An Acad Bras Cienc (2020) 92(1): e20180496 DOI 10.1590/0001-3765202020180496
Anais da Academia Brasileira de Ciências | *Annals of the Brazilian Academy of Sciences*
Printed ISSN 0001-3765 I Online ISSN 1678-2690
www.scielo.br/aabc | www.fb.com/aabcjournal

BIOLOGICAL SCIENCES

Genetic diversity and structuring in the arapaima (Osteoglossiformes, Osteoglossidae) population reveal differences between the Amazon and the Tocantins-Araguaia basins

FABRÍCIA NOGUEIRA, PÉRICLES S. DO RÊGO, HELDER QUEIROZ, PAULO VENERE, EDUARDO S. VARELA, IRACILDA SAMPAIO, HORACIO SCHNEIDER & JULIANA ARARIPE

Abstract: Arapaima is a widely-distributed fish of enormous economic importance in the Amazon region. In the present study, a total of 232 specimens were sampled, 121 from five sites in the Amazon basin and 111 from five sites in the Tocantins-Araguaia basin. The analyses investigated fragments of the Cytochrome b, Control Region, Cytochrome Oxidase I, NADH dehydrogenase subunit 2 and seven loci microsatellites. The analyses revealed the existence of two mitochondrial lineages within the general area, with no haplotypes shared between basins, and genetic variability significantly higher in the Amazon than in the Tocantins-Araguaia basin. Two divergent, but sympatric mitochondrial lineages were found in the Amazon basin, but only a single lineage in the Tocantins-Araguaia basin. The existence of these two mitochondrial lineages indicates that past events, probably occurring during the Pleistocene, resulted in the separation of the populations of this species and molded its evolutionary history, which is reflected directly in its mitochondrial DNA. The analysis of the arapaima population structure identified distinct levels of diversity within the distribution of the species, indicating specific geographic regions that will require special attention for the development of conservation and management strategies.

Key words: Amazon Basin, Tocantins-Araguaia Basin, Lineages, structure population, conservation.

INTRODUCTION

The arapaima, *Arapaima gigas* Schinz, 1822, is one of the largest freshwater fish in the world, reaching up to 3 m in total length and weighing over 200 kg (Nelson et al. 2016). This species prefers lentic habitats, such as floodplains and lakes (Castello 2008). This important fishery resource has a long history of exploitation, as shown by records of fishery landings at the main ports of the Amazon region. For this reason, the arapaima was included in Appendix II of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), a list of threatened species that may become extinct unless trade is strictly controlled (CITES 2017). However, the International Union for Conservation of Nature (IUCN) considers the species to be data deficient for the reliable assessment of extinction risk (IUCN 2017).

Population and phylogeographic studies of the arapaima have identified different levels of structuring in the populations within the Amazon and Tocantins-Araguaia basins, although the samples analyzed in these studies were not adequate for conclusive comparisons of the

genetic structure of the two basins (Hrbek et al. 2005, 2007, Araripe et al. 2013). This comparison is hampered by the small sample size, especially from the Tocantins-Araguaia basin (Hrbek et al. 2005, 2007, Araripe et al. 2013), or the restriction of the samples to only one basin (Vitorino et al. 2015, 2017, Fazzi-Gomes et al. 2017a). While the arapaima is classified as a sedentary fish (Castello 2008), phylogeographic studies based on mitochondrial DNA markers have demonstrated high levels of gene flow, indicating that they form a continuous population (Hrbek et al. 2005). However, studies using codominant markers with high mutation rates (i.e. microsatellites) contradict this continuity, and indicate the existence of different levels of structuring at different spatial scales (Araripe et al. 2013, Fazzi-Gomes et al. 2017a, Vitorino et al. 2017).

Genetic studies in different areas of the Amazon region have provided important insights into the variation among populations within the geographic range of the species, which is fundamental to the management of this fishery resource (Escobar et al. 2015, Fazzi-Gomes et al. 2017b). The genetic differentiation of arapaima populations observed at some points in the distribution of the species (Vitorino et al. 2015, 2017, Watson et al. 2016, Fazzi-Gomes et al. 2017a) may be related to the distinct characteristics of the floodplains and drainage systems of each basin, which may have a direct influence on the migration patterns of the species (Castello 2008). The Amazon basin is a complex system composed of large tracts of tropical forest interspersed with rivers, lakes and channels (Ayres 2006). By contrast, the Tocantins-Araguaia basin encompasses a mosaic of habitats, including the Cerrado savannas of the central Brazilian plateau, lowland Amazon rainforest, and transitional environments between these two biomes (Mérona et al. 2010). The rivers of this basin have transparent water and relatively narrow floodplains (Latrubesse & Stevaux 2002, Junk et al. 2011) characterized by an extremely rapid annual flood pulse, which connects the floodplain lakes for only a short period each year (Aquino & Latrubesse 2008). This contrasts considerably with the prolonged flood and ebb periods observed in the Amazon basin (Castello 2008, Ramalho et al. 2009).

