UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciências Biológicas

Departamento de Genética, Ecologia e Evolução

Programa de Pós-graduação em Genética

Izabela Santos Mendes

DNA AMBIENTAL *METABARCODING* APLICADO AO MANEJO E CONSERVAÇÃO DE PEIXES

Belo Horizonte

2023

Izabela Santos Mendes

DNA AMBIENTAL *METABARCODING* APLICADO AO MANEJO E CONSERVAÇÃO DE PEIXES

Tese apresentada ao Programa de pós-graduação em Genética da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção de título de Doutora em Genética

Orientador: Dr. Daniel Cardoso de Carvalho
Coorientadora: Dra. Naiara Guimarães Sales

Belo Horizonte
2023

043 Mendes, Izabela Santos.
DNA ambiental metabarcoding aplicado ao manejo e conservação de peixes [manuscrito] / Izabela Santos Mendes. – 2023.
150 f. : il. ; 29,5 cm.

Orientador: Dr. Daniel Cardoso de Carvalho. Coorientadora: Dra. Naiara Guimarães Sales.
Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa de Pós-Graduação em Genética.

1. Genética. 2. Peixes. 3. Biodiversidade. 4. Monitoramento. 5. DNA Ambiental. I. Carvalho, Daniel Cardoso de. II. Sales, Naiara Guimarães. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 575

Ficha catalográfica elaborada pela Bibliotecária Fabiane C M Reis. CRB/6 - 2680

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa de Pós-Graduação em Genética

ATA DE DEFESA DE TESE

ATA DA DEFESA DE TESE	**173/2023** **entrada** **1º/2018**
Izabela Santos Mendes	**CPF: 113.371.276-20**

Às oito horas do dia **31 de março de 2023**, reuniu-se a Comissão Examinadora de Tese, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: **"DNA ambiental metabarcoding aplicado ao manejo e conservação de peixes"**, requisito para obtenção do grau de Doutora em **Genética.** Abrindo a sessão, o Presidente da Comissão, **Daniel Cardoso de Carvalho**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa da candidata. Logo após, a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Prof./Pesq.	Instituição	CPF	Indicação
Daniel Cardoso de Carvalho	PUC	034.375.066-09	APROVADA
Fernanda Antunes Carvalho	UFMG	045.665.426-70	APROVADA
Carla Martins Lopes	USP	299.322.458-10	APROVADA
André Batista Nobile	UNESP	305.751.978-46	APROVADA
Jonathan Stuart Ready	UFPA	745.826.071-87	APROVADA
Naiara Guimarães Sales	Universidade de Salford	099.679.846-35	APROVADA

Pelas indicações, a candidata foi considerada: APROVADA

O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 31 de março de 2023.

Daniel Cardoso de Carvalho

Fernanda Antunes Carvalho

Carla Martins Lopes

André Batista Nobile

Jonathan Stuart Ready

Naiara Guimarães Sales

Documento assinado eletronicamente por **Carla Martins Lopes**, **Usuária Externa**, em 31/03/2023, às 11:35, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **André Batista Nobile**, **Usuário Externo**, em 31/03/2023, às 11:52, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Jonathan Stuart Ready**, **Usuário Externo**, em 03/04/2023, às 13:09, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Daniel Cardoso de Carvalho**, **Usuário Externo**, em 05/04/2023, às 10:38, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Naiara Guimaraes Sales**, **Usuário Externo**, em 06/04/2023, às 05:48, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Fernanda Antunes Carvalho**, **Professora do Magistério Superior**, em 06/04/2023, às 10:10, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2189672** e o código CRC **58CA7C52**.

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Biológicas
Programa de Pós-Graduação em Genética

FOLHA DE APROVAÇÃO

"DNA ambiental metabarcoding aplicado ao manejo e conservação de peixes"

Izabela Santos Mendes

Tese aprovada pela banca examinadora constituída pelos Professores:

Daniel Cardoso de Carvalho
PUC

Fernanda Antunes Carvalho
UFMG

Carla Martins Lopes
USP

André Batista Nobile
UNESP

Jonathan Stuart Ready
UFPA

Naiara Guimarães Sales

Universidade de Salford

Belo Horizonte, 31 de março de 2023.

Documento assinado eletronicamente por **Carla Martins Lopes**, **Usuária Externa**, em 31/03/2023, às 11:35, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **André Batista Nobile**, **Usuário Externo**, em 31/03/2023, às 11:53, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Jonathan Stuart Ready**, **Usuário Externo**, em 03/04/2023, às 13:09, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Daniel Cardoso de Carvalho**, **Usuário Externo**, em 05/04/2023, às 10:38, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Naiara Guimaraes Sales**, **Usuário Externo**, em 06/04/2023, às 05:48, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Fernanda Antunes Carvalho**, **Professora do Magistério Superior**, em 06/04/2023, às 10:10, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2189737** e o código CRC **05C0BBDC**.

Referência: Processo nº 23072.218506/2023-20 SEI nº 2189737

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), que financiou as bolsas de estudos no país e no exterior, através do Programa de Internacionalização CAPES PrInt Edital 41/2017.

Agradeço aos membros da banca pela participação e pelas contribuições, sugestões e comentários.

Agradeço a CEMIG pelo recurso investido no Projeto de Pesquisa e Desenvolvimento P&D GT 0635. E para além do recurso, agradeço o que o projeto me proporcionou, novas conexões profissionais e pessoais. Nesse caso, agradeço a todos os profissionais envolvidos nesse projeto direta ou indiretamente, em especial a Dra. Raquel Loures pela maestria em gerenciar um projeto complexo e inovador e à Andreza, pelo apoio, disponibilidade, por tornar todas as burocracias mais leves e ter se tornado uma amiga.

Agradeço aos meus orientadores e aqueles que me guiaram profissionalmente durante esse processo. Ao prof. Daniel Carvalho, pela orientação ao longo dos anos. A minha coorientadora e grande inspiração, prof. Naiara Sales, sem ela, eu não teria topado fazer um Doutorado e sem dúvidas, ela foi uma das pessoas que me ajudou a seguir e chegar até aqui. Ao prof. Allan McDevitt, meu supervisor na Inglaterra, por ter me recebido tão bem e por sua inspiradora vontade de fazer acontecer, de forma humana, mas muito firme e competente. E por último, mas não menos importante, ao prof. Paulo Pompeu, que apesar de não constar nos papéis, me orientou durante esses anos, uma orientação na vida de Doutorando, que vai muito além da escrita da tese.

Agradeço ao LGC (Laboratório de Genética da Conservação), que por muitos anos considerei minha primeira e segunda casas. Passei os melhores e não tão bons momentos da minha vida por lá. Foram 10 anos de uma experiência desafiadora, mas engrandecedora, que carregarei comigo para sempre. Agradeço a cada um que tive o prazer de dividir um pouco da vida de laboratório e partilhar as angústias e conquistas. Em especial, ao Júnio, David, Alyce, Bruna, Amanda e Gabriel.

Agradeço ao McCostia Lab por ter me recebido tão bem e ter me proporcionado uma experiência muito além das minhas expectativas. Em especial, a Holly, por ter me ajudado tanto e ter pegado na minha mão e me guiado por um mundo que ainda era quase que desconhecido por mim, o da Bioinformática.

Agradeço ao Laboratório de Ecologia de Peixes da UFLA, coordenado pelo prof. Paulo, por serem receptivos, me ensinarem tanto ao longo desses anos e pela parceria. Em especial, não poderia deixar de agradecer ao Chico, por me ensinar dentro e fora de campo. Está pra nascer alguém com mais paciência em ensinar, como ele. Ao Ivo, por segurar as pontas e ser o herói e conseguir resolver os maiores pepinos rsrs. Agradeço a paciência, carinho e a amizade que plantamos ao longo desses anos. Agradeço a Tatau, por deixar o processo mais leve e por enlouquecermos juntas com tantos documentos e *deadlines* rsrs.

Ao GPEPAD, coordenado pelo prof. Gilmar Bastos, por toda colaboração e compreensão ao longo dos anos. Criamos fortes laços que foram reforçados ainda mais em um dos momentos mais desafiadores, que foi a pandemia. Nos apoiamos profissionalmente e pessoalmente. São conexões que levo para minha vida, certamente.

Aos meus mestres, que ao longo dos anos me ensinaram grande parte do que eu sei hoje.

Aos professores do PPG Genética, em especial, aos profs. Anderson Myoshi e Frederico Soriani pelo apoio, principalmente nessa reta final, que foi fundamental. Aos funcionários dos PPG Genética e Biologia de Vertebrados da UFMG e PUC Minas, por tudo que fizeram ao longo desses anos. Meu agradecimento especial a Cledma e ao Rogério, por não medirem esforços para ajudar e apoiar os alunos. E à Raíssa, sem ela, definitivamente eu não teria ido ao exterior e ter tido a experiência que eu tive. Ela “lutou” bravamente e conseguiu com que fosse possível a minha saída do país.

A minha família, não tenho palavras que sejam capazes de expressarem minha gratidão e amor. Aos meus pais e a minha boadrasta por serem meu porto seguro. Não há sentimento melhor do que saber que você tem um colo para recorrer e eu tenho os melhores. Aos meus irmãos, Victor e Laura, as duas pessoas que eu me viro e reviro para vê-los felizes e a quem faço questão de ser sempre melhor e servir como inspiração, como uma boa irmã mais velha. Agradeço aos meus familiares por todo o apoio durante esse processo.

Aos meus amigos, que durante esse processo foram essenciais, sendo apoio nas horas desafiadoras e vibrando com as minhas conquistas. Em especial, às filhas da PUC, que me proporcionam as melhores emoções e aventuras, aos meus amigos de infância, do Fuentes, e ao Mohammed, amigo que fiz durante a minha estadia na Inglaterra e que dividiu os melhores e piores momentos que tive por lá.

Ao Daniel Teixeira, pelo companheirismo, parceria e paciência comigo. Sou uma pessoa e uma pesquisadora melhor graças a você. Te admiro como cientista, pesquisador, parceiro e companheiro. Que nossa parceria seja leve e perdure por longos e felizes anos. À família do Daniel, agradeço o amor e carinho, o incentivo e por se tornarem minha família também, minha gratidão.

E por último, mas não menos importante, ao Jequitinhonha e toda a sua comunidade, que viabilizou a realização desse projeto de diversas maneiras. Foram as pessoas mais acolhedoras e carinhosas que eu tive o prazer de conhecer e sem dúvida, são a minha maior saudade. Criei laços e tive experiências incríveis. O Jequi foi calmaria para mim quando eu mais precisei e o maior combustível para reafirmar o que faço de melhor. Minha maior gratidão. Em especial, ao Neguim por ser um ser humano incrível e uma das mentes mais sábias sobre os peixes do Jequi e também a sua família que tive o prazer de conhecer. A Dona Edna por proporcionar conforto e carinho de mãe quando estávamos fora de casa. E ao Salim e Mangela, pela melhor comida do mundo e pelo carinho e acolhimento a cada campanha. Lélis, como chamo Lelivéldia, é pequena em tamanho, mas tem um dos maiores espaços no meu coração e na minha história. E claro, a todos aqueles que tive o prazer de conhecer. Eternamente grata.

“Aprendi através da experiência amarga a suprema lição: controlar minha ira e torná-la como o calor que é convertido em energia. Nossa ira controlada pode ser convertida numa força capaz de mover o mundo”.

Mahatma Gandhi

RESUMO

A região Neotropical compreende uma das maiores diversidades de peixes de água doce do mundo. Entretanto, essa biodiversidade enfrenta níveis sem precedentes de impactos antrópicos. Dessa forma, a avaliação da biodiversidade é essencial para a definição das mais adequadas estratégias de manejo e conservação. Essa avaliação depende da detecção confiável e da identificação precisa das espécies; assim, métodos adicionais, associados à identificação taxonômica tradicional, estão sendo cada vez mais implementados em todo o mundo, como é o caso do DNA ambiental *metabarcoding*. Entretanto, apesar do aumento exponencial de publicações de eDNA *metabarcoding*, a maioria dos estudos tem sido conduzido em regiões temperadas e em áreas razoavelmente acessíveis e poucos estudos são conduzidos em regiões megadiversas, como a região Neotropical, especialmente em regiões afetadas por represamentos, um dos maiores estressores para a comunidade de peixes de água doce. Atualmente, o entendimento de como o eDNA *metabarcoding* poderá ser um método bem estabelecido para avaliação da biodiversidade e monitoramento ecológico em reservatórios é relativamente limitado e aspectos metodológicos como a falta de conhecimento da biodiversidade local e ausência de sequências no banco de dados referência, vieses de *primer* e as diferentes formas de identificação taxonômica (MOTUs x ASVs) precisam ser melhor avaliados e compreendidos antes de aplicar essa técnica na identificação da comunidade de peixes em larga escala. Nessa tese, primeiramente, desenvolvemos um conjunto de *primers mini-barcodes*, baseado em um banco de dados referência customizado. Posteriormente, validamos esse conjunto de marcadores moleculares em uma comunidade simulada de espécies de peixes neotropicais e por fim, aplicamos a metodologia do eDNA *metabarcoding* em um reservatório Neotropical, comparando os resultados obtidos aos dados de redes de emalhar, representando os métodos tradicionais de pesca. Coletivamente, os resultados dessa tese evidenciam a importância de um banco de dados referência customizado para a identificação correta e confiável do DNA ambiental. Além disso, nossos resultados demonstraram uma diferença significativa na comunidade de peixes, demonstrando o papel complementar do eDNA *metabarcoding* no monitoramento de peixes.

Palavras-chave: eDNA, inventariamento, monitoramento, 12S, ictiofauna, região neotropical.

ABSTRACT

The Neotropical region comprises one of the greatest diversity of freshwater fish in the world. However, this biodiversity faces unprecedented levels of anthropic impacts. Thus, the assessment of biodiversity is pivotal for defining the most appropriate management and conservation strategies. This assessment depends on reliable detection and accurate identification of species, thus, additional methods, associated with traditional taxonomic identification, are being increasingly implemented worldwide, as environmental DNA metabarcoding. However, despite the exponential increase in eDNA metabarcoding publications, most studies have been conducted in temperate regions and in reasonably accessible areas, and few studies are conducted in megadiverse regions, such as the Neotropics, especially in regions affected by damming, one of the major stressors for the freshwater fish community. Currently, the understanding of how eDNA metabarcoding can be a well-established method for biodiversity assessment and ecological monitoring in reservoirs is relatively limited and methodological aspects such as lack of knowledge of local biodiversity and absence of sequences in the reference database, biases of primer and the different forms of taxonomic identification (MOTUs x ASVs) need to be better evaluated and understood before applying this technique in large-scale fish community identification. In this thesis, we developed a set of mini-barcodes primers, based on a customized reference database. Moreover, we validated this set of molecular markers in different mock communities and finally, we applied the eDNA metabarcoding methodology in a Neotropical reservoir, comparing the results obtained with data from gillnets, representing traditional fishing methods. Collectively, the results of this thesis highlight the importance of a customized reference database for the correct and reliable identification of species. Furthermore, our results demonstrated the complementary role of eDNA metabarcoding in fish monitoring.

Keywords: eDNA, inventory, monitoring, 12S, ichthyofauna, neotropical region.

LISTA DE FIGURAS

CAPÍTULO I

Figure 1. (**a**) DNA sequence alignment showing the annealing positions (< < < < < < < <) for NeoFish_3 forward and reverse primers, V05F_898 (used for reference database construction), and for two others previously developed 12S markers (MiFishU and Teleo1). Base pairs highlighted in bold represent positions where the full-length database starts (317 bp) and ends (851 bp) (**b**) Mismatches and primer sites from the full 12S alignment (565 bp) considering 132 sequences of 67 fish species. Mismatches equal to 10 represent gap sites.

Figure 2. Screening for the best target region for mini-barcode across the 12S full alignment (565 bp) of 67 species using the sliding window analyses (SWAN) from SPIDER package. We analyzed (1) high mean K2P distance; (2) few zero pairwise non-conspecific distances; (3) high proportion of clades shared between the Neighbor-joining tree from the 12S full-length barcode and the tree constructed using only data from selected windows; and (4) high sum of diagnostic nucleotides. Rectangles represent the ideal region for primer design based on best indices of all criteria.

Figure 3. Bayesian phylogenetic ultrametric trees for 12S (**a**) full barcode and (**b**) mini-barcode for all species analyzed. MOTUs are represented by different sizes in accordance with each different species delimitation methods (I—GMYC, II—bPTP, III—ABGD, and IV—interspecific genetic distances thresholds). Green, blue and red colors represent MOTUs with a single species, multiple species, and different MOTUs for the same species, respectively.

Figure 4. Threshold optimization for species delimitation for 12S (**a**) full length and (**b**) mini-barcode, showing the false positive (light grey) and false negative (dark grey) rate of fish species identification as pre-set thresholds change. Cumulative error is the sum of false positives + false negatives. For the full length, the percentages with lowest cumulative error (six) are between 0.4% and 0.55%. For the mini-barcode, the lowest cumulative error (also six) is within 0.55% and 1%.

Figure 5. Barcode gap line plot for the 132 fish specimens. For each specimen in the dataset, the grey bars represent the gap between the highest intraspecific distance (bottom of the bar) and the lowest interspecific distance (top of the bar) representing the barcoding gap range. The red bar represents a specimen which these values overlap (intraspecific is higher than interspecific) meaning there is no barcoding gap.

Figure S1. Neighbor-Joining tree built with 132 samples of the 12S region (a) full length (565 bp) and (b) mini-barcode (193 bp) from 67 species.

Figure S2. Environmental DNA amplification, sequence alignment, and Sanger DNA sequencing chromatogram of a sample obtained from an aquarium containing multiple specimens of the cichlid *G. brasiliensis*. Positive eDNA amplifications obtained using NeoFish_3 primer set with eDNA dilutions 1, 1:10, and 1:20 are depicted in the agarose electrophoresis gel. The recovered eDNA sequence was compared with a reference sequence from Genbank (KU531434.1) and with the 12S sequence obtained from the cichlid tissue (*G_brasiliensis*).

CAPÍTULO II

Figure 1. Study framework (A) and description of analysed mock communities including different species composition and DNA input, markers evaluated, and proxies used for biodiversity assessment (B). All mock communities built using species from the São Francisco River Basins (SFRB) and Jequitinhonha River Basin (JQRB). (1) normalised São Francisco River mock community (SFmc) comprising 23 species from the SFRB with the same concentration of DNA (10 ng/µl) and (2) SFmc skewed using different DNA concentrations. (3) São Francisco and Jequitinhonha combined mock community (SFJQmc) built with 38 unique species from the normalised Jequitinhonha River mock community (JQmc) JQmc and SFmc. (4) normalised Jequitinhonha River mock community (JQmc) composed of 23 species from the JQRB built using equal concentrations of DNA. (5) JQmc skewed: mock community composed of 23 species from the JQRB built with skewed concentrations of DNA.

Figure 2. Numbers of ASVs, MOTUs and assigned species for all mock communities and markers. Bars represent the values for each category. Markers are identified by colours: NeoFish - green, MiFish - blue, Teleo - red. The dotted lines highlight the number of species included in each mock community. Number of species DNA added in each mock community: SFJQmc= 38; SFmc= 23; JQmc=17

Figure 3. Fold variation between input DNA and relative read abundance (RRA) for SFmc, JQmc & SFJQmc species. Bars represent the range of the fold change between the input DNA and RRA for each species and marker. Undetected species are marked by *. Light red spaces indicate species absent in the mock communities amplified with Teleo.

Figure 4. Correlations between relative read abundance (RRA) and correspondent input DNA added in each mock community. Each dot represents one species in a mock community. Markers are identified by colours: NeoFish - green, MiFish - blue, Teleo - red.

Figure S1. The LGC 12Sdb.

CAPÍTULO III

Figure 1. Map of the IRR displaying sampling sites of both eDNA (+) and fishing gill nets (o). Yellow circles distributed in the red area demonstrating the sampled areas. The green bar represents the dam location.

Figure 2. Accumulation curves for eDNA metabarcoding considering the number of MOTUs (A) and the number of species (B) for the number of samples.

Figure 3. Stack barplot representing the number of sequences reads relative to the frequency of A) orders and B) families by eDNA samples at the three sampling locations through the eight campaigns.

Figure 4. Stack barplot representing the number of sequences reads relative to the frequency of species from A) Characiformes, B) Serrasalmidae, C)

Cyprinodontiformes, and D) Cichliformes by eDNA samples at the three sampling locations through the eight campaigns.

Figure 5. MOTU richness (α-diversity) distributed along the sampling sites (IR10 in yellow, IR2 in blue and IR3 in red) and sampling events.

Figure 6. MOTU richness (α-diversity) and Principal Coordinates Analysis (PCoA) of β-diversity of sampling locations (Jaccard distance).

Figure 7. Principal Coordinates Analysis (PCoA) of seasons and years (Jaccard distance).

Figure 8. Principal Coordinates Analysis (PCoA) of months and sampling sites (Jaccard distance).

Figure 9. Species accumulation curves for gillnets (red line), eDNA metabarcoding (green line) and both methods together (black line). Grey shading represents the standard deviation of the expected number of species per sampling method.

Figure 10. Taxa and species richness along the IRR for eDNA metabarcoding (A) and gill nets (B).

Figure 11. Venn diagram (blue for eDNA and grey for gillnets) and PCoAs of methods (Jaccard distance).

Figure 12. Principal Coordinates Analysis (PCoA) of years (Jaccard distance) considering both methods.

Figure S1. Principal Coordinates Analysis (PCoA) of β-diversity of sampling locations (Jaccard distance) for eDNA metabarcoding (A) and gill nets (B).

Figure S2. Phyogenetic tree for Prochilodus spp. and Hoplias spp.

LISTA DE TABELAS

CAPÍTULO I

Table 1. Primers designed for short amplifications of the 12S sequences obtained from the major Neotropical fish orders.

Table 2. Primer-BLAST results using the NeoFish_3 primer set and previously developed 12S markers (MiFishU—Miya et al., 2015; and Teleo1—Valentini et al., 2016). Numbers correspond to hits recovered for each taxonomic group.

Table S1. List of all species used to build the 12S reference database and the number of specimens for each species.

Table S2. Species used with each mini-barcode primer set for *in vitro* validation.

CAPÍTULO II

Table S1. List of species of the Conservation Genetics Lab 12S Sequence Database (LGC12Sdb)

Table S2. Species summary used in the mock communities, the originary basin, the LGCdb ID, the input DNA (ng), the total DNA in skewed pool (ng), the percentage on respective skewed pool, the percentage on respective normalized pool, percentage on the SFJQ pool and the presence on the pool (green for presence and red for absence)

Table S3. Primers used for the rRNA 12S gene amplification

Table S4. Primers used in the second PCR of the library's preparation

Table S5A. Mock community containing 40 species from both river basins combined, using equal input DNA concentrations (normalized)

Table S5B. Mock community containing 23 species from São Francisco River basin, using both normalized and skewed input DNA concentrations

Table S5C. Mock community containing 17 species from Jequitinhonha River basin, using both normalized and skewed input DNA concentrations

Table S6. Read counts along all pipeline steps

Table S7. Summary of the Pearson's correlation coefficient between RRA and input DNA concentration for each mock community

CAPÍTULO III

Table 1. Samples collected by campaign, season, sampling site and the number of samples collected at each site.

Table 2. Species distribution in the IRR by eDNA, according to each campaign conducted during three years.

Table S1. List of samples including primers and tags used.

Table S2. Count of reads after each filtering step.

Table S3. List of all MOTUs recovered for all treatments.

Table S4. Species detected by eDNA metabarcoding, its relative read abundance per sampling campaign.

Table S5. Summary of PERMANOVA results (R_2-effect sizes and significance level) showing the effect of sampling season, year, month and site on MOTU diversity recovery.

Table S6. Species detected by eDNA metabarcoding and gill nets and its status.

Table S7. Species identified applying the minimum identity of 0.97.

Table S8. Species richness by eDNA metabarcoding recovered at each sampling site during the sampling campaigns.

SUMÁRIO

1. INTRODUÇÃO GERAL 21

2. OBJETIVO GERAL 26

3. JUSTIFICATIVA 26

4. CAPÍTULO I - New 12S metabarcoding primers for enhanced Neotropical freshwater fish biodiversity assessment (Scientifis Reposts, 2020) 28

4.1 Objetivos específicos 28

5. CAPÍTULO II - DNA metabarcoding of mock communities highlights potential biases when assessing Neotropical fish diversity (submitted on Molecular Ecology Resources, 2023) 41

5.1 Objetivos específicos 41

5.2 INTRODUCTION 45

5.3 MATERIAL AND METHODS 48
5.3.1 Building a local 12S reference database 48
5.3.2 Assemblage of mock communities 49
5.3.4 Bioinformatic analyses 50

5.4 RESULTS 52
5.4.1 12S reference database 52
5.4.2 Data quality control and processing 53
5.4.3 Fold variation between input DNA and RRA 55
5.4.4 Correlation between input DNA and RRA 57

5.5 DISCUSSION 58

REFERENCES 62

Supplementary material 69

6. CAPÍTULO III - ENVIRONMENTAL DNA METABARCODING OF FISH COMMUNITIES IN A NEOTROPICAL DAM RESERVOIR: A comparison of eDNA metabarcoding to traditional survey methods of fish from a Neotropical dam 80

6.1 Objetivos específicos 80

6.2 INTRODUCTION 83

6.3 MATERIALS AND METHODS 86
6.3.1 Study Area 86
6.3.2 Traditional sampling methods 87
6.3.3 Local reference database 87
6.3.4 eDNA sampling and processing 88
6.3.5 Bioinformatic analyses 90
6.3.6 Statistical analyses 91

6.4 RESULTS 93
6.4.1 Library quality and raw data 93
6.4.2 Taxonomic assignment 93
6.4.3 Biodiversity patterns (Species richness and β-diversity) 102
6.4.4 Comparison with established fishing methods 105

6.5 DISCUSSION 110
6.5.1 Taxonomic assignment by eDNA metabarcoding 110
6.5.2 Comparison between eDNA metabarcoding and gill nets 113

REFERENCES 117

Supplementary material 126

7. DISCUSSÃO GERAL 140

8. CONSIDERAÇÕES FINAIS 143

REFERÊNCIAS BIBLIOGRÁFICAS 144

1. INTRODUÇÃO GERAL

A biodiversidade Neotropical de peixes de água doce é considerada a mais diversa do mundo, representando cerca de 30% de toda a biodiversidade ictiológica mundial (Reis et al., 2016). Entretanto, atualmente, essa diversidade enfrenta níveis de destruição antropogênica sem precedentes (Brauer et al., 2016), levando ao declínio populacional e resultando na maior taxa de extinção entre os vertebrados (Burkhead 2012; WWF 2020). Dentre esses impactos, a construção de barragens é uma das ameaças mais emergentes para as espécies de peixes de água doce. Esses barramentos foram construídos ao longo da maioria dos grandes rios (Nilsson et al., 2005), obstruindo as rotas migratórias de muitas espécies (Agostinho et al., 2008), impedindo-as de chegar a áreas críticas de desova e alimentação. Além disso, os represamentos também alteram as condições do hábitat a montante e a jusante, alterando o fluxo natural dos rios (Pelicice et al., 2021; Xingyuan et al., 2023). Estratégias eficazes de conservação para a biodiversidade de peixes de água doce requerem uma compreensão de seus padrões de distribuição, tendências populacionais e ameaças subjacentes. No entanto, existem lacunas de conhecimento significativas no estado de conservação dessas comunidades. Dessa forma, o monitoramento ambiental é fundamental para garantir a obtenção de dados sobre as comunidades de peixes, especialmente em ecossistemas altamente impactados por atividades antrópicas. É primordial que o monitoramento seja realizado de forma regular e sistemática, permitindo que as informações coletadas possam ser usadas para aprimorar e definir as mais adequadas ações para manejo e conservação (Kelly et al., 2014; Sales et al., 2018) e, assim, minimizar os impactos sobre o meio ambiente.

A avaliação da biodiversidade depende da identificação precisa e confiável das espécies, assim, métodos complementares aos métodos tradicionais de monitoramento, baseados na utilização do DNA, principalmente, estão sendo cada vez mais implementados em todo o mundo. Entretanto, apesar de ser um país megadiverso, o Brasil ainda está em uma fase incipiente em termos de uso de abordagens baseadas em DNA para o monitoramento da biodiversidade. Isso se deve em parte à falta de investimento em pesquisa e tecnologia, bem como à falta de incentivos para a adoção de novas tecnologias pelos órgãos governamentais responsáveis pela gestão da biodiversidade (Sales et al., 2018). Além disso, os países tropicais são em geral menos desenvolvidos economicamente, o que associado aos custos de infraestrutura e reagentes necessários resultam em baixa geração de

recursos humanos especializados (Zinger et al., 2020; Haddad et al., 2022). Atualmente, a avaliação da biodiversidade de peixes baseia-se na amostragem utilizando métodos tradicionais de pesca (e.g. redes de emalhar, arrasto, tarrafa) que dependem extensivamente da captura ou observação dos indivíduos (Bonar et al., 2009), sendo economicamente dispendiosas, seletivas, além de apresentarem baixa efetividade na detecção de espécies raras e elusivas e aquelas em estágios iniciais de vida (Becker et al., 2015; Sales et al., 2018, 2021). Dessa forma, desenvolver técnicas inovadoras que possibilitem a avaliação da biodiversidade utilizando métodos não invasivos pode auxiliar na detecção e no monitoramento da biodiversidade (Deiner et al., 2017).

Uma alternativa promissora aos métodos tradicionais de biomonitoramento de ecossistemas aquáticos no mundo, tanto de ambientes marinhos como de água doce, é a abordagem baseada na identificação de DNA obtido de amostras ambientais (e.g solo, sedimento, ar, água, etc.), denominado DNA ambiental (eDNA) (Ficetola et al., 2008; Thomsen et al., 2012a Bohmann et al., 2014; Boussarie et al., 2018; Deiner et al., 2017; Hänfling et al., 2016; Thomsen & Willerslev, 2015). Essas amostras de eDNA são amplificadas utilizando iniciadores conservados para regiões previamente definidas, informativas para o táxon de interesse, e são submetidas ao sequenciamento de alto rendimento (*High Throughput Sequencing* - HTS), o que permite a identificação taxonômica de múltiplas espécies, sem a necessidade de isolamento ou captura dos organismos alvo, metodologia conhecida como, eDNA *metabarcoding*. Adicionalmente, a abordagem do eDNA *metabarcoding* pode complementar e romper as limitações dos métodos convencionais (Rees et al., 2014), identificando diferentes espécies e aumentando a resolução da identificação taxonômica (Thomsen et al., 2012a, b; Valentini et al., 2016; Deiner et al., 2017; Keck et al., 2022). Melhores estimativas da distribuição de espécies vulneráveis, e realizadas de forma não invasiva, facilitam o desenvolvimento de políticas e permitem direcionar de forma eficiente os esforços de gestão em todos os hábitats (Kelly et al., 2014; Thomsen e Willerslev, 2015).

Apesar do aumento de estudos utilizando o eDNA *metabarcoding* nos últimos anos, três potenciais limitações foram destacadas por estudos utilizando o eDNA *metabarcoding* para avaliar peixes em diferentes ecossistemas aquáticos: (1) a seleção de *primers* universais adequados, (2) a cobertura de bancos de dados de

referência e a (3) sensibilidade de *pipelines* de bioinformática (Collins et al., 2021; Schenekar et al., 2020; Weigand et al., 2019).

Para a análise do eDNA *metabarcoding*, é necessário a utilização de *primers* que amplifiquem regiões curtas de DNA (<200pb), tendo em vista a dificuldade de amplificação de fragmentos longos de DNA extraído de amostras ambientais, material que normalmente já se encontra altamente degradado (Willerslev et al., 2003; Hansen et al., 2006), especialmente em regiões com temperaturas e radiação solar mais altas associadas ao aumento da turbidez, como é o caso da região Neotropical (Barnes et al., 2014; Matheson et al., 2014; Pilliod et al., 2014). Além disso, esses fragmentos do DNA devem ser conservados o suficiente nas extremidades para permitir o alinhamento dos *primers*, e diferentes o bastante para serem informativas sobre a espécie de origem. Para identificação de animais, os *primers* mais frequentemente utilizados se baseiam em *loci* mitocondriais, como o citocromo oxidase subunidade I (COI ou cox1) (Hebert et al., 2003), marcador padrão definido pelo Consórcio do Barcode of Life (CBOL). No entanto, os *barcodes* baseados no gene COI não podem ser usados com segurança para a análise de eDNA *metabarcoding*. Além de se basearem em *amplicons* mais longos (~650pb), o caráter conservativo do gene COI, que o torna útil para identificação de espécies através de *barcodes*, dificulta o desenvolvimento de *primers* específicos para *metabarcoding* (Clarke et al., 2014; Deagle et al., 2014; Sharma et al., 2014). Além disso, os *primers* desenvolvidos para amplificação de *amplicons* menores são projetados com um alto grau de bases degeneradas (Leray et al., 2013; Marquina et al., 2019), resultando na amplificação de uma gama diversificada de táxons (Elbrecht & Leese, 2017), mas também uma porcentagem significativa de leituras de sequências de táxons não-alvo (Hajibabaei et al., 2019; Collins et al., 2019).

Alternativamente, outras regiões mitocondriais, como o RNA ribossomal 12S (rRNA), têm fornecido resultados eficientes para a identificação molecular de várias espécies, incluindo espécies de peixes (Miya et al., 2015; Valentini et al., 2016; Deiner et al., 2017). Dentre os *primers* mais utilizados em estudos de eDNA *metabarcoding* de peixes, os mais comumente utilizados são aqueles desenhados por Myia e colaboradores (2015), o MiFish, e Teleo, padronizado por Valentini e colaboradores (2016). Entretanto, ambos os *primers* foram desenhados ou validados baseados em espécies de peixes de regiões temperadas. Além disso, os estudos sobre comunidades de peixes usando o eDNA *metabarcoding* estão concentrados na

Europa, América e Japão (Valdez-Moreno et al., 2019) e demonstraram diferenças na fauna de peixes entre as regiões, utilizando conjuntos de *primers* universais. Portanto, ainda é necessário demonstrar a aplicabilidade e generalidade de *primers* universais projetados para peixes de água doce quando usados em estudos conduzidos em diferentes ambientes (Hu et al., 2022). Dessa forma, questionamentos sobre qual o conjunto de *primers* poderia amplificar de forma confiável o eDNA de todas as espécies de peixes de uma bacia hidrográfica megadiversa e ter resolução suficiente para identificá-los e quais análises computacionais são mais adequadas para detectar essas espécies simultaneamente, ainda permanecem em aberto.

No eDNA *metabarcoding*, como a designação taxonômica é realizada através da comparação dos *amplicons* com sequências de DNA referência, um banco de dados de referência incompleto pode comprometer a aplicação efetiva dessa metodologia na identificação de peixes (Mendoza et al., 2015). Uma cobertura abrangente de táxons conhecidos em bancos de dados de referência é necessária para a detecção completa e precisa das espécies (Andersen et al., 2019; Weigand et al., 2019). Os bancos de dados baseados no gene COI são geralmente mais completos do que os bancos de dados direcionados aos genes mitocondriais 12S e 16S do rRNA. No entanto, como a região COI demonstrou ser menos adequada para os trabalhos de eDNA *metabarcoding* de peixes, é necessário expandir as bibliotecas referências baseadas em regiões mais adequadas, como 12S e 16S. (Collins et al., 2019; Deagle et al., 2014). Adicionalmente, estudos realizados na região Neotropical reforçam que bancos de dados de referência insuficientes e a falta de conhecimento da biodiversidade de peixes local (como espécies não descritas ou crípticas), podem comprometer a eficácia do eDNA *metabarcoding* na avaliação da biodiversidade megadiversa neotropical de peixes (Sales et al., 2018; Cilleros et al., 2019; Sales et al., 2020; Jackman et al., 2021). Dessa forma, Xiong e colaboradores (2022) tem sugerido a utilização de bancos de dados referências customizados para cada região estudada, avaliando listas de espécies locais para a identificação das sequências de eDNA e, posteriormente, a utilização de bancos de dados públicos para a designação taxonômica de sequências que não puderam ser identificadas, a fim de checar espécies crípticas ou não-nativas.

Por fim, o terceiro aspecto limitador na aplicação do eDNA *metabarcoding* para designação taxonômica de peixes é o processamento dos dados por análises de bioinformática, visto que diferentes variáveis podem influenciar nas estimativas de

biodiversidade de peixes (Evans et al., 2017), e por isso esses dados devem ser interpretados com cautela. Por exemplo, os estudos de eDNA *metabarcoding* na região Neotropical têm se baseado principalmente em técnicas de agrupamento baseadas em MOTUs (Molecular Operational Taxonomic Units), com um limite de similaridade de sequência de 97% usado para identificação de espécies (Cilleros et al., 2019, Sales et al., 2021). Entretanto, estudos recentes têm demonstrado a importância de incorporar abordagens distintas, como por exemplo: as *zero-radius* OTUs (ZOTUs) e as Amplicon Sequence Variants (ASVs) para descrever a diversidade recuperada pelas análises de eDNA *metabarcoding* (Dal Pont et al., 2021). Recentemente, Antich e colaboradores (2021) destacaram a importância de combinar procedimentos de tratamento das leituras ao considerar o fragmento do gene COI, destacando a necessidade de combinar as abordagens de *denoising* e *clustering*, com escolha adequada dos parâmetros. No entanto, uma avaliação de ambos os métodos usando comunidades simuladas para determinar os verdadeiros benefícios de cada abordagem ainda não foi aplicada a outros marcadores, como o 12S, em comunidades de peixes megadiversos neotropicais.

Nesse sentido, o presente trabalho propôs desenvolver metodologia molecular baseada na abordagem do eDNA *metabarcoding* para a detecção e monitoramento de espécies de peixes brasileiros, utilizando como modelo a comunidade de peixes do Reservatório da Usina Hidrelétrica de Irapé (IRR), como parte do Projeto de Pesquisa e Desenvolvimento da CEMIG, uma das principais concessionárias de energia elétrica do Brasil, o P&D GT 635. A UHE Irapé está localizada na bacia do Rio Jequitinhonha, possui uma área alagada de 137,16 Km^2, com capacidade de geração de 360,00 MW, e é considerada a barragem mais alta do Brasil e uma das mais altas do mundo, com 208 metros. A UHE Irapé foi construída em 2006 e interrompe o curso do Rio Jequitinhonha pouco depois do seu entroncamento com o Rio Itacambiruçu. Essa bacia se destaca tanto pela grande proporção de espécies ameaças e endêmicas, como pela escassez de estudos (Rosa e Lima, 2008; Pugedo et al., 2016). O reservatório da UHE Irapé foi escolhido para o desenvolvimento desse estudo, pois nele são realizados monitoramentos de peixes de forma tradicional desde 2011, fornecendo informações complementares para comparações com os resultados que serão encontrados com a técnica proposta.

Para tal, esse trabalho foi dividido em três capítulos, em que objetivamos: (1) desenvolver uma biblioteca referência e um conjunto de *primers* baseados no

marcador de RNA ribossomal 12S a partir da biblioteca de sequências de DNA referência customizada para a bacia do rio Jequitinhonha, (2) entender melhor o uso do *metabarcoding* na detecção da diversidade de peixes neotropicais, avaliando a influência de variáveis específicas na composição da amostra, usando comunidades simuladas de peixes e por fim, (3) realizar a identificação taxonômica de espécies de água doce no reservatório da Usina Hidrelétrica de Irapé (IRR), e comparar os resultados de eDNA *metabarcoding* com métodos de captura tradicionais estabelecidos, de forma a expandir o conhecimento sobre o uso do eDNA *metabarcoding* como método de monitoramento de peixes em reservatórios. A compreensão destes fatores permitirão a expansão do entendimento da dinâmica do eDNA nesses ambientes, fornecendo uma melhor visão sobre a sua captura e análise, e aprimorando as inferências espaço-temporais de presença de espécies.

2. OBJETIVO GERAL

Desenvolver metodologia molecular baseada na abordagem do DNA ambiental (eDNA) metabarcoding para a detecção e monitoramento de espécies de peixes no Reservatório da Usina Hidrelétrica de Irapé (IRR).

3. JUSTIFICATIVA

Programas de monitoramento da ictiofauna vêm sendo desenvolvidos por várias concessionárias de energia, entretanto, vários desses não são avaliados quanto a sua eficácia para o estabelecimento de medidas mitigadoras dos impactos causados pelos barramentos de usinas hidrelétricas. Com a ausência desta avaliação crítica da eficiência dos métodos de biomonitoramento na detecção de espécies não é possível ter uma real perspectiva da relação custo-benefício dos monitoramentos e da efetividade das técnicas utilizadas. Atualmente, técnicas tradicionais de pesca são as metodologias estabelecidas na região Neotropical para monitoramento da ictiofauna, entretanto, o seu alto custo e a necessidade de capturar os indivíduos são limitações que estão sendo cada vez mais questionadas.

Nesse sentido, este trabalho visou a avaliação e a aplicação do eDNA *metabarcoding*, uma técnica incipiente no Brasil, para monitoramento da ictiofauna. Essa metodologia, bem estabelecida e aceita pelas agências de monitoramento em países do Hemisfério Norte (e.g., Great crested newts – Reino Unido; European Water Framework Directive – União Européia) poderá fornecer informações importantes sobre a dinâmica da comunidade de peixes de água doce e avaliar o impacto de ações

antrópicas, como introdução de espécies e construção de barragens. Além disso, o biomonitoramento de espécies de peixes usando ferramentas moleculares também pode contribuir para estabelecer esquemas de monitoramento de longo prazo e obter informações de áreas frequentemente negligenciado devido à falta de acessibilidade.

Entretanto, a eficácia do eDNA *metabarcoding* em um contexto neotropical ainda precisa ser explorado, especialmente devido a alguns aspectos metodológicos como a falta de conhecimento da biodiversidade local e ausência de sequências no banco de dados referência, vieses de *primer* e as diferentes formas de identificação taxonômica (MOTUs x ASVs), antes de aplicar essa técnica na categorização da comunidade de peixes em larga escala.

O reservatório de Irapé foi escolhido para o desenvolvimento deste estudo, pois além de se já realizar o monitoramento de peixes da forma tradicional, como condicionante ambiental, forneceu informações prévias para as comparações com os novos resultados.

4. CAPÍTULO I - New 12S metabarcoding primers for enhanced Neotropical freshwater fish biodiversity assessment (Scientific Reports, 2020)

4.1 Objetivos específicos

- Desenvolver biblioteca referência do RNA ribossomal 12S;
- Desenvolver um conjunto de *primers* específicos para espécies de peixes neotropicais de água doce;
- Avaliar a resolução taxonômica dos fragmentos amplificados pelos conjuntos de *primers* desenvolvidos usando Métodos bayesianos e baseados em distância;
- Determinar o *threshold* de distância genética para delimitação de espécies peixes usando a região 12S, delimitada pelos *primers* NeoFish_3;
- Realizar testes *in vitro* para validar os conjuntos de *primers* desenvolvidos.

www.nature.com/scientificreports

scientific reports

Check for updates

OPEN

New 12S metabarcoding primers for enhanced Neotropical freshwater fish biodiversity assessment

David T. Milan[1,5], Izabela S. Mendes[1,2,5], Júnio S. Damasceno[1], Daniel F. Teixeira[1,2], Naiara G. Sales[3,4] & Daniel C. Carvalho[1,2,5]✉

The megadiverse Neotropical fish fauna lacks a comprehensive and reliable DNA reference database, which hampers precise species identification and DNA based biodiversity assessment in the region. Here, we developed a mitochondrial 12S ribosomal DNA reference database for 67 fish species, representing 54 genera, 25 families, and six major Neotropical orders. We aimed to develop mini-barcode markers (i.e. amplicons with less than 200 bp) suitable for DNA metabarcoding by evaluating the taxonomic resolution of full-length and mini-barcodes and to determine a threshold value for fish species delimitation using 12S. Evaluation of the target amplicons demonstrated that both full-length library (565 bp) and mini-barcodes (193 bp) contain enough taxonomic resolution to differentiate all 67 fish species. For species delimitation, interspecific genetic distance threshold values of 0.4% and 0.55% were defined using full-length and mini-barcodes, respectively. A custom reference database and specific mini-barcode markers are important assets for ecoregion scale DNA based biodiversity assessments (such as environmental DNA) that can help with the complex task of conserving the megadiverse Neotropical ichthyofauna.

Assessing biodiversity in species-rich regions is fundamental for environmental conservation as anthropogenic activities are drastically increasing the rate of biodiversity loss and changing ecosystem functioning[1]. Freshwater ecosystems are currently considered a priority target for biodiversity conservation due to the reported massive decline (i.e. ~ 83% since 1970) in species richness[2]. Therefore, aquatic ecosystem assessment and biomonitoring programs are conducted to provide data on fish species conservation status and community changes[3]. However, these programs are mostly based on traditional assessment methods (e.g. netting, trawling) and depend extensively on capture or observation[4], which may be inefficient or cause harmful impacts to the environment and the biological communities[5]. Hence, developing alternative tools to monitor biodiversity is pivotal to inform conservation and management strategies[6].

A promising alternative to traditional aquatic ecosystem assessment and biomonitoring methods is a DNA-based approach, which can complement or even be more efficient than traditional methods[7,8]. Further, it is possible to obtain DNA mixtures from environmental samples (e.g. water and sediment) without first isolating target organisms (environmental DNA – eDNA). After extraction, these samples can be subjected to high-throughput sequencing (HTS) to identify the presence of multiple species, namely DNA metabarcoding[9,10].

