Genetics and Molecular Biology, 43, 1, e20190131 (2020)

DOI: https://doi.org/10.1590/1678-4685-GMB-2018-0131

Research Article
Genomics and Bioinformatics

Genetic evidence for a species complex within the piranha *Serrasalmus maculatus* (Characiformes, Serrasalmidae) from three Neotropical river basins based on mitochondrial DNA sequences

Thaís Souto Bignotto[1], Thiago Cintra Maniglia[2], Vivian Nunes Gomes[3], Isadora Janolio de Oliveira[3], Carlos Sérgio Agostinho[4], Sônia Maria Alves Pinto Prioli[3,5] and Alberto José Prioli[3]

[1]*Universidade Estadual do Oeste do Paraná (Unioeste), Centro de Engenharias e Ciências Exatas, Grupo de Pesquisas em Recursos Pesqueiros e Limnologia (Gerpel), Toledo, PR, Brazil.*
[2]*Universidade Tecnológica Federal do Paraná (UTFPR), Toledo, PR, Brazil.*
[3]*Universidade Estadual de Maringá (UEM), Núcleo de Pesquisas em Limnologia, Ictiologia e Aquicultura, Maringá, PR, Brazil.*
[4]*Universidade Federal do Tocantins (UFT), Núcleo de Estudos Ambientais (Neamb), Porto Nacional, TO, Brazil.*
[5]*Universidade Estadual de Maringá (UEM), Departamento de Biotecnologia, Genética e Biologia Celular, Maringá, PR, Brazil.*

Abstract

Mitochondrial molecular markers (DNA sequences of D-loop, cytochrome b and cytochrome c oxidase I) were employed to characterize populations of the piranha *Serrasalmus maculatus* from Upper Paraná, Upper Paraguay and Tocantins River basins. D-loop sequences of *S. maculatus* population from Paraná-Paraguay River basin exhibited tandem repeats of short motifs (12 base pairs) and variable numbers depending on specimens, accounting for length variation. Concatenated mitochondrial sequences suggested that *S. maculatus* encompasses different mitochondrial DNA lineages. Although sampling was restricted to three river basins, phylogenetic analysis clearly indicated that the species currently recognized as *S. maculatus* presents high genetic variability. Maximum likelihood and Bayesian analysis clustered *S. maculatus* populations according to their locations. However, the highest genetic differentiation was identified between populations from Paraná-Paraguay system and Tocantins River basin. Three species delimitation analyses (PTP, GMYC, and ABGD) suggested that there are at least two species among the analyzed populations. The analysis of the mitochondrial sequences evidenced genetic differentiation among populations corresponding to related, but different species, suggesting that at least *S. maculatus* from the Tocantins River and Paraná-Paraguay River basins are most likely different species. Therefore, *S. maculatus* should be considered a species complex with morphologically cryptic diversity. An integrative revision is suggested.

Keywords: Cryptic species, species delimitation, D-loop, cytochrome b, cytochrome c oxidase I.

Received: May 9, 2018; Accepted: March 8, 2019.

Introduction

For many years, piranhas have fascinated scientists around the world primarily because of their biological and evolutionary characteristics. However, the systematics of piranhas has confused ichthyologists (Fink and Machado-Allison, 2001; Jégu and Santos, 2001), and several studies revealed great ecological diversity in this group of fish (Fink and Machado-Allison, 1992; Machado-Allison and Fink, 1995a, b). Freeman *et al.* (2007) stated that the taxonomy and systematics of piranhas, as well as of other species of Serrasalmidae, Characiformes, are complex and many questions remain unanswered. *Serrasalmus* is one of the most diverse but taxonomically problematic genera of piranhas. Among the several difficulties to define species correctly, there is the probable occurrence of species complexes in the piranhas *Serrasalmus maculatus* Kner 1858 and *S. rhombeus* (L. 1766) (see for example, Freeman *et al.*, 2007).

Serrasalmus maculatus (Characiformes, Serrasalmidae) is a widely distributed piranha species, occurring naturally in the Amazon and Paraná-Paraguay River basins. The species had its karyotype described by specimens from different localities of the basins they occur (Cestari and Galetti

Send correspondence to Thaís Souto Bignotto. Universidade Estadual do Oeste do Paraná (Unioeste), Centro de Engenharias e Ciências Exatas, Grupo de Pesquisas em Recursos Pesqueiros e Limnologia (Gerpel), Campus de Toledo, R. da Faculdade, 645, 85.903-000, Toledo, PR, Brazil. E-mail: thais.bignotto@unioeste.br.

Jr, 1992; Martins-Santos *et al.*, 1994; Nakayama *et al.*, 2000; Centofante *et al.*, 2002; Nakayama *et al.*, 2002, 2012). All specimens presented a diploid number of 60 chromosomes and multiple nucleolus organizer regions (NORs) located on the short arm of the acrocentric chromosomes. However, high intraspecific chromosome diversity was reported by the observation of seven distinct karyotypes, three of them occurring in the Paraná-Paraguay system and the remaining in the Amazon basin, with two of them in sympatry (Nakayama *et al.*, 2000). The presence of cytotypes found in both sympatry and allopatry for *S. maculatus* suggests that either the species presents wide karyotype heterogeneity or there is a complex of species whose morphological characteristics are very similar (Nakayama *et al.*, 2000). The high diversity of karyotypes allied to a wide geographical distribution of the species, enabled researchers to suggest that *S. maculatus* may represent a complex of cryptic species (Cestari and Galetti Jr, 1992; Nakayama *et al.*, 2000; Hubert *et al.*, 2006).

The correct species delimitation is of extreme importance not only for systematics, but also for ecology, biogeography, comparative biology, and conservation areas (Frankham *et al.*, 2012; Hernández *et al.*, 2015). Molecular methods, such as DNA sequencing, have arisen over recent decades as invaluable tools for identifying biodiversity that may not be evident by traditional morphology-based taxonomy and systematics (Bickford *et al.*, 2007; Larson *et al.*, 2016). Although molecular approaches should not replace traditional taxonomy and systematics, these methods do offer considerable power to clarify cases of convergent evolution or complex evolutionary histories (Larson *et al.*, 2016). Thus, the elucidation of the taxonomic status of *S. maculatus*, in addition to its genetic characterization, is useful to reveal intra and interspecific genetic differences, producing relevant information on biodiversity and evolutionary history of populations.

Different segments of mitochondrial DNA may be used as tools for the correct identification of species. For this reason, the aim of this study was to characterize natural populations of the *S. maculatus* piranha from Paraná-Paraguay and Tocantins hydrographic River basins using a multi-gene molecular approach [mitochondrial sequences of the control region, D-loop, and the genes cytochrome c oxidase subunit I (*coI*), and cytochrome b (*cytb*)], allied to three molecular-based species delimitation methods in order to detect the existence of a complex of cryptic species. In addition, specimens of *Serrasalmus* sp. from Tocantins River basin were included in the analysis in order to verify the genetic relationship with *S. maculatus*.

