1 **Title:**

2 **eDNA in a bottleneck: obstacles to fish metabarcoding studies in megadiverse freshwater**

3 **systems**

4

5 **Authors:**

6 Jake M. Jackman[1], Chiara Benvenuto[1], Ilaria Coscia[1], Cintia Oliveira Carvalho[2], Jonathan S.

7 Ready[2], Jean P. Boubli[1], William E. Magnusson[3], Allan D. McDevitt[1]* and Naiara Guimarães

8 Sales[1,4]*

9

10 **Addresses:**

11 [1]*Environment and Ecosystem Research Centre, School of Science, Engineering and Environment,*

12 *University of Salford, Salford, M5 4WT, UK*

13 [2]*Centro de Estudos Avançados de Biodiversidade, Instituto de Ciências Biológicas, Universidade*

14 *Federal do Pará, Belém, Brazil*

15 [3]*Coordenação de Biodiversidade, Instituto Nacional de Pesquisas da Amazônia, Manaus,*

16 *Amazonas, Brazil*

17 [4]*CESAM - Centre for Environmental and Marine Studies, Departamento de Biologia Animal,*

18 *Faculdade de Ciências da Universidade de Lisboa, Lisbon, Portugal*

19

20 ***Corresponding authors:**

21 Naiara Guimarães Sales, naiarasl@gmail.com

22 Allan McDevitt, a.mcdevitt@salford.ac.uk

23

24 **Running title:** Obstacles to eDNA surveys in megadiverse systems

25

26 **Keywords:** Amazon, barcoding gap, freshwater, MiFish, Neotropics, reference database,

27 taxonomic resolution

28

ABSTRACT

The current capacity of environmental DNA (eDNA) to provide accurate insights into the biodiversity of megadiverse regions (e.g., the Neotropics) requires further evaluation to ensure its reliability for long-term monitoring. In this study, we first evaluated the taxonomic resolution capabilities of a short fragment from the 12S rRNA gene widely used in fish eDNA metabarcoding studies, and then compared eDNA metabarcoding data from water samples with traditional sampling using nets. For the taxonomic discriminatory power analysis, we used a specifically curated reference dataset consisting of 373 sequences from 264 neotropical fish species (including 47 newly generated sequences) to perform a genetic distance-based analysis of the amplicons targeted by the MiFish primer set. We obtained an optimum delimitation threshold value of 0.5% due to lowest cumulative errors. The barcoding gap analysis revealed only a 50.38% success rate in species recovery (133/264), highlighting a poor taxonomic resolution from the targeted amplicon. To evaluate the empirical performance of this amplicon for biomonitoring, we assessed fish biodiversity using eDNA metabarcoding from water samples collected from the Amazon (Adolpho Ducke Forest Reserve and two additional locations outside the Reserve). From a total of 84 identified Molecular Operational Taxonomic Units (MOTUs), only four could be assigned to species level using a fixed threshold. Measures of α-diversity analyses within the Reserve showed similar patterns in each site between the number of MOTUs (eDNA dataset) and species (netting data) found. However, β-diversity revealed contrasting patterns between the methods. We therefore suggest that a new approach is needed, underpinned by sound taxonomic knowledge, and a more thorough evaluation of better molecular identification procedures such as multi-marker metabarcoding approaches and tailor-made (i.e., order-specific) taxonomic delimitation thresholds.

2

INTRODUCTION

The need for advancing our understanding of the world's biodiversity increases in parallel with the acceleration of anthropogenic impacts on the planet's ecosystems. To implement strategies to minimise the effects of human impacts, understanding compositions of species assemblages within ecosystems is paramount (Morris, 2010). This task is particularly difficult when investigating megadiverse regions of the world such as the Neotropics, which harbour an extremely large diversity of living organisms. The Amazon basin, for example, is estimated to hold the highest diversity of freshwater fish found anywhere on the planet (Albert & Reis, 2011). To date, it has been documented that 2,406 species belonging to 514 genera and 56 families of fish inhabit the tributaries of the Amazon River, with many more not yet described (Jézequel et al., 2020). Undoubtedly, this region serves as a biodiversity hotspot, with Amazonian fishes representing ~15% of all freshwater fish species described worldwide (Jézequel et al., 2020; Leroy et al., 2019).

Due to the increase in anthropogenic impacts in Neotropical rivers (e.g., pollution, siltation, mining and damming), there is a growing danger that this rich biodiversity will be lost before it can be fully described (Alho, Reis & Aquino, 2015; Agostinho, Thomaz & Gomes, 2005). This emphasizes the urgency of accurate and rapid biodiversity assessments throughout the region. Although years of biomonitoring in the Neotropical region have been conducted, inventories of fish fauna remain incomplete (Frota, Depra, Petenucci & Graça, 2016) demonstrating the need for improvements in biodiversity assessment methods through a more integrative approach aimed at circumnavigating traditional sampling limitations.

A powerful addition to biodiversity surveying is the application of DNA-based approaches. The use of environmental DNA (eDNA; i.e., DNA extracted from environmental samples such as water or soil; Taberlet et al., 2012) for biomonitoring is now widespread, particularly within freshwater ecosystems (Hering et al., 2018; Senapati et al., 2018). Advances in next-generation sequencing (NGS) have unlocked the potential use of eDNA metabarcoding for monitoring whole communities within specific taxonomic groups (e.g., fishes; Miya, Gotoh & Sado, 2020). Recent studies display its efficiency in different aquatic environments and show how it compares favourably to, or even outperforms, traditional sampling methods in terms of species detections (McDevitt et al., 2019; McElroy et al., 2020) and facilitates investigations into patterns of extirpation, invasive species detection, and dynamics of species richness (Lacoursière-Roussel et al., 2018; Sales et al., 2021). Despite the increase of eDNA surveys in megadiverse systems,

83   several limiting factors prevent its full application (Cilleros et al., 2019; Doble el al., 2020; Sales
84   et al., 2021). Although challenges associated with the collection and preservation of samples have
85   already been addressed (Sales et al., 2019), obstacles in taxonomic assignments remain
86   understudied.

87   The ability to detect species is reliant on the reference library used to assign retrieved
88   sequences to species level and on the robustness of the taxonomic resolution of targeted gene
89   fragments (Sassoubre, Yamahara, Gardner, Block & Boehm, 2016). As outlined by the
90   International Barcode of Life initiative (Hebert, Cywinska, Ball & Dewaard, 2003), databases
91   targeting mitochondrial cytochrome oxidase subunit I (COI) are generally more complete than
92   other gene regions such as the mitochondrial 12S and 16S rRNA genes. However, the COI region
93   has been shown to be less suitable for eDNA metabarcoding work for vertebrates (due to inflated
94   detections of non-target organisms), therefore substantiating the need to explore the use of more
95   suitable gene regions such as 12S and 16S and to expand their respective databases (Collins et al.,
96   2019; Deagle, Jarman, Coissac, Pompanon & Taberlet, 2014).

97   In this context, a set of universal PCR primers for metabarcoding eDNA from fishes has
98   been developed targeting a length of around 163-185 bp of the 12S rRNA gene region (MiFish;
99   Miya et al., 2015). This widely used primer set has been pivotal in describing fish communities on
100   a truly global scale (Miya, Gotoh & Sado, 2020), and has also provided meaningful information
101   for species-rich rivers (Ahn et al., 2020). Despite the efforts made by global barcoding initiatives
102   towards the development of more comprehensive reference databases, in most circumstances these
103   databases remain far from complete, especially for the currently commonly used 12S
104   mitochondrial gene region (Doble et al., 2020; Sales et al., 2020; Weigand et al., 2019).
105   Furthermore, the proportion of species sequenced is lower in species-rich regions, such as the
106   Neotropics with poorly sampled habitats and taxa. As it stands now, only a limited number of fish
107   species can be found in DNA databases, hindering the potential of eDNA metabarcoding as a
108   biomonitoring method in these regions (Cilleros et al., 2019; Sales et al., 2019, 2021).