DNA metabarcoding is a powerful tool for biodiversity assessment that has been widely used for several purposes and different taxonomic groups, including identification and quantification of neotropical ichthyoplankton[11,12], stomach-content analysis of a ray species[13], and identification of wasp species using and comparing the Sanger and HTS methods[14]. Furthermore, environmental sampling (eDNA) has been successfully used for molecular identification of several vertebrate groups in temperate regions[15–17], monitoring of endangered species such as freshwater fish in Australia and turtles in the United States[18,19], and improved detection

[1]Conservation Genetics Lab, Postgraduate Program in Vertebrate Biology, Pontifical Catholic University of Minas Gerais, PUC Minas, Belo Horizonte, Brazil. [2]Postgraduate Program in Genetics, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Brazil. [3]Ecosystems and Environment Research Centre, School of Environment and Life Sciences, University of Salford, Salford, UK. [4]CESAM - Centre for Environmental and Marine Studies, Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Lisbon, Portugal. [5]These authors contributed equally: David T. Milan, Izabela S. Mendes and Daniel C. Carvalho. ✉email: carvalho.lgc@gmail.com

over traditional assessment methods for monitoring the invasive American bullfrog in France[20]. Additionally, Reid et al. (2019) highlight the use of eDNA as one of the main conservation and management tools for dealing with the emerging threats for freshwater biodiversity. However, the potential of DNA metabarcoding to monitor vertebrate communities remains poorly explored in the Neotropical region, and few studies have been conducted to date[11,21–23].

The relative lack of DNA-based monitoring in the Neotropics may be due to some constraints that hamper its full application, such as incomplete taxonomic assignment due to the lack of reference sequences[21,22]. Therefore, the construction of a curated and complete reference molecular database is vital for efficient application of DNA-based methods towards biodiversity assessment in megadiverse realms. In the absence of reference sequences, taxonomic assignment is hindered, restricting the analyses to the use of Molecular Operational Taxonomic Units (MOTUs) and often only allowing assignments up to the family level, limiting ecological conclusions. The need for short amplicon length, due to DNA degradation in environmental DNA samples[24], and for avoiding amplification of non-target taxa (e.g. invertebrates) are other pitfalls for sound DNA-based ecological monitoring, especially in biodiverse environments such as the Neotropics[21].

A large dataset built using the DNA barcoding marker sensu stricto (i.e. use of ~ 600 base pairs (bp) of the mitochondrial cytochrome oxidase subunit I—COI) combined with traditional morphological techniques has contributed to the improvement of reference databases and to a better assessment of the Neotropical megadiverse ichthyofauna[25–28]. However, usage of the COI gene for macro-organism DNA metabarcoding analyses has proven to be difficult due to non-target amplification of bacteria and small microeukaryotes, which is inherent to the use of COI in eDNA samples[29].

The 12S and 16S ribosomal RNA genes (rRNA) have been widely used as alternative markers and have provided efficient results for molecular detection of several species through eDNA metabarcoding, including fishes[7,21,30,31]. For instance, Miya et al. (2015)[32] developed a 12S set of universal PCR primers for eDNA metabarcoding (MiFish) by targeting a hypervariable region with 163–185 bp from whole mitogenomes of 880 fishes, mostly subtropical marine species. Another 12S primer set commonly used in metabarcoding studies, Teleo1, was designed to amplify a region shorter than 100 bp based on 117 standard Teleostei (bony fish) sequences of the European Molecular Biology Laboratory—European Nucleotide Archive database[7]. However, there is also a need to use human blocking primers to avoid cross amplification. These markers were successfully applied in eDNA studies of high-diversity environments within the Neotropical freshwater ichthyofauna[21,22] but without any previous analysis of marker taxonomic resolution or species detection efficiency.

Therefore, before applying DNA metabarcoding in the Neotropics, the development and validation of molecular markers that can provide a reliable and robust taxonomic assignment is highly recommended. To this end, MacDonald & Sarre (2017)[33] suggested a framework for the development and validation of taxon-specific primers for eDNA metabarcoding analyses in ecological studies. This framework includes the construction of a reference database and its phylogenetic evaluation, primer design, and the in silico and in vitro evaluation of primer specificity, sensitivity, and utility in the laboratory.

Here, we developed a reference database targeting the 12S rRNA, using an *in-silico* approach, designed three new mini-barcoding 12S primer sets based on our reference database, and evaluated their phylogenetic resolution. The taxonomic resolution of full-length and mini-barcodes for species delimitation were compared using Bayesian and distance-based methods. In addition, we determined the genetic distance threshold value for fish species delimitation using the targeted 12S region for both mini and full-length libraries. In vitro tests were also conducted to validate our new 12S mini-barcode marker. Our custom reference database and new primer sets may be an alternative to previously developed markers to target Neotropical freshwater biodiversity and assist in the complex task of monitoring and conserving such diverse ichthyofauna.

Materials and methods

Tissue samples collection. We used fin clips from 67 fish species collected prior to this study and stored in 100% ethanol at − 4 °C at the Conservation Genetics Lab—PUC Minas. The specimens are from the São Francisco River Basin (Brazil), and the tissue samples and vouchers were previously used to build a DNA reference Barcode database using 650 bp of the Cytochrome oxidase I gene (ICMBIO collection permit number: 37298-1) from which the barcode data indicated cryptic species that would result in greater number of molecular taxonomic units. We followed the taxonomic classification obtained by Carvalho et al. (2011)[28] through morphological and DNA barcoding for all fish. Additional information regarding DNA extraction, amplification, and sequencing is provided in the Supplementary Material (page 1).

Sequences analyses. DNA was extracted using a salting-out protocol adapted from Aljanabi & Martinez (1997)[34]. Polymerase Chain Reactions (PCR) of the 12S rRNA gene were performed in a PCR thermal cycler (Veriti, Life Biosystems) using 10.0 µl solution composed of 7.0 µl of ultrapure water (Promega), 1.0 µl of 10X buffer containing 2.5 mM $MgCl_2$, 1.0 µl of template DNA, 0.3 µl of dNTP (10 mM) (Invitrogen), 0.25 µl of each primer (10 µM), and 0.2 µl of Taq DNA polymerase (5U/ µl) (PHT). In order to amplify a fragment of ca. 600 bp of the 12S region (namely full-length region), we used the V05F_898 (5′-AAACTCGTGCCAGCCACC-3′) and teleoR (5′-CTTCCGGTACACTTACCATG-3′) primer sets presented in Thomsen et al. (2016)[35]. The thermal cycle profile consisted of initial denaturation at 95 °C (2 min), then 35 cycles of denaturation at 95 °C (1 min), primers annealing at 57 °C (30 s) and extension at 72 °C (1 min), and final extension at 72 °C (7 min). The amplicons were visualized in agarose 1% electrophoresis before DNA sequencing.

All samples were sequenced bi-directionally. The DNA sequencing reaction was performed using a Big Dye Terminator v.3.1 (Applied Biosystems) commercial kit in a reaction with a 10.0 µl final volume that consisted of: 1 µl of PCR amplified product, 1 µl of primer (10uM), 1 µl of Pre-Mix solution (Big Dye Terminator), 1.5 µl of

Buffer 5X, and 5.5 µl of ultrapure water. The DNA sequencing reaction was performed in a Veriti thermocycler (Life Biosystems) with the following conditions: denaturation at 96 °C (2 min), then 35 cycles of denaturation at 96 °C (30 s), annealing at 50 °C (15 s), and extension at 60 °C (4 min). The samples were precipitated with EDTA (125 mM) and ethanol (100%) and washed with 70% ethanol. The sequencing plates were dried at 95 °C for eight minutes. DNA sequences were obtained in an ABI 3500 Genetic Analyzer (Life Technologies) automatic DNA analyzer.

The 12S consensus sequences (contigs) were obtained using DNA Baser v.3.5.4 software and aligned using ClustalW[36], after trimming ambiguous ends. MEGA v7.0 software[37] was used to estimate all genetic distances (intraspecific, intrageneric, intrafamilial, and interspecific) using the Kimura two-parameter (K2P) nucleotide evolution model[38] and to construct dendrograms using the Neighbor-joining (NJ) method[39], with 10,000 bootstrap pseudoreplicates[40], showing only well-supported clade values (> 80%).

Design and screening for best annealing primer sites. Three mini-barcode primer sets (NeoFish_1, NeoFish_2, and NeoFish_3) were designed to anneal to highly conserved flanking regions targeting variable sequences based on the alignment of all 12S DNA sequences obtained, which included 132 sequences corresponding to 67 species (19 species with only one specimen), ranging from one to three specimens/species (mean of 1.97 sequences per species). We used PRIMER3 software, implemented in Geneious v.4.8.5 (Kearse et al. 2012[41]—https://www.geneious.com) to find the best primer sites based on the 12S reference database by applying default parameters but restricting an amplicon length to shorter than 250 bp. Primers were designed with a 20%–80% guanine-cytosine (GC) content and a melting temperature between 57 and 63 °C. The best primer set was chosen based on an in vitro test, and it was then used for further analyses.

Evaluation of the newly developed primer sets was performed using a sliding window analysis (SWAN)[42] conducted in the SPIDER package[43] in R (version 3.6.1)[44], which possesses useful analyses for determining ideal regions for mini-barcode design[45]. The *slideAnalyses* function was used to generate windows of 70 bp, which were shifted along the length of the 12S alignment in 10 bp intervals to evaluate regions of: (1) high mean K2P distance; (2) few zero pairwise non-conspecific distances; (3) high proportion of clades shared between the Neighbor-joining tree from the 12S full-length barcode and the tree constructed using only data from selected windows; and (4) high sum of diagnostic nucleotides.

Using the Primer Map analysis we check for overlapping amplification target regions of our newly developed mini-barcode primer set with previously developed 12S markers (i.e. MiFishU[32], Teleo1[7] and V05F_898[35]). The complete 12S rRNA sequence from *Prochilodus costatus* mitogenome (952 bp—GenBank number NC_027690) was used as a template.

To compare the non-target organism amplifications between our mini-barcode primer set to previously developed 12S markers (i.e. MiFishU and Teleo1), we performed in silico PCR using Primer-BLAST[46]. For primer specificity stringency options, we allowed at least three mismatches to unintended targets, including at least three mismatches within the last five base pairs at the 3′ end, a maximum target size of 400 bp and an annealing temperature of 60 °C.

Species delimitation analyses of the full-length and mini-barcode reference database. The MOTUs were obtained to assess the taxonomic resolution of the full-length library (565 bp) and 193 bp mini-barcode (618–851 bp of *P. costatus* 12S complete sequence) from the trimmed 12S full-length reference by applying four single-locus species delimitation analyses. Two of these analyses were conducted using the Bayesian methods of Generalized Mixed Yule-Coalescent (GMYC)[47] and Bayesian implementation of Poisson Tree Process (bPTP)[48]. Two other analyses used genetic distance-based methods: Automatic Barcode Gap Discovery (ABGD)[49] and species delimitation threshold defined by threshold optimization analysis in SPIDER package. Each analysis was conducted as described below.

For GMYC, an ultrametric tree was generated for each marker by the Bayesian Phylogenetic reconstruction in BEAST[50] and used as the input file. The relaxed lognormal distribution and the Birth and Death process as tree priors were used as clock models. The GTR + G + I model was used as nucleotide evolution model for 12S full-length and mini barcodes, and the Markov Chain (MCMC) procedure was used with 50 × 106 and 150 × 106 generations for 12S mini and full barcodes, respectively, sampling one tree every 104 generations. Convergence was indicated by Tracer v1.6[51] with estimated sample sizes (ESS) superior to 200. An appropriate number of trees (first 10%) from each run was discarded as burn-in and the MCMC samples were generated using the maximum clade credibility (MCC) topology in TreeAnnotator v1.4.7[52] and visualized in FigTree v1.4.3. The annotated trees were submitted for GMYC analysis in R with the Splits package (Species Limits by Threshold Statistics; https://r-forge.r-project.org/projects/splits) and a single threshold strategy using default scaling parameters.

We used the bPTP model in the bPTP web server (https://species.h-its.org/ptp/) under default parameters to delimitate the MOTUs. bPTP does not require an ultrametric gene tree and uses, instead, a Newick tree as the input file with branch lengths representing the number of nucleotide substitutions. We used Newick trees generated in MEGA7 as input files, using a Neighbor-joining method and the TN93 + G evolution model, which was chosen as the best evolutionary model in MEGA.

ABGD was applied to automatically group species into partitions indicating the molecular taxonomic resolution of the 12S database. ABGD first uses a range of prior intraspecific divergences to divide the data into groups based on a statistically inferred barcode gap and then recursively applies the same procedure to the groups obtained in the first step. ABGD analysis was performed using a web interface (https://wwwabi.snv.jussieu.fr/public/abgd/) with a relative gap width value of X = 0.8, while the other parameter values employed defaults. Assignments for intraspecific divergence (P-distances) between 0.001 and 0.100 were recorded[49].

Primer set	Primer name	Amplicon Length (bp)	Primer sequence (5′–3′)
NeoFish_1	NeoFish_1F	177	GCCGTCGCAAGCTTACCCTGT
	NeoFish_1R	177	GTGTGCGCGTCTCAGAGCCT
NeoFish_2	NeoFish_2/3F	184	CGCCGTCGCAAGCTTACCCT
	NeoFish_2R	184	GCGGTGTGTGCGCGTCTCAG
NeoFish_3	NeoFish_2/3F	193	CGCCGTCGCAAGCTTACCCT
	NeoFish_3R	193	AGTGACGGGCGGTGTGTGC

Table 1. Primers designed for short amplifications of the 12S sequences obtained from the major Neotropical fish orders.

Threshold optimization analysis (SPIDER package) was conducted using the *threshVal* and *threshID* functions. A genetic distance-based species delimitation analysis was estimated using threshold values determined by the *threshVal* function. This function shows the number of true positive, true negative, false negative, and false positive, rate of fish species identification, together with the cumulative error (i.e. the sum of false positives and false negatives) using a range of threshold values based on K2P genetic distances. These estimated interspecific genetic distance thresholds were applied as the best cut-off values to delimitate species, as there are no previous references delimiting cut-off values for the 12S marker, unlike the COI gene (the 2% standard threshold defined by Ward, 2009[53]). Then, we used the distance threshold defined by *threshVal* in the *threshID* function. The *threshID* function assigns four possible results for each sequence in the dataset: "correct", "incorrect", "ambiguous", and "no id", where "correct" means that all matches within the threshold of the query are the same species and "no ID" shows that no matches were found to any individual within the threshold. Specimens identified as "no ID" were put in individual MOTUs and "correct" ones were put alongside their peers.

In addition, two distance-based analyses were performed (also using SPIDER) to identify taxa with low taxonomic resolution with the mini-barcode: barcoding gap and nearest Neighbor. Singleton sequences (19) were excluded. Detailed information about these analyses is provided in Supplementary Material (page 2).

In vitro tests: evaluation of primer efficiency. To evaluate the efficiency of our mini-barcodes primers in amplifying DNA extracted from fish tissue samples and environmental samples, we conducted two tests. The first in vitro test consisted of PCR amplification and sequencing of the 12S mini-barcode region using the three newly developed primer sets with 16 fish species (22 samples). These samples had been previously used to develop the 12S reference database and represent the six major neotropical orders. PCR conditions for this test consisted of initial denaturation at 95 °C for 1 min, then 35 cycles of denaturation at 95 °C for 30 s, primers annealing at 60 °C for 30 s and extension at 72 °C for 1 min, and final extension at 72 °C for 7 min. DNA sequencing was conducted the same way as in the reference database construction section. For the second test, a water sample collected in an 80-L aquarium containing multiple individuals of pearl cichlid (*Geophagus brasiliensis*) was used to conduct an eDNA experiment to evaluate the potential use of our newly developed marker to detect fish DNA extracted from the environment (detailed information is provided in Supplementary Material, page 6). Experimental procedures followed the principles established by the Brazilian College of Animal Experimentation (COBEA) and approved by the Ethics Committee of the Pontifícia Universidade Católica de Minas Gerais (CEU PUC Minas – permit number: 021/2017).

Results

Custom reference database construction of full-length 12S. We sequenced 132 specimens from 67 fish species representing 54 genera, 25 families, and six orders: Characiformes (60.5% of species), Siluriformes (26%), Cyprinodontiformes (4.5%), Perciformes (4.5%), Gymnotiformes (3%), and Synbranchiformes (1.5%), with an average of 1.97 individuals per species (Supplementary Table S1). The 12S contigs were 565 bp long after trimming the ambiguous ends and had a nucleotide composition of 31.81% adenine, 26.84% cytosine, 20.4% guanine, and 20.95% thymine.

Design and screening for best annealing primer sites. We aimed for conserved primer sites from the 12S full-length library (565 bp) and designed three primer sets with amplicons ranging from 171 to 193 bp, namely NeoFish_1, NeoFish_2, and NeoFish_3 (Table 1). All three primer sets recovered by Primer3 software targeted a similar amplicon region, differing by few base-pairs. The amplicon region started at position 639 and ends at the position 831 of the12S rRNA region of *Prochilodus costatus* (GenBank accession number NC_027690) (Fig. 1a). Primer Map showed that our target amplicon region does not overlap with other sets previously designed for the 12S region (i.e. Teleo1 and MiFishU); however, the primer NeoFish_3R uses almost the same annealing site as Teleo1-F (Fig. 1a). According to SWAN analyses conducted in SPIDER, the region that recovered the best indices of all criteria to design primers is within the 320 to 500 bp of the 12S full-length database, due to the higher sum of diagnostic nucleotides and congruence of NJ trees, as well as lower proportion of zero non-conspecifics (Fig. 2). The mean K2P distance of each window was highest at the beginning of our alignment, between 0 to 100 bp, but was also high within 320 to 400 bp range. Moreover, our chosen target region (~ from nucleotide 320 to 500 bp in Fig. 2) was surrounded by conserved regions with a low frequency of

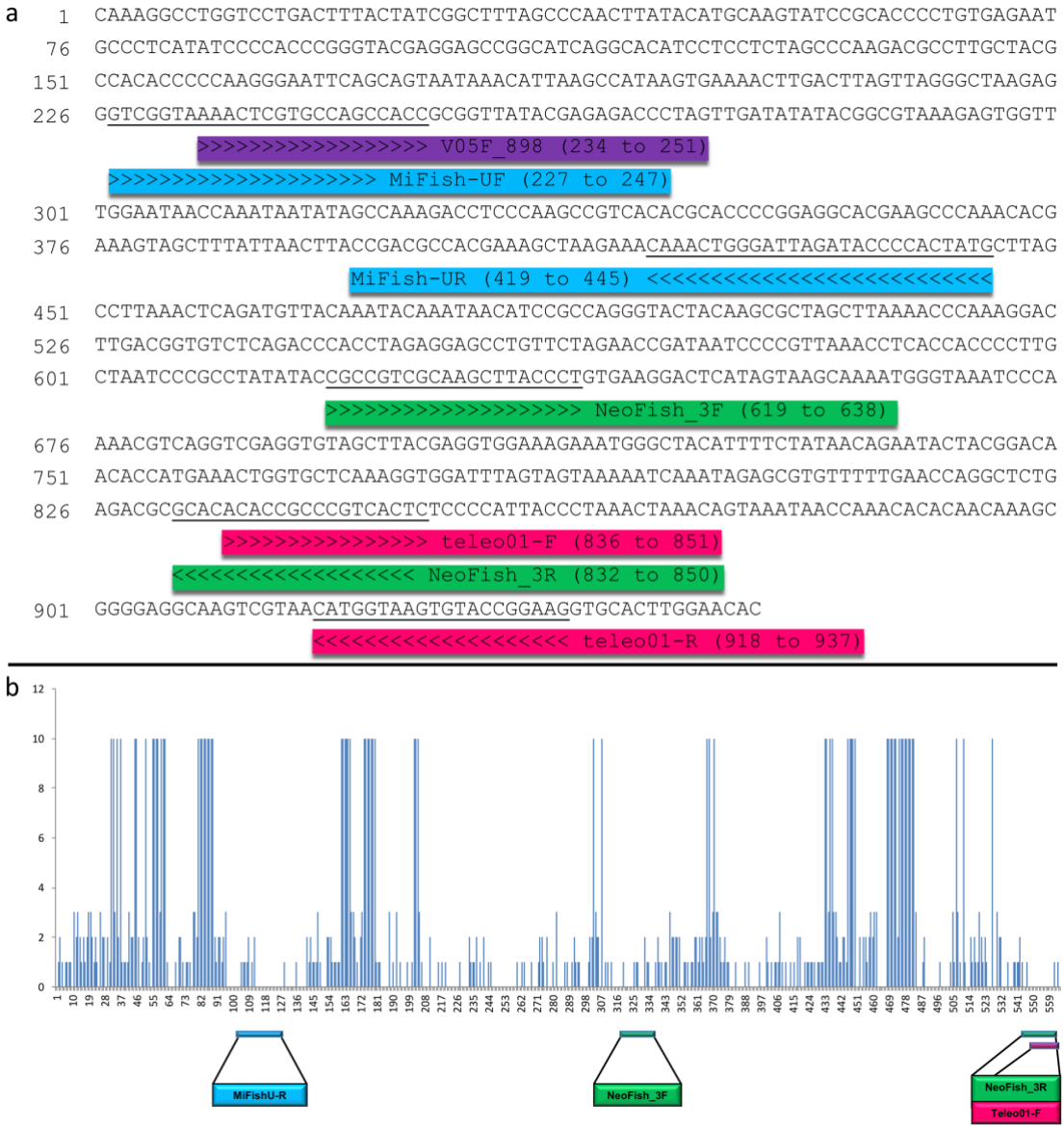


Figure 1. (**a**) DNA sequence alignment showing the annealing positions (< < < < < < < <) for NeoFish_3 forward and reverse primers, V05F_898 (used for reference database construction), and for two others previously developed 12S markers (MiFishU and Teleo1). Base pairs highlighted in bold represent positions where the full-length database starts (317 bp) and ends (851 bp) (**b**) Mismatches and primer sites from the full 12S alignment (565 bp) considering 132 sequences of 67 fish species. Mismatches equal to 10 represent gap sites.

mismatches per primer, thus making it potentially useful to design group-specific primers in accordance with Primer3 choice of primer sites used by NeoFish_1, NeoFish_2, and NeoFish_3 (Fig. 1b).

Primer specificity analysis using Primer-BLAST showed that primer set NeoFish_3 had a lower amplification rate of non-targeted organisms such as bacteria, arthropods, mollusks, and mammalian species (including *Homo sapiens*) when compared to previously developed 12S primer sets (MiFishU and Teleo1) (Table 2). However, NeoFish_3 also showed cross amplification with birds, Lepidosauria, and amphibian sequences similar to previously published 12S markers (Table 2).

Species delimitation analyses of the full-length reference database. Intraspecific genetic K2P divergences ranged from 0% to 2.06% (mean: 0.12%), 0% to 8.88% (mean: 0.92%) for intrageneric comparisons,

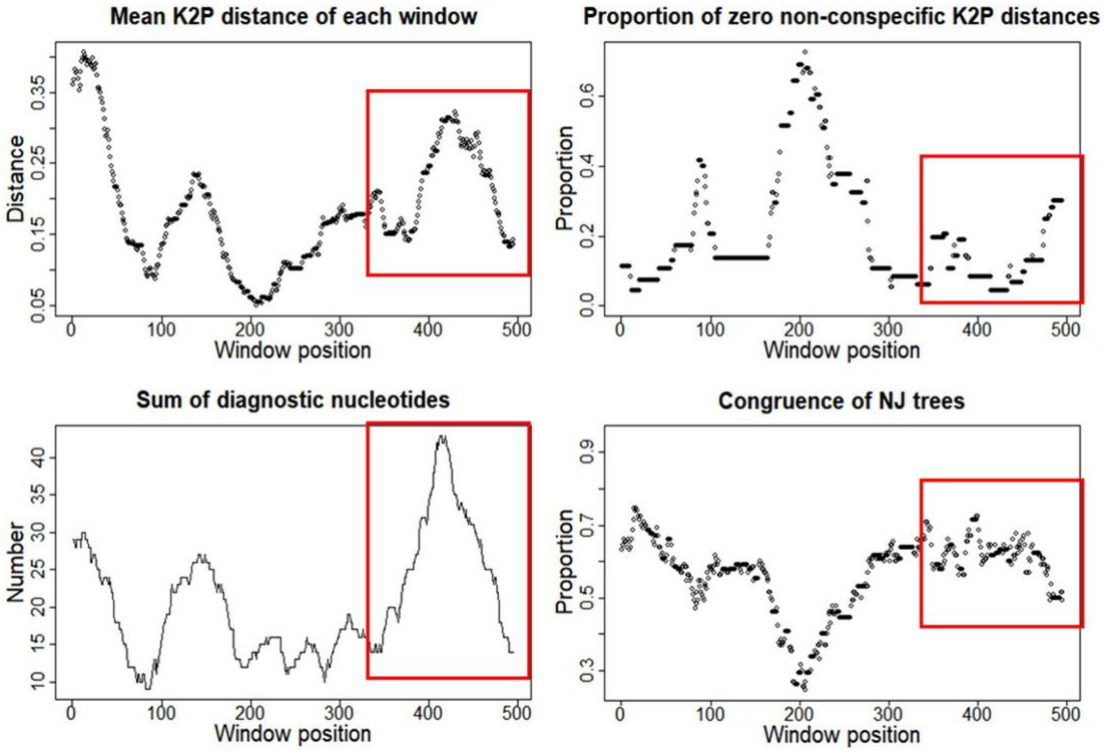


Figure 2. Screening for the best target region for mini-barcode across the 12S full alignment (565 bp) of 67 species using the sliding window analyses (SWAN) from SPIDER package. We analyzed (1) high mean K2P distance; (2) few zero pairwise non-conspecific distances; (3) high proportion of clades shared between the Neighbor-joining tree from the 12S full-length barcode and the tree constructed using only data from selected windows; and (4) high sum of diagnostic nucleotides. Rectangles represent the ideal region for primer design based on best indices of all criteria.

Taxonomic groups	Primers sets		
	NeoFish_3	Teleo1	MiFishU
Bacteria	0	0	0
Arthropoda	0	33	16
Mollusca	0	62	15
Fish	> 1000	> 1000	> 1000
Amphibia	> 1000	> 1000	> 1000
Testudines	211	260	211
Crocodilia	57	50	1
Lepidosauria	> 1000	999	> 1000
Birds	> 1000	> 1000	> 1000
Mammalia	0	> 1000	> 1000
Homo sapiens	0	> 1000	865

Table 2. Primer-BLAST results using the NeoFish_3 primer set and previously developed 12S markers (MiFishU—Miya et al., 2015; and Teleo1—Valentini et al., 2016). Numbers correspond to hits recovered for each taxonomic group.

and 0% to 9.97% (mean: 2.82%) for intrafamilial comparisons. Interspecific genetic distances ranged from 0.41% (*Prochilodus argenteus* vs *P. costatus*) to 32.33% (*Astyanax bimaculatus* vs *Synbranchus marmoratus*). The NJ dendrogram generated with all specimens showed species-specific branches (Supplementary Fig. S1).When considering species delimitation based on Bayesian methods, GMYC detected between 68 and 71 MOTUs (Fig. 3aI) and a threshold time of − 0.008, indicating the time before which all nodes reflect speciation events and after

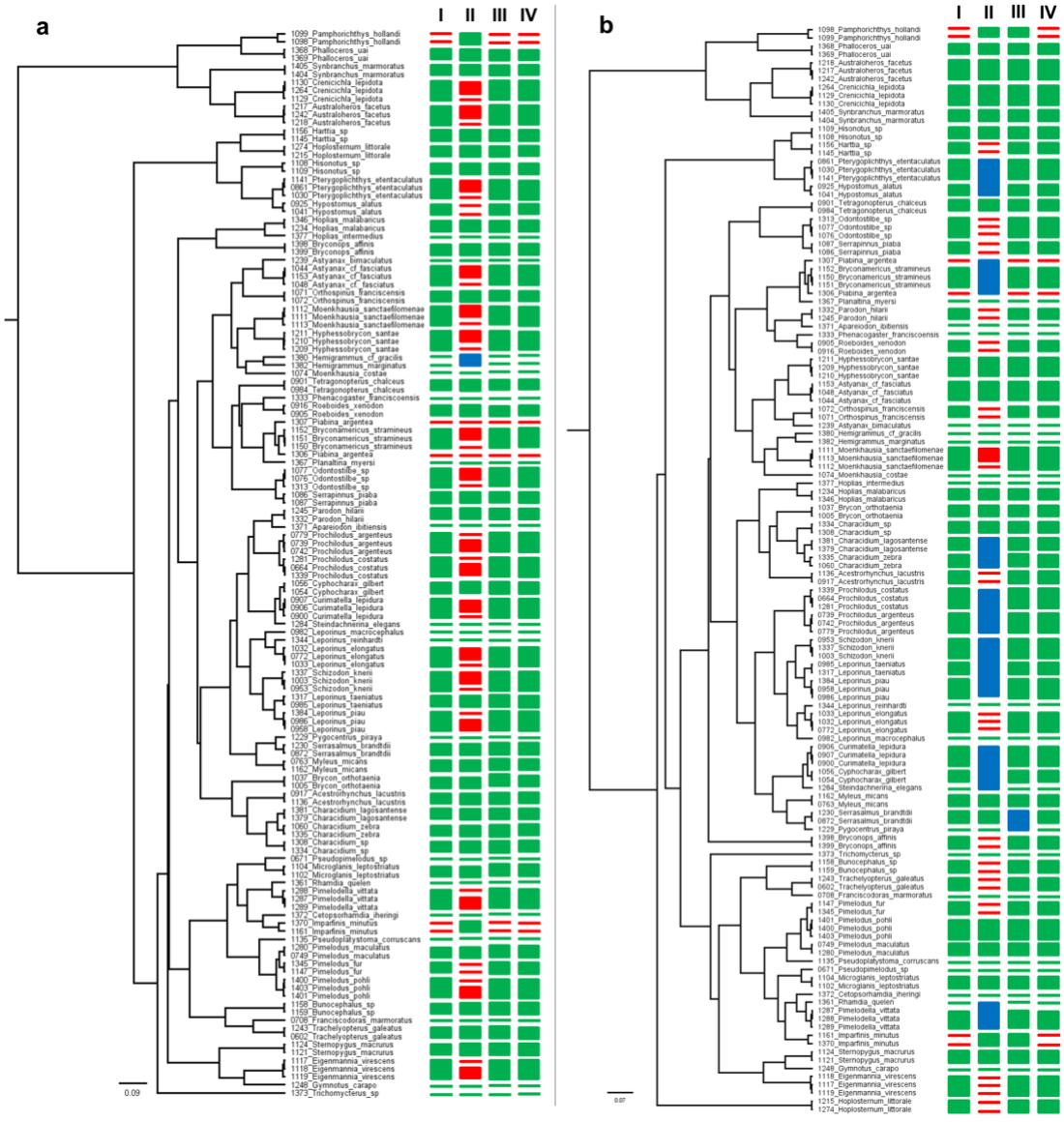


Figure 3. Bayesian phylogenetic ultrametric trees for 12S (**a**) full barcode and (**b**) mini-barcode for all species analyzed. MOTUs are represented by different sizes in accordance with each different species delimitation methods (I—GMYC, II—bPTP, III—ABGD, and IV—interspecific genetic distances thresholds). Green, blue and red colors represent MOTUs with a single species, multiple species, and different MOTUs for the same species, respectively.

which all nodes reflect coalescent events. Maximum likelihood (ML) for the null model was 745.7335 and ML for GMYC model was 788.7312. The ML for the null model revealed the likelihood score of the model that considers that all the sequences belong to the same species, and the likelihood score of the model that splits the sequences into different species. In our case, it is highly significant ($P = 0$), indicating that there is more than one species in our sample. The bPTP revealed 86 MOTUs using a ML approach (Fig. 3aII), with branch support ranging from 0.487 to 1.

Species delimitation based on genetic distance using ABGD analysis detected between 57 and 70 MOTUs when varying the prior maximal distance from $P = 0.021$ to $P = 0.001$, respectively, using the simple distance (p-distance) (Fig. 3aIII). Four partitions, with prior maximal intraspecific distances ranging from 0.001 to 0.004 recovered 70 groups. Two partitions recovered 69 MOTUs, with prior maximal distances of 0.007 and 0.013.

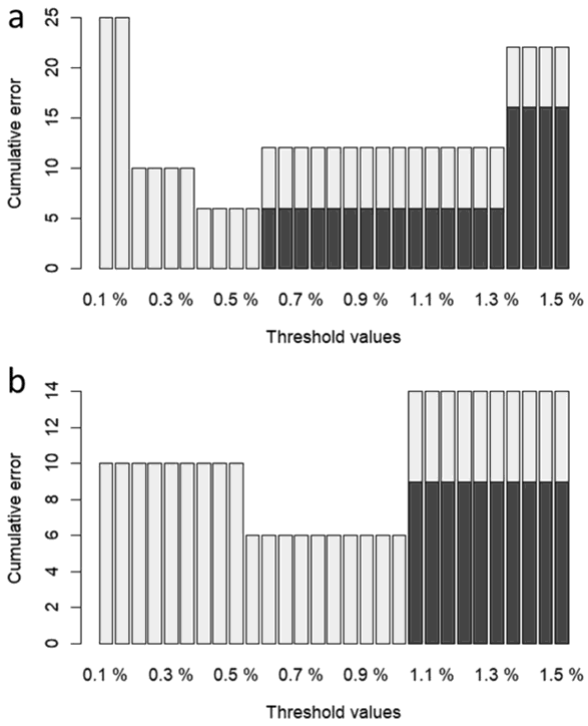


Figure 4. Threshold optimization for species delimitation for 12S (**a**) full length and (**b**) mini-barcode, showing the false positive (light grey) and false negative (dark grey) rate of fish species identification as pre-set thresholds change. Cumulative error is the sum of false positives + false negatives. For the full length, the percentages with lowest cumulative error (six) are between 0.4% and 0.55%. For the mini-barcode, the lowest cumulative error (also six) are within 0.55% and 1%.

The ABGD partition of 70 groups (Fig. 3.aIII) of delimitated species was in agreement with the NJ clusters (Supplementary Fig. S1).

The threshold analysis for species delimitation identified ranged from 0.4% up to 0.55% as the intraspecific values with the lowest number of cumulative errors (six). We used 0.4% as it is the most conservative percentage (Fig. 4a). Using the 0.4% threshold for species delimitation analysis (*threshID*), we recovered 70 MOTUs (Fig. 3aIV) within the 67 morpho-species previously identified by Carvalho et al. (2011). The overestimated three MOTUs are *Imparfinis minutus*, *Piabina argentea,* and *Pamphorichthys hollandi*.

Species delimitation analyses of the mini-barcodes (193 bp). The mini-barcode intraspecific genetic distance ranged from 0.0% to 1.14% (mean: 0.08%%), while the interspecific distances ranged from 1.14% to 49.03% (mean: 17.67%). Distance values for intrageneric ranged from 0.0% to 9.32% (mean: 1.03%) and intrafamilial ranged from 0.0% to 10.87% (mean: 3.37%). The NJ dendrogram generated with all specimens showed species-specific clades (Supplementary Fig. S1).

In silico species delimitation analyses based on Bayesian approaches, GMYC, and bPTP were able to recover all 67 species previously identified using traditional morphology-based identification. GMYC model recovered 70 genetic MOTUs (interval 68–71) (Fig. 3bI). The threshold time was −0.005 and the ML for the null and GMYC model were 780.9562 and 821.8, respectively. The bPTP analysis revealed a total of 76 MOTUs (Fig. 3bII), with branch support ranging from 0.091 to 0.994.

ABGD was able to recover 59 to 67 groups when varying the prior maximal distance from $P = 0.021$ to $P = 0.001$. Five partitions, with prior maximal intraspecific distances ranging from 0.001 to 0.007, recovered 67 groups within our 12S mini-barcode database (Fig. 3bIII). The ABGD partition of 67 groups could delimitate most species in agreement with the NJ clusters (Supplementary Fig. S1); however, one group combined more than one species: (1) *Pygocentrus piraya* and *Serrasalmus brandtii* even though interspecific divergences could clearly differentiate this species (3.51%).

The intraspecific values with the lowest number of cumulative errors in the threshold analysis for species delimitation (six) were 0.55% up to 1%. We used 0.55%, which is the most conservative percentage in this case (Fig. 4b) and recovered 70 MOTUs (Fig. 3bIV) with this value. The overestimated three MOTUs (identified as "no ID" by *threshID* function) are *Imparfinis minutus*, *Piabina argentea,* and *Pamphorichthys hollandi*.

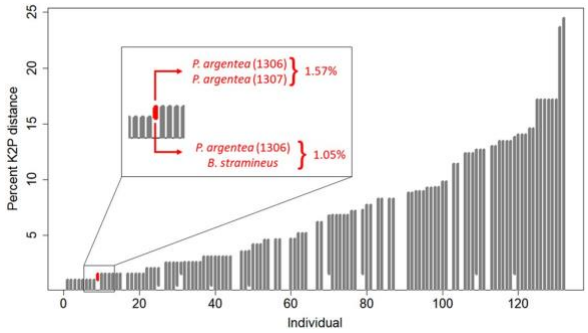


Figure 5. Barcode gap line plot for the 132 fish specimens. For each specimen in the dataset, the grey bars represent the gap between the highest intraspecific distance (bottom of the bar) and the lowest interspecific distance (top of the bar) representing the barcoding gap range. The red bar represents a specimen which these values overlap (intraspecific is higher than interspecific) meaning there is no barcoding gap.

Distance-based analyses performed using SPIDER showed similar results. In nearest-neighbor analysis, 99.2% of the sequences (112 out of 113—excluding 19 singletons) were correctly clustered, with only *P. argentea* (1306) being incorrectly clustered as nearest-neighbor of *Bryconammericus stramineus*. Barcoding gap analysis successfully recovered all species, since no overlap of intra and interspecific divergence was observed, except for *P. argentea* (1306) that has an intraspecific divergence of 1.57% with *P. argentea* (1307) and interspecific distance of 1.05% with *B. stramineus* specimens (Fig. 5).

Discussion

We developed and curated a reference database for 67 fish species, belonging to 54 genera that are widespread across the Neotropical realm, and used it to develop a 12S mini-barcode marker and estimate a genetic distance threshold value for Neotropical fish species delimitation. Having a reference database associated with mini-barcode primer sets specific for Neotropical species is an important asset for DNA metabarcoding, especially when analyzing eDNA samples from such megadiverse fauna[21,22].

The taxonomic resolution of 12S full and mini barcodes libraries provided enough molecular polymorphism to differentiate all 67 morpho-species. Moreover, the 12S full-length barcode (ca. 565 bp) was sufficient to discriminate all 70 MOTUs, which was in accordance with previous molecular (COI based) identifications of the same specimens[28]. Interestingly, the mini-barcode region's (i.e. 193 bp—NeoFish_3) taxonomic resolution performed similarly to the full-length database, providing the same number of MOTUs when applying the GMYC and genetic distances thresholds analyses (70 MOTUs). The other analyses of the mini-barcode dataset overestimated the number of MOTUs (bPTP with 76) or underestimated it (ABGD with 67 MOTUs).

When performing genetic distance threshold analysis using the full-length library, we obtained a threshold value (0.40%, Fig. 4a) similar to our mini-barcode region (0.55%, Fig. 4b). Fish species delimitation threshold values based on the 12S region are an important reference for future studies using this marker, but they may need to establish a priori reference value when interpreting genetic distance data, such as the 2% widely used for COI[53]. Although we have analyzed several genera from all major Neotropical fish taxa, it is important to note that its value will be more robust and better reflect the real divergence between species when more species are added to our reference database.

Species delimitation and taxonomic resolution analyses revealed the potential of NeoFish_3 amplicons to reliably identify species, since there was no relevant disparity between full-length and mini barcode libraries for these analyses. Similar results were obtained for the COI gene, as a comparison between full-length and mini barcodes, especially when it was used in degraded samples. This demonstrates that the latter is informative for species-level sorting of: (1) major eukaryotic groups and archival specimens[45]; (2) moth and wasp museum specimens[54], and; (3) several bird species[55]. However, few congeneric species have been analyzed in this study, and thus, to overcome this putative drawback, future analyses should include a higher number of species from the same genus to provide even more robust results.

SWAN analysis showed that the target NeoFish_3 amplicon would be the best region for taxonomic differentiation of species since it recovered the best indices in all established criteria (Fig. 2). However, we did not analyze the whole 12S gene of all species to proper compare the NeoFish_3 to other previously used amplicons (MifishU and Teleo1) using characteristics such as taxonomic resolution and best primer site. The target 12S rRNA gene region used to build our reference database represents approximately 60% of the 12S full-length gene (952 bp) (Fig. 1a) and includes only a small fragment of the 12S region amplified by the MiFishU marker and also the initial region of the forward Teleo1 (Fig. 1b).

In vitro tests showed that the newly developed NeoFish_3 marker is efficient and thus, was able to amplify the target region of the 12S rRNA gene from 22 tissue DNA extracts and environmental DNA recovered from an aquarium containing one fish species (Supplementary Table S1; Fig. S1). However, further evaluation of amplification success with samples obtained from Neotropical river basins using a DNA metabarcoding approach

for a whole fish community is recommended, as different types of environmental samples will vary in patterns of DNA degradation and exposure to inhibitors[33]. Although 67 fish species represent a low percentage of the Neotropical freshwater fish species, they nevertheless account for the main Neotropical orders, since we include DNA of species from Characiformes, Cyprinodontiformes, Gymnotiformes, Perciformes, Siluriformes, and Synbranchiformes.

Amplification of non-target organisms has been previously reported as a drawback of universal eDNA available primer sets that led to the use of human blocking primers to avoid cross amplification. When comparing amplification of non-target taxa to previously designed primers sets (Teleo1 and MiFishU), a better specificity of NeoFish_3 was detected with our in silico PCR analysis. For Teleo1 and MiFishU the amplification rate for Mammalia, including *Homo sapiens*, was over 1000 sequences (Table 2), while the NeoFish_3 had no cross amplification of these. Moreover, when using the Teleo1 and MiFishU markers to assess fish communities diversity in French Guiana[21] and Japan[31], both papers report amplification of DNA from insects and mammals when analyzing eDNA samples. Such untargeted amplification and detection in eDNA studies may hamper the identification of rare species since it may consume most of the DNA sequences obtained[29,56]. However, before assuming that NeoFish_3 outperformed other 12S mini-barcode markers, in situ tests would be needed to check if there would indeed be lower amplification of non-targeted species.

Herein, we applied a powerful framework for the development and validation of a fish-specific primer set together with a custom reference database aimed at DNA metabarcoding analysis in the Neotropical realm. Species delimitation analyses strongly suggest that even when using a short region of the 12S mitochondrial region, we could discriminate each taxon to the species level. In addition, we were able to set an interspecific distance-based threshold for species delimitation that would be helpful throughout bioinformatics metabarcoding short reads analysis. Thus, our custom reference database and mini-barcodes markers are an important asset for an ecoregion scale DNA based biodiversity evaluation, such as eDNA metabarcoding, that can help with the complex task of conserving the megadiverse Neotropical ichthyofauna.

Data availability

The newly generated sequences are available at GenBank under accession numbers MG755503 – MG755639.

Received: 29 May 2020; Accepted: 18 September 2020
Published online: 21 October 2020

References

1. Cardinale, B. J. *et al.* Biodiversity loss and its impact on humanity. *Nature* **486**, 59–67 (2012).
2. WWF. *Living Planet Report - 2018: Aiming higher.* (WWF International, 2018).
3. Kelly, R. P. *et al.* Harnessing DNA to improve environmental management. *Science* **344**, 1455–1456 (2014).
4. Bonar, S. A., Hubert, W. A. & Willis, D. W. Standard methods for sampling North American freshwater fishes (2009).
5. Wheeler, Q. D., Raven, P. H. & Wilson, E. O. Taxonomy: impediment or expedient?. *Science (New York, NY)* **303**, 285 (2004).
6. Kelly, R. P., Port, J. A., Yamahara, K. M. & Crowder, L. B. Using environmental DNA to census marine fishes in a large mesocosm. *PLoS ONE* **9**, e86175 (2014).
7. Valentini, A. *et al.* Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Mol. Ecol.* **25**, 929–942 (2016).
8. McDevitt, A. D. *et al.* Environmental DNA metabarcoding as an effective and rapid tool for fish monitoring in canals. *J. Fish Biol.* **95**, 679–682 (2019).
9. Taberlet, P., Coissac, E., Hajibabaei, M. & Rieseberg, L. H. Environmental DNA. *Mol. Ecol.* **21**, 1789–1793 (2012).
10. Deiner, K. *et al.* Environmental DNA metabarcoding: transforming how we survey animal and plant communities. *Mol. Ecol.* **26**, 5872–5895 (2017).
11. Nobile, A. B. *et al.* DNA metabarcoding of neotropical ichthyoplankton: enabling high accuracy with lower cost. *Metabarcoding Metagenomics* **3**, e35060 (2019).
12. Mariac, C. *et al.* Metabarcoding by capture using a single COI probe (MCSP) to identify and quantify fish species in ichthyoplankton swarms. *PLoS ONE* **13**, e0202976 (2018).
13. Leray, M., Meyer, C. P. & Mills, S. C. Metabarcoding dietary analysis of coral dwelling predatory fish demonstrates the minor contribution of coral mutualists to their highly partitioned, generalist diet. *PeerJ* **3**, e1047 (2015).
14. Shokralla, S. *et al.* Massively parallel multiplex DNA sequencing for specimen identification using an IlluminaMiSeq platform. *Sci. Rep.* **5**, 9687 (2015).
15. Kitano, T., Umetsu, K., Tian, W. & Osawa, M. Two universal primer sets for species identification among vertebrates. *Int. J. Legal Med.* **121**, 423–427 (2007).
16. Stoeckle, M. Y., Soboleva, L. & Charlop-Powers, Z. Aquatic environmental DNA detects seasonal fish abundance and habitat preference in an urban estuary. *PLoS ONE* **12**, e0175186 (2017).
17. Sales, N. G. *et al.* Fishing for mammals: landscape-level monitoring of terrestrial and semi-aquatic communities using eDNA from riverine systems. *J. Appl. Ecol.* **57**, 707–716 (2020).
18. Bylemans, J. *et al.* An environmental DNA-based method for monitoring spawning activity: a case study, using the endangered Macquarie perch (*Macquaria australasica*). *Methods Ecol. Evol.* **8**, 646–655 (2017).
19. De Souza, L. S., Godwin, J. C., Renshaw, M. A. & Larson, E. Environmental DNA (eDNA) detection probability is influenced by seasonal activity of organisms. *PLoS ONE* **11**, e0165273 (2016).
20. Dejean, T. *et al.* Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog *Lithobates catesbeianus*. *J. Appl. Ecol.* **49**, 953–959 (2012).
21. Cilleros, K. *et al.* Unlocking biodiversity and conservation studies in high-diversity environments using environmental DNA (eDNA): a test with Guianese freshwater fishes. *Mol. Ecol. Resour.* **19**(1), 27–46. https://doi.org/10.1111/1755-0998.12900 (2018).
22. Sales, N. G., Wangensteen, O. S., Carvalho, D. C. & Mariani, S. Influence of preservation methods, sample medium and sampling time on eDNA recovery in a neotropical river. *Environ. DNA* **1**(2), 119–130. https://doi.org/10.1002/edn3.14 (2019).
23. Sales, N. G. *et al.* Assessing the potential of environmental DNA metabarcoding for monitoring Neotropical mammals: a case study in the Amazon and Atlantic Forest, Brazil. *Mamm. Rev.* **50**, 221–225 (2020).
24. Dejean, T. *et al.* Persistence of environmental DNA in freshwater ecosystems. *PLoS ONE* **6**, e23398 (2011).
25. Gomes, L. C., Pessali, T. C., Sales, N. G., Pompeu, P. S. & Carvalho, D. C. Integrative taxonomy detects cryptic and overlooked fish species in a neotropical river basin. *Genetica* **143**, 581–588 (2015).

