MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2015) **15**, 1226–1237 doi: 10.1111/1755-0998.12385

DNA barcoding of freshwater ichthyoplankton in the Neotropics as a tool for ecological monitoring

W. FRANTINE-SILVA,* S. H. SOFIA,* M. L. ORSI† and F. S. ALMEIDA*
*Departamento de Biologia Geral, Universidade Estadual de Londrina, Centro de Ciências Biológicas, Campus Universitário, Rodovia Celso Garcia Cid, PR 445 km 380, 86057-970, Londrina-PR, Brasil, †Departamento de Biologia Animal e Vegetal, Universidade Estadual de Londrina, Centro de Ciências Biológicas, Campus Universitário, Rodovia Celso Garcia Cid, PR 445 km 380, 86057-970, Londrina-PR, Brasil

Abstract

Quantifying and classifying ichthyoplankton is one of the most effective ways of monitoring the recruitment process in fishes. However, correctly identifying the fish based on morphological characters is extremely difficult, especially in the early stages of development. We examined ichthyoplankton from tributaries and reservoirs along the middle stretch of the Paranapanema River, one of the areas most impacted by hydroelectric projects in the Neotropics. Matching DNA sequences of the COI gene (628–648 bp) allowed us to identify 99.25% of 536 samples of eggs (293) and larvae (243) subjected to BOLD-IDS similarity analysis with a species-level threshold of 1.3%. The results revealed 37 species in 27 genera, 15 families and four orders, some 23.8% of documented fish species in the Paranapanema River. Molecular identification meant that we could include data from egg samples that accounted for about 30% of the species richness observed. The results in this study confirm the efficacy of DNA barcoding in identifying Neotropical ichthyoplankton and show how the data produced provide valuable information for preparing plans for conserving and managing inland waters.

Keywords: conservation, eggs, fish recruitment, upper Paraná River

Received 13 November 2014; revision received 30 January 2015; accepted 2 February 2015

Introduction

In many Neotropical countries, hydroelectric power plants (HPPs) are primary energy sources. For instance, in Brazil, 76.9% of the 592 Thw produced in 2013 were generated by HPPs (ANEEL 2013). In South America, most of these HPPs are located in the upper Paraná River basin, the region most impacted by dams in the Neotropics (Agostinho *et al.* 2008). The Paranapanema River is currently fragmented by 11 HPPs forming a complex system of cascade reservoirs. The dams impact fish fauna in many ways, primarily by blocking migration routes and disrupting the natural environment as a result of flood flow regulation, both of which impair reproduction and recruitment processes. These adverse conditions put the sustainability of fish populations at serious risk (Agostinho *et al.* 2004; Sanches *et al.* 2006; Antonio *et al.* 2007).

Studies on ichthyoplankton provide important information on the reproductive biology of a species, as well as reproduction sites and preferred times, possible migration routes and population recruitment success rates (Baumgartner *et al.* 2004; Bialetzki *et al.* 2005; Reynalte-Tataje *et al.* 2011). This information is extremely important for ecological monitoring, analysing environmental impacts and developing management and conservation plans. However, it is difficult to identify ichthyoplankton at species level, even for experienced taxonomists. Most species' descriptions are based on adult subjects (Pegg *et al.* 2006; Ko *et al.* 2013), and many diagnostic characters are not fully developed in the early development stages (Nakatani *et al.* 2001). This applies in particular to eggs, which usually account for around 80% of ichthyoplankton samples (Baumgartner *et al.* 2004; Reynalte-Tataje *et al.* 2012b). Analysing the composition of ichthyoplankton assemblages is further complicated by the enormous diversity of fish in the Neotropics. Within some families (Characidae and Loricariidae, for example), even classifying and identifying adult individuals can be a very complex process (Armbruster 2004; Oliveira *et al.* 2011).

To facilitate the discrimination and identification of animal species, Hebert *et al.* (2003a) proposed a DNA barcoding method based on the low intraspecific

Correspondence: Fernanda Simões de Almeida, Fax: +55-43-3371-4191; E-mail: fernandasa@uel.br

variation and high interspecific variation of a small fragment of 648 bp at the 5′ end of the gene cytochrome *c* oxidase, subunit I (COI or COX1, Hebert *et al.* 2003b). Since then, an extensive and unique database (BOLD: The Barcode of Life Data System, http://www.barcodinglife.org and http://www.boldsystems.org) has been set up to store this and other gene sequences for species' molecular identification (Ratnasingham & Heber 2007). By early 2014, over 2.4 million specimens had been deposited in the BOLD collection, including 13 535 fish species (Ratnasingham & Heber 2007).

Several studies have demonstrated the success of DNA barcoding for discriminating and identifying the species of adult specimens (Hebert *et al.* 2004; Valdez-Moreno *et al.* 2009; Ward *et al.* 2009), despite the megadiverse fauna produced by recent radiation, as is the case with Neotropical ichthyofauna (Carvalho *et al.* 2011; Pereira *et al.* 2013).

The aim of this study was to analyse the effectiveness of DNA barcoding in identifying fish eggs and larvae sampled in the middle stretch of the Paranapanema River, with a view to validating barcoding as a legitimate way of identifying the eggs and larvae of fish from the upper Paraná river basin, and assessing the value of information on the reproduction and recruitment of resident fish populations and distribution patterns in the ichthyoplankton community.

Material and methods

Study area and ichthyoplankton sampling

Sampling locations were selected along the lower-middle stretch of the Paranapanema River. Samples were taken between November 2012 and March 2013 in the hydrographic area of the Canoas I and II reservoirs and at five sites in the region of influence of the Capivara reservoir: three sites at the mouth of the Tibagi River (Taquara River, Apertados River and Congonhas River) and two in the main channel of the Cinzas River (Fig. 1).

Conical nets (0.5 mm mesh, 1.6 m long) were used for sampling, in conjunction with a mechanical flowmeter for measuring the amount of water filtered. The flowrate measurement was then used to determine the density of eggs and larvae according to Nakatani *et al.* (2001), standardizing egg and larvae density to a volume of 10 m^3. The nets were kept at around 20 cm under the water surface until the end of the sampling period. Samples were immediately immersed in 98% alcohol on the sampling site (three parts alcohol to one part sample) and then stored at −20 °C. In the laboratory, samples were sorted and eggs separated from larvae. To reduce and standardize the number of barcoded samples, up to thirty samples of each larval and egg morphotype from each sampling site were randomly selected for genetic analysis.