In this study, we present the results of the first genetic analysis of the arapaima that encompasses comprehensive samples from both the Amazon and Tocantins-Araguaia basins, with the primary aim of characterizing the genetic variation between these basins. These analyses contribute to the understanding of the current population structure and genetic diversity of arapaima, as well as historical patterns.

MATERIALS AND METHODS

Sampling

To analyze the population structure of the arapaima, a total of 232 specimens were collected, 121 from five sites in the Amazon basin (AM) and 111 from five sites in the Tocantins-Araguaia basin (TA) (Figure 1 and Table I). The sampling sites were distributed throughout the two study basins, and the most distant sampling points (Iquitos and Quatro Bocas) were approximately 5000 km apart, following the main course of the rivers. Muscle tissue, fin and scale samples were fixed in 96% ethanol, catalogued, and deposited in the tissue collection of the Genetic and Conservation Laboratory on the Bragança campus of the Federal University of Pará, in northern Brazil.

All the specimens were obtained from dead individuals obtained directly from local artisanal fisheries, which precludes the need for approval by the Animal Ethics Committee. The transportation of the tissue samples was authorized by the

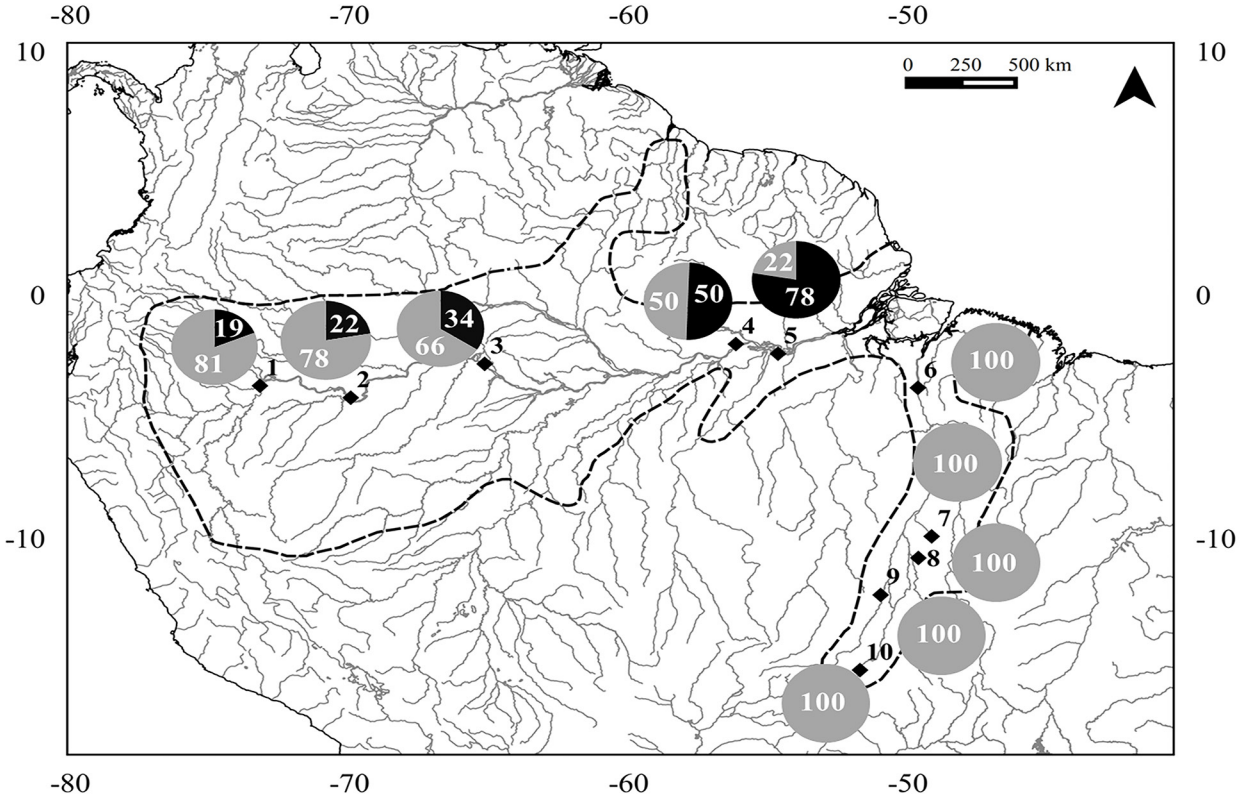

Figure 1. Map of the approximate natural distribution of Arapaima (dashed line) showing the localities sampled for this study. 1-Iquitos, 2-Letícia, 3-Mamirauá , 4-Jurutí, 5-Santarém, 6-Tucuruí, 7-Caseara, 8-Lagoa da Confusão, 9-Novo Santo Antônio and 10-Quatro Bocas. Graphical representation of the Amazon Only (black) and Amazon/ Araguaia-Tocantins (gray) lineages. The numbers inside the circles represent the percentage of the members of each lineage.

Brazilian Environment Ministry (process number 02001.007554/2005-76 IBAMA/MMA).

Laboratory procedures

The genomic material was extracted with proteinase K and washed with phenol/ chloroform and isopropanol for precipitation (Sambrook and Russell 2001), or by using a DNA isolation kit (Wizard Genomic DNA Purification Kit - Madison, WI, USA). Four regions of the mitochondrial genome were amplified by PCR: Cytochrome b (*Cytb*), the Control Region, Cytochrome