26. Pugedo, M. L., de Andrade Neto, F. R., Pessali, T. C., Birindelli, J. L. O. & Carvalho, D. C. Integrative taxonomy supports new candidate fish species in a poorly studied neotropical region: the Jequitinhonha River Basin. *Genetica* **144**, 341–349 (2016).
27. Ramirez, J. L. *et al.* Revealing hidden diversity of the underestimated NeotropicalIchthyofauna: DNA barcoding in the recently described genus *Megaleporinus* (Characiformes: Anostomidae). *Front. Genet.* **8**, 1–11 (2017).
28. Carvalho, D. C. *et al.* Deep barcode divergence in Brazilian freshwater fishes: the case of the São Francisco River basin. *Mitochondrial DNA* **22**, 80–86 (2011).
29. Collins, R. A. *et al.* Non-specific amplification compromises environmental DNA metabarcoding with COI. *Methods Ecol. Evol.* **10**, 1985–2001 (2019).
30. Shaw, J. L. A. *et al.* Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system. *Biol. Conserv.* **197**, 131–138 (2016).
31. Yamamoto, S. *et al.* Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. *Sci. Rep.* **7**, 40368 (2017).
32. Miya, M. *et al.* MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *R. Soc. Open Sci.* **2**, 150088 (2015).
33. MacDonald, A. J. & Sarre, S. D. A framework for developing and validating taxon-specific primers for specimen identification from environmental DNA. *Mol. Ecol. Resour.* **17**, 708–720 (2017).
34. Aljanabi, S. M. & Martinez, I. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Res.* **25**, 4692–4693 (1997).
35. Thomsen, P. F. *et al.* Environmental DNA from seawater samples correlate with trawl catches of subarctic deepwater fishes. *PLoS ONE* **11**, e0165252 (2016).
36. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
37. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
38. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120 (1980).
39. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees'. *Mol. Biol. Evol.* **4**, 406–425 (1987).
40. Felsenstein, J. Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. *Evolution (N. Y.)* **35**, 1229–1242 (1981).
41. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
42. Proutski, V. & Holmes, E. SWAN: sliding window analysis of nucleotide sequence variability. *Bioinformatics* **14**, 467–468 (1998).
43. Brown, S. D. J. *et al.* Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Mol. Ecol. Resour.* **12**, 562–565 (2012).
44. R Core Team. *R: A Language and Environment for Statistical Computing* (2020).
45. Meusnier, I. *et al.* A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics* **9**, 214 (2008).
46. Ye, J. *et al.* Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinform.* **13**, 134 (2012).
47. Fujisawa, T. & Barraclough, T. G. Delimiting species using single-locus data and the generalized mixed yule coalescent approach: a revised method and evaluation on simulated data sets. *Syst. Biol.* **62**, 707–724 (2013).
48. Zhang, J., Kapli, P., Pavlidis, P. & Stamatakis, A. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* **29**, 2869–2876 (2013).
49. Puillandre, N., Lambert, A., Brouillet, S. & Achaz, G. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Mol. Ecol.* **21**, 1864–1877 (2012).
50. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
51. Rambaut, A., Suchard, M. A., Xie, D. & Drummond, A. J. *Tracer 1.6* http://beast.bio.ed.ac.uk/tracer (2014).
52. Rambaut, A. & Drummond, A. J. *TreeAnnotator, version 1.7. 5*. Available beast. bio. ed. ac. uk/TreeAnnotator (accessed 15 April 2010) (2012).
53. Ward, R. D. DNA barcode divergence among species and genera of birds and fishes. *Mol. Ecol. Resour.* **9**, 1077–1085 (2009).
54. Hajibabaei, M. *et al.* A minimalist barcode can identify a specimen whose DNA is degraded. *Mol. Ecol. Notes* **6**, 959–964 (2006).
55. Yu, H.-J. & You, Z.-H. Comparison of DNA truncated barcodes and full-barcodes for species identification. in *International Conference on Intelligent Computing* 108–114 (Springer, 2010).
56. Harper, L. R. *et al.* Environmental DNA (eDNA) metabarcoding of pond water as a tool to survey conservation and management priority mammals. *Biol. Conserv.* **238**, 108225 (2019).

Acknowledgements

This study was financially supported by CEMIG (P&D GT0635), FAPEMIG, and CNPq. DM was supported by a master's fellowship from FAPEMIG (5236/15). DC is grateful to a CNPq fellowship (306155/2018-4). NGS is grateful to FCT/MCTES for the financial support to CESAM (UIDP/50017/2020+UIDB/50017/2020), through national funds.

Author contributions

D.T.M., I.S.M., J.S.D., and D.F.T. carried out research in the lab. D.T.M., I.S.M., and N.G.S. carried out bioinformatic data analysis. D.C.C. supervised and obtained funding. D.T.M., I.S.M., and D.C.C. wrote the manuscript with significant contributions from all the authors. D.T.M. and I.S.M. contributed equally to this study. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-74902-3.

Correspondence and requests for materials should be addressed to D.C.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020

5. CAPÍTULO II - DNA metabarcoding of mock communities highlights potential biases when assessing Neotropical fish diversity (submitted on Molecular Ecology Resources, 2023)

5.1 Objetivos específicos

- Entender melhor o uso do *metabarcoding* na detecção da diversidade de peixes neotropicais pelo uso de três conjuntos de *mini-barcodes*;
- Avaliar a influência da proporção de *input* de DNA em amostras de comunidades simuladas na identificação de espécies;
- Avaliar a influência da escolha do marcador molecular na detecção de espécies em comunidades simuladas;
- Comparar o desempenho da identificação taxonômica usando ASV/MOTUs e a atribuição de leituras de DNA a espécies usando uma biblioteca referência customizada;
- Avaliar a complementaridade de *primers* na detecção de espécies;
- Investigar espécies preferencialmente amplificadas pelos *primers* testados.

Molecular Ecology Resources

MOLECULAR ECOLOGY
RESOURCES

DNA metabarcoding of mock communities highlights potential biases when assessing Neotropical fish diversity

Journal:	*Molecular Ecology Resources*
Manuscript ID	Draft
Manuscript Type:	Resource Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Hilário, Heron; Pontifícia Universidade Católica de Minas Gerais, Programa de Pós Graduação em Biologia dos Vertebrados Mendes, Izabela; UFMG, Genetics Sales, Naiara; Universidade de Lisboa, Departamento de Biologia Carvalho, Daniel; Pontifícia Universidade Católica de Minas Gerais, Programa de Pós-graduação em Zoologia de Vertebrados
Keywords:	Biodiversity, 12S gene, ASV, MOTU, freshwater fish, bioinformatics

SCHOLARONE™
Manuscripts

DNA metabarcoding of mock communities highlights potential biases when assessing Neotropical fish diversity

Heron Oliveira Hilário[1*], Izabela Santos Mendes[1,3*], Naiara Guimarães Sales[2], Daniel Cardoso Carvalho[1,3]

[1] Conservation Genetics Lab, Postgraduate Program in Vertebrate Biology, Pontifical Catholic University of Minas Gerais, PUC Minas, Belo Horizonte, Brazil

[2] School of Science, Engineering and Environment, University of Salford, Crescent, Salford, UK

[3] Postgraduate Program in Genetics, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Brazil

[*]HOH and ISM should be considered joint first author

Paper submitted on Molecular Ecology Resources (currently under review).

ABSTRACT

Despite the increasing popularity of DNA metabarcoding in the assessment of aquatic ecosystems using fish eDNA or ichthyoplankton, challenges have hampered its broader application in the Neotropical freshwaters. Using five mock communities composed of fish species from two Neotropical river basins, we evaluated the influence of DNA concentration and choice of mitochondrial 12S molecular markers (MiFish, NeoFish and Teleo) on species detection and Relative Read Abundance (RRA) using DNA metabarcoding. Of the three 12S markers analysed, only MiFish detected all species from all mock communities. The performance of a taxonomy-free approach using ASV/MOTUs was not as precise as assigning DNA reads to species using a curated 12S library that includes approximately 100 fish species, since more than one ASV/MOTU was observed for the same specimen. Thus, here we showcase the importance of a custom reference database to allow precise assignment of Neotropical fish species in metabarcoding studies and that the RRA is dependent on community composition, marker and DNA concentration. We highlight the importance of controlled experiments using known species communities before large investments are made in assessing biodiversity using non-invasive methods that apply DNA metabarcoding.

Keywords: Biodiversity, 12S marker, ASV, MOTU, freshwater fish, bioinformatics

5.2 INTRODUCTION

Metabarcoding has been recognised as a powerful tool for non-invasive biomonitoring and management of biodiversity (Deiner et al., 2017; Boivin-Delisle et al., 2021), including megadiverse realms such as the Amazonian ichthyofauna (Castelo et al., 2022). The use of universal primers and high-throughput sequencing have enabled the assessment of biodiversity and community composition for target taxa, from many types of environmental samples, unravelling patterns linked to environmental variables or human impacts (Deiner et al., 2017; Elbrecht & Leese, 2015; Hänfling et al., 2016; Riaz et al., 2011).

Despite the increasing popularity of this method in the assessment of aquatic ecosystems using fish eDNA or ichthyoplankton pools (Cilleros et al., 2018, Sales et al. 2019, Sales et al. 2021ab), the vast majority of fish metabarcoding studies are still concentrated in temperate regions, in well-characterised and reasonably accessible environments (Handley et al., 2019; McDevitt et al., 2019). Considering the highly diverse Neotropical ichthyofauna, appropriate implementation of metabarcoding biomonitoring is hindered by a lack of knowledge of the local biodiversity (e.g., undescribed or cryptic species), primer biases, and the incompleteness of reference databases (Sato et al., 2017, Sales et al., 2018, Jackman et al., 2021). Additionally, the multiple bioinformatic pipelines used for sequence processing and taxonomic identification (Cilleros et al., 2018, Sales et al., 2019, Jackman et al., 2021) may also influence the ultimate resolution of species detected in a given sample (Majaneva et al., 2018).

Since its adoption as the marker of choice for DNA barcoding of animal species, a vast amount of sequence, taxonomic and metadata information has been amassed on the cytochrome oxidase subunit 1 (cox1 or COI), especially through the efforts of the ‘Consortium for the Barcode of Life’ (Porter & Hajibabaei, 2018). Despite this knowledge, the internal low conservation of COI markers hampers the development of shorter mini-barcodes, and usually requires primer cocktails to enable balanced polymerase chain reaction (PCR) amplification of all target fish species, which also results in side amplification of bacteria and other non-target taxa (Collins et al., 2019). In addition, amplification of the COI DNA barcoding region of approximately 650 bp is not always compatible with metabarcoding surveys considering that the most widely used sequencing platforms output short reads, which cover amplicons of 250 to

500 bp, leading to increased difficulty in analysing larger fragments (Kocher et al., 2017).

Ideally, DNA markers used for metabarcoding must also correspond to a highly polymorphic locus and also facilitate the design of highly specific primers (Deagle et al., 2014). Concerning fish metabarcoding, the 12S rRNA mitochondrial gene has emerged as an alternative marker to COI, leading to the development of many primers targeting this region (Zhang et al., 2020). For instance, Miya et al. designed the MiFish markers for fish eDNA amplification, based on 880 mitochondrial genomes of mostly marine fishes (2015), while Valentini et al. (2016) developed the Teleo marker, using all fish sequences from the EMBL (European Molecular Biology Laboratory). Together, these two primer sets represented the marker of choice in over 52% of recent studies involving fish eDNA (Xiong et al., 2022). More recently, aiming to investigate Neotropical fish biodiversity, Milan et al. (2020) developed the NeoFish marker, based on the 12S region of 132 fishes of the Neotropical São Francisco River Basin. These three markers target the 12S gene in adjacent regions and were initially considered able to contain enough genetic variation to allow the differentiation of most fish species. However, Mifish and Teleo have already been demonstrated to fail in the detection of some fish taxa (Zhang et al., 2020, Jackman et al., 2021) and, to this date, no comprehensive and comparative evaluation of their amplification or taxonomic classification for fish from megadiverse Neotropical rivers has been conducted, which hinders informed decisions regarding their suitability for different study systems.

Regarding the processing of sequencing data and taxonomic assignment, eDNA studies of the Neotropics have mostly relied on clustering techniques based on MOTUs (Molecular Operational Taxonomic Units), with the application of a 97% sequence similarity threshold for species identification (e.g., Cilleros et al., 2018, Sales et al., 2021). More recently, studies have started to demonstrate the need of including distinct approaches (i.e., ZOTUs, ASVs) to describe the recovered eDNA diversity (Dal Pont et al., 2021). Clustering methods merge a set of sequences into meaningful biological identities, which are hence considered as a proxy for species. Alternatively, the Amplicon Sequence Variants (ASVs) (or ESVs, Exact Sequence Variants) have been proposed to provide a finer resolution, allowing the detection of single-nucleotide differences over the sequenced gene region (Callahan et al., 2017). Sometimes interpreted as intraspecific variability, which can be a desirable output of eDNA, ASVs may provide additional ecological signals that can be extrapolated from when

considering species (García-García et al., 2019). This sequence diversity might provide an increased resolution but could also be prone to harbour sequencing errors (e.g. errors originating from PCR amplification and sequencing artefacts).

Recently, the importance of combining both denoising and clustering procedures has been raised by Antich et al. (2021) when considering the standard barcode fragment of COI. Yet, to our knowledge, an evaluation of applying both methods using mock communities to uncover the real advantages of each approach has still not been applied to other markers (i.e., 12S gene) targeting the Neotropical megadiverse fish communities. The strengths and weaknesses of the molecular markers used for eDNA metabarcoding can be revealed through the investigation of mock communities which enables the identification not only of species that failed to be amplified but also the extent to which biased amplification occurs.

In this study, we aim to better understand the use of metabarcoding in the detection of Neotropical fish diversity by evaluating the influence of the following features of sample composition, using fish mock communities: (1) how the proportion of input DNA and (2) molecular marker choice influences the detection of species in mock communities and (3) to compare the performance of taxonomy-free approach using ASV/MOTUs and assignment of DNA reads to species using a curated library (Figure 1). Moreover, a robust database for the 12S region was built to enable metabarcoding studies of eastern shield Neotropical species that can be used to evaluate primers complementarity on species detection and to investigate preferentially amplified species. Our results highlighted the importance of controlled experiments using known species communities and the availability of a robust reference database for metabarcoding studies, especially in megadiverse realms.

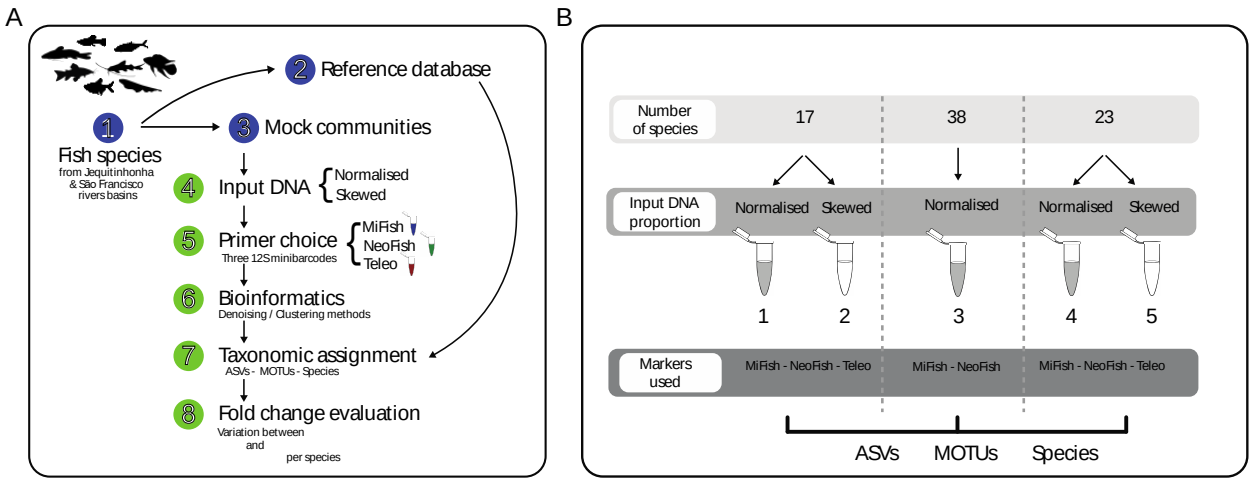


Figure 1. Study framework (A) and description of analysed mock communities including different species composition and DNA input, markers evaluated, and proxies used for biodiversity assessment (B). All mock communities built using species from the São Francisco River Basins (SFRB) and Jequitinhonha River Basin (JQRB). (1) normalised São Francisco River mock community (SFmc) comprising 23 species from the SFRB with the same concentration of DNA (10 ng/µl) and (2) SFmc skewed using different DNA concentrations. (3) São Francisco and Jequitinhonha combined mock community (SFJQmc) built with 38 unique species from the normalised Jequitinhonha River mock community (JQmc) JQmc and SFmc. (4) normalised Jequitinhonha River mock community (JQmc) composed of 23 species from the JQRB built using equal concentrations of DNA. (5) JQmc skewed: mock community composed of 23 species from the JQRB built with skewed concentrations of DNA.

5.3 MATERIAL AND METHODS

5.3.1 Building a local 12S reference database

A custom reference database was developed using previously molecularly identified specimens using DNA Barcoding and traditional taxonomy (Carvalho et al., 2011; Pugedo et al., 2016) deposited at the Conservation Genetics Lab collection (PUC Minas University). The 12S reference library was built using a DNA fragment of 600 bp obtained by Sanger DNA sequencing as described in Milan et al., 2020. The targeted 12S fragment includes the region used by three 12S markers: (1) NeoFish (~190 bp, Milan et al., 2020); (2) MiFish (~170 bp, Miya et al., 2015); and (3) Teleo (~70 bp, Valentini et al. 2016).

The Conservation Genetics Lab 12S Sequence Database (LGC 12Sdb hereafter) comprises 187 sequences representing 99 fish species (Table S1) collected from the Jequitinhonha (JQ) and São Francisco (SF) River Basins. Considering the Jequitinhonha River Basin, 40 species are represented by 53 sequences obtained in this study, (complete description of the fish collection, sample storage and processing available in Supplementary Material). From the São Francisco River Basin, 70 fish species are represented by 134 sequences (corresponding to sequences obtained in

this study and combined with data retrieved from Milan et al., 2020). The 12S sequences were deposited into GenBank and are available under the accession numbers: XXXXXX–XXXXXX.

5.3.2 Assemblage of mock communities

DNA extraction was carried out using samples of the same tissues used for the construction of the LGC 12Sdb, with the Wizard® Genomic DNA Purification Kit (Promega). DNA integrity was assessed using 2% agarose gel electrophoresis, and DNA concentration was determined using QubitTM 1x dsDNA HS assay (Thermo Fisher).

All mock communities were constructed using known ratios of input DNA, including DNA from fish specimens also included in the aforementioned reference database (Figure 1). Normalised mock communities contained all DNA extracts pooled at a concentration of 10 ng/µl DNA per analysed species, whereas skewed mock communities contained the original DNA concentration obtained from the tissue DNA extractions (corresponding to between 0.26% - 51.58% of the total DNA, Table S2). A total of five mock communities were analysed, including two composed of 17 species collected from the Jequitinhonha River Basin (JQRB), namely JQmc (Table S2), of which one was constructed using equal concentrations of DNA (JQmc normalised) and the other one based on skewed concentrations of DNA (JQmc skewed). Two other mock communities were constructed comprising 23 species from the São Francisco River Basin (SFRB) (SFmc, Table S2), also including one normalised (SFmc normalised) and one skewed version (SFmc skewed). A fifth mock community contained 38 unique species used to build the SFmc and JQmc combined and was normalised (SFJQmc, Table S2).

5.3.3 Amplification, library preparation and sequencing

The JQmc was amplified with all three markers (NeoFish, MiFish and Teleo) (Table S3); SFmc and SFJQmc were both amplified only with NeoFish and MiFish primers because they have similar amplicon sizes and are suitable for combining in a single HTS Illumina run. Each mock community was amplified independently in three replicates for each marker. Each sample/primer combination was first amplified in 25µl PCR triplicates using the high-fidelity AmpliTaq Gold (Applied Biosystems). To account for possible tag jumping and contamination, one positive control, not included in any

of the analysed mock communities, was used including a marine cartilaginous fish DNA (*Carcharinus glaucus*).

Twelve libraries were prepared using the three markers, NeoFish, MiFish and Teleo. Markers were designed with 5' Illumina library construction adaptors, and a two-step PCR procedure was used for library preparation using the Nextera XT kit (Illumina) following the manufacturer's instructions. In the first step, a fragment of the mitochondrial 12S rRNA gene was amplified using the MiFish, NeoFish, and Teleo markers. The first PCR reaction was performed on a 25 µl final volume, using AmpliTaq Gold, with 2 µl DNA. The thermal cycles for this step were as follows: initial denaturation at 95°C for 2 minutes, followed by 35 cycles of denaturation at 95°C for 1 minute, annealing at 60°C (NeoFish), 50°C (MiFish) and 55°C (Teleo) for 30 seconds and elongation at 72°C for 1 minute, followed by final elongation at 72°C for 7 minutes. PCR products were purified using Agencourt Ampure beads (Beckman Coulter) according to the manufacturer's instructions and used as templates for the second PCR. The Illumina sequencing adaptors plus the 8-bp identifier indices were added in the subsequent PCR using a forward and reverse fusion primer (Table S4). The second PCR reaction was performed on a 25 µl final volume, containing 10µl 2.51×KAPA HiFi HotStart ReadyMix (KAPA Biosystems, Wilmington, WA, USA), 5 µl of each primer, and 5 µl of the first PCR products. The thermal cycles of the second PCR were: initial denaturation at 95°C for 3 minutes, followed by eight cycles of denaturation at 98°C for 20 seconds, annealing and elongation combined at 72°C for 15 seconds, with a final elongation at 72°C for 5 minutes. DNA concentration was estimated using the Qubit 1x dsDNA HS assay kit and a Qubit fluorometer (Thermo Fisher Scientific). Libraries were pooled in equimolar concentrations and sequenced on the MiniSeq platform (Illumina) using the Mid-output Kit and 2×150 bp paired-end sequencing following the manufacturer's instructions.

5.3.4 Bioinformatic analyses

All bioinformatics analyses were conducted in R v.4.1.1 (R Core Team, 2021), using a pipeline based on DADA2 v1.16.0 (Callahan et al., 2016) and phyloseq v1.34.0 (McMurdie & Holmes, 2013). R scripts are fully available on Github (https://github.com/heronoh/fish_eDNA). Raw sequence data is available at https://zenodo.org/.

In brief, reads were submitted to the removal of undetermined bases and quality filtering (Q-scores >= 30). Only reads containing the expected primer sequence corresponding to each sample were kept for downstream analysis. Error correction, read-pair merging, and chimaera identification and removal were performed using the default settings of DADA2 functions. No length truncation was performed since primer removal automatically clips uninformative regions, but the resulting ASVs beyond the expected amplicon length range for each marker were discarded (NeoFish: < 185 & > 200; MiFish: < 165 & > 180; Teleo: < 60 & > 75). Subsequently, identified ASVs were clustered into MOTUs using SWARM v3.1.0 (Mahé et al., 2014 & 2015), using the fastidious option and d=1. Taxonomic assignments were conducted using the DADA2 RDP classifier and the LGC 12Sdb. Additionally, all ASVs were searched against local NCBInr using an automated BLASTn 2.10.1+ function with minimum similarity and minimum coverage (-perc_identity 95 and -qcov_hsp_perc 95). After taxonomic assignment and in order to curate and correct partial assignments and unexpected species, phylogenetic trees were built for each marker using the correspondent ASVs and the marker amplicon region for all database species. Sequences were aligned using MUSCLE and a Maximum-likelihood tree was generated using MEGAX.

The relative read abundance (RRA) was obtained for each ASV, dividing the number of copies of the ASVs by the sum of the number of copies of all ASVs in the sample. A test of significance was conducted using an analysis of similarity (ANOSIM) to compare primer's depiction of both normalised and skewed versions of the same mock community, and the relationship between RRA and input DNA concentration for each sample was further assessed through Pearson's correlation coefficient. Linear regression models between the proportion of input DNA and RRA were generated for each primer, considering each species separately, to evaluate biases on amplification (Tables S5A to S5C). It was not possible to evaluate the difference between normalised and skewed mock communities since the narrow distribution of the input DNA on the normalised versions prevented reliable regressions.

We then used the actual proportion of the corresponding species' DNA in the respective mock community to calculate a fold change between DNA input and the resulting RRA for each species, using the formula (frac{num}{denom}) if (num>denom), and as (frac{-denom}{num}) if (num<denom), and the resulting values were used to generate bar plots and boxplots.

The total counts of ASVs, MOTUs and correctly identified species were used to compare diversity estimates for each mock community and primer. A fold change between -1 and 1 corresponds to a 1:1 relative proportion of input DNA and sequence counts on samples. A fold change < -1 indicates a lower abundance of the species among the sequences than its actual abundance in the mock community input DNA. Fold change values >1 indicate that the species is more represented on the sequences. For example, a fold change of 10 indicates that the species was ten times more abundant on the ASVs than it was on the input DNA. An optimal primer pair would recover fold change values close to zero, with a minimum variation between taxa. It is important to highlight that all the markers we targeted here occur in multiple copies in each cell which can skew read abundances (Krehenwinkel et al. 2017). Hence, fold change values of zero will rarely be found, even in the absence of amplification bias. However, as amplification bias amplifies exponentially with PCR cycle number, its effect on fold change should be more severe than simple copy number variation.

5.4 RESULTS

5.4.1 12S reference database

The updated 12S reference database (LGC 12Sdb) is comprised of 187 specimens from the São Francisco (SF) and Jequitinhonha (JQ) River Basins, with sequences ranging from 568 to 721 bp (663 bp on average), representing 73 genera, 28 families, 9 orders, and 99 species; 15 species of which are shared between the basins and are represented by different specimens from each drainage. Considering the known ichthyofauna, the species diversity of the database corresponds to approximately 75% for the JQ and 46% for the SF River basins. The database encompasses the amplicon regions of the three 12S markers NeoFish, MiFish, and Teleo but, for the latter two, some species are partially represented after trimming low-quality bases on the ends of the alignment.

Concerning the species used in the mock communities, the database covers 100% of the NeoFish amplicon, an average of 86% of the MiFish amplicon total length (min: 69.6%, max: 97.24%, on the 5’ end), and 88.89% of the Teleo amplicon (min: 81.49%, max: 97.23%, on the 3’ end) (Figure S1). These reference sequences, despite their length, allowed the correct identification of almost all species included in the mock communities and thus, the absence of full sequences did not hinder correct species

assignment. The exceptions were for two species pairs: *Prochilodus argenteus* and *P. hartii*, and *Hoplias brasiliensis* and *H. intermedius*, which shared identical amplicon sequences for all analysed markers despite complete amplicon sequencing. Therefore, due to this hindrance in the taxonomic resolution, each species pair was considered as a single entity (*P. argenteus*/*P. hartii* and *H. brasiliensis*/*H. intermedius*) leading to a total of 38 molecular identifiable species. Thus, for all further analyses (primer performance; RRA vs input DNA, and others) each of these species' pairs was considered as a single entity.

5.4.2 Data quality control and processing

A total of 2.66 million raw reads were obtained for the five mock communities (1.34 million - NeoFish, 0.92 million - MiFish, and 0.39 million - Teleo). After raw data processing to remove undetermined bases and primers, read merging, chimaera removal, and error correction, a total of 1.57 million sequences were retained for the three analysed markers: 0.60 million for NeoFish (45%), 0.65 million for MiFish (70%), and 0.31 million for Teleo (79%). Read counts along all pipeline steps are detailed in Table S6.

3.3 Primers performance on species detection

Out of the three 12S markers analysed, only MiFish was able to detect all species from all mock communities. The NeoFish marker was able to detect 34 of 38 species in the SFJQmc (89.5%), 20 of 23 in the SFmc (86.96%), and 16 of 17 in the JQmc (94.11%) (Figure 2). NeoFish presented a poor performance on the Pseudopimelodidae, Pimelodidae and Heptapteridae families, failing to detect four species *Imparfinis minutus*, *Microglanis leptostriatus*, *Pimelodus pohli*, and *Steindachneridion amblyurum* in all mock communities. Teleo, detected 16 out of 17 species (94.11%) in the normalised JQmc, failing to detect *Prochilodus costatus*; and 15 out of 17 species in the skewed JQmc, missing *P. costatus* and *A. lacustris*.

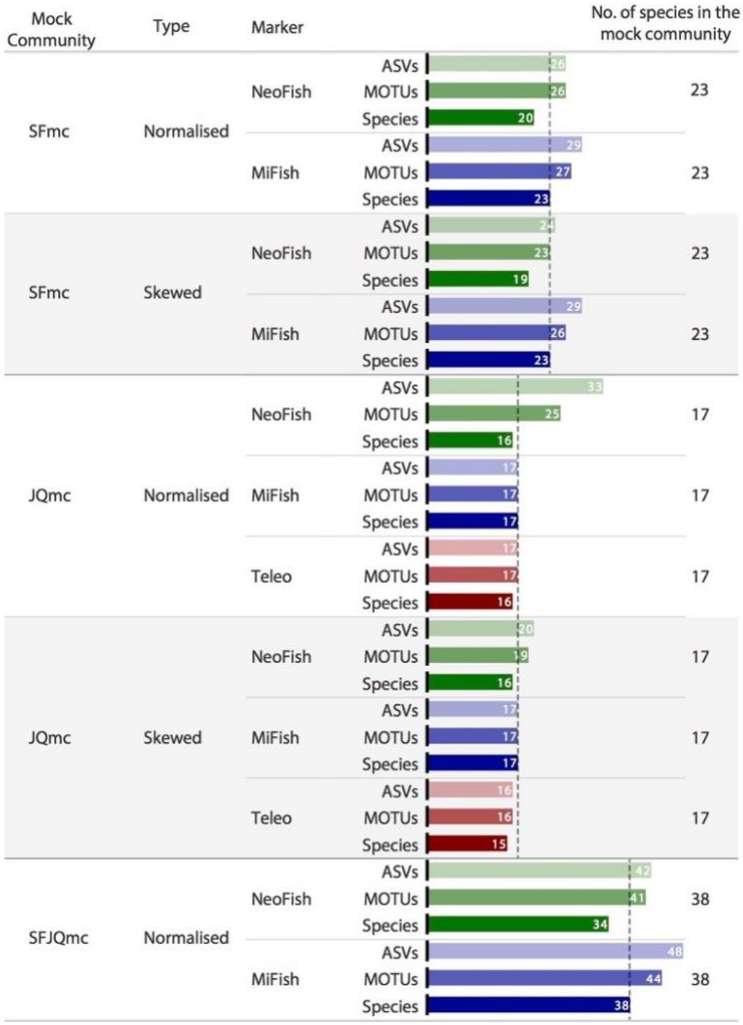


Figure 2. Numbers of ASVs, MOTUs and assigned species for all mock communities and markers. Bars represent the values for each category. Markers are identified by colours: NeoFish - green, MiFish - blue, Teleo - red. The dotted lines highlight the number of species included in each mock community. Number of species DNA added in each mock community: SFJQmc= 38; SFmc= 23; JQmc=17

Concerning ASVs, for the 38 species present in the SFJQmc, MiFish recovered 48 ASVs while NeoFish retrieved 42 ASVs (Figure 2). Considering the 23 species in the SFmc, MiFish recovered 29 ASVs in both normalised and skewed mock communities, while NeoFish recovered 26 and 24 ASVs in the normalised and in skewed versions, respectively (Figure 2). For the normalised and skewed JQmc, both MiFish and Teleo retrieved 17 ASVs, and NeoFish recovered 33 ASVs in the normalised, but only 20 ASVs in the skewed version (Figure 2). P. maculatus, from SFmc, was the species with the highest number of ASVs generated with MiFish, with three in the normalized version and four in the skewed. For NeoFish, *Tetragonopterus chalceus* presented four to five ASVs, and *M. garmani* had one to six ASVs recovered.

For the Teleo marker, *Hypostomus nigrolineatus* and *Delturus brevis* had two ASVs each (Table S5A to S5C).

5.4.3 Fold variation between input DNA and RRA

MiFish detected some species with mean fold change lower than -10x (e.g., *Microglanis leptostriatus*, *Pterygoplichthys etentaculatus*, *Pamphorichtys hollandi* and *Myleus micans*). The species with the lowest fold change value were *Myleus micans* (< -50x) and *Pamphorichthys hollandi* (< -200x). In contrast, *Moenkhausia costae* had a fold increase of 13.9x, and *Pimelodus maculatus* had an increase of ~5x. For *Gymnotus carapo*, fold change varied from -1.12x to 9x; *Pterygoplichthys etentaculatus*, between -7.5x and -12.2x; and *Tetragonopterus chalceus*, between -1.16x and 40x (Figure 3).

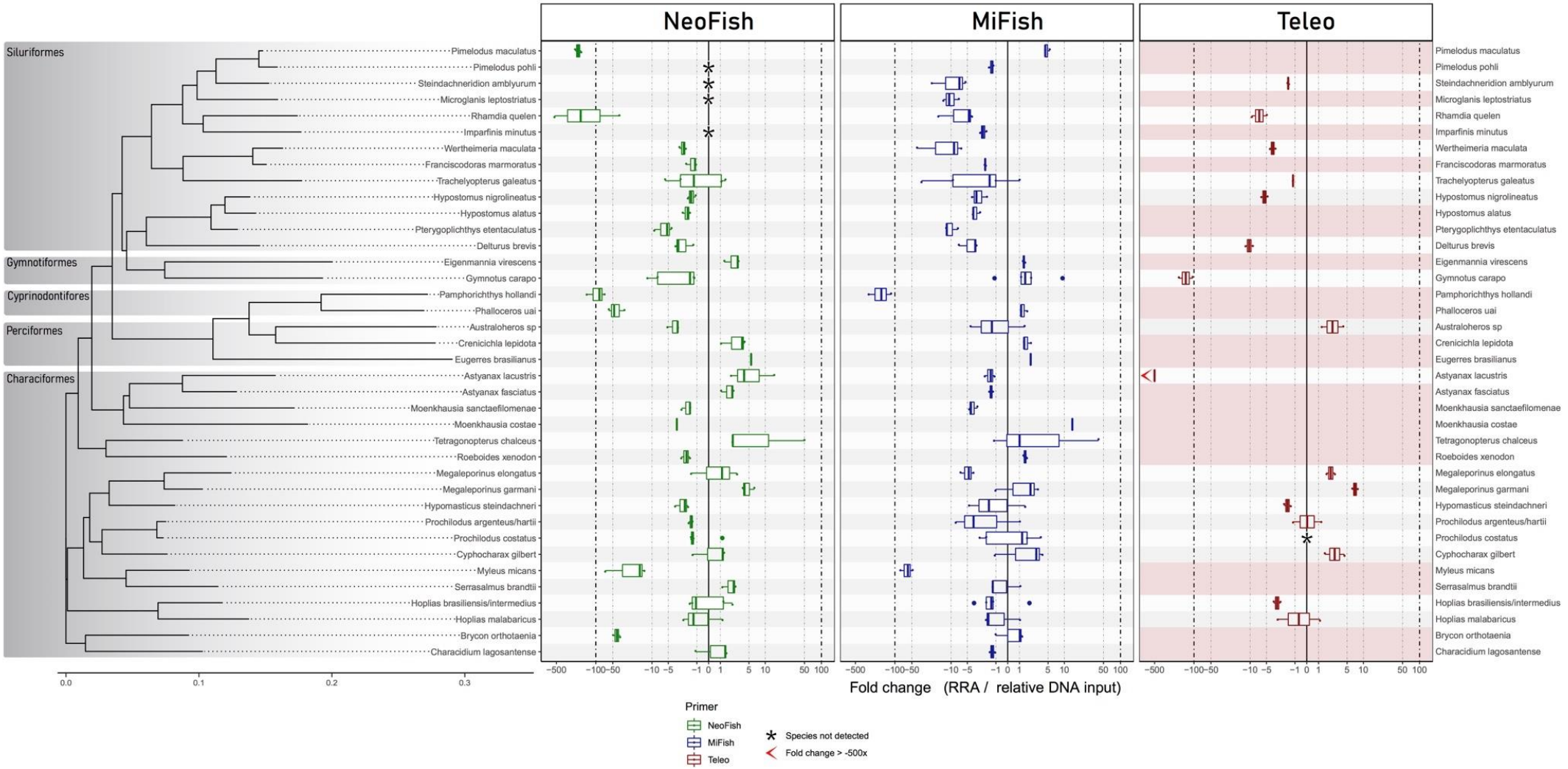


Figure 3. Fold variation between input DNA and relative read abundance (RRA) for SFmc, JQmc & SFJQmc species. Bars represent the range of the fold change between the input DNA and RRA for each species and marker. Undetected species are marked by *. Light red spaces indicate species absent in the mock communities amplified with Teleo.

The Teleo marker detected Astyanax lacustris in the normalised JQmc with an extremely low RRA (0.0000246%; fold change < -500x) but was completely missed on the skewed pool. Other underrepresented species for this marker were: *Gymnotus carapo* (~130x fold decrease), *Delturus brevis* (~10x fold decrease), *Rhamdia quelen* (~7.5x fold decrease), and *Hypostomus nigrolineatus* (~5x fold decrease). On the other hand, *Megaleporinus garmani* was overrepresented, with a fold increase of ~7x (Figure 3).

The three primers showed similar fold change profiles for some species, such as *R. quelen*, *W. maculata*, *T. galeatus*, *H. nigrolineatus*, *D. brevis* and *H. steindachneri*, which were all under amplified. Conversely, *M. garmani* and *C. gilbert* were overamplified by all primers. The species missed by NeoFish presented mild negative fold variation with MiFish, between -2 & -10. The only species not detected by Teleo (*P. costatus*) was not under or overamplified by MiFish and NeoFish.

5.4.4 Correlation between input DNA and RRA

Comparisons between input DNA and correspondent RRA showed a significant correlation only for the skewed JQmc amplified by NeoFish, indicating a moderate correlation, with r2 = 0.419 (p-value=0.0123; Figure 4). For the other markers and mock communities, no significant correlations were obtained, but MiFish had similar regression patterns for SFmc and JQmc, presenting the closest profiles for regression lines. NeoFish had different profiles for the mock communities, with a less steep regression line for the SFmc. When mock communities were evaluated in combination (i.e., SFJQmc), no significant correlation was observed (Table S7).

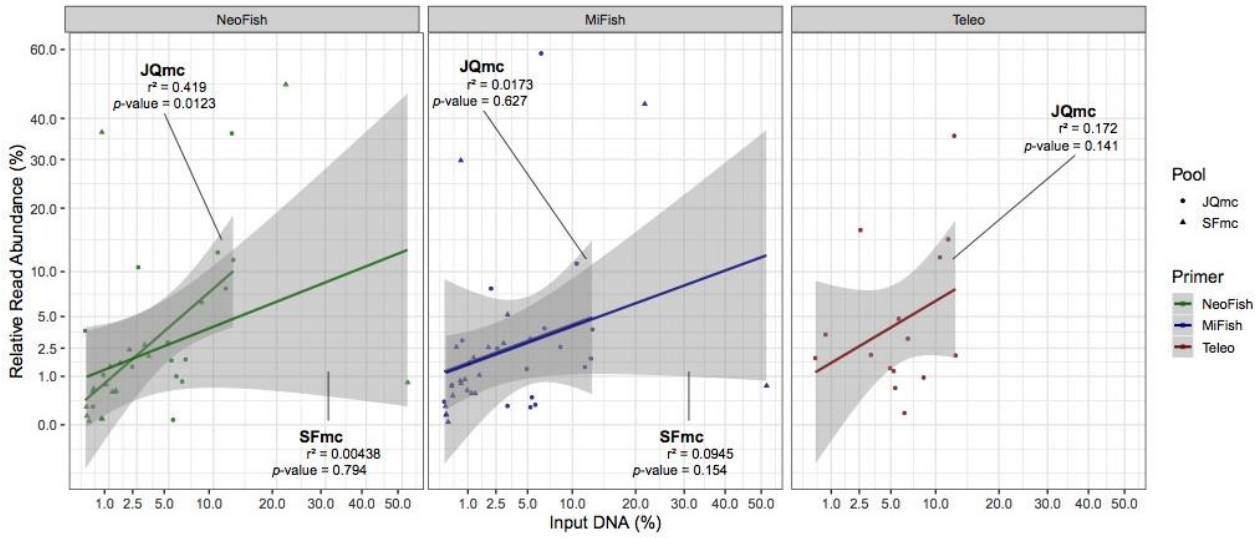


Figure 4. Correlations between relative read abundance (RRA) and correspondent input DNA added in each mock community. Each dot represents one species in a mock community. Markers are identified by colours: NeoFish - green, MiFish - blue, Teleo - red.

5.5 DISCUSSION

Despite a sharp increase in DNA metabarcoding studies, their potential to reveal biodiversity patterns in the Neotropics remains largely unexplored (Zinger et al., 2020). Here, we present a 12S fish reference database for approximately one hundred fish species that allows a broader range of species assignment using non-invasive molecular methods. We showcased the influence of community composition, proportion of input DNA, molecular marker, and taxonomy free vs species assignment caveats on species detection using mock communities.

Although all three 12S molecular markers (MiFish, Neofish and Teleo) allowed amplification and detection of most species from the mock communities, MiFish had the best performance and was the only marker able to identify most species (Figure 2). These findings are in agreement with Polanco et al. (2021), who compared the performance of MiFish and Teleo using in-silico PCR, and environmental samples from a tropical and temperate origin. They reported that both markers had dropouts and recommended the use of multiple primers to increase species detection probability using combinations that allow amplification of all target species. However, Neotropical fish biodiversity is far more complex than the mock communities used in the present work. For instance, the SF River Basin hosts at least 300 fish species (Barbosa et al., 2017). Therefore, it is important to consider that the identification outcomes can differ from different species assemblages and, in order to investigate less characterised and diverse environments such as the Neotropical rivers, the use of multiple and

complementary markers should be considered, especially in preliminary eDNA studies where target biodiversity is not known.

None of the 12S short amplicons used in this study presented enough molecular resolution to discriminate between the congeners species *Prochilodus argenteus*/*hartii*, and *Hoplias brasiliensis*/*intermedius*. However, it is known that these genera have issues concerning their taxonomy (Rosso et al., 2018), and hybrids exist (Sales et al., 2018). Moreover, Santos et al. (2021) reported that the ribosomal 12S and 16S genes have the lowest sequence variation in the whole mitochondrial genome of *Prochilodus* species. Therefore, other loci could be evaluated in the future to further improve these species resolution issues, if necessary, its discrimination.

The NeoFish marker, recently developed to detect Neotropical fish (Milan et al 2020) did not performed well to detect Siluriformes species, especially from the Heptapteridae family. These species share a divergent base at the 3’ end of the forward primer that may explain such poor performance. In order to overcome this limitation and with the use of our custom database, we propose a simple redesign by displacing the NeoFish forward primer a single base towards the 5’ (NeoFishF_v2: 5'-CCGCCGTCGCAAGCTTACCC-3'). It is known that even universal primers designed using large databases such as MiFish often require different versions to be able to recover and identify all species (Miya et al., 2015). For instance, when Miya and Sado (2019) observed that MiFish primer did not amplify DNA of ten species of sea sculpins from four genera *Pseudoblennius*, *Furcina*, *Ocynectes*, and *Vellitor* (Scorpaeniformes: Cottidae), which are common reef-associated fish species along the Pacific coast of central Japan, the authors suggested the use of optimised primers to accommodate the sequence variation of these taxa. Further, a series of later experiments provided reasonable results that more accurately reflected the regional fish community compositions. The effectiveness of such a modification in the NeoFish marker should be tested in a future study.