DNA extraction, COI gene amplification and sequencing

For total DNA extraction, eggs and larvae were placed in 96-well plates with 100 μL (eggs) or 200 μL (larvae) of extraction buffer. Each 200 μL of extraction buffer contained 200 mg chelex100 (Bio-Rad), 0.1 μg/mL proteinase K (Invitrogen) and ultrapure water to make up the volume. After sealing, the plates were vortexed for 10 s and then placed in a thermal cycler at 63 °C for 55 min and heated to 90 °C for 5 min. The plates were then stored in a freezer at −20 °C. Cytochrome *c* oxidase subunit 1 (COI, 648 bp) was amplified using the primers described by Ward *et al.* (2005) (FishF1-5′TCA ACC AAC CAC AAA GAC ATT GGC AC-3′ and FishR1-5′TAG ACT TCT GGG TGG CCA AAG AAT CA-3′) in an MJ Research PTC-100 thermocycler, with initial denaturation of 5 min at 94 °C, 35 cycles at 94 °C for 30 s, 54 °C for 30 s and 72 °C for 1 min, and a final extension of 10 min at 72 °C. The amplification reactions were performed in a final volume of 10 μL with 1× PCR Master Mix (Promega), 0.4 μM FishF1 and FishR1 primers, 15 ng DNA and water to make up the volume. Amplification products were purified by adding 0.5 μL Illustra Exo-Star 1-Step PCR Clean Up Kit (Thermo Fisher Scientific, Waltham, MA, USA). The purified samples were sequenced in 10 μL reactions containing 1 μL 1× BigDye Buffer (400 mM Tris-HCl pH 9.0 and 10 mM MgCl2), 2 μL BigDye Terminator v. 3.1. Cycle Sequencing Kit (Applied Biosystems, Carlsbad, CA, USA), 2 μM FishF1 or FishR1 primer (separate reactions) and water to make up the volume. The product of the reactions was sequenced bidirectionally and analysed in an ABI-PRISM 3500 XL automated sequencer (Applied Biosystems).

Data analysis

Sequences from the PCR products of each sample were combined to form a consensus using Electropherogram Quality Analysis software (http://asparagin.cenargen.embrapa.br/phph/, Togawa & Brigido 2003). They were then aligned and manually edited using MEGA v5.0 (Tamura *et al.* 2011). After editing, the sequences were entered in the BOLD database (Ratnasingham & Heber 2007) and compared to public sequences using the BOLDIDS tool (http://www.boldsystems.org/) to check the similarity of the new sequences against existing database sequences. Identification was based on the genetic distance threshold and associated similarity threshold, as well as genetic distance distribution.

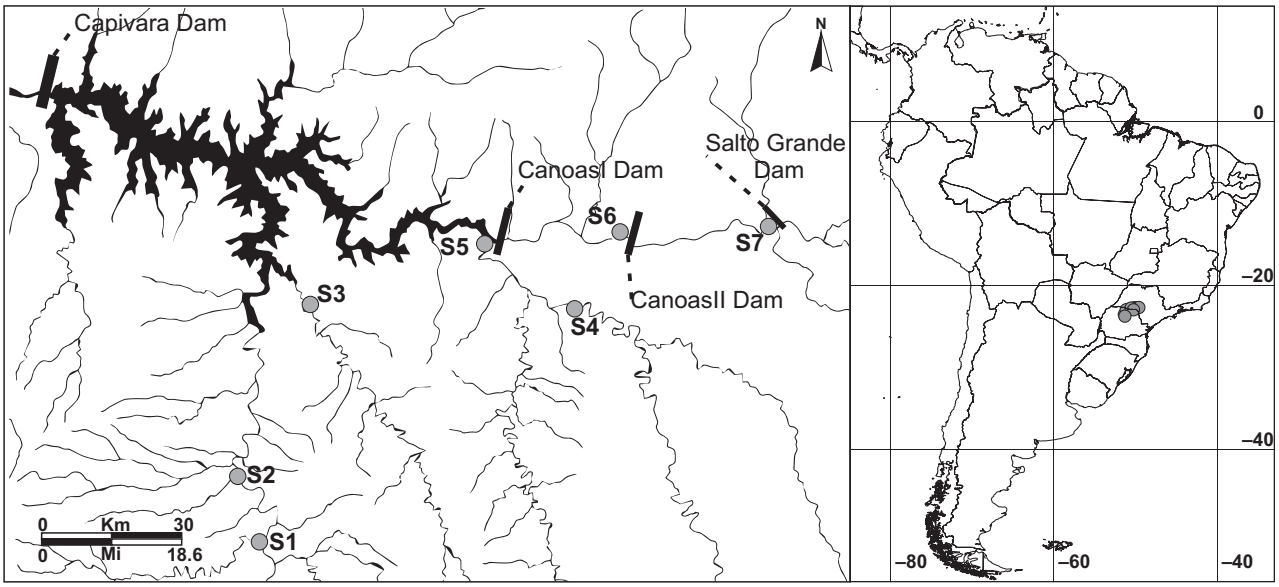

Fig. 1 Distribution of sampling points along the middle stretch of the Paranapanema River basin. Points in grey delimit the sampling sites: S1 – Taquara River (23°30′48.49″S; 50°57′15.02″O); S2 – Apertados River (23°23′28.05″S; 50°59′45.08″W); S3 – Congonhas River (23°04′04.31″S; 50°51′32.56″); S4 – middle Cinzas River (23° 4′33.31″S 50°21′47.22″O); S5 – lower Cinzas River (22°57′12.12″S 50°31′40.24″ O); S6 – Canoas I (22°56′17.10″S; 50°17′21.32″O); and S7 – Canoas II (22°54′58.29″S 49°59′32.83″O). The bold black lines denote HPP dams on the Paranapanema River.

First, matches with 99% similarity were tagged with a reference label and retained in data sets with their respective reference sequences. Sequences with 99–95% matches were tagged with the genus name and included in a data set for analysing genetic distance based on the Kimura 2-parameter model (K2P, Kimura 1980) in MEGA v5.0 (Tamura *et al.* 2011). An intraspecific genetic distance threshold of 1.3% was used to complement species identification and a 6.8% intragenus genetic distance threshold to complement genus identification where species-level identification was not possible. Both genetic distance thresholds were based on the average COI genetic distance for fish in the upper Paraná River basin (Pereira *et al.* 2013).

Finally, to assess identification reliability, neighbour-joining (NJ) trees were built based on the K2P model, including samples, references and all other sequences from the upper Paraná River basin available in the Fishes from Upper Parana River (FUPR) project (see data accessibility section, Pereira *et al.* 2013). All NJ-K2P trees and genetic distance analyses were performed using MEGA v5.0, and the robustness of NJ-K2P trees evaluated by the bootstrap method with 1000 replicates.

Results

Genetic distance and molecular identification

Throughout the study period, 6897 samples were collected comprising 5849 eggs (84.8%) and 1048 larvae (15.2%). In the laboratory, all samples were sorted and categorized as egg or larva morphotypes to reduce and standardize the sequenced samples, resulting in 536 samples for genetic analysis. Based on the similarity between the sample sequences and the adult sequences available in the database, it was possible to identify 535 of the 536 samples (99.81%) at species level. Among the 536 samples subjected to molecular analysis, KM897165 was the only sequence with similarity values lower than 99%, hampering a direct match for this sample at species level. For the remaining 535 samples, the similarity values obtained after comparing analysed sequences with existing database adult sequences remained above 99% (see Table S1, Supporting information).

