

# DNA metabarcoding of Neotropical ichthyoplankton: Enabling high accuracy with lower cost

André B. Nobile<sup>1</sup>, Diogo Freitas-Souza<sup>1</sup>, Francisco J. Ruiz-Ruano<sup>2,3</sup>, Maria Lígia M. O. Nobile<sup>1</sup>, Gabriela O. Costa<sup>1</sup>, Felipe P. de Lima<sup>1</sup>, Juan Pedro M. Camacho<sup>2</sup>, Fausto Foresti<sup>1</sup>, Claudio Oliveira<sup>1</sup>

<sup>1</sup> Laboratório de Biologia e Genética de Peixes, Departamento de Morfologia, Unesp, Botucatu, Brazil.

<sup>2</sup> Evolutionary Genetics group, Department of Genetics, University of Granada, Granada, Spain.

<sup>3</sup> Present address: Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden.

Corresponding author: André B. Nobile (andrenobile@hotmail.com)

Academic editor: Filipe Costa | Received 8 April 2019 | Accepted 29 July 2019 | Published 5 September 2019

## Abstract

Knowledge of ichthyoplankton dynamics is extremely important for conservation management as it can provide information about preferential spawning sites, reproductive period, migratory routes and recruitment success, which can be used to guide management and conservation efforts. However, identification of the eggs and larvae of Neotropical freshwater fish is a difficult task. DNA barcodes have emerged as an alternative and highly accurate approach for species identification, but DNA barcoding can be time-consuming and costly. To solve this problem, we aimed to develop a simple protocol based on DNA metabarcoding, to investigate whether it is possible to detect and quantify all species present in a pool of organisms. To do this, 230 larvae were cut in half, one half was sequenced by the Sanger technique and the other half was used to compose six arrays with a pool of larvae that were sequenced using a next-generation technique (NGS). The results of the Sanger sequencing allowed the identification of almost all larvae at species level, and the results from NGS showed high accuracy in species detection, ranging from 83% to 100%, with an average of 95% in all samples. No false positives were detected. The frequency of organisms in the two methods was positively correlated (Pearson), with low variation among species. In conclusion, this protocol represents a considerable advance in ichthyoplankton studies, allowing a rapid, cost-effective, quali-quantitative approach that improves the accuracy of identification.

## Key Words

DNA Barcoding, Fisheries management, Ichthyofauna, Upper Paraná River

## Introduction

Studies concerning ichthyoplankton can provide novel information about preferential spawning sites, reproductive period, migratory routes, and recruitment success (Baumgartner et al. 2004, Bialetzki et al. 2005, Reynalte-Tataje et al. 2012). These are extremely important, as they provide useful knowledge for conservation management, such as the delimitation of protected areas (de Silva et al. 2015; Silva et al. 2017). However, morphological identification of eggs and larvae have several limitations, because, in the early stages, the morphological diagnos-

tic characters are not yet fully formed (Baumgartner et al. 2004, Reynalte-Tataje et al. 2012). Such limitations restrict efforts to gather information about the life history strategies of most species, including endangered species, thus hampering proposals for targeted management actions.

The use of DNA barcodes in ichthyoplankton improved the quality of identifications, both for eggs and larvae, and emerged as an effective alternative for traditional identification methods (Becker et al. 2015, Frantini-Silva et al. 2015). Nevertheless, despite its high accuracy, when applied in extensive studies with thousands of eggs and larvae, DNA barcoding can be time-consum-

ing and costly, because each organism has to be individually processed and sequenced (Evans et al. 2016, Elbrecht and Leese 2017, Mariac et al. 2018).

To solve this problem, DNA metabarcoding, based on next generation sequencing (NGS) could offer an alternative approach, allowing a reduction of labor time as well as costs. Although Neotropical ichthyofauna metabarcoding studies are scarce (Maggia et al. 2017, Mariac et al. 2018, Sales et al. 2019), this technique has been utilized in several ecological studies on ichthyoplankton and also to identify gut contents (Cruaud et al. 2017, Guillerault et al. 2017, Maggia et al. 2017, Divoll et al. 2018, Barbato et al. 2019).

Here we describe a novel DNA metabarcoding protocol for ichthyoplankton using NGS technology and test the accuracy of this protocol in identifying all species present in an array. We also test whether the proposed approach is valid for estimating the relative frequency of species present in a sample.

## Material and methods

### Field sampling

The larvae used in this experiment come from another study using DNA barcoding for species identification. Ichthyoplankton sampling was carried out in the Mogi-Guaçu river, an important tributary of the left bank of the Grande River, Upper Paraná River Basin. Samples were taken from two sites: 1: 21°35.441'S, 47°57.244'W (Dec/16 and Jan/17) and site 2: 21°30.168'S, 48°2.534'W (Dec/15, Nov/16, Dec/16, and Jan/17) (Table 1), located in the northeast region of the state of São Paulo, Brazil. The sites are about 15 km apart from each other in a straight line.