Oxidase I (*COI*) and NADH dehydrogenase subunit 2 (*ND2*). The *Cytb* was sequenced using primers adapted from Bossuyt and Milinkovitch (2000), and AgiProf (5'-TTTTAACTCCCACCCTTAACTCC-3') and AgiPher (5'GGTCCGTCTTAACATCTTCAGTG-3') (Lima I., unpublished data) were used to sequence the Control Region. The primers LIICOIF3 (Brito et al. 2015) and FishR1 (Ward et al. 2005) were used to sequence *COI*, while Obi198L (5'-TACATTCGCCAGCTCCCAC-3') and ObiAsn1H (5'GGAAGCTCGTTGGTTGGAGC-3') were used for *ND2*. The PCR protocol consisted of initial denaturation at 95°C for 3 minutes, followed by 35 cycles at 94°C for 30 seconds, annealing for 30 seconds, and extension at 72°C for 1 minute, with a final extension at 72°C for 7 minutes. The annealing temperature was 56°C for the *ND2*, 57°C for *Cytb*, and 58°C for *COI*. The annealing temperature for the Control Region was 62°C, with a decrease of 0.3°C after each cycle.

Table I. Genetic diversity indices for *Arapaima* based on an analysis of the concatenated (*Cytb* and Control Region) dataset.

Sites	Localities	Geographical coordinates	N	S	H	EH	SEH (%)	*h*	π
1	Iquitos	3°43'31.35" S 73°12'18.59" W	16	15	7	3	31	0.833	0.00413
2	Letícia	4°12'56.25" S 69°59'14.74" W	18	14	6	3	28	0.758	0.00378
3	Mamirauá	2°49'35.06" S 65°9'26.35" W	38	20	11	7	24	0.853	0.00478
4	Jurutí	2º0'50.12" S 56º9'36.23" W	22	20	11	7	73	0.857	0.00579
5	Santarém	2°8'15.28" S 54°44'33.445" W	27	22	12	8	63	0.920	0.00470
AM	**Amazon basin**	**-**	**121**	**41**	**35**	**35**	**100**	**0.937**	**0.00552**
6	Tucuruí	3°48'29.91" S 49°38'37.20" W	35	2	3	2	26	0.417	0.00035
7	Caseara	9º54'34.20" S 49º8'53.19" W	9	3	4	3	33	0.583	0.00067
8	Lagoa da Confusão	10º47'46.38" S 49º37'23.72"W	11	0	1	0	0	0.000	0.00000
9	Novo Santo Antônio	12°18'56.11" S 50°58'24.78" W	25	3	4	3	52	0.597	0.00055
10	Quatro Bocas	15º23'57.52" S 51º42'57.54" W	31	0	1	1	100	0.000	0.00000
TA	**Tocantins-Araguaia basin**	**-**	**111**	**9**	**10**	**10**	**100**	**0.668**	**0.00092**
	Total	-	232	48	45	-	-	0.907	0.00448

N (number of specimens analyzed), S (number of variable sites), H (number of haplotypes), EH (number of exclusive haplotypes), SEH% (percentage of specimens with exclusive haplotypes), h (haplotype diversity), π (nucleotide diversity), AM (Amazon basin) and TA (Tocantins-Araguaia basin).

