

## Molecular Phylogeny Inferred from the Concatenated Genes of Two Neotropical Catfish Species and Implications for Conservation

Josi M Ponzetto<sup>1\*</sup>, Anderson L Alves<sup>2</sup>, Eduardo S Varela<sup>3</sup>, Luciana CV Villela<sup>3</sup>, Alexandre R Caetano<sup>4</sup> and Eduardo Leonardecz<sup>1</sup>

<sup>1</sup>Postgraduate Program in Evolutionary Genetics and Molecular Biology, Centre for Biological and Health Sciences, Federal University of São Carlos (UFSCar), São Carlos, SP, Brazil

<sup>2</sup>Embrapa Products and Markets, SPM, Campinas, SP, Brazil

<sup>3</sup>Embrapa Fishing and Aquaculture, CNPASA, Palmas, TO, Brazil

<sup>4</sup>Embrapa Genetic Resources and Biotechnology, CENARGEN, Brasília, DF, Brazil

\*Corresponding author: Ponzetto JM, Postgraduate Program in Evolutionary Genetics and Molecular Biology, Centre for Biological and Health Sciences, Federal University of São Carlos (UFSCar), São Carlos, SP, Brazil, Tel: (+55) 19 98316-0662; E-mail: [josiponzetto@gmail.com](mailto:josiponzetto@gmail.com)

Received date: February 02, 2017; Accepted date: February 28, 2017; Published date: March 10, 2017

Copyright: © 2017 Ponzetto JM, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### Abstract

The Neotropics host the most diverse ichthyofauna in the world, with catfish species forming one of the most diverse groups in the region. Nuclear (RAG1) and mitochondrial (ATPase and Cytb) markers were analyzed to identify genetic variability in populations of *Pseudoplatystoma reticulatum* and *Pseudoplatystoma corruscans* from the La Plata and Sao Francisco Basins. Bayesian topology identified the division of *P. corruscans* into two main clades. One of these clades was formed of samples from the Sao Francisco Basin and the other was formed of samples from the Parana+Paraguay Basins. *P. reticulatum* was grouped together without any clear geographic or taxonomic patterns in Bayesian topology. While only a few common nuclear haplotypes were widely identified in both species, there was great variability in the mitochondrial sequences. The genetic and geographical distance correlations were tested using the Mantel permutation, which detected no significant relationships. The results of the present study suggest a panmictic population for both species (excluding *P. corruscans* in the Sao Francisco Basin, which is suggested as a new species), with the greatest diversity concentrated in the region covered by the Pantanal biome, and the lowest diversity in Mogi Guacu, in the Parana Basin. These findings support the establishment of public conservation policies and provide information regarding genetic diversity and population differentiation patterns for these ecological and economically important species.

**Keywords** Genetic diversity; *Pseudoplatystoma*; DNA sequences; Conservation

### Introduction

The Neotropics host the most diverse ichthyofauna in the world, including 46% of all freshwater fish species [1,2]. Catfish are one of the most diverse groups of such freshwater species, and include the Pimelodidae family [3]. Within this family, the main genera are *Pseudoplatystoma*, *Brachyplatystoma*, *Pharactocephalus*, and *Zungaro* [4]. *Pseudoplatystoma* Bleeker, 1862, includes species known to perform complex migrations between rivers, lakes and floodplains [5-7]. According to Buitrago-Suárez and Burr [8] this genus is endemic to the South American continent, and occurs widely in important watershed rivers.

*Pseudoplatystoma corruscans* Spix & Agassiz, 1829, and *Pseudoplatystoma reticulatum* Eigenmann & Eigenmann, 1889, are among the largest catfish in South America [9]. They occur in sympatry in the La Plata Basin, which includes the basins of the Paraguay, Parana and Uruguay Rivers (covering Brazil, Argentina, Paraguay and Uruguay) [3,8]. In other areas only one species occurs, such as *P. reticulatum* in the Amazon Basin [3,8] and *P. corruscans* in the São Francisco River Basin. Although these migratory species have a high commercial value, and are important in the river basins where they occur, basic biological information about them remains scarce [8,10].

There is a lack of genetic studies, with most focusing on the taxonomy and systematic data of the genus [8,10-13]. Despite this lack of information, some studies have identified genetic variability data for populations of species of this genus [10,14,15]. Homing behavior has been suggested for *P. corruscans* [15] and *P. reticulatum* [14], from the La Plata basin, which is evidenced by the strong genetic structure of these populations. Batista and Alves-Gomes [16], studying the *Brachyplatystoma rousseauxii* catfish species collected from the Amazon River, also suggested the possibility of homing movements. However, other studies of migratory fish in South America have suggested that different basins are inhabited by panmictic populations [17,18].

Genetic variation is essential for the potential survival of a species [19]. Verifying the genetic structure of natural populations through important tools such as genetics allows the distribution of genetic variability to be identified, enabling the effective management and conservation of wild populations [19-23].

According to Barthem and Goulding [24] many stocks of these catfish have decreased in recent years. The overfishing of catfish such as *P. corruscans* and *P. reticulatum*, combined with the pollution of aquatic environments, dam building and other harmful events, are important factors in the decline of wild populations [6,25]. The present study analyzed nuclear and mitochondrial markers to identify the genetic variability of distinct populations of *P. reticulatum* and *P. corruscans* from the La Plata and Sao Francisco Basins.

## Materials and Methods

### Sample collection and DNA amplification

*Pseudoplatystoma* specimens were collected between November and February, in the breeding season. The fish were caught with gillnets at eight distinct locations in the Parana-Paraguay Basin and one site in the Sao Francisco Basin, distributed across four Brazilian states (Table 1, Figure S1, Supplementary material), and were collected for tissue collection and subsequently released back into the river. All samples were identified in accordance with Buitrago Suarez and Burr [8]. The numbers of samples genetically analyzed were 74 for ATPase8/6 (ATPase), 72 for Cytochrome b (Cytb), and 58 for RAG1Pse (RAG1)

for *Pseudoplatystoma corruscans* (Table 1); and 60 for ATP, 55 for Cyt b, and 43 for RAG1 for *Pseudoplatystoma reticulatum* (Table 1). In addition, one sample of *Pseudoplatystoma tigrinum* and two of *Pseudoplatystoma punctifer* from the Amazon Basin were analyzed for the Cyt B gene (Table 1). Some sequences from GenBank were included in the analyses to confirm the monophyly of *P. corruscans* and *P. reticulatum*. These were *Pseudoplatystoma fasciatum* (GU593111 and GU593110), *Pseudoplatystoma magdaleniatum* (GU593155), *Pseudoplatystoma orinocoense* (GU593133 and GU593134), *P. metaense* (GU593154), *Rhamdia laticauda* (AF287412, ATPase 8/6; AF287456, Cytb, used as outgroup) and *Zungaro* sp. (GU593241, used as outgroup).