Collecting took place with a conical-cylindrical net with a mesh size of 0.5 mm. Samples were immediately fixed in 96% ethanol. In the laboratory, eggs and larvae were separated from other materials present in the samples (e.g., leaves, sediment, and sticks) under a stereomicroscope and kept in 96% ethanol until the molecular procedure.

### Sampling and experiment setup

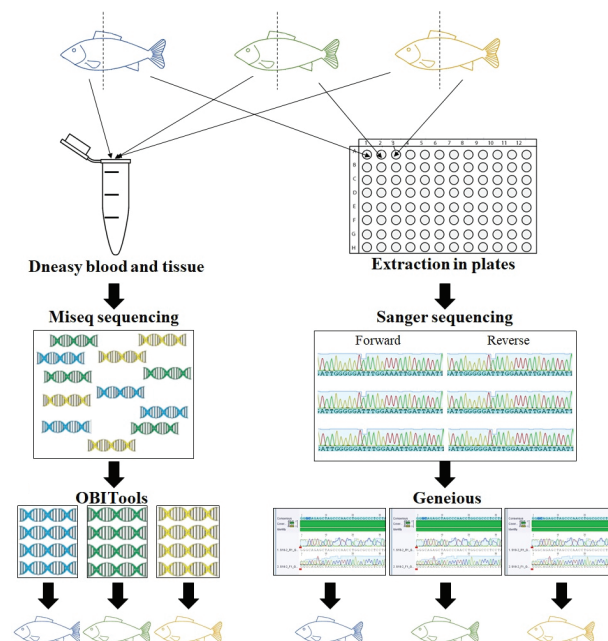
In this study, we used 230 larvae distributed randomly across six arrays, with the number of organisms per array ranging from 36 to 44. However, each sample contained larvae of only one site and month (Table 1). Each larva was cut into halves; one half was subject to gDNA extraction and subsequent Sanger sequencing, while the other half was pooled and homogenized with other larvae from the same sample for combined DNA extraction and subsequent metabarcoding using next-generation sequencing (NGS) (Figure 1).

### Sanger sequencing

DNA extraction for Sanger sequencing was made following the protocol proposed by Ivanova et al. (2006). Cy-

**Table 1.** Number of larvae in each sample and their respective sites and month of capture.

Sample number	Number of larvae	Site	Month
S1	36	1	Jan/17
S2	42	1	Dec/16
S3	26	2	Dec/16
S4	38	2	Jan/17
S5	44	2	Nov/16
S6	44	2	Dec/15
<b>Total</b>	<b>230</b>		



**Figure 1.** Experimental setup of this study for Sanger and metabarcoding analysis of ichthyoplankton.

tochrome C Oxidase Subunit I (COI) primers FishF1 and FishR1 (Ward et al. 2005) were selected because these have a good species recovery in the hydrographic basin where the study occurred. PCR conditions were: 1 min at 96 °C followed by 35 cycles of 10 s at 96 °C, 5 s at 50 °C and 4 min at 60 °C. The PCR product was sequenced using a Big Dye 3.1 Terminator kit (Applied Biosystems) in an ABI Prism 3130 (Applied Biosystems), generating two complementary sequences of 654 bp.

### DNA metabarcoding

DNeasy Blood and Tissue Kit (Qiagen) was used for DNA extraction from bulk samples following the manufacturer's protocol. Two PCR runs were performed: the first to amplify the COI region, using the same primer pair used for the Sanger sequencing coupled with Illumina MiSeq adapter, and the second to attach a multiplex identifier (MID).

In the first PCR, each reaction contained, 2 µl of DNA template, 2.5 µl buffer, 1 µl MgCl<sub>2</sub> (50mM), 1 µl dNTP (2 mM), 0.5 µl forward adapter+FishF1 primer (5 mM), 0.5 µL reverse adapter+FishR1 primer (5 mM), 0.4 µl Taq

Polymerase (5 U/μl) (Phonutria), and 17.1 μl of molecular biology grade water in a final volume of 25 μl. PCR conditions were 95 °C for 3 min, 35 cycles of 95 °C for 30 s, 52 °C for 45 s, and 68 °C for 60 s; and a final extension at 68 °C for 10 min. We purified PCR products through NucleoSpin Gel and PCR Clean-up from Macherey-Nagel. In the second PCR, the library construction was made, where amplicons were dual indexed with MID's by Macrogen, using the Nextera XT Index Kit and following the instructions present in 16S Metagenomic Sequencing Preparation. Amplicons were sequenced on the Illumina MiSeq platform, generating paired-end reads of 301 nucleotides in length.

## Data analyses

For sequences generated with the Sanger method we used the Geneious Pro v4.8.5 software (Kearse et al. 2012), where forward and reversed files were combined to assemble a consensus sequence. All these consensus sequences were then exported as fasta files to be matched against the same database comprising the Illumina reads.