Materials and Methods

Sample collection

Four specimens of *S. maculatus* were collected in the Upper Paraná River basin (at sampling sites along the Upper Paraná River Floodplain, Baía River, and Garças lagoon), seven in the Upper Paraguay River basin (Manso River), and seven in the Tocantins River basin (Tocantins River) (Figure 1 and Table 1). Additionally, four specimens morphologically identified as *Serrasalmus* sp. of occurrence in the Tocantins River basin were collected and included in the analyses to be confronted with *S. maculatus* specimens. A sample of the red piranha *Pygocentrus nattereri* Kner 1858 was used as outgroup in the analysis. Piranha specimens of Paraná-Paraguay and Tocantins River basins were taxonomically identified by C.S. Pavanelli and C.S. Agostinho, respectively.

Specimens were anaesthetized and subsequently sacrificed by clove oil overdose, according to Griffiths (2000). Samples of muscle tissue were taken from each individual, fixed in commercial 96% ethyl alcohol, and kept in individual flasks. They were then stored in a freezer at -20 °C. Specimens were deposited in the Ichthyological Collection of the Center for Research in Limnology, Ichthyology and Aquaculture (Nupélia) of the State University of Maringá (UEM), and at the Laboratory of Ichthyology and Systematics of the Federal University of Tocantins (UFT) (Table 1). The study was approved by the UEM's Committee of Ethics on Animal Care (protocol number 123/2010).

DNA extraction, PCR amplification, and DNA sequencing

Total DNA was extracted from muscle tissue according to methodology based on phenol/chloroform (Oliveira *et al.*,

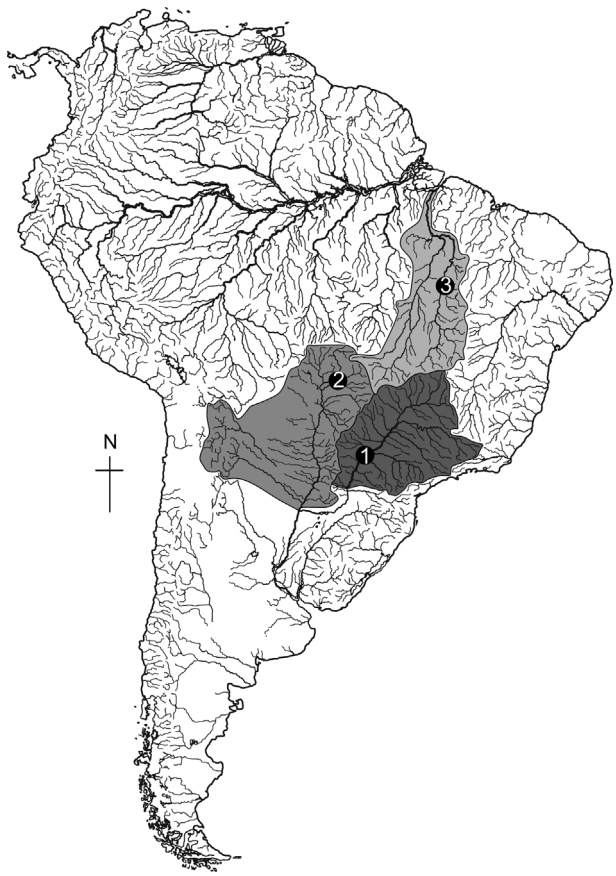

Figure 1 - Sampling locations of *Serrasalmus maculatus* and *Serrasalmus* sp. in Brazil. Numbers correspond to the local sample: 1. Upper Paraná River basin; 2. Upper Paraguay River basin; 3. Tocantins River basin.

Table 1 - Specimens of *Serrasalmus maculatus*, *Serrasalmus* sp. and *Pygocentrus nattereri* (outgroup) analyzed in the present study, and their respective GenBank accession numbers for mitochondrial DNA sequences of cytochrome b (*cytb*), cytochrome c oxidase I (*coI*) and control region D-loop. Code: species name and sampling sites abbreviation; N: number of analyzed specimens; Voucher: NUP – number of catalogue at Ichthyological Collection of the Nupelia (Center for Research in Limnology, Ichthyology and Aquaculture)/State University of Maringá (UEM), UNT - number of catalogue at Laboratory of Ichthyology and Systematic of Federal University of Tocantins (UFT).

Species	Code	Sampling sites	N	Voucher	GenBank Accession No.		
					cytb	*col*	D-loop
S. maculatus	SmcPY	Upper Paraguay basin (Manso River)	7	NUP 884	KP256436-442	KP256372-378	KP998542-548
S. maculatus	SmcPR	Upper Paraná basin (Garças Lagoon)	1	NUP 4208	KP998540	KP998541	KP998549
S. maculatus	SmcPR	Upper Paraná basin (Baía River)	2	NUP 4208	KP256447; 449	KP256383; 385	KP998550; 551
S. maculatus	SmcPR	Upper Paraná basin (Floodplain)	1	NUP 4208	KP256454	KP256390	KP998552
S. maculatus	SmcTO	Tocantins River	7	UNT 8175	KP256455-461	KP256391-397	KP998553-559
Serrasalmus sp.	SrsTO	Tocantins River	4	—	KP998566-569	KP998570-573	KP998560-563
P. nattereri	NtrPY	Upper Paraguay basin (Manso River)	1	NUP 886	KP256488	KP256424	KP998565

2006). After DNA quantification, fragments of the mitochondrial genome were amplified via polymerase chain reaction (PCR), from total DNA samples. PCR amplification conditions were based on Prioli *et al.* (2002). Three segments of mitochondrial DNA were used: control region (D-loop) and the cytochrome b (*cytb*) and cytochrome c oxidase subunit I (*coI*) genes. The pair of primers H16498 (5'- CCT GAA GTA GGA ACC AGA TG -3'; Meyer *et al.*, 1990) and L14841 (5'-CCA TCC AAC ATC TCA GCA TGA TGA AA-3'; Kocher *et al.*, 1989) were used to amplify by PCR a segment of almost 1,700 base pairs (bp) that included both D-loop region and *cytb* gene. However, PCR amplification and sequencing of some *S. maculatus* specimens was not possible with this set of primers. In this specific case, a second pair of primers was used, H16498 and L D-loop M (5'-WAA GCR TCG GTC TTG TAA WCC -3', Cronin *et al.*, 1993, with modifications), resulting in a fragment ranging from 550 to 650 bp, approximately. Finally, primers H7152 (5'-CAC CTC AGG GTG TCC GAA RAA YCA RAA -3'; Ivanova *et al.*, 2007) and L6448-F1 (5'- TCA ACC AAC CAC AAA GAC ATT GGC AC -3'; Ward *et al.*, 2005) were used to amplify a partial sequence of the *coI* gene, approximately 700 bp long.