109   Besides the well-known obstacles posed by incomplete reference databases, the taxonomic
110   resolution of target amplicons can hinder inferences of species occurrence. For example, the
111   current fixed general threshold used for species assignment (e.g., > 97% similarity - Sales et al.,
112   2021; >98% similarity - Marques et al., 2020) assumes the existence of a barcoding gap (i.e.,
113   presence of a gap between the highest intraspecific and the lowest interspecific variation within

114   the analysed taxonomic group; Meyer & Paulay, 2005). The accuracy of selected markers in

115   detecting species relies on the separation between the intra and inter-specific divergences, and the

116   greater the overlap between these variations, the less effective DNA barcoding becomes (Meyer

117   & Paulay, 2005). In the absence of a barcoding gap, the use of a general threshold may lead to an

118   inaccurate taxonomic assignment, over splitting some taxa while lumping together others. This

119   issue is particularly important in highly biodiverse areas, especially where the proportion of

120   recently diverged species and cryptic species is relatively high. In this regard, Sales et al. (2021)

121   have highlighted issues of low taxonomic resolution with the widely used 12S MiFish marker,

122   unable to assign members of the genus *Prochilodus* to the species level. This matter then raises

123   conservation issues as native species (e.g., *P. hartii*) could be wrongfully assigned to congeneric

124   invasive species (e.g., *P. argenteus*). eDNA metabarcoding represents a technological leap for

125   characterizing and assessing biodiversity (Petruniak, Bradley, Kelly & Hanner, 2020), but these

126   obstacles can represent a bottleneck to its application in highly biodiverse regions (Sales et al.,

127   2020, 2021).

128   In this study, we aim to assess the use of eDNA metabarcoding as a tool to estimate fish

129   biodiversity in a megadiverse Neotropical system, by directly comparing it to data obtained by

130   netting surveys carried out at the same time and sites. We adopted a step by step, integrative

131   approach using newly generated and existing sequence data from a wide range of neotropical fish

132   species, eDNA from water samples and netting data. We first performed a genetic distance-based

133   analysis to investigate the optimum delimitation values based on the 12S MiFish primers, followed

134   by an assessment of the performance of the delimitation values through a barcoding gap analysis.

135   We then analysed water samples collected from the Brazilian Amazon and we compared measures

136   of α-diversity (species richness) and β-diversity (change in species composition among locations)

137   generated from eDNA and traditional netting data collected in the same sampling sites.

138 **METHODS AND MATERIALS**

139 *Evaluation of taxonomic resolution power of the 12S rRNA fragment*

140 In order to improve taxonomic assignment, 47 fish species (Table S1) were sequenced and

141 included in a customised reference database for Neotropical fishes. Tissue samples were provided

142 by the Grupo de Investigação Biológica Integrada (GIBI) tissue collection, located at the

143 Universidade Federal do Pará (UFPA; Belém, Brazil). Fragments of the mitochondrial 12S rRNA

144 gene were obtained using PCR (conducted using MiFish primers, following the same protocols

145 used for eDNA samples described below) and Sanger sequenced. Consensus sequences were

146 obtained with Geneious v8.1 (Kearse et al., 2012).

147 Several analyses were conducted to verify the taxonomic resolution of 12S rRNA

148 fragments targeted by the MiFish metabarcoding primers. For this stage of the analysis we created

149 a combined enhanced dataset comprising 264 neotropical fish species, belonging to eight orders

150 (Table S2) from existing GenBank and newly sequenced data (373 barcodes in total). Threshold

151 optimization analysis was conducted using the SPIDER package in R v3.5.1 (R Core Team 2019)

152 through the *threshOpt* function. This returns the false positive and negative rates of identification

153 accuracy for different threshold values as well as providing the total cumulative errors (false

154 positive + false negative). When applying a range of thresholds, this function allows the

155 optimization of values aiming to minimise the error rates. The default threshold for this function

156 is set to 0.01 (1%), however, this can be changed using the *threshVal* function, and here were

157 included values ranging from 0.001 to 0.03 (0.1% to 3%). Using the thresholds generated from the

158 *threshVal* function, a genetic-based delimitation analysis was performed using K2P genetic

159 distances (Kimura, 1980). The threshold estimates were applied as the best delimitation values to

160 estimate species. Milan et al. (2020) suggested threshold values ranging from 0.4% to 0.55% for a

161 different fragment of the 12S rRNA gene. Currently, no delimitating reference values for the

162 MiFish 12S marker used here have been published, thus we established distance thresholds, which

163 were then used within the *threshID* function. The *threshID* function assigns four possible results

164 for each sequence in the dataset: "correct", "incorrect", "ambiguous", and "no ID". The "correct"

165 results suggest that all matches within the threshold value of the query are the same species and

166 "no ID" shows that no matches were found to any individual within the threshold.

167 SPIDER was also used as a means to investigate the presence/absence of the "barcoding gap" by

168 identifying the furthest intraspecific distance among the same species, using the *maxInDist()*

6

169   function and the closest non-conspecific using the *nonConDist()* function. The occurrence of no

170   barcoding gap is represented by a zero or negative distance as a result of the maximum intraspecific

171   distance being subtracted from the minimum interspecific distance.

172

173   ***Study sites and sample collection***

174   The eDNA component of the study was conducted in six different sites located in the Brazilian

175   Amazon, with four sites inserted inside the Adolpho Ducke Forest Reserve (sites B1-B4; Fig. 1;

176   Table S3) and two outside (A and C). The Reserve is a designated 100 km^2 area protecting

177   continuous and non-isolated rainforest established by the National Institute of Amazon Research

178   (INPA, for a more detailed description see Supp. Information). The Ducke Reserve represents one

179   of the first sites of the Brazilian Long-Term Ecological Research Program (PELD) and the

180   Biodiversity Research Program (PPBio), run by the Brazilian Ministry of Science and Technology.

181   A long-term study by Zuanon et al. (2015) has revealed that the streams of the Ducke Reserve

182   comprise an estimated 70 species of fish from seven taxonomic orders: Characiformes,

183   Siluriformes, Gymnotiformes, Perciformes (from which now Cichliformes have been separated),

184   Cyprinodontiformes, and Synbranchiformes. The most diverse taxonomic order in the reserve is

185   the Characiformes, comprising six families and 24 species followed by the Siluriformes with seven

186   families and 17 species, Gymnotiformes with four families and 12 species, Perciformes with two

187   families and 13 species, Cyprinodontiformes with one family and three species, and

188   Synbranchiformes with one family and one species.

189   Two main drainage basins have there headwaters near the centre of Ducke Reserve, one on

190   the western side of the reserve that flows to the black waters of the Rio Negro and one on the

191   eastern side that flows to the white (sediment laden) waters of the Amazon River. Sampling within

192   the reserve was carried out on the north-west side, along four tributaries of Acará. Three sites were

193   on unnamed third-order streams (B1-B3; Fig. 1) and one was on Barro Branco, a second-order

194   stream (B4; Fig 1). Net sampling and eDNA sampling protocols were both carried out in January

195   2019. Additionally, for the eDNA analysis, two more sites were sampled opportunistically outside

196   the Reserve (A and C; Fig. 1), one on Aturiá stream, which is a northern tributary of the Rio Negro

197   (A; Fig. 1), and the other at "the meeting of the waters" where the black waters from the Rio Negro

198   and the white waters of the Solimões (upper Amazon) form the Amazon River (C; Fig. 1).

7

199   For the eDNA metabarcoding, from each of the four streams within Ducke Reserve, three

200   water replicates were taken from three positions (from the left bank, middle and right bank) along

201   a transect at 0m, 25m, and 50m using 500 ml water bottles resulting in a total of nine replicates

202   per stream. Two water-sample replicates were collected at site A and three at site C (no netting

203   was performed at these sites). At the start of each sampling period, a field blank was collected,

204   totaling four field blanks overall. Samples were collected prior to the start of the netting sampling

205   to avoid the risk of contamination. Water samples were manually filtered using a syringe and

206   Sterivex enclosed filters (0.45μm, Merck Millipore) and kept cool until transport to the UK. In

207   total, 51 samples were analysed (including 41 water samples, four field blanks, two extraction

208   blanks, four PCR negative controls, Table S4). The eDNA extraction, amplification of the 12S

209   rRNA fragment using the MiFish primer set, and library preparation were conducted following the

210   procedure described in Sales et al. (2021).

211   Netting was conducted in the same four sites in the Reserve (sites B1-B4) straight after

212   water was collected for eDNA analysis (McDevitt et al., 2019), spanning a range of 50m. Each

213   stream was sampled following the rapid assessment protocol RAPELD (Magnusson et al., 2005).

214   Two nets (5 mm mesh size) were placed at the edge of the sampling area to prevent fish accessing

215   and exiting the study area. An additional net was used to subdivide the full lenght in segments. A

216   total of three sub-stretches of ~16 meters each moving upstream were covered. The fishes were

217   collected with nets and hand sieves (2 mm mesh size) then stored in buckets to allow for individual

218   identification until they were released back into the stream. Morphological identification of

219   collected specimens was conducted by CB following Zuanon et al. (2015).