The amplification reaction consisted of 12.75 µl of H_2O, 4 µl of dNTPs (1.25 mM), 5 µl of buffer (5X buffer+$MgCl_2$), 1 µl of each primer (50 ng/µl), 0.25 µl of Taq DNA polymerase (5 U/µl) and 1 µl of genomic DNA (100 ng/µl). The amplification products were purified using 20% PEG 8000 according to a protocol modified from Paithankar and Prasad (1991). The samples were then sequenced using the Big Dye Terminator kit (ABI Prism Dye Terminator Cycle Sequencing Reading Reaction-PE, Applied Biosystems, Carlsbad, CA, USA). The reaction products were precipitated with 70% ethanol, resuspended in formamide and injected into an ABI 3500XL automated sequencer (Applied Biosystems).

Data analysis

The sequences obtained through these procedures were inspected visually, edited and aligned using BIOEDIT 7.2.5 (Hall 1999). The *Cytb* and Control Region fragments were used to identify patterns of diversity and genetic structure. As new *Arapaima* species have been proposed recently (Stewart 2013a, b), the *ND2* and *COI* genes were sequenced to ensure that all the samples analyzed were derived from the same taxon. The *ND2* fragments allowed for the verification of the genetic divergence levels proposed to differentiate the species of the family Osteoglossidae (Mu et al. 2012), while the *COI* sequences allowed for the evaluation of genetic divergence using the parameters applied in DNA barcoding (Hebert et al. 2003).

Genetic diversity was analyzed by calculating the number of variable sites (S) and haplotypes (H), and the haplotype (*h*) and nucleotide (π) diversity, in DnaSP 5.1 (Librado and Rozas 2009). The number of exclusive haplotypes (EH) and the percentage of specimens with exclusive haplotypes per site (SEH %) were also analyzed. Differences between the two basins in nucleotide and haplotype diversity and in the number of

haplotypes were compared with *t* tests run in the R 3.4.2 software (R Development Core Team 2017).

To evaluate the relationships among the haplotypes a haplotype network was constructed using the concatenated *Cytb* and Control Region data in the HAPLOVIEWER 4.2 software (Salzburger et al. 2011) based on the maximum-likelihood method available in PHYML 3.0 (Guindon et al. 2010) and employing the HKY+I substitution model, which was selected in JMODELTEST 2.1.1 (Darriba et al. 2012) by the Akaike Information Criterion, AIC (Akaike 1974).

The genetic differentiation among populations from the two study basins was evaluated by estimating pairwise F*st* indices. The correlation between genetic distances (measured as pairwise F*st* values) and geographical distances (km along the course of the main river) was evaluated using the Mantel test run in ARLEQUIN 3.5.1.2 (Excoffier & Lischer 2010) with 10,000 permutations at $p < 0.05$. The degree of genetic differentiation between the basins was evaluated by the Analysis of Molecular Variance based on the Φ-statistic (AMOVA) run in ARLEQUIN 3.5.1.2 (Excoffier & Lischer 2010) with 10,000 permutations at $p < 0.05$.

Seven microsatellite loci in 39 specimens from the two basins were obtained from the study of Araripe et al. (2013). Observed and expected heterozygosis were obtained from Arlequin 3.11 (Excoffier & Lischer 2010), while the *a posteriori* probability of the number of stocks was estimated using a Bayesian approach, run in Structure 2.3.4 (Pritchard et al. 2000), which does not require the prior definition of a structure to be tested. The analyses had a burn-in of 100,000, followed by one million replicates of the Monte Carlo Markov Chain (MCMC), the ancestral model with population admixture, and the correlated allele frequency (CAF) model. The number of stocks was defined based on the approach proposed by Evanno et al. (2005), run in Structure Harvest (Earl & vonHoldt 2012), with the results of the concatenated interactions being obtained in CLUMPP (Jakobsson & Rosenberg 2007) and plotted in DISTRUCT (Rosenberg 2004).