Using mock communities with different input DNA proportions, we could simulate real eDNA sampling scenarios where rare target species may present low DNA concentration, and thus may be proportionally under amplified. In the present study, both NeoFish and Teleo detected less species in the skewed versions of the mock communities, retrieving one species less than the normalised versions. These species were present at lower than 0.6% of the input DNA in the mock skewed communities (*Pimelodus maculatus*, with 0.58% on skewed SFmc, not detected by

NeoFish, and *Astyanax lacustris*, with 0.26% on skewed JQmc, not detected by Teleo). Therefore, the use of deeper sequencing, with more reads per sample, is advised to enable rare species detection, as proposed by Van der Loos & Nijland (2021), who also highlighted the importance of mock communities as positive controls to enable the evaluation of the quantitative signal and to define correction factors.

Despite some species presenting similar amplification profiles and fold changes by all three markers, other species behaved differently for each molecular marker, being under or overrepresented on the yielded reads, or even not amplified, a pattern known as “PCR dropouts” (Miya et al., 2020). This contrasting amplification profile between different markers is known to result from many factors, but primer annealing leading to biased amplification would be a plausible explanation.

This is the first study to estimate fold changes from input DNA to RRA for fish species using mock communities of different species compositions. Our results demonstrated that fold changes from input DNA to RRA are dependent on the primer/species pair. Understanding the primer interaction with the species, both in silico through mismatch detection and PCR and in vitro through fold change evaluation, could enable calibration of field experiments with the application of corrections for species/primers with known under and over-amplification behaviour. Similarly, Duke and Burton (2020) and Sard et al. (2019) found generalised differences between primer pairs and their ability to detect species in a mock community. Future studies could tease apart the effects of amplification bias from other variables that may impact detection thresholds. In fact, McLaren et al. (2019) developed a model for correcting RRA abundances of samples by the use of reference mock communities of known composition and pointed out that RRA outcome of a given species depends not only on its amplification profile by the chosen marker, but also on the amplification profiles of the other species in the mixture.

Both ASV and MOTU measurements slightly overestimated richness, with values higher than the original number of species presented in the mock communities, since a single specimen yielded multiple ASVs/MOTUs. Thus, if ASVs or MOTUs were used as a proxy to species alpha biodiversity that estimation would be inflated. Therefore, the use of taxonomy free metabarcoding, which has been proposed in the investigation of highly diverse and poorly characterised species assemblies (Mächler et al., 2021), as an alternative tool to enable ecological assessments of megadiverse Neotropical fish fauna should be used with caution. The sequence diversity

represented on these proxies may arise from many sources, including intraspecific marker diversity as well as errors incorporated in PCR and sequencing steps (Tsuji et al., 2020). Moreover, in our study, the mock community from the same basin amplified with the same marker in normalised and skewed configurations yielded different numbers of ASVs and MOTUs but had the identical species assigned. Further, this discrepancy was more pronounced in the mock community with highest number of species in its composition.

The cut-off necessary to remove putative contaminations and/or tag jumping (RRA < = 0.02%) resulted in the loss of three species for the NeoFish marker, and one for both MiFish and Teleo (Tables S5A to S5C). Although the use of minimum cut-off values for RRA have been proposed to mitigate contaminations within eDNA metabarcoding studies (Shirazi et al., 2021), this practice is subjective and can result in the removal of species that are present in the environment.

Thus, here we showcased the importance of a custom reference database to allow precise assignment of Neotropical fish species in metabarcoding studies, and that the RRA is dependent on community composition, marker and DNA concentration. Moreover, since species concentration influences amplification and the number of ASVs/MOTUs recovered for a single species, studies focusing on estimating species abundance using metabarcoding must consider these methodological caveats. The poor performance of NeoFish marker to identify some target taxa highlights the importance of controlled experiments using known species communities before large investments are made in assessing biodiversity using non-invasive methods such as environmental DNA.

REFERENCES

Andrade Neto, F. R. (2010). Estado atual do conhecimento sobre a fauna de peixes da bacia do Jequitinhonha. INSTITUTO ESTADUAL DE FLORESTAS—MG, 23.

Antich, A., Palacin, C., Wangensteen, O. S., & Turon, X. (2021). To denoise or to cluster, that is not the question: optimizing pipelines for COI metabarcoding and metaphylogeography. BMC bioinformatics, 22, 1-24.

Barbosa, J. M., Soares, E. C., Cintra, I. H. A., Hermann, M., & Araújo, A. R. R. (2017). Perfil da ictiofauna da bacia do rio São Francisco. Acta of Fisheries and Aquatic resources, 5(1), 70-90.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Wheeler, D. L. (2007). GenBank Nucleic Acids Res. 35. D21-D25.

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., ... & De Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. Trends in ecology & evolution, 29(6), 358-367.

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

Boussarie, G., Bakker, J., Wangensteen, O. S., Mariani, S., Bonnin, L., Juhel, J. B., ... & Mouillot, D. (2018). Environmental DNA illuminates the dark diversity of sharks. Science advances, 4(5), eaap9661.

Brauer, C. J., Hammer, M. P., & Beheregaray, L. B. (2016). Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. Molecular Ecology, 25(20), 5093-5113.

Burkhead, N. M. (2012). Extinction rates in North American freshwater fishes, 1900–2010. BioScience, 62(9), 798-808.

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME journal, 11(12), 2639-2643.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nature methods, 13(7), 581-583.

Cardoso, Y. P., Rosso, J. J., Mabragaña, E., González-Castro, M., Delpiani, M., Avigliano, E., ... & Díaz de Astarloa, J. M. (2018). A continental-wide molecular approach unraveling mtDNA diversity and geographic distribution of the Neotropical genus Hoplias. PLoS One, 13(8), e0202024.

Carvalho, D. C. (2022). Ichthyoplankton DNA metabarcoding: Challenges and perspectives. Molecular Ecology, 31(6), 1612-1614.

Castello, L., & Macedo, M. N. (2016). Large-scale degradation of Amazonian freshwater ecosystems. Global change biology, 22(3), 990-1007.

Cilleros, K., Valentini, A., Allard, L., Dejean, T., Etienne, R., Grenouillet, G., ... & 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., ... & Mariani, S. (2019). Non-specific amplification compromises environmental DNA metabarcoding with COI. Methods in Ecology and Evolution, 10(11), 1985-2001.

de Carvalho, D. C., Oliveira, D. A., Pompeu, P. S., Leal, C. G., Oliveira, C., & Hanner, R. (2011). Deep barcode divergence in Brazilian freshwater fishes: the case of the São Francisco River basin. Mitochondrial Dna, 22(sup1), 80-86.

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., ... & Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular ecology, 26(21), 5872-5895.

Duke, E. M., & Burton, R. S. (2020). Efficacy of metabarcoding for identification of fish eggs evaluated with mock communities. Ecology and evolution, 10(7), 3463-3476.

Fonseca, V. G. (2018). Pitfalls in relative abundance estimation using eDNA metabarcoding.

Fujisawa, T., & Barraclough, T. G. (2013). Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. Systematic biology, 62(5), 707-724.

García-García, N., Tamames, J., Linz, A. M., Pedrós-Alió, C., & Puente-Sánchez, F. (2019). Microdiversity ensures the maintenance of functional microbial communities under changing environmental conditions. The ISME journal, 13(12), 2969-2983.

Govender, A., Singh, S., Groeneveld, J., Pillay, S., & Willows-Munro, S. (2022). Experimental validation of taxon-specific mini-barcode primers for metabarcoding of zooplankton. Ecological Applications, 32(1), e02469.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A., & Baird, D. J. (2011). Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS one, 6(4), e17497.

Hänfling, B., Lawson Handley, L., Read, D. S., Hahn, C., Li, J., Nichols, P., ... & Winfield, I. J. (2016). Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. Molecular ecology, 25(13), 3101-3119.

Hansen, A. J., Mitchell, D. L., Wiuf, C., Paniker, L., Brand, T. B., Binladen, J., ... & Willerslev, E. (2006). Crosslinks rather than strand breaks determine access to ancient DNA sequences from frozen sediments. Genetics, 173(2), 1175-1179.

Hayami, K., Sakata, M. K., Inagawa, T., Okitsu, J., Katano, I., Doi, H., ... & Minamoto, T. (2020). Effects of sampling seasons and locations on fish environmental DNA metabarcoding in dam reservoirs. Ecology and Evolution, 10(12), 5354-5367.

Hebert, P. D., Ratnasingham, S., & De Waard, J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society of London. Series B: Biological Sciences, 270(suppl_1), S96-S99.

Hulsen, T., de Vlieg, J., & Alkema, W. (2008). BioVenn–a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC genomics, 9(1), 1-6.

Jackman, J. M., Benvenuto, C., Coscia, I., Oliveira Carvalho, C., Ready, J. S., Boubli, J. P., ... & Guimarães Sales, N. (2021). eDNA in a bottleneck: Obstacles to fish metabarcoding studies in megadiverse freshwater systems. Environmental DNA, 3(4), 837-849.

Kelly, R. P., Port, J. A., Yamahara, K. M., & Crowder, L. B. (2014). Using environmental DNA to census marine fishes in a large mesocosm. PloS one, 9(1), e86175.

Kocher, A., De Thoisy, B., Catzeflis, F., Huguin, M., Valière, S., Zinger, L., ... & Murienne, J. (2017). Evaluation of short mitochondrial metabarcodes for the identification of Amazonian mammals. Methods in Ecology and Evolution, 8(10), 1276-1283.

Krehenwinkel, H., Wolf, M., Lim, J. Y., Rominger, A. J., Simison, W. B., & Gillespie, R. G. (2017). Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. Scientific reports, 7(1), 17668.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular biology and evolution, 33(7), 1870-1874.

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

Lim, N. K., Tay, Y. C., Srivathsan, A., Tan, J. W., Kwik, J. T., Baloğlu, B., ... & Yeo, D. C. (2016). Next-generation freshwater bioassessment: eDNA metabarcoding with a conserved metazoan primer reveals species-rich and reservoir-specific communities. Royal Society open science, 3(11), 160635.

Littleford-Colquhoun, B. L., Freeman, P. T., Sackett, V. I., Tulloss, C. V., McGarvey, L. M., Geremia, C., & Kartzinel, T. R. (2022). The precautionary principle and dietary DNA metabarcoding: Commonly used abundance thresholds change ecological interpretation.

Mächler, E., Walser, J. C., & Altermatt, F. (2021). Decision-making and best practices for taxonomy-free environmental DNA metabarcoding in biomonitoring using Hill numbers. Molecular Ecology, 30(13), 3326-3339.

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

Mahé, F., Rognes, T., Quince, C., De Vargas, C., & Dunthorn, M. (2015). Swarm v2: highly-scalable and high-resolution amplicon clustering. PeerJ, 3, e1420.

Majaneva, M., Diserud, O. H., Eagle, S. H., Boström, E., Hajibabaei, M., & Ekrem, T. (2018). Environmental DNA filtration techniques affect recovered biodiversity. Scientific reports, 8(1), 4682.

McLaren, M. R., Willis, A. D., & Callahan, B. J. (2019). Consistent and correctable bias in metagenomic sequencing experiments. Elife, 8, e46923.

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

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

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

Miya, M., Sado, T., & eDNA Methods Standardization Committee. (2019). Multiple species detection using MiFish primers. Environmental DNA sampling and experimental manual version, 2, 55-92.

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., ... & Iwasaki, W. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. Royal Society open science, 2(7), 150088.

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

Polanco F, A., Richards, E., Flück, B., Valentini, A., Altermatt, F., Brosse, S., ... & Pellissier, L. (2021). Comparing the performance of 12S mitochondrial primers for fish environmental DNA across ecosystems. Environmental DNA, 3(6), 1113-1127.

Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., ... & Vogler, A. P. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Systematic biology, 55(4), 595-609.

Porter, T. M., & Hajibabaei, M. (2018). Automated high throughput animal CO1 metabarcode classification. Scientific Reports, 8(1), 4226.

Pugedo, M. L., de Andrade Neto, F. R., Pessali, T. C., Birindelli, J. L. O., & Carvalho, D. C. (2016). Integrative taxonomy supports new candidate fish species in a poorly studied neotropical region: the Jequitinhonha River Basin. Genetica, 144(3), 341-349.

Puillandre N, Lambert A, Brouillet S, Achaz G (2012) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Mol Ecol 21: 1864–1877.

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

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R., & Gough, K. C. (2014). The detection of aquatic animal species using environmental DNA–a review of eDNA as a survey tool in ecology. Journal of applied ecology, 51(5), 1450-1459.

Reis, R. E., Albert, J. S., Di Dario, F., Mincarone, M. M., Petry, P., & Rocha, L. A. (2016). Fish biodiversity and conservation in South America. Journal of fish biology, 89(1), 12-47.

Ritter, C. D., Dal Pont, G., Stica, P. V., Horodesky, A., Cozer, N., Netto, O. S. M., ... & Pie, M. R. (2022). Wanted not, wasted not: Searching for non-target taxa in environmental DNA metabarcoding by-catch. Environmental Advances, 7, 100169.

Rivera, S. F., Rimet, F., Vasselon, V., Vautier, M., Domaizon, I., & Bouchez, A. (2022). Fish eDNA metabarcoding from aquatic biofilm samples: Methodological aspects. Molecular Ecology Resources, 22(4), 1440-1453.

Rosa, R. S., & Lima, F. C. (2008). Os peixes brasileiros ameaçados de extinção. Livro vermelho da fauna brasileira ameaçada de extinção, 2, 9-19.

Rosso, J. J., González-Castro, M., Bogan, S., Cardoso, Y. P., Mabragaña, E., Delpiani, M., & de Astarloa, J. M. D. (2018). Integrative taxonomy reveals a new species of the Hoplias malabaricus species complex (Teleostei: Erythrinidae). Verlag Dr. Friedrich Pfeil.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution, 4(4), 406-425.

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

Sales, N. G., Pessali, T. C., Andrade Neto, F. R., & Carvalho, D. C. (2018). Introgression from non-native species unveils a hidden threat to the migratory Neotropical fish Prochilodus hartii. Biological Invasions, 20, 555-566.

Sales, N. G., Pessali, T. C., Andrade Neto, F. R., & Carvalho, D. C. (2018). Introgression from non-native species unveils a hidden threat to the migratory Neotropical fish Prochilodus hartii. Biological Invasions, 20, 555-566.

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

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

Santos, R. P., Melo, B. F., Yazbeck, G. M., Oliveira, R. S., Hilário, H. O., Prosdocimi, F., & Carvalho, D. C. (2021). Diversification of Prochilodus in the eastern Brazilian

Shield: Evidence from complete mitochondrial genomes (Teleostei, Prochilodontidae). Journal of Zoological Systematics and Evolutionary Research, 59(5), 1053-1063.

Sard, N. M., Herbst, S. J., Nathan, L., Uhrig, G., Kanefsky, J., Robinson, J. D., & Scribner, K. T. (2019). Comparison of fish detections, community diversity, and relative abundance using environmental DNA metabarcoding and traditional gears. Environmental DNA, 1(4), 368-384.

Sato, H., Sogo, Y., & Yamanaka, H. (2017). Usefulness and limitations of sample pooling for environmental DNA metabarcoding of freshwater fish communities. Scientific Reports, 7(1), 1-12.

Shirazi, S., Meyer, R. S., & Shapiro, B. (2021). Revisiting the effect of PCR replication and sequencing depth on biodiversity metrics in environmental DNA metabarcoding. Ecology and Evolution, 11(22), 15766-15779.

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. Molecular Ecology21: 1789-1793.

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA–An emerging tool in conservation for monitoring past and present biodiversity. Biological conservation, 183, 4-18.

Thomsen, P. F., Kielgast, J. O. S., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., ... & Willerslev, E. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. Molecular ecology, 21(11), 2565-2573.

Thomsen, P. F., Kielgast, J., Iversen, L. L., Møller, P. R., Rasmussen, M., & Willerslev, E. (2012). Detection of a diverse marine fish fauna using environmental DNA from seawater samples.

Thomsen, P. F., Møller, P. R., Sigsgaard, E. E., Knudsen, S. W., Jørgensen, O. A., & Willerslev, E. (2016). Environmental DNA from seawater samples correlate with trawl catches of subarctic, deepwater fishes. PloS one, 11(11), e0165252.

Tsuji, S., Miya, M., Ushio, M., Sato, H., Minamoto, T., & Yamanaka, H. (2020). Evaluating intraspecific genetic diversity using environmental DNA and denoising approach: A case study using tank water. Environmental DNA, 2(1), 42-52.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., ... & Dejean, T. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Molecular ecology, 25(4), 929-942.

Van der Loos, L. M., & Nijland, R. (2021). Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. Molecular Ecology, 30(13), 3270-3288.

Wagih, O. (2017). ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics, 33(22), 3645-3647.

Wang, S., Yan, Z., Hänfling, B., Zheng, X., Wang, P., Fan, J., & Li, J. (2021). Methodology of fish eDNA and its applications in ecology and environment. Science of the Total Environment, 755, 142622.

Willerslev, E., Hansen, A. J., Binladen, J., Brand, T. B., Gilbert, M. T. P., Shapiro, B., ... & Cooper, A. (2003). Diverse plant and animal genetic records from Holocene and Pleistocene sediments. Science, 300(5620), 791-795.

Wright, E. S. (2016). Using DECIPHER v2. 0 to analyze big biological sequence data in R. R J., 8(1), 352.

Xiong, F., Shu, L., Gan, X., Zeng, H., He, S., & Peng, Z. (2022). Methodology for fish biodiversity monitoring with environmental DNA metabarcoding: the primers, databases and bioinformatic pipelines. Water Biology and Security, 100007.

Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. Bioinformatics, 29(22), 2869-2876.

Zinger, L., Donald, J., Brosse, S., Gonzalez, M. A., Iribar, A., Leroy, C., ... & Lopes, C. M. (2020). Advances and prospects of environmental DNA in neotropical rainforests. Advances in ecological research, 62, 331-373.

Supplementary material

Supplementary Figures

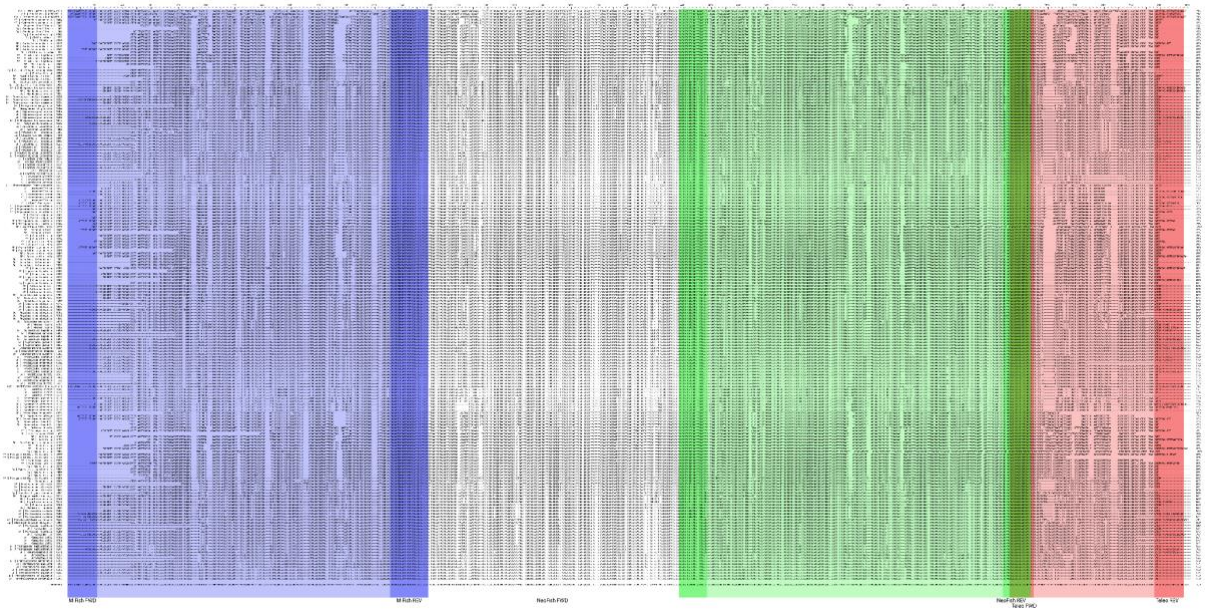


Figure S1. The LGC 12Sdb.

Supplementary Tables

Table S1: List of species of the Conservation Genetics Lab 12S Sequence Database (LGC12Sdb)

Genus	Species	Specimen ID	Basin	Reference
Acestrorhynchus	*Acestrorhynchus lacustris*	917	SF	Milan *et al.*, 2020
Acestrorhynchus	*Acestrorhynchus lacustris*	1136	SF	Milan *et al.*, 2020
Acinocheirodon	*Acinocheirodon melanogramma*	1547	Jq	This study
Anchoviella	*Anchoviella lepidentostole*	2912	Jq	This study
Apareiodon	*Apareiodon ibitiensis*	1371	SF	Milan *et al.*, 2020
Astyanax	*Astyanax cf_fasciatus*	1153	SF	Milan *et al.*, 2020
Astyanax	*Astyanax cf_fasciatus*	1044	SF	Milan *et al.*, 2020
Astyanax	*Astyanax cf_fasciatus*	1048	SF	Milan *et al.*, 2020
Astyanax	*Astyanax lacustris*	1239	SF	Milan *et al.*, 2020
Astyanax	*Astyanax aff_fasciatus*	5693	Jq	This study
Astyanax	*Astyanax aff_fasciatus*	5694	Jq	This study
Astyanax	*Astyanax aff_fasciatus*	5695	Jq	This study
Astyanax	*Astyanax cf_lacustris*	5672	Jq	This study
Astyanax	*Astyanax lacustris*	7880	Jq	This study
Australoheros	*Australoheros facetus*	1217	SF	Milan *et al.*, 2020
Australoheros	*Australoheros facetus*	1218	SF	Milan *et al.*, 2020
Australoheros	*Australoheros facetus*	1242	SF	Milan *et al.*, 2020
Australoheros	*Australoheros sp*	6584	Jq	This study
Brycon	*Brycon orthotaenia*	1037	SF	Milan *et al.*, 2020
Brycon	*Brycon orthotaenia*	1005	SF	Milan *et al.*, 2020
Bryconamericus	*Bryconamericus stramineus*	1150	SF	Milan *et al.*, 2020

Bryconamericus	*Bryconamericus stramineus*	1151	SF	Milan *et al.*, 2020
Bryconamericus	*Bryconamericus stramineus*	1152	SF	Milan *et al.*, 2020
Bryconops	*Bryconops affinis*	1398	SF	Milan *et al.*, 2020
Bryconops	*Bryconops affinis*	1399	SF	Milan *et al.*, 2020
Bunocephalus	*Bunocephalus sp*	1158	SF	Milan *et al.*, 2020
Bunocephalus	*Bunocephalus sp*	1159	SF	Milan *et al.*, 2020
Cetopsorhamdia	*Cetopsorhamdia iheringi*	1372	SF	Milan *et al.*, 2020
Characidium	*Characidium lagosantense*	1381	SF	Milan *et al.*, 2020
Characidium	*Characidium lagosantense*	1379	SF	Milan *et al.*, 2020
Characidium	*Characidium sp*	1308	SF	Milan *et al.*, 2020
Characidium	*Characidium sp*	1334	SF	Milan *et al.*, 2020
Characidium	*Characidium zebra*	1060	SF	Milan *et al.*, 2020
Characidium	*Characidium zebra*	1335	SF	Milan *et al.*, 2020
Cichla	*Cichla kelberi*	2932	Jq	This study
Cichla	*Cichla sp*	6590	Jq	This study
Cichlasoma	*Cichlasoma facetum*	292	Jq	This study
Colossoma	*Colossoma macropomum*	2871	Jq	This study
Corydoras	*Corydoras sp*	1568	Jq	This study
Crenicichla	*Crenicichla lepidota*	1264	SF	Milan *et al.*, 2020
Crenicichla	*Crenicichla lepidota*	1129	SF	Milan *et al.*, 2020
Crenicichla	*Crenicichla lepidota*	1130	SF	Milan *et al.*, 2020
Curimatella	*Curimatella lepidura*	900	SF	Milan *et al.*, 2020
Curimatella	*Curimatella lepidura*	906	SF	Milan *et al.*, 2020
Curimatella	*Curimatella lepidura*	907	SF	Milan *et al.*, 2020
Cyphocharax	*Cyphocharax gilbert*	1054	SF	Milan *et al.*, 2020
Cyphocharax	*Cyphocharax gilbert*	1056	SF	Milan *et al.*, 2020
Cyphocharax	*Cyphocharax gilbert*	6586	Jq	This study
Delturus	*Delturus brevis*	1936	Jq	This study
Eigenmannia	*Eigenmannia virescens*	1117	SF	Milan *et al.*, 2020
Eigenmannia	*Eigenmannia virescens*	1118	SF	Milan *et al.*, 2020
Eigenmannia	*Eigenmannia virescens*	1119	SF	Milan *et al.*, 2020
Eugerres	*Eugerres brasilianus*	2943	Jq	This study
Franciscodoras	*Franciscodoras marmoratus*	708	SF	Milan *et al.*, 2020
Geophagus	*Geophagus brasiliensis*	303	Jq	This study
Geophagus	*Geophagus brasiliensis*	304	Jq	This study
Gymnotus	*Gymnotus carapo*	1248	SF	Milan *et al.*, 2020
Gymnotus	*Gymnotus carapo*	1631	Jq	This study
Gymnotus	*Gymnotus carapo*	1610	Jq	This study
Harttia	*Harttia sp*	1145	SF	Milan *et al.*, 2020
Harttia	*Harttia sp*	1156	SF	Milan *et al.*, 2020
Harttia	*Harttia garavelloi*	439	Jq	This study
Hemigrammus	*Hemigrammus cf_gracilis*	1380	SF	Milan *et al.*, 2020
Hemigrammus	*Hemigrammus marginatus*	1382	SF	Milan *et al.*, 2020
Hisonotus	*Hisonotus sp*	1108	SF	Milan *et al.*, 2020
Hisonotus	*Hisonotus sp*	1109	SF	Milan *et al.*, 2020
Hoplias	*Hoplias intermedius*	1377	SF	Milan *et al.*, 2020
Hoplias	*Hoplias malabaricus*	1234	SF	Milan *et al.*, 2020
Hoplias	*Hoplias malabaricus*	1346	SF	Milan *et al.*, 2020
Hoplias	*Hoplias brasiliensis*	1772	Jq	This study

Hoplias	*Hoplias malabaricus*	7822	Jq	This study
Hoplias	*Hoplias sp*	6630	Jq	This study
Hoplosternum	*Hoplosternum littorale*	1215	SF	Milan *et al.*, 2020
Hoplosternum	*Hoplosternum littorale*	1274	SF	Milan *et al.*, 2020
Hoplosternum	*Hoplosternum littorale*	2643	Jq	This study
Hyphessobrycon	*Hyphessobrycon santae*	1209	SF	Milan *et al.*, 2020
Hyphessobrycon	*Hyphessobrycon santae*	1210	SF	Milan *et al.*, 2020
Hyphessobrycon	*Hyphessobrycon santae*	1211	SF	Milan *et al.*, 2020
Hypomasticus	*Hypomasticus steindachneri*	2984	Jq	This study
Hypomasticus	*Hypomasticus steindachneri*	5664	Jq	This study
Hypomasticus	*Hypomasticus steindachneri*	7826	Jq	This study
Hypostomus	*Hypostomus alatus*	1041	SF	Milan *et al.*, 2020
Hypostomus	*Hypostomus alatus*	925	SF	Milan *et al.*, 2020
Hypostomus	*Hypostomus nigrolineatus*	7893	Jq	This study
Hypostomus	*Hypostomus sp*	1711	Jq	This study
Hypostomus	*Hypostomus sp*	5650	Jq	This study
Hypostomus	*Hypostomus sp*	5703	Jq	This study
Hypostomus	*Hypostomus sp*	5704	Jq	This study
Imparfinis	*Imparfinis minutus*	1161	SF	Milan *et al.*, 2020
Imparfinis	*Imparfinis minutus*	1370	SF	Milan *et al.*, 2020
Leporinus	*Leporinus macrocephalus*	982	SF	Milan *et al.*, 2020
Leporinus	*Leporinus piau*	958	SF	Milan *et al.*, 2020
Leporinus	*Leporinus piau*	986	SF	Milan *et al.*, 2020
Leporinus	*Leporinus piau*	1384	SF	Milan *et al.*, 2020
Leporinus	*Leporinus reinhardti*	1344	SF	Milan *et al.*, 2020
Leporinus	*Leporinus taeniatus*	985	SF	Milan *et al.*, 2020
Leporinus	*Leporinus taeniatus*	1317	SF	Milan *et al.*, 2020
Leporinus	*Leporinus taeniatus*	1553	Jq	This study
Lycengraulis	*Lycengraulis grossidens*	2918	Jq	This study
Megaleporinus	*Megaleporinus elongatus*	772	SF	Milan *et al.*, 2020
Megaleporinus	*Megaleporinus elongatus*	1032	SF	Milan *et al.*, 2020
Megaleporinus	*Megaleporinus elongatus*	1033	SF	Milan *et al.*, 2020
Megaleporinus	*Megaleporinus elongatus*	1762	Jq	This study
Megaleporinus	*Megaleporinus elongatus*	1761	Jq	This study
Megaleporinus	*Megaleporinus elongatus*	6633	Jq	This study
Megaleporinus	*Megaleporinus garmani*	4290	Jq	This study
Microglanis	*Microglanis leptostriatus*	1104	SF	Milan *et al.*, 2020
Microglanis	*Microglanis leptostriatus*	1102	SF	Milan *et al.*, 2020
Moenkhausia	*Moenkhausia costae*	1074	SF	Milan *et al.*, 2020
Moenkhausia	*Moenkhausia sanctaefilomenae*	1113	SF	Milan *et al.*, 2020
Moenkhausia	*Moenkhausia sanctaefilomenae*	1111	SF	Milan *et al.*, 2020
Moenkhausia	*Moenkhausia sanctaefilomenae*	1112	SF	Milan *et al.*, 2020
Moenkhausia	*Moenkhausia costae*	7812	Jq	This study
Myleus	*Myleus micans*	1162	SF	Milan *et al.*, 2020
Myleus	*Myleus micans*	763	SF	Milan *et al.*, 2020
Odontostilbe	*Odontostilbe sp*	1076	SF	Milan *et al.*, 2020
Odontostilbe	*Odontostilbe sp*	1077	SF	Milan *et al.*, 2020
Odontostilbe	*Odontostilbe sp*	1313	SF	Milan *et al.*, 2020
Oligosarcus	*Oligosarcus macrolepis*	7690	Jq	This study

Orthospinus	*Orthospinus franciscensis*	1072	SF	Milan *et al.*, 2020
Pamphorichthys	*Pamphorichthys hollandi*	1099	SF	Milan *et al.*, 2020
Pamphorichthys	*Pamphorichthys hollandi*	1098	SF	Milan *et al.*, 2020
Pareiorhaphis	*Pareiorhaphis stephanus*	4196	Jq	This study
Parodon	*Parodon hilarii*	1245	SF	Milan *et al.*, 2020
Parodon	*Parodon hilarii*	1332	SF	Milan *et al.*, 2020
Phalloceros	*Phalloceros uai*	1368	SF	Milan *et al.*, 2020
Phalloceros	*Phalloceros uai*	1369	SF	Milan *et al.*, 2020
Phalloceros	*Phalloceros sp*	1551	Jq	This study
Phenacogaster	*Phenacogaster franciscoensis*	1333	SF	Milan *et al.*, 2020
Piabina	*Piabina argentea*	1306	SF	Milan *et al.*, 2020
Piabina	*Piabina argentea*	1307	SF	Milan *et al.*, 2020
Pimelodella	*Pimelodella vittata*	1287	SF	Milan *et al.*, 2020
Pimelodella	*Pimelodella vittata*	1288	SF	Milan *et al.*, 2020
Pimelodella	*Pimelodella vittata*	1289	SF	Milan *et al.*, 2020
Pimelodus	*Pimelodus fur*	1147	SF	Milan *et al.*, 2020
Pimelodus	*Pimelodus fur*	1345	SF	Milan *et al.*, 2020
Pimelodus	*Pimelodus maculatus*	1280	SF	Milan *et al.*, 2020
Pimelodus	*Pimelodus maculatus*	749	SF	Milan *et al.*, 2020
Pimelodus	*Pimelodus pohli*	1400	SF	Milan *et al.*, 2020
Pimelodus	*Pimelodus pohli*	1401	SF	Milan *et al.*, 2020
Pimelodus	*Pimelodus pohli*	1403	SF	This study
Planaltina	*Planaltina myersi*	1367	SF	Milan *et al.*, 2020
Prionace	*Prionace glauca*	AF448008.1	Out	NCBI genbank
Prochilodus	*Prochilodus argenteus*	739	SF	Milan *et al.*, 2020
Prochilodus	*Prochilodus argenteus*	742	SF	Milan *et al.*, 2020
Prochilodus	*Prochilodus argenteus*	779	SF	Milan *et al.*, 2020
Prochilodus	*Prochilodus costatus*	1339	SF	Milan *et al.*, 2020
Prochilodus	*Prochilodus costatus*	664	SF	Milan *et al.*, 2020
Prochilodus	*Prochilodus costatus*	1281	SF	Milan *et al.*, 2020
Prochilodus	*Prochilodus costatus*	NC027690.1	SF	NCBI genbank
Prochilodus	*Prochilodus argenteus*	5612	Jq	This study
Prochilodus	*Prochilodus costatus*	2860	Jq	This study
Prochilodus	*Prochilodus hartii*	1765	Jq	This study
Prochilodus	*Prochilodus hartii*	7033	Jq	This study
Pseudopimelodus	*Pseudopimelodus sp*	671	SF	Milan *et al.*, 2020
Pseudoplatystoma	*Pseudoplatystoma corruscans*	1135	SF	Milan *et al.*, 2020
Pterygoplichthys	*Pterygoplichthys etentaculatus*	1141	SF	Milan *et al.*, 2020
Pterygoplichthys	*Pterygoplichthys etentaculatus*	861	SF	Milan *et al.*, 2020
Pterygoplichthys	*Pterygoplichthys etentaculatus*	1030	SF	Milan *et al.*, 2020
Pygocentrus	*Pygocentrus piraya*	1229	SF	Milan *et al.*, 2020
Rhamdia	*Rhamdia quelen*	1361	SF	Milan *et al.*, 2020
Rhamdia	*Rhamdia aff_quelen*	5645	Jq	This study
Rhamdia	*Rhamdia cf_jequitinhonha*	3054	Jq	This study
Roeboides	*Roeboides xenodon*	916	SF	Milan *et al.*, 2020
Schizodon	*Schizodon knerii*	953	SF	Milan *et al.*, 2020
Schizodon	*Schizodon knerii*	1003	SF	Milan *et al.*, 2020
Schizodon	*Schizodon knerii*	1337	SF	Milan *et al.*, 2020
Serrapinnus	*Serrapinnus piaba*	1086	SF	Milan *et al.*, 2020

Serrapinnus	*Serrapinnus piaba*	1087	SF	Milan *et al.*, 2020
Serrasalmus	*Serrasalmus brandtii*	1230	SF	Milan *et al.*, 2020
Serrasalmus	*Serrasalmus brandtii*	872	SF	Milan *et al.*, 2020
Serrasalmus	*Serrasalmus brandtii*	7890	Jq	This study
Steindachneridion	*Steindachneridion amblyurum*	2996	Jq	This study
Steindachnerina	*Steindachnerina elegans*	1284	SF	Milan *et al.*, 2020
Steindachnerina	*Steindachnerina elegans*	231	Jq	This study
Sternopygus	*Sternopygus macrurus*	1121	SF	Milan *et al.*, 2020
Sternopygus	*Sternopygus macrurus*	1124	SF	Milan *et al.*, 2020
Synbranchus	*Synbranchus marmoratus*	1404	SF	Milan *et al.*, 2020
Synbranchus	*Synbranchus marmoratus*	1405	SF	Milan *et al.*, 2020
Tetragonopterus	*Tetragonopterus chalceus*	901	SF	Milan *et al.*, 2020
Tetragonopterus	*Tetragonopterus chalceus*	984	SF	Milan *et al.*, 2020
Trachelyopterus	*Trachelyopterus galeatus*	1243	SF	Milan *et al.*, 2020
Trachelyopterus	*Trachelyopterus galeatus*	602	SF	Milan *et al.*, 2020
Trachelyopterus	*Trachelyopterus cf_galeatus*	5675	Jq	This study
Trichomycterus	*Trichomycterus sp*	1373	SF	Milan *et al.*, 2020
Wertheimeria	*Wertheimeria maculata*	7817	Jq	This study
Wertheimeria	*Wertheimeria maculata*	7906	Jq	This study

Table S2: Species summary used in the mock communities, the originary basin, the LGCdb ID, the input DNA (ng), the total DNA in skewed pool (ng), the percentage on respective skewed pool, the percentage on respective normalized pool, percentage on the SFJQ pool and the presence on the pool (green for presence and red for absence)

Originary Basin	Full Database name	Species	LGCdb ID	Input DNA (ng/µl)	Total DNA in Skewed pool (ng)	Percentage on respective skewed pool	Percentage on respective Normalized pool	Percentage on the SFJQ pool	Presence on pool		
									SF	JQ	SFJQ
SF	*Franciscodoras marmoratus 0708 SF*	*Franciscodoras marmoratus*	0708	0.137	47.89	0.29%	4.35%	2.33%			
SF	*Tetragonopterus chalceus 0901 SF*	*Tetragonopterus chalceus*	0901	0.350	47.89	0.73%	4.35%	2.33%			
SF	*Roeboides xenodon 0916 SF*	*Roeboides xenodon*	0916	0.600	47.89	1.25%	4.35%	2.33%			
SF	*Brycon orthotaenia 1037 SF*	*Brycon orthotaenia*	1037	0.354	47.89	0.74%	4.35%	2.33%			
SF	*Hypostomus alatus 1041 SF*	*Hypostomus alatus*	1041	0.550	47.89	1.15%	4.35%	2.33%			
SF	*Pamphorichthys hollandi 1099 SF*	*Pamphorichthys hollandi*	1099	0.168	47.89	0.35%	4.35%	2.33%			
SF	*Microglanis leptostriatus 1104 SF*	*Microglanis leptostriatus*	1104	0.148	47.89	0.31%	4.35%	2.33%			
SF	*Moenkhausia sanctaefilomenae 1113 SF*	*Moenkhausia sanctaefilomenae*	1113	0.634	47.89	1.32%	4.35%	2.33%			
SF	*Eigenmannia virescens 1117 SF*	*Eigenmannia virescens*	1117	0.502	47.89	1.05%	4.35%	2.33%			
SF	*Pterygoplichthys etentaculatus 1141 SF*	*Pterygoplichthys etentaculatus*	1141	0.145	47.89	0.30%	4.35%	2.33%			
SF	*Astyanax fasciatus 1153 SF*	*Astyanax fasciatus*	1153	0.735	47.89	1.53%	4.35%	2.33%			
SF	*Imparfinis minutus 1161 SF*	*Imparfinis minutus*	1161	0.467	47.89	0.98%	4.35%	2.33%			
SF	*Myleus micans 1162 SF*	*Myleus micans*	1162	24.700	47.89	51.58%	4.35%	2.33%			
SF	*Serrasalmus brandtii 1230 SF*	*Serrasalmus brandtii*	1230	0.223	47.89	0.47%	4.35%	2.33%			
SF	*Trachelyopterus galeatus 1243 SF*	*Trachelyopterus galeatus*	1243	0.423	47.89	0.88%	4.35%	2.33%			
SF	*Gymnotus carapo 1248 SF*	*Gymnotus carapo*	1248	1.590	47.89	3.32%	4.35%	2.33%			
SF	*Crenicichla lepidota 1264 SF*	*Crenicichla lepidota*	1264	0.213	47.89	0.44%	4.35%	2.33%			
SF	*Pimelodus maculatus 1280 SF*	*Pimelodus maculatus*	1280	0.276	47.89	0.58%	4.35%	2.33%			
SF	*Prochilodus costatus 1339 SF*	*Prochilodus costatus*	1339	0.971	47.89	2.03%	4.35%	2.33%			
SF	*Phalloceros uai 1368 SF*	*Phalloceros uai*	1368	0.339	47.89	0.71%	4.35%	2.33%			
SF	*Hoplias intermedius 1377 SF*	*Hoplias intermedius*	1377	10.400	47.89	21.72%	4.35%	2.33%			
SF	*Characidium lagosantense 1381 SF*	*Characidium lagosantense*	1381	1.450	47.89	3.03%	4.35%	2.33%			
SF	*Pimelodus pohli 1403 SF*	*Pimelodus pohli*	1403	2.510	47.89	5.24%	4.35%	2.33%			
JQ	*Gymnotus carapo 1631 Jq*	*Gymnotus carapo*	1631	29.200	461.77	6.32%	5.56%	2.33%			
JQ	*Megaleporinus elongatus 1762 Jq*	*Megaleporinus elongatus*	1762	58.000	461.77	12.56%	5.56%	2.33%			
JQ	*Prochilodus hartii 1765 Jq*	*Prochilodus hartii*	1765	43.700	461.77	9.46%	5.56%	2.33%			
JQ	*Hoplias brasiliensis 1772 Jq*	*Hoplias brasiliensis*	1772	22.700	461.77	4.92%	5.56%	2.33%			
JQ	*Delturus brevis 1936 Jq*	*Delturus brevis*	1936	39.200	461.77	8.49%	5.56%	2.33%			
JQ	*Prochilodus costatus 2860 Jq*	*Prochilodus costatus*	2860	10.100	461.77	2.19%	5.56%	2.33%			
JQ	*Eugerres brasilianus 2943 Jq*	*Eugerres brasilianus*	2943	1.85	0.00	0.00%	0.00%	2.33%			
JQ	*Hypomasticus steindachneri 2984 Jq*	*Hypomasticus steindachneri*	2984	30.900	461.77	6.69%	5.56%	2.33%			
JQ	*Steindachneridion amblyurum 2996 Jq*	*Steindachneridion amblyurum*	2996	15.300	461.77	3.31%	5.56%	2.33%			
JQ	*Megaleporinus garmani 4290 Jq*	*Megaleporinus garmani*	4920	11.900	461.77	2.58%	5.56%	2.33%			

JQ	*Prochilodus argenteus 5612 Jq*	*Prochilodus argenteus*	5612	10.300	461.77	2.23%	5.56%	2.33%			
JQ	*Rhamdia quelen 5645 Jq*	*Rhamdia quelen*	5645	25.000	461.77	5.41%	5.56%	2.33%			
JQ	*Trachelyopterus galeatus 5675 Jq*	*Trachelyopterus galeatus*	5675	26.500	461.77	5.74%	5.56%	2.33%			
JQ	*Australoheros sp 6584 Jq*	*Australoheros sp*	6584	2.070	461.77	0.45%	5.56%	2.33%			
JQ	*Cyphocharax gilbert 6586 Jq*	*Cyphocharax gilbert*	6586	3.590	461.77	0.78%	5.56%	2.33%			
JQ	*Moenkhausia costae 7812 Jq*	*Moenkhausia costae*	7812	4.03	0.00	0.00%	0.00%	2.33%			
JQ	*Wertheimeria maculata 7817 Jq*	*Wertheimeria maculata*	7817	24.300	461.77	5.26%	5.56%	2.33%			
JQ	*Hoplias malabaricus 7822 Jq*	*Hoplias malabaricus*	7822	48.800	461.77	10.57%	5.56%	2.33%			
JQ	*Astyanax lacustris 7880 Jq*	*Astyanax lacustris*	7880	1.210	461.77	0.26%	5.56%	2.33%			
JQ	*Hypostomus nigrolineatus 7893 Jq*	*Hypostomus nigrolineatus*	7893	59.000	461.77	12.78%	5.56%	2.33%			

Table S3: Primers used for the rRNA 12S gene amplification

Primer	**Orientation**	**Sequence**	**Annealing temperature**	**Expected amplicon length**	**Reference**
MiFish	Forward	5′-GTCGGTAAAACTCGTGCCAGC-3’,	60°C	> 185 & < 200	Milan *et al.,* 2020
	Reverse	5′-A CATAGTGGGGTATCTAATCCCAGTTTG-3′			
NeoFish	Forward	5′-CGCCGTCGCAAGCTTACCCT-3′	60°C	> 165 & < 180	Miya *et al.*, 2015
	Reverse	5′-AGTGACGGGCGGTGTGTGC-3′			
Teleo	Forward	5′-ACACCGCCCGTCACTCT-3′	55°C	> 60 & < 75	Valentini *et al.,* 2016
	Reverse	5′- ACTTCCGGTACACTTACCATG-3′			
Illumina adapter	Forward	5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3′			
	Reverse	5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3′			
MiFish, NeoFish and Teleo primers have the respective Illumina adapter concatenated to the 5’ end.					