All amplified fragments were longer than 600 bp, and sequences showed good quality, with no evidence of insertions, deletions or stop codons. The distribution of K2P genetic distances between different taxonomic levels shows a genetic distance increasing in proportion to the taxonomic level, ranging from 0% to 2.82% (mean 0.17%, SE 0.01%) for intraspecific comparisons and 3.99% to 17.15% (mean 11.93%, SE 0.016%) for congeneric comparisons. In addition to genetic distance analysis, the NJ-K2P tree (Fig. 2) shows clearly defined end groups, with bootstrap values higher than 80% and low distance between samples within the clusters formed by the species. Results for genetic distance for all 1249 reference sequences from upper Paraná River basin and 536 subject sequences are graphically summarized in the full

K2P-NJ tree, including all 1785 sequences, confirming correct distribution of the sequences tagged by similarity comparison (see data accessibility). In terms of nearest neighbour distance (NND), the lowest distance between taxa was 3.9%, observed in *Hypostomus regani* and *Hypostomus strigaticeps*.

Ichthyoplankton: density, abundance and taxonomic composition

A highest density of eggs (972.2 egg/m³) was observed in the middle stretch of the Cinzas River (site S4, Fig. 1). The Taquara River (site S2, Fig. 1) exhibited the lowest density, and no eggs were found at the site (see Fig. 3-D). Taking the data from molecular identification in isolation, the similarity analysis of 536 samples compared with the sequences of adults available in the BOLD database identified 37 species in 27 genera, 15 families and four orders (Table 1), around 23.8% of 155 species documented for the Paranapanema River (Duke Energy 2008), from nine morphotypes (three egg and six larva morphotypes) classified based on morphological criteria (Fig. 4).

The highest number of species (21) was identified in the order Characiformes, for both for eggs and larvae, followed by Siluriformes with 11 species and Perciformes with three species. There was only one sample in the order Gymnotiformes. Based on the molecular identification data, the most abundant order was the Siluriformes (53.15% of the total), followed by Characiformes (41.61%). Gymnotiformes and Perciformes together accounted for 5.16% of the total samples analysed.

Discussion

Egg and larva detection efficiency

Previous studies using molecular approaches to analyse the composition of ichthyoplankton assemblages were conducted almost exclusively in marine environments (Pegg *et al.* 2006; Valdez-Moreno *et al.* 2010; Ko *et al.* 2013). Recently, Loh *et al.* (2014) used this approach for identifying ichthyoplankton in inland waters in Australia. However, our study provides the first report on the distribution and taxonomic composition of ichthyoplankton assemblages in the continental waters of a megadiverse Neotropical region based on the DNA barcoding.

One of the most comprehensive methods for species identification entails analysing sequence similarity (Ratnasingham & Heber 2007; Ward *et al.* 2009). Although the similarity and genetic distance approaches may introduce some bias (DeSalle *et al.* 2005; Taylor & Harris 2012), we found both approaches effective in the data set. Similarity values compared to the sequences in the BOLD Systems database (≥99%; see Table S1, Supporting information) showed just how efficient DNA barcoding can be in analysing ichthyoplankton samples in the Neotropics, identifying 100% of the 536 samples analysed at

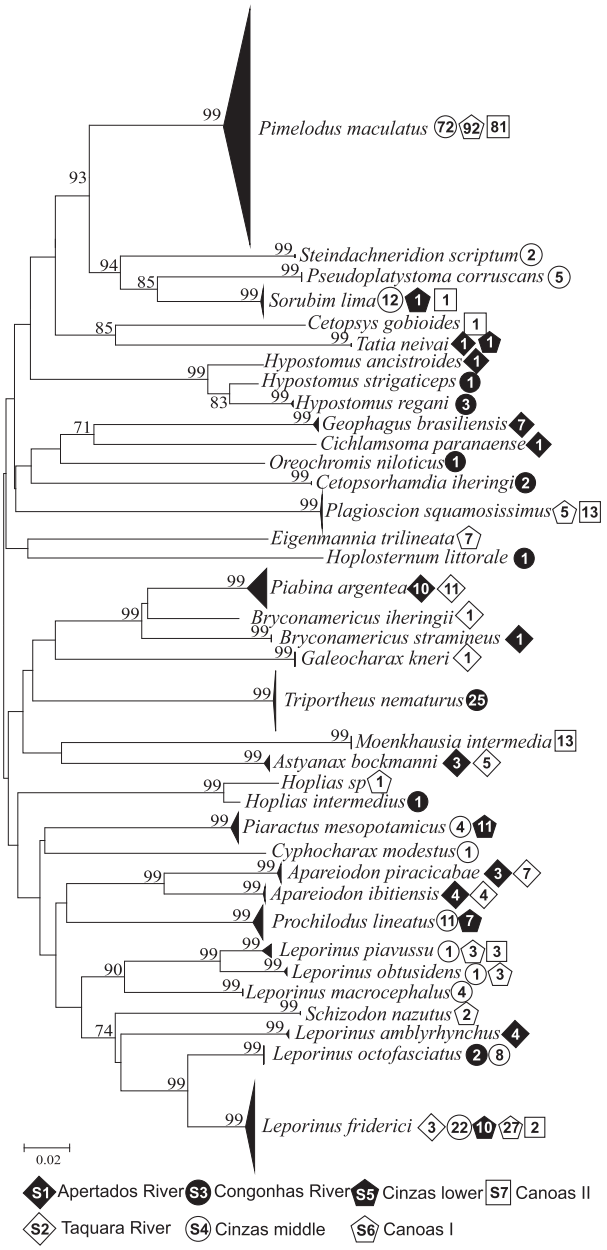

Fig. 2 K2P genetic distance. Genetic distance between the sequences of the 536 samples of eggs and larvae from Paranapanema River basin, estimated based on the Kimura 2-parameters grouped by the neighbour-joining method with 1000 bootstraps using the MEGA v5.0 software model. Branch sustainability values are shown above the nodes at each bifurcation. Values below 80% are hidden. Terminal taxa corresponding to individuals of the same species are grouped. Group names in italics indicate a match for the lowest possible taxonomic level. The symbols and figures before the taxa indicate the number of individuals of each taxon identified at each site.