Fragments were amplified by independent PCR assays in order to be sequenced and analyzed. The reaction mix consisted of Tris-KCl buffer (20 mM Tris-HCl, pH 8.4, 50 mM KCl), 1.5 mM $MgCl_2$, 2.5 μM of each primer, 0.1 mM of each dNTP, 2.5 U of *Taq* DNA polymerase, 15 ng of DNA and filtered/deionized water (Milli-Q) for a final volume of 25 μL. Amplifications of fragments were performed in a thermocycler, programmed for different temperature profiles depending on the set of primers used. The thermal profiles used for amplification of the D-loop and *cytb* regions were as follows: an initial step of 4 min at 94 °C, followed by 40 cycles of 15 s at 94 °C, 30 s at 59–61 °C, and 2 min at 72 °C, with an additional last step of 10 min at 72 °C. For the *coI* region, an initial step of 2 min at 94 °C was followed by 35 cycles of 30 s at 94 °C, 40 s at 52–55 °C, and 1 min at 72 °C, with an final step of 10 min at 72 °C.

The amplification efficiency was confirmed on 1% agarose gel. Samples were then purified with polyethylene glycol (PEG), according to Rosenthal *et al.* (1993). The purified PCR fragments were once more amplified unidirectionally with primers H16498 or L D-loop M for the D-loop region, L14841 for the *cytb* region, and L6448-F1 for the *coI* region. Approximately 50 ng of DNA from the final product of each PCR reaction were used directly in sequencing reactions with the DYEnamic ET Dye Terminator Kit (Amersham Biosciences) in a MegaBACE 1000 automatic sequencer (Amersham Biosciences) according to the manufacturer's instructions. All sequences were deposited in GenBank (Table 1).

Data and phylogenetic analyses

The nucleotide sequences were aligned with Clustal Omega software (Sievers *et al.*, 2011) and manually edited in the BioEdit Sequence Alignment Editor 7.0.1 (Hall, 1999). Polymorphic sites, number of parsimony informative sites, and nucleotide compositions were obtained with MEGA7 software (Kumar *et al.*, 2016). The number of haplotypes was assessed with DNAsp 6.10 software (Rozas *et al.*, 2017). Fixation indices (F_{ST}) and analysis of molecular variance, AMOVA, were calculated using Arlequin 3.11 (Excoffier *et al.*, 2005).

Three independent single-gene phylogenies were constructed using the D-loop, *coI*, and *cytb* mtDNA regions, excluding any redundant sequences. After checking for congruence among tree topologies derived from the single-gene phylogenies, analyses were based on the concatenated sequences of the D-loop, *coI*, and *cytb* sequences, resulting in a single large alignment. The best-fit model of nucleotide evolution was estimated by PartitionFinder 2.1 software (Lanfear *et al.*, 2012). Phylogenetic analyses were based on both maximum likelihood (ML) and Bayesian (BA) approaches. First and second codon positions of the coding regions *coI* and *cytb*, and the third codon position of the *coI* and *cytb* together with the noncoding D-loop region, were used as two different partitions in the concatenated analyses, as defined by the PartitionFinder 2.1.

Best-scoring ML trees were estimated for each dataset using the raxmlGUI software (Silvestro and Michalak,

2012), using rapid bootstrap algorithm, autoMRE function for resamplings, and the substitution model and partition set, previously defined by the PartitionFinder 2.1. Bayesian trees were calculated using the uncorrelated lognormal relaxed-clock model implemented in BEAST 1.8.2 with an input file generated in BEAUti 1.8.0 (Drummond *et al.*, 2012). The Yule process of speciation, which assumes a constant speciation rate among lineages, was applied as a tree prior. Each analysis ran for 10,000,000 generations with a sample frequency of 1,000. The final trees were calculated after 10% of burn-in. Length of burn-in was determined by examination of traces in Tracer 1.6 (Rambaut *et al.*, 2014). Support for nodes was determined using posterior probabilities (PP; calculated by BEAST).

Species delimitation

Three methods for species delimitation were used to identify the specific boundaries in *S. maculatus* and *Serrasalmus* sp.: (i) the Poisson tree process model (PTP; Zhang *et al.*, 2013); (ii) the General Mixed Yule Coalescent method (GMYC, single and multiple threshold algorithm) of Pons *et al.* (2006), and (iii) the Automated Barcode Gap Discovery (ABGD) method of Puillandre *et al.* (2012). These methods were applied to unique haplotypes (redundant sequences were excluded) for the *coI* dataset only, in order to preserve consistence in the relation of this fragment and to specify Molecular Operational Taxonomic Units (MOTUs) (i.e., DNA barcoding; Hebert *et al.*, 2003; Larson *et al.*, 2016). Unlike ABGD that uses detection of the `barcode gap' in the distribution of genetic pairwise distances, GMYC and PTP use a phylogenetic input tree from which the fit of speciation and coalescent processes are modeled to delineate MOTUs (Tang *et al.*, 2014; Larson *et al.*, 2016).

The ABGD method was conducted on the online server http://wwwabi.snv.jussieu.fr/public/abgd with the default parameters and the Kimura model (K80) of nucleotide substitution. The PTP model was implemented on the server http://species.h-its.org using the best-scoring ML tree constructed with the raxmlGUI, via the aforementioned protocol. The GMYC was implemented using the ultrametric tree based on the Bayesian inference constructed in BEAST, as mentioned above. Tracer 1.6 software (Rambaut *et al.*, 2014) was used to check for chain convergence and the effective sampling size (ESS > 200). The identification of significant clusters was implemented in RStudio software (2016) by using the *splits* package (Ezard, 2009). K2P distance based on *coI* sequences within and between principal clusters defined by PTP, ABGD, and GMYC species delimitation methods were obtained with MEGA7.

Results

Data and phylogenetic analyses of mitochondrial DNA sequences

PCR amplification of the D-loop region, with primers H16498 and L D-loop M, resulted in fragments of different sizes, ranging from 550 to 650 bp, approximately (Figure 2). *Serrasalmus maculatus* population from Tocantins River basin presented the smallest fragments, with approximately 550 bp, while individuals from Upper Paraná and Upper Paraguay River basins presented fragments between 580 bp and 650 bp, characterizing a length polymorphism of the D-loop region. The observed size variation for the D-loop region in *S. maculatus* was mainly due to tandem repeats identified at the 5' extremity of the D-loop (H-strand), nearby the $tRNA^{Pro}$ gene, and were exclusive to *S. maculatus* population from Paraná-Paraguay system (Table 2). Thus, the presence or absence of these repetitive regions in the D-loop enabled the characterization of *S. maculatus* populations of Paraná-Paraguay River basins or Tocantins River, respectively. Repeated motifs had 12 bp in size and three to five repetitions (Table 2).