Table S4: Primers used in the second PCR of the library’s preparation

Orientation	**Sequencing adapter sequence**
Forward	5′-AATGATACGGCGACCACCGAGATCTACA-index-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′
Reverse	5′-CAAGCAGAAGACGGCATACGAGAT-index-TGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3′

Table S5A. Mock community containing 40 species from both river basins combined, using equal input DNA concentrations (normalized)

Expected species		São Francisco & Jequitinhonha mock community Normalized DNA input						
			NeoFish			MiFish		
		input DNA (%)	RRA (%)	Num ASVs	Num OTUs	RRA (%)	Num ASVs	Num OTUs
1	*Astyanax fasciatus*	2.33	5.585	2	1	1.409	2	2
2	*Astyanax lacustris*	2.33	4.898	1	1	1.079	1	1
3	*Australoheros sp*	2.33	0.715	1	1	0.538	1	1
4	*Brycon orthotaenia*	2.33	0.063	1	1	2.404	2	2
5	*Characidium lagosantense*	2.33	3.421	2	2	1.460	1	1
6	*Crenicichla lepidota*	2.33	8.557	1	1	3.384	1	1
7	*Cyphocharax gilbert*	2.33	2.820	1	1	2.110	1	1
8	*Delturus brevis*	2.33	0.647	1	1	0.323	1	1
9	*Eigenmannia virescens*	2.33	6.943	3	3	3.087	1	1
10	*Eugerres brasilianus*	2.33	12.817	1	1	5.046	1	1
11	*Franciscodoras marmoratus*	2.33	2.155	1	1	1.075	1	1
12	*Gymnotus carapo*†	4.66†	3.649	2	2	4.141	2	2
13	*Hoplias brasiliensis/ intermedius* ‡	4.66‡	4.328	1	1	2.352	2	**2**
14	*Hoplias malabaricus*	2.33	1.767	1	1	1.156	1	1
15	*Hypomasticus steindachneri*	2.33	1.029	1	1	0.503	1	1
16	*Hypostomus alatus*	2.33	1.394	1	1	0.597	1	1
17	*Hypostomus nigrolineatus*	2.33	1.222	1	1	0.575	1	1
18	*Imparfinis minutus*	2.33	**0**	**0**	**0**	0.847	1	1
19	*Megaleporinus elongatus*	2.33	1.500	1	1	0.496	1	1
20	*Megaleporinus garmani*	2.33	8.825	1	1	4.986	1	1
21	*Microglanis leptostriatus*	2.33	**0**	**0**	**0**	0.216	1	1
22	*Moenkhausia costae*	2.33	0.688	1	1	32.448	3	1
23	*Moenkhausia sanctaefilomenae*	2.33	1.397	1	1	0.507	1	1
24	*Myleus micans*	2.33	0.170	1	1	0.048	1	1
25	*Pamphorichthys hollandi*	2.33	0.027	1	1	0.008	1	1
26	*Phalloceros uai*	2.33	0.075	1	1	2.681	1	1
27	*Pimelodus maculatus*	2.33	0.010	1	1	10.297	3	2
28	*Pimelodus pohli*	2.33	**0**	**0**	**0**	1.886	2	1
29	*Prochilodus argenteus/hartii* **	4.66**	2.560	1	1	1.231	1	1
30	*Prochilodus costatus**	4.66*	3.536	1	1	2.35	2	2
31	*Pterygoplichthys etentaculatus*	2.33	0.539	1	1	0.195	1	**1**
32	*Rhamdia quelen*	2.33	0.061	1	1	0.499	1	1
33	*Roeboides xenodon*	2.33	1.394	1	1	3.201	1	1
34	*Serrasalmus brandtii*	2.33	5.659	1	1	1.780	1	1
35	*Steindachneridion amblyurum*	2.33	**0**	**0**	**0**	0.430	1	1
36	*Tetragonopterus chalceus*	2.33	5.373	4	4	1.993	1	1
37	*Trachelyopterus galeatus**	4.66*	5.036	1	1	2.895	1	1
38	*Wertheimeria maculata*	2.33	1.087	1	1	0.358	1	1
TOTAL without contaminations			**34**	**42**	**41**	**38**	**48**	**44**
39	*Acestrorhynchus lacustris*	0	0.00597	1	1	0.00477	1	1
40	*Geophagus brasiliensis*	0	0.00896	1	1	0	0	0
41	*Pseudoplatystoma corruscans*	0	0	0		0.00953	1	1
TOTAL with contaminations			**36**	**44**	**43**	**40**	**50**	**46**

†These species have twice the standard DNA input because they are represented by one specimen from each of the São Francisco & Jequitinhonha mock communities

‡These species were grouped since their amplicons cannot be distinguished by both MiFish and NeoFish amplicons.

Table S5B. Mock community containing 23 species from São Francisco River basin, using both normalized and skewed input DNA concentrations

Expected species		São Francisco mock community													
		Normalized DNA input							Skewed DNA input						
			NeoFish			MiFish				NeoFish			MiFish		
		input DNA (%)	RRA (%)	Num ASVs	Num OTUs	RRA (%)	Num ASVs	Num OTUs	input DNA (%)	RRA (%)	Num ASVs	Num OTUs	RRA (%)	Num ASVs	Num OTUs
1	*Astyanax fasciatus*	4.35	9.783	1	1	3.412	1	1	1.53	1.629	1	1	1.044	1	1
2	*Brycon orthotaenia*	4.35	0.103	1	1	5.457	1	1	0.74	0.015	1	1	0.724	2	2
3	*Characidium lagosantense*	4.35	6.908	2	2	3.186	1	1	3.03	2.707	2	2	2.833	1	1
4	*Crenicichla lepidota*	4.35	17.894	1	1	9.536	1	1	0.44	0.467	1	1	0.633	1	1
5	*Eigenmannia virescens*	4.35	13.490	3	3	6.115	1	1	1.05	1.445	1	1	1.632	1	1
6	*Franciscodoras marmoratus*	4.35	3.745	1	1	2.067	1	1	0.29	0.138	1	1	0.143	1	1
7	*Gymnotus carapo*	4.35	3.517	1	1	5.010	1	1	3.32	1.992	1	1	5.165	1	1
8	*Hoplias intermedius*	4.35	5.575	1	1	3.435	2	2	21.72	49.289	1	1	43.811	2	2
9	*Hypostomus alatus*	4.35	2.261	1	1	1.136	1	1	1.15	0.455	1	1	0.417	1	1
10	*Imparfinis minutus*	4.35	**0**	**0**	**0**	1.825	1	1	0.98	**0**	**0**	**0**	0.500	1	1
11	*Microglanis leptostriatus*	4.35	**0**	**0**	**0**	0.316	1	1	0.31	**0**	**0**	**0**	0.043	1	1
12	*Moenkhausia sanctaefilomenae*	4.35	2.631	1	1	1.018	1	1	1.32	0.489	1	1	0.418	1	1
13	*Myleus micans*	4.35	0.263	1	1	0.074	1	1	51.58	0.757	1	1	0.646	1	1
14	*Pamphorichthys hollandi*	4.35	0.030	1	1	0.025	1	1	0.35	0.005	1	1	0.003	1	1
15	*Phalloceros uai*	4.35	0.074	1	1	7.696	1	1	0.71	0.015	1	1	0.817	1	1
16	*Pimelodus maculatus*	4.35	0.024	1	1	23.411	4	2	0.58	**0**	**0**	**0**	2.564	3	1
17	*Pimelodus pohli*	4.35	**0**	**0**	**0**	3.250	1	1	5.24	**0**	**0**	**0**	3.120	1	1
18	*Prochilodus costatus*	4.35	2.775	1	1	1.540	2	2	2.03	2.393	1	1	2.541	2	2
19	*Pterygoplichthys etentaculatus*	4.35	0.827	1	1	0.356	1	1	0.3	0.033	1	1	0.040	1	1
20	*Roeboides xenodon*	4.35	2.109	1	1	7.575	1	1	1.25	0.458	1	1	1.918	1	1
21	*Serrasalmus brandtii*	4.35	11.561	1	1	4.700	1	1	0.47	0.554	1	1	0.362	1	1
22	*Tetragonopterus chalceus*	4.35	9.900	4	4	4.395	1	1	0.73	36.460	5	4	29.734	2	1
23	*Trachelyopterus galeatus*	4.35	6.470	1	1	4.411	1	1	0.88	0.688	1	1	0.873	1	1
TOTAL without contaminations			**20**	**26**	**26**	**23**	**29**	**27**		**19**	**24**	**23**	**23**	**29**	**26**
24	*Astyanax lacustris*	0	0.004	1	1	0	0	0	0	0	0	0	0	0	0
25	*Delturus brevis*	0	0	0	0	0	0	0	0	0	0	0	0.002	1	1
26	*Moenkhausia costae*	0	0	0	0	0.006	1	1	0	0	0	0	0	0	0
27	*Acestrorhynchus lacustris*	0	0.010	1	1	0.012	1	1	0	0	0	0	0.005	1	1
28	*Curimatella lepidura*	0	0.010	1	1	0.002	1	1	0	0.012	1	1	0.009	1	1
29	*Geophagus brasiliensis*	0	0.020	1	1	0	0	0	0	0	0	0	0	0	0
30	*Leporinus reinhardti*	0	0.010	1	1	0	0	0	0	0	0	0	0	0	0
31	*Planaltina myersi*	0	0.010	1	1	0	0	0	0	0	0	0	0	0	0
32	*Pseudoplatystoma corruscans*	0	0	0	0	0	0	0	0	0	0	0	0.006	1	1
TOTAL with contaminations			**26**	**32**	**32**	**26**	**32**	**30**		**20**	**25**	**24**	**26**	**33**	**30**

Table S5C. Mock community containing 17 species from Jequitinhonha River Basin, using both normalized and skewed input DNA concentrations

	Expected species	Jequitinhonha mock community																		
		Normalized DNA input										Skewed DNA input								
		input DNA (%)	NeoFish			MiFish			Teleo			input DNA (%)	NeoFish			MiFish			Teleo		
			RRA (%)	Num ASVs	Num OTUs	RRA (%)	Num ASVs	Num OTUs	RRA (%)	Num ASVs	Num OTUs		RRA (%)	Num ASVs	Num OTUs	RRA (%)	Num ASVs	Num OTUs	RRA (%)	Num ASVs	Num OTUs
1	*Astyanax lacustris*	5.56	22.360	4	2	3.700	1	1	0.0000246	1	1	0.26	3.770	1	1	0.230	1	1	**0**	**0**	**0**
2	*Australoheros sp*	5.56	1.070	1	1	4.030	1	1	7.200	1	1	0.45	0.140	1	1	0.670	1	1	1.900	1	1
3	*Cyphocharax gilbert*	5.56	3.950	2	2	16.140	1	1	8.960	1	1	0.78	1.050	1	1	3.030	1	1	3.470	1	1
4	*Delturus brevis*	5.56	1.760	2	2	1.590	1	1	0.460	1	1	8.49	6.390	1	1	2.580	1	1	0.950	2	2
5	*Gymnotus carapo*	5.56	0.460	1	1	12.340	1	1	0.030	1	1	6.32	0.800	1	1	58.790	1	1	0.060	1	1
6	*Hoplias brasiliensis*	5.56	3.870	2	2	3.850	1	1	2.150	1	1	4.92	2.890	1	1	1.330	1	1	1.370	1	1
7	*Hoplias malabaricus*	5.56	2.270	1	1	3.180	1	1	1.870	1	1	10.57	12.660	1	1	11.060	1	1	11.960	1	1
8	*Hypomasticus steindachneri*	5.56	2.960	1	1	8.630	1	1	3.980	1	1	6.69	1.820	1	1	3.960	1	1	3.150	1	1
9	*Hypostomus nigrolineatus*	5.56	3.590	2	1	2.920	1	1	1.170	2	2	12.78	11.570	2	1	3.870	1	1	2.050	1	1
10	*Megaleporinus elongatus*	5.56	6.400	2	2	1.460	1	1	9.990	1	1	12.56	36.200	3	3	1.870	1	1	35.590	1	1
11	*Megaleporinus garmani*	5.56	35.050	6	2	17.500	1	1	43.930	1	1	2.58	10.570	2	2	2.480	1	1	16.140	1	1
12	*Prochilodus argenteus/hartii*‡	11.12‡	7.87	2	2	11.510	1	1	9.450	1	1	11.69‡	7.910	1	1	1.420	1	1	14.670	1	1
13	*Prochilodus costatus*	5.56	4.060	3	3	9.740	1	1	**0**	**0**	**0**	2.19	1.440	1	1	7.900	1	1	**0**	**0**	**0**
14	*Rhamdia quelen*	5.56	0.030	1	1	1.370	1	1	1.130	1	1	5.41	0.010	1	1	0.320	1	1	0.580	1	1
15	*Steindachneridion amblyurum*	5.56	**0**	**0**	**0**	0.790	1	1	3.190	1	1	3.31	**0**	**0**	**0**	0.150	1	1	2.080	1	1
16	*Trachelyopterus galeatus*	5.56	1.960	1	1	0.600	1	1	4.780	1	1	5.74	1.000	1	1	0.170	1	1	4.780	1	1
17	*Wertheimeria maculata*	5.56	2.320	1	1	0.630	1	1	1.680	1	1	5.26	1.760	1	1	0.130	1	1	1.230	1	1
TOTAL without contaminations			**16**	**33**	**25**	**17**	**17**	**17**	**16**	**17**	**17**		**16**	**20**	**19**	**17**	**17**	**17**	**15**	**16**	**16**
18	*Acinocheirodon melanogramma*	0	0	0	0	0.010	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
19	*Brycon orthotaenia*	0	0.0000213	1	1	0	0	0	0	0	0	0	0	0	0	0.010	1	1	0	0	0
20	*Moenkhausia costae*	0	0	0	0	0.0000406	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
21	*Serrasalmus brandtii*	0	0.020	1	1	0.010	1	1	0.030	1	1	0	0.020	1	1	0.010	1	1	0.010	1	1
TOTAL with contaminations			**20**	**35**	**28**	**21**	**20**	**20**	**17**	**18**	**18**		**18**	**21**	**20**	**19**	**19**	**19**	**16**	**17**	**17**

‡ These species were grouped since their amplicons cannot be distinguished by MiFish, NeoFish or Teleo amplicons.

Table S6: Read counts along all pipeline steps

Sample ID	Primer	N-cleaned	Filtered	Raw FWD	Raw REV	Denoised FWD	Denoised REV	Merged	Non-Chimeric
Skewed JQmc	MiFish	221501	216073	257970	257970	215788	215613	215386	165947
	NeoFish	120679	118671	172630	172630	118162	118375	117876	56513
	Teleo	226919	225973	229486	229486	225723	225831	225583	193654
Normalized JQmc	MiFish	166421	162219	167689	167689	161314	161435	160560	98453
	NeoFish	640011	629879	649695	649695	627377	628176	625703	234543
	Teleo	147991	147341	167693	167693	147070	147305	147043	122012
Normalized SFJQmc	MiFish	183644	176948	183740	183740	176237	176253	175607	146830
	NeoFish	175318	171921	175402	175402	170583	171100	169795	100492
Skewed SFmc	MiFish	153438	148570	154286	154286	148371	148228	148068	117579
	NeoFish	165008	160603	167221	167221	160232	160334	159990	110570
Normalized SFmc	MiFish	155743	150146	156541	156541	149566	149461	148984	122992
	NeoFish	175914	171914	178463	178463	170993	171199	170300	105146

Table S7: Summary of the Pearson's correlation coefficient between RRA and input DNA concentration for each mock community

Primer	Mock Community	r.squared	adj.r. squared	sigma	statistic	p.value	df	logLik	AIC	BIC	deviance	df.residual	nobs
NeoFish	JQmc	0.419	0.371	0.0336	8.67	0.0123*	1	28.7	-51.4	-49.5	0.0136	12	14
NeoFish	SFmc	0.00438	-0.0578	0.115	0.0704	0.794	1	14.5	-22.9	-20.2	0.212	16	18
MiFish	JQmc	0.0173	-0.0529	0.0419	0.247	0.627	1	29.1	-52.3	-50	0.0245	14	16
MiFish	SFmc	0.0945	0.0514	0.102	2.19	0.154	1	20.9	-35.7	-32.3	0.22	21	23
Teleo	JQmc	0.172	0.103	0.0367	2.49	0.141	1	27.5	-48.9	-47	0.0162	12	14
NeoFish	SF JQmc	0.0326	0.0101	0.0852	1.45	0.235	1	48.0	-90.0	-84.6	0.312	43	45
MiFish	SF JQmc	0.0702	0.0496	0.0835	3.40	0.0718	1	51.0	-96.1	-90.5	0.314	45	47

6. CAPÍTULO III - ENVIRONMENTAL DNA METABARCODING OF FISH COMMUNITIES IN A NEOTROPICAL DAM RESERVOIR: A comparison of eDNA metabarcoding to traditional survey methods of fish from a Neotropical dam

6.1 Objetivos específicos

- Amplificar um fragmento de ~170bp do marcador 12S do eDNA obtido de amostras de água coletadas na IRR;
- Sequenciar ~170bp do marcador 12S do eDNA obtido de amostras de água coletadas na IRR;
- Fazer a identificação taxonômica de espécies de peixes em amostras ambientais através da comparação das sequências de DNA obtidas com a biblioteca de referência customizada;
- Comparar a detecção de espécies fornecida pelo eDNA *metabarcoding* e pelo método tradicional de pesca, utilizando redes de emalhar;
- Testar o potencial do eDNA *metabarcoding* como ferramenta de monitoramento em reservatórios para melhorar o conhecimento das comunidades de peixes de água doce.

ENVIRONMENTAL DNA METABARCODING OF FISH COMMUNITIES IN A NEOTROPICAL DAM RESERVOIR:

A comparison of eDNA metabarcoding to traditional survey methods of fish from a Neotropical dam

Izabela Santos Mendes[1,3], Naiara Guimarães Sales[2], Allan McDevitt[4], Paulo dos Santos Pompeu[5], Francisco de Andrade Neto[5], Heron Oliveira Hilário[1], Daniel Cardoso Carvalho[1,3]

[1] Conservation Genetics Lab, Postgraduate Program in Vertebrate Biology, Pontifical Catholic University of Minas Gerais, PUC Minas, Belo Horizonte, Brazil

[2] School of Science, Engineering and Environment, University of Salford, Crescent, Salford, UK

[3] Postgraduate Program in Genetics, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Brazil

[4]Atlantic Technological University, Galway-Mayo Institute of Technology, Ireland

[5]Federal University of Lavras, Minas Gerais, Brazil

ABSTRACT

Environmental DNA (eDNA) metabarcoding has emerged as a promising tool for monitoring fish diversity in a range of diverse aquatic realms. However, eDNA-based monitoring of fish communities found in hydroelectric dam reservoirs has received relatively limited attention, especially in megadiverse regions such as the Neotropics. Herein, we evaluate eDNA metabarcoding as a biodiversity monitoring tool together with capture-based fishing methods, in order to compare their efficacy and analyse their complementarity in the characterization of Neotropical fish communities. We obtained 147 water samples from three different sample points across the Irapé hydroelectric power plant reservoir (IRR) in eight sampling events, over three years (2019-2021). The ribosomal RNA 12S mitochondrial marker (~170 bp) was amplified and using a curated 12S reference library, the sequences were associated with 115 Molecular Operational Taxonomic Units (MOTU), of which at least 30 were assigned to species level, represented by five taxonomic orders, 16 families and 25 genera., although 11 of the 30 species identified (~36%) could not be detected to species level. Our results indicate that eDNA metabarcoding was able to identify a higher number of species solely (n=23) when compared to the traditional capture-based fishing gear (n=7). Considering both methodologies together, at least 36 species were identified. A significant difference in fish communities was recovered when compared the species detected by both methods (PERMANOVA, R2=0.140, p=0.001). However, no significant effect of sampling season was observed for both methods together. Overall, the number of sequence reads and fish biomass/catch in the IRR reservoir were not correlated. Therefore, our study suggests that eDNA metabarcoding may be used as a complementary tool to traditional capture-based fishing methods for the investigation and biomonitoring of fish diversity in hydropower dam reservoirs.

6.2 INTRODUCTION

Global biodiversity is decreasing rapidly due to environmental changes and human activities, such as climate change, habitat degradation, biological invasion, and overexploitation (Brauer et al., 2016). Freshwater ecosystems, one of the most fragile ecosystems on Earth, currently face unprecedented levels of anthropogenic destruction (Reid et al., 2019) caused by pollution, deforestation and habitat loss and fragmentation. Such impacts strongly affect freshwater species, leading to population declines, resulting in the highest rate of extinction among vertebrates (Burkhead 2012). For instance, the impact of dams is reported as one of the main stressors on freshwater fishes (He et al., 2018), as it blocks migration routes for the migratory species, reduces access to fish spawning areas and causes habitat loss and degradation in downstream and upstream reaches (Liermann et al., 2012). Because of this, fish population data is essential for defining the most appropriate actions for conservation and management (Kelly et al., 2014; Sales et al., 2018), which is ultimately an important issue for protecting freshwater ecosystems (Elliott et al., 2010).

Traditional capture-based fishing methods (e.g., netting, fishing) may be biased and are frequently expensive (Hickley and Starkie 1985; Bonar et al. 2009). This is even more relevant to large aquatic systems, such as Neotropical basins, where sampling fish communities are highly complex, time-consuming, and laborious (Hickley and Starkie 1985; Sutherland 2006). More recently, DNA shed from fishes and other organisms via metabolic waste, damaged tissue, or dead skin cells has been detected in various aquatic environments, including ponds, rivers and marine waters (environmental DNA or eDNA; Ficetola et al. 2008; Jerde et al. 2011; Kelly et al., 2014) and it has shown great promise for detecting individual species presence and distribution in water bodies through to characterising whole communities using high-throughput DNA sequencing known as eDNA metabarcoding (Minamoto et al. 2012; Goldberg et al. 2013; Laramie et al. 2015).

Since eDNA metabarcoding has been shown to be a reliable non-invasive and cost-effective alternative for traditional fish surveys (Valentini et al. 2016; Deiner et al. 2017; Yamamoto et al. 2017; McDevitt et al., 2019), it has now been widely used in fisheries management and fish diversity monitoring in freshwater ecosystems (Ruppert et al., 2019, Zhang et al., 2020). For example, it has been applied to: detect invasive, rare, or threatened species (Jerde et al., 2013, Mahon et al., 2013, Yamanaka and Minamoto, 2016, Wilcox et al., 2013) for determining biomass within an ecosystem

(Doi et al., 2017); for monitoring movements and spawning activity of fish (Erickson et al., 2016, Yamanaka and Minamoto, 2016); and for studying population genetics (Sigsgaard et al., 2016). Moreover, the use of fish eDNA facilitates investigations into patterns of extirpation, invasive species detection, and dynamics of species richness (Lacoursière-Roussel et al. 2018; Sales et al., 2021).

Recent studies have demonstrated its efficiency in different aquatic environments and show how it compares favourably to, or even outperforms, traditional sampling methods in terms of species detection (McDevitt et al., 2019; McElroy et al., 2020). Agreement between how many and which species are detected has ranged from nearly identical (Olds et al., 2016) to very disparate (Cilleros et al., 2019). Moreover, the comparison between eDNA metabarcoding and traditional approaches underlines that these approaches can be considered complementary, by targeting different species, sampling greater diversity, and increasing taxonomic resolution (Deiner et al., 2017; Jerde et al., 2019, Keck et al., 2022). Despite its considerable ability to detect fish species in aquatic systems, most of the studies are applied to study fish communities in lakes and rivers, as well as coastal fish communities (Xiong et al., 2022). Currently, there is a lack of a broad understanding of how eDNA metabarcoding will be a well-established method for fish monitoring in reservoirs. The application of eDNA analysis for biodiversity assessment and ecological monitoring in reservoirs is relatively limited (but see Lim et al. 2016, Perez et al. 2017, Hayami et al. 2020). For example, Lim et al. (2016) concluded that eDNA could be useful for freshwater bioassessment in reservoirs by applying eDNA analysis using universal metazoan primers, and Hayami et al. (2020) investigated the seasonal change of fish communities in reservoirs using eDNA metabarcoding. More recently, Dal Pont et al. (2021) demonstrate the analytical efficiency of the metabarcoding approach for monitoring fish species in a fish pass system, in Brazil, reinforcing the need to supplement the reference databases with additional information on the species present, particularly in the megadiverse regions, such as Brazilian basins, as there are many species yet to be described (Cilleros et al., 2019; Sales et al., 2019, 2021; Jackman et al., 2021; Dal Pont et al., 2021).

Brazil has a huge hydropower generating capacity and power companies are obligated by law to conduct regular fish monitoring programs (Política Nacional de Meio Ambiente, Law 6.938/1981, Resolução Conama number 001/1986; Resolução Conama number 237/1997). Thus, areas of the dam construction have some of the

most comprehensive knowledge of fish assemblage composition in comparison with other Brazilian regions and therefore offer an ideal opportunity to compare taxonomic surveys with molecular approaches. In this sense, here, we conducted an eDNA survey of freshwater fish species in the Irapé Hydroelectric Power Plant reservoir (IRR), the deepest reservoir in Brazil and the second-highest dam in Latin America (Cachapuz, 2006; Godinho & Loures, 2017) and compared eDNA metabarcoding results with established traditional capture methods, i.e., gill nets, through three years of sampling. The aims of this study were to: (1) assess the suitability of eDNA metabarcoding as a biomonitoring tool and (2) evaluate the accuracy of the eDNA data through a comparison with traditional capture data and determine the similarities and differences between the two methods.

6.3 MATERIALS AND METHODS

6.3.1 Study Area

This study was conducted in the Irapé Hydroelectric Power Plant reservoir (IRR), located in the Jequitinhonha River basin, northern Minas Gerais state, Southeast Brazil (16°44′15″S, 42°34′30″W) (Figure 1). The IRR, operating since 2016, is formed by the dam built along the Jequitinhonha and Itacambiruçu rivers. It is a large accumulation reservoir with 100 km of extension and 280 days of water residence time (Silva et al., 2020). The 142.95 km^2 reservoir is formed by a 205-metre-high dam, the highest in Brazil and the second highest in Latin America (Cachapuz, 2006; Godinho & Loures, 2017).

The Jequitinhonha River basin is inserted in a transition area among two biodiversity hotspots (Cerrado and Atlantic Forest) and is characterised by tropical climate and environmental heterogeneity. Climatic characteristics of the Jequitinhonha River basin vary from semi-arid to humid, with an average rainfall of 600–1600 mm irregularly distributed throughout the year (Ferreira & Silva, 2012). The rainy season is concentrated in 6 months (October to March) (Ferreira & Silva, 2012; Gonçalves, 1997), with the December–February quarter responsible for more than 50% of total precipitation (Gonçalves, 1997).

Considering its biodiversity, the Jequitinhonha River basin is characterised by a high endemism (67% of its fish species are endemic, e.g., *Wertheimeria* sp., *Delturus* sp.; Reis et al., 2016, Vono & Birindelli, 2007) and a significant number of endangered species (*Steindachneridium amblyurum, Rhamdia jequitinhonha*, *Nematocharax venustus*) (Rosa & Lima, 2008). The known fish diversity of this catchment included 114 described fish species (including 16 introduced species; Andrade-Neto, 2010, Sales et al., 2021). Compared to neighbouring basins, Jequitinhonha has long been seen as a low-biodiversity ecosystem. Although its reduced species richness had been linked to historical geological and geographical features (Andrade-Neto, 2010), more recent factors to explain low biodiversity in the catchment, such as the lack of adequate surveys and anthropogenic activities (e.g., mining, exploitation of fisheries, and deforestation) have been considered as the geological history of the Jequitinhonha is very similar to that of adjacent basins (e.g., Doce river, Mucuri river). Moreover, the Jequitinhonha is particularly affected by the impact of dams on the main river course and tributaries, and recent studies highlighted the occurrence of introduced, undescribed and cryptic species (Pugedo et al., 2016, Sales et al., 2018), leading to

the description of several new species (Jerep et al., 2016; Nielsen et al., 2017; Zawadzki et al., 2016).

6.3.2 Traditional sampling methods

Eight sampling campaigns were conducted from May 2019 to August 2021 during rainy and dry seasons, across three sampling areas along the reservoir were sampled including the main river Jequitinhonha (IR10), the water intake located upstream of the dam (IR2) and its main tributary, the Itacambiruçu River (IR3) (Figure 1). Gillnets were used to sample fish species at the IRR. On the same day and after eDNA sampling at each sampling area, 30 linear metres of gill nets of different mesh sizes (stretch mesh; 2.4, 3, 4, 5, 6, 7, 8, 10, 12, 14 and 16 cm) were used as such nets can capture fish of a very broad size range. Nets were deployed in the river margins at each sampling site. Fishing effort was 24 hr per sampling site and the nets were checked every 12 h, at 6 pm and 6 am during three days. Catch per unit effort (CPUE) and biomass per unit effort (BPUE) were calculated from the survey.

Fish were anaesthetized with clove oil, humanely killed, fixed in 10% formalin and identified using the taxonomic identification books/catalogues and by experts in the field. Voucher specimens were deposited in the ichthyological collection of the Laboratory of Fish Ecology of the Federal University of Lavras (UFLA). This study was approved by the institutional ethics committee (collection permission nº 72631).

6.3.3 Local reference database

In order to improve taxonomic assignment, the custom reference database developed by Hilário et al. (2023) was used together with Salford's Lab databases (RDMyia-2022). Briefly, 496 sequences, corresponding to 294 fish species, were sequenced and included in a customised reference database for Neotropical fishes, based on rRNA 12S.

6.3.4 eDNA sampling and processing

eDNA sampling was applied at the same period as gillnet sampling. For eDNA collection, three sample sites in the three sampling areas along the reservoir were sampled (Figure 1).

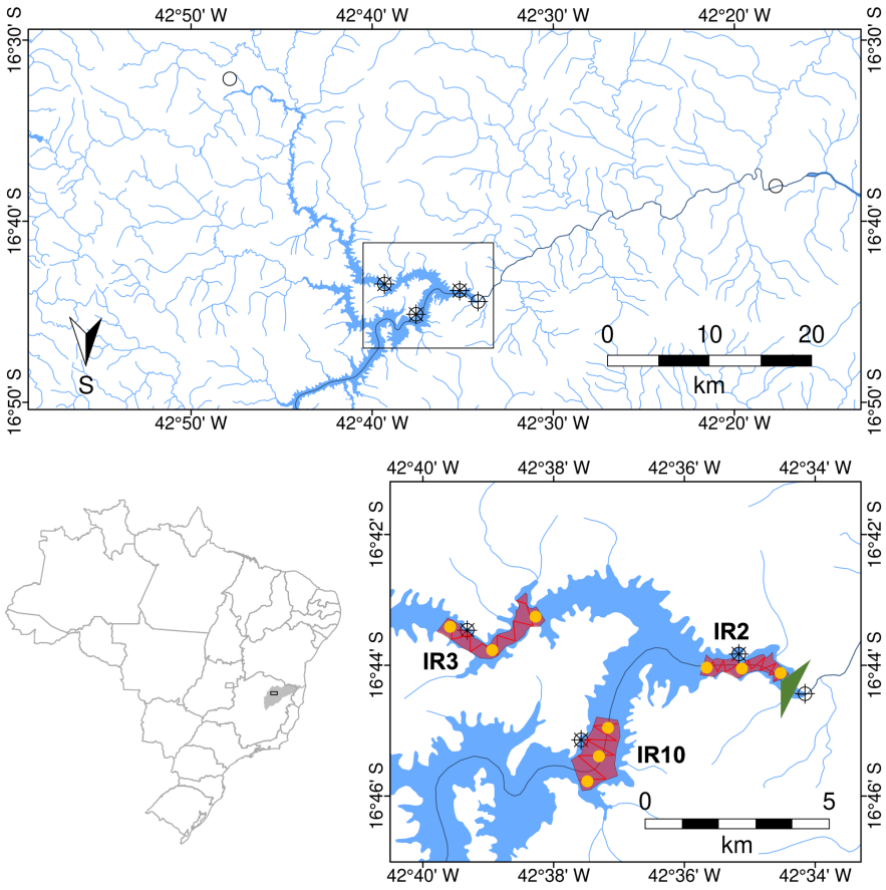


Figure 1. Map of the IRR displaying sampling sites of both eDNA (+) and fishing gill nets (o). Yellow circles distributed in the red area demonstrating the sampled areas. The green bar represents the dam location.

From each of the sampling points within the IRR, three surface water replicates were taken from three different sites (from the left side, middle and right side) using 500 ml water bottles resulting in a total of nine replicates per reservoir stretch (9× 500 ml), 27 replicates per campaign (except for the campaign conducted in November/2019, when we collected more than 54 replicates, that is, 18 samples). Then, after all the campaigns, the sum of samples collected was 81, which corresponded to 243 replicates (Table 1). In order to prevent contamination, a contamination control approach was used in both fieldwork and laboratory operations. Then, samples were collected before the start of the netting sampling and disposable collection bottles were used. Disposable gloves were changed at all times and surfaces

of equipment used were cleaned with water after each use and after being treated with a 50% bleach solution for 10 minutes. In order to check for potential contamination during the filtration stage, filtration blanks were run between each sample site right before the next filter. Moreover, extraction and PCR controls were additionally included to test potential cross-contamination in the laboratory. Water samples were kept on ice before filtering and were filtered approximately 5-6 hours after collection using the EZ-fit manifold and the automatic vacuum pump (Merck Millipore) and 0.45µm Mixed Cellulose Membrane (MCE, Merck Millipore). All samples were kept at −20°C in microcentrifuge tubes containing silica beads (Majaneva et al., 2018) until extraction procedures.

Table 1. Samples collected by campaign, season, sampling site and the number of samples collected at each site

Campaign	Sampling site	Date of sampling	Season	N of samples	N of replicates
01	IR10	2019/05	Dry	3	9
	IR2	2019/05	Dry	3	9
	IR3	2019/05	Dry	3	9
02	IR10	2019/08	Dry	3	9
	IR2	2019/08	Dry	3	9
	IR3	2019/08	Dry	3	9
03	IR10	2019/11	Rainy	6	18
	IR2	2019/11	Rainy	6	18
	IR3	2019/11	Rainy	6	18
04	IR10	2020/02	Rainy	3	9
	IR2	2020/02	Rainy	3	9
	IR3	2020/02	Rainy	3	9
05	IR10	2020/11	Rainy	3	9
	IR2	2020/11	Rainy	3	9
	IR3	2020/11	Rainy	3	9
06	IR10	2021/02	Rainy	3	9
	IR2	2021/02	Rainy	3	9
	IR3	2021/02	Rainy	3	9
07	IR10	2021/05	Dry	3	9
	IR2	2021/05	Dry	3	9
	IR3	2021/05	Dry	3	9
08	IR10	2021/08	Dry	3	9
	IR2	2021/08	Dry	3	9
	IR3	2021/08	Dry	3	9

eDNA extraction from the filters was conducted using the DNeasy PowerWater Kit (Qiagen), with minor modifications: (1) addition of 500 µL more of the lysis buffer, (2) addition of 20 µL of Proteinase K and (3) heating the elution buffer to 55 degrees Celsius prior to adding it to the samples. Samples were amplified with the MiFish primer

set (~170bp of the rRNA 12S, Miya et al., 2015), using a single-step protocol to minimise bias in individual reactions according to the protocol described in Sales et al. (2019). Each sample corresponded to each sample site was tagged with a primer set that was designed with seven-base sample-specific oligo-tags and a variable number (2-4) of leading Ns to increase variability in amplicon sequences (Table S1). PCRs were replicated three times for each sample. The PCR reaction consisted of a total volume of 20 µL including 10 µl AmpliTaq Gold™ 360 Master Mix (1X; Applied Biosystems); 0.16 µl of BSA (5mg); 1 µl of each of the two primers (5 µM); 5.84 µl of ultra-pure water and 2 µl of eDNA template. The PCR profile included an initial denaturing step of 95°C for 10 min, 40 cycles of 95°C for 30s, 60°C for 45s, and 72°C for 30s and a final extension step of 72°C for 5 min. Amplifications were confirmed by electrophoresis in a 1.5% agarose gel stained with SYBR™ Safe DNA Gel Stain (KASVI). Library preparation was conducted using a Collibri™ PS DNA Library Prep Kit (Thermo Fisher), a PCR-free library preparation kit. Size selection was performed using 1.8x Agencourt AMPure XP (Beckman Coulter), libraries were then quantified by Qubit fluorometer (Invitrogen) following the manufacturer's protocol and pooled in equimolar concentrations along with 5% PhiX (MiniSeq Mid Output Kit 300-cycles, Illumina). The libraries were run at final molarity of 1.25 pM in the Illumina MiniSeq platform run.

6.3.5 Bioinformatic analyses

The metabarcoding pipeline used for data analysis was based on the OBITools software suite (Boyer et al., 2016) following the protocol described in Sales et al. (2021). FastQC was used to assess the quality of the reads and a length filter (command *obigrep*) was used to select fragments of 140–190 bp and to remove reads with ambiguous bases, followed by a clustering method using SWARM at d = 1 (Mahé et al., 2014). The taxonomic assignment was performed using *ecotag* with a custom reference database built including all 12S sequences available from GenBank and the local reference database built in this study. *Ecotag* function works in two phases: firstly, it performs a search of the assigned reference database to locate the sequence with the highest overall similarity to the query sequence; then, the similarity value obtained from the first step is set as the threshold for searches of additional sequences, equal to or lower than that of the threshold value within the assigned database.

Stringent filtering steps were applied to the final dataset to remove Molecular Operational Taxonomic Units (MOTUs)/reads originating from sequencing errors or possible contamination to avoid false positives. All non-fish reads were removed from the dataset and MOTUs that were likely to have been carried over from contamination. To control for putative cross-contamination, the maximum number of reads recorded in the controls obtained during all sampling and process stages (field collection, DNA extraction, and DNA amplification) were subtracted from all samples. The final taxonomic assignment was conducted according to current fixed general thresholds: MOTUs were assigned at the species level when matching the reference sequence with > 97% similarity as performed in previous studies in the Neotropics (Cilleros et al., 2019; Sales et al., 2019, 2021; Jackman et al., 2021), at genus level with 95%–97% similarity, at the family level with 90%–95% similarity, and the highest taxonomic level of order was attributed to MOTUs with less than 90% similarity matching the reference sequences. MOTUs well represented but showing <97% similarity to references were presented and discussed as putatively belonging to species still absent in the reference database.

6.3.6 Statistical analyses

6.3.6.1 Species richness and β-diversity

All statistical analyses were performed in R v3.5.1 (https://www.R-project.org/). Sample replicates were pooled for each site (n=9 replicates per site, i.e., one sample for IR2/IR3/IR10 each) prior to the following statistical analyses.

For the analysis of the diversity contained within the eDNA dataset (MOTU richness/α-diversity and β-diversity), the data were analysed with a presence/absence approach as suggested by Li et al. (2018). α-diversity (MOTU richness) was estimated as the number of MOTUs assigned to species level, at each sample site. As reported by other eDNA-based studies (Sales et al., 2021), the use of MOTUs as a proxy for species is common, however, it could be highly biased, as it could be influenced by the existence of cryptic species and the universal thresholds used in bioinformatic analyses (Pawlowski et al., 2018), especially when we are dealing with the megadiversity of the Neotropical region. Alternatively, when the analyses are based just on the species identified by the association of the MOTUs to the reference database, we are subjected to the underestimation of the richness, because of the

incompleteness of the reference database and the low taxonomic resolution of the fragment analysed. In this sense, in order to overcome these limitations, we focused on both richness approaches. Two statistical methods were employed to estimate the species richness of the study area. Firstly, accumulation curves for the species/MOTUs richness were done with the "*specaccum*" function using the vegan package (Oksanen et al., 2016). In the second approach, we used the Chao I index in order to estimate the total number of species/MOTUs in the survey area. β-diversity was obtained by applying the Jaccard distance using the vegan package version 2.5-2 using the "*vegdist*" command (Oksanen et al., 2013). For combined eDNA and traditional method curves, a species was considered present if it was found in either gillnets or eDNA metabarcoding. To visualise the relationships amongst sampling sites, we obtained PCoA plots using the β-diversity matrix ("*cmdscale*" command). To examine the potential effect of season (dry *vs* rainy), year (2019 *vs* 2020 *vs* 2021) and sampling sites (IR10 *vs* IR2 *vs* IR3) on community composition, a Permutational Multivariate Analysis of Variance (PERMANOVA) was then performed with 1000 permutations, applying the Jaccard dissimilarity index through the function *'adonis'* (vegan 2.5–2 R package).

6.3.6.2 Comparison with established fishing methods

We compared the performance of the eDNA approach and established fishing survey methods in two ways. First, accumulation curves for the species richness comparison were done with the "*specaccum*" function using the vegan package (Oksanen et al., 2016). eDNA and fishing samples were categorised according to the sampling site (i.e., IR10, IR2 and IR3), month (i.e., February, May, August and November) and year of collection (i.e., 2019, 2020 and 2021). Then, for all the statistical analyses, we considered each sample site (n=3) for each campaign (n=8) as a sampling unit numbered from one to 24. The taxonomic identification of fish assemblages and their relative abundances revealed by eDNA and established survey methods were exhibited by drawing species bar plots and Venn diagrams. Second, Spearman's rank correlation was used to investigate the correlation between the number of fish captured per unit of sampling effort (CPUE) and relative sequence abundances (RRA).

To examine the potential effect of methods (eDNA *vs* gillnets) and year (2019 *vs* 2020 *vs* 2021) on community composition, a PERMANOVA was then performed with 1000 permutations, applying the Jaccard dissimilarity index through the function *'adonis'*. We also used principal coordinates analysis (PCoA) to visualise the fish communities depicted by eDNA metabarcoding and gill nets. After standardising the data on the total proportion ("*decostand*" function from the vegan package), we compared the different communities using a Jaccard distance ("*vegdist*" function from the vegan package) while grouping by those approaches.

6.4 RESULTS

6.4.1 Library quality and raw data

A total of 30,111,673 reads were obtained in five Illumina runs after trimming, merging, and length filtering during bioinformatic analysis, including 213 water samples, representing 147 samples (including 63 water samples and 84 filtration, extraction and PCR blanks). 18,899,406 reads were kept after initial quality filtering and removal of chimaeras (Table S2). Then, after applying a subsequent conservative filtering step (retaining only reads taxonomically assigned to Actinopterygii, and removal of MOTUs containing less than five reads), the final dataset comprised 12,362,655 reads which were designated to 115 MOTUs (Table S2 and S3). The number of retained MOTUs varied considerably between filtered and unfiltered datasets and for several species, more than one MOTU was also recovered (Table S3).

6.4.2 Taxonomic assignment

As a result of the taxonomic assignment, a total of five different taxonomic orders (i.e., Characiformes, Cichliformes, Cyprinodontiformes, Siluriformes, and Gymnotiformes), 16 families, 25 genera and 30 fish "species" were identified (Table S3). When considering the sampling effort, the accumulation curves based on the number of MOTUs (Figure 2A) showed that with an increased sampling effort, eDNA metabarcoding could depict a higher fish diversity. On the other hand, when considering the number of species, the curve seems to be closer to the plateau (Figure 2B). From the 115 MOTUs identified, only one was identified to family level with the

fixed general threshold, whereas 38 were assigned solely at the genus level and 76 could be attributed to the species level (Table S3).

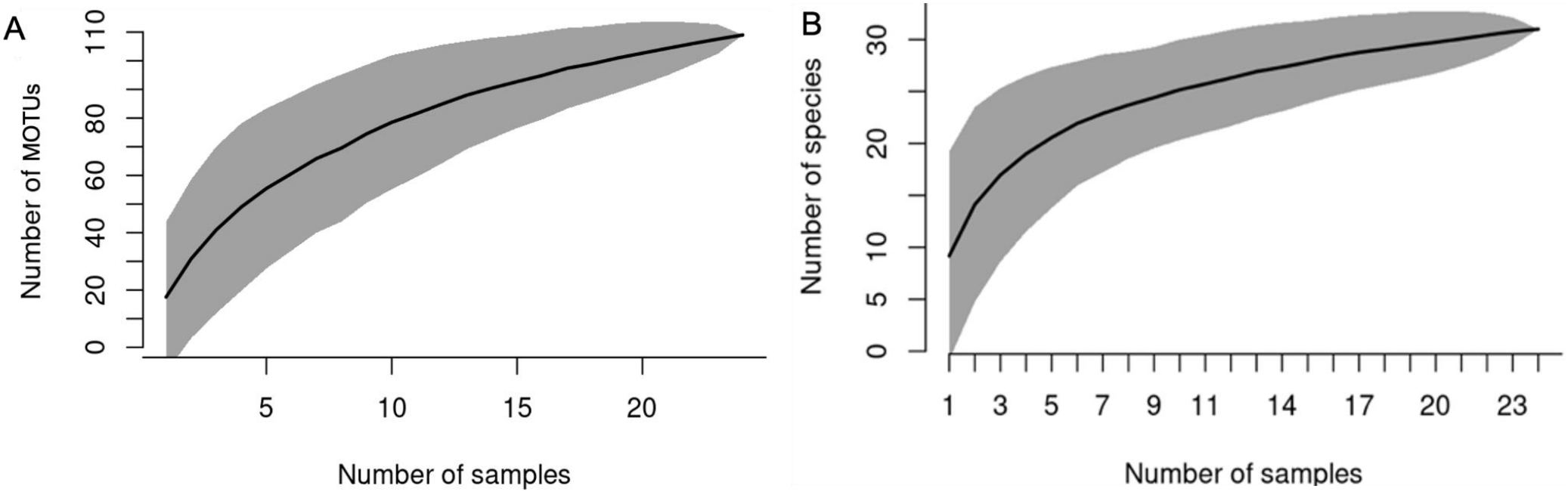


Figure 2. Accumulation curves for eDNA metabarcoding considering the number of MOTUs (A) and the number of species (B) for the number of samples.

A visualisation of the fish community composition was obtained based on the taxonomic identification at the order and family levels (Figure 3). As expected, the assemblage composition depicted by the stacked bar plots (Figure 3A) demonstrated a higher overall read count for Characiformes and Siluriformes. Furthermore, the stacked bar plots evidenced the fish diversity recovered from eDNA samples due to the occurrence of 16 families among the 115 analysed MOTUs (Figure 3B): Characiformes (8), Siluriformes (5), Cichliformes (1), Cyprinodontiformes (1) and Gymnotiformes (1).