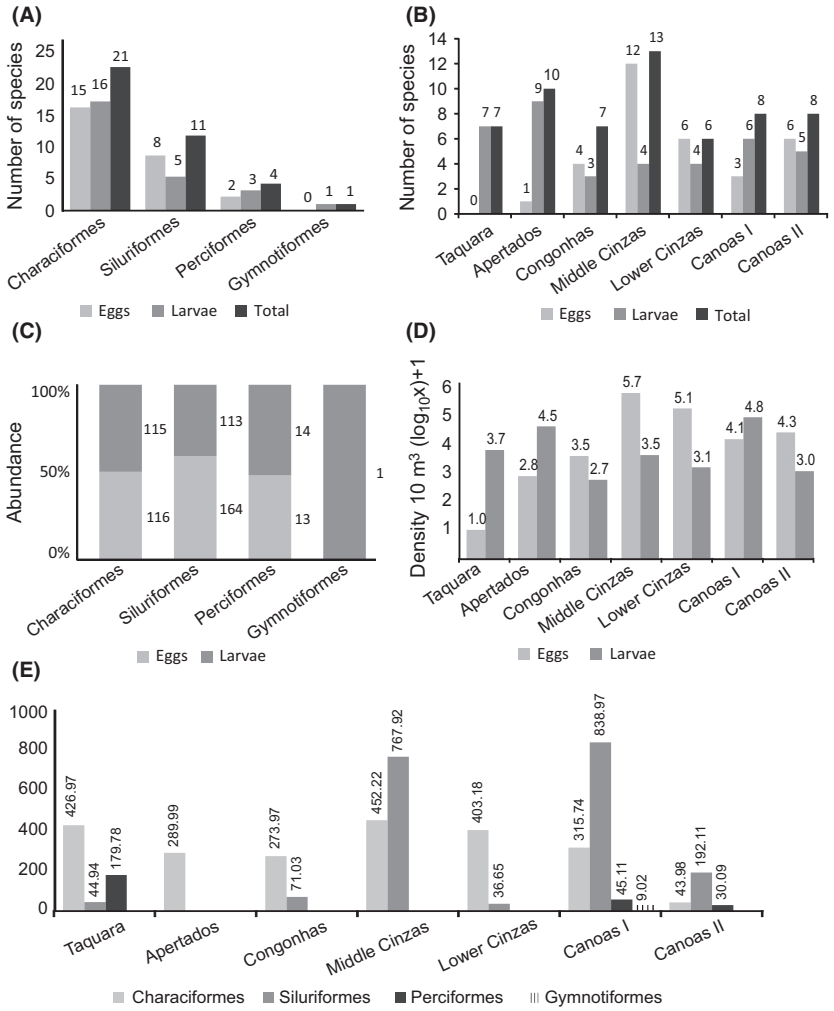

Fig. 3 Density and number of species in the samples of eggs and larvae at each site along the middle Paranapanema River. (A) – Number of species for eggs and larvae for each order; (B) – number of species identified (eggs and larvae) for each site; (C) – relative abundance of eggs and larvae for each order; (D) – density of eggs and larvae for each site (all samples taken into account for data not related to molecular identification); and (E) – relative frequency of samples for each order.

genus level and 99.81% at species level. Previous studies using this approach show a detection efficiency of around 85% for analysing ichthyoplankton in marine environments (Valdez-Moreno *et al.* 2010; Ko *et al.* 2013), and 92–100% for Neotropical adult fish (Valdez-Moreno *et al.* 2009; Pereira *et al.* 2011a, 2013). Furthermore, our data reveal efficient identification of the taxa that are not often morphologically identified among ichthyoplankton samples, such as species of the families Anostomidae, Characidae and Loricariidae (Bialetzki *et al.* 2005; Baumgartner *et al.* 2008), unambiguously identifying species of the families sampled in our study (Fig. 2).

Although the analysis of sequence similarity is a fast approach, using more than one approach (for example, the tree-based method combined with a distance threshold) is recommended to minimize false positives (Ross *et al.* 2008). The NJ-K2P tree (Fig. 2) shows well-defined clusters with high sustainability (>80%). These results, along with the nonoverlapping of genetic distances (max. 2.2% intraspecific and min. 4.99% interspecific) in relation to the intraspecific distance patterns for fish from the upper Paraná River basin (1.3% average distance, Pereira *et al.* 2013), indicate that the clusters formed by each species previously identified by similarity in fact correspond to distinct taxonomic units.

Despite the effectiveness of species identification based on similarity, one sample could not be identified at species level. Sequence KM897165 (*Hoplias* sp.) was identified only at genus level, as the similarity value compared to the sequences in the database was below the similarity threshold (99%) at species level (BOLD best match = 96.27% – *Hoplias intermedius* GB-JN988905.1; see Table S1, Supporting information, Fig. 2). The nearest neighbour distance for this sequence was 3.6%, above the average distance found among species of the upper Paraná River (1.3%, Pereira *et al.* 2013). Although *Hoplias intermedius* is recognized as a single species within the group *Hoplias lacerdae*, which is endemic to the upper Paraná River basin (Oyakawa & Mattox 2009), Pereira *et al.* (2013) observed high intraspecific variations for the COI