220

Bioinformatic analysis

222 The bioinformatic analysis was completed using the *OBITools* metabarcoding package (Boyer

223 et al., 2016), following the protocol described by Sales et al. (2021). *FastQC* was used to assess

224 the quality of the reads and a length filter (command *obigrep*) was used to select fragments of 140-

225 190 bp and to remove reads with ambiguous bases. SWARM was then used to compute sequence

226 differences between aligned pairs of amplicons to delineate MOTUs (Mahé, Rognes, Quince, de

227 Vargas & Dunthorn, 2014). The taxonomic assignment was conducted using the *ecotag* command,

228 which works in two phases: initially, it performs a search of the assigned reference database to

229 locate the sequence with the highest overall similarity to the query sequence; then the similarity

230 value obtained from the first step is set as the threshold for searches of additional sequences, equal
231 to or lower than that of the threshold value within the assigned database. Stringent filtering steps
232 were applied to the final dataset to remove Molecular Operational Taxonomic Units
233 (MOTUs)/reads originating from sequencing errors or contamination to avoid false positives for
234 the library (Table S3). To reach this target, all non-fish reads were removed from the dataset,
235 including non-target species (e.g., human and domestic species reads) and MOTUs that were likely
236 to have been carried over from contamination. To remove putative contaminants, the maximum
237 number of reads recorded in the controls (field collection, DNA extraction, and PCR blanks) were
238 removed from all samples. Finally, all MOTUs with <10 reads were removed from the final
239 dataset. The final taxonomic assignment was conducted according to current fixed general
240 thresholds: MOTUs were assigned at species level when matching the reference sequence with
241 >97% similarity as performed in previous studies in the Neotropics (Sales et al., 2020, 2021), at
242 genus level with 95-97% similarity, at family level with 90-95% similarity, and the highest
243 taxonomic level of order was attributed to MOTUs with less than 90% similarity matching the
244 reference sequences.

245

246 ***Data analyses***

247 Given the difficulties of taxonomic assignment without complete reference databases, species
248 identification was not possible from eDNA metabarcoding data and therefore MOTUs as opposed
249 to species were used for all subsequent analyses. Replicates were pooled (nine water samples per
250 site for the Ducke streams, two water samples for Aturiá and three for the Solimões river) before
251 the following statistical analyses.

252 The MetacodeR package version 0.3.4 (Foster, Sharpton & Grünwald, 2017) was used to
253 analyse the overall taxonomic diversity of the final eDNA dataset. A heat tree displaying the
254 overall sample reads was produced to display the patterns of MOTU distribution of the eDNA data
255 per taxonomic family. To visualise the magnitude of uncertainty of taxonomic assignments, a
256 schematic phylogenetic tree was adapted from Betancur-R et al. (2017) representing the orders,
257 families, genera and species detected by eDNA metabarcoding with the respective number of
258 MOTUs assigned to each taxonomic group.

259 For the analysis of the diversity contained within the eDNA dataset (MOTU richness/α-
260 diversity and β-diversity), the data were analysed with a presence/absence approach as suggested

261   by Li et al. (2018). The α-diversity for the eDNA data (richness) was calculated as the total number

262   of MOTUs found in each sample site and the β-diversity was obtained by the Jaccard dissimilarity

263   index using the *vegdist* function in the vegan 2.5-2 package (Oksanen et al., 2013). Principal

264   Coordinates Analysis (PCoA) was then used to investigate the relationship between distance and

265   sites generated through the *cmdscale* function in the β-diversity matrix. The α-diversity analysis

266   for the netting dataset was calculated as the total number of identified species found in each sample

267   site and the β-diversity analysis was performed using the same method applied to the eDNA

268   dataset. The results of the separate eDNA and netting β-diversity analyses were then superimposed

269   to produce a final figure.

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292 **RESULTS**

293 *Evaluation of taxonomic resolution of the 12S rRNA fragment*

294 Analysis of the taxonomic delimitation power of the 12S fragment targeted by the MiFish primers

295 indicated a high rate of ambiguous identification (13.40% - 23.59%) for values ranging from 1%

296 to 3%, with higher rates associated with increased threshold values. Optimum threshold analysis

297 identified 0.1% up to 0.5% as intraspecific values with the lowest number of cumulative errors

298 (Fig. 2A). Barcoding gap analysis revealed an extensive overlap between intraspecific and

299 interspecific divergences, and only 133 species out of 264 were successfully recovered due to the

300 absence of a barcoding gap (Fig. 2B). The overlap between intraspecific and interspecific genetic

301 distances was particularly evident for species still lacking a full taxonomic description (e.g.,

302 *Trichomycterus* spp., *Hypostomus* spp. and *Harttia* spp.). Still, high divergence values (>3%) were

303 also found for several other species including the Amazonian cichlids *Heros severus* (10.31%) and

304 *Aequidens metae* (8.21%), and the loricariids *Hypostomus affinis* (12.3%), *H. plecostomus* (3.64%)

305 and *Harttia carvalhoi* (3.57%).

306

307 *eDNA data analysis*

308 A total of 4,416,267 reads were obtained after trimming, merging, and length filtering during

309 bioinformatic analysis (Table S5). Considering only MOTUs belonging to Actinopterygii, a final

310 dataset containing 3,838,166 reads and 84 MOTUs was used for downstream analyses. A total of

311 seven different taxonomic orders of fish were identified as a result of the taxonomic assignment:

312 Characiformes, Cichliformes, Siluriformes, Gobiiformes, Synbranchiformes, Gymnotiformes, and

313 Cypriniformes (Figs 3 and 4; Table S6). Taxonomic assignment was poor for the 84 recovered

314 MOTUs; only four were identified to species level with the fixed general threshold, whereas 41

315 were assigned solely at the family level and 37 could only be attributed to the order level (Fig. 4).

316 From the MOTUs identified to species level, one is known to occur in the Ducke Reserve (*Hoplias*

317 *malabaricus*) and two have the genus present in this area (*Aequidens* and *Synbranchus*) and one

318 has been detected only in the Aturiá (Site A; *Phreatobius* sp.). A MOTU was assigned to *Aequidens*

319 *metae* (0.976 similarity) and *Aequidens pallidus* has been recorded from the Reserve. Another

320 MOTU was assigned to *Synbranchus marmoratus* (0.976 similarity) but for the studied area, an as

321 yet undescribed congeneric species (*Synbranchus* sp.) has been reported. A visualization of the

322 fish community structure was obtained based on the taxonomic identification at the family level

323　due to the difficulties of species identification. The assemblage structure depicted by the heat tree

324　(Fig. 4) demonstrated a higher overall MOTU count for Characiformes (especially for the

325　Anostomidae and Erythrinidae families) and Cichliformes (family Cichlidae). Furthermore, the

326　heat trees evidenced the fish diversity recovered from eDNA samples due to the occurrence of 16

327　families among the 84 analysed MOTUs.

328

Alpha and beta diversity

330　The netting dataset revealed 18 different species in total (Figs 5A and 5B) with nine species for

331　B1, 14 for B2, six for B3, and 11 for B4 (Figs 5B). Species richness for the eDNA data revealed a

332　total of 32 unique MOTUs for the Ducke Reserve streams and a total of 46 MOTUs across the

333　Ducke Reserve streams (B1: five MOTUs; B2: 19 MOTUs; B3: eight MOTUs and B4: 14 MOTUs;

334　Fig. 6A). The eDNA dataset also revealed a total of 55 MOTUs for the Aturiá stream, 50 of which

335　are unique to this sampling location with five of the MOTUs being shared with the other sampling

336　locations. Three MOTUs (all of which are unique) were found in the Solimões (upper Amazon)

337　river (Fig. S1).

338　　　β-diversity patterns inferred from Principal Coordinate Analysis (PCoA) showed a

339　distinction between the Ducke Reserve streams and Aturiá and Solimões (Fig. S1). β-diversity also

340　shows a clear distinction between the Aturiá stream and Solimões river. Within the Ducke Reserve,

341　the four streams shared a total of 16 MOTUs and the β-diversity analysis shows a slightly more

342　clustered pattern as opposed to the other two sampling locations in which Solimões shares no

343　MOTUs with any other sampling location and the Aturiá shares only five MOTUs with the streams

344　of the Ducke Reserve. Using only the sampling locations within the Ducke Reserve also

345　demonstrated the distinction between the species compositions with considerable ordination

346　distance between each point (Fig. 6B). Only sampling points B3 and B4 were slightly clustered,

347　indicating that they share a more similar composition to each other compared to B1 and B2.

348　Otherwise, large ordination distances separate all sampling locations within the Ducke Reserve.

349　β-diversity analysis using the netting dataset again reveals marked distinctions between species

350　compositions for all four analysed locations (Fig. 6B). However, in this case locations B1 and B2

351　showed the most similar compositions obtained through the netting data. β-diversity for the netting

352　dataset did not show similar patterns compared with that obtained from the eDNA dataset (Fig.