Regarding Characiformes, we observed that this order represented from ~68% (Feb/2020) to ~94% (Aug/2021) of the species richness recovered from all campaigns. Together with the Gymnotiformes, represented solely by *Gymnotus carapo*, Characiformes and Gymnotiformes were the two orders observed in all campaigns. Within Characiformes, Characidae was the most representative family (~43 to 82%), followed by the Anostomidae (~0.0002 to 31%, Figure 3).

The Siluriformes order was the second most representative, varying from ~0.13% (Aug/2021) to ~30% (Feb/2020) and was not detected in August/2019 (Figure 3A). Within Siluriformes, Auchenipteridae, represented solely by *Trachelyopterus striatulus*, was the most representative family and was detected from Feb/2020 to Aug/2021 (Figure 3A, Table S4).

The representativeness of Cichliformes, Cyprinodontiformes and Gymnotiformes was similar, as these orders were represented solely by a single family (Figure 3B, Table S4). The Cichliformes was represented by the Cichlidae family and varied from ~0.2% (Nov/2020) to ~13% (May/2021) and was not detected in May/2019 (Figure 3A, Table S4). The Poeciliidae family represented the Cyprinodontiformes and was identified from Nov/2019 (~7.7%) to May/2021 (~0.3%) (Figure 3A, Table S4). The Gymnotiformes was represented by the Gymnotidae family and was most frequently detected in May/2021 (~9.0%) but was less abundant in Feb/2020 (~1.1%) (Figure 3A, Table S4).

We also observed that 2020 was the only year that all campaigns were able to identify the five orders recovered in this study (Figure 3A, Table S4). In 2021, all five orders were recovered in Feb and May/2021. However, the Cyprinodontiformes were not found in Aug/2021 (Figure 3A, Table S4). Within 2019, Nov/2019 was the only campaign where all orders were recovered. Species representing Characiformes,

Gymnotiformes and Siluriformes were recovered in May/2019 and Characiformes, Gymnotiformes and Cichliformes were recovered in Aug/2019 (Figure 3A, Table S4).

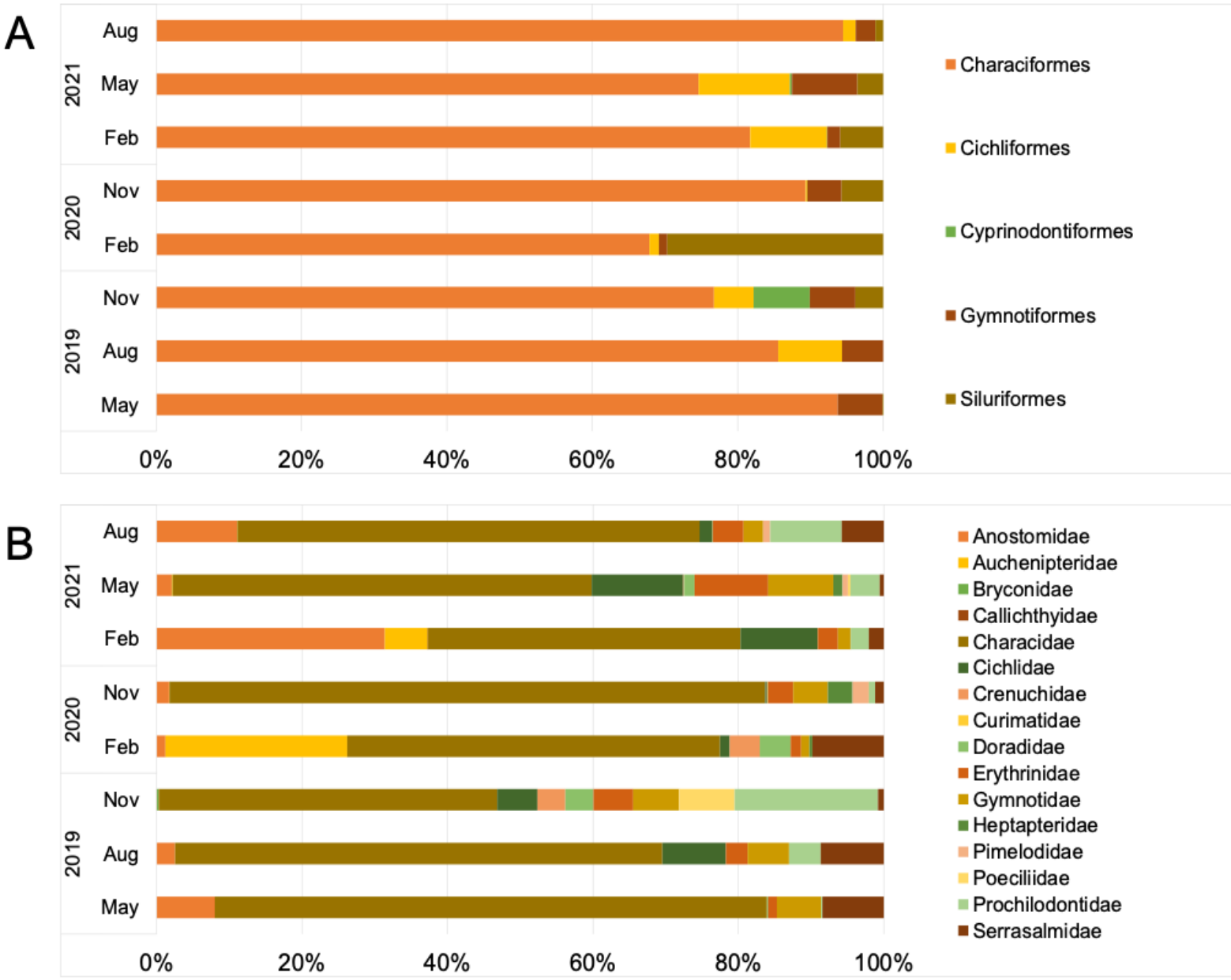


Figura 3. Stack barplot representing the number of sequences reads relative to the frequency of A) orders and B) families by eDNA samples at the three sampling locations through the eight campaigns.

When considering species richness per order, we observed that, as expected, the Characiformes was the richest order with eight families representing 16 species (Table S4, Figure 4A), followed by Siluriformes, with five families corresponding to seven species (Table Sx, Figure 4B). The orders Cichliformes, Cyprinodontiformes and Gymnotidormes identified one single family each and four, two and one fish species, respectively (Table S4, Figure 4C, D).

Within Characiformes, the richest family was Characidae with four species identified: *Astyanax lacustris*, *Astyanax* sp., *Hasemania* sp. and *Moenkhausia costae*. *M. costae* was the most representative species within Characidae and represented from ~25% (Feb/2020) to ~74% (Nov/2020), followed by *A. lacustris* that was most identified in May/2019 (~27.6%) and less identified in May/2021 (~1.4%) (Table S4, Figure 4A). On the other hand, *Hasemania* sp. was the species that were less representative and represented from ~0.008% (Nov/2019) to ~0.015% (Nov/2020) (Table S4, Figure 4A).

Within Siluriformes, the richest family was Pimelodidae with three species identified: *Pimelodus maculatus*, *Pimelodus* sp. and *Pseudoplatystoma reticulatum*. *P. reticulatum* was the most frequent and representative species within Pimelodidae and appeared in three campaigns and represented from ~0.72% (May/2021) to ~1.44% (Nov/2020), followed by *P. maculatus* that was identified in Nov/2020 (~0.79%) and *Pimelodus* sp., recovered in Aug/2021 (~0.033%) (Table S4, Figure 4B).

Within Cichliformes, the most frequent species from Cichlidae was *Australoheros facetus*, being recorded from Aug/2019 to Aug/2021, followed by *Coptodon zilli*, identified in three campaigns (Feb/2020, Feb/2021 and May/2021) as *Geophagus brasiliensis*, identified in three campaigns (Aug/2019, Nov/2019 and Nov/2020) and *Oreochromis* sp. identified in Nov/2019 and Feb/2021 (Table S4, Figure 4C).

Within Cyprinodontiformes, *Poecilia reticulata* was identified more frequently, from Feb/2020 to May/2021, although its representativeness was lower (~0.0001% to ~0.34%) when compared to *Phalloceros* sp. (~7.7%) identified only in Nov/2019 (Table S4, Figure 4D).

Although only one species from Gymnotiformes was detected, *Gymnotus carapo* was recovered during all campaigns and its representativeness varied from ~1.15% (Feb/2020) to ~8.98% (May/2021) (Table S4).

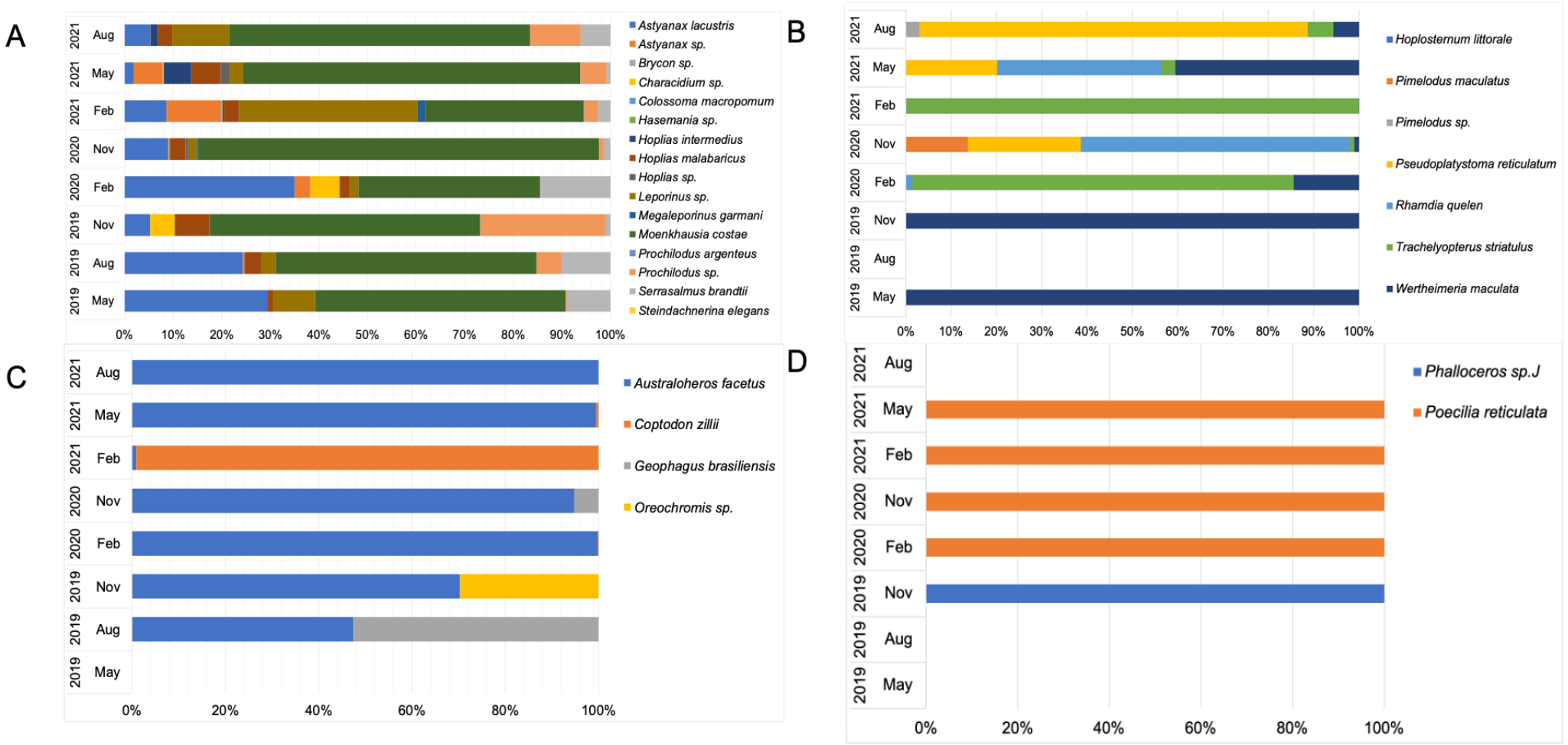
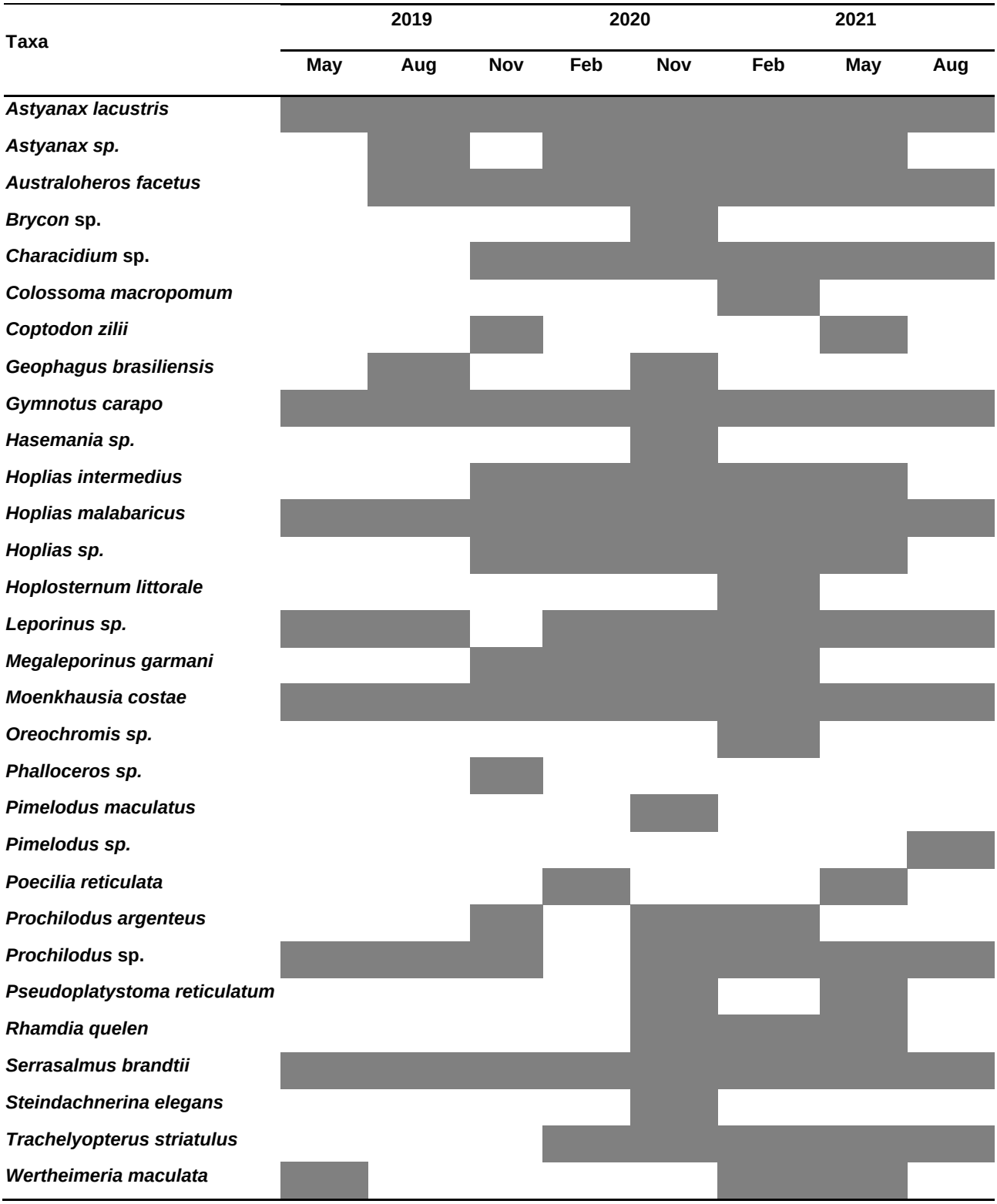


Figura 4. Stack barplot representing the number of sequences reads relative to the frequency of species from A) Characiformes, B) Serrasalmidae, C) Cyprinodontiformes, and D) Cichliformes by eDNA samples at the three sampling locations through the eight campaigns.

Considering that our reference database does not cover all species recovered for the Jequitinhonha River basin, some of the species detected might not correspond exactly to the species belonging to this basin as the taxonomic assignment is performed based on the nearest neighbour species (e.g., *Trachelyopterus striatulus* may refer to a closely related species of the same genus, as *Trachelyopterus galeatus*). In these cases, we referenced those species only up to the genus level. Moreover, species that were not previously reported for this basin are marked with an asterisk.

Comparing the data obtained for all sampling periods (Table 2), five species were detected: *Astyanax lacustris*, *Gymnotus carapo*, *Hoplias malabaricus*, *Moenkhausia costae*, *Serrasalmus brandtii*, whilst *Brycon* sp., *Colossoma macropomum**, *Hasemania* sp., *Hoplosternum littorale*, *Oreochromis* sp.*, *Phalloceros* sp., *Pimelodus maculatus**, *Pimelodus* sp.*, and *Steindachnerina elegans* were detected only in one out of the eight sampling acts (Table 2). Several taxa (e.g., *Astyanax* sp., *Australoheros facetus*, *Characidium* sp., *Leporinus* sp.) were detected in most of the sampling campaigns, and therefore seem to have a broad seasonal distribution in the IRR.

Table 2. Species distribution in the IRR by eDNA, according to each campaign conducted during three years.



6.4.3 Biodiversity patterns (Species richness and β-diversity)

During the first campaign (May/2019), according to the data obtained, the site upstream the dam (IR2) harboured the highest MOTU richness (n=14), followed by locations IR3 (n=6) and IR10 (n=4, Figure 5). The lowest number of MOTUs (n=1) was recovered for IR10 and IR2 from the second and fourth campaigns (Aug/19 and Feb/20, respectively). Regarding the second campaign, the IR3 site recovered the highest number of MOTUs (n=15), followed by IR2 (n=6) and IR10 (n=1). The third campaign was characterised for detecting a great number of MOTUs which were associated with a lower number of species, with IR2 being the richest site (n=31), followed by IR10 (n=30) and IR3 (n=24). During February 2020, we observed a huge discrepancy between the number of recovered MOTUs between sampling sites, with IR10 being the richest site (n=43) and IR2 recovering the lowest number of MOTUs (n=1, Figure 5). Considering the fifth campaign, IR2 recovered the highest MOTUs richness (n=33), followed by IR3 (n=31) and IR10 (n=28). We observed that IR3 hosted the major number of MOTUs in February 2021 (n=52), followed by IR2 (n=33) and IR10 (n=7). In May 2021, we observed a balance between the MOTUs recovered in each of the three sites, with 14 MOTUs recovered in the IR2 and 15 MOTUs for IR10 and IR3. Finally, during the last campaign, IR3 harboured the highest richness (n=13), followed by IR10 (n=6) and IR2 (n=2) (Figure 5).

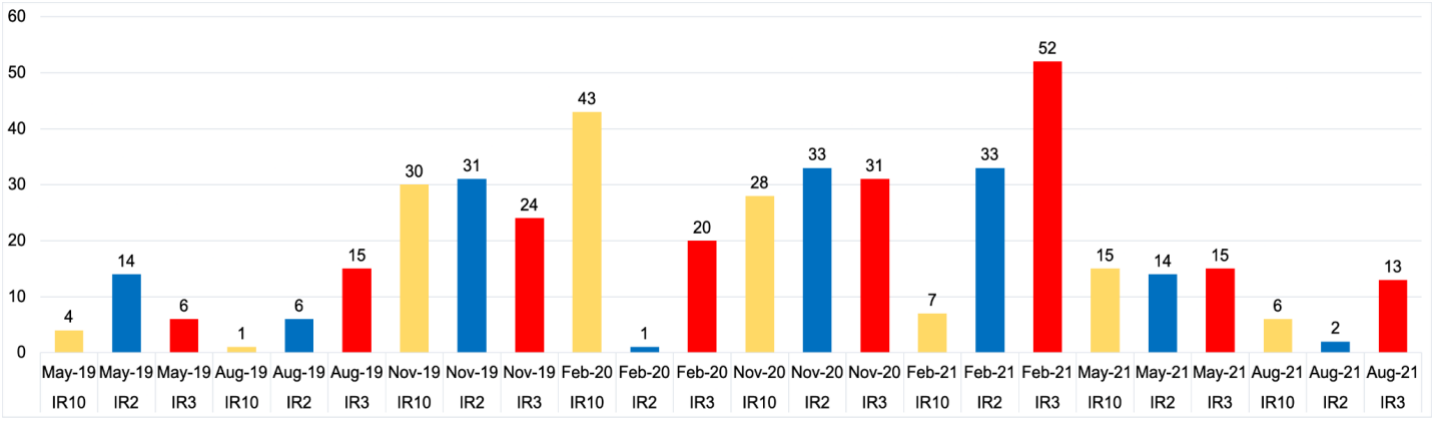


Figure 5. MOTU richness (α-diversity) distributed along the sampling sites (IR10 in yellow, IR2 in blue and IR3 in red) and sampling events.

Comparing the distribution of the MOTUs richness for each of the sampling sites, we observe that there is a greater range for IR2 (max=43, min=1) and IR10 (max=33, min=1), compared to IR3 (max=52, min=6). IR10 recovered the lowest median (μ=11), followed by IR2 (μ=14) and IR3 (μ=20) (Figure 6). Overall, the IR3 site showed the smallest range and recovered the highest mean of number of MOTUs ($\bar{X}$=24) followed by IR10 ($\bar{X}$=18.2) and IR2 ($\bar{X}$=17.1) (Figure 6). Moreover, as demonstrated by the PCoA (Figure 6), the highest dissimilarities observed between sampling sites were obtained for the first (May/2019) and second campaigns (Aug/2019). On the other hand, higher similarities were recovered across sites for 2020 and 2021 (Figure 6).

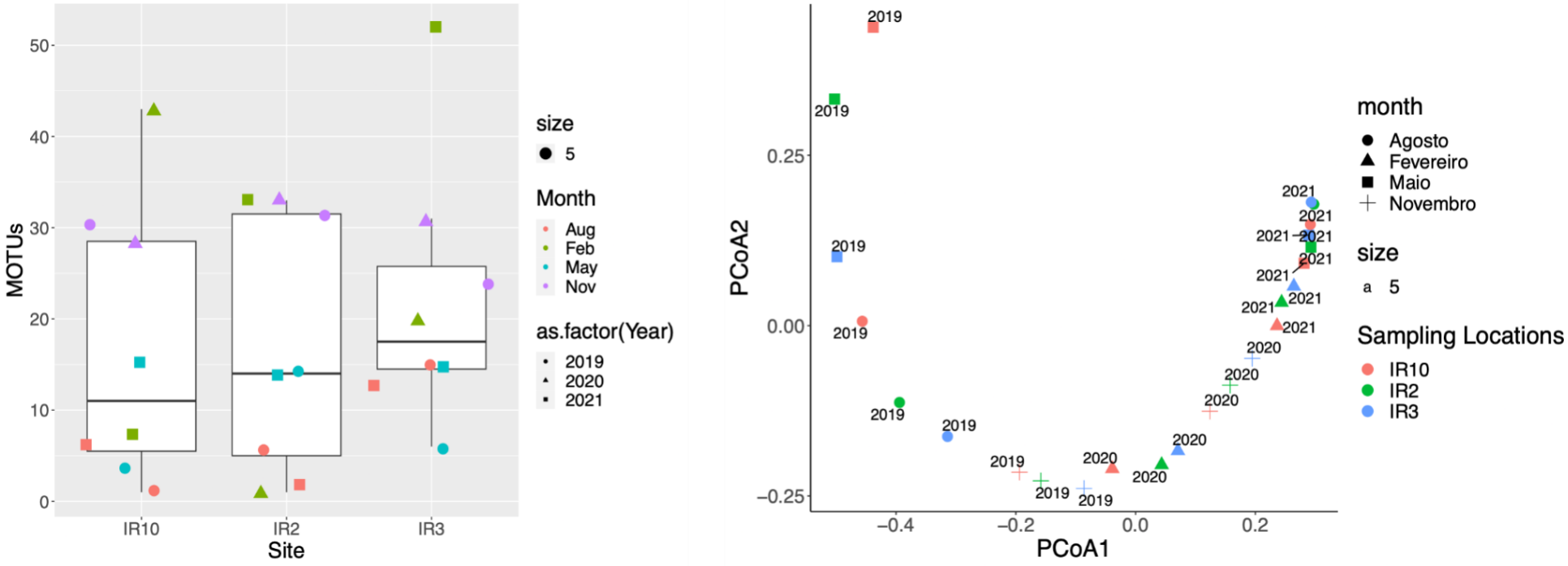


Figure 6. MOTU richness (α-diversity) and Principal Coordinates Analysis (PCoA) of β-diversity of sampling locations (Jaccard distance).

All results of the PERMANOVA analyses (Jaccard, p<0.05), including effect size (R2) and significance (p-value) are summarised in Table S5. The influence of the season on species diversity recovery was small and non-significant between samples collected (dry *vs* rainy, R2 = 4.5%, p=0.377), with a great overlap between “dry” and “rainy” fish communities detected (Figure 7). On the other hand, a significant effect was detected when considering “year” as variable (2019 *vs* 2020 *vs* 2021) with a greater dispersion observed in 2020 (Figure 7), although a small influence on species diversity was detected (R2 = 0.071, p=0.013, Table S5).

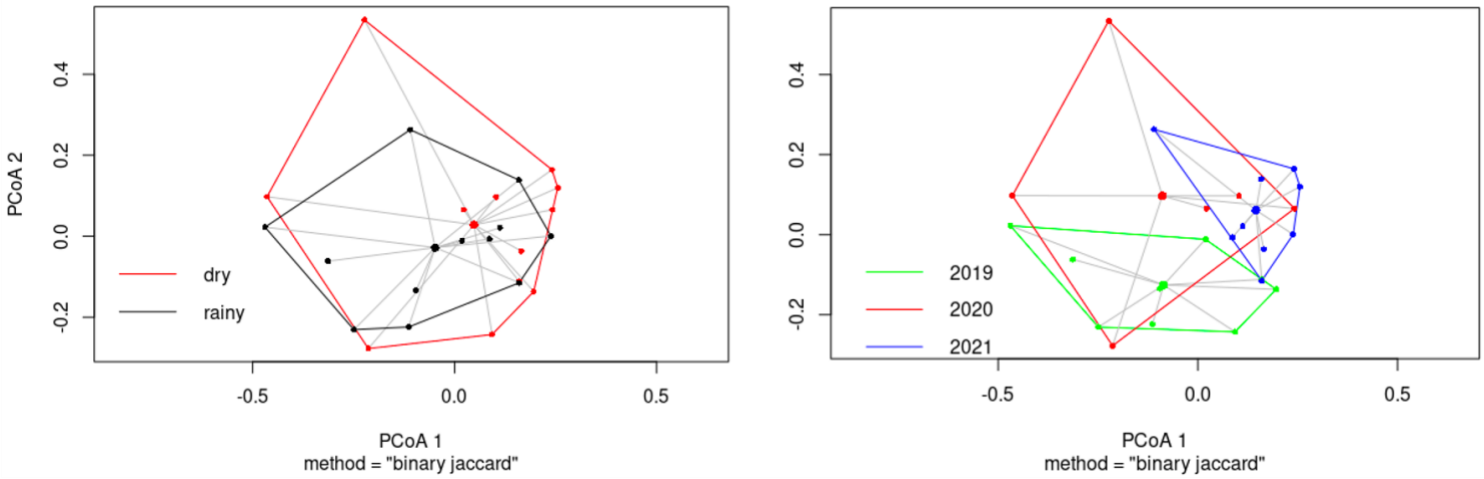


Figure 7. Principal Coordinates Analysis (PCoA) of seasons and years (Jaccard distance).

The influence of “month” (February *vs* May *vs* August *vs* November) on species richness recovery was higher but non-significant when compared to the other variables (R2=15.4%; p=0.078, Figure 8). Similar results were detected when considering the sampling sites as a variable (IR10 *vs* IR2 *vs* IR3), as non-significant results were recovered from this analysis (R2 = 7.7, p=0.761, Table Sx). PCoAs also agreed with these results, as we observed a great overlap between species richness recovered when “month” and “year” were tested (Figure 8).

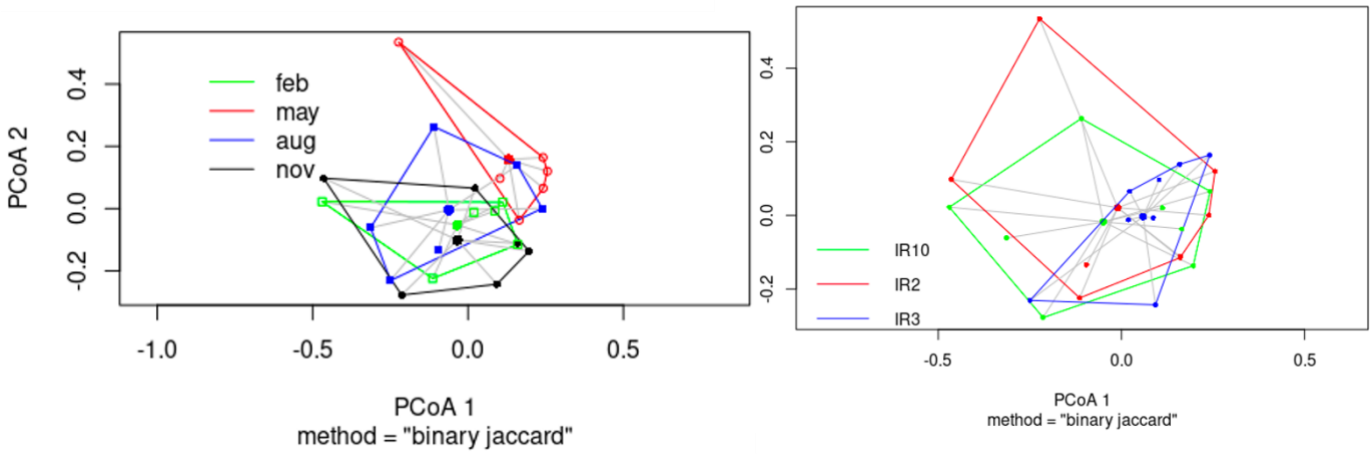


Figure 8. Principal Coordinates Analysis (PCoA) of months and sampling sites (Jaccard distance).

6.4.4 Comparison with established fishing methods

Morphological identification of fish captured by surface gillnets allowed the identification of a total of 13 species (Figure 9). Accumulation curve for gillnets showed that this method reached a plateau after sampling the three sampling sites during all campaigns (Figure 9). By considering the taxonomic assignment conditions, a large diversity of fish was observed during the whole survey in terms of identified orders (n=5), families (n=16), genera (n=25) and species (n=36) for both techniques taken together (Table S6). Although the eDNA metabarcoding was able to identify more species (n=30) than the gillnets (at least five orders, 16 families and 23 species more), this technique did not reach a plateau, as the gillnets, thus indicating that sampling effort was not sufficient to detect all fish present in the community. When considering both methods together, the species accumulation curve was similar to the eDNA metabarcoding curve, except by the number of species detected (n=36) (Figure 9).

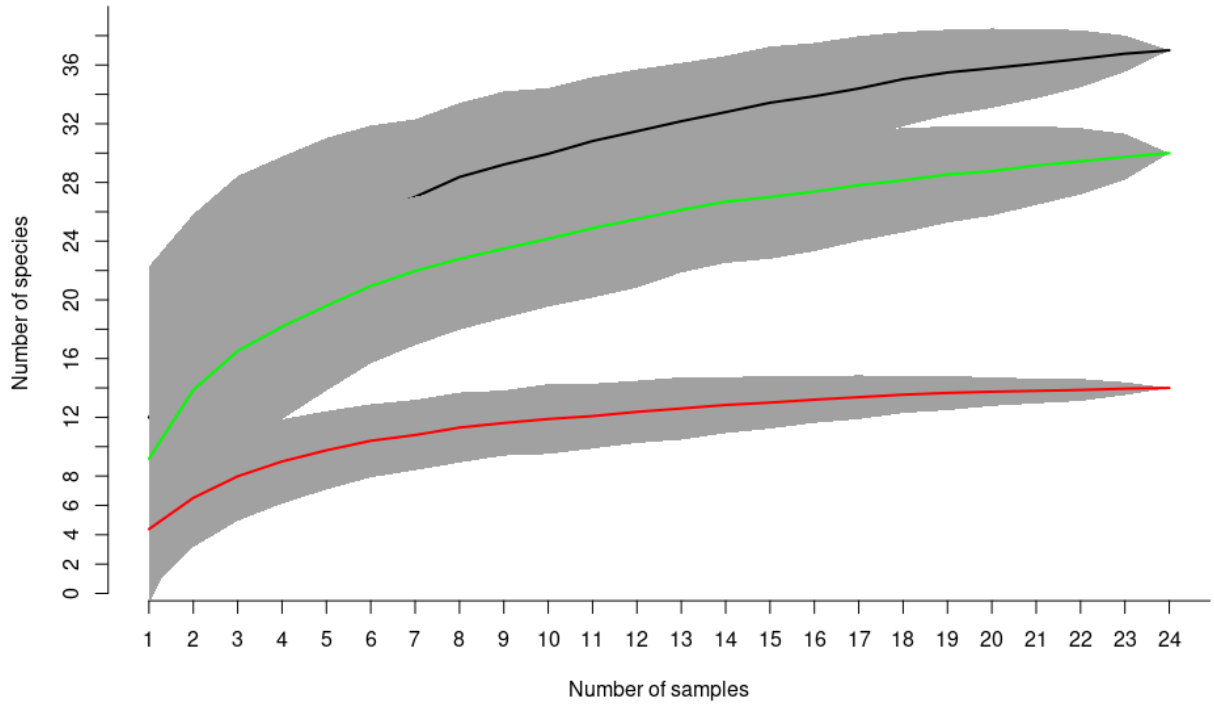


Figure 9. Species accumulation curves for gillnets (red line), eDNA metabarcoding (green line) and both methods together (black line). Grey shading represents the standard deviation of the expected number of species per sampling method.

Regarding α and β-diversities performed based on species richness for gillnets, we observed a more visible pattern considering that most of the sampling units were grouped together (Figure S1). The most dissimilar ones, that is, the sampling units that presented the most different compositions and abundances were characterised by IR3 according to the PCoA (Figure S1). Still, the richest sampling site across all sites was the major Jequitinhonha's tributary, the IR3, which recovered the highest mean of species richness ($\bar{X}$=4.8), followed by IR2 ($\bar{X}$=4) and IR10 ($\bar{X}$=3) (Figure 10A).

On the other hand, when considering eDNA metabarcoding β-diversity was dissimilar between all sampling units and no pattern was detected (Figure S1). Comparing the range of species richness for each of the sampling sites, we observe the greater variation for IR3 (max=13, min=5). Overall, as shown by the species richness, the IR3 site recovered the highest mean number of species ($\bar{X}$=8.1) followed by IR10 ($\bar{X}$=5.5) and IR2 ($\bar{X}$=5) (Figure x10B, Table S8).

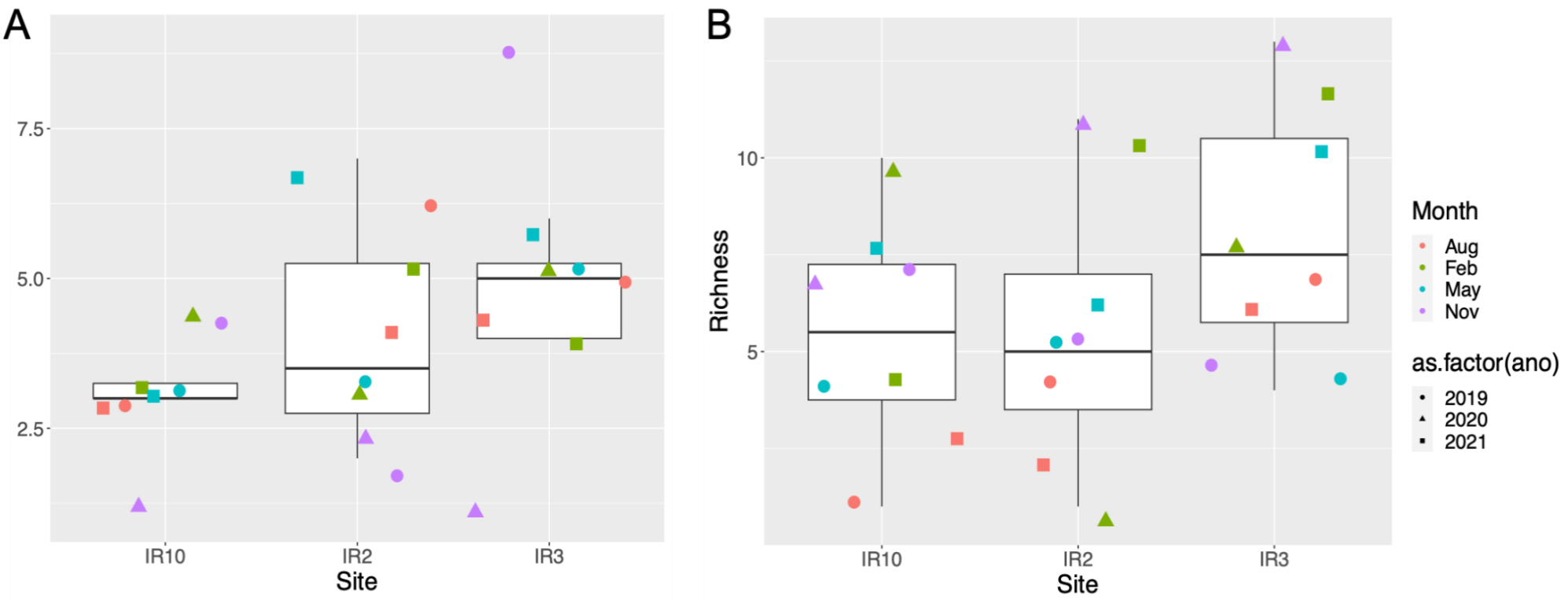


Figure 10. Taxa and species richness along the IRR for eDNA metabarcoding (A) and gill nets (B).

The Venn diagram showed that the overlaps of species detected by eDNA and gills nets represents 18.9% of the fish biodiversity (Figure 11). While the eDNA metabarcoding recovered 64.9% of the fish diversity detected, the gillnets were able to identify 19.3% of the fish species. Reinforcing the results from the Venn diagram, the PCoA based on β-diversity comparing both methods showed a clear and significance difference between the fish community detected by each monitoring method, with no overlap between species richness recovered by both methods (R2 = 14.0%, p = 0.001; Figure 11).

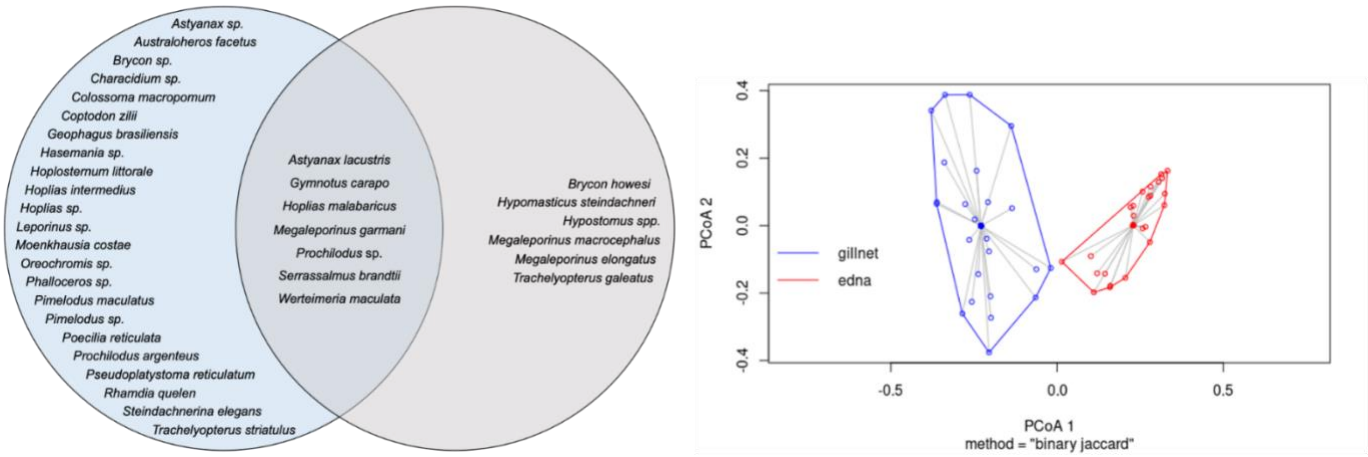


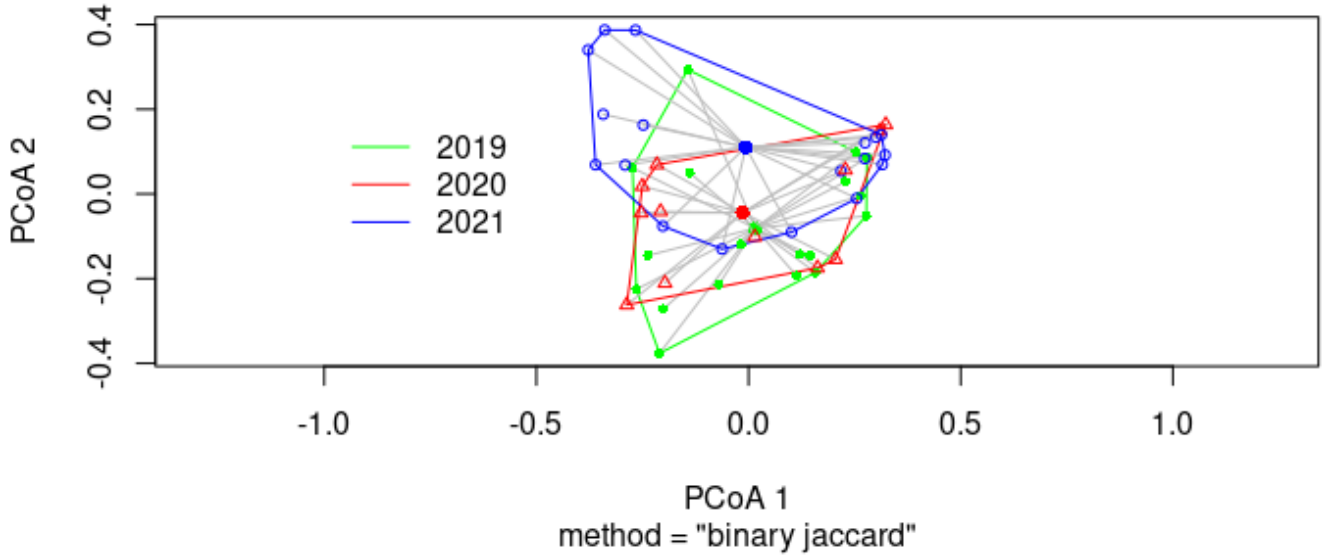
Figure 11. Venn diagram (blue for eDNA and grey for gillnets) and PCoAs of methods (Jaccard distance).

The relative number of sequencing read counts (RRA) for fish species poorly correlate with gillnet counts, when considering RRA for each species sampled in each site in each campaign and also when we summed all RRAs for each campaign and species (S = 242002, p = 0.8339).

Considering the gillnets and eDNA metabarcoding results together, the influence of “season” on species diversity recovery was small (around 1.62% variance explained), and non-significant between samples collected (dry *vs* rainy, p=0.868, Table S5). The influence of “month” on species diversity recovery was higher when compared with the other variables (around 15.4% variance explained), however, non-significant between samples collected (February *vs* May *vs* August *vs* November, p=0.078, Table S5). On the other hand, when considering the sampling site as a variable (IR10 *vs* IR2 *vs* IR3), a lower and non-significant influence on species diversity was detected (R2 = 0.038, p=0.695, Table S5). On the other hand, similar to the eDNA metabarcoding results analysed solely, a significant effect was detected when considering “year” as variable, although a small influence on species diversity was

detected (R2 = 0.037, p=0.012, Table S5), with a great overlap between years sampled (Figure 12).

Figure 12. Principal Coordinates Analysis (PCoA) of years (Jaccard distance) considering both methods.



6.5 DISCUSSION

Taxa richness is considered the elementary metric of community structure, then, most eDNA-based studies focus on its description. Many studies have been conducted in order to compare eDNA metabarcoding and established fishing methods results in numerous aquatic habitats (Czeglédi et al., 2021; Shu et al., 2022; Li et al., 2022). Most of these studies report that eDNA-based surveys perform well in detecting taxa, and, in many cases outperform traditional survey methods (e.g., Bylemans et al., 2017; Goutte et al., 2020; Zou et al., 2020), especially, when we have the ideal conditions, such as complete reference database, and appropriate markers. In this study, we evaluated the performance of metabarcoding to describe freshwater fish communities through the comparison of eDNA data (number of sequence reads/species) to that of gillnet captures in a Neotropical hydroelectric impoundment environment. We found that eDNA metabarcoding allowed the detection of several species at most of sampling locations where they were missed by the gillnet survey. On the other hand, the reverse also happened but to a lesser degree. Interestingly, eDNA metabarcoding and gillnet captures showed a significant difference in fish community assemblage, demonstrating the complementary role of eDNA metabarcoding in the monitoring of fishes.