Locality	City/State	Whatershed	<i>P. corruscans</i> (Pintado) <sup>a</sup>			<i>P. reticulatum</i> (Cachara) <sup>a</sup>			<i>P. punctifer</i> <sup>a</sup>	<i>P. tigrinum</i> <sup>a</sup>
			Number of sequenced specimen/ ATPase code	Number of sequenced specimen/ Cytb code	Number of sequenced specimen/ RAG1 code	Number of sequenced specimen/ ATPase code	Number of sequenced specimen/ Cytb code	Number of sequenced specimen/ RAG1 code	Number of sequenced specimen/ Cytb code	Number of sequenced specimen/ Cytb code
Miranda river	Miranda/MS	BAP	8/A(n)MP	9/C(n)MP	5/R(n)MP	8/A(n)MC	6/C(n)MC	5/R(n)MC	-	-
Aquidauana river	Aquidauana/MS	BAP	11/A(n)AP	10/C(n)AP	9R(n)AP	7/A(n)CA	4/C(n)CA	4/R(n)CA	-	-
Paraguay river	Corumbá/MS	BAP	6/APCor(n)	4/CPCOR(n)	4/RPcor(n)	16/ACcor(n)	17/CCOR(n)	11/RCCor(n)	-	-
Paraguay river	Caceres/MT	BAP	10/APMT(n)	10/CPMT(n)	9/RPMT(n)	8/ACMT(n)	6/CCMT(n)	4/RCCMT(n)	-	-
Cuiabazinho river	Cuiaba/MT	BAP	7/APCZ(n)	8/CPCZ(n)	6/RPCZ(n)	11/ACCZ(n)	10/CCZ(n)	10/RCCZ(n)	-	-
Mogi Guacu river	Pirassununga/SP	BAPR	3/A(n)PMG	6/C(n)PMG	2/R(n)PMG	3/A(n)CMG	2/C(n)CMG	3/R(n)CMG	-	-
Cuiaba river	Barão de Melgaco/MT	BAP	11/A(n)PBM	14/C(n)PBM	15/R(n)PBM	7/A(n)CBM	10/C(n)CBM	6/R(n)CBM	-	-
Sao Francisco river	Pirapora/MG	BSF	15/ASF(n)	10/CSF(n)	7/RSF(n)	-	-	-	-	-
Ivinhema river	Ivinhema/MS	BAPR	3/A(n)IV	1/C(n)IV	1/R(n)IV	-	-	-	-	-
Ucayali river	Pucallpa-Peru	BA	-	-	-	-	-	-	-	1C(n)PC
Rio Branco river	Boa Vista-RR	BA	-	-	-	-	-	-	1C(n)RB	-
Amazon river	Iquitos-Peru	BA	-	-	-	-	-	-	2C(n)IQ	-
TOTAL AMPLIFIED			74	72	58	60	55	43	3	1

**Table 1:** Places collection, number samples and codes used to identify the specimens of *Pseudoplatystoma* genus sequenced for each marker; <sup>a</sup>Names according to Buitrago-Suárez and Burr; <sup>†</sup>BAP (Paraguay Basin); <sup>‡</sup>BAPR (Parana Basin); <sup>§</sup>BSF (Sao Francisco Basin); <sup>¶</sup>BA (Amazon Basin).

Genomic DNA was isolated from the fins of each specimen, which were preserved in 95% ethanol using the DNeasy® Blood & Tissue Kit (Qiagen), in accordance with the manufacturer's instructions. The

product was then quantified in a NanoDrop 2000 (Thermo Scientific, Wilmington, USA) spectrophotometer at 260 nm. The mitochondrial genes Cytochrome b (1109 bp) and ATPase 8/6 (839 bp), and the

nuclear gene RAG1Pse (770 bp) were analyzed (Table 2). The gene fragments were amplified via PCR (polymerase chain reaction) in a final reaction volume of 30.6 µL, containing 25.0 µL of Mix PCR (Invitrogen)\*, 1.0µL of each primer (10 µM), 3.0 µL of DNA (50 ng) and 0.6 µL of MgCl<sub>2</sub> (50 ng).\* The Mix solution contained Taq DNA Polymerase with Platinum® Taq Antibody 22 U/mL, Tris-HCl 22 mM, KCl 55 mM, MgCl<sub>2</sub> 1.65 mM, dDNTP (dideoxynucleotide triphosphate) 220 µM, stabilizers, glycerol, and blue tracking dye. Polymerase chain reaction (PCR) amplification for the mitochondrial genes was as follows: 94°C for 2 mins, 35 cycles of 94°C for 45 s, annealing temperature (Table 2) for 45 s, 68°C for 90 s, and 72°C for 5 min, while for the nuclear genes 94°C for 5 min, 35 cycles of 94°C for

30 s, annealing temperature (Table 2) for 35 s, 72°C for 40 s, and 72°C for 5 min was applied. The amplified DNA was purified with the Exosap-IT (USB Corporation) enzyme and subsequently sequenced by MacroGen Inc. (Korea) and Beckman Coulter Genomics (France). Sequences were independently aligned using the Clutal W [26] algorithm in Bioedit v 7.1 [27] under default parameters and the alignments were inspected for problems. A concatenated dataset was generated for pairwise alignments using the Geneious v5.5 software package [28]. The mitochondrial DNA sequences were also translated using Geneious v5.5 [28] to check for the unexpected occurrence of stop codons.