Metabarcoding data were analyzed following a bioinformatic protocol described in the Suppl. material 1. As the overlapping is not possible due to the length of PCR product (654 bp), only the forward file was used (Cruaud et al. 2017). This file was selected by checking the sequencing quality with FASTQC (Andrews 2010) (Suppl. material 2). Next, a quality trimming with Trimmomatic (Bolger et al. 2014) was performed, and we filtered out reads with a minimum quality of Q20 in at least the first 250 nt. We also cut the sequences up to the length of 250 nucleotides since the quality of final region was low (Suppl. material 2). For this trimming we used the Fastx-Trimmer software ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)).

After this step, we used the OBITools toolkit version 1.01 22 (<http://metabarcoding.org/obitools>; Boyer et al. 2016) for DNA metabarcoding analysis. We first used the “obiuniq” command to dereplicate the sequences, to reduce the size file and computation time. In this step, all of the identical DNA sequences were grouped into a single sequence, scoring their abundance for comparison with the Sanger data. Subsequently, we used the “obigrep” command to remove sequences with less than 50 reads.

We constructed a custom database by downloading sequences from Project – FUPR Fishes from Upper Parana River, Brazil, present in the BOLD System (Ratnasingham and Hebert 2007), since this project was executed in the same river basin as the current study. Also, we downloaded sequences from RefSeq (<https://www.ncbi.nlm.nih.gov/refseq/>), using keywords “fish”, “fishes”, and “COI”, and then filtered the results only for “bony fishes”. For the first step of identification we used the BLASTN software (Altschul et al. 1990) to indicate which sequences had the greatest similarity to those obtained via Sanger and metabarcoding.

As a final filtering step, all sequences generated in the present study (NGS + Sanger) were aligned with those

sequences present in a custom database that had similarity greater than 95% with one of the sequencing methods, using the Muscle (Edgar 2004) aligner implemented in Geneious 4.8.5 (Kearse et al. 2012). Next, we calculated the genetic distances among and within the species, using the Kimura-2-Parameter (K2P) distance model (Kimura 1980). To provide a graphic representation of the patterning of divergence among species, a neighbor-joining (NJ) dendrogram of K2P distances was created using the MEGA v 7.0 software (Kumar et al. 2016) for each sample. Those sequences (NGS and Sanger) that presented divergence greater than 2% in the K2P analysis compared to the sequences downloaded from the BOLD System database were removed from the analysis of abundance (Pereira et al. 2013).

To validate the efficacy of metabarcoding, statistical analyses were conducted to evaluate whether this method has a quali-quantitative resolution. To achieve our qualitative aim, from the genetic distance (K2P) and neighbor-joining (NJ) dendrogram, we estimated richness ( $S$  = number of species) per sample in each method. Then, the richness between methods was compared to estimate the rate of species detection by metabarcoding, using the Sanger data as a reference.

To achieve our quantitative analysis, as the exact number of larvae per sample was known, the relative frequency per species (or taxon) was estimated for the Sanger data. This frequency compared with the value of the relative frequency (number of reads assigned to a taxon or species in relation to the total) obtained metabarcoding. This comparison was made per sample individually (e.g., Sanger sample 1 × NGS 1 sample). After this, we applied a Pearson correlation to test if the frequencies of Sanger and NGS showed a significant positive correlation and a permutational analysis of variance (PERMANOVA) to evaluate significant differences between relative abundances of Sanger and NGS data.

## Results

### DNA sequencing

Of the 230 larvae used in this experiment (Suppl. material deposited in: <https://doi.org/10.6084/m9.figshare.6726956>) 226 were retained for the analysis with Sanger sequencing. Four larvae were removed, due to low quality in Sanger sequencing and genetic divergence above 2% in the alignment with those downloaded sequences. Results from the Sanger sequencing showed a richness of 29 taxa, belonging to 12 families and three orders, varying from six species in sample S2 to 13 species in sample S1 (Table 2). Only three taxa were assigned to family or subfamily level.

NGS yielded 2,511,656 reads in the eight samples (Suppl. material deposited in: <https://doi.org/10.6084/m9.figshare.6726956>) and, after the bioinformatic analysis, 633,224 reads were kept. In total, 28 taxa were identified

**Table 2.** Species composition and relative abundance of analyzed samples. S= Sanger sequences; NGS= Next generation sequencing metabarcoding; \* false negative.