After sequence editing and trimming, partial sequences of *coI* (548 bp), *cytb* (592 bp), and D-loop (369 to 414 bp) were concatenated, resulting in an alignment of 1,509 to 1,554 bp in length. Considering *S. maculatus* and *Serrasalmus* sp. (excluding outgroup), 120 variable sites were identified (30 in *coI*, 28 in *cytb*, and 62 in D-loop), containing 102 parsimony-informative (27 in *coI*, 24 in *cytb*, and 51 in D-loop), besides several indels (insertions/deletions) occurring only in the D-loop region. The nucleotide frequencies in the in-group were A = 26.4%, T = 26.4%, C = 30.8%, and G = 16.4%. Sixteen haplotypes were identified in the concatenated sequences (9 in *coI*, 8 in *cytb*, and 16 in D-loop). Only one haplotype (Hap_12) was common to different specimens of *S. maculatus* (samples SmcTO 65, 69, 70, 73, 74, 80) and *Serrasalmus* sp. (sample SrsTO 60), both from Tocantins River basin, i.e., these species shared a mitochondrial haplotype.

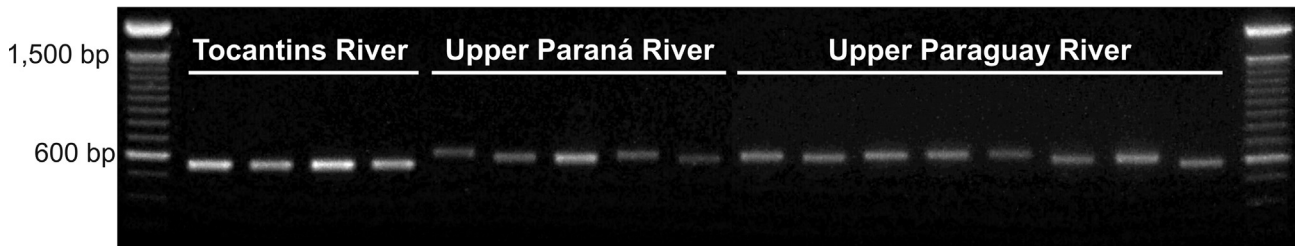

Figure 2 - PCR amplification products of the mitochondrial DNA hypervariable region (control region, D-loop) evidencing the length polymorphism in populations of *Serrasalmus maculatus* from Tocantins, Upper Paraná, and Upper Paraguay River basins. First and last columns contain the standard molecular size ladder 100 bp (Invitrogen).

Table 2 - Tandem repeats observed on the D-loop mitochondrial region of *Serrasalmus maculatus* populations from the Upper Paraguay (SmcPY) and Upper Paraná (SmcPR) River basins. Letters in bold highlight differences among specimens.

Sample	Repetitive motifs	Number of repetitions	Length of Tandem Repetition Region
SmcPY51, 52, 58	GGC**A**CCCCACA**T**	4	48 bp
SmcPY53, 61	GGC**A**CCCCACA**T**	5	60 bp
SmcPY55	GGC**A**CCCCACA**C**	5	60 bp
SmcPY69	GGC**A**CCCCACA**T**	3	36 bp
SmcPR12, 04	GGC**G**CCCCACA**T**	5	60 bp
SmcPR08, 15	GGC**G**CCCCACA**T**	4	48 bp

Most of the observed genetic variability (88.96%) was found among populations of *S. maculatus* and *Serrasalmus* sp. from the Upper Paraná, Upper Paraguay, and Tocantins River basins, according to results of AMOVA (Table 3). Fixation indices F_{ST} revealed that the three populations of *S. maculatus* differed from each other (Table 4). It is evident that both *S. maculatus* and *Serrasalmus* sp. from Tocantins River basin are distinct in comparison to *S. maculatus* populations from Upper Paraná and Upper Paraguay River basins due to values of F_{ST} higher than 0.879. On the other hand, *Serrasalmus* sp. was not differentiated from *S. maculatus* population from the Tocantins River, with $F_{ST} = 0.097$.

ML and BA phylogenetic trees were congruent regarding the formation of the haplogroups of *S. maculatus* and *Serrasalmus* sp. based on the concatenated sequences (Figure 3) and each of the single gene/region analyses (Figure S1). Most of the results indicated *S. maculatus*, as well as *Serrasalmus* sp., as a monophyletic group, except in the BA analyses using the concatenated sequences and the D-loop region. The clustering of *S. maculatus* specimens according to their localities was supported by high bootstrap and posterior probability values in the ML and BA analysis, respectively. Genetic differentiation between populations from Paraná-Paraguay system and Tocantins River was evident. *Serrasalmus maculatus* populations from the Upper Paraná River basin and the Upper Paraguay River were also separated in the dendrograms, but to a lesser degree. Specimens of *Serrasalmus* sp. were grouped in the same clade of *S. maculatus* from Tocantins River, with no genetic differentiation among these species, since there is haplotype sharing as previously described.

Table 3 - Analysis of molecular variance (AMOVA) for populations of *Serrasalmus maculatus* and *Serrasalmus* sp. from Paraná-Paraguay and Tocantins River basins.

Variation source	df	Sum of squares	Components of variation	Percentage of variation
Between populations	3	646.977	39.299	88.96*
Within populations	18	87.750	4.875	11.04
Total	21	734.727	44.174	
Fixation index (F_{ST}) = 0.890				

*$P < 0.01$

Table 4 - Fixation index F_{ST} (below diagonal) obtained by the concatenated mitochondrial sequences (cytochrome c Oxidase I, cytochrome b and D-loop) and mean values of K2P distances obtained from the Cytochrome c Oxidase I gene between and within (above and diagonal) populations of *Serrasalmus maculatus* and *Serrasalmus* sp. from the Upper Paraná (PR), Upper Paraguay (PY) and Tocantins (TO) River basins.

	1	**2**	**3**	**4**
1. *S. maculatus* (PY)	0.34%	1.73%	3.77%	3.77%
2. *S. maculatus* (PR)	0.676*	0.12%	4.66%	4.66%
3. *S. maculatus* (TO)	0.912*	0.953*	0.00%	0.00%
4. *Serrasalmus* sp. (TO)	0.879*	0.923*	0.097^{ns}	0.00%

*$P < 0.05$; $^{ns}P > 0.05$

Species delimitation

Results of ABGD, PTP, and GMYC obtained with *coI* sequences are summarized in Figure 4. Both ABGD and PTP methods resulted in the delimitation of two MOTUs, whereas GMYC recovered three MOTUs. In the first case, groups were defined as: (1) specimens of *S. maculatus* and *Serrasalmus* sp. from Tocatins River basin, and (2) specimens of *S. maculatus* from the Paraná-Paraguay system. The three MOTUs defined by GMYC approach included: (1) specimens of *S. maculatus* and *Serrasalmus* sp. from the Tocantins River, (2) specimens of *S. maculatus* from the Upper Paraná River, and (3) *S. maculatus* from the Upper Paraguay River.