Primers	Sequences of primers	Ta (°C)	References
Gludg-F	TGA CTT GAA RAA CCA YCG TTG	55	Palumbi et al. [49]
H15915-R	AA CTG CAG TCA TCT CCG GTT TAC AAG AC		Irwin et al. [50]
L8331-F	AA GCR TYR GCC TTT TAA GC	55	Perdices et al. [51]
H9236-R	GTT AGT GGT CAK GGG CTT GGR TC		
Rag1Pse-F	GCT GGC AGA CAA GTG GAT CT	58	Carvalho-Costa et al. [12]
Rag1Pse-R	AGG CCA TCT AGA AGC CTT GC		

**Table 2:** Sequences of primers used in the samples amplification. Ta: Annealing temperature; F: Forward; R: Reverse.

### Phylogenetic and population analyses

The patterns of the phylogenetic variations of the *Pseudoplatystoma corruscans* and *Pseudoplatystoma reticulatum* populations were analyzed by combining phylogenetic and population analyses. Phylogenetic analyses were estimated separately for each data set (ATPase, Cytb and RAG1) using the distance (neighbor joining, NJ) method, carried out in MEGA v. 6.0 [29], and for the concatenated alignment data set (ATPase+Cytb+RAG1) using the Bayesian inference (BI) method, in MrBayes 3.1.2 [30]. A Bayesian Markov chain Monte Carlo reaction was carried out using MrBayes 3.1.2 [30] on the CIPRES supercomputing cluster [31]. MrBayes was programmed to run for 200 million generations using four independent MCMC chains (nchain=4, representing two parallel runs with two cold and two hot chains each; temperature parameters set to default), with sampling every 1000 steps. Gaps were treated as missing data. The distribution of log likelihood scores was examined to determine stationary for each search and to decide if extra runs were required to achieve convergence, using the Tracer 1.6 program [32]. Initial trees estimated prior to convergence were discarded as part of a burn-in procedure. The posterior distribution and maximum clade credibility tree were constructed in Tree Annotator v1.8.1 (available as part of the BEAST package) [33] from the runs after eliminating 50% of the trees for burn-in to avoid the possibility of the inclusion of random sub-optimal trees. Finally, the trees were visualized using the Fig Tree v.1.4.2 software [34].

Partition Finder v1.1.0 [35] was used to determine the best molecular evolution model and partition scheme for each gene under the Bayesian information criterion (BIC). This implements the relaxed clustering algorithm, with the default setting to check the top 10% of schemes, based on those expected to present the greatest improvement. For *P. corruscans*, a generalized time reversible model [36], with the rate heterogeneity of the remainder modeled by gamma distribution

(GTR+Gamma), was identified as the best molecular evolution model for the first and second genes (ATPase and Cytb, respectively). A F81 model [37] with a proportion of invariable sites and the rate heterogeneity of the remainder modeled by gamma distribution (F81+I+Gamma) was identified as the best model for the molecular evolution for the third gene (RAG1). For *P. reticulatum*, a Hasegawa-Kishono-Yano model [38] with a proportion of invariable sites and the rate heterogeneity of the remainder modeled by gamma distribution (HKY+I+Gamma), was identified as the best molecular evolution model for the three genes. Phylogenetic hypotheses were inferred from the data set with Bayesian inference (B).

In addition to these analyses, independent phylogenetic analysis for *P. corruscans* and all other species of the *Pseudoplatystoma* genus was performed for the Cytb gene using the Bayesian inference (BI) method, in MrBayes 3.1.2 [30] on the CIPRES supercomputing cluster [31]. Partition Finder v1.1.1 [35] was used to determine codon-specific models of molecular evolution for the Cytb gene under the Bayesian information criterion (BIC). A Hasegawa-Kishono-Yano model [38], with a proportion of invariable sites and the rate heterogeneity of the remainder modelled by propinv distribution (HKY+I), was identified as the best molecular evolution model for the first codon position. A generalized time reversible model [36], with the rate heterogeneity of the remainder modelled by gamma distribution (GTR+Gamma), and was identified as the best molecular evolution model for the second codon position. And a Kimura [39] model, with a proportion of invariable sites and the rate heterogeneity of the remainder modelled by propinv distribution (K80+I), was identified as the best molecular evolution model for the second codon position. Phylogenetic hypotheses were also inferred from the data set with Bayesian inference (B). All other parameters were as described above.

All subsequent analyses were performed independently for each gene for the *P. corruscans* and *P. reticulatum* populations. Sequence

divergence (p-uncorrected distance) was assessed using the MEGA v. 6.0 tools [29]. Haplotype and nucleotide diversity were calculated using DNASP v.5.10.01 [40]. We used a median-joining algorithm in Network 4.6.1.1 (<http://fluxus-engineering.com>) to examine the relationships between haplotypes. Population structure and genetic variation were characterized and compared using Arlequin v.3.5 [41]. Analysis of Molecular Variance (AMOVA) was used to assess the population configuration and geographical pattern of the subdivision. To achieve this, the populations were analyzed and grouped in three forms: all populations together in one single group; populations separated into two groups, with Group I composed of the samples from the Paraguay Basin and Group II composed of the populations from the Parana Basin (Mogi Guacu and Ivinhema), with no a population from the Sao Francisco Basin; and the populations separated into two different groups, with Group III consisting of all the samples from the Parana-Paraguay Basins, while group IV was composed of the populations from the Sao Francisco Basin. Global and pairwise estimates of genetic differentiation were examined using FST, incorporating haplotype frequency and relatedness. The correlations of pairwise multilocus FST (significant) genetic and geographical distances were tested using the Mantel permutation procedure [42] and significances were tested with 2,000 permutations.

## Results

### Molecular characterization

A total of 839 bp were analyzed from 74 *Pseudoplatystoma corruscans* individuals to test for the mitochondrial ATPase gene, while for the Cytb gene, 1109 bp were analyzed from 72 *P. corruscans* individuals, and for the RAG1 gene 770 bp were analyzed from 58 *P. corruscans* individuals (Table 1). For the concatenated data set 2718 bp were analyzed from 89 different specimens. A total of 413 variable polymorphic sites were found for the concatenated data set (ATPase+Cytb+RAG1).

A total of 839 bp were analyzed from 60 *Pseudoplatystoma reticulatum* individuals for the mitochondrial ATPase gene. For the Cytb gene, 1109 bp were analyzed from 55 *P. reticulatum* individuals, and for the RAG1 gene 770 bp were analyzed from 43 *P. reticulatum* individuals (Table 1). For the concatenated data set 2718 bp were analyzed from 71 different specimens. A total of 189 variable polymorphic sites were found for the concatenated data set (ATPase+Cytb+RAG1).

Only the RAG1 gene had sequences with gaps, with 11 gaps identified (starting: 194 bp to *P. reticulatum* and 602 bp to *P. corruscans*) when aligned with other *Pseudoplatystoma* sequences.