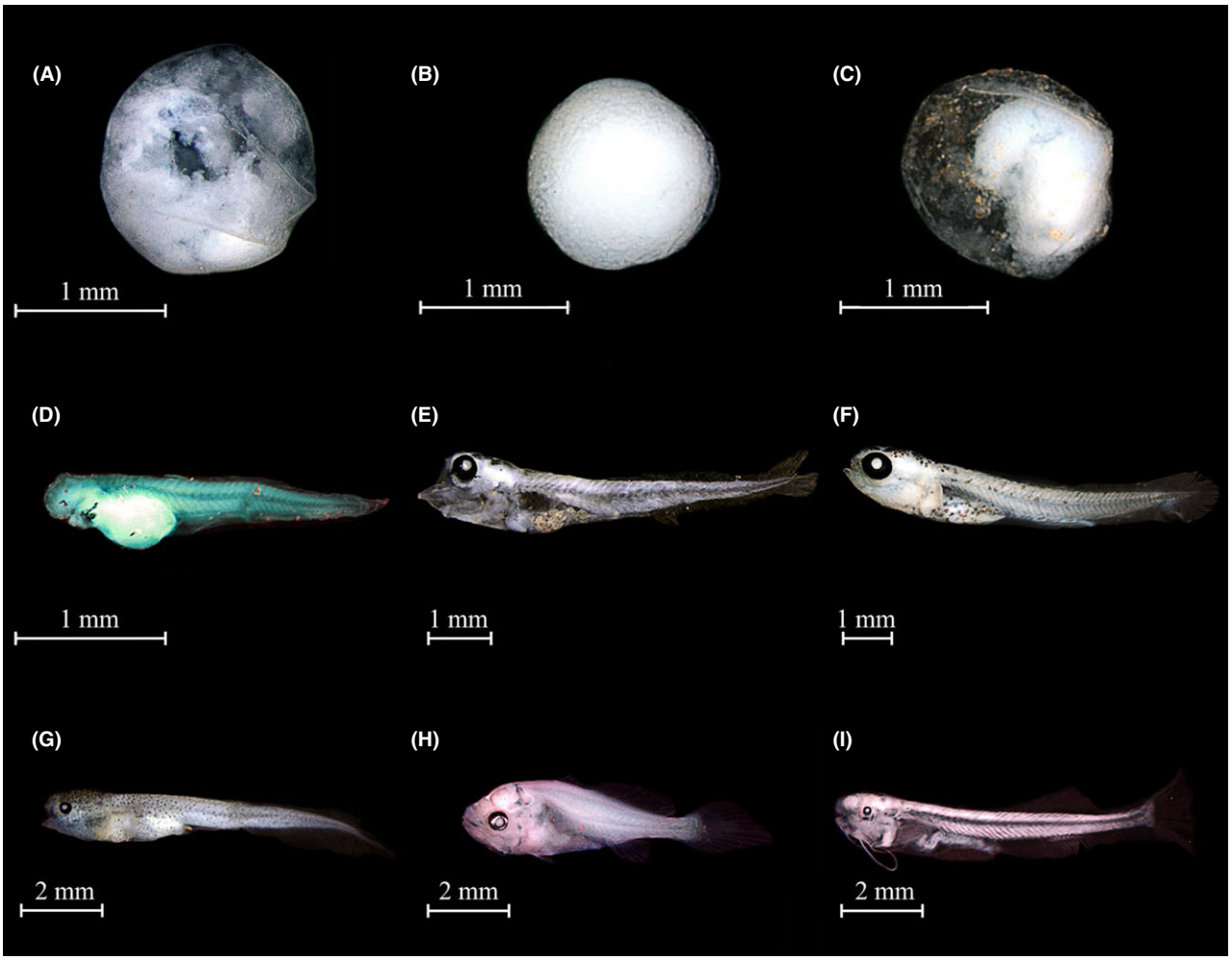

Fig. 4 Morphotypes of eggs and larvae identified prior to molecular identification. The first three tables correspond to eggs and were classified based on the vitellus distribution pattern. (A) – morphotype 1, diffuse vitellus (PDCAI101); (B) – morphotype 2 (PDCAI233), vitellus filling the whole of the yolk; (C) – morphotype 3 (PDCAI105), vitellus accumulated at a specific point. Larvae morphotypes. (D) – morphotype 4 (PDCON07), poorly differentiated samples; (E) – morphotype 5 (PDCAI144); (F) – morphotype 6 (PDCON33); (G) – morphotype 7 (PDCAI161); (H) – morphotype 8 (PDCAI146); (I) – morphotype 9 (PDCAI158). Larvae in stages of development among those exhibited in [E, F, G, H and I] and D were omitted from this table.

gene within this species, indicating the possibility of more than one taxonomic unit in the species *H. intermedius*. Furthermore, it is interesting that some authors report the existence of complex species in the genus *Hoplias* due to the high karyotypic diversity within the group (Vicari *et al.* 2006; Ferreira *et al.* 2007). This result may point to a gap in DNA barcode reference library for the *Hoplias* species, and further investigation is essential to ascertain the actual taxonomic status of *Hoplias* species in the upper Paraná River basin.

Significant COI gene intraspecific diversity was also observed in *Pimelodus maculatus*. This species showed levels of intraspecific genetic distance ranging from 0% to 2.2% (mean 0.41%, SE 0.02%). Although these values may suggest that there is more than one taxonomic unit for *P. maculatus*, our results do not support this hypothesis. The NJ-K2P tree (Fig. 3) shows the assemblages of individuals of *P. maculatus* from different sites, and the full K2P-NJ tree, including 1785 sequences from 254 species of the upper Paraná River basin (see data accessibility), indicates an important distance between congeners. Furthermore, studies evaluating the genetic structure of this species based on nuclear markers have reported a total absence or only low levels of genetic structure in *P. maculatus* throughout the upper Paraná River basin (Almeida *et al.* 2003; Ribolli *et al.* 2012), indicating that there is gene flow along the basin among these populations. Studies based on barcoding usually involve only a few samples (5–10, Hebert *et al.* 2003b; Ward *et al.* 2009; Valdez-Moreno *et al.* 2010; Pereira *et al.* 2013), and this

Table 1 Distribution of eggs and larvae along the middle Paranapanema River. The absolute abundance (A) and relative abundance (%) are shown for each identified species, as well as the absolute frequency and density in 10 m^3 (number in parentheses) for larvae (L) and eggs (E) by sampling site based on DNA barcoding data

			UHE Capivara										HPP Canoas I		HPP Canoas II	
			Apertados river		Taquara river		Congonhas river		Cinzas middle		Cinzas lower					
Species	A	%	L	E	L	E	L	E	L	E	L	E	L	E	L	E
Apareiodon ibitiensis	8	1.49	4 (89.9)		4 (36.3)											
Apareiodon piracicabae	10	1.87	3 (67.5)		7 (63.5)											
Astyanax bockmanni	8	1.49	1 (22.5)	2 (45)	5 (45.4)											
Bryconamericus iheringii	1	0.19			1 (9.1)											
Bryconamericus stramineus	4	0.75	3 (67.5)					1 (10.2)								
Cetopsis gobioides	1	0.19													1 (2.4)	
Cetopsorhamdia iheringi	2	0.37						2 (20.3)								
Cichlasoma paranaense	1	0.19	1 (22.5)													
Cyphocharax modestus	1	0.19								1 (8.6)						
Eigenmannia trilineata	1	0.19											1 (9.1)			
Galeocharax knerii	8	1.49			1 (9.1)											7 (16.3)
Geophagus brasiliensis	7	1.31	7 (157.4)													
Hoplias intermedius	1	0.19						1 (10.2)								
Hoplias sp.	1	0.19												1 (9.1)		
Hoplosternum littorale	1	0.19					1 (10.2)									
Hypostomus ancistroides	1	0.19		1 (22.5)												
Hypostomus regani	3	0.56					3 (30.5)									
Hypostomus strigaticeps	1	0.19					1 (10.2)									
Leporinus amblyrhynchus	4	0.75	4 (89.9)													

Table 1 (Continued)

			UHE Capivara													
			Apertados river		Taquara river		Congonhas river		Cinzas middle		Cinzas lower		HPP Canoas I		HPP Canoas II	
Species	A	%	L	E	L	E	L	E	L	E	L	E	L	E	L	E
*Leporinus elongatus**	4	0.75							1 (8.6)				3 (27.1)			
*Leporinus friderici**	64	11.9			3 (27.2)				8 (68.3)	14 (119.5)	3 (36.7)	7 (85.6)	27 (243.6)		2 (4.7)	
*Leporinus macrocephalus**	4	0.75								4 (34.2)						
*Leporinus obtusidens**	7	1.31								1 (8.6)			3 (27.1)		1 (2.4)	2 (4.7)
Leporinus octofasciatus	10	1.87						2 (20.3)	2 (17.1)	6 (51.2)						
Moenkhausia intermedia	7	1.31													3 (7)	4 (9.3)
Oreochromis niloticus	1	0.19						1 (10.2)								
Piabina argentea	21	3.92	10 (224.8)		11 (99.7)											
*Piaractus mesopotamicus**	16	2.99								5 (42.7)	4 (48.9)	7 (85.6)				
*Pimelodus maculatus**	245	45.7							20 (170.7)	52 (443.7)			51 (460.1)	41 (369.9)	36 (83.4)	45 (104.2)
Plagioscion squamosissimus	18	3.36											5 (45.2)			13 (30.1)
*Prochilodus lineatus**	18	3.36								11 (93.9)	1 (12.3)	6 (73.4)				
*Pseudoplatystoma corruscans**	5	0.93								5 (42.7)						
*Schizodon nasutus**	2	0.37												2 (18.1)		
*Sorubim lima**	14	2.61								12 (102.4)		1 (12.3)				1 (2.4)
*Steindachneridion scriptum**	2	0.37										2 (24.5)				
Tatia neivai	2	0.37	1 (22.5)							1 (8.6)						
Triportheus nematurus	32	5.97						25 (253.7)		2 (17.1)	3 (36.7)	2 (24.5)				
F_{Total}			34	3	32	0	5	32	31	114	11	25	90	44	43	72
D_{Total} (m^3)			764.05	67.42	290	0	50.74	324.71	264.51	972.7	134.4	305.44	811.9	396.9	99.53	166.7

*Species' average or long-run productive displacement according to Agostinho *et al.* (2003).

kind of approach can lead to bias (Meyer & Paulay 2005). In our study, 245 samples were identified as *P. maculatus*, providing a broad sampling of this species at different sites. Our results therefore suggest that some species, such as *P. maculatus*, may have higher levels of diversity for this gene and that a cautious approach should be taken to cases of above-average COI genetic divergence.