6.5.1 Taxonomic assignment by eDNA metabarcoding

As expected, compared to species richness (n=30, Figure 2), MOTU richness overestimated the diversity (n=115, Figure 2). Although MOTUs are frequently employed as a proxy for species, the relationship between these two types of diversity classification is not precise (Sales et al., 2021). Two factors may have a significant impact on MOTU richness, the presence of cryptic species and the thresholds used in bioinformatic analyses (Pawlowski et al., 2018), leading to its overestimation. For instance, Hilário et al. (2023) reported an overestimated MOTU richness, with values higher than the original number of fish species even when using mock communities. Moreover, the inflated number of MOTUs for a same species might evidence the natural intraspecific variability (García-García et al., 2019; Sales et al., 2021). Here, we observed that the most common species depicted by eDNA metabarcoding, *Moenkhausia costae*, presented a total of 15 MOTUs (Table S7). High haplotype diversity was also reported for a congener species, *Moenkhausia bonita*, collected in the upper Paraná River floodplain (Mota et al., 2018). Here, we also stress that although taxonomy free metabarcoding has been advocated in the exploration of very

diverse and poorly characterised species assembly (Mächler et al., 2021) to enable ecological analyses of the megadiverse Neotropical fish fauna, it should be employed with caution (Hilário et al., 2023). On the other hand, the underestimation of richness, when using MOTU diversity, could be caused by the incompleteness of the reference database. For instance, fish diversity depicted by the analyses shows that a hidden diversity might be present, especially for the order Characiformes, as many families appear to comprise several MOTUs (Table S3).

Moreover, 11 of the 30 species (~36%) identified by eDNA metabarcoding could not be detected to species level, despite using a large reference database (486 sequences from 294 species). Moreover, due to incomplete inventories and the requirement for taxonomic revision for a number of groups, estimates of fish diversity in Brazilian inland waters are currently tentative. Even in the highly researched hydrographic basins, such as the upper Paraná River basin, new species are described every year (Agostinho et al., 2008). Like many other catchments in the Neotropical region, Jequitinhonha has not been thoroughly explored, and we still do not fully know its fish biodiversity. It is reported that reference databases not only encompass a low number of species included but most of these taxa are represented by just one sequenced specimen, hampering taxonomic identifications (Jackman et al., 2021). Another major impediment to the expansion of the reference data is the prevalence of efforts toward the construction of databases targeting very specific and short fragments, and more importantly, the failure to make these sequences fully available in public repositories (Marques et al., 2021). This study emphasizes that the expansion of eDNA studies, particularly in megadiverse areas, is heavily reliant on improved collaborative efforts aimed at creating more publicly available datasets.

As reported by previous studies, molecular identification through eDNA metabarcoding might be challenging due to numerous factors, beyond cryptic diversity and the incompleteness of the reference database, such as the lack of resolution of phylogenetic markers (Yu et al. 2012; Carew et al. 2013; Eiler et al. 2013). According to genetic-distance-threshold analyses, an ideal threshold value of ~0.5% was reported by Milan et al. (2020) and Jackman et al. (2021) when evaluating the taxonomic resolution of the rRNA 12S gene. The 12S MiFish primers have obviously been used effectively in many research and are legitimately still in widespread use (Miya et al., 2020), however it should be emphasised that it has been shown that in particular systems, these primers may not be able to reliably identify closely related

species (Sales et al., 2021). As we were conservative in delimiting species using the 97% threshold, the number of species detected is surely underestimated and many congeneric species might have been clustered together as one single species due to the low taxonomic resolution of the fragment analysed. Here, most of the MOTUs of *Prochilodus* sp. could not be taxonomically assigned to species level due to the low taxonomic resolution of the 12S fragment used here. Despite the identification of *Prochilodus argenteus,* it represents an invasive species for this basin and a low genetic divergence from the endemic species *P. hartii* is already known for this genus and these species (Melo et al., 2018; Santos et al., 2021; Figure S2). In this case, other markers could be evaluated in the future to overcome these limitations and further improve these species resolution issues. Regarding *Hoplias*, as its species tend to be morphologically similar, taxonomic confusion is known to cause misidentifications and erroneous identifications might also be present in the reference databases. For instance, Guimarães *et al.* (2022) using DNA barcoding analyses also detected misidentifications within *H. brasiliensis* from the Jequitinhonha River, which is more likely to be *H. intermedius*. Therefore, here, we cannot affirm that there is a lack of taxonomic resolution from the 12S gene and it is recommended to sequence more specimens of *H. brasiliensis* to better understand the taxonomic resolution within *Hoplias* species.

Moreover, eDNA metabarcoding allowed the identification of species not expected to occur in the target area, which makes it an important tool for tracking species invasions, known as one of the most damaging anthropogenic activities (Deiner et al., 2017). Almost 30% of the taxa detected by eDNA are represented by non-indigenous species, including commercially important species, such as tilapia species (*Oreochromis* sp. and *Coptodon zilli*), which is already introduced for fishery purposes in several Brazilian basins and it is known by its invasive potential. Although, *Coptodon zillii* is frequently reported in North America (Cassemiro et al., 2018), closely related species have been reported to the Jequitinhonha River basin (*Oreochromis niloticus*, *Tilapia* sp.) and neighbouring drainages (*Tilapia rendalli* – Doce river basin). Regarding the native species, eDNA allowed the detection of endemic species, such as *Geophagus brasiliensis*, *Gymnotus carapo* and *Wertheimeria maculata*. Furthermore, a notable result obtained by eDNA included the detection of species rarely reported in traditional sampling studies (e.g., *Moenkhausia costae*; Andrade Neto, 2010) in all sites analysed (Table 2) and, suggested the occurrence of species

that was recorded more than a decade ago (Godinho 2007), that is the case of the surubim – *Pseudoplatystoma reticulatum* (Table 2 and Figure 11).

6.5.2 Comparison between eDNA metabarcoding and gill nets

eDNA metabarcoding is reported to outperform traditional sampling in deep and large lakes where traditional fishing gears fail to represent community structure properly (Civade et al., 2016; Hänfling et al., 2016; Lawson Handley et al., 2019). Our results showed that the number of fish species revealed by eDNA metabarcoding surpassed that revealed by gillnets, confirming earlier findings. Moreover, species accumulation curves recovered by gill nets showed that this method reached a plateau, that is, even with more sampling effort the diversity detected by this technique was totally sampled (Figure 9). However, it does not mean that there is no more diversity to be sampled as the eDNA metabarcoding curve showed a high diversity to be detected with more sampling campaigns (Figure 9). For instance, the species *Australoheros facetus* and *Moenkhausia costae* were detected through eDNA, but not by netting, conducted at the same time. These species have been previously identified in the studied area (Sales et al., 2021) and this highlights the randomness of net sampling, which could be due to the limited spatial and temporal scales covered by the netting survey, or, in this case, even by species biology and behaviour. *A. facetus* and *M. costae* are species of small size; hence, they can easily elude netting. These results highlight that the use of multiple types of sampling gear ensures that distinct groups of fishes are sample due to the inherent bias of all established fishing methods (Kubečka et al. 2009).

Nevertheless, despite the large number of species detected by eDNA metabarcoding, we observed some overlap between taxa sets identified by the eDNA and traditional methods. Both eDNA metabarcoding and the traditional methods detected taxa that were unique to one of the sampling methods (Figure 11, Table S6). It was noteworthy that eDNA metabarcoding did not cover all fish species detected by conventional fishing methods, such as *Brycon howesi* and *Hypomasticus steindachneri*. There are several potential explanations for the failure of eDNA metabarcoding to detect certain fish species, such as the performance of the primers used (Takeuchi et al. 2019), quality and completeness of the reference database (Callahan et al. 2017), and bias of PCR against low-abundance sequences (Ficetola et al. 2015, Hänfling et al. 2016, Li et al. 2018). For instance, in the case of *B. howesi*,

considered rare and threatened with extinction, is not represented by a reference sequence, which could be the cause of the non-identification of this species. However, it is important to highlight that we were able to identify its congener, *Brycon* sp. (Figure 11, Table S6).

Overall, the number of sequence reads and fish biomass/catch in the IRR reservoir were not correlated, suggesting that read count data are unreliable to estimate abundance based on net catches, at least in this system. Although the relationships between sequence abundances and fish population occurrences remain obscured and might be confounded by various biological factors and local environmental settings, such as temperature (Jo et al. 2019), pH, and UV-B radiation (Strickler et al. 2015), a positive significant correlation between eDNA data and observed species has been revealed in several aquatic ecosystems (Hänfling et al., 2016; Lacoursière-Roussel et al., 2016; Thomsen et al., 2016). However, many of these studies obtained accurate population estimates using standardised sampling methods and/or long-term data. We hypothesised that the lack of correlation could be due to the difference between the fish communities recovered by both methods. Generally, due to the relative stability of the habitat shape (Zhang et al. 2020) and the mixture of water, observed in small lakes, the fish assemblages in these environments tend to be more homogenous (Evans et al. 2017, Thomsen et al. 2012). On the other hand, fish communities in large lentic systems, such as, dam reservoirs, are more likely to be heterogeneous (Hänfling et al. 2016), probably caused by the characteristics related to hydrodynamics and fish ecology (Zhang et al. 2020). In our study, the spatial distribution of fish communities in the IRR revealed by eDNA showed a heterogeneous pattern that agreed with the fish distribution pattern of large lentic systems described above. The IRR is a large artificial lake and its elongated drainage basin indicates that the elongated nature of the lake system created more variable habitat features and formed many microenvironments for distinct fish groups.

We also observed contrasting patterns recovered by β-diversity analysis for both eDNA metabarcoding and traditional methods. While gill net data showed similarities in compositions with slight clustering of sites IR10 and IR2, eDNA data suggest major differences between each sampling location and campaigns (Figure S1). These results can be explained by the selective nature of the traditional fishing methods, and characteristics, such as the species size and behaviour might affect the

susceptibility of species capture, creating a bias toward species that are easier to capture, producing species-selective results (Holubová et al., 2019).

As for the eDNA metabarcoding results solely (p=0.377), no significant effect of sampling season was observed for both methods together. Here, the comparison between the dry and rainy seasons showed a non-significant influence of seasonality on community composition (p=0.868). These results might suggest that freshwater fish assemblages in dam reservoirs may not vary significantly between dry and wet seasons. These results are different from previous studies that have demonstrated compositional changes in accordance with seasonal variations (Sigsgaard et al., 2017; Hayami et al., 2020), including a study conducted in the same river basin (Sales et al., 2021). However, most of these studies were conducted in lotic habitats, such as rivers and streams. An important factor to take into consideration is that our study was conducted in a dam reservoir. The loss of native fish and an increase in nonnative fish from damming a river can result in biotic homogeneity due to a loss of fish variety (Olden et al., 2006; Petesse and Petrere, 2012; Pelicice and Agostinho, 2008; Pompeu et al., 2012), since dams have a significant negative impact on the environment (i.e., altering the physical and ecological characteristics of the habitats, such as changes in water flow, nutrient dynamics, water quality, and temperature). Our study encompasses an area that is highly influenced by the IRR and our results evidence a lower species richness (n=30) in comparison to the actual number of species estimated for Jequitinhonha River basin (~110 species). T hus our data might be related to the homogenization of species communities expected to be found in this type of environment. Moreover, the ecology of DNA might play an important role regarding this matter, as eDNA molecules could be more diluted in the water column decreasing the detectability of some species (e.g., rare or less abundant species). Then, future studies should aim to evaluate important factors related to the ecology of eDNA in dam reservoirs.

In summary, our study demonstrates that the performance of eDNA metabarcoding can be complementary to that of traditional survey methods for fish species surveillance in reservoirs. Due to the inability of eDNA metabarcoding to record all fish species revealed by established fishing methods, it cannot completely replace traditional capture-based methods at present. However, it will be an important tool for monitoring for future dam projects, both pre- and post-construction. Finally, despite providing important initial information regarding species occurrence and

distribution, the data provided here should be used with caution. The effectiveness of this innovative method in recovering eDNA of rare and elusive species from the environment is widely recognized, however, some drawbacks might be discussed before drawing final conclusions, such as the ecology of eDNA including the understanding of the origin, persistence, transport and degradation of eDNA within these systems (Barnes and Turner, 2016).

REFERENCES

Agostinho, A. A., Pelicice, F. M., & Gomes, L. C. (2008). Dams and the fish fauna of the Neotropical region: impacts and management related to diversity and fisheries. *Brazilian journal of biology*, *68*, 1119-1132.

Andrade Neto, F. R. (2010). Estado atual do conhecimento sobre a fauna de peixes da bacia do Jequitinhonha. *INSTITUTO ESTADUAL DE FLORESTAS—MG*, 23.

Barnes, M. A., & Turner, C. R. (2016). The ecology of environmental DNA and implications for conservation genetics. *Conservation genetics*, *17*(1), 1-17.

BONAR, A.; HUBERT, A.; WILLIS, W. Standard methods for sampling North American freshwater fishes. 2009.

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.

BRAUER, C. J., HAMMER, M. P., & BEHEREGARAY, L. B. Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. Mol Ecol 25, 5093-5113 (2016).

BURKHEAD, N. M. Extinction rates in North American freshwater fishes, 1900–2010. BioScience 62, 798-808 (2012).

Bylemans, J., Furlan, E. M., Pearce, L., Daly, T., & Gleeson, D. M. (2016). Improving the containment of a freshwater invader using environmental DNA (eDNA) based monitoring. *Biological Invasions*, *18*, 3081-3089.

Cachapuz, P. B. D. B. (2006). Usinas da Cemig: a história da eletricidade em Minas e no Brasil, 1952-2005. *Centro da Memória da Eletricidade no Brasil: Rio de Janeiro, Brazil*.

Callahan, T. J., Bauck, A. E., Bertoch, D., Brown, J., Khare, R., Ryan, P. B., ... & Kahn, M. G. (2017). A comparison of data quality assessment checks in six data sharing networks. *EGEMs*, *5*(1).

Carew, M. E., Pettigrove, V. J., Metzeling, L., & Hoffmann, A. A. (2013). Environmental monitoring using next generation sequencing: rapid identification of macroinvertebrate bioindicator species. *Frontiers in zoology*, *10*, 1-15.

Cassemiro, F. A. S., Bailly, D., da Graça, W. J., and Agostinho, A. A. (2018). The invasive potential of tilapias (Osteichthyes, Cichlidae) in the Americas. Hydrobiologia, 817(1), 133– 154. doi: 10.1007/s10750-017-3471-1.

CILLEROS, K., VALENTINI, A., ALLARD, L., DEJEAN, T., ETIENNE, R., GRENOUILLET, G., ... & 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, v. 19, n. 1, p. 27-46.

Civade, R., Dejean, T., Valentini, A., Roset, N., Raymond, J. C., Bonin, A., ... & Pont, D. (2016). Spatial representativeness of environmental DNA metabarcoding signal for fish biodiversity assessment in a natural freshwater system. *PloS one*, *11*(6), e0157366.

Czeglédi, I., Sály, P., Specziár, A., Preiszner, B., Szalóky, Z., Maroda, Á., ... & Erős, T. (2021). Congruency between two traditional and eDNA-based sampling methods in characterising taxonomic and trait-based structure of fish communities and community-environment relationships in lentic environment. *Ecological Indicators*, *129*, 107952.

DAL PONT, G., DUARTE RITTER, C., AGOSTINIS, A. O., STICA, P. V., HORODESKY, A., COZER, N., ... & PIE, M. R. Monitoring fish communities through environmental DNA metabarcoding in the fish pass system of the second largest hydropower plant in the world. Scientific Reports, v. 11, n. 1, p. 23167, 2021.

DEINER, K., BIK, H. M., MÄCHLER, E., SEYMOUR, M., LACOURSIÈRE-ROUSSEL, A., ALTERMATT, F., ... & BERNATCHEZ, L. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Mol ecol 26, 5872-5895, 2017.

DOI, H., INUI, R., AKAMATSU, Y., KANNO, K., YAMANAKA, H., TAKAHARA, T., & MINAMOTO, T. Environmental DNA analysis for estimating the abundance and biomass of stream fish. Freshwater Biology, v. 62, n. 1, p. 30–39, jan. 2017.

Eiler, A., Drakare, S., Bertilsson, S., Pernthaler, J., Peura, S., Rofner, C., ... & Lindström, E. S. (2013). Unveiling distribution patterns of freshwater phytoplankton by a next generation sequencing based approach. *PLoS One*, *8*(1), e53516.

Erickson, R. A., Rees, C. B., Coulter, A. A., Merkes, C. M., McCalla, S. G., Touzinsky, K. F., ... & Amberg, J. J. (2016). Detecting the movement and spawning activity of bigheaded carps with environmental DNA. *Molecular Ecology Resources*, *16*(4), 957-965.

Evans, N. T., Li, Y., Renshaw, M. A., Olds, B. P., Deiner, K., Turner, C. R., ... & Pfrender, M. E. (2017). Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. *Canadian Journal of Fisheries and Aquatic Sciences*, *74*(9), 1362-1374.

Ferreira, V. O., & Silva, M. M. (2012). O clima da bacia do rio Jequitinhonha, em Minas Gerais: subsídios para a gestão de recursos hídricos. *Revista Brasileira de Geografia Física*, *5*(2), 302-319.
Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology letters*, *4*(4), 423-425.

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., ... & Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular ecology resources*, *15*(3), 543-556.

García-García, N., Tamames, J., Linz, A. M., Pedrós-Alió, C., & Puente-Sánchez, F. (2019). Microdiversity ensures the maintenance of functional microbial communities under changing environmental conditions. *The ISME journal*, *13*(12), 2969-2983.

Godinho, A. L., & Loures, R. C. (2017). Risk of fish death at Brazilian hydropower plants. *Risk Assessment of Fish Death at Hydropower Plants in Southeastern Brazil'.(Eds RC Loures and AL Godinho.) Vol*, *1*, 19-36.

Goldberg, C. S., Sepulveda, A., Ray, A., Baumgardt, J., & Waits, L. P. (2013). Environmental DNA as a new method for early detection of New Zealand mudsnails (Potamopyrgus antipodarum). *Freshwater Science*, *32*(3), 792-800.

GOLDBERG, C. S., TURNER, C. R., DEINER, K., KLYMUS, K. E., THOMSEN, P. F., MURPHY, M. A., ... & TABERLET, P. Critical considerations for the application of environmental DNA methods to detect aquatic species. Methods in Ecology and Evolution, v. 7, n. 11, p. 1299–1307, 22 nov. 2016.

Gonçalves, R. D. N. (1997). Diagnóstico ambiental da Bacia do Rio Jequitinhonha: Diretrizes gerais para a ordenação territorial. *Salvador: IBGE*.

Goutte, A., Molbert, N., Guérin, S., Richoux, R., & Rocher, V. (2020). Monitoring freshwater fish communities in large rivers using environmental DNA metabarcoding and a long-term electrofishing survey. *Journal of Fish Biology*, *97*(2), 444-452.

HAJIBABAEI, M., SHOKRALLA, S., ZHOU, X., SINGER, G. A., & BAIRD, D. J. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS one 6, e17497, 2011.

HÄNFLING, B., LAWSON HANDLEY, L., READ, D. S., HAHN, C., LI, J., NICHOLS, P., ... & WINFIELD, I. J. Environmental DNA metabarcoding of lake fish communities reflects long term data from established survey methods. Mol ecol 25, 3101-3119, 2016.

HAYAMI, K., SAKATA, M. K., INAGAWA, T., OKITSU, J., KATANO, I., DOI, H., ... & MINAMOTO, T. Effects of sampling seasons and locations on fish environmental DNA metabarcoding in dam reservoirs. Ecology and Evolution, v. 10, n. 12, p. 5354–5367, 6 jun. 2020.

He, F. et al. Freshwater megafauna diversity: Patterns, status and threats. Divers. Distrib. 24, 1395– 1404 (2018).

HICKLEY, P.; STARKIE, A. Cost effective sampling of fish populations in large water bodies. **Journal of fish biology**, v. 27, p. 151-161, 1985.

Hilário, H., Mendes, I., Sales, N., & Carvalho, D. (2023). DNA metabarcoding of mock communities highlights potential biases when assessing Neotropical fish diversity.

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

JACKMAN, J. M., BENVENUTO, C., COSCIA, I., CARVALHO, C., READY, J. S., BOUBLI, J. P., ... & SALES, N. eDNA in a bottleneck: Obstacles to fish metabarcoding studies in megadiverse freshwater systems. Environmental DNA, v. 3, n. 4, p. 837–849, 2021.

JERDE, C. L., CHADDERTON, W. L., MAHON, A. R., RENSHAW, M. A., CORUSH, J., BUDNY, M. L., ... & LODGE, D. M. Detection of Asian carp DNA as part of a Great Lakes basin-wide surveillance program. Canadian Journal of Fisheries and Aquatic Sciences, v. 70, n. 4, p. 522–526, 2013.

Jerde, C. L., Mahon, A. R., Chadderton, W. L., & Lodge, D. M. (2011). “Sight-unseen” detection of rare aquatic species using environmental DNA. *Conservation letters*, *4*(2), 150-157.

JERDE, C. L.; WILSON, E. A.; DRESSLER, T. L. Measuring global fish species richness with <scp>eDNA</scp> metabarcoding. Molecular Ecology Resources, v. 19, n. 1, p. 19–22, 30 jan. 2019.

JEREP, F. C., Camelier, P., & Malabarba, L. R. (2016). Serrapinnus zanatae, a new species from the rio Jequitinhonha basin, Minas Gerais State, Brazil (Teleostei: Characidae: Cheirodontinae). *Ichthyological Exploration of Freshwaters*, *26*(4), 289-298.

JO, T., Murakami, H., Yamamoto, S., Masuda, R., & Minamoto, T. (2019). Effect of water temperature and fish biomass on environmental DNA shedding, degradation, and size distribution. *Ecology and evolution*, *9*(3), 1135-1146.

KECK, François et al. Meta-analysis shows both congruence and complementarity of DNA and eDNA metabarcoding to traditional methods for biological community assessment. Molecular Ecology, v. 31, n. 6, p. 1820-1835, 2022.

KELLY, R. P., PORT, J. A., YAMAHARA, K. M., & CROWDER, L. B. Using environmental DNA to census marine fishes in a large mesocosm. PloS one 9, e86175, 2014.

KUBEČKA, Jan et al. The true picture of a lake or reservoir fish stock: A review of needs and progress. Fisheries Research, v. 96, n. 1, p. 1-5, 2009.
LARAMIE, Matthew B.; PILLIOD, David S.; GOLDBERG, Caren S. Characterizing the distribution of an endangered salmonid using environmental DNA analysis. Biological Conservation, v. 183, p. 29-37, 2015.

Lawson Handley, L., Read, D. S., Winfield, I. J., Kimbell, H., Johnson, H., Li, J., ... & Hänfling, B. (2019). Temporal and spatial variation in distribution of fish environmental DNA in England’s largest lake. *Environmental DNA*, *1*(1), 26-39.

Li, H., Yang, F., Zhang, R., Liu, S., Yang, Z., Lin, L., & Ye, S. (2022). Environmental DNA metabarcoding of fish communities in a small hydropower dam reservoir: a comparison between the eDNA approach and established fishing methods. *Journal of Freshwater Ecology*, *37*(1), 341-362.

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

Liermann, C. R., Nilsson, C., Robertson, J. & Ng, R. Y. Implications of dam obstruction for global freshwater fish diversity. Bioscience 62, 539–548 (2012).

LIM, N. K., TAY, Y. C., SRIVATHSAN, A., TAN, J. W., KWIK, J. T., BALOĞLU, B., ... & YEO, D. C. Next-generation freshwater bioassessment: eDNA metabarcoding with a conserved metazoan primer reveals species-rich and reservoir-specific communities. Royal Society Open Science, v. 3, n. 11, p. 160635, nov. 2016.

Mächler, E., Walser, J. C., & Altermatt, F. (2021). Decision-making and best practices for taxonomy-free environmental DNA metabarcoding in biomonitoring using Hill numbers. *Molecular Ecology*, *30*(13), 3326-3339.

MAHON, A. R., JERDE, C. L., GALASKA, M., BERGNER, J. L., CHADDERTON, W. L., LODGE, D. M., ... & NICO, L. G. Validation of eDNA Surveillance Sensitivity for Detection of Asian Carps in Controlled and Field Experiments. PLoS ONE, v. 8, n. 3, p. e58316, 5 mar. 2013.

Majaneva, M., Diserud, O. H., Eagle, S. H., Boström, E., Hajibabaei, M., & Ekrem, T. (2018). Environmental DNA filtration techniques affect recovered biodiversity. *Scientific reports*, *8*(1), 4682.

Marques, V., Milhau, T., Albouy, C., Dejean, T., Manel, S., Mouillot, D., & Juhel, J. B. (2021). GAPeDNA: Assessing and mapping global species gaps in genetic databases for eDNA metabarcoding. Diversity and Distributions, 27(10), 1880-1892.

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

MCELROY, M. E., DRESSLER, T. L., TITCOMB, G. C., WILSON, E. A., DEINER, K., DUDLEY, T. L., ... & JERDE, C. L. Calibrating Environmental DNA Metabarcoding to Conventional Surveys for Measuring Fish Species Richness. Frontiers in Ecology and Evolution, v. 8, n. August, p. 1–12, 28 ago. 2020.

Melo, B. F., Dorini, B. F., Foresti, F., & Oliveira, C. (2018). Little divergence among mitochondrial lineages of Prochilodus (Teleostei, Characiformes). *Frontiers in Genetics*, *9*, 107.

MILAN, D. T., MENDES, I. S., DAMASCENO, J. S., TEIXEIRA, D. F., SALES, N. G., & CARVALHO, D. C. New 12S metabarcoding primers for enhanced Neotropical freshwater fish biodiversity assessment. Scientific Reports, v. 10, n. 1, p. 17966, 21 dez. 2020.

Minamoto, T., Yamanaka, H., Takahara, T., Honjo, M. N., & Kawabata, Z. I. (2012). Surveillance of fish species composition using environmental DNA. *Limnology*, *13*, 193-197.

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., ... & Iwasaki, W. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Royal Society open science*, *2*(7), 150088.

Mota, T. F., Fabrin, T., Depra, G. C., Gasques, L. S., OLIVEIRA, A. V., Pavanelli, C. S., ... & Prioli, A. J. (2018). Molecular characterization of Moenkhausia (Pisces: Characiformes) populations with different lateral line developmental levels. Anais da Academia Brasileira de Ciências, 90, 2815-2825.

Murienne, J., Cantera, I., Cerdan, A., Cilleros, K., Decotte, J. B., Dejean, T., ... & Brosse, S. (2019). Aquatic eDNA for monitoring French Guiana biodiversity. *Biodiversity data journal*, *7*.

NIELSEN, D.T.B.; PESSALI, T. C.; DUTRA, G. M. A new annual fish of the genus Simpsonichthys (Cyprinodontiformes: Cynolebiidae) from the upper Rio Jequitinhonha basin, Brazil. Zootaxa. V. 4263, P. 165-412, 2017.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... & Stevens, M. H. H. (2016). Eduard Szoecs and Helene Wagner. vegan: Community Ecology Package. *R package version 2.4*, *1*.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R. B., ... & Wagner, H. (2013). Community ecology package. *R package version*, *2*(0), 321-326.

OLDS, B. P., JERDE, C. L., RENSHAW, M. A., LI, Y., EVANS, N. T., TURNER, C. R., ... & LAMBERTI, G. A. Estimating species richness using environmental DNA. Ecology and Evolution, v. 6, n. 12, p. 4214–4226, 30 jun. 2016.
PAWLOWSKI, J.; LEJZEROWICZ, F.; ESLING, P. Next-Generation Environmental Diversity Surveys of Foraminifera: Preparing the Future. The Biological Bulletin, v. 227, n. 2, p. 93–106, out. 2014.

Pelicice, F. M., & Agostinho, A. A. (2008). Fish-passage facilities as ecological traps in large neotropical rivers. *Conservation biology*, *22*(1), 180-188.

PEREZ, C. R., BONAR, S. A., AMBERG, J. J., LADELL, B., REES, C., STEWART, W. T., ... & ROBINSON, A. T. Comparison of American Fisheries Society (AFS) Standard Fish Sampling Techniques and Environmental DNA for Characterizing Fish Communities in a Large Reservoir. North American Journal of Fisheries Management, v. 37, n. 5, p. 1010–1027, 5 out. 2017.

Petesse, M. L., & Petrere Jr, M. (2012). Tendency towards homogenization in fish assemblages in the cascade reservoir system of the Tietê river basin, Brazil. *Ecological Engineering*, *48*, 109-116.

PUGEDO, M. L., ANDRADE NETO, F. R., PESSALI, T. C., BIRINDELLI, J. L. O., & CARVALHO, D. C. Integrative taxonomy supports new candidate fish species in a poorly studied neotropical region: the Jequitinhonha River Basin. Genetica 144, 341-349, 2016.

RADINGER, J., BRITTON, J. R., CARLSON, S. M., MAGURRAN, A. E., ALCARAZ-HERNÁNDEZ, J. D., ALMODÓVAR, A., ... & GARCÍA-BERTHOU, E. Effective monitoring of freshwater fish. Fish and Fisheries, v. 20, n. 4, p. faf.12373, 29 maio 2019.

REID, Andrea J. et al. Emerging threats and persistent conservation challenges for freshwater biodiversity. **Biological Reviews**, v. 94, n. 3, p. 849-873, 2019.

REIS, R. E., ALBERT, J. S., DI DARIO, F., MINCARONE, M. M., PETRY, P., & ROCHA, L. A. Fish biodiversity and conservation in South America. Journal of fish biology, v. 89, n. 1, p. 12-47, 2016.

RUPPERT, K. M.; KLINE, R. J.; RAHMAN, M. S. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. Global Ecology and Conservation, v. 17, p. e00547, jan. 2019.

SALES, N. G., MARIANI, S., SALVADOR, G. N., PESSALI, T. C., & CARVALHO, D. C. Hidden diversity hampers conservation efforts in a highly impacted Neotropical river system. Frontiers in genetics, v. 9, p. 271, 2018.

SALES, N. G., WANGENSTEEN, O. S., CARVALHO, D. C., & MARIANI, S. Influence of preservation methods, sample medium and sampling time on eDNA recovery in a neotropical river. Environmental DNA, v. 1, n. 2, p. 119-130, 2019.

SALES, N. G., WANGENSTEEN, O. S., CARVALHO, D. C., DEINER, K., PRÆBEL, K., COSCIA, I., ... & MARIANI, S. Space-time dynamics in monitoring neotropical fish communities using eDNA metabarcoding. Science of the Total Environment, v. 754, p. 142096, 2021.

Santos, R. P., Melo, B. F., Yazbeck, G. M., Oliveira, R. S., Hilário, H. O., Prosdocimi, F., & Carvalho, D. C. (2021). Diversification of Prochilodus in the eastern Brazilian Shield: Evidence from complete mitochondrial genomes (Teleostei, Prochilodontidae). *Journal of Zoological Systematics and Evolutionary Research*, *59*(5), 1053-1063.

Shu, L., Chen, S., Li, P., & Peng, Z. (2022). Environmental DNA Metabarcoding Reflects Fish DNA Dynamics in Lentic Ecosystems: A Case Study of Freshwater Ponds. *Fishes*, *7*(5), 257.

Sigsgaard, E. E., Nielsen, I. B., Bach, S. S., Lorenzen, E. D., Robinson, D. P., Knudsen, S. W., ... & Thomsen, P. F. (2016). Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature ecology & evolution*, *1*(1), 0004.

SIGSGAARD, E. E., NIELSEN, I. B., BACH, S. S., LORENZEN, E. D., ROBINSON, D. P., KNUDSEN, S. W., ... & THOMSEN, P. F. Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. Nature Ecology & Evolution, v. 1, n. 1, p. 0004, 21 jan. 2017.

Strickler, K. M., Fremier, A. K., & Goldberg, C. S. (2015). Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. Biological Conservation, 183, 85–92. https://doi.org/10.1016/j.biocon.2014.11.038

SUTHERLAND, William J. Predicting the ecological consequences of environmental change: a review of the methods. **Journal of Applied Ecology**, p. 599-616, 2006.

Takeuchi, A., Sado, T., Gotoh, R. O., Watanabe, S., Tsukamoto, K., & Miya, M. (2019). New PCR primers for metabarcoding environmental DNA from freshwater eels, genus Anguilla. *Scientific Reports*, *9*(1), 1-11.

THOMSEN, P. F., KIELGAST, J. O. S., IVERSEN, L. L., WIUF, C., RASMUSSEN, M., GILBERT, M. T. P., ... & WILLERSLEV, E. Monitoring endangered freshwater biodiversity using environmental DNA. Mol ecol 21, 2565-2573, 2012a.

THOMSEN, P. F., KIELGAST, J., IVERSEN, L. L., MØLLER, P. R., RASMUSSEN, M., & WILLERSLEV, E. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. PLoS one 7, e41732.

Thomsen, P. F., Møller, P. R., Sigsgaard, E. E., Knudsen, S. W., Jørgensen, O. A., & Willerslev, E. (2016). Environmental DNA from seawater samples correlate with trawl catches of subarctic, deepwater fishes. *PloS one*, *11*(11), e0165252.
THOMSEN, P. F., MØLLER, P. R., SIGSGAARD, E. E., KNUDSEN, S. W., JØRGENSEN, O. A., & WILLERSLEV, E. Environmental DNA from seawater samples correlate with trawl catches of subarctic, deepwater fishes. PloS one 11, e0165252, 2016.

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

VALENTINI, A., TABERLET, P., MIAUD, C., CIVADE, R., HERDER, J., THOMSEN, P. F., ... & DEJEAN, T. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Molecular Ecology, v. 25, n. 4, p. 929–942, 2016.

Vono, V., & Birindelli, J. L. (2007). Natural history of Wertheimeria maculata, a basal doradid catfish endemic to eastern Brazil (Siluriformes: Doradidae). *Ichthyological Exploration of Freshwaters*, *18*(2), 183.

WILCOX, T. M., MCKELVEY, K. S., YOUNG, M. K., JANE, S. F., LOWE, W. H., WHITELEY, A. R., & SCHWARTZ, M. K. Robust Detection of Rare Species Using Environmental DNA: The Importance of Primer Specificity. PLoS ONE, v. 8, n. 3, p. e59520, 26 mar. 2013.

XIONG, F., SHU, L., GAN, X., ZENG, H., HE, S., & PENG, Z. Methodology for fish biodiversity monitoring with environmental DNA metabarcoding: The primers, databases and bioinformatic pipelines. Water Biology and Security, v. 1, n. 1, p. 100007, fev. 2022.

Yamamoto, S., Masuda, R., Sato, Y., Sado, T., Araki, H., Kondoh, M., ... & Miya, M. (2017). Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. *Scientific reports*, *7*(1), 1-12.

YAMANAKA, H.; MINAMOTO, T. The use of environmental DNA of fishes as an efficient method of determining habitat connectivity. Ecological Indicators, v. 62, p. 147–153, 2016.

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

Zawadzki, C. H., Roxo, F. F., & Da Graça, W. J. (2016). Hisonotus pachysarkos, a new species of cascudinho from the rio Ivaí basin, upper rio Paraná system, Brazil (Loricariidae: Otothyrinae). *Ichthyological Exploration of Freshwaters*, *26*(4), 373-383.

ZHANG, S., LU, Q., WANG, Y., WANG, X., ZHAO, J., & YAO, M. Assessment of fish communities using environmental DNA: Effect of spatial sampling design in lentic systems of different sizes. Molecular Ecology Resources, v. 20, n. 1, p. 242–255, 15 jan. 2020.

Zou, K., Chen, J., Ruan, H., Li, Z., Guo, W., Li, M., & Liu, L. (2020). eDNA metabarcoding as a promising conservation tool for monitoring fish diversity in a coastal wetland of the Pearl River Estuary compared to bottom trawling. *Science of the Total Environment*, *702*, 134704.

Supplementary material

Figures

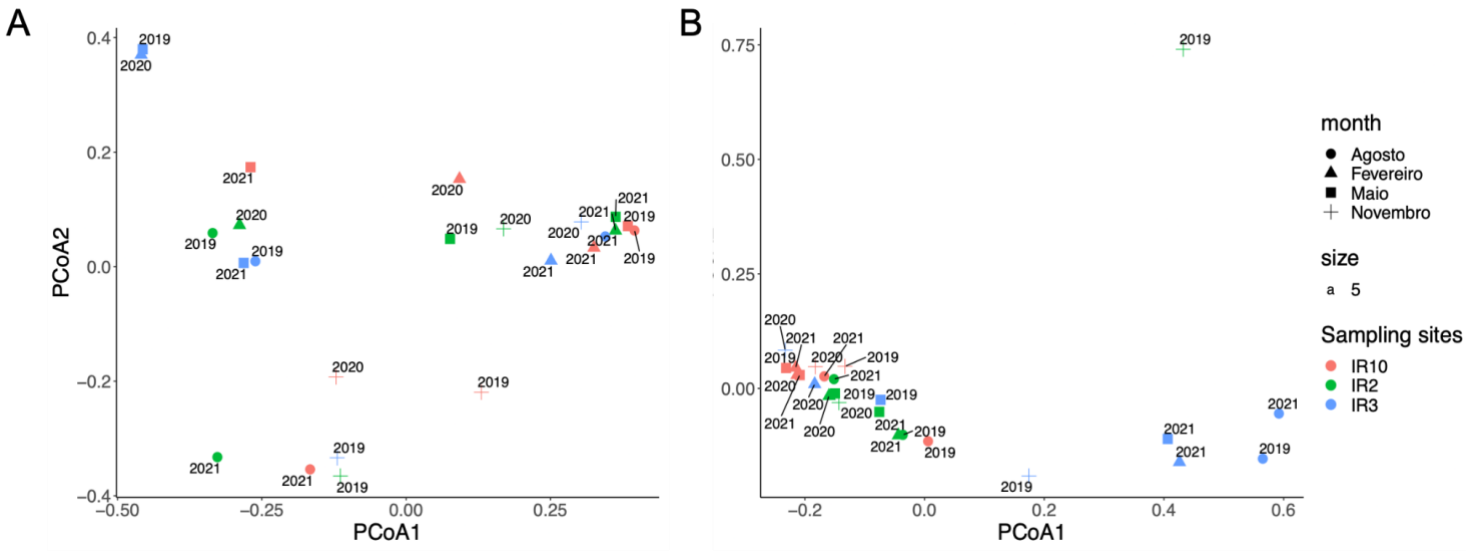


Figure S1. Principal Coordinates Analysis (PCoA) of β-diversity of sampling locations (Jaccard distance) for eDNA metabarcoding (A) and gill nets (B).

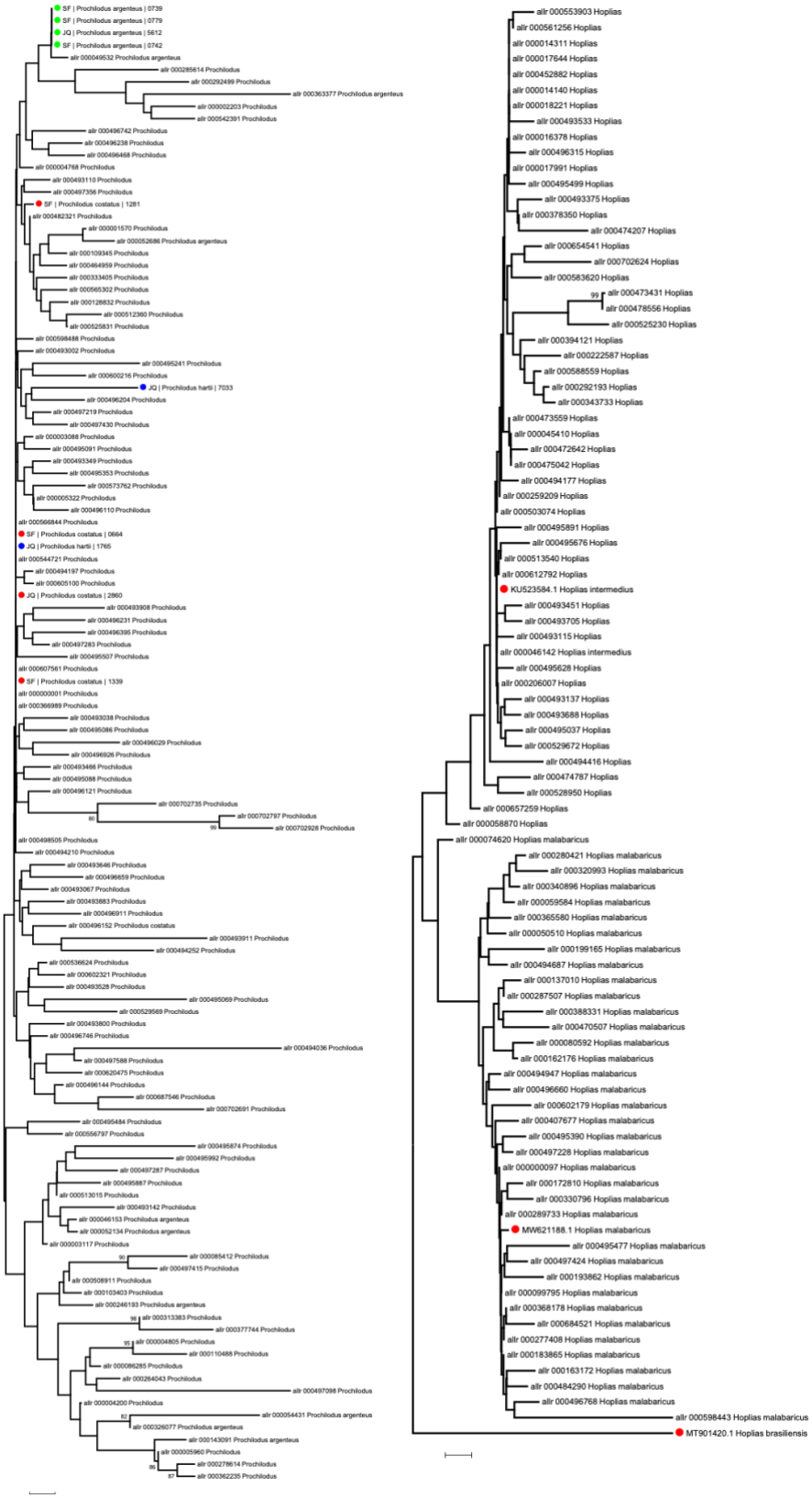


Figure S2. Phylogenetic tree for *Prochilodus* spp. and *Hoplias* spp.

Tables

Table S1. List of samples including primers and tags used.