The mean values of inferred genetic distance based on the K2P model were also sufficient to discriminate populations (Table 4). Within each population, these values were low (ranging from 0 to 0.34%), but at least five times higher when comparing the genetic distances between populations (from 1.73 to 4.66%; see Table 4). Populations from the Upper Paraná River and the Upper Paraguay River were different based on low K2P distance (1.73%). However, when comparing populations of *S. maculatus* from Paraná-Paraguay system with those from Tocantins River, the genetic differentiation was higher, ranging from 3.77 to 4.66%. The same values were observed between *Serrasalmus* sp. and *S. maculatus* from Paraná-Paraguay system. Moreover, no genetic differentiation was detected between *Serrasalmus* sp. and *S. maculatus* both from Tocantins River (0.00%) (Table 4).

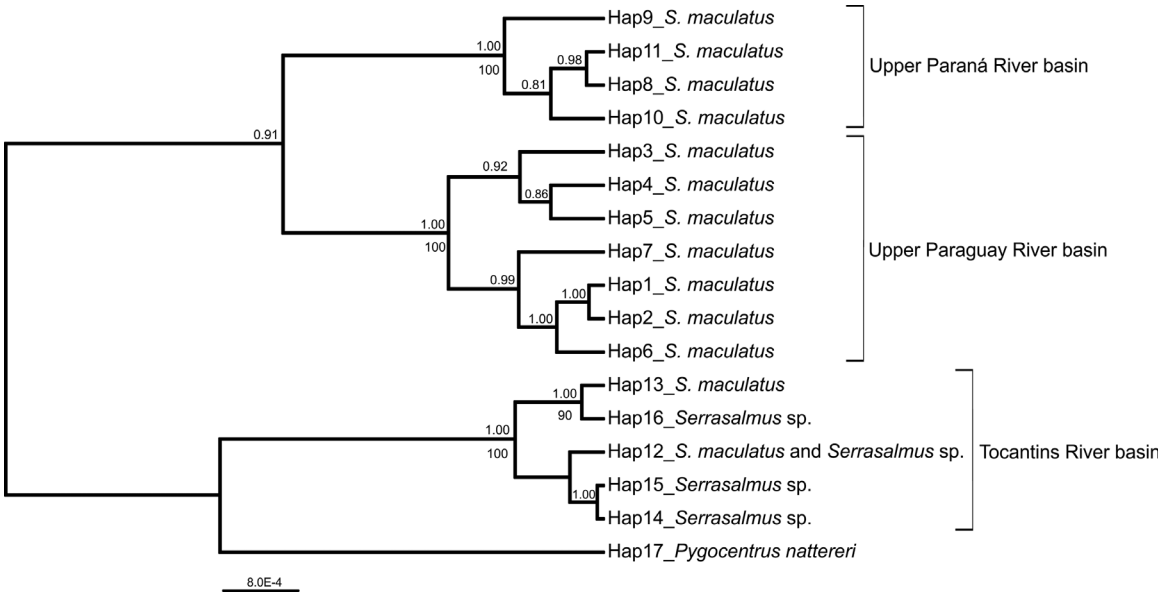

Figure 3 - Bayesian phylogenetic tree for *Serrasalmus maculatus* and *Serrasalmus* sp. from Paraná-Paraguay and Tocantins River basins based on concatenated nucleotide sequences of the mitochondrial regions cytochrome c oxidase I, cytochrome b and D-loop. Values near branches indicate Bayesian (posterior probability, PP; above) and maximum likelihood (bootstrap; below) support values for each node. Sixteen haplotypes (Hap) were recovered for *Serrasalmus* specimens. *Pygocentrus nattereri* (Hap17) was included as outgroup.

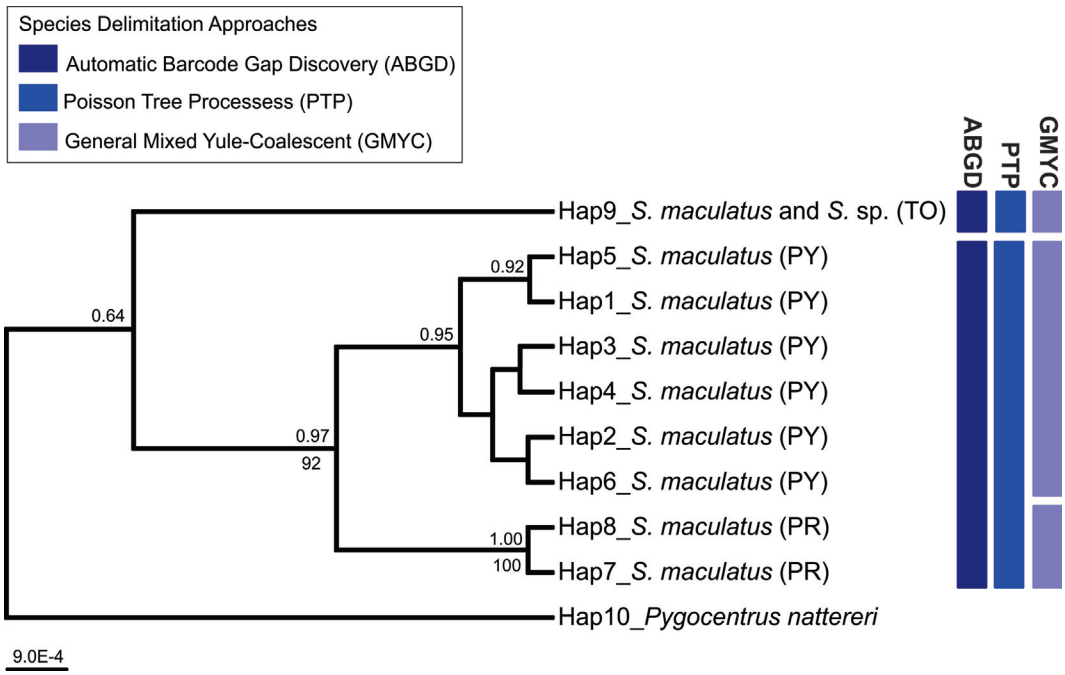

Figure 4 - Species delimitation analyses based on the cytochrome c oxidase I (*coI*) sequences of *Serrasalmus maculatus* and *Serrasalmus* sp. from Upper Paraná (PR), Upper Paraguay (PY), and Tocantins (TO) River basins, using Automatic Barcode Gap Discovery (ABGD), Poisson Tree Processes (PTP), and General Mixed Yule Coalescent (GMYC) methods. Bayesian (posterior probability, PP; above) and maximum likelihood (bootstrap; below) support values for each node in the Bayesian phylogenetic *coI* tree. Nine haplotypes (Hap) were recovered for *Serrasalmus* specimens. *Pygocentrus nattereri* (Hap10) was used as outgroup.