353　6B).

12

DISCUSSION

354

355 Understanding community compositions and species distributions within a target area is a crucial
356 step to inform conservation management strategies. This requires cost-effective and robust
357 methods for surveying biodiversity in a timely manner and eDNA metabarcoding has emerged as
358 a promising method for surveying whole ecosystems (Lacoursière-Roussel et al., 2018; Murienne
359 et al., 2019; Valdez-Moreno et al., 2019). As the use of genetic approaches to biodiversity
360 monitoring relies on accurate, precise species identifications, the lack of appropriate taxonomic
361 resolution and, consequently, detailed reference databases prevents the use of eDNA for
362 biomonitoring from reaching its full potential. Herein, we demonstrate how the combination of
363 these factors can significantly hinder the application of eDNA metabarcoding in monitoring fish
364 communities in a megadiverse Neotropical system.

365 There is a knowledge gap regarding the optimum threshold for taxonomic assignments, as
366 well as the taxonomic resolution/discriminative power of the gene fragments currently used in
367 metabarcoding studies, particularly within megadiverse regions (Milan et al., 2020; Servis, Reid,
368 Timmers, Stergioula & Naro-Maciel, 2020; Sales et al., 2021). Here, by performing a genetic-
369 distance-threshold analysis, an optimum threshold value of 0.5% was obtained as a result of the
370 lowest total cumulative errors (Fig. 2A). This value is in agreement with a previous study
371 evaluating the taxonomic resolution of full length and mini-barcodes targeting the same gene
372 (Milan et al., 2020, 0.55% for mini-barcodes), but contrasts with the standard threshold values (2-
373 3%) currently widely applied for fish species delimitation based on various 12S rRNA fragments.
374 Thresholds around 2% have been used in DNA barcoding studies targeting the mitochondrial COI
375 region to delimit fish species and this has generally been extrapolated to eDNA surveys. However,
376 using a fixed threshold (e.g., 2%) as a blanket value for studies in poorly described, species-rich
377 areas can lead to incorrect conclusions regarding species richness. Consequences of this have been
378 highlighted by Sales et al. (2018) where 306 fish DNA barcodes from the COI region showed that
379 over one fifth had a mean intraspecific genetic divergence higher than 2%, flagging numerous
380 possibilities of potential new MOTUs, cryptic species or errors originating from previous
381 morphological identification of species.

382 For DNA barcodes to be considered effective and accurate, there should be a separation
383 between intraspecific and interspecific divergences in the analysed marker, referred to as the
384 "barcoding gap" (Meyer & Paulay, 2005). Analysis of the combined dataset investigating the

13

385   presence of a barcoding gap revealed an extensive overlap between intraspecific variation and

386   interspecific distances, leading to only 50% of species being recovered (Fig. 2B). The occurrence

387   of no barcoding gap represents high levels of intra- and interspecific divergence within the

388   analysed dataset. This suggests multiple possibilities: the presence of potential cryptic species (a

389   common phenomenon for Neotropical fishes; Melo, Ochoa, Vari & Oliveira, 2016; Pugedo, de

390   Andrade Neto, Pessali, Birindelli & Carvalho, 2016; Sales et al., 2018) errors within the reference

391   database originating from poor taxonomy of previously identified species (Locatelli, McIntyre,

392   Therkildsen & Baetscher, 2020); or poor taxonomic resolution of the marker used (Meyer &

393   Paulay, 2005). Considering the high diversity of freshwater fish species in the Neotropics and

394   consequently broad variation between intraspecific and interspecific genetic distances, caution

395   should be taken when using universal cut-off values for species delimitation. As evidenced by the

396   barcoding gap analysis, we stress that the application of a general threshold is not advisable, and

397   values should be optimized considering the delimitation power of target fragments and the focal

398   taxa. In order to achieve a robust taxonomic assignment, a thoughtful choice of primers is required.

399   Although the 12S MiFish primers have undoubtedly been successfully applied in many studies and

400   justifiably remain widely used (see Miya et al., 2020), it should be noted that it has been

401   demonstrated that these primers might not provide an unambiguous identification of closely related

402   species in some systems (Doble et al., 2020; Sales et al., 2021). Greater species level assignments

403   can be achieved through the use of more specific primer sets (Doble et al., 2020), and/or by

404   applying a multi-marker approach, taking into account that different markers/locus may show

405   distinct taxonomic coverage and divergence ranges (Zhang et al., 2020).

406   Despite using a large dataset (373 sequences from 264 species), the results provided remain

407   limited due to the lack of complete reference databases for Neotropical fish species. Not only are

408   a low number of species are included but also, most of these are represented by just one sequenced

409   specimen, hampering sound analyses regarding intraspecific genetic divergence and limiting

410   further investigations. Another major impediment to the expansion of the reference data is the

411   prevalence of efforts towards the construction of databases targeting very specific and short

412   fragments, and more importantly, the lack of effort in making these sequences fully available in

413   public depositories. For example, for Neotropical fish species, customised reference databases

414   have been constructed targeting different fragments of the 12S gene (Cilleros et al. 2019; Milan et

415   al., 2020; Sales et al., 2021) but in some cases only the fragment being analysed for eDNA has

416 been made publicly available, even when a larger fragment of the gene has been sequenced for the

417 reference database (Cilleros et al., 2019). This study highlights that the expansion of eDNA studies

418 in megadiverse areas depends on better collaborative efforts focused on building more publicly

419 available datasets.

420 The eDNA sampling successfully identified seven orders and 17 families of fish (Figs 3

421 and 4). If we consider the >99.5% threshold for species assignment based on the optimum

422 threshold analyses presented above, only *Phreatobius* spp. would be assigned to species level.

423 When considering a fixed general threshold of >97% for species assignment, only four species

424 could be assigned (Fig. 4; Table S5). Two of these might represent different congeneric species

425 (i.e., genera *Aequidens* and *Sybranchus*), possibly due to the lack of taxonomic resolution of the

426 marker used (Yu et al., 2012). The potential for over and underestimations within our eDNA results

427 is likely. We generated MOTUs using SWARM, aiming for an estimate of the overall biodiversity.

428 While many MOTUs correspond to true biological species, some might be the result of PCR or

429 sequencing errors, unidentified because of an incomplete reference database (Marques et al.,

430 2019). The ambiguous identification values obtained (85-100%) may result in both the

431 overestimation of true diversity (Reeder & Knight, 2009; Morgan et al., 2013) while bioinformatic

432 clustering may have pooled closely related species, underestimating their number within certain

433 taxonomic groups (Huse, Welch, Morrison & Sogin, 2010). This significantly hinders the

434 information that we can draw from the eDNA dataset. Therefore, we used the information up to

435 family level to depict the overall diversity detected in the eDNA data in which ecological

436 inferences can be made. Despite underperforming eDNA accuracy regarding species

437 identification, we could still draw important ecological information from the dataset. The high

438 number of Characiformes MOTUs per site falls in line with previous studies in this region

439 (Birindelli, Britski & Ramirez, 2020; Sales et al., 2021).

440 Species richness analysis of the eDNA and netting dataset within the Ducke Reserve

441 reveals a similar pattern (Fig. 6A). Although sampling using eDNA detected slightly more MOTUs

442 than species detected through the netting survey, it can be assumed that the 32 unique MOTUs

443 obtained from the eDNA data might represent an overestimation of true diversity obtained (as

444 outlined above). Species richness estimates obtained from netting surveys (Fig. 5) are reliable (but

445 not infallible), given that each individual is assigned to a species with high confidence based on

446 its morphological features (except for cryptic species). Nevertheless, while netting surveys

447   produce accurate data, they cannot be applied to the required spatial scale in Neotropical basins.

448   Furthermore, their success and reliability depend on the ability/expertise of the surveyor (labour

449   intensive), the accessibility of the target area (challenging environments), and selectivity based on

450   the deployed techniques' ability to capture target species. Yet, when optimal conditions are met

451   (complete reference database, and appropriate markers to name a few), eDNA will outperform

452   traditional sampling, for its ability to detect species that are missed by the fishing gear deployed.

453   In this study, the order Synbranchiformes, known to occur in the sampled area (Zuanon et al.,

454   2015), was detected through eDNA, but not netting conducted at the same time. This species has

455   been previously identified in the studied area and this highlights the randomness of net sampling,

456   which could be due to the limited spatial and temporal scales covered by the netting survey, or, in

457   this case, even by species behaviour. *Synbranchus* are nocturnal fish and can even survive buried

458   up to three months, hence they can easily elude netting (Prestes-Carneiro & Béarez, 2017).

459   β-diversity analysis for both eDNA and traditional methods yielded contrasting patterns.