DNA barcoding helped highlight an interesting taxonomic case: the formation of two well-defined clusters of *Piabina argentea*, one restricted to the Apertados River (site S1) and the other to the Taquara River (Fig. 1, Fig. S2, Supporting information), with an average genetic distance of 1.73% (SE 0.05%, Fig. S3, Supporting information). In contrast to *P. maculatus*, *P. argentea* does not share haplotypes at different sites, indicating possible population structuring. The existence of more than one biological unit of *P. argentea* has been suggested in previous studies. Pereira *et al.* (2011b) studied COI gene distance patterns in *P. argentea* along the upper Paraná River basin and found six well-structured clusters, suggesting that each cluster corresponds to a different biological unit. Morphological and nuclear DNA analyses are required to observe actual taxonomic status and the level of gene flow between the two clusters found.

The similarity-based approach proved adequate for identifying eggs and larvae from upper Parana River basin, except in the case of *Hoplias* sp. The effectiveness of the similarity approach was in part due to the broad reference library for upper Paraná River basin, including around 254 of the 310 species already documented (Langeani *et al.* 2007; Pereira *et al.* 2013). Other studies have observed that a less exhaustive reference library could restrict the application of molecular identification (Valdez-Moreno *et al.* 2010; Hubert *et al.* 2015). However, cases such as *Hoplias* sp. show that work should continue on updating the upper Parana River basin reference library.

Species richness based on molecular analysis

One of the main advantages of using molecular tools for identifying ichthyoplankton is that samples can be analysed irrespective of development stage. The shortage of morphological characters available for classifying samples can result in an error rate of 97% in identification at species level (Ko *et al.* 2013). In our study, DNA barcoding allowed us to identify 37 species, 26 genera, seven families and four orders based on nine morphotypes (three egg and six larva morphotypes – see Fig. 4).

In addition to highlighting significant larval diversity (26 spp., Table 1), DNA barcoding revealed 24 species of egg, including 11 species (29.7% of total) that were not found in larva samples. Data from egg samples are often assessed solely for the purpose of determining spawning areas by analysing variations in site density, given that specific morphological identification is not possible for eggs (Baumgartner *et al.* 2004; Kipper *et al.* 2011; Reynalte-Tataje *et al.* 2012a). The results of our study show that molecular identification allows egg density variations to be analysed for each species individually. For example, eggs of *Steindachneridion scriptum*, an endangered species (Oyakawa & Menezes 2011), were found exclusively in the middle stretch of the Cinzas River (Fig 1, site S4), suggesting that this site is critical for species conservation.

A comparison of species richness at different sites also highlights the importance of including eggs in the taxonomic composition of ichthyoplankton assemblages. The sampling site located in the middle stretch of the Cinzas River (site S4, Fig. 1) had the highest number of species (13 spp.). Twelve species were found in egg samples and nine in larva samples. In the egg samples, eight species with some level of reproductive displacement (Agostinho *et al.* 2003) were observed (Table 1). Species with this type of behaviour are usually the most affected by the fragmentation caused by building dams (Antonio *et al.* 2007). This is even more important in view of the high frequency of eggs in ichthyoplankton samples. In a study of the upper stretch of the Paraná River and the Itaipu hydroelectric plant reservoir, Baumgartner *et al.* (2004) observed that 60% of the ichthyoplankton sampled consisted of eggs. Reynalte-Tataje *et al.* (2012a) reported a figure of 94.6% eggs in total samples taken from the Uruguay River basin.

Density, abundance and taxonomic composition

Correct molecular identification provides a more accurate idea of the spawning, foraging and growth sites of larval fish species. Some egg and larva distribution patterns were observed based on molecular identification data. In general, assemblages of eggs and larvae at the study sites consisted mainly of Characiformes and Siluriformes. The same distribution pattern was observed by Vianna & Nogueira (2008) in the Cinzas River, and by Orsi (2010) in the Capivara reservoir, both in the upper Paraná River basin. Characiformes were predominant in lotic regions, followed by Siluriformes and Perciformes in semilotic and lentic regions.

In our study, Characiformes were most abundant in all lotic regions (Apertados, Taquara and Congonhas rivers and the lower stretch of the Cinzas River). Siluriformes were most abundant only in the middle stretch of the Cinzas River. Similar results were obtained in the Canoas I and Canoas II reservoirs. At these three sites, capture density was strongly influenced by the high sample frequency of *Pimelodus maculatus* (both eggs and larvae), whose abundance may be related in part to its

reproductive strategy, characterized by a shorter reproductive displacement, high yield of oocytes and high fertility (Agostinho *et al.* 2003; Orsi 2010). As reported by Orsi (2010), species with these characteristics were more successful in colonizing the Capivara reservoir.

Another possible example of successful colonization is *Leporinus friderici*, the second most abundant species (11.93%), with similar reproductive characteristics to those of *P. maculatus*. These two species were the most abundant in the Canoas I and II reservoirs. *L. friderici* and *P. maculatus* together accounted for 57.63% of the total individuals sampled at all sites. However, the high capture density of these species may be the result of an agglomeration of individuals close to the lotic stretches of the Canoas I and II reservoirs, due to the absence of major tributaries. More accurate studies are required to evaluate the stability of these densities.

The lowest frequencies and capture densities were obtained for Gymnotiformes and Perciformes, in line with the results of previous studies (Baumgartner *et al.* 2004; Bialetzki *et al.* 2005; Reynalte-Tataje *et al.* 2012a). For both orders, this low frequency may be associated with reproductive behaviours (nest-building and rearing activities) as described for some species of the cichlid family (Goodwin *et al.* 1998). Low frequency is also found among species of Gymnotiformes (*Eigenmannia* spp., Crampton & Hopkins 2005) and some species of Siluriformes, such as Loricariidae (*Hypostomus* spp., Suzuki *et al.* 2000) and Erythrinidae (*Hoplias* spp., Prado *et al.* 2006), hampering the capture of these species. In terms of Characiformes species, an ongoing theme in ichthyoplankton studies due to their egg and larva characteristics (Vono *et al.* 2002; Kipper *et al.* 2011), the lowest densities were recorded for *Bryconamericus iheringii* and *Cyphocharax*.

In fact, smaller species (except for *Cetopsis gobiodes*, *Eigenmannia trilineata* and *Moenkhausia intermedia*) showed higher concentrations in smaller tributaries, such as the Taquara, Apertados and Congonhas rivers (Table 1). This pattern highlights the importance of conserving smaller tributaries, as they serve as breeding sites for these species. Furthermore, reservoir and tributary lotic regions (highest intake of water) were characterized by the presence and accumulation of individuals belonging to species with medium and long reproductive displacement (Table 1), following the pattern observed by Vianna & Nogueira (2008) and Reynalte-Tataje *et al.* (2012b). The same authors stress that the breeding activities of migratory species are influenced by very specific and uncommon limnological conditions. Thus, sites with the minimum conditions required for these species to reproduce should be regarded as sites of high conservation priority.

