. *et al.* Living planet report 2018: Aiming higher. (2018).

580 62. Díaz, S. M. *et al.* The global assessment report on biodiversity and ecosystem services:
581 Summary for policy makers. (2019).

582 63. Dudgeon, D. Asian river fishes in the Anthropocene: threats and conservation challenges
583 in an era of rapid environmental change. *J. Fish Biol.* **79**, 1487–1524 (2011).

584 64. Dudgeon, D. Multiple threats imperil freshwater biodiversity in the Anthropocene. *Curr.*
585 *Biol.* **29**, R960–R967 (2019).

586 65. He, F. *et al.* Disappearing giants: a review of threats to freshwater megafauna. *Wiley*
587 *Interdiscip. Rev. Water* **4**, e1208 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

588 66. Agostinho, A. A., Thomaz, S. M. & Gomes, L. C. Threats for biodiversity in the
589 floodplain of the Upper Paraná River: effects of hydrological regulation by dams. (2018).

590 67. Santana, M. L., Carvalho, F. R. & Teresa, F. B. Broad and fine-scale threats on threatened
591 Brazilian freshwater fish: variability across hydrographic regions and taxonomic groups.
592 *Biota Neotrop.* **21**, (2021).

593 68. Matthews, W. J. *Patterns in freshwater fish ecology*. (Springer Science & Business Media,
594 2012).

595 69. de Oliveira Bueno, E., Alves, G. J. & Mello, C. R. Hydroelectricity water footprint in
596 Parana hydrograph region, Brazil. *Renew. Energy* **162**, 596–612 (2020).

597 70. Camacho Guerreiro, A. I., Amadio, S. A., Fabre, N. N. & da Silva Batista, V. Exploring
598 the effect of strong hydrological droughts and floods on populational parameters of
599 Semaprochilodus insignis (Actinopterygii: Prochilodontidae) from the Central Amazonia.
600 *Environ. Dev. Sustain.* **23**, (2021).

601 71. Jespersen, H., Rasmussen, G. & Pedersen, S. Severity of summer drought as predictor for
602 smolt recruitment in migratory brown trout (Salmo trutta). *Ecol. Freshw. Fish* **30**, 115–
603 124 (2021).

604 72. Pool, T. K., Grenouillet, G. & Villéger, S. Species contribute differently to the taxonomic,
605 functional, and phylogenetic alpha and beta diversity of freshwater fish communities.
606 *Divers. Distrib.* **20**, 1235–1244 (2014).

607 73. Oliveira, E. F. de, Goulart, E. & Minte-Vera, C. V. Fish diversity along spatial gradients
608 in the Itaipu Reservoir, Paraná, Brazil. *Brazilian J. Biol.* **64**, 447–458 (2004).

609 74. Daga, V. S. *et al.* Homogenization dynamics of the fish assemblages in Neotropical
610 reservoirs: comparing the roles of introduced species and their vectors. *Hydrobiologia*
611 **746**, 327–347 (2015).

612 75. Vitule, J. R. S. Introdução de peixes em ecossistemas continentais brasileiros: revisão,
613 comentários e sugestões de ações contra o inimigo quase invisível. *Neotrop. Biol.*
614 *Conserv.* **4**, 111–122 (2009).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

615 76. Mariac, C. *et al.* Species□level ichthyoplankton dynamics for 97 fishes in two major river
616 basins of the Amazon using quantitative metabarcoding. *Mol. Ecol.* (2021).

617 77. Jackman, J. M. *et al.* eDNA in a bottleneck: obstacles to fish metabarcoding studies in
618 megadiverse freshwater systems. *Environ. DNA* (2021).

619 78. Bessey, C. *et al.* Maximizing fish detection with eDNA metabarcoding. *Environ. DNA* **2**,
620 493–504 (2020).

621 79. Evans, N. T. *et al.* Fish community assessment with eDNA metabarcoding: effects of
622 sampling design and bioinformatic filtering. *Can. J. Fish. Aquat. Sci.* **74**, 1362–1374
623 (2017).

624 80. Prodan, A. *et al.* Comparing bioinformatic pipelines for microbial 16S rRNA amplicon
625 sequencing. *PLoS One* **15**, e0227434 (2020).

626 81. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST.
627 *Bioinformatics* **26**, 2460–2461 (2010).

628 82. Elbrecht, V. & Leese, F. Can DNA-based ecosystem assessments quantify species
629 abundance? Testing primer bias and biomass—sequence relationships with an innovative
630 metabarcoding protocol. *PLoS One* **10**, e0130324 (2015).

631 83. Pawluczyk, M. *et al.* Quantitative evaluation of bias in PCR amplification and next-
632 generation sequencing derived from metabarcoding samples. *Anal. Bioanal. Chem.* **407**,
633 1841–1848 (2015).

634 84. Holman, L. E., Chng, Y. & Rius, M. How does eDNA decay affect metabarcoding
635 experiments? *Environ. DNA* (2021).

636 85. Edgar, R. C. UNCROSS2: identification of cross-talk in 16S rRNA OTU tables. *BioRxiv*
637 400762 (2018).

638 86. MacArthur, R. H. *Geographical ecology: patterns in the distribution of species*.
639 (Princeton University Press, 1984).

640 87. Leray, M. & Knowlton, N. Random sampling causes the low reproducibility of rare
641 eukaryotic OTUs in Illumina COI metabarcoding. *PeerJ* **5**, e3006 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.17.456687; this version posted August 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

642 88. Team, Q. D. QGIS geographic information system. *Open Source Geospatial Found. Proj.*
643 *Versão* **2**, (2015).