Family	Species	S1		S2		S3		S4		S5		S6	
		S	NGS	S	NGS	S	NGS	S	NGS	S	NGS	S	NGS
Characiformes													
Anostomidae	<i>Leporellus vittatus</i>							5.6	10.4				
	<i>Leporinus friderici</i>	13.9	23.2					2.8	1.5	9.3	11.6	2.3	1.2
	<i>Leporinus paranensis</i>											2.3	3.7
	<i>Leporinus striatus</i>	5.6	2.5										
	<i>Megaleporinus obtusidens</i>			7.3	*	3.8	3.3			9.3	0.8	2.3	0.1
	<i>Megaleporinus piavussu</i>							2.8	0.3				
	<i>Schizodon nasutus</i>			2.4	2.1								
Characidae	<i>Astyanax lacustris</i>	11.1	0.1					5.6	0.3	4.7	*		
	<i>Astyanax schubarti</i>	2.8	*					5.6	*				
	Cheirodontinae	11.1	4.3	2.4	0.05	3.8	0.6	2.8	2.0	4.7	1.6		
	Cheirodontinae 1					3.8	6.2						
	<i>Hyphessobrycon eques</i>							5.6	2.0				
	<i>Steindachnerina insculpta</i>	2.8	0.6										
Erythrinidae	<i>Hoplias malabaricus</i>									2.3	2.2		
Parodontidae	<i>Apareiodon affinis</i>			2.4	0.1								
Prochilodontidae	<i>Prochilodus lineatus</i>							2.8	15.2	4.7	9.2		
Gymnotiformes													
Sternopigidae	<i>Eigenmannia trilineata</i>											2.3	5.0
Siluriformes													
Doradidae	<i>Rhinodoras dorbignyi</i>	2.8	1.4										
Heptapteridae	Heptapteridae	5.6	2.2							2.3	1.9		
	<i>Iheringichthys labrosus</i>											2.3	4.7
	<i>Pimelodella meeki</i>	5.6	2.4					8.3	2.1				
Loricariidae	<i>Rhinelepis aspera</i>	2.8	16.9										
Pimelodidae	<i>Pimelodus maculatus</i>	27.8	36.0	70.7	95.0	69.2	58.1	52.8	61.3	32.6	49.0	86.4	85.2
	<i>Pimelodus microstoma</i>	2.8	6.3			11.5	21.1						
	<i>Pinirampus pirinampu</i>					3.8	3.9						
	<i>Pseudoplatystoma corruscans</i>			14.6	2.7	3.8	6.9			16.3	8.6	2.3	0.1
	<i>Pseudoplatystoma reticulatum</i>									4.7	1.3		
	<i>Sorubim lima</i>							2.8	3.0	9.3	13.8		
Pseudopimelodidae	<i>Pseudopimelodus mangurus</i>	5.6	4.2					2.8	1.8				

with metabarcoding NGS out of the 29 taxa identified by the Sanger sequencing. Only *Astyanax schubarti* was not identified by metabarcoding in any of the samples (Table 2).

### Species detection (qualitative approach)

No false positives were observed in this study (Table 2); all the taxa retrieved by metabarcoding were also found in Sanger sequencing data. However, in four out of the six samples, one taxon was not detected by NGS: S1 and S4 – *Astyanax schubarti*, S2 – *Megaleporinus obtusidens*, S5 – *Astyanax lacustris*. In the other two samples, all taxa detected by Sanger sequencing were also detected by NGS (S3 and S6). The taxa detection rate was higher than 83% in all samples, with 100% of detection in four samples and higher than 90% in three others (mean, 95%).