460   While eDNA data showed similarities in compositions in the Ducke Reserve streams with slight

461   clustering of sites B1 and B2, netting data suggest major differences between each sampling

462   location (Fig. 6B). Barriers that hinder traditional surveying stemming from species size (affecting

463   susceptibility of capture) and behavioural traits (i.e. shoaling/solitary, fast/slow, exposed/hidden)

464   can be bypassed when analysing biodiversity on a genetic level (Evans, Shirey, Wieringa, Mahon

465   & Lamberti, 2017). When utilising traditional methods, particularly in understudied areas, these

466   barriers can create a bias towards species that are easier to capture, producing species-selective

467   results (Holubová, Čech, Vašek & Peterka, 2019). A barrier that poses a great hindrance to

468   biodiversity monitoring, and that is shared by both methods, is the presence of cryptic species

469   (Beng & Corlett, 2020). Discerning between cryptic species may be difficult when species are

470   morphologically similar and may result in the incorrect identification of a specimen. However,

471   despite MOTUs being used as a proxy for species, and showing a positive correlation when used

472   for taking ecological measures (e.g., β-diversity, Marques et al., 2019; Sales et al., 2021), results

473   from the eDNA dataset should be used with caution as highlighted above. Therefore, prior to using

474   these measures, eDNA metabarcoding and netting surveys merits further investigation in this

475   Reserve.

476   Although eDNA metabarcoding potentially offers a means to assess biodiversity on a

477   larger, more time-efficient scale, ensuring accuracy in results is critical. Evidently, eDNA

478   metabarcoding as a biomonitoring tool in the Neotropical region is in its infancy, highlighted by

479   the lack of appropriate reference databases. Herein, we argue that the commonly adopted threshold

480   of >97% to assign MOTUs to species level is not optimal in megadiverse, understudied regions

481   due to the likelihood of false positive or negative assignments. While this study shows that limited,

482   albeit reliable, ecological inferences of biodiversity can be made with eDNA metabarcoding in

483   this region (based on MOTU richness for example), it also highlights the need for significant

484   investment in research aimed at improving availability of reference databases. We advocate for

485   more transparency and collaboration within the research community, and recommend moving

486   towards building whole mitochondrial genomes of specimens/species to identify multiple/other

487   mini-barcodes and investigate order-specific taxonomic delimitation thresholds for future eDNA

488   metabarcoding surveys of fishes and other vertebrates in the Neotropics (Milan et al., 2020; Sales

489   et al., 2020, 2021).

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

NGS, ADM, CB and IC conceived, designed and acquired funding for the study. CB carried out the eDNA sampling and netting surveys. NGS performed the eDNA-based laboratory work and COC and JSR performed the species barcoding. NGS and JMJ carried out the bioinformatic analyses. JMJ, NGS, CB, IC and ADM analysed the data. JMJ, NGS and ADM wrote the manuscript, with all authors contributing to editing and discussions.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known personal relationships or competing financial interests that could have influenced the work conducted in this study.

REFERENCES

Agostinho A.A., Thomaz S.M., Gomes L.C. (2005). Conservação da biodiversidade em águas continentais do Brasil. Megadiversidade 1, 70–78.

Ahn H., Kume M., Terashima Y., Ye F., Kameyama S., Miya M., Kasai A. (2020). Evaluation of fish biodiversity in estuaries using environmental DNA metabarcoding. PLoS One, 15(10), e0231127.

Albert J.S., Reis R. (2011). Historical biogeography of Neotropical freshwater fishes. University of California Press.

Alho C.J., Reis R.E., Aquino P.P. (2015). Amazonian freshwater habitats experiencing environmental and socioeconomic threats affecting subsistence fisheries. Ambio, 44(5), 412-425.

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

Betancur-R R., Wiley E.O., Arratia G., Acero A., Bailly N., Miya M., Orti G. (2017). Phylogenetic classification of bony fishes. BMC Evolutionary Biology, 17, 162.

Birindelli J.L., Britski H.A., Ramirez J.L. (2020). A new endangered species of Megaleporinus (Characiformes: Anostomidae) from the Rio de Contas basin, eastern Brazil. Journal of Fish Biology, 96(6), 1349-1353.

Boyer F., Mercier C., Bonin A., Le Bras Y., Taberlet P., Coissac E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. Molecular Ecology Resources, 16(1), 176-182.

Cilleros K., Valentini A., Allard L., Dejean T., Etienne R., Grenouillet G., Iribar A., Taberlet P., Vigouroux, R., Brosse S. (2019). Unlocking biodiversity and conservation studies in high-diversity environments using environmental DNA (eDNA): A test with Guianese freshwater fishes. Molecular Ecology Resources, 19(1), 27-46.

Collins R.A., Bakker J., Wangensteen O.S., Soto A.Z., Corrigan L., Sims D.W., Genner M.J., Mariani S. (2019). Non-specific amplification compromises environmental DNA metabarcoding with COI. Methods in Ecology and Evolution, 10(11), 1985-2001.

Deagle B.E., Jarman S.N., Coissac E., Pompanon F., Taberlet P. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. Biology Letters, 10(9), 20140562.

Doble C.J., Hipperson H., Salzburger W., Horsburgh G.J., Mwita C., Murrell D.J., Day J.J. (2020). Testing the performance of environmental DNA metabarcoding for surveying highly diverse tropical fish communities: A case study from Lake Tanganyika. Environmental DNA, 2(1), 24-41.

Evans N.T., Shirey P.D., Wieringa J.G., Mahon A.R., Lamberti G.A. (2017). Comparative cost and effort of fish distribution detection via environmental DNA analysis and electrofishing. Fisheries, 42(2), 90-99.

567   Foster Z.S., Sharpton T.J., Grünwald N.J. (2017) Metacoder: an R package for visualization and
568   manipulation of community taxonomic diversity data. PLoS Computational Biology, 13, e1005404.

569   Frota A., Depra G.D.C., Petenucci L.M., Graça W.J.D. (2016). Inventory of the fish fauna from
570   Ivaí River basin, Paraná State, Brazil. Biota Neotropica, 16(3), e20150151.

571   Hebert P.D., Cywinska A., Ball S.L., Dewaard J.R. (2003). Biological identifications through DNA
572   barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences, 270, 313-321.

573   Hering D., Borja A., Jones J.I., Pont D., Boets P., Bouchez A., Kelly M. (2018). Implementation
574   options for DNA-based identification into ecological status assessment under the European Water
575   Framework Directive. Water Research, 138, 192– 205.

576   Holubová M., Čech M., Vašek M., Peterka J. (2019). On the use of a visual census in surveying
577   fish communities in lentic water bodies. Ecological Indicators, 105, 1-5.

578   Huse S.M., Welch D.M., Morrison H.G., Sogin M.L. (2010). Ironing out the wrinkles in the rare
579   biosphere through improved OTU clustering. Environmental Microbiology, 12(7), 1889–1898.

580   Jézéquel C., Tedesco P.A., Bigorne R., Maldonado-Ocampo J.A., Ortega H., Hidalgo M., Martens
581   K., Torrente-Vilara G., Zuanon J., Acosta A., Agudelo E. (2020). A database of freshwater fish species of
582   the Amazon Basin. Scientific data, 7, 96..

583   Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Thierer T. (2012).
584   Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis
585   of sequence data. Bioinformatics, 28(12), 1647-1649.

586   Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through
587   comparative studies of nucleotide sequences. Journal of Molecular Evolution, 16(2), 111-120.

588   Kubečka J., Godø O.R., Hickley P., Prchalová M., Říha M., Rudstam L., Welcomme R. (2012).
589   Fish sampling with active methods. Fisheries Research, 123-124, 1-3.

590   Lacoursière-Roussel A., Howland K., Normandeau E., Grey E.K., Archambault P., Deiner K.,
591   Lodge D.M., Hernandez C., Leduc N., Bernatchez, L. (2018). eDNA metabarcoding as a new surveillance
592   approach for coastal Arctic biodiversity. Ecology and Evolution, 8(16), 7763-7777.

593   Leroy B., Dias M.S., Giraud E., Hugueny B., Jézéquel C., Leprieur F., Tedesco P.A. (2019). Global
594   biogeographical regions of freshwater fish species. Journal of Biogeography, 46(11), 2407-2419.

595   Li Y., Evans N.T., Renshaw M.A., Jerde C.L., Olds B.P., Shogren A.J., Pfrender M.E. (2018).
596   Estimating fish alpha-and beta-diversity along a small stream with environmental DNA metabarcoding.
597   Metabarcoding and Metagenomics, 2, e24262.