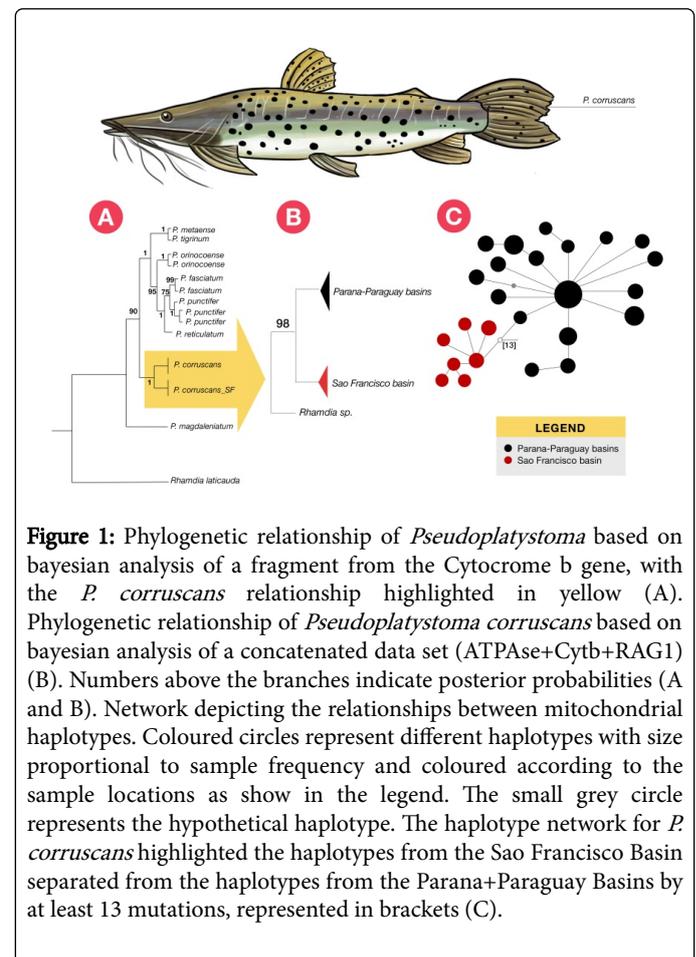
The molecular characterization data of the Cytb gene for *P. punctifer* and *P. tigrinum* are not shown.

### Phylogenetic analysis

Independent analysis of the mitochondrial and nuclear data presented similar topologies to the concatenated data set (Supplementary Figures S2-S4).

The concatenated data set revealed the topology to be partitioned into two main clades (I and II) of *P. corruscans* individuals. Clade I was composed of samples from the Sao Francisco Basin and clade II consisted of the other samples (Parana+Paraguay Basins). The analysis

was supported by high posteriori probability values (Figure 1b, Supplementary figure S5A).



**Figure 1:** Phylogenetic relationship of *Pseudoplatystoma* based on bayesian analysis of a fragment from the Cytocrome b gene, with the *P. corruscans* relationship highlighted in yellow (A). Phylogenetic relationship of *Pseudoplatystoma corruscans* based on bayesian analysis of a concatenated data set (ATPase+Cytb+RAG1) (B). Numbers above the branches indicate posterior probabilities (A and B). Network depicting the relationships between mitochondrial haplotypes. Coloured circles represent different haplotypes with size proportional to sample frequency and coloured according to the sample locations as show in the legend. The small grey circle represents the hypothetical haplotype. The haplotype network for *P. corruscans* highlighted the haplotypes from the Sao Francisco Basin separated from the haplotypes from the Parana+Paraguay Basins by at least 13 mutations, represented in brackets (C).

Low sequence divergence (p-distance) within and between all the populations of *P. corruscans* was observed, ranging from 0.1 to 0.4% with all the genes analyzed separately. Divergence ranged from 1.5 to 1.7% for all the other populations for both the mitochondrial DNA, other than when the Sao Francisco population was analyzed.

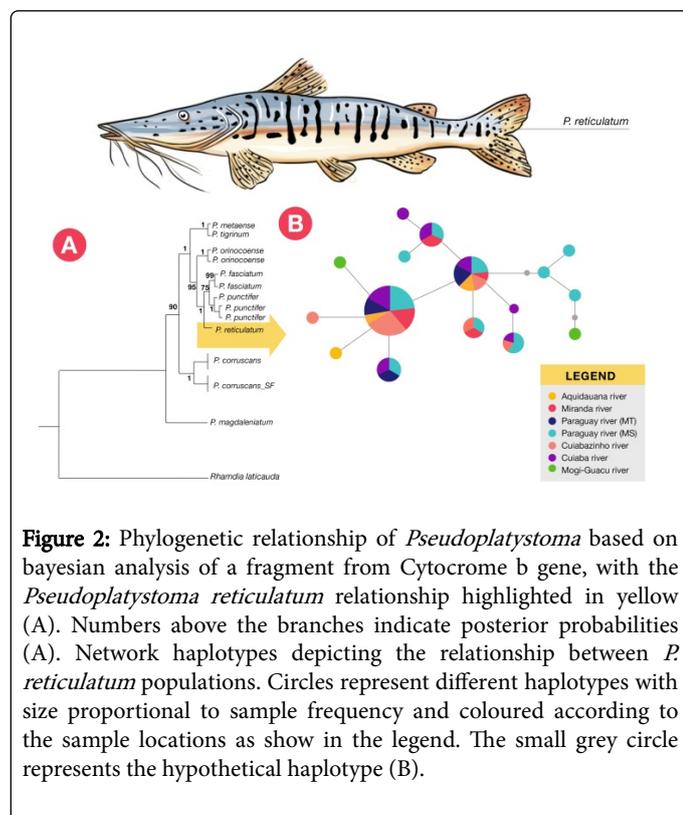
The pattern of genetic divergence between the *P. corruscans* populations (Parana+Paraguay Basins X Sao Francisco Basin) was confirmed with Cytochrome b by Bayesian analysis to the *Pseudoplatystoma* genus (Figure 1a). The differentiation between Clade I and Clade II, as obtained for the concatenated data set, was well supported, with a high posterior probability value (Supplementary Figure 1b, S6).

The concatenated data set showed that specimens from the different populations of *P. reticulatum* remained grouped together in the topology, without any clear geographic or taxonomic pattern (Figure S5B, Supplementary material). There was no resolution within *P. reticulatum* at either the population or subspecies levels.

Low sequence divergence (p-distance) was observed within and between all populations of *P. reticulatum*, ranging from 0.1 to 0.8% with all the genes analyzed separately.

The genetic divergence between the *P. reticulatum* populations, without any clear geographic or taxonomic pattern, was confirmed

with Cytochrome b by Bayesian analysis to the *Pseudoplatystoma* genus (Supplementary Figure 2a, S6).