### Relative abundance (quantitative approach)

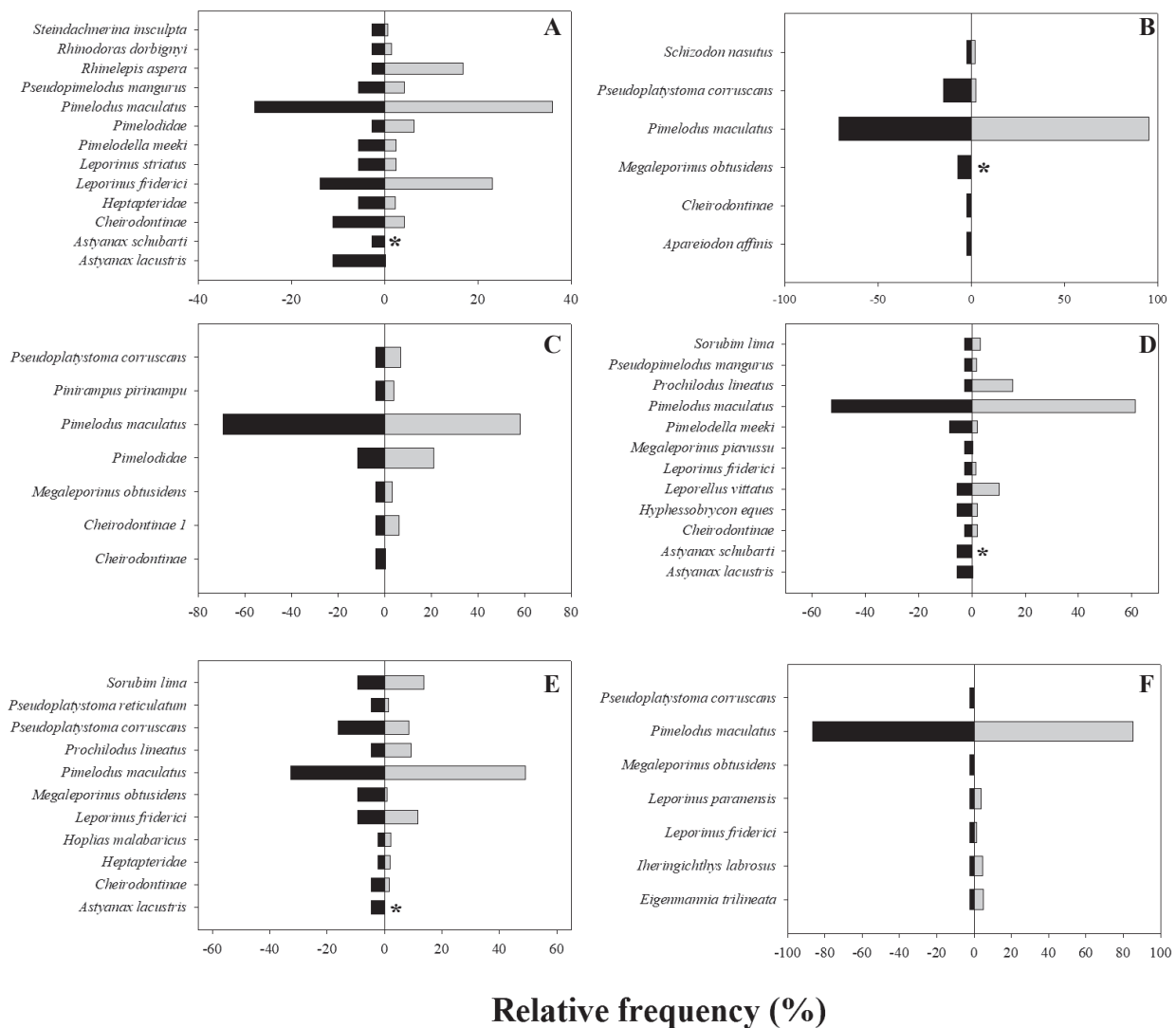
The correlation between Sanger and NGS methods showed few differences in all samples. We detected a considerable difference between the two methods in only some cases: *Rhinelepis aspera*, Cheirodontinae sp.1, and

*Astyanax lacustris*, in sample 1 and *Prochilodus lineatus* and *Astyanax lacustris* in sample 4. Nevertheless, the samples showed non-significant differences between the relative abundances of the two methods in PERMANOVA analysis (Pseudo-F = 0.39293; P(perm) = 0.838), denoting that, despite some deviations in the values, the quantification of organisms with the NGS method can be considered reliable here (Table 3; Figure 2). Also, in the same way, all samples showed higher values of Pearson's correlation in relative frequency between Sanger and NGS methods, varying from 0.776 (sample S1) to 0.998 (sample S6) (Table 3; Figure 2).

**Table 3.** Species richness per sample and method of sequencing, detection rate, and Pearson correlation (r) between Sanger and NGS methods.

Sample	Richness		Detection rate NGS/Sanger	Pearson correlation
	Sanger	NGS		
S1	13	12	92.3	0.776
S2	6	5	83.3	0.987
S3	7	7	100.0	0.976
S4	12	11	91.7	0.955
S5	11	10	90.9	0.914
S6	7	7	100.0	0.998





**Figure 2.** Graphic representation of species abundance proportion (%) between Sanger (x-axis, left side) and NGS (x-axis, right side) for analyzed samples. A: sample 1, B: sample 2, C: sample 3, D: sample 4, E: sample 5, F: sample 6. \* = false negative.

## Discussion

### Species identification

Despite some failures in species detection, we can consider that the use of NGS is functional for COI-based ichthyoplankton species detection, with an average detection rate higher than 95%, which is similar to previous findings that compared Sanger and NGS (Cruaud et al. 2017). Also, when compared to morphology-based methods of egg and larvae identification, detection rates were even more satisfactory, as more than 50% of the larvae caught in natural populations cannot be identified due to the absence of diagnostic characters (newly hatched larvae) or damage due to fixation process (Bialecki et al. 2005; da Silva et al. 2015; Silva et al. 2017).

With our present approach, most individuals were identified to the species level, demonstrating that even us-

ing only the forward reads, the information was sufficient for species identification. Only three taxa were assigned to family or subfamily level only, probably because they are cryptic species, not described or with sequences not deposited in any database. This observation reinforces the idea that more sequences of fish species must be deposited in databases, especially in river basins with rich fish fauna as the Upper Paraná River. Another positive aspect is that, compared with DNA barcoding (Arroyave and Stiassny 2014, Becker et al. 2015, Frantini-Silva et al. 2015), our method can trigger significant advances in the quality of ecological studies in fish, as observed in some recent studies with metabarcoding (Guilleraut et al. 2017, Maggia et al. 2017, Mariac et al. 2018, Barbato et al. 2019) and, mainly, to ensure significant reduction in labor time and cost per specimen reaching 80% of reduction (Shokralla et al. 2015).