The timing of the split among the clades of the *A. gigas* was estimated using only the *Cytb* data. The GTR+G substitution model was selected for this analysis by JMODELTEST 2.1.1 (Darriba et al. 2012). A mutation rate of 1.5% per million years was considered (Zardoya & Doadri 1999), and the data were modeled using a relaxed molecular clock. The probability of each node was computed by Bayesian Inference in BEAST 1.7 (Drummond et al. 2012), with runs of 200 million generations, and one tree being sampled every 10,000 generations, with 10% of the generations being discarded as burn-in. The nodes were considered to be well-resolved when the *a posteriori* probability was higher than 0.9. The species *Heterotis niloticus* (AB035240, NC015081, FJ890318) and *Osteoglossum bicirrhosum* (AB043025, NC003095, AB035238) were employed as outgroups.

The analysis of the distribution of pairwise genetic differences (the mismatch distribution) was run according to the sudden expansion model (Rogers & Harpending 1992). Subsequently, BEAST 1.7 (Drummond et al. 2012) was used to produce a Bayesian Skyline Plot (BSP) to test for historical fluctuations in population size, considering the evolutionary models selected for this analysis by JMODELTEST 2.1.1 (Darriba et al. 2012), which were the JC (first partition), HKY+I (second partition), and HKY (third partition) models for *Cytb*, and HKY+I for the Control Region. These analyses were run for $1x10^8$ generations with genealogies being sampled every 10,000 generations, in a strict molecular clock model, with the first 10% of the generations being discarded as burn-in. The strict molecular clock was calibrated based on

an intraspecific mutation rate of 1.5% per million years estimated for *Cytb* (Zardoya & Doadri 1999). To confirm that the study samples belong to the same taxon, indices of divergence were calculated from 1000 bootstrap pseudoreplicates and the Kimura 2-parameter model (K2P), run in MEGA 6 (Tamura et al. 2013). The indices of genetic divergence recorded in the arapaima were compared with those described previously for congeneric species of Osteoglossidae (Mu et al. 2012), while the divergence of the *COI sequences* was analyzed based on the criteria of the DNA barcoding approach. The sequences with the following accession numbers were used for comparisons of the divergence indices: for *ND2: O. bicirrhosum* JQ337773 to JQ337780, *O. ferreirai* JQ337781 to JQ337785, *Scleropages formosus* JQ337787 to JQ337806 and *S. leichardti* JQ337810 and JQ337811; for *COI*: *S. leichardti* HM006989, *S. jardinii* KY123527 and KY123529 and *S. formosus* KY123492 and KY123493.

RESULTS

A total of 1,241 base pairs (567 bps for *Cytb* and 674 bps for the Control Region) of the mitochondrial genome were obtained from the 121 arapaima specimens collected from the Amazon basin and the 111 specimens from the Tocantins-Araguaia basin. GenBank accession numbers range from MH830284 to MH830294 for *Cytb*, MH830244 to MH830283 for Control Region segment haplotypes. The two fragments were concatenated for population inferences, producing a database of 232 sequences with 45 haplotypes defined by 48 variable sites. None of the haplotypes were shared between the river basins, and the data presented a high level of haplotype diversity (0.907) and a moderate level (0.00448) of nucleotide diversity (Table I).

Comparisons between sampling sites revealed high levels of genetic variability in the Amazon basin, with haplotype diversity varying between 0.758 and 0.920 and nucleotide diversity between 0.00378 (Letícia) and 0.00579 (Jurutí). Sites in the Tocantins-Araguaia basin exhibited moderate to low levels of genetic variability, with haplotype diversity ranging from zero to 0.597 and nucleotide diversity from zero (Lagoa da Confusão and Quatro Bocas) to 0.00067 (Caseara) (Table I). The levels of genetic variability were significantly higher in the Amazon basin (mean h = 0.847 and π = 0.463%) than in the Tocantins-Araguaia basin (mean h = 0.319, t = 3.862, df = 4.316, $p < 0.05$ and π = 0.031%, t = 11.706, df = 8, $p < 0.05$) (Table I).