In conclusion, the results obtained confirm the high success rates for molecular identification already reported in other studies, extending the field of application to ichthyoplankton samples from inland waters of Neotropics. DNA barcoding proved especially useful for identifying taxa that are difficult to classify in ichthyoplankton samples, such as species of Anostomidae, Characidae and Loricariidae. Eggs can be analysed in ichthyoplankton assemblages, detecting taxa that are found only in egg samples. The similarity-based approach proved adequate for identifying eggs and larvae from upper Parana River basin; however, work on updating libraries should continue to cover all species. It is therefore strongly recommended that molecular approaches to ichthyoplankton studies be included in future research proposals aimed at enhancing accuracy of results and providing more exact information for preparing detailed conservation plans.

Acknowledgements

The authors wish to thank the team of Laboratório de Ecologia de Peixes e Invasões Biológicas (LEPIB) and Edson Santana for their help with sampling and selecting the material; Duke Energy for supporting this project; CAPES for the fellowship grant to the first author; and Universidade Estadual de Londrina for providing facilities and technical support. Silvia H. Sofia is a research fellow with the Brazilian Council for Scientific and Technological Development (CNPq).

References

Agostinho AA, Gomes LC, Suzuki HI, Julio HF Jr (2003) Migratory fishes of the Upper Paraná River Basin Brazil. In: *Migratory Fishes of South America: Biology, Fisheries and Conservation Status* (eds Carolsfeld J, Harvey B, Ross C & Baer A), p. 361. International Development Research Centre, Ottawa, Ontario.

Agostinho AA, Gomes LC, Verı S, Okada EK (2004) Flood regime, dam regulation and fish in the Upper Paraná River: effects on assemblage attributes, reproduction and recruitment. *Fish Biology and Fisheries*, **14**, 11–19.

Agostinho AA, Pelicice FM, Gomes LC (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.

Almeida FS, Sodré LMK, Contel EPB (2003) Population structure analysis of *Pimelodus maculatus* (Pisces, Siluriformes) from the Tietê and Paranapanema Rivers (Brazil). *Genetics and Molecular Biology*, **26**, 301–305.

ANEEL (2013) BIG - Banco de Informações de Geração. Available from http://www.aneel.gov.br/aplicacoes/capacidadebrasil/capacidadebrasil.cfm.

Antonio RR, Agostinho AA, Pelicice FM *et al.* (2007) Blockage of migration routes by dam construction: can migratory fish find alternative routes? *Neotropical Ichthyology*, **5**, 177–184.

Armbruster JW (2004) Phylogenetic relationships of the suckermouth armoured catfishes (Loricariidae) with emphasis on the Hypostominae and the Ancistrinae. *Zoological Journal of the Linnean Society*, **141**, 1–80.

Baumgartner G, Nakatani KK, Gomes LC *et al.* (2004) Identification of spawning sites and natural nurseries of fishes in the upper Paraná River, Brazil. *Environmental Biology of Fishes*, **71**, 115–125.

Baumgartner G, Nakatani K, Gomes LC *et al.* (2008) Fish larvae from the upper Paraná River : do abiotic factors affect larval density? *Neotropical Ichthyology*, **6**, 551–558.

Bialetzki 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.

Carvalho D, Denise AAO, Pompeu PS *et al.* (2011) Deep barcode divergence in Brazilian freshwater fishes: the case of the São Francisco River basin. *Mitochondrial DNA*, **22**, 80–86.

Crampton W, Hopkins C (2005) Nesting and paternal care in the weakly electric fish *Gymnotus* (Gymnotiformes: Gymnotidae) with descriptions of larval and adult electric organ discharges of two. *Copeia*, **1**, 48–60.

DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1905–1916.

Duke Energy (2008) *Peixes do rio Paranapanema*. Horizonte Geográfico, São Paulo.

Ferreira IA, Bertollo LAC, Martins C (2007) Comparative chromosome mapping of 5S rDNA and 5SHindIII repetitive sequences in Erythrinidae fishes (Characiformes) with emphasis on the Hoplias malabaricus "species complex". *Cytogenetic and Genome Research*, **118**, 78–83.

Goodwin NB, Balshine-Earn S, Reynolds JD (1998) Evolutionary transitions in parental care in cichlid fish. *Proceedings of the Royal Society B Biological Sciences*, **265**, 2265–2272.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B Biological Sciences*, **270**, 313–321.

Hebert PDN, Ratnasingham S, DeWaard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B Biological Sciences*, **270**, 96–99.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *PNAS*, **101**, 14812–14817.

Hubert N, Espiau B, Meyer C, Planes S (2015) Identifying the ichthyoplankton of a coral reef using DNA barcodes. *Molecular Ecology Resources*, **15**, 57–67.

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.

Kipper D, Bialetzki A, Santin M (2011) Composição taxonômica da assembleia de larvas de peixes no reservatório de Rosana, Rio Paranapanema, Brasil Introdução. *Biota Neotropica*, **11**, 421–426.

Ko H, Wang Y, Chiu T *et al.* (2013) Evaluating the accuracy of morphological identification of larval fishes by applying DNA barcoding. *PLoS ONE*, **8**, 3–9.

Langeani F, Castro RMC, Oyakawa OT *et al.* (2007) Diversidade da ictiofauna do Alto Rio Paraná: composição atual e perspectivas futuras. *Biota Neotropica*, **7**, 181–197.

Loh WKW, Bond P, Ashton KJ, Roberts DT, Tibbetts IR (2014) DNA barcoding of freshwater fishes and the development of a quantitative qPCR assay for the species-specific detection and quantification of fish larvae from plankton samples. *Journal of Fish Biology*, **85**, 307–328.

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, **3**, e422.

Nakatani K, Agostinho AA, Baumgartner G *et al.* (2001) *Ovos e larvas de peixes de água doce: desenvolvimento e manual de identificação*. EDUEM, Maringá.

Oliveira C, Avelino GS, Abe KT *et al.* (2011) Phylogenetic relationships within the speciose family Characidae (Teleostei: Ostariophysi: Characiformes) based on multilocus analysis and extensive ingroup sampling. *BMC Evolutionary Biology*, **11**, 275–300.

Orsi ML (2010) *Estratégias reprodutivas de peixes: Estratégia reprodutiva de peixes da região média baixa do rio Paranapanema, Reservatório de Capivara*. Edgard Blucher, São Paulo.

Oyakawa OT, Mattox GMT (2009) Revision of the Neotropical trahiras of the *Hoplias lacerdae* species-group (Ostariophysi : Characiformes: Erythrinidae) with descriptions of two new species. *Neotropical Ichthyology*, **7**, 117–140.

Oyakawa OT, Menezes NA (2011) Checklist dos peixes de água doce do Estado de São Paulo, Brasil. *Biota Neotropica*, **11**, 19–31.

Pegg GG, Sinclair B, Briskey L, Aspden WJ (2006) MtDNA barcode identification of fish larvae in the southern Great Barrier Reef, Australia. *Scientia Marina*, **70**, 7–12.