Discussion

Phylogenetic analyses

Serrasalmus maculatus is a piranha species broadly distributed throughout the Amazon and Paraná-Paraguay rivers. In addition to this extensive geographic distribution, the species presents pronounced karyotypic variability (Cestari and Galetti Jr, 1992; Martins-Santos *et al.*, 1994; Nakayama *et al.*, 2000; Centofante *et al.*, 2002; Nakayama *et al.*, 2002). The cytotypes found in this species are associated with

their hydrographic basins, characterizing differences among populations, which allows the assumption that *S. maculatus* may actually comprise a complex of cryptic species (Cestari and Galetti Jr, 1992; Nakayama *et al.*, 2000; Hubert *et al.*, 2006). Similarly, our phylogenetic analyses provided further evidence that *S. maculatus* includes morphologically cryptic diversity.

Our results were also in agreement with the karyotype differences previously reported between populations of *S. maculatus* from the Amazon basin and Paraná-Paraguay system (Nakayama *et al.*, 2000, 2002; Centofante *et al.*, 2002). The mitochondrial molecular markers used in this study were efficient in recognizing the wide genetic differentiation between *S. maculatus* populations of Tocantins River and Paraná-Paraguay system. Genetic distance (K2P), F_{ST} values, and species delimitation results were consistent in order to demonstrate the genetic differentiation of these two populations.

The genetic differentiation between populations of *S. maculatus* from the Tocantins River and the Paraná-Paraguay system was also revealed by characteristics found specifically in the mitochondrial DNA control region (D-loop). In this segment, a complex pattern of variation involving several indels and tandem repeats were also identified. The majority of indels enabled the characterization of populations, but the sharper difference was observe when *S. maculatus* from Tocantins River was compared with the population of the Paraná-Paraguay system (data not presented). Tandem repetitions were also identified in the D-loop sequences, evidencing again the distinctiveness of the populations, since these repetitions were present exclusively in *S. maculatu*s from the Paraná-Paraguay system. Ortí *et al.* (2008) also detected the same type of repetitions in a specimen of *S. maculatus* from the Uruguay River. These tandem repeats typically occur at the 5' or 3' extremities of the D-loop, where replication of the mtDNA begins or ends, respectively (Nesbo *et al.*, 1999), and may be related to size variations within and between individuals or species (White and Martin, 2009).

Our results indicate that there is a limited connectivity between populations of *S. maculatus* of the Tocantins River and Paraná-Paraguay system, suggesting that they are evolutionarily independent lineages. The isolation of *S. maculatus* populations may have started with the formation of the Amazon and Paraná-Paraguay basins, 10 million years ago (Ma) (Hubert and Renno, 2006), by vicariance. However, Lundberg *et al.* (1998) reported that headwater catchments of the Paraná system by the Amazon system continued even after the establishment of these two basins. Montoya-Burgos (2003) and Hubert and Renno (2006) also identified probable routes of dispersion between the Madeira and Guaporé River basins and the Paraguay River. It is possible that presently there still exists a communication between basins during rainy periods, with a consequent exchange of faunas in the region. Several studies confirm the possibility of connection between rivers of the Amazon and Paraná-Paraguay basins (Garda and Cannatella, 2007; Antunes *et al.*, 2010; Aquino and Schaefer, 2010). Consequently, both dispersion and vicariance events may have influenced the differentiation of *S. maculatus* populations. According to the values of genetic divergence observed, it seems more plausible that populations have interrupted gene flow at a time following the separation of these two hydrographic basins.

The difference between *S. maculatus* populations of the Upper Paraguay River and Upper Paraná River was observed both at cytogenetic (Cestari and Galetti Jr, 1992) and molecular levels (present study). Additionally, Machado *et al.* (2018) refuted the monophyly of *S. maculatus* and reported that the population of *S. maculatus* from the upper Paraná River formed a distinct lineage from the population of the lower Paraná River. Similarly, in the present study, the two populations were genetically differentiated from each other; nevertheless, with relatively low values. In addition, the presence of regions containing tandem repeats in the D-loop sequences, presented only in the *S. maculatus* specimens of the Upper Paraná and Upper Paraguay Rivers, suggests that these populations had a common origin. Previous to the construction of the Itaipu reservoir in 1982, the Upper Paraná River was isolated from the rest of the Paraná-Paraguay system by the Sete Quedas falls, a natural geographic barrier that prevented the free dispersion of fish, mainly upstream (Menezes, 1972). Although, considering the downstream course, migration of individuals may occur with some frequency. Therefore, two events could explain the low values of genetic differentiation observed among the *S. maculatus* populations from the Upper Paraná River and the Upper Paraguay River: eventual downstream migrations and the incorporation of the subpopulation from the Itaipu region to the Upper Paraná River population.

Serrasalmus species occurring in the Tocantins River basin (*S. maculatus* and *Serrasalmus* sp.; Lucinda *et al.*, 2007) presented no significant difference based on the mtDNA sequences investigated in this study. Although *S. maculatus* and *Serrasalmus* sp. from Tocantins River are morphologically different from each other, mainly considering the color pattern, both species demonstrated shared mitochondrial haplotype, resulting in a single clade in dendrograms; besides, they presented low F_{ST} index and no K2P genetic distance, indicating that the taxonomic validity of these species in the Tocantins River basin should be revised. One possibility would be that lineages did not have sufficient time to reach the condition of evolutionarily independent entities. Alternatively, hybridization events, with *S. maculatus* as the maternal parent, may be promoting genetic homogenization of the two populations; or even, the combination of these factors.

Nuclear molecular markers should be used to indicate the occurrence of interspecific hybrids between *S. maculatus* and *Serrasalmus* sp. in the Tocantins River basin. Although there are no reports on the occurrence of interspecific natural hybrids of piranhas in the literature, Hubert *et al.* (2008) obtained evidence of former introgressions followed by hybridization between *Serrasalmus* sp. and *S. compressus*, and between *Serrasalmus* sp. and *S. hollandi*, sympatric species of the Upper Madeira River. Under these circumstances, a

taxonomic revision of *Serrasalmus* sp. and *S. maculatus* from the Tocantins River basin is suggested.

Species delimitation

The three analyses of species delimitation based on *coI* sequences (ABGD, PTP, and GMYC) presented two possible scenarios for limits of species in *S. maculatus* and *Serrasalmus* sp.: i) each of the three populations comprises a different species or, ii) at least the population from Tocantins River and the population from Paraná-Paraguay system belonging to two distinct species. Although the methods used to delimit species did not achieve consensus for the numbers of MOTUs (ABGD and PTP = 2 MOTUs; GMYC = 3 MOTUs), the results are not in conflict, since the divisions among MOTUs defined by ABGD and PTP were also recovered by GMYC.

Differences in mitochondrial nucleotide sequences have been used for distinguishing species for more than 30 years (Avise, 2004). Hebert *et al.* (2003) suggested that a region of the *coI* gene is appropriate as a tool in the identification of animals at the species level (DNA barcode). A standard sequence threshold of 10× the mean intraspecific variation for the group under study was proposed (Hebert *et al.*, 2004). Consequently, a divergence 10× greater than the mean of the intraspecific variation would be indicative of a new species. With this limit applied to our data (0.15% average intraspecific variation), the 10× threshold (1.53%) would establish each of the three *S. maculatus* populations analyzed as distinct species (see Table 4).