Overall, 35 of the 45 haplotypes identified were exclusive to the Amazon basin, and 10 were exclusive to the Tocantins-Araguaia basin (Supplementary Material - Table SI). Three of the haplotypes from the Amazon basin (H1, H20 and H22) were the most frequent in this area, and were present at different sites. The most frequent haplotype (H36) in the Tocantins-Araguaia basin was recorded at all sites except Quatro Bocas, which had a unique and exclusive haplotype (Supplementary Material - Table SI). The number of haplotypes per site was significantly higher in the Amazon basin (mean H = 9.4) than in the Tocantins-Araguaia basin (mean H = 2.6, t = 4.907, df = 8, $p < 0.05$; Table I).

The haplotype network revealed the presence of two haplotype groups differentiated by nine mutational events (Figure 2), recovered by the phylogenetic inferences. The reciprocally monophyletic groups were classified as mitochondrial lineages (Figure 3). The two lineages were referred to as the Amazon Only Lineage and the Amazon/Tocantins-Araguaia Lineage. The two mitochondrial lineages were found in sympatry in the Amazon basin, although

only one lineage was present in the Tocantins-Araguaia basin (Figure 1).

The mismatch distribution forms multimodal curve over the geographic range of the study species (data not shown). A similar pattern was observed when the analysis included only the Amazon populations, confirming the presence of the two lineages in the Amazon basin. By contrast, a unimodal curve was obtained for the data from the Tocantins-Araguaia basin, which is consistent with the presence of a single lineage in this area. The mean estimate of the time of the split between the mitochondrial *A. gigas* lineages was approximately 564 (692–335) thousand years ago (Kyr), which corresponds to the middle Pleistocene (Figure 3). The diversity observed in the Amazon Only and the Amazon/Tocantins-Araguaia lineages indicates that they coalesced at approximately 383 (476–175) Kyr and 454 (519-259) Kyr ago, respectively. The evaluation of historic changes in arapaima population size revealed a scenario of relative stability over time, which was upheld when the two lineages were analyzed (Supplementary Material - Figure S1).

The analysis of the genetic differentiation between the sites of the two study basins, based on the F*st* values, found significant values for all the pairs, with genetic differentiation varying between 0.223 and 0.748 (Table II). Significant values for this index were recorded between some population pairs within both the Amazon and Tocantins-Araguaia basins. The result of the Mantel test was not significant ($p > 0.05$), indicating that the observed genetic differentiation (F*st*) cannot be accounted for by the geographic distances among samples. The genetic differentiation between the two basins was also confirmed by the value obtained from the AMOVA with 37.42% of the variation being found between basins ($\Phi_{CT} = 0.374$, $p < 0.05$), 16.46% of the variation occurring among the populations within a given basin ($\Phi_{SC} = 0.263$, $p > 0.05$), and 46.11% of the variation occurring within populations ($\Phi_{ST} = 0.538$, $p < 0.05$).

The genetic divergence between the two lineages was 0.9%, whereas the divergence within each lineage was 0.2% for the concatenated *Cytb* and Control Region markers. To verify whether this level of differentiation is compatible with the intraspecific variation found in the two lineages, 600 bp of the *COI* gene (GenBank accession numbers: MH830240- to MH830243) and 764 bp of the *ND2* gene (GenBank accession numbers: MH830294 to MH830297) were sequenced in 22 arapaima specimens representing the two lineages. In this analysis, the mean divergence found between the lineages was 0.3% for *ND2* and 0.4% for *COI*. Within the Amazon Only lineage, divergence was 0.1% for *ND2* and zero for *COI*, whereas no variation whatsoever was found within the Amazon/Tocantins-Araguaia lineage in either marker.

The alleles identified by Araripe et al. (2013) from seven specimens of the Amazon Only Lineage (from the Mamirauá and Santarém) and 32 specimens of the Amazon/Tocantins-Araguaia Lineage (from the Mamirauá, Santarém, and Tucuruí) found no significant difference between observed and expected heterozygosity. The Bayesian analysis of these loci revealed the presence of two distinct genetic groups (K = 2; mean Ln Prob = -550.140), although they were present in each of the two mitochondrial lineages identified (Supplementary Material – Figure S2). The methodological limitations of Evanno's estimate of the number of population units mean that the possibility of the existence of a single stock (K = 1) cannot be excluded, although the profile of the contribution of the genetic clusters to each individual in our database does not support this conclusion (K=1; mean Ln Prob = -588.430).