Pereira LHG, Maia GMG, Hanner R, Foresti F, Oliveira C (2011a) DNA barcodes discriminate freshwater fishes from the Paraíba do Sul River Basin, São Paulo, Brazil. *Mitochondrial DNA*, **22**, 71–79.

Pereira LHG, Pazian MF, Hanner R, Foresti F, Oliveira C (2011b) DNA barcoding reveals hidden diversity in the Neotropical freshwater fish *Piabina argentea* (Characiformes: Characidae) from the Upper Paraná Basin of Brazil. *Mitochondrial DNA*, **22**, 87–96.

Pereira LHG, Hanner R, Foresti F, Oliveira C (2013) Can DNA barcoding accurately discriminate megadiverse Neotropical freshwater fish fauna? *BMC Genetics*, **14**, 20–34.

Prado CPA, Gomiero LM, Froehlich O (2006) Spawning and parental care in *Hoplias malabaricus* (Teleostei, Characiformes, Erythrinidae) in the southern pantanal, Brazil. *Brazilian Journal of Biology*, **66**, 697–702.

Ratnasingham S, Heber PDN (2007) BOLD: The Barcode of Life Data System (www.bacodingoflife.org). *Molecular Ecology Notes*, **7**, 355–364.

Reynalte-Tataje DA, Nakatani K, Fernandes R, Agostinho AA, Bialetzki A (2011) Temporal distribution of ichthyoplankton in the Ivinhema River (Mato Grosso do Sul State/Brazil): influence of environmental variables. *Neotropical Ichthyology*, **9**, 427–436.

Reynalte-Tataje DA, Agostinho AA, Bialetzki A *et al.* (2012a) Spatial and temporal variation of the ichthyoplankton in a subtropical river in Brazil. *Environmental Biology of Fishes*, **94**, 403–419.

Reynalte-Tataje DA, Nuñer A, Nunes M *et al.* (2012b) Spawning of migratory fish species between two reservoirs of the upper Uruguay River, Brazil. *Neotropical Ichthyology*, **10**, 829–835.

Ribolli J, Manoel C, De Melo R, Zaniboni-filho E (2012) Genetic characterization of the Neotropical catfish *Pimelodus maculatus* (Pimelodidae, Siluriformes) in the Upper Uruguay River. *Genetics and Molecular Biology*, **35**, 40–49.

Ross HA, Murugan S, Li WLS (2008) Testing the reliability of genetic methods of species identification via simulation. *Systematic Biology*, **57**, 216–230.

Sanches PV, Nakatani K, Bialetzki A *et al.* (2006) Flow regulation by dams affecting ichthyoplankton: the case of the Porto Primavera Dam, Paraná River, Brazil. *River Research and Applications*, **22**, 555–565.

Suzuki H, Agostinho AA, Winemiller KO (2000) Relationship between oocyte morphology and reproductive strategy in loricariid catfishes of the Paraná River, Brazil. *Journal of Fish Biology*, **57**, 791–807.

Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Taylor HR, Harris WE (2012) An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. *Molecular Ecology Resources*, **12**, 377–388.

Togawa RC, Brigido MM (2003) PHPH?: Web based tool for simple electropherogram quality analysis. In: *1st International Conference on Bioinformatics and Computational Biology - IcoBiCoBi, p. 1. Ribeirão Preto, Brasil*. Available from http://asparagin.cenargen.embrapa.br/phph/.

Valdez-Moreno M, Vásquez-Yeomans L, Elías-Gutiérrez M, Ivanova NV, Hebert PDN (2010) Using DNA barcodes to connect adults and early life stages of marine fishes from the Yucatan Peninsula, Mexico: potential in fisheries management. *Marine and Freshwater Research*, **61**, 665–671.

Valdez-Moreno M, Ivanova NV, Elias-Gutiérrez M, Balderas SC, Hebert PDN (2009) Probing diversity in freshwater fishes from Mexico and Guatemala with DNA barcodes. *Journal of Fish Biology*, **74**, 377–402.

Vianna NC, Nogueira MG (2008) Ichthyoplankton and limnological factors in the Cinzas River – an alternative spawning site for fishes in the middle Paranapanema River basin, Brazil. *Acta Limnologica Brasiliensia*, **20**, 139–151.

Vicari M, Pazza R, Artoni R (2006) Cytogenetics and biogeography: considerations about the natural origin of *Hoplias malabaricus* (Characifor-

mes, Erythrinidae) on the Iguaçu River. *Brazilian Archives of Biology and Technology*, **49**, 297–303.

Vono V, Silva LGM, Maia BP, Godinho HP (2002) Biologia reprodutiva de três espécies simpátricas de peixes neotropicais: *Pimelodus maculatus* Lacépède (Siluriformes, Pimelodidae), *Leporinus amblyrhynchus* Garavello & Britski e *Schizodon nasutus* Kner (Characiformes, Anostomidae) do recém-formado Reservatório. *Revista Brasileira de Zoologia*, **19**, 819–826.

Ward RD, Zemlak TS, Innes BH *et al*. (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1847–1857.

Ward RD, Hanner R, Hebert PDN (2009) The campaign to DNA barcode all fishes, FISHBOL. *Journal of Fish Biology*, **74**, 329–356.

W.F.S., F.S.A. and M.L.O designed the project. M.L.O. and W.F.S collected the specimens. W.F.S. performed the genetic data analysis. W.F.S., F.S.A. and S.H.S drafted the study.

Data Accessibility

DNA sequences: GenBank Accession no. KM897137-KM897672.

Full K2P-NJ tree, phylogenetic data and final DNA assembly: TreeBASE study Accession no. 16630, Reviewer access URL: http://purl.org/phylo/treebase/phylows/study/TB2:S16630?x-access-code=aa2492f86d28c473053f3341cc5da570&format=html.

BOLD projects: PDCAI (Accession nos. PDCAI001-13 to PDCAI134-13); PDCAP (Accession nos. PDCAP001-14 to PDCAP287-14); PDCII (Accession no. PDCII115-14) – container FELPR. BOLD annexed project FUPR, Accession nos.: BAST525-12, FUPR001-09 to FUPR1468-10.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 K2P genetic divergence. Distribution of K2P genetic divergence within species and within genus levels. Histogram plots showing the distribution of normalized divergence for species (pink) against the genus divergences (green).

Fig. S2 Dendrogram for all *Pimelodus maculatus*. NJ dendrogram shows graphic representation of K2P genetic distance from 245 taxon sharing haplotypes among different sample sites. Node values = bootstrap test. Codes before hyphen is BOLD sequence ID.

Fig. S3 Dendrogram for all *Piabina argentea*. NJ-K2P tree shows graphic representation of genetic distance from 21 taxon with two distinct haplotypes between two different localities. Node values = bootstrap test. Branch length is equal to K2P distance between nodes. Distance between clusters is equal to the sum of branch lengths. Codes before hyphen are BOLD sequence ID.

Table S1 Distances within species and distance to Nearest Neighbor Species comparison.

Table S2 Summary of K2P genetic divergence within different taxonomic levels from 536 analyzed specimens.