Phylogenetic analysis, in combination with estimates of species delimitation, suggests that *S. maculatus* includes morphologically cryptic diversity. The data obtained in this study strongly indicate that the populations currently identified as *S. maculatus* from the Paraná-Paraguay and Tocantins River basins are different species. The differences highlight that populations of *S. maculatus* remained isolated geographically long enough for speciation to occur. Thus, there is strong evidence that *S. maculatus* constitutes a complex of at least two morphologically similar species in the hydrographic basins of the Tocantins and Paraná-Paraguay Rivers. Since the type-locality of *S. maculatus* is the Guaporé River basin (Jégu and Santos, 2001), it would be necessary the inclusion of specimens of this basin in future studies to determine if the denomination *S. maculatus* should be restricted to the Tocantins River basin or to the Paraná-Paraguay system. Therefore, *S. maculatus* should be treated as a complex of species distributed in several regions of South America. Increasing sampling efforts of populations from other hydrographic basins could reveal other haplogroups corresponding to new species. It is possible that populations currently identified as *S. maculatus* comprise a complex of several species.

Acknowledgments

The authors are grateful to Nupélia/UEM for technical, scientific, and financial support and to E. K. Okada for sampling carried out in the Manso River. T. S. Bignotto thanks CAPES for the scholarship grants. Part of this study was supported by grants from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-PELD).

Conflict of interest

The authors declare no conflict of interest.

Author contributions

TSB, AJP, SMAPP, CSA conceived and designed the study; TSB, AJP, SMAPP, CSA collected the samples; TSB, AJP, TCM, VNG, IJO conducted the experiments and analyzed the data; TSB, AJP, SMAPP, TCM, VNG, IJO have contributed to the manuscript writing; all authors read and approved the final version.

References

Antunes RSP, Gomes VN, Prioli SMAP, Prioli RA, Julio Jr. HF, Prioli LM, Agostinho CS and Prioli AJ (2010) Molecular characterization and phylogenetic relationships among species of the genus *Brycon* (Characiformes: Characidae) from four hydrographic basins in Brazil. Genet Mol Biol 9:674-684.

Aquino AE and Schaefer SA (2010) Systematics of the genus *Hypoptopoma* Günther, 1868 (Siluriformes, Loricariidae). Bull Am Mus Nat Hist 336:1-110.

Avise JC (2004) Molecular markers, natural history and evolution. 2nd edition. Sinauer Associates Inc. Publishers, Sunderland, 541 pp.

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK and Das I (2007) Cryptic species as a window on diversity and conservation. Trends Ecol Evol 22:148-155.

Centofante L, Porto JIR and Feldberg E (2002) Chromosomal polymorphism in *Serrasalmus spilopleura* Kner, 1858 (Characidae, Serrasalminae) from Central Amazon basin. Caryologia 55:37-45.

Cestari MM and Galetti Jr PM (1992) Chromosome studies of *Serrasalmus spilopleura* (Characidae, Serrasalminae) from the Paraná-Paraguay rivers: Evolutionary and cytotaxonomic considerations. Copeia 1:108-112.

Cronin MA, Spearman WJ, Wilmot RL, Patton JC and Bickham JW (1993) Mitochondrial DNA variation in chinook (*Oncorhynchus tshawytscha*) and chum salmon (*O. keta*) detected by restriction enzyme analysis of Polymerase Chain Reaction (PCR) products. Can J Fish Aquat Sci 50:708-715.

Drummond AJ, Suchard MA, Xie D and Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29:1969-1973.

Excoffier L, Laval G and Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evol Bioinform Online 1:47-50.

Fink WL and Machado-Allison A (1992) Three new species of piranhas from Brazil and Venezuela (Teleostei: Characiformes). Ichthyol Explor Freshwaters 3:55-71.

Fink WL and Machado-Allison A (2001) *Serrasalmus hastatus*, a new species of piranha from Brazil, with comments on *Serrasalmus altuvei* and *Serrasalmus compressus* (Teleostei, Characiformes). Occas Pap Mus Zool Univ Mich 730:1-18.

Frankham R, Ballou JD, Dudash MR, Eldridge MDB, Fenster CB, Lacy RC, Mendelson III JR, Porton IJ, Ralls K and Ryder OA (2012) Implications of different species concepts for conserving biodiversity. Biol Conserv 153:25-31.

Freeman B, Nico LG, Osentoski M, Jelks HL and Collins TM (2007) Molecular systematic of Serrasalmidae: deciphering the iden-

tities of piranha species and unraveling their evolutionary histories. Zootaxa 1484:1-38.

Garda A and Cannatella DC (2007) Phylogeny and biogeography of paradoxical frogs (Anura, Hylidae, Pseudae) inferred from 12S and 16S mitochondrial DNA. Mol Phylogenet Evol 44:104-114.

Griffiths SP (2000) The use of clove oil as an anaesthetic and method for sampling intertidal rockpool fishes. J Fish Biol 57:1453–1464.

Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95-98.

Hebert PDN, Ratnasingham S and Waard JR (2003) Barcoding animal life: Cytochrome *c* oxidase subunit 1 divergences among closely related species. Proc Biol Sci 270:596-599.

Hebert PDN, Stoeckle MY, Zemlak TS and Francis CM (2004) Identification of birds through DNA barcodes. PLoS Biol 2:1657-1663.

Hernández CL, Ortega-Lara A, Sánchez-Garcés GC and Alford MH (2015) Genetic and morphometric evidence for the recognition of several recently synonymized species of Trans-Andean *Rhamdia* (Pisces: Siluriformes: Heptapteridae). Copeia 103:563-579.

Hubert N and Renno JF (2006) Historical biogeography of South American freshwater fishes. J Biogeogr 33:1414-1436.

Hubert N, Duponchelle F, Nuñez J, Riveira R and Renno JF (2006) Evidence of reproductive isolation among closely related sympatric species of *Serrasalmus* (Ostariophysii, Characidae) from the Upper Madeira River, Amazon, Bolivia. J Fish Biol 69:31-51.

Hubert N, Torrico JP, Bonhomme F and Renno, JF (2008) Species polyphyly and mtDNA introgression among three *Serrasalmus* sister-species. Mol Phylogenet Evol 46:375-381.

Ivanova NV, Zemlak TS, Hanner RH and Hebert RH (2007) Universal primer cocktails for fish DNA barcoding. Mol Ecol Notes 7:555-548.

Jégu M and Santos GM (2001) Mise au point à propos de *Serrasalmus spilopleura* Kner, 1858 et réhabilitation de *S. maculatus* Kner, 1858 (Characidae: Serrasalminae). Cybium 25:119-143.

Kocher TD, Thomas WK, Meyer A, Edwards SV, Paabo S, Villablanca FX and Wilson AC (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc Natl Acad Sci U S A 86:6196-6200.

Kumar S, Stecher G and Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870-1874.