Table II. Genetic differentiation matrix (pairwise Fst) based on the concatenated mitochondrial (*Cytb* and Control Region) data set (below the diagonal) and the geographical distances (km) between *Arapaima* populations.

Fst / Geographical Distances	IQ	LT	MM	JR	ST	TC	CS	LC	NS	QB
IQ	-	476	1342	2584	2775	3828	4886	4880	5055	5481
LT	0.037	-	866	2108	2299	3352	4410	4404	4579	5055
MM	0.027	0.059	-	1242	1433	2486	3544	3538	3713	4139
JR	0.143*	0.132*	0.039	-	191	1244	2302	2296	2471	2897
ST	0.382*	0.402*	0.247*	0.142*	-	1053	2111	2105	2280	2706
TC	0.399*	0.535*	0.407*	0.543*	0.691*	-	1058	1052	1227	1643
CS	0.223*	0.362*	0.289*	0.378*	0.563*	0.136*	-	245	414	826
LC	0.254*	0.399*	0.304*	0.406*	0.587*	0.063	0.09	-	261	673
NS	0.366*	0.495*	0.385*	0.501*	0.652*	0.324*	0.261*	0.296*	-	412
QB	0.652*	0.707*	0.563*	0.652*	0.748*	0.897*	0.922*	1.000*	0.807*	-

Iquitos (IQ), Letícia (LT), Mamirauá (MM), Jurutí (JR), Santarém (ST), Tucuruí (TC), Caseara (CS), Lagoa da Confusão (LC), Novo Santo Antônio (NS) and Quatro Bocas (QB). *p<0.05.

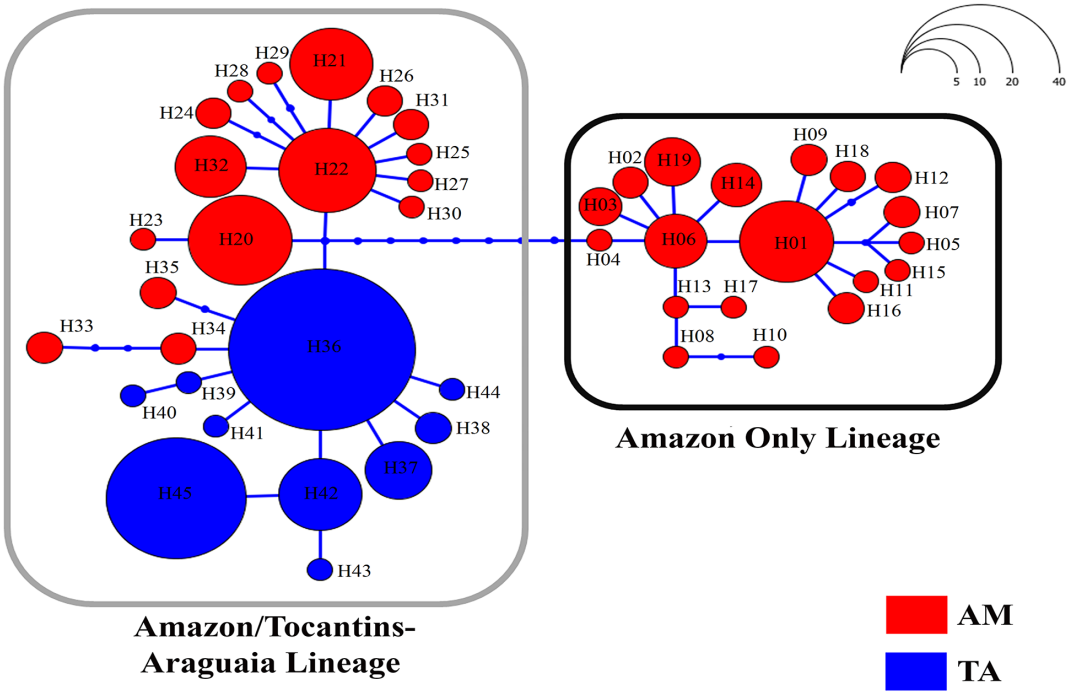

Figure 2. Genetic relationships among the Arapaima haplotypes. Haplotype network produced by the Maximum Likelihood method, based on the concatenated Cytb and Control Region data set. The Amazon populations are represented by red circles, and the Tocantins-Araguaia populations, by blue circles. The Amazon Only Lineage is outlined in black, and the Amazon/Tocantins-Araguaia Lineage, in gray.