## Population analysis

Although few common nuclear haplotypes were widely found in all the *Pseudoplatystoma corruscans* populations, there was greater variability in the mitochondrial sequences. For ATPase, a total of 11 haplotypes (H) were found, with three present in 62 of the 74 individuals. The most frequent haplotypes were H1, shared by 36 individuals (all populations, except the Sao Francisco population). For Cytb, a total of 25 haplotypes (H) were found, with one haplotype present in 41 of the 72 individuals. The most frequent haplotypes were H8, shared by 31 individuals (all populations, except the Sao Francisco population). For *P. corruscans* both genes, ATPase and Cytb, showed the haplotypes from the Sao Francisco River as being separated from the others by at least 13 mutations (Figure 1c). The phylogenetic networks depicting the relationships between the nuclear haplotypes and the haplotype distribution among samples showed that nuclear diversity is not geographically structured. The structuration of the Sao Francisco population was not identified for the RAG1 gene. For the nuclear gene, three RAG1 haplotypes were found, with one identified in 56 of the 58 specimens, and two haplotypes identified in one specimen each (not shown).

The phylogenetic networks depicting the relationships among the *Pseudoplatystoma reticulatum* population show that diversity is not geographically structured. For ATPase, a total of 16 haplotypes were found (H), with two present in 43 of the 60 individuals. The most frequent haplotypes were H2, shared by 32 individuals (all populations). For Cytb, a total of 16 haplotypes (H) were found, with two present in 31 of the 55 individuals. The most frequent haplotypes

were H3, shared by 18 individuals (all populations, except Mogi-Guaçu). For the nuclear gene RAG1, a total of seven haplotypes (H) were found, with three haplotypes present in 39 of the 43 individuals. The most frequent haplotypes were H2, shared by 25 individuals (all populations) (not shown).

The FST indexes were obtained pairwise for all populations and for each mtDNA, to investigate the genetic structure between the populations of each species. The FST index values of *P. corruscans*, except for the Sao Francisco population, exhibited low to high values ranging from 0 to 0.4 for both genes. Statistically significant values ( $p < 0.05$ ) were observed among some geographically distant populations (Table 3). With the Sao Francisco population, however, the values increased from 0.5 to 0.7 for the ATPase gene and from 0.07 to 0.36 for the Cytb gene, while statistically significant values ( $p < 0.05$ ) were observed between the Sao Francisco population and all the other populations of both genes, except between the Sao Francisco and Ivinhema populations for Cytb (Table 3). The Mantel test detected no significant relationship based on the FST index ( $r^2=41.42$ ,  $P>0.05$  for ATPase and  $r^2=37.88$ ,  $P>0.05$  for Cytb).

Analysis of Molecular Variance (AMOVA) was performed and considered all the populations together (without structuration) and then separated into two forms according to geographic distribution: Group I was composed of the samples from the Paraguay Basin while Group II was composed of the populations from the Parana Basin (Mogi-Guaçu and Ivinhema). Group III was composed of all the samples from the Parana-Paraguay Basins, while group IV was composed of the populations from the Sao Francisco Basin. Only Groups I and II were tested for *P. reticulatum*; however Group II did not include the Ivinhema population samples.

When analyses were conducted considering all the populations of *P. corruscans* together, AMOVA revealed a total genetic variance of 0.36 for ATPase and 0.4 for Cytb, with most of the variance found within populations. When analyses were performed considering the putative existence of two groups (groups I and II), hierarchical AMOVA revealed a total genetic variance of 0.26 for ATPase and 0.36 for Cytb, with most of the variance also found within populations. The variance between groups I and II was zero. This value indicates that there was no genetic difference between the groups. The analyses considering the putative existence of the other two groups (groups III and IV) in a hierarchic AMOVA revealed a total genetic variance of 0.51 for ATPase and 0.45 for Cytb, with most of the variance was also found within populations (Table 4).

The population pairwise FST values of *P. reticulatum* exhibited low to moderate values ranging from 0 to 0.19 for both genes, while no statistically significant values were observed (not shown). Values that were not significant ( $P>0.05$ ) showed no structuration in the populations of *P. reticulatum*. AMOVA considering all populations of *P. reticulatum* together revealed a total genetic variance of 0.34 for ATPase and 0.42 for Cytb, with the most of the variance within populations. The variability among populations was therefore zero. When the analyses were conducted considering the putative existence of two groups (Groups I and II), the hierarchical AMOVA revealed a total genetic variance of 0.35 for ATPase and 0.48 for Cytb, with the most variance within populations. The variances among populations within groups were zero (Table 4).

Haplotype Diversity was higher for Cytb b for both species, but was also high for the two genes analyzed, ranging from 0.7 to 0.83.

	Sao Francisco	Aquidauana	Miranda	Paraguay_MT	Paraguay_MS	Cuiabazinho	Cuiaba	Ivinhema	Mogi Guacu
Sao Francisco	-	0.07778*	0.11581*	0.16667*	0.10674*	0.20511*	0.36441*	0.06667	0.21379*
Aquidauana	0.54226*	-	0.03029	0.02116	0.04952	0.0387	0.13758	0	0.04
Miranda	0.50552*	0.00968	-	0.13062	0.14286	0.20422*	0.35451*	0.0625	0.21386*
Paraguay_MT	0.50951*	0	0.08841	-	0.11629	0	0.09193	0	0
Paraguay_MS	0.50000*	0.12134	0	0.19547*	-	0.14504	0.33597*	0	0.15655
Cuiabazinho	0.65896*	0.0461	0.27924*	0	0.41945*	-	0.00514	0	0
Cuiaba	0.64005*	0	0.22700*	0.01313	0.38281*	0	-	0	0
Ivinhema	0.56146*	0.00758	0	0.11056	0	0.42268*	0.33698	-	0
Mogi Guacu	0.70092*	0.01635	0.28189	0	0.43103	0	0	0.5	-

**Table 3:** F<sub>ST</sub> (above diagonal) for Cytb and FST (below diagonal) for ATPase, values among pairwise *Pseudoplatystoma corruscans* populations; \*P<0.05 (significant values).

<i>Pseudoplatystoma corruscans</i>				<i>Pseudoplatystoma reticulatum</i>			
Groups	Source variation of	Variance componentes (ATP/Cytb)	Percentage variation (ATP/Cytb)	Source variation of	Variance componentes (ATP/Cytb)	Percentage variation (ATP/Cytb)	
Without structuration	Among populations	0.12/0.05	34.14/12.99	Among populations	0/0	0/0	
	Within populations	0.24/0.35	65.86*/87.01*	Within populations	0.34/0.42	99.9/99.9	
Groups I and II	Among groups	0/0	0/0	Among groups	0.01/0.06	4.77/14.43	
	Among populations within groups	0.01/0.03	3.19*/7.19*	Among populations within groups	0/0	0/0	
	Within populations	0.25/0.33	96.81*/92.81*	Within populations	0.34/0.42	95.23/85.57	
Groups III and IV	Among groups	0.24/0.07	46.43/14.78	-	-	-	
	Among populations within groups	0.03/0.03	6.88*/7.52*				
	Within populations	0.24/0.35	46.69*/77.70*				

**Table 4:** Molecular variance analyses (AMOVA); \*Significant variation (P<0.05).

## Discussion

When performed independently for each gene, mtDNA genealogy exhibited greater resolution than nDNA in terms of well-supported clades, and the concatenated data set (ATPase+Cytb+RAG1) displayed similar topologies and supports to the independent mitochondrial analysis for both species analyzed. Due to the low mutational rate of nuclear genes compared to mitochondrial genes, differentiation between the Sao Francisco Basin and other basins was not identified for the nuclear genes for *Pseudoplatystoma corruscans*. According to Avise [43], this may be due to the fact that lineages take longer to be completely sorted for autosomal loci.