Lanfear R, Calcott B, Ho SYW and Guindon S (2012) PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol 29:1695–1701.

Larson ER, Castelin M, Williams BW, Olden JD and Abbott CL (2016) Phylogenetic species delimitation for crayfishes of the genus *Pacifastacus*. PeerJ 4:e1915.

Lucinda PHF, Freitas IS, Soares AB, Marques EE, Agostinho CS and Oliveira RJ (2007) Fish, Lajeado Reservoir, rio Tocantins drainage, State of Tocantins, Brazil. Check List 3:70-83.

Lundberg JG, Marshall LG, Guerrero J, Horton B, Malabarba MCSL and Wesselingh F (1998) The stage for Neotropical fish diversification: a history of tropical South American Rivers. In: Malabarba LR, Reis RE, Vari RP, Lucena ZM and Lucena CAS (eds) Phylogeny and Classification of Neotropical Fishes. EdiPUCRS, Porto Alegre. pp 13-48.

Machado-Allison A and Fink WL (1995a) Sinopsis de las especies de la subfamilia Serrasalminae presentes en la cuenca del Orinoco: Claves, diagnosis e ilustraciones. Universidad Central de Venezuela, Caracas.

Machado-Allison A and Fink WL (1995b) Los peces Caribes de Venezuela: Diagnosis, claves, aspectos ecologicos y evolutivos. Serie peces de Venezuela. Universidade Central de Venezuela, Caracas, 90 p.

Machado V, Collins RA, Ota RP, Andrade MC, Farias IP and Hrbek T (2018) One thousand DNA barcodes of piranhas and pacus reveal geographic structure and unrecognised diversity in the Amazon. Sci Rep 8:8387.

Martins-Santos IC, Julio Jr HF and Santos SJ (1994) Chromosome study of two species of the genus *Serrasalmus* (Characidae, Serrasalminae) from the Paraná River. Cytologia 59:175-181.

Menezes NA (1972) Distribuição e origem da fauna de peixes de água doce das grandes bacias fluviais do Brasil. In: Comissão Internacional da Bacia Paraná-Uruguai (ed) Poluição e piscicultura. Universidade de São Paulo, Faculdade de Saúde Pública; Secretaria da Agricultura, Instituto de Pesca, São Paulo, pp 73-78.

Meyer A, Kocher TD, Basasibwaki P and Wilson AC (1990) Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. Nature 347:550-553.

Montoya-Burgos JI (2003) Historical biogeography of the catfish genus *Hypostomus* (Siluriformes: Loricariidae), with implications on the diversification of Neotropical ichthyofauna. Mol Ecol 12:1855-1867.

Nakayama CM, Porto JIR and Feldberg E (2000) Ocorrência de dois citótipos em *Serrasalmus spilopleura* Kner, 1858 (Characiformes, Serrasalmidae) da região de confluência dos rios Negro e Solimões, Amazonas, Brasil. Acta Amaz 30:149-154.

Nakayama CM, Porto JIR and Feldberg E (2002) A comparative cytogenetic study of five piranha species (*Serrasalmus*, Serrasalminae) from the Amazon basin. Genetica 114:231-236.

Nakayama CM, Feldberg E and Bertollo LAC (2012) Karyotype differentiation and cytotaxonomic considerations in species of Serrasalmidae (Characiformes) from the Amazon basin. Neotrop Ichthyol 10:53-58.

Nesbo CL, Fossheim T, Vollestad LA and Jakobsen KS (1999) Genetic divergence and phylogeographic relationships among European perch (*Perca fluviatilis*) populations reflect glacial refugia and postglacial colonization. Mol Ecol 8:1387-1404.

Oliveira AV, Prioli AJ, Prioli SMAP, Bignotto TS, Júlio Jr HF, Carrer H, Agostinho CS and Prioli LM (2006) Genetic diversity of invasive and native *Cichla* (Pisces: Perciformes) populations in Brazil with evidence of interspecific hybridization. J Fish Biol 69:260-277.

Ortí G, Sivasundar A, Dietz K and Jégu M (2008) Phylogeny of the Serrasalmidae (Characiformes) based on mitochondrial DNA sequences. Genet Mol Biol 31:343-351.

Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Durand P, Hazell S, Kamoun S, Sumlinw D and Vogler AP (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst Biol 55:595-609.

Prioli SMAP, Prioli AJ, Júlio Jr HF, Pavanelli CS, Oliveira AV, Carrer H, Carraro DM and Prioli LM (2002) Identification of *Astyanax altiparanae* in the Iguaçu River, Brazil, based on mitochondrial DNA and RAPD markers. Genet Mol Biol 25:421-430.

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

Rosenthal A, Coutelle O and Craxton M (1993) Large-scale of DNA sequencing templates by microtitre format PCR. Nucleic Acids Res 21:173-174.

Rozas J, Ferrer-Mata A, Sánchez-Delbarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE and Sánchez-Gracia A (2017) DnaSP 6: DNA Sequence polymorphism analysis of large datasets. Mol Biol Evol 34:3299-3302.

Sievers F, Wilm A, Dineen DG, Gibson TJ, Karplus K, Li W, Lopez R, Mcwilliam H, Remmert M, Söding J *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539.

Silvestro D and Michalak I (2012) RaxmlGUI: A graphical front-end for RAxML. Org Divers Evol 12:335–337.

Tang CQ, Humpreys AM, Fontaneto D and Barraclough TG (2014) Effects of phylogenetic reconstruction method on the robustness of species delimitation using single-locus data. Methods Ecol Evol 5:1086-1094.

Ward RD, Zemlak SZ, Innes BH, Last PR and Hebert PDN (2005) DNA barcoding Australia's fish species. Philos Trans R Soc Lond B Biol Sci 360:1847-1857.

White MM and Martin HR (2009) Structure and conservation of tandem repeats in the mitochondrial DNA control region of the least brook lamprey (*Lampetra aepyptera*). J Mol Evol 68:715-723.

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

Internet Resources

Ezard T, Fujisawa T and Barraclough TG (2009) SPLITS: SPecies' LImits by Threshold Statistics, R package version 1.0-18/r45, http://R-Forge.R-project.org/projects/splits/ (accessed 10 September 2017).

Rambaut A, Suchard MA, Xie D and Drummond AJ (2014) Tracer v1.6, http://tree.bio.ed.ac.uk/software/tracer/ (accessed 10 September 2017).

Rstudio Team (2016) RStudio: Integrated Development for R. RStudio, http://www.rstudio.com/ (accessed 10 September 2017)

Supplementary material

The following online material is available for this article:
Figure S1 - Individual Bayesian phylogenetic trees for *Serrasalmus maculatus* and *Serrasalmus* sp. from Upper Paraná (PR), Upper Paraguay (PY), and Tocantins (TO) River basins based on nucleotide sequences of the mitochondrial regions Cytochrome c oxidase I (*coI*), Cytochrome b (*cytb*) and control region (D-loop).

Associate Editor: Fabrício Rodrigues dos Santos