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

601   Magnusson W.E., Lima A.P., Luizão R., Luizão F., Costa F.R., Castilho C.V.D., Kinupp V.F.
602   (2005). RAPELD: a modification of the Gentry method for biodiversity surveys in long-term ecological
603   research sites. Biota Neotropica, 5(2), 19-24.

604   Mahé F., Rognes T., Quince C., de Vargas C., Dunthorn M. (2014). Swarm: robust and fast
605   clustering method for amplicon-based studies. PeerJ, 2, e593.

606   Marques V., Guérin P.É., Rocle M., Valentini A., Manel S., Mouillot D., Dejean T. (2020). Blind
607   assessment of vertebrate taxonomic diversity across spatial scales by clustering environmental DNA
608   metabarcoding sequences. Ecography, 43(12), 1779-1790.

609   McDevitt A.D., Sales N.G., Browett S.S., Sparnenn A.O., Mariani S., Wangensteen O.S.,
610   Benvenuto C. (2019). Environmental DNA metabarcoding as an effective and rapid tool for fish monitoring
611   in canals. Journal of Fish Biology, 95(2), 679-682.

612   McElroy M.E., Dressler T.L., Titcomb G.C., Wilson E.A., Deiner K., DudleyT.L., Lamberti G.A.
613   (2020). Calibrating environmental DNA metabarcoding to conventional surveys for measuring fish species
614   richness. Frontiers in Ecology and Evolution, 8, 276.

615   Melo B.F., Ochoa L.E., Vari R.P., Oliveira C. (2016). Cryptic species in the Neotropical fish genus
616   Curimatopsis (Teleostei, Characiformes). Zoologica Scripta, 45(6), 650-658.

617   Meyer C.P., Paulay G. (2005). DNA barcoding: error rates based on comprehensive sampling.
618   PLoS Biology,. 3(12), e422.

619   Milan D., Mendes I., Damasceno J., Teixeira D., Sales N., Carvalho D. (2020). New 12S
620   metabarcoding primers for enhanced Neotropical freshwater fish biodiversity assessment. Scientific
621   Reports 10, 17966.

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

625   Miya M., Sato Y., Fukunaga T., Sado T., Poulsen J.Y., Sato K., Minamoto T., Yamamoto S.,
626   Yamanaka H., Araki H., Kondoh M. (2015). MiFish, a set of universal PCR primers for metabarcoding
627   environmental DNA from fishes: detection of more than 230 subtropical marine species. Royal Society
628   Open Science, 2(7), 150088.

629   Morgan M.J., Chariton A.A., Hartley D.M., Hardy C.M. (2013). Improved inference of taxonomic
630   richness from environmental DNA. PLoS One, 8(8), e71974.

631   Morris R.J. (2010). Anthropogenic impacts on tropical forest biodiversity: a network structure and
632   ecosystem functioning perspective. Philosophical Transactions of the Royal Society B: Biological Sciences,
633   365(1558), 3709-3718.

634   Murienne J., Cantera I., Cerdan A., Cilleros K., Decotte J.B., Dejean T., Brosse S. (2019). Aquatic
635   eDNA for monitoring French Guiana biodiversity. Biodiversity Data Journal, 7, e37518.

636   Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P.R., O'hara R.B., Oksanen M.J.
637   (2013). Package 'vegan'. Community ecology package, version, 2(9), 1-295.

638   Petruniak J., Bradley D., Kelly J.M., Hanner R.H. (2020). Commentary: integrating environmental
639   DNA into applied ecological practice. Journal of Environmental Studies and Sciences. doi:
640   10.1007/s13412-020-00638-1.

641   Prestes-Carneiro G., Béarez P. (2017). Swamp-eel (Synbranchus spp.) fishing in Amazonia from
642   pre-Columbian to present times. Journal of Ethnobiology, 37(3), 380-397.

643   Pugedo M.L., de Andrade Neto F.R., Pessali T.C., Birindelli J.L.O., Carvalho D.C. (2016).
644   Integrative taxonomy supports new candidate fish species in a poorly studied neotropical region: the
645   Jequitinhonha River Basin. Genetica, 144, 341-349.

646   Reeder J., Knight R. (2009). The 'rare biosphere': A reality check. Nature Methods, 6(9), 636–637.

647   Sales N.G., Kaizer M.D.C., Coscia I., Perkins J.C., Highlands A., Boubli J.P., Magnusson W.E.,
648   Da Silva M.N.F., Benvenuto C., McDevitt A.D. (2020). Assessing the potential of environmental DNA
649   metabarcoding for monitoring Neotropical mammals: a case study in the Amazon and Atlantic Forest,
650   Brazil. Mammal Review, 50(3), 221-225.

651   Sales N.G., Wangensteen O.S., Carvalho D.C., Mariani S. (2019). Influence of preservation
652   methods, sample medium and sampling time on eDNA recovery in a neotropical river. Environmental
653   DNA, 1(2), 119-130.

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

657   Sales N.G., Mariani S., Salvador G.N., Pessali T.C., Carvalho D.C. (2018). Hidden diversity
658   hampers conservation efforts in a highly impacted Neotropical river system. Frontiers in Genetics, 9, 271.

659   Sassoubre L.M., Yamahara K.M., Gardner L.D., Block B.A., Boehm A.B. (2016). Quantification
660   of environmental DNA (eDNA) shedding and decay rates for three marine fish. Environmental Science &
661   Technology, 50(19), 10456-10464.

662   Senapati D., Bhattacharya M., Kar A., Chini D.S., Das B.K., Patra B.C. (2018). Environmental
663   DNA (eDNA): A promising biological survey tool for aquatic species detection. Proceedings of the
664   Zoological Society, 72, 211– 228.

665   Servis J.A., Reid B.N., Timmers M.A., Stergioula V., Naro-Maciel E. (2020). Characterizing coral
666   reef biodiversity: genetic species delimitation in brachyuran crabs of Palmyra Atoll, Central Pacific.
667   Mitochondrial DNA Part A, 31(5), 178-189.

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

670   Valdez-Moreno M., Ivanova N.V., Elías-Gutiérrez M., Pedersen S.L., Bessonov K., Hebert P.D.
671   (2019). Using eDNA to biomonitor the fish community in a tropical oligotrophic lake. PLoS One, 14(4),
672   e0215505.

673   Weigand H., Beermann A.J., Čiampor F., Costa F.O., Csabai Z., Duarte S., Strand M. (2019). DNA
674   barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and
675   recommendations for future work. Science of the Total Environment, 678, 499-524.

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

679   Zhang S., Lu Q., Wang Y., Wang X., Zhao J., Yao M. (2020). Assessment of fish communities
680   using environmental DNA: Effect of spatial sampling design in lentic systems of different sizes. Molecular
681   Ecology Resources, 20(1), 242-255.

682   Zuanon J., Medonça F.P., Espírito Santo H., Dias M.S., Galuch A.V., Akama A. (2015). Guia de
683   peixes da Reserva Adolpho Ducke.