The concatenated data set (ATPase+Cytb+RAG1) and Bayesian topology with Cytb exhibited topologies partitioned into two main clades (I and II) for *P. corruscans* individuals. Clade I was composed of samples from the Sao Francisco Basin, while clade II is integrated by

other samples (Parana+Paraguay Basins), with high posterior probability values (Figures 1a and 1b, Supplementary Figures S5a and S6). Carvalho-Costa et al. [12] also showed geographically distinct clades for *P. corruscans* between the Sao Francisco and Parana-Paraguay-Uruguay Basins. Carvalho et al. [13] recently considered *P. corruscans* from the Sao Francisco Basin as distinct Evolutionarily Significant Units (ESUs). The results of the present study are congruent with those relating to the differentiation of the Sao Francisco Basin and the Parana-Paraguay Basins, with data suggesting that *P. corruscans* from the Sao Francisco Basin is a new species. Morphological confirmation, however, is required.

In addition to topology analysis, the genetic differentiation between *P. corruscans* populations was identified by grouping the populations in three different forms (first: considering all the populations together; second: group I from Paraguay Basin and group II from Parana Basin; third: group III from the Parana+Paraguay Basins and group IV from

Sao Francisco Basin). The  $F_{ST}$  index observed among all the populations (except Sao Francisco) of *P. corruscans* revealed low to high values (0 to 0.4), according to Wright [44], for both mitochondrial genes, suggesting the occurrence of a low to high genetic structure (Table 3). Although statistically significant values have previously been found between some geographically distant populations, there was no difference when analysis was performed based on the putative existence of two groups (Group I: Paraguay Basin and Group II: Parana Basin). The variance between groups I and II was zero (Table 4). This division was expected, since as these two groups were separated by an important physical barrier, the Guaira Falls, until 1983, and are currently divided by the Itaipu Dam. As observed by Borghetti et al. [45] it cannot be determined if these findings are a consequence of the fish ladder constructed at the Itaipu hydroelectric complex. On the other hand, the  $F_{ST}$  indexes of all the pairwise populations for each mtDNA, calculated among the Sao Francisco and other populations, had mean values higher than 0.4, with a significant genetic structure among all the populations, except between the Sao Francisco and Ivinhema populations for the Cytb gene (Table 3). When the analyses were conducted considering the putative existence of the two other groups (Group III: Paraguay+Parana Basins and Group IV: Sao Francisco Basin), the variance between the groups revealed high, but not significant P-values. There were few connections between the river basins (Parana-Paraguay and Sao Francisco Basins), and geological records indicate that rivers such as the Sao Francisco have an older geomorphological history than rivers such as the Parana-Paraguay [46]. In addition to the small number of connections between the river basins investigated, geological evidence suggests allopatric population differentiation, as detected herein for *P. corruscans*.

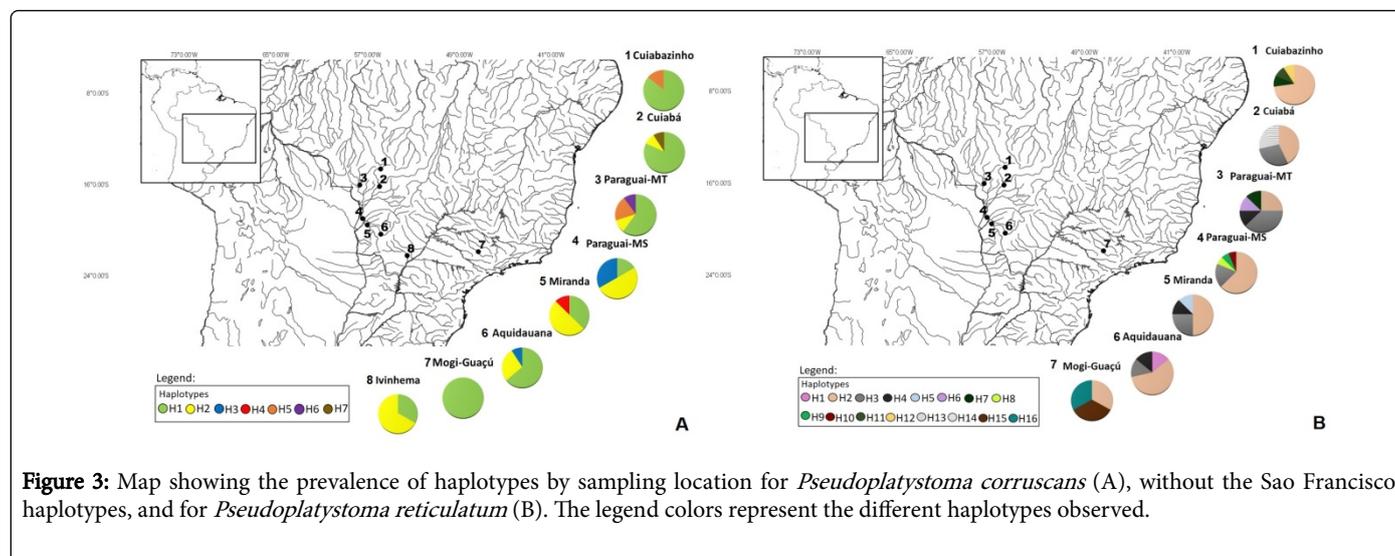
The concatenated data set showed that specimens from the different populations of *P. reticulatum* remained grouped together in topological analysis, without any clear geographic or taxonomic pattern. There was no resolution within *P. reticulatum* at the population or subspecies levels (Supplementary Figure S5b). The  $F_{ST}$  population pairwise values for *P. reticulatum* showed low to moderate values, according to Wright [44], ranging from 0 to 0.19, for both genes (not shown). No statistically significant values ( $P > 0.05$ ) were observed. The lack of significant P-values suggests that there is no structuration between the populations. When the analyses were conducted considering the putative existence of two groups (Groups I and II, without the Ivinhema population), hierarchical AMOVA revealed low variance among groups, and values that were not significant ( $P > 0.05$ ). Therefore, although the use of  $F_{ST}$  values in population differentiation is still the subject of debate, in terms of conservation, evaluating the existence of subdivisions between populations instead of a single population is recommended, avoiding depletion of genetic variation [47].