Library	Samples	Tags	Primer forward	Primer reverse
RUN2	IR1-1-0m_mi	S502:N706	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	FB1	CACGTAT:GAGCTAT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	FB2	TCTCTTG:AAACGGC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	EB1	TATCGCA:TATCGCA	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	EB2	TATCGCA:TATCATT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	PB1	TATCATT:TCTCTTG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	PB2	TCTCTTG:CGGAAAC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	IR10 1 0m	AGCACAT:AGCACAT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	IR10 4 0m	CGGAAAC:TCTCTTG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	IR10 7 0m	TAGCGTG:TAGCGTG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	IR10 7 0m	AAAGACC:AAAGACC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	IR2 1 0m	CGGAAAC:CGGAAAC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	IR2 1 0m	AAAGACC:GAGGCCG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	IR2 4 0m	TCTCTTG:TCTCTTG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	IR2 4 0m	TGTGGGT:GAGCTAT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	IR2 4 0m	GAGCTAT:AGCACAT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	IR2 7 0m	TCTCTTG:GAGCTAT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	IR2 7 0m	TCTCTTG:TAGCGTG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	IR2 7 0m	GAGGCCG:GAGGCCG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	IR3 4 0m	TCTCTTG:AAAGACC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	IR3 7 0m	CGGAAAC:AGCACAT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN3	IR3 7 0m	CGGAAAC:AAACGGC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	positive_control	CTGCATA:CTGCATA	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	extraction_blank_1	AAACTTT:CGGAAAC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	extraction_blank_2	AAACGGC:GCGGGAG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	extraction_blank_3	AAACGGC:TATCGCA	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	extraction_blank_4	AAACGGC:TATCATT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	extraction_blank_5	CTGCATA:AAAGACC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	extraction_blank_6	TGTTATG:ATCCCGG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	extraction_blank_7	GAGGCCG:AAACGGC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	extraction_blank_8	GAGGCCG:GAGGCCG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	extraction_blank_9	CTCCTGA:CTCCTGA	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	extraction_blank_10	TGTTCAC:GAGGCCG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	extraction_blank_11	CGGAAAC:TGTGGGT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	extraction_blank_12	CTCCTGA:TCTCTTG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	extraction_blank_13	AAACTTT:ATCCCGG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	extraction_blank_14	TATCTAC:AAAGCAT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	extraction_blank_15	CACGTAT:CACGTAT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG

RUN4	extraction_blank_16	TGTTCAC:CTCCTGA	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	extraction_blank_17	ATCCCGG:AAACTTT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	extraction_blank_18	CTGCATA:GGTACCC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	extraction_blank_19	GAGGCCG:TAGCGTG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	filtration_blank_1	CGGAAAC:CGGAAAC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	filtration_blank_2	TGTGGGT:TGTGGGT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	filtration_blank_3	AGCACAT:AGCACAT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	filtration_blank_4	TATCGCA:TATCGCA	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	filtration_blank_5	TCTCTTG:TCTCTTG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	filtration_blank_6	TATCATT:TATCATT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	filtration_blank_7	TAGCGTG:TAGCGTG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	filtration_blank_8	GCGGGAG:GCGGGAG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	filtration_blank_9	GCGGGAG:TATCATT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	filtration_blank_10	AAACGGC:AAACGGC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	filtration_blank_11	CACGTAT:GAGCTAT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	filtration_blank_12	CACGTAT:AGCACAT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	filtration_blank_13	CACGTAT:CGGAAAC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	filtration_blank_14	CTGCATA:CGGAAAC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	filtration_blank_16	GGTACCC:TATCATT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	filtration_blank_15	TGTTATG:TGTGGGT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	filtration_blank_17	ACATTAT:AAACGGC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_1	CTCCTGA:GTTAGCA	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_10	GGTACCC:AAAGCAT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_11	GGTACCC:GGTACCC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_12	TATCATT:TCTCTTG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_13	TATCATT:TATCGCA	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_14	TATCATT:CGGAAAC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_15	TATCATT:GCGGGAG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_16	TGTTATG:ACATTAT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_17	ACATTAT:TGTTATG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_18	TGTTCAC:TGTTCAC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_19	TGTTCAC:GAGCTAT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_2	CTCCTGA:TGTTCAC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_20	GAGGCCG:GGTACCC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_21	AAACGGC:GTTAGCA	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_22	GAGCTAT:GAGCTAT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_23	GAGCTAT:TAGCGTG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_24	TAGCGTG:TCTCTTG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_3	GTTAGCA:CTCCTGA	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_4	GTTAGCA:TATCATT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_5	AAAGACC:AAAGACC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_6	ATCCCGG:TATCGCA	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG

RUN4	pcr_blank_7	ATCCCGG:TATCTAC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_8	TATCTAC:ATCCCGG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	pcr_blank_9	AAAGCAT:TCTCTTG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	IR10_mai19	AAAGCAT:TATCTAC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	IR10_jus_ago21	TGTGGGT:AAACGGC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	IR10_0m_fev20	TGTTCAC:ACATTAT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	IR10_0m_fev21	ATCCCGG:ATCCCGG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	IR10_ago21	TATCTAC:CTCCTGA	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	IR10_0m_mai21	AAACGGC:TGTGGGT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	IR2_0m_ago19	TAGCGTG:GCGGGAG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	IR2_0m_fev20	AAAGCAT:GTTAGCA	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	IR2_0m_mai21	AAAGACC:TGTGGGT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	IR2_0m_fev21	GTTAGCA:AAACTTT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	IR2_0m_ago21	ACATTAT:TATCTAC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	IR2_7_0m_mai19	TGTTATG:CTGCATA	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	IR3_0m_ago19	GAGGCCG:CACGTAT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	IR3_0m_fev21	TAGCGTG:TATCGCA	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	IR3_mai19	AAACTTT:GTTAGCA	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	IR3_0m_mai21	CACGTAT:GCGGGAG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	IR3_0m_ago21	TGTTCAC:AAAGCAT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	IR3_7_0m_mai21	ATCCCGG:AAACGGC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN4	IR3_7_0m_fev20	AAAGCAT:GGTACCC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN5	branco_extracao_dez21	TATCATT:TATCTAC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN5	branco_filtragem_dez21	TCTCTTG:ATCCCGG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN5	branco_pcr_dez21	TCTCTTG:AAAGCAT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN5	be_05_08_21	GTTAGCA:TATCTAC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN5	be_12_07_21	CTCCTGA:AAAGCAT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN5	be_13_07_21	CTCCTGA:GGTACCC	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN5	be_15_07_21	GTTAGCA:ATCCCGG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN5	bf_18_05_21	GTTAGCA:CTGCATA	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN5	bf_19_05_21	GTTAGCA:TGTTATG	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
RUN5	IR10_fev21	CTCCTGA:AAACTTT	GTCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	Extraction Blank	TCAGGTGA:TCAGGTGA	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	Extraction Blank	TCAGTTTC:TCAGTTTC	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	Extraction Blank	TGATAACC:TGATAACC	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	Extraction Blank	GTAGTTAC:GTAGTTAC	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	Extraction Blank	CGTGGCAA:CGTGGCAA	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	Extraction Blank	TTGTACAT:TTGTACAT	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	Extraction Blank	TTGGATGA:TTGGATGA	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR1-Negativo (amtule)	CTGGTTCC:CTGGTTCC	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR2-Control	GGCGTATC:GGCGTATC	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR3-Control	TCATAGAT:TCATAGAT	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG

SALF	Blank	AAAGCTAA:AAAGCTAA	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR1-1-Sup	TTATAGTC:TTATAGTC	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR1-1-Sup	CGAGGCCG:CGAGGCCG	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR1-1-Sup	GAGTACGG:GAGTACGG	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	positive_control	CCGTAATG:CCGTAATG	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR1-3-Sup	TCAGGTGA:GTAGTTAC	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR1-3-Sup	TCAGTTTC:CGTGGCAA	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR1-3-Sup	TGATAACC:TTGTACAT	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	positive_control	GTAGTTAC:TTGGATGA	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR1-Sup	TTATAGTC:TTGTATCC	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR1-5-Sup	GAGTACGG:CGATAAGT	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IRI-5-Sup	TTGTATCC:TGATATGA	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IRI-5-Sup	GAGCAGTA:CCGTAATG	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR1-7-Sup	TCAGTTTC:CTGGTTCC	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR1-7-Sup	TGATAACC:GGCGTATC	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR1-7-Sup	GTAGTTAC:TCATAGAT	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR2-1-Sup	CGTGGCAA:AAAGCTAA	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR2-1-Sup	TTGTACAT:TTATAGTC	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR2-1-Sup	TTGGATGA:CGAGGCCG	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR2-7-Sup	CGATAAGT:GGATATCT	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR2-7-Sup	TGATATGA:TCGTCATC	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	positive_control	CCGTAATG:TCAGGTGA	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR2-7-Sup	ACCCAACG:TCAGTTTC	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR3-1-Sup	GAGTAGCC:TGATAACC	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR3-1-Sup	GAGTATAA:GTAGTTAC	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR3-1-Sup	GGATATCT:CGTGGCAA	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR3-3-Sup	GGCGTATC:TGATATGA	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR3-3-Sup	TCATAGAT:CCGTAATG	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR3-3-Sup	AAAGCTAA:ACCCAACG	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR3-5-Sup	ACCCAACG:CGTGGCAA	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR3-5-Sup	GAGTAGCC:TTGTACAT	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR3-5-Sup	GAGTATAA:TTGGATGA	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR3-7-Sup	CTGGTTCC:ACCCAACG	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR3-7-Sup	GGCGTATC:GAGTAGCC	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	IR3-7-Sup	TCATAGAT:GAGTATAA	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG
SALF	positive_control	CGATAAGT:CGTGGCAA	GCCGGTAAAACTCGTGCCAGC	CATAGTGGGGTATCTAATCCCAGTTTG

Table S2: Count of reads after each filtering step.

Filtering step	ALLRUNS
reads before filtering	18909076
reads after removing PC	18359474
reads after removing blanks	18359474
reads after removing non-fish	17765064
reads after >=0.97 and < 5 reads total	12354785

Table S3. List of all MOTUs recovered for all treatments.

rank	best_identity	order_name	family_name	Scientific_name
genus	1	Characiformes	Prochilodontidae	*Prochilodus*
species	1	Characiformes	Serrasalmidae	*Serrasalmus brandtii*
species	1	Characiformes	Characidae	*Moenkhausia costae*
species	0.9821428571428571	Characiformes	Characidae	*Astyanax lacustris*
species	0.9767441860465116	Gymnotiformes	Gymnotidae	*Gymnotus carapo*
genus	0.9941176470588235	Characiformes	Anostomidae	*Leporinus*
species	0.9825581395348837	Gymnotiformes	Gymnotidae	*Gymnotus carapo*
species	0.9880239520958084	Characiformes	Erythrinidae	*Hoplias malabaricus*
species	0.9885714285714285	Characiformes	Characidae	*Moenkhausia costae*
species	0.9715909090909091	Characiformes	Characidae	*Moenkhausia costae*
genus	0.9767441860465116	Characiformes	Prochilodontidae	*Prochilodus*
species	1	Siluriformes	Pimelodidae	*Pimelodus maculatus*
genus	0.9767441860465116	Characiformes	Prochilodontidae	*Prochilodus*
species	0.9885714285714285	Characiformes	Characidae	*Moenkhausia costae*
genus	0.9824561403508772	Characiformes	Prochilodontidae	*Prochilodus*
species	0.9715909090909091	Characiformes	Characidae	*Moenkhausia costae*
genus	0.9883040935672515	Characiformes	Prochilodontidae	*Prochilodus*
species	0.9771428571428571	Characiformes	Characidae	*Moenkhausia costae*
species	0.9882352941176471	Characiformes	Serrasalmidae	*Serrasalmus brandtii*
species	0.9942528735632183	Siluriformes	Pimelodidae	*Rhamdia quelen*
genus	0.9880239520958084	Characiformes	Erythrinidae	*Hoplias*
genus	0.9760479041916168	Characiformes	Erythrinidae	*Hoplias*
species	1	Siluriformes	Pimelodidae	*Pseudoplatystoma reticulatum*
species	0.9771428571428571	Characiformes	Characidae	*Moenkhausia costae*
species	0.9771428571428571	Characiformes	Characidae	*Moenkhausia costae*
genus	0.9764705882352941	Characiformes	Anostomidae	*Leporinus*
species	0.9883040935672515	Siluriformes	Doradidae	*Wertheimeria maculata*
species	0.9704142011834319	Characiformes	Characidae	*Astyanax lacustris*

genus	0.988023952095808	Characiformes	Erythrinidae	*Hoplias*
species	0.9940828402366686	Characiformes	Crenuchidae	*Characidium sp.*
species	1	Characiformes	Erythrinidae	*Hoplias intermedius*
species	1	Characiformes	Prochilodontidae	*Prochilodus argenteus*
species	0.971264367816092	Siluriformes	Pimelodidae	*Rhamdia quelen*
species	0.9940828402366686	Cichliformes	Cichlidae	*Australoheros facetus*
species	1	Characiformes	Anostomidae	*Megaleporinus garmani*
genus	0.988095238095238	Characiformes	Characidae	*Astyanax*
species	0.988023952095808	Characiformes	Bryconidae	*Brycon sp.*
species	0.976331360946746	Characiformes	Characidae	*Astyanax lacustris*
species	0.970414201183432	Characiformes	Characidae	*Astyanax lacustris*
species	1	Siluriformes	Auchenipteridae	*Trachelyopterus striatulus*
genus	0.976331360946746	Characiformes	Characidae	*Hasemania*
species	0.982558139534884	Characiformes	Curimatidae	*Steindachnerina elegans*
species	1	Cichliformes	Cichlidae	*Geophagus brasiliensis*
species	0.970059880239521	Characiformes	Erythrinidae	*Hoplias malabaricus*
species	0.970238095238095	Characiformes	Characidae	*Astyanax lacustris*
genus	0.970238095238095	Characiformes	Characidae	*Astyanax*
species	1	Cyprinodontiformes	Poeciliidae	*Poecilia reticulata*
species	0.982352941176471	Characiformes	Serrasalmidae	*Serrasalmus brandtii*
family	0.976470588235294	Characiformes	Anostomidae	*Anostomidae*
genus	0.970414201183432	Characiformes	Characidae	*Astyanax*
species	0.970238095238095	Characiformes	Characidae	*Astyanax lacustris*
genus	0.976608187134503	Characiformes	Anostomidae	*Leporinus*
species	0.970414201183432	Cichliformes	Cichlidae	*Australoheros facetus*
species	0.982456140350877	Characiformes	Serrasalmidae	*Serrasalmus brandtii*
species	0.982352941176471	Characiformes	Serrasalmidae	*Serrasalmus brandtii*
species	0.982352941176471	Characiformes	Anostomidae	*Megaleporinus garmani*
species	0.976047904191617	Characiformes	Erythrinidae	*Hoplias malabaricus*
species	0.970238095238095	Characiformes	Characidae	*Astyanax lacustris*
genus	0.982456140350877	Characiformes	Prochilodontidae	*Prochilodus*
species	0.982352941176471	Characiformes	Anostomidae	*Megaleporinus garmani*
species	0.970930232558139	Characiformes	Serrasalmidae	*Serrasalmus brandtii*
species	0.977142857142857	Characiformes	Characidae	*Moenkhausia costae*
species	0.982857142857143	Characiformes	Characidae	*Moenkhausia costae*
species	0.970059880239521	Characiformes	Erythrinidae	*Hoplias malabaricus*
species	0.970414201183432	Characiformes	Crenuchidae	*Characidium sp.*
species	0.976047904191617	Characiformes	Erythrinidae	*Hoplias malabaricus*
genus	0.988023952095808	Characiformes	Erythrinidae	*Hoplias*
species	0.982857142857143	Characiformes	Characidae	*Moenkhausia costae*
genus	0.988023952095808	Characiformes	Erythrinidae	*Hoplias*
species	0.982248520710059	Perciformes	Serranidae	*Epinephelus lanceolatus*
species	0.976331360946746	Characiformes	Characidae	*Astyanax lacustris*
species	0.977142857142857	Characiformes	Characidae	*Moenkhausia costae*
species	1	Characiformes	Serrasalmidae	*Colossoma macropomum*

species	0.976047904191617	Cichliformes	Cichlidae	*Coptodon zillii*
genus	0.988095238095238	Characiformes	Characidae	*Astyanax*
species	1	Siluriformes	Callichthyidae	*Hoplosternum littorale*
genus	1	Cichliformes	Cichlidae	*Oreochromis*
genus	0.976470588235294	Characiformes	Anostomidae	*Leporinus*
species	0.982857142857143	Characiformes	Characidae	*Moenkhausia costae*
species	0.970588235294118	Characiformes	Characidae	*Astyanax lacustris*
genus	0.970414201183432	Characiformes	Characidae	*Astyanax*
species	0.971428571428571	Characiformes	Characidae	*Moenkhausia costae*
genus	0.970414201183432	Characiformes	Characidae	*Astyanax*
genus	0.970760233918129	Characiformes	Anostomidae	*Leporinus*
species	0.982456140350877	Characiformes	Serrasalmidae	*Serrasalmus brandtii*
species	0.982857142857143	Characiformes	Characidae	*Moenkhausia costae*
species	0.970930232558139	Siluriformes	Auchenipteridae	*Trachelyopterus striatulus*
species	0.976744186046512	Characiformes	Serrasalmidae	*Serrasalmus brandtii*
genus	1	Siluriformes	Pimelodidae	*Pimelodus*
genus	0.970930232558139	Characiformes	Anostomidae	*Leporinus*
species	0.976331360946746	Characiformes	Crenuchidae	*Characidium sp.*
genus	0.982456140350877	Characiformes	Prochilodontidae	*Prochilodus*
genus	0.988304093567251	Characiformes	Prochilodontidae	*Prochilodus*
genus	0.982352941176471	Characiformes	Anostomidae	*Leporinus*
species	0.970760233918129	Characiformes	Crenuchidae	*Characidium sp.*
species	0.971264367816092	Siluriformes	Pimelodidae	*Rhamdia quelen*
genus	0.976190476190476	Characiformes	Characidae	*Astyanax*
genus	0.988304093567251	Characiformes	Prochilodontidae	*Prochilodus*
species	0.994047619047619	Cyprinodontiformes	Poeciliidae	*Phalloceros sp.J*
species	0.970238095238095	Characiformes	Characidae	*Astyanax lacustris*
genus	0.988023952095808	Characiformes	Erythrinidae	*Hoplias*
species	0.970930232558139	Gymnotiformes	Gymnotidae	*Gymnotus carapo*
genus	0.982456140350877	Characiformes	Prochilodontidae	*Prochilodus*
genus	0.976608187134503	Characiformes	Prochilodontidae	*Prochilodus*
genus	0.976608187134503	Characiformes	Prochilodontidae	*Prochilodus*
genus	0.976047904191617	Characiformes	Erythrinidae	*Hoplias*
species	0.982248520710059	Characiformes	Crenuchidae	*Characidium sp.*
species	0.970414201183432	Characiformes	Crenuchidae	*Characidium sp.*
genus	0.976608187134503	Characiformes	Prochilodontidae	*Prochilodus*
species	0.976331360946746	Characiformes	Crenuchidae	*Characidium sp.*
species	0.970414201183432	Characiformes	Crenuchidae	*Characidium sp.*
species	0.982248520710059	Characiformes	Crenuchidae	*Characidium sp.*
genus	0.988439306358382	Characiformes	Prochilodontidae	*Prochilodus*

Table S4. Species detected by eDNA metabarcoding, its relative read abundance per sampling campaign.

Taxa	2019			2020		2021		
	May	Aug	Nov	Feb	Nov	Feb	May	Aug
Characiformes	**93,75166**	**85,54445**	**76,71663**	**67,80493**	**89,33637**	**81,72794**	**74,57374**	**94,46199**
Anostomidae	8,02444	2,57971	0,00021	1,23182	1,76944	31,38715	2,11701	11,04015
Leporinus sp.	8,02444	2,57971		1,2267	1,52692	30,01874	2,11701	11,04015
Megaleporinus garmani			0,00021	0,00503	0,24129	1,36585		
Bryconidae			0,33365		0,00214			
Brycon sp.					0,00214			
Characidae	75,91255	66,97916	46,5397	51,23555	81,87199	42,99378	57,61409	63,57859
Astyanax lacustris	27,5268	20,83902	4,04608	23,75406	8,05081	7,06958	1,37088	5,05564
Astyanax sp.		0,27457		2,16155	0,0678	9,26944	4,50472	
Hasemania sp.			0,00074		0,01479	0,01044		
Moenkhausia costae	48,38575	45,86557	42,49288	25,31994	73,7386	26,64432	51,7385	58,52295
Crenuchidae			3,85157	4,09792	0,11763	0,07546	0,15377	0,00022
Characidium sp.			3,85157	4,09792	0,11763	0,07546	0,15377	0,00022
Curimatidae					0,01742			
Steindachnerina elegans					0,01742			
Erythrinidae	1,21152	2,9583	5,50432	1,41123	3,4625	2,75639	10,07452	4,22144
Hoplias intermedius			0,06766	0,00698	0,08106	0,20927	4,25669	1,43525
Hoplias malabaricus	1,21152	2,9583	5,33853	1,39488	2,89505	2,54697	4,46507	2,78619
Hoplias sp.			0,09814	0,00937	0,48639	0,00014	1,35276	
Prochilodontidae	0,17521	4,36573	19,68908		0,86647	2,43208	4,03831	9,83263
Prochilodus argenteus			0,05141		0,0391	0,06677		
Prochilodus sp.	0,17521	4,36573	19,63767		0,82736	2,36531	4,03831	9,83263
Serrasalmidae	8,42794	8,66155	0,7981	9,82843	1,22879	2,08308	0,57604	5,78897
Colossoma macropomum						0,0449		
Serrasalmus brandtii	8,42794	8,66155	0,7981	9,82843	1,22879	2,03818	0,57604	5,78897
Cichliformes		**8,74763**	**5,47353**	**1,33259**	**0,19878**	**10,52572**	**12,55121**	**1,75411**
Cichlidae		8,74763	5,47353	1,33259	0,19878	10,52572	12,55121	1,75411
Australoheros facetus		4,14835	3,84633	1,33041	0,18839	0,09367	12,48452	1,75411
Coptodon zillii				0,00217		10,12897	0,0667	
Geophagus brasiliensis		4,59928	0,00273		0,01039			
Oreochromis sp.			1,62446			0,00192		
Cyprinodontiformes			**7,69251**	**0,00008**	**0,00039**	**0,06665**	**0,33091**	
Poeciliidae			7,69251	0,00008	0,00039	0,06665	0,33091	
Phalloceros sp.J			7,69251					
Poecilia reticulata				0,00008	0,00039	0,06665	0,33091	
Gymnotiformes	**6,11366**	**5,70792**	**6,27118**	**1,1423**	**4,72038**	**1,74796**	**8,97491**	**2,68083**
Gymnotidae	6,11366	5,70792	6,27118	1,1423	4,72038	1,74796	8,97491	2,68083
Gymnotus carapo	6,11366	5,70792	6,27118	1,1423	4,72038	1,74796	8,97491	2,68083
Siluriformes	**0,13468**		**3,84615**	**29,72009**	**5,74409**	**5,93173**	**3,56923**	**1,10308**
Auchenipteridae				25,01918	0,05692	5,92373	0,10976	0,0624
Trachelyopterus striatulus				25,01918	0,05692	5,92373	0,10976	0,0624
Callichthyidae						0,0058		
Hoplosternum littorale						0,0058		

Doradidae	0,13468	3,84615	4,27851	0,05711	0,00059	1,44559	0,062
Wertheimeria maculata	0,13468	3,84615	4,27851	0,05711	0,00059	1,44559	0,062
Heptapteridae			0,4224	3,40961	0,00161	1,29412	
Rhamdia quelen			0,4224	3,40961	0,00161	1,29412	
Pimelodidae				2,22044		0,71975	0,97868
Pimelodus maculatus				0,7853			
Pimelodus sp.							0,03274
Pseudoplatystoma reticulatum				1,43514		0,71975	0,94594

Table S5. Summary of PERMANOVA results (R_2-effect sizes and significance level) showing the effect of sampling season, year, month and site on MOTU diversity recovery.

Dataset	Variable	df	Sum of Sqs	F	R2	p
eDNA+gillnets	Season	1	0.2998	0.7561	0.01617	0.868
	Year	1	0.6873	1.7712	0.03708	0.012*
	Month	3	1.2211	1.2134	0.15399	0.078
	Site	2	0.7089	0.8946	0.03824	0.695
	Method	1	2.5987	7.4994	0.14018	0.001*
eDNA	Season	1	0.3573	1.0380	0.04506	0.377
	Year	1	0.5657	1.6902	0.07134	0.013*
	Month	3	1.2211	1.2134	0.15399	0.078
	Site	2	0.6151	0.8829	0.07756	0.761
Gillnets	Season	1	0.2411	0.6727	0.02967	0.919
	Year	1	0.7963	2.3906	0.09801	0.001*
	Month	3	1.3275	1.302	0.16339	0.057
	Site	2	0.7079	1.0021	0.08713	0.433

Table S6. Species detected by eDNA metabarcoding and gill nets and its status.

Taxa	Status	eDNA	Gill nets
Astyanax_lacustris	native	■	■
*Astyanax*_sp	native	■	
Australoheros_facetus	non-native	■	
*Brycon*_sp	not described yet	■	
Brycon_howesi	native		■
*Characidium*_sp	not described yet	■	
Colossoma_macropomum	non-native	■	
Coptodon_zilii	introduced	■	
Geophagus_brasiliensis	native	■	
Gymnotus_carapo	native	■	■
*Hasemania*_sp	native	■	
Hoplias_intermedius	introduced	■	
Hoplias_malabaricus	native	■	■
*Hoplias*_sp	not described yet	■	
Hoplosternum_littorale	non-native	■	
Hypomasticus_steindachneri	native		■
Hypostomus_spp	native		■
*Leporinus*_sp	not described yet	■	
Megaleporinus_macrocephalus	not described yet	■	■
Megaleporinus_elongatus	native		■
Megaleporinus_garmani	native	■	■
Moenkhausia_costae	introduced	■	
*Oreochromis*_sp	introduced	■	
*Phalloceros*_sp	native	■	
Pimelodus_maculatus	non-native	■	
*Pimelodus*_sp	not known for this basin	■	
Poecilia_reticulata	introduced	■	
Prochilodus_argenteus	introduced	■	
*Prochilodus*_sp	native	■	■
Pseudoplatystoma_reticulatum	introduced	■	
Rhamdia_quelen	native	■	
Serrasalmus_brandtii	introduced	■	■
Steindachnerina_elegans	native	■	
Trachelyopterus_galeatus	native		■
Trachelyopterus_striatulus	native	■	
Wertheimeria_maculata	native	■	■

Table S7. Species identified applying the minimum identity of 0.97

Species	N MOTUs
Astyanax lacustris	10
Astyanax sp.	6
Australoheros fascetus	2
Brycon sp.	1
Characidium sp.	9
Colossoma macropomum	1
Coptodon zilli	1
Geophagus brasiliensis	1
Gymnotus carapo	3
Hasemania sp.	1
Hoplias intermedius	1
Hoplias malabaricus	5
Hoplias sp.	7
Hoplosternum littorale	1
Leporinus sp.	7
Megaleporinus garmani	3
Moenkhuasia costae	15
Oreochromis sp.	1
Phalloceros sp. J	1
Pimelodus maculatus	1
Pimelodus sp.	1
Poecilia reticulata	1
Prochilodus argenteus	1
Prochilodus sp.	1
Pseudoplatystoma reticulata	1
Rhamdia quelen	3
Serrassalmus brandtii	8
Steindachnerina elegans	1
Trachelyopterus striatulus	2
Wertheimeria maculata	1

Table S8. Species richness by eDNA metabarcoding recovered at each sampling site during the sampling campaigns.

Site	Month	Year	eDNA Richness	Gill nets Richness
IR10	May	2019	4	3
IR2	May	2019	5	3
IR3	May	2019	4	5
IR10	Aug	2019	1	3
IR2	Aug	2019	4	6
IR3	Aug	2019	7	5
IR10	Nov	2019	7	4
IR2	Nov	2019	5	2
IR3	Nov	2019	5	9
IR10	Feb	2020	10	4
IR2	Feb	2020	1	3
IR3	Feb	2020	8	5
IR10	Nov	2020	7	1
IR2	Nov	2020	11	2
IR3	Nov	2020	13	1
IR10	Feb	2021	4	3
IR2	Feb	2021	10	5
IR3	Feb	2021	12	4
IR10	May	2021	8	3
IR2	May	2021	6	7
IR3	May	2021	10	6
IR10	Aug	2021	3	3
IR2	Aug	2021	2	4
IR3	Aug	2021	6	4

7. DISCUSSÃO GERAL

Apesar do aumento exponencial de publicações de eDNA *metabarcoding*, a maioria dos estudos tem sido conduzido em regiões temperadas (Valdéz-Moreno et al., 2019; McDevitt et al., 2019) e em áreas razoavelmente acessíveis. Entretanto, poucos estudos são conduzidos em regiões megadiversas, como a região Neotropical (Cilleros et al., 2019; Sales et al., 2021), especialmente em regiões afetadas por represamentos (Dal Pont et al., 2021), um dos maiores estressores para a comunidade de peixes de água doce (He et al., 2018). Atualmente, o entendimento de como o eDNA *metabarcoding* poderá ser um método bem estabelecido para avaliação da biodiversidade e monitoramento ecológico em reservatórios neotropicais é relativamente limitado e se concentram em sua maioria na Ásia (Hayami et al., 2020; Li et al., 2022; Ma et al., 2022; Xu et al., 2022).

Devido a escassa aplicação em regiões que possuem elevada biodiversidade e condições ambientais diferentes (e.g. temperaturas mais elevadas, acidez e turbidez da água, entre outros), aspectos metodológicos como a falta de conhecimento da biodiversidade local e ausência de sequências no banco de dados referência (Sales et al., 2018, Jackman et al., 2021), vieses de *primer* e as diferentes formas de identificação taxonômica (MOTUs x ASVs; Cilleros et al., 2019, Sales et al., 2019, Jackman et al., 2021) precisam ser melhor avaliados e compreendidos antes de aplicar essa técnica na identificação da comunidade de peixes em larga escala. Portanto, antes de aplicar o eDNA *metabarcoding* em regiões megadiversas, como as bacias hidrográficas neotropicais, é altamente recomendável o desenvolvimento e validação de marcadores moleculares que possam fornecer uma atribuição taxonômica confiável e robusta. MacDonald & Sarre (2017) sugeriram um fluxo de trabalho para o desenvolvimento e validação de *primers* específicos para análises de *metabarcoding* em estudos ecológicos. Essa estrutura inclui a construção de um banco de dados de referência e sua avaliação filogenética, *design* de primer e avaliação *in silico* e *in vitro* da especificidade e sensibilidade do *primer*. Dessa forma, nesse trabalho desenvolvemos uma biblioteca referência de sequências do gene rRNA 12S, desenhamos um conjunto de *primers* específicos para espécies da região Neotropical, avaliamos a utilização desse *primer* em amostras de comunidades simuladas de peixes neotropicais e finalmente, aplicamos a metodologia do eDNA *metabarcoding* em um reservatório Neotropical, comparando os dados obtidos aos dados de redes de emalhar, representando os métodos tradicionais de pesca.

A biodiversidade de espécies varia muito entre as regiões geográficas (Abell et al., 2008; Olson et al., 2001; Spalding et al., 2007), complicando ainda mais o processo de seleção de *primers* mais adequados para identificação taxonômica, já que o desempenho dos marcadores para *metabarcoding* varia dependendo da composição de espécies no local de amostragem. Dessa forma, primeiramente, desenhamos três conjuntos de *primers* baseados em uma biblioteca referência customizada, compreendendo 67 espécies de peixes neotropicais. O *primer* NeoFish_3, que amplifica, ~190 pares de bases (pb), foi avaliado e demonstrou ter resolução taxonômica suficiente para delimitar as espécies do banco de dados referência desenvolvido e customizado para esse estudo, fornecendo o mesmo número de MOTUs ao aplicar análises bayesianas e de distâncias genéticas (70 MOTUs). Além disso, fomos capazes de definir um limite baseado em distância interespecífica para a delimitação de espécies que seria útil em análises de bioinformática. Assim, nosso banco de dados de referência personalizado e os *primers mini-barcodes* desenhados são um recurso importante para uma avaliação da biodiversidade em larga escala baseada em DNA, como o eDNA *metabarcoding*, que pode ajudar na conservação da ictiofauna neotropical.

Apesar disso, prevê-se que alguns *primers* sejam mais afetados por vieses de amplificação, e certas espécies podem até permanecer não detectadas devido a grandes incompatibilidades do *primer*-molde. Dessa forma, é de grande importância a condução de experimentos controlados, usando comunidades simuladas de espécies-alvo, para a avaliação de *primers mini-barcodes* em relação a influência de algumas características de amostras. Nesse sentido, posteriormente ao desenvolvimento dos *primers*, testamos o primer desenvolvido pelo nosso grupo, o NeoFish_3, junto a dois conjuntos de *primers* também amplamente utilizados em estudos de eDNA *metabarcoding*, o MiFish (~170pb; Myia et al., 2015) e o Teleo1 (~100pb; Valentini et al., 2016), testamos a influência dos seguintes fatores em amostras de comunidades de peixes simuladas: (1) a proporção de DNA de entrada, (2) a escolha do marcador molecular na detecção de espécies em comunidades simuladas e (3) a comparação do desempenho da abordagem livre de identificação molecular usando ASV/MOTUs e a atribuição de leituras de DNA para espécies usando uma biblioteca curada e customizada, aproximadamente, 100 espécies de peixes neotropicais. Assim, nesse trabalho também mostramos que a abundância relativa de leituras (RRA) é dependente da composição da comunidade, do marcador

utilizado e da concentração de DNA na amostra. Além disso, como a concentração das espécies influencia sua amplificação e o número de ASVs/MOTUs recuperados para uma única espécie, estudos com foco na abundância estimada de espécies usando *metabarcoding* devem considerar essas ressalvas metodológicas. Surpreendentemente, o *primer* desenvolvido pelo nosso grupo, o NeoFish_3, demonstrou alguns falhas para identificar alguns táxons-alvo, principalmente de espécies da ordem dos Siluriformes. Isso destaca a importância de experimentos controlados usando comunidades de espécies conhecidas antes que grandes investimentos sejam feitos na avaliação da biodiversidade usando métodos não invasivos, como eDNA *metabarcoding*. Assim, para a condução das próximas análises do nosso trabalho e aplicação do eDNA *metabarcoding* na bacia do Jequitinhonha, utilizamos o *primer* MiFish, que apresentou capacidade para detecção de todas as espécies do banco de dados referência customizado.

Muitos estudos tem sido realizados em diversos ambientes aquáticos, mas poucos estão focados em entender a capacidade do eDNA *metabarcoding* em detectar espécies quando comparado aos métodos tradicionais de captura de peixes, especialmente em regiões megadiversas, como a região neotropical. Neste estudo, avaliamos o desempenho do eDNA *metabarcoding* para descrever a comunidade de peixes de água doce por meio da comparação de dados de eDNA (número de leituras de sequências de DNA/espécies) com os dados de capturas por rede de espera em um ambiente de represamento hidrelétrico neotropical.

Nesse estudo, utilizamos um banco de dados referência mais robusto (486 sequências, representando 294 espécies), entretanto, 11 das 30 espécies identificadas pelo eDNA *metabarcoding* não puderam ser identificadas em nível de espécie, evidenciando que a falta de um banco de dados de referência abrangente continua sendo uma das principais limitações dessa técnica, como já reportado em estudos anteriores (Cilleros et al., 2018; Sales et al., 2019). Além disso, a recuperação de uma grande quantidade de MOTUs para um único táxon demonstra a possibilidade desse método em evidenciar uma possível diversidade intraspecífica e deve ser melhor investigada.

Nossos resultados também demonstraram que o eDNA *metabarcoding* permitiu a detecção de espécies que não foram detectadas pelo levantamento com redes de espera. Por outro lado, o inverso também aconteceu, mas com um menor número de espécies. Curiosamente, ambas metodologias mostraram uma diferença significativa

na comunidade de peixes, demonstrando o papel complementar do eDNA *metabarcoding* no monitoramento de peixes. Existem várias explicações possíveis para a falha do eDNA *metabarcoding* em detectar certas espécies de peixes, incluindo desempenho do *primer* (Takeuchi et al., 2019), qualidade e integridade do banco de dados de referência (Callahan et al. 2017) e viés de PCR em amplificar sequências que estão em baixa abundância (Ficetola et al., 2015, Hänfling et al., 2016).

8. CONSIDERAÇÕES FINAIS

Os dados aqui apresentados mostram o potencial do uso do eDNA *metabarcoding* para fornecer informações adicionais sobre a biodiversidade e contribuir para a conservação e manejo de uma comunidade de peixes neotropicais de água doce. Entretanto, embora esses dados forneçam informações preliminares importantes sobre a ocorrência e distribuição das espécies, eles devem ser usados com cautela.

A eficácia deste método inovador na recuperação de DNA de espécies de peixes em é amplamente reconhecida. Até fevereiro de 2022, a literatura contava com mais de 120 estudos publicados com foco no uso do método do eDNA *metabarcoding* para monitorar as composições das comunidades de peixes de diversos ambientes aquáticos (Leese et al., 2018; McElroy et al., 2020; McClenaghan et al., 2020). Entretanto, algumas limitações ainda devem ser discutidas antes de tirar conclusões finais.

Destacamos ao longo desse estudo que o melhor entendimento da ecologia do eDNA, que inclui a compreensão da origem, persistência, transporte e degradação do eDNA dentro desses sistemas (Barnes et al., 2016) permitirá a interpretação mais robusta dos dados obtidos. E como reportado por diversos estudos anteriormente como um dos três principais limitadores na aplicação dessa metodologia, um banco de dados de sequências referências mais robusto e completo é necessário, principalmente em regiões megadiversas, em que a biodiversidade é subestimada como reportado para a nossa área de estudo (Andrade-Neto, 2010, Sales et al., 2021).

REFERÊNCIAS BIBLIOGRÁFICAS

ABELL, R. et al. Freshwater ecoregions of the world: a new map of biogeographic units for freshwater biodiversity conservation. BioScience, v. 58, n. 5, p. 403-414, 2008.

ALMOND, R. E., GROOTEN, M., & PETERSON, T. (2020). Living Planet Report 2020-Bending the curve of biodiversity loss. World Wildlife Fund.

ANDERSEN, Jeremy C. et al. Categorization of species as native or nonnative using DNA sequence signatures without a complete reference library. Ecological applications, v. 29, n. 5, p. e01914, 2019.

ANDRADE NETO, F. R. (2010). Estado atual do conhecimento sobre a fauna de peixes da bacia do Jequitinhonha. INSTITUTO ESTADUAL DE FLORESTAS—MG, 23.

BARNES, Matthew A. et al. Environmental conditions influence eDNA persistence in aquatic systems. Environmental science & technology, v. 48, n. 3, p. 1819-1827, 2014.

BECKER, R. A. et al. DNA barcoding and morphological identification of neotropical ichthyoplankton from the Upper Paraná and São Francisco. Journal of Fish Biology, v. 87, n. 1, p. 159-168, 2015.

BOHMANN, Kristine et al. Environmental DNA for wildlife biology and biodiversity monitoring. Trends in ecology & evolution, v. 29, n. 6, p. 358-367, 2014.

BONAR, S. A. et al. Standard methods for sampling North American freshwater fishes. 2009.

BOUSSARIE, Germain et al. Environmental DNA illuminates the dark diversity of sharks. Science advances, v. 4, n. 5, p. eaap9661, 2018.

BRAUER, C. J. et al. Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. Mol Ecol 25, 5093-5113 (2016).

BURKHEAD, N. M. Extinction rates in North American freshwater fishes, 1900–2010. BioScience 62, 798-808 (2012).

CILLEROS, Kévin et al. Unlocking biodiversity and conservation studies in high-diversity environments using environmental DNA (eDNA): A test with Guianese freshwater fishes. Molecular ecology resources, v. 19, n. 1, p. 27-46, 2019.

CLARKE, Laurence J. et al. Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. Molecular ecology resources, v. 14, n. 6, p. 1160-1170, 2014.

COLLINS, Rupert A. et al. Non-specific amplification compromises environmental DNA metabarcoding with COI. Methods in Ecology and Evolution, v. 10, n. 11, p. 1985-2001, 2019.

COLLINS, Rupert A. et al. Meta-Fish-Lib: a generalised, dynamic DNA reference library pipeline for metabarcoding of fishes. Journal of Fish Biology, v. 99, n. 4, p. 1446-1454, 2021.

DAL PONT, G., DUARTE RITTER, C., AGOSTINIS, A. O., STICA, P. V., HORODESKY, A., COZER, N., ... & PIE, M. R. Monitoring fish communities through environmental DNA metabarcoding in the fish pass system of the second largest hydropower plant in the world. Scientific Reports, v. 11, n. 1, p. 23167, 2021.

DEINER, Kristy et al. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Mol ecol 26, 5872-5895 (2017).

ELBRECHT, Vasco; LEESE, Florian. Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. Frontiers in Environmental Science, v. 5, p. 11, 2017.

EVANS, Nathan T. et al. Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. Canadian Journal of Fisheries and Aquatic Sciences, v. 74, n. 9, p. 1362-1374, 2017.

FICETOLA, Gentile F. et al. Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Molecular ecology resources, v. 15, n. 3, p. 543-556, 2015.

FICETOLA, Gentile Francesco et al. Species detection using environmental DNA from water samples. Biology letters, v. 4, n. 4, p. 423-425, 2008.

HADDAD, Célio FB et al. From genes to ecosystems: a synthesis of amphibian biodiversity research in Brazil. Biota Neotropica, v. 22, 2022.

HAJIBABAEI, Mehrdad et al. Watered-down biodiversity? A comparison of metabarcoding results from DNA extracted from matched water and bulk tissue biomonitoring samples. PloS one, v. 14, n. 12, p. e0225409, 2019.

HAJIBABAEI, M. et al. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS one 6, e17497 (2011).

HÄNFLING, B. et al. Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. Mol ecol 25, 3101-3119 (2016).

HANSEN, A.J. et al. Crosslinks rather than strand breaks determine access to ancient DNA sequences from frozen sediments. Genetics 173, 1175-1179 (2006).

HAYAMI, Kana et al. Effects of sampling seasons and locations on fish environmental DNA metabarcoding in dam reservoirs. Ecology and Evolution, v. 10, n. 12, p. 5354-5367, 2020.

HE, F. et al. Freshwater megafauna diversity: Patterns, status and threats. Divers. Distrib. 24, 1395– 1404 (2018).

HEBERT, P.D.N et al. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc R Soc Lond B Biol Sci 270, S96-S99 (2003).

HU, Wenjing et al. Comparison of fish communities using environmental DNA metabarcoding and capture methods in a freshwater lake: A new set of universal PCR primers. Fisheries Research, v. 253, p. 106365, 2022.

JACKMAN, Jake M. et al. eDNA in a bottleneck: Obstacles to fish metabarcoding studies in megadiverse freshwater systems. Environmental DNA, v. 3, n. 4, p. 837-849, 2021.

KECK, François et al. Meta-analysis shows both congruence and complementarity of DNA and eDNA metabarcoding to traditional methods for biological community assessment. Molecular Ecology, v. 31, n. 6, p. 1820-1835, 2022.

KELLY, R.P. et al. Using environmental DNA to census marine fishes in a large mesocosm. PloS one 9, e86175 (2014).

Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, et al. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Front Zool. 2013;10. pmid:23496925

LI, Hai et al. Environmental DNA metabarcoding of fish communities in a small hydropower dam reservoir: a comparison between the eDNA approach and established fishing methods. Journal of Freshwater Ecology, v. 37, n. 1, p. 341-362, 2022.

MA, Xiuhui et al. eDNA Metabarcoding Analysis of the Composition and Spatial Patterns of Fish Communities in the Sanbanxi Reservoir, China. Sustainability, v. 14, n. 20, p. 12966, 2022.

MACDONALD, Anna J.; SARRE, Stephen D. A framework for developing and validating taxon-specific primers for specimen identification from environmental DNA. Molecular ecology resources, v. 17, n. 4, p. 708-720, 2017.

MARQUINA, Daniel; ANDERSSON, Anders F.; RONQUIST, Fredrik. New mitochondrial primers for metabarcoding of insects, designed and evaluated using in silico methods. Molecular Ecology Resources, v. 19, n. 1, p. 90-104, 2019.

MATHESON, Carney D. et al. Assessing PCR inhibition from humic substances. The Open Enzyme Inhibition Journal, v. 3, n. 1, p. 38-45, 2010.

MCDEVITT, Allan D. et al. Environmental DNA metabarcoding as an effective and rapid tool for fish monitoring in canals. Journal of fish biology, v. 95, n. 2, p. 679-682, 2019.

MIYA, M.i et al. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. Royal Soc open sci 2, 150088 (2015).

NILSSON, Christer et al. Fragmentation and flow regulation of the world's large river systems. Science, v. 308, n. 5720, p. 405-408, 2005.

OLSON, D. M. et al. Terrestrial Ecoregions of the World: A New Map of Life on EarthA new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. BioScience, v. 51, n. 11, p. 933-938, 2001.

PELICICE, Fernando M. et al. Human impacts and the loss of Neotropical freshwater fish diversity. Neotropical Ichthyology, v. 19, 2021.

PILLIOD, David S. et al. Factors influencing detection of eDNA from a stream-dwelling amphibian. Molecular ecology resources, v. 14, n. 1, p. 109-116, 2014.

PUGEDO, M.L. et al. Integrative taxonomy supports new candidate fish species in a poorly studied neotropical region: the Jequitinhonha River Basin. Genetica 144, 341-349 (2016).

ROSA, R.S. & LIMA, F.C.T. Os peixes brasileiros ameaçados de extinção. Livro vermelho da fauna brasileira ameaçada de extinção 2, 1-278 (2008).

SALES, Naiara G. et al. Hidden diversity hampers conservation efforts in a highly impacted Neotropical river system. Frontiers in genetics, v. 9, p. 271, 2018.

SALES, Naiara G. et al. Influence of preservation methods, sample medium and sampling time on eDNA recovery in a neotropical river. Environmental DNA, v. 1, n. 2, p. 119-130, 2019.

SALES, Naiara Guimarães et al. Introgression from non-native species unveils a hidden threat to the migratory Neotropical fish Prochilodus hartii. Biological Invasions, v. 20, n. 3, p. 555-566, 2018.

SALES, Naiara Guimarães et al. Space-time dynamics in monitoring neotropical fish communities using eDNA metabarcoding. Science of the Total Environment, v. 754, p. 142096, 2021.

SCHENEKAR, Tamara et al. Reference databases, primer choice, and assay sensitivity for environmental metabarcoding: Lessons learnt from a re-evaluation of an eDNA fish assessment in the Volga headwaters. River Research and Applications, v. 36, n. 7, p. 1004-1013, 2020.

Sharma P, Kobayashi T. Are “universal” DNA primers really universal? J Appl Genet. 2014;55(4):485–96. pmid:24839163

SPALDING, Mark D. et al. Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. BioScience, v. 57, n. 7, p. 573-583, 2007.

TABERLET, P.et al. Environmental DNA. Mol ecol 21, 1789-1793 (2012).

TAKEUCHI, A., SADO, T., GOTOH, R. O., WATANABE, S., TSUKAMOTO, K., & MIYA, M. (2019). New PCR primers for metabarcoding environmental DNA from freshwater eels, genus Anguilla. Scientific Reports, 9(1), 1-11.

THOMSEN, P. et al. Monitoring endangered freshwater biodiversity using environmental DNA. Mol ecol 21, 2565-2573 (2012a).

THOMSEN, P. F. et al. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. PLoS one 7, e41732 (2012b).

THOMSEN, P.F. & WILLERSLEV, E. Environmental DNA-An emerging tool in conservation for monitoring past and present biodiversity. Biological Conservation 183, 4-18 (2015).

VALDEZ-MORENO, Martha et al. Using eDNA to biomonitor the fish community in a tropical oligotrophic lake. PLoS One, v. 14, n. 4, p. e0215505, 2019.

VALENTINI, A. et al. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Mol Ecol 25, 929-942 (2016).

WEIGAND, Hannah et al. DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. Science of the Total Environment, v. 678, p. 499-524, 2019.

WILLERSLEV, E. et al. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. Science 300, 791-795 (2003).

XINGYUAN, Zhang; FAWEN, Li; YONG, Zhao. Impact of changes in river network structure on hydrological connectivity of watersheds. Ecological Indicators, v. 146, p. 109848, 2023.

XU, Longfei et al. Environmental DNA Captures Variations in Fish Assemblages with Distance from Dams in Karst Reservoirs. Water, v. 15, n. 1, p. 73, 2022.

ZEPEDA MENDOZA, Marie Lisandra; SICHERITZ-PONTEN, Thomas; GILBERT, M. Thomas P. Environmental genes and genomes: understanding the differences

and challenges in the approaches and software for their analyses. Briefings in Bioinformatics, v. 16, n. 5, p. 745-758, 2015.

ZINGER, Lucie et al. Advances and prospects of environmental DNA in neotropical rainforests. Advances in ecological research, v. 62, p. 331-373, 2020.