684

685

686

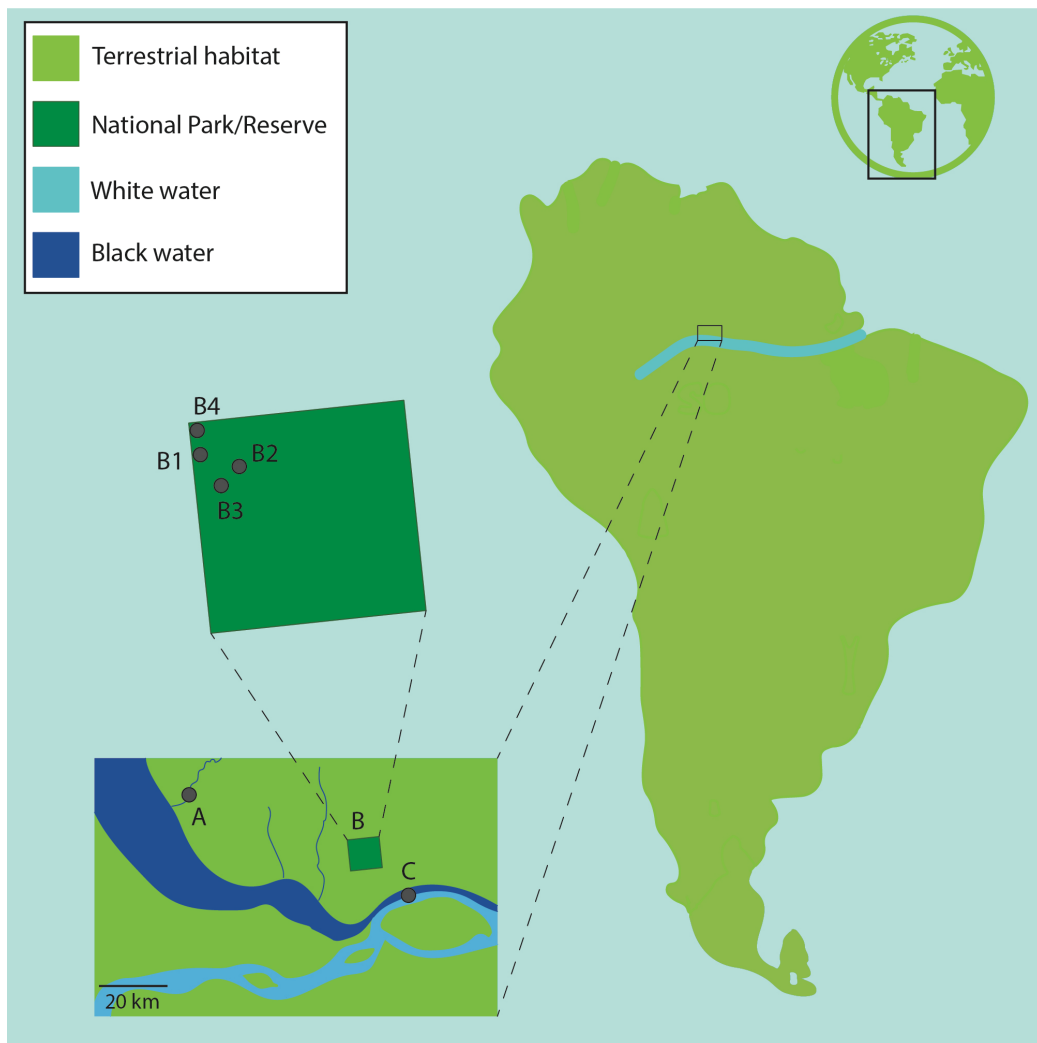687

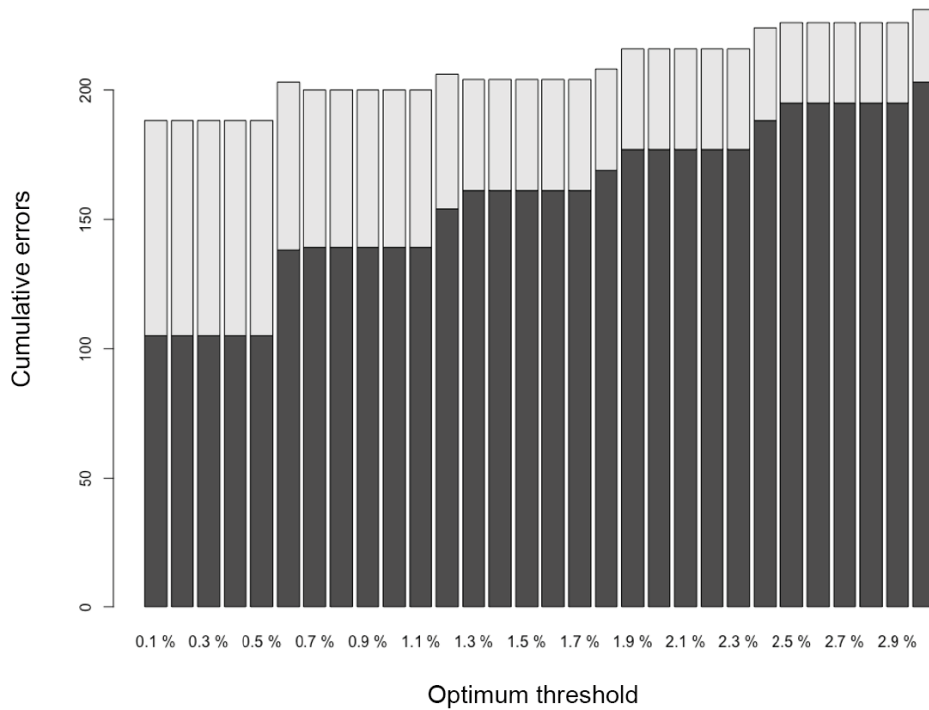688

689

690

691

692

693

694

695

696

Figures
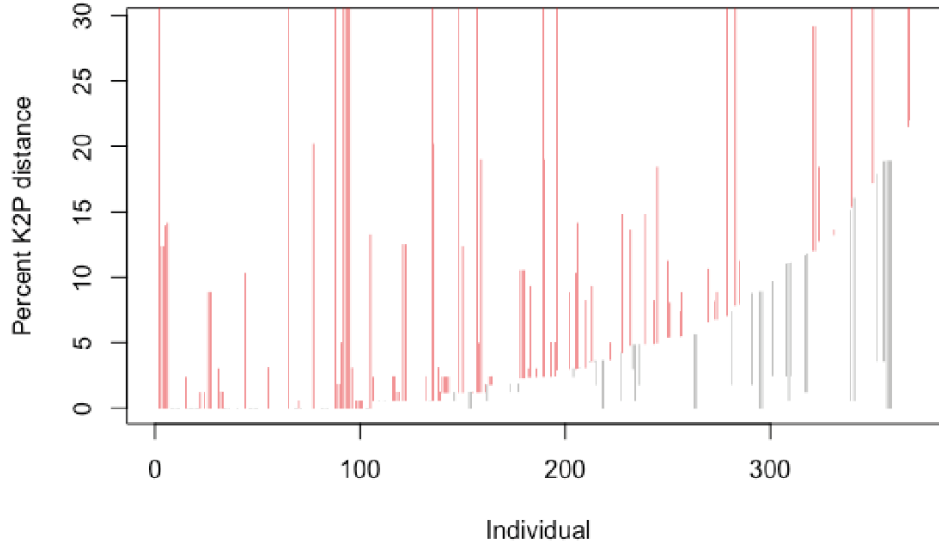

Figure 1. Sampling sites located in the Adolpho Ducke Forest Reserve (B1-B4), and the two additional sites located outside the Reserve (A: Aturiá stream, and C: confluence between the Solimões and Negro rivers - *"the meeting of the waters"*).

Figure 2. Optimum thresholds (A) and barcoding gap (B) analyses. Barplot (A) shows the false positive (light grey) and false negative (dark grey) rate of identification of species within the curated reference database at thresholds ranging from 0.1% to 3%. Line plot (B) of the barcode gap analyses for the 264 species (373 sequences) within the curated reference database. For each

709   individual in the dataset, the grey lines represent the furthest intraspecific distance (bottom of line

710   value), and the closest interspecific distance (top of line value) (i.e., presence of a barcode gap).

711   The red lines show where this relationship is reversed (i.e., no barcode gap).

712
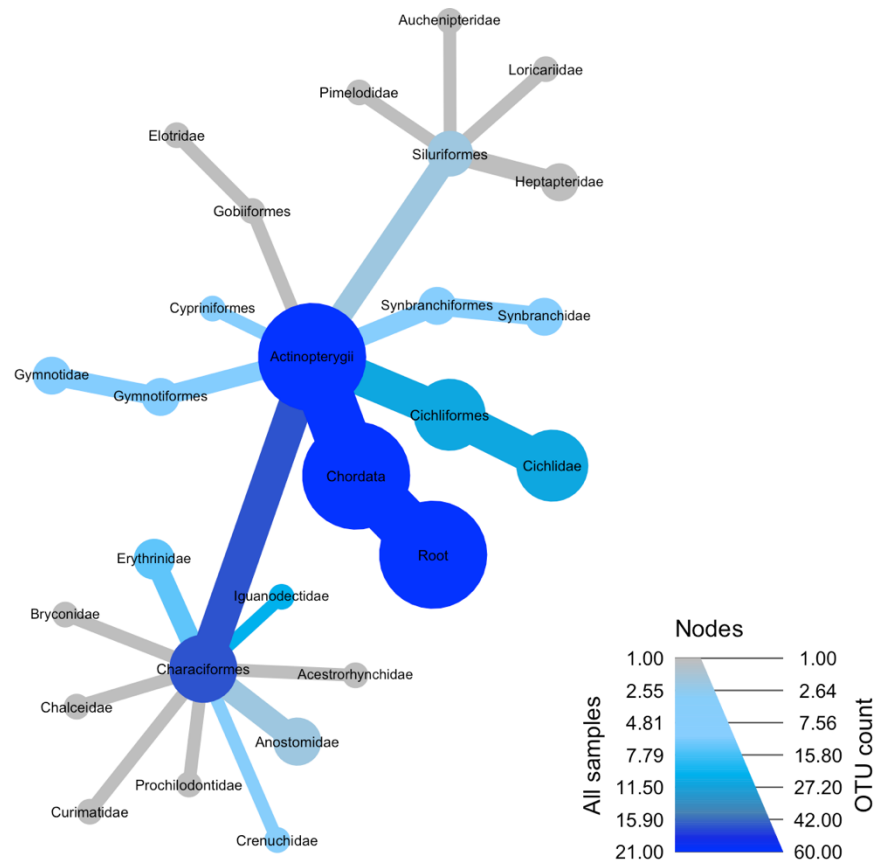
713

714 **Figure 3**. Heat tree displaying the fish diversity recovered for all sampling locations using eDNA