To confirm these findings the haplotype network was observed. For *P. corruscans* a nested clade displayed the haplotypes from the Sao Francisco River as separated from other populations by at least 13 mutations (Figure 1c). *P. reticulatum* showed no structuration for the nested clade, showing that its diversity is not geographically structured (Figure 2b). All the results obtained for both species and both mitochondrial DNA genes corroborate those seen in the topologies and the genetic diversity values.

The data of the present study partially agree with those obtained by studies using microsatellite groups [15], as while there was evidence of significant structure among some *P. corruscans* populations, this was not identified for other populations (Table 3) in the present study, even

between populations collected in the breeding season. According to Pereira et al. [15] genetic population studies involving migratory species should be conducted only with specimens collected in tributaries during their reproductive migration to avoid collecting mixed populations in feeding areas. The migration flows during the year are based around remaining in the river beds from February to September (feeding season), and then in October begin upstream migration (breeding season) [7,48]. However, the results of the present study disagree with the findings of Abreu et al. [14] and Pereira et al. [15] that corroborated the occurrence of homing in *P. corruscans* and *P. reticulatum*, respectively, since in our analysis there was no evidence of structure for such populations. The possibility exists that this study may not have identified subtle structures in some populations of these species. This may be due to sampling error, with the possible collection of individuals from different populations, even samples being collected in breeding season, the relatively low number of individuals sampled in some locations, which would not be enough to identify interpopulation differences, or the markers used. However, it is possible that this lack of genetic differentiation in some populations is related to the panmictic nature of the population. Migratory movements common to these species may cause a lack of significant differentiation among some geographically close populations, which can also explain the high diversity found in these populations. According to Laikre et al. [19] the factor that most contributes to increasing intrapopulation diversity is the immigration of individuals through the gene flow. The high genetic variability observed in the present study for both species (0.7 to 0.83) may be related to contact between the populations. As the *P. reticulatum* and *P. corruscans* species are migratory [49-51] and perform major migrations during their life cycles, the occurrence of genetic flow is possible, even among geographically distant populations.

The geographic distribution of the haplotypes observed (Figure 3) suggests that the greatest diversity is concentrated in the region covered by the Pantanal biome in the Paraguay Basin, while the lowest diversity is in Mogi-Guacu in the Parana Basin, for both the species analyzed. Additionally, genetic diversity comparison found that *P. corruscans* is a more endangered species in terms of conservation, as while both species presented similar distribution, the effects of predatory fishing, artificial hybridization and fish that have escape from cultivated stocks, mean that effects on the genetic diversity of natural populations are more drastically observed in *P. corruscans* (seven haplotypes, without the Sao Francisco haplotypes) (Figure 3a) than in *P. reticulatum* (16 haplotypes) (Figure 3b). In this way, the establishment of a Germplasm Bank for these catfishes can define localities based on the present data. We suggest that capture should be performed in the Pantanal region (Paraguay-Mato Grosso) for both species, with additional sampling in Cuiabá, Paraguay-Mato Grosso do Sul and Miranda for *P. corruscans* and in the other locations to sample specific haplotypes. These data indicate a different strategy for Germplasm bank formation, or the use of the population as breeders in aquaculture. This information can therefore support the establishment of public policies for the conservation and restoration of the remaining biodiversity, especially in localities with less diversity, and the addition of programs to maintain the diversity observed in biomes such as the Pantanal. The results of the present study provide information about genetic diversity and the patterns of population differentiation for these ecological and economically important species, and represents an advance for future management and conservation studies.



**Figure 3:** Map showing the prevalence of haplotypes by sampling location for *Pseudoplatystoma corruscans* (A), without the Sao Francisco haplotypes, and for *Pseudoplatystoma reticulatum* (B). The legend colors represent the different haplotypes observed.

## Acknowledgments

This research was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (the São Paulo State Research Foundation) (FAPESP) (2012/03553-8). We would like to thank the Instituto Chico Mendes de Conservação da Biodiversidade (the Chico Mendes Biodiversity Preservation Institute) (ICMBio) for the sampling permission granted through the Sistema de Autorização e Informação em Biodiversidade (the Biodiversity Authorization and Information System) (SISBIO).