Despite some differences in obtaining DNA samples, both environmental DNA (eDNA) and DNA metabarcoding from bulk samples are subject to false negatives and false positives (Ficetola Gentile et al. 2015, 2016; Divoll et al. 2018). However, in the present study, only false negatives were observed. False negatives can lead to losses in mitigating measures of threatened species, due to the non-detection of target species, which might lead to underestimation of populations and non-implementation of mitigation measures, causing injury to these species (Ficetola Gentile et al. 2015, 2016; Divoll et al. 2018). False negatives are mainly those related to inhibition issues, primer bias, PCR or sequencing errors, and bioinformatics analysis (Ficetola Gentile et al. 2015, 2016; Divoll et al. 2018).

### Relative abundance

Quali- and quantitative studies involving metabarcoding and eDNA have increased significantly in recent years (Hänfling et al. 2016; Valentini et al. 2016; Evans and Lamberti 2017; Yamamoto et al. 2017; Nakagawa et al. 2018; Mariac et al. 2018). However, the reliable quantification of organisms remains a problem (Evans et al. 2016; Kelly et al. 2017; Hansen et al. 2018; Mariac et al. 2018). eDNA studies have shown a positive correlation between species-specific DNA concentrations in biomass and abundance in eDNA (Takahara et al. 2012; Maruyama et al. 2014; Doi et al. 2015; Evans et al. 2016). In this way, by focusing on developing a low-cost yet efficient protocol, here we show that PCR-based methods can be used to determine relative abundance in studies of eggs and larvae.

## Conclusions

Our results clearly show that the NGS approach on bulk samples can detect ichthyoplankton diversity at the species level, and also gives good estimates of relative abundance of the larvae, thus allowing reliable environmental monitoring at reduced cost and labor (Hansen et al. 2018). These results are extremely important since a great part of ichthyoplankton remains undefined with the current technology, impairing management actions. By applying metabarcoding protocols, the identification of all sampled organisms is possible, even those that occur in low density, such as endangered species. Considering ecological aspects, the use of metabarcoding in ichthyoplankton studies to identify fish larvae and eggs can provide several advances, such as precise determination of spawning sites of migratory and/or threatened species, and early detection of invasive species, which can help and guide management actions (Almeida et al. 2018).

## Acknowledgements

This research received financial support from FAPESP grants 2015/19025-9, 2017/12758-6 (ABN),

CNPq 165830/2015-8 (ABN), 150344/2016 (FPL), 141526/2015-7 (DFS), 144265/2017-6 (GOC), and Capes 180517/2018-01 (MLMON). We wish to thank Centro Nacional de Pesquisa e Conservação de Peixes Continentais – CEPTA, Pirassununga-SP, José Augusto Senhorini and Claudio Bock for help in the field and Jorge Doña by advised about the bioinformatic analysis.

## References

- Almeida FS, Frantini-Silva W, Lima SC, Garcia DAZ, Orsi ML (2018) DNA barcoding as a useful tool for identifying non-native species of freshwater ichthyoplankton in the neotropics. *Hydrobiologia* 817: 111–119. <https://doi.org/10.1007/s10750-017-3443-5>
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Arroyave J, Stiasny MLJ (2014) DNA barcoding reveals novel insights into pterygophagy and prey selection in distichodontid fishes (Characiformes: Distichodontidae). *Ecology and Evolution* 4: 4534–4542. <https://doi.org/10.1002/ece3.1321>
- Barbato M, Kovacs T, Coleman MA, Broadhurst MK, de Bruyn M (2019) Metabarcoding for stomach-content analyses of Pygmy devil ray (*Mobula kuhlii* cf. *eregoodootenkee*): Comparing tissue and ethanol preservative-derived DNA. *Ecology and Evolution*. <https://doi.org/10.1002/ece3.4934>
- Baumgartner G, Nakatani K, Gomes L, Bialezki A, Sanches P (2004) Identification of spawning sites and natural nurseries of fishes in the upper Paraná River, Brazil. *Environmental Biology of Fishes* 71: 115–125. <https://doi.org/10.1007/s10641-004-0098-z>
- Becker RA, Sales NG, Santos GM, Santos GB, Carvalho DC (2015) DNA barcoding and morphological identification of neotropical ichthyoplankton from the Upper Paraná and São Francisco. *Journal of Fish Biology* 87: 159–168. <https://doi.org/10.1111/jfb.12707>
- Bialezki A, Nakatani K, Sanches PV, Baumgartner G, Gomes LC (2005) Larval fish assemblage in the Baía River (Mato Grosso do Sul State, Brazil): temporal and spatial patterns. *Environmental Biology of Fishes* 73: 37–47. <https://doi.org/10.1007/s10641-004-3795-3>
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- 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: 176–182. <https://doi.org/10.1111/1755-0998.12428>
- Cruaud P, Rasplus J-Y, Rodríguez LJ, Cruaud A (2017) High-throughput sequencing of multiple amplicons for barcoding and integrative taxonomy. *Scientific Reports* 7: 41948. <https://doi.org/10.1038/srep41948>
- Divoll TJ, Brown VA, Kinne J, McCracken GF, O’Keefe JM (2018) Disparities in second-generation DNA metabarcoding results exposed with accessible and repeatable workflows. *Molecular Ecology Resources*. <https://doi.org/10.1111/1755-0998.12770>
- Doi H, Uchii K, Takahara T, Matsushashi S, Yamanaka H, Minamoto T (2015) Use of droplet digital PCR for estimation of fish abundance