715 metabarcoding. The blue colouration represents diversity identified from water samples, the darker

716 the shade of blue, the more MOTUs detected for that taxonomic order.
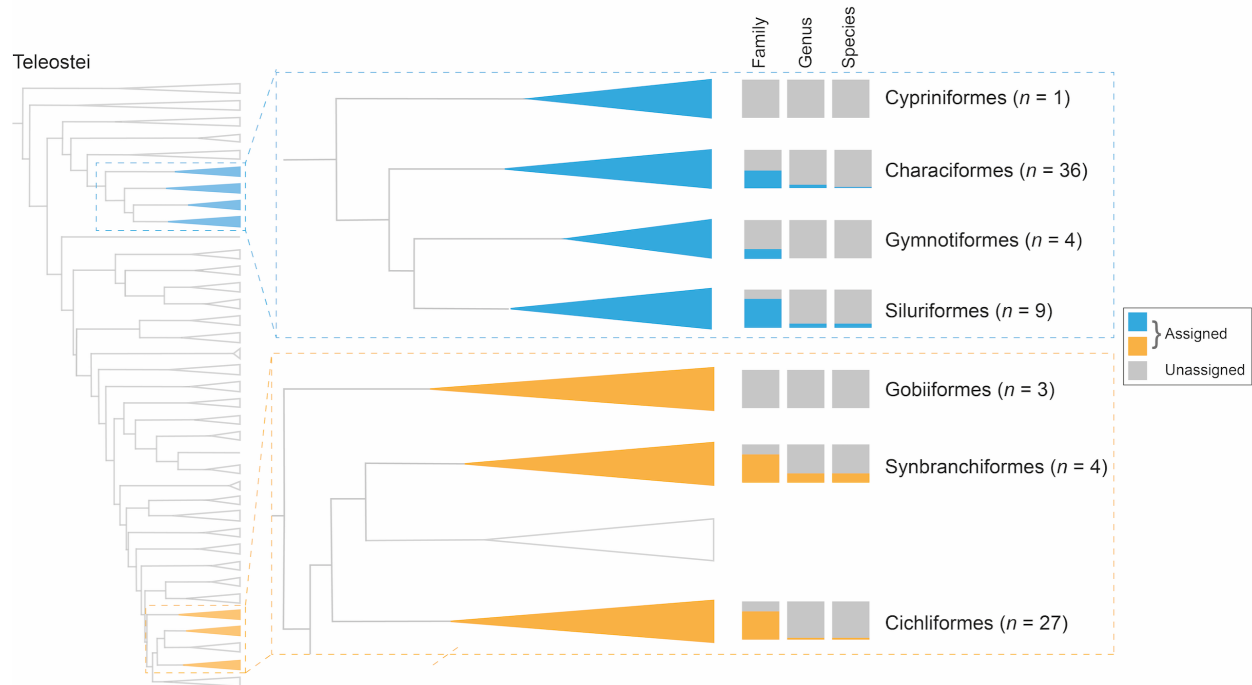
717

718

719

720

721

722

723

724

725

726

727

Figure 4. Represented orders within Teleostei (tree adapted from Betancur-R et al. 2017) to which the number of identified Molecular Operational Taxonomic Units (MOTUs) obtained from environmental DNA can be proportionally assigned to family, genus and species within each order.
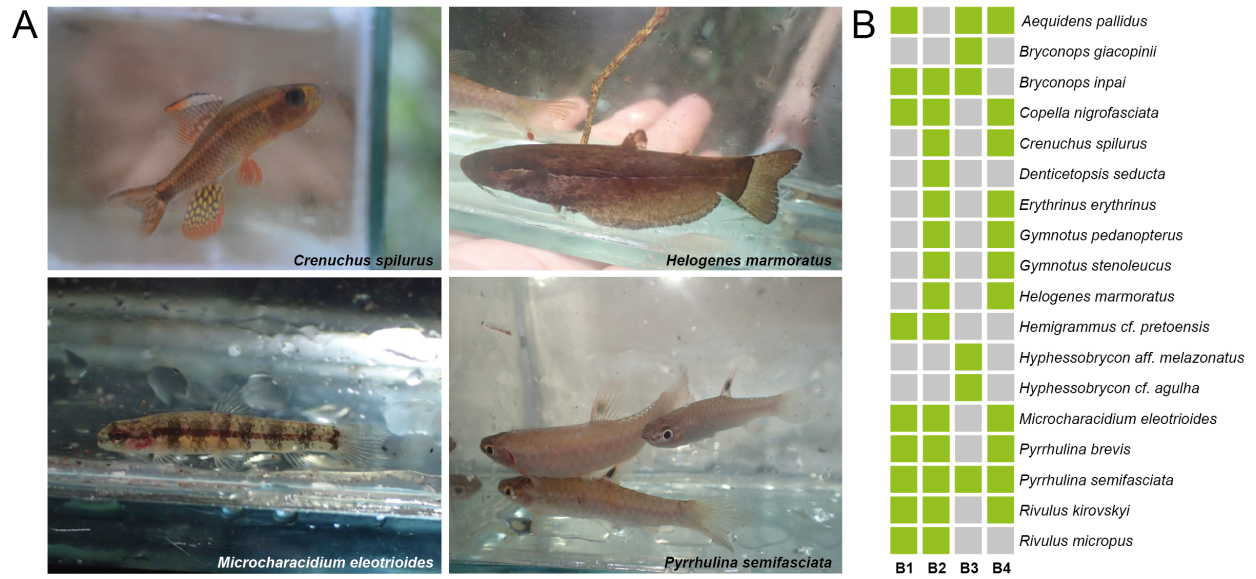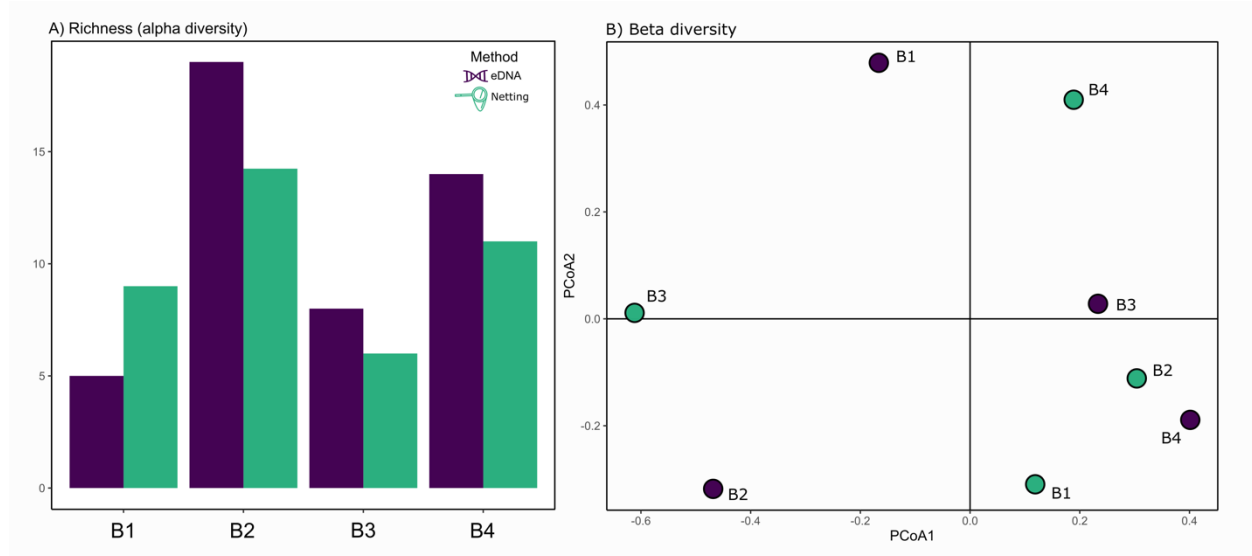
Figure 5. Examples of four species caught by netting in the Ducke Reserve (A), and presence (green) and absence (grey) of each of the 18 species caught in the four streams (B1-B4) within the Reserve (B).

Figure 6. Species richness (A) and β-diversity inferred from separate PCoAs (B) based on eDNA MOTU data (purple) and netting data (green) from locations in the Ducke Reserve. See Fig. 1 for sampling locations.