## References

- Vari RP, Malabarba LR (1998) Neotropical Ichthyology: an overview. In: Malabarba LR, Reis RE, Vari RP, Lucena ZMS, Lucena CAS (eds.) Phylogeny and Classification of Neotropical Fishes. Porto Alegre: Edipucrs pp: 1-11.
- Reis RE, Kullander SO, Ferraris C (2003) Check List of freshwater Fishes of South and Central America (CLOFFSCA). Porto Alegre, Brasil: Edipucrs.
- Lundberg JG, Littmann MW (2003) Family Pimelodidae. In: Reis, R. E.; Kullander SO, Ferraris-JR CJ (Org) Check List of the freshwater fishes of South and Central America, Porto Alegre, pp: 432-446.
- Ferreira EJJ, Zuanon JAS, Santos GM (1998) Peixes Comerciais do Médio Amazonas: Região de Santarém, Pará. IBAMA, Brasília.
- Loubens G, Panfili J (2000) Biologie de *Pseudoplatystoma fasciatum* et *Pseudoplatystoma tigrinum* (Teleostei: Pimelodidae) dans le bassin du Mamoré (Amazonie Bolivienne). Ichthyological Exploration of Freshwaters 11: 13-34.
- Carolsfeld J, Harvey B, Ross C, Baer A (2003) Migratory Fishes of South America: Biology, Fisheries and Conservation Status. World Fisheries Trust/World Bank/ IDRC, Victoria.
- Godinho AL, Kynard B, Godinho HP (2007) Migration and spawning of female surubim (*Pseudoplatystoma corruscans*, Pimelodidae) in the São Francisco river, Brazil. Environ Biol Fish 80: 421-433.
- Buitrago-Suárez UA, Burr BM (2007) Taxonomy of the catfish genus *Pseudoplatystoma* Bleeker (Siluriformes: Pimelodidae) with recognition of eight species. Zootaxa 1512: 1-38.
- Mateus LAF, Penha JMF (2007) Dinâmica populacional de quatro espécies de grandes bagres na bacia do rio Cuiabá, Pantanal norte, Brasil (Siluriformes, Pimelodidae). Rev Bras Zool 24: 87-98.
- Bignotto TS, Prioli AJ, Prioli SMAP, Maniglia TC, Boni TA, et al. (2009) Genetic divergence between *Pseudoplatystoma corruscans* and *Pseudoplatystoma reticulatum* (Siluriformes: Pimelodidae) in the parana river basin. Braz J Biol 69: 681-689.
- Torrico JP, Hubert N, Desmarais E, Duponchelle F, Nuñez Rodriguez J, et al. (2009) Molecular phylogeny of the genus *Pseudoplatystoma* (Bleeker, 1862): biogeographic and evolutionary implications. Mol Phylogenet Evol 51: 588-594.
- Carvalho-Costa LF, Piorski NM, Willis SC, Galetti Jr PM, Ortí G (2011) Molecular systematics of the neotropical shovelnose catfish genus *Pseudoplatystoma* Bleeker 1862 based on nuclear and mtDNA markers. Molecular Phylogenetics and Evolution 59: 177-194.
- Carvalho DC, Oliveira DAA, Beheregaray LB, Torres RA (2012) Hidden genetic diversity and distinct evolutionarily significant units in a commercially important Neotropical apex predator, the catfish *Pseudoplatystoma corruscans*. Conserv Genet 13: 1671-1675.
- Abreu MM, Pereira LHG, Vila VB, Foresti F, Oliveira C (2009) Genetic variability of two populations of *Pseudoplatystoma reticulatum* from the Upper Paraguay River Basin. Genet Mol Biol 32: 868-873.
- Pereira LH, Foresti F, Oliveira C (2009) Genetic Structure of the migratory catfish *Pseudoplatystoma corruscans* (Siluriformes: Pimelodidae) suggests homing behavior. Ecol Fresh Fish 18: 215-225.
- Batista JS, Alves-Gomes JA (2006) Phylogeography of *Brachyplatystoma rousseauxii* (Siluriformes - Pimelodidae) in the Amazon Basin offers preliminary evidence for the first case of "homing" for an Amazonian migratory catfish. Genetics and Molecular Research 5: 723-740.
- Sivasundar A, Bermingham E, Ortí G (2001) Population structure and biogeography of migratory freshwater fishes (Prochilodus: Characiformes) in major South American rivers. Mol Ecol 10: 407-417.
- Carvalho-Costa LF, Hatanaka T, Galetti PM Jr (2008) Evidence of lack of population substructuring in the Brazilian freshwater fish *Prochilodus costatus*. Genet Mol Biol 31: 377-380.
- Laikre L, Antunes A, Alexandrino P (1999) Conservation genetic management of brown trout (*Salmo trutta*) in Europe. In: Laikre L (ed.) "Troutconcert": EU Fair CT97-3882. Stockholm University, Sweden.
- Hansen MM, Kenchington E, Nielsen EE (2001) Assigning individual fish to populations using microsatellite DNA markers. Fish Fish 2: 93-112.
- Schwartz MK, Luikart G, Waples RS (2007) Genetic monitoring as a promising tool for conservation and management. Trends Ecol Evolut 22: 25-33.
- Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. Nature 464: 697-709.
- Zhang Z, Glove KA, Wennevik V, Svasand T, Sorvik AGE, et al. (2013) Genetic analysis of Atlantic salmon captured in a netting station reveals multiple escapement events from commercial fish farms. Fish Manag Ecol 20: 42-5.

24. Barthem R, Goulding M (2007) Um Ecosistema Inesperado: A Amazônia Revelada Pela Pesca. Amazon Conservation Association, Lima, p: 240.
25. Mello PH, Venturieri RLL, Honji RM, Moreira RG (2009) Threatened fishes of the world: *Pseudoplatystoma corruscans* (Agassiz, 1829) (Siluriformes: Pimelodidae). *Env Biol Fish* 85: 359-360.
26. Thompson JD, Higgins DG, Gibrion TJ (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, Position specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673-4680.
27. 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.
28. <http://www.geneious.com>
29. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30: 2725-2729.
30. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572-1574.
31. Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees in Proceedings of the Gateway Computing Environments Workshop (GCE). New Orleans, LA, pp: 1-8.
32. <http://beast.bio.ed.ac.uk/Tracer>
33. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7 *Mol Biol Evol* 29: 1969-1973.
34. <http://beast.bio.ed.ac.uk/TreeAnnotator>
35. Lanfear R, Calcott B, Ho SYW, Guindon S (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol* 29: 1695-1701.
36. Tavaré S (1986) Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. *Lectures on Mathematics in the Life Sciences. American Mathematical Society* 17: 57-86.
37. Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* 17: 368-376.
38. Hasegawa M, Kishino H, Yano T (1985) Dating of human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 22: 160-174.
39. Kimura M (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16: 111-120.
40. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452.
41. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res* 10: 564-567.
42. Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* 27: 209-220.
43. Avise JC (2000) *Phylogeography: The History and Formation of Species*, first ed. Harvard University Press, Cambridge.
44. Wright S (1978) *Evolution and the genetics of populations: variability within and among natural populations*. University of Chicago Press: Chicago, USA.
45. Borghetti JR, Nogueira VSG, Borghetti NRB, Canzi C (1994) The fish ladder at the Itaipu Binacional hydroelectric complex on the Parana River, Brazil. *Regul. rivers. Research & Management* 9: 127-130.
46. Lima FCT, Ribeiro AC (2011) Continental scale tectonic controls of biogeography and ecology. In: Albert JS, Reis RE (eds) *Historical biogeography of Neotropical freshwater fishes*. University of California Press, Berkeley pp: 145-164.
47. Laikre L, Palm S, Ryman N (2005) Genetic population Structure of fishes: implications for coastal zone management. *Ambio* 34: 11-119.
48. Resende EK (2003) Migratory fishes of the Paraguay-Parana' Basin excluding the Upper Parana' Basin. In: Carolsfeld J, Harvey B, Ross C, Baer A eds *Migratory fishes of South America. Biology, fisheries and conservation status*. Victoria, BC, Canada: World Fisheries Trust, pp: 99-156.
49. Palumbi S, Martin A, Romano S, McMillan WO, Stice L, et al. (2002) *The Simple Fool's Guide to PCR*. University of Hawaii Press, Hawaii.
50. Irwin DM, Kocher TD, Wilson AC (1991) Evolution of cytochrome b gene in mammals. *J Mol Evol* 2: 13-34.
51. Perdices A, Bermingham E, Montilla A, Doadrio I (2002) Evolutionary history of the genus *Rhamdia* (Teleostei: Pimelodidae) in Central America. *Mol Phylogenet Evol* 25: 172-189.