- and biomass in environmental DNA surveys. *PLoS ONE* 10(3): e0122763. <https://doi.org/10.1371/journal.pone.0122763>
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Elbrecht V, Leese F (2017) Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. *Frontiers in Environmental Science* 5. <https://doi.org/10.3389/fenvs.2017.00011>
- Evans NT, Lamberti GA (2017) Freshwater fisheries assessment using environmental DNA: A primer on the method, its potential, and shortcomings as a conservation tool. *Fisheries Research*. <https://doi.org/10.1016/j.fishres.2017.09.013>
- Evans NT, Olds BP, Renshaw MA, Turner CR, Li Y, Jerde CL, Mahon AR, Pfrender ME, Lamberti GA, Lodge DM (2016) Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Molecular Ecology Resources* 16: 29–41. <https://doi.org/10.1111/1755-0998.12433>
- Ficetola Gentile F, Pansu J, Bonin A, Coissac E, Giguët-Covex C, De Barba M, Gielly L, Lopes CM, Boyer F, Pompanon F, Rayé G, Taberlet P (2015) Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources* 15: 543–556. <https://doi.org/10.1111/1755-0998.12338>
- Ficetola Gentile F, Taberlet P, Coissac E (2016) How to limit false positives in environmental DNA and metabarcoding? *Molecular Ecology Resources* 16: 604–607. <https://doi.org/10.1111/1755-0998.12508>
- Frantini-Silva W, Sofia SH, Orsi ML, Almeida FS (2015) DNA barcoding of freshwater ichthyoplankton in the Neotropics as a tool for ecological monitoring. *Molecular Ecology Resources* 15: 1226–1237. <https://doi.org/10.1111/1755-0998.12385>
- Guillerault N, Bouletreau S, Iribar A, Valentini A, Santoul F (2017) Application of DNA metabarcoding on faeces to identify European catfish *Silurus glanis* diet. *Journal of Fish Biology* 90: 2214–2219. <https://doi.org/10.1111/jfb.13294>
- Hänfling B, Lawson Handley L, Read DS, Hahn C, Li J, Nichols P, Blackman RC, Oliver A, Winfield IJ (2016) Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Molecular Ecology* 25: 3101–3119. <https://doi.org/10.1111/mec.13660>
- Hansen BK, Bekkevold D, Clausen LW, Nielsen EE (2018) The sceptical optimist: challenges and perspectives for the application of environmental DNA in marine fisheries. *Fish and Fisheries*. <https://doi.org/10.1111/faf.12286>
- Ivanova NV, Dewaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes* 6: 998–1002. <https://doi.org/10.1111/j.1471-8286.2006.01428.x>
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Kelly RP, Closek CJ, O'Donnell JL, Kralj JE, Shelton AO, Samhoury JF (2017) Genetic and manual survey methods yield different and complementary views of an ecosystem. *Frontiers in Marine Science* 3. <https://doi.org/10.3389/fmars.2016.00283>
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16: 111–120. <https://doi.org/10.1007/BF01731581>
- Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular evolutionary genetics analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* 33: 1870–1874. <https://doi.org/10.1093/molbev/msw054>
- Maggia ME, Vigouroux Y, Renno JF, Duponchelle F, Desmarais E, Nunez J, García-Dávila C, Carvajal-Vallejos FM, Paradis E, Martin JF, Mariac C (2017) DNA metabarcoding of Amazonian ichthyoplankton swarms. *PLoS ONE* 12(1): e0170009. <https://doi.org/10.1371/journal.pone.0170009>
- Mariac C, Vigouroux Y, Duponchelle F, García-Dávila C, Nunez J, Desmarais E, Renno JF (2018) Metabarcoding by capture using a single COI probe (MCSP) to identify and quantify fish species in ichthyoplankton swarms. *PLoS ONE* 13(9): e0202976. <https://doi.org/10.1371/journal.pone.0202976>
- Maruyama A, Nakamura K, Yamanaka H, Kondoh M, Minamoto T (2014) The release rate of environmental DNA from juvenile and adult fish. *PLoS ONE* 9(12): e114639. <https://doi.org/10.1371/journal.pone.0114639>
- Nakagawa H, Yamamoto S, Sato Y, Sado T, Minamoto T, Miya M (2018) Comparing local- and regional-scale estimations of the diversity of stream fish using eDNA metabarcoding and conventional observation methods. *Freshwater Biology*. <https://doi.org/10.1111/fwb.13094>
- Pereira LH, Hanner R, Foresti F, Oliveira C (2013) Can DNA barcoding accurately discriminate megadiverse Neotropical freshwater fish fauna? *BMC Genetics* 14: 20. <https://doi.org/10.1186/1471-2156-14-20>
- Ratnasingham S, Hebert PDN (2007) BARCODING: bold: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Molecular Ecology Notes* 7: 355–364. <https://doi.org/10.1111/j.1471-8286.2007.01678.x>
- Reynalte-Tataje DA, Agostinho AA, Bialetzki A, Hermes-Silva S, Fernandes R, Zaniboni-Filho E (2012) Spatial and temporal variation of the ichthyoplankton in a subtropical river in Brazil. *Environmental Biology of Fishes* 94: 403–419. <https://doi.org/10.1007/s10641-011-9955-3>
- Sales NG, Wangenstein OS, Carvalho DC, Mariani S (2019) Influence of preservation methods, sample medium and sampling time on eDNA recovery in a neotropical river. *Environmental DNA: edn3.14*. <https://doi.org/10.1002/edn3.14>
- Shokralla S, Porter TM, Gibson JF, Dobosz R, Janzen DH, Hallwachs W, Golding GB, Hajibabaei M (2015) Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Scientific Reports* 5: 9687. <https://doi.org/10.1038/srep09687>
- Silva CB, Dias JD, Bialetzki A (2017) Fish larvae diversity in a conservation area of a neotropical floodplain: influence of temporal and spatial scales. *Hydrobiologia* 787: 141–152. <https://doi.org/10.1007/s10750-016-2953-x>
- Silva JC, Rosa RR, Galdioli EM, Soares CM, Domingues WM, Verissimo S, Bialetzki A (2017) Importance of dam-free stretches for fish reproduction: the last remnant in the Upper Paraná River. *Acta Limnologica Brasiliensia* 29. <https://doi.org/10.1590/s2179-975x10216>
- da Silva PS, Makrakis MC, Miranda LE, Makrakis S, Assumpção L, Paula S, Dias JHP, Marques H (2015) Importance of reservoir tributaries to spawning of migratory fish in the Upper Paraná River. *River Research and Applications* 31: 313–322. <https://doi.org/10.1002/rra.2755>
- Takahara T, Minamoto T, Yamanaka H, Doi H, Kawabata Z (2012) Estimation of fish biomass using environmental DNA. *PLoS ONE* 7: e35868. <https://doi.org/10.1371/journal.pone.0035868>

- Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, Bellemain E, Besnard A, Coissac E, Boyer F, Gaboriaud C, Jean P, Poulet N, Roset N, Copp GH, Geniez P, Pont D, Argillier C, Baudoin J-M, Peroux T, Crivelli AJ, Olivier A, Acqueberge M, Le Brun M, Moller PR, Willerslev E, Dejean T (2016) Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology* 25: 929–942. <https://doi.org/10.1111/mec.13428>
- Ward RD, Zemplak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360: 1847–1857. <https://doi.org/10.1098/rstb.2005.1716>
- Yamamoto S, Masuda R, Sato Y, Sado T, Araki H, Kondoh M, Minamoto T, Miya M (2017) Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. *Scientific Reports* 7: 40368. <https://doi.org/10.1038/srep40368>

### Supplementary material 1

#### Step-by-step bioinformatic analysis, using sample S1 as model

Authors: André B. Nobile, Diogo Freitas-Souza, Francisco J. Ruiz-Ruano, Maria Lígia Marques de Oliveira Nobile, Gabriela Omura Costa, Felipe Pontieri de Lima, Juan Pedro M. Camacho, Fausto Foresti, Claudio Oliveira

Data type: Bioinformatic protocol

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited. Link: <https://doi.org/10.3897/mbmg.3.35060.suppl1>

### Supplementary material 2

#### NGS quality checking by FASTQC

Authors: André B. Nobile, Diogo Freitas-Souza, Francisco J. Ruiz-Ruano, Maria Lígia Marques de Oliveira Nobile, Gabriela Omura Costa, Felipe Pontieri de Lima, Juan Pedro M. Camacho, Fausto Foresti, Claudio Oliveira

Data type: NGS quality control

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited. Link: <https://doi.org/10.3897/mbmg.3.35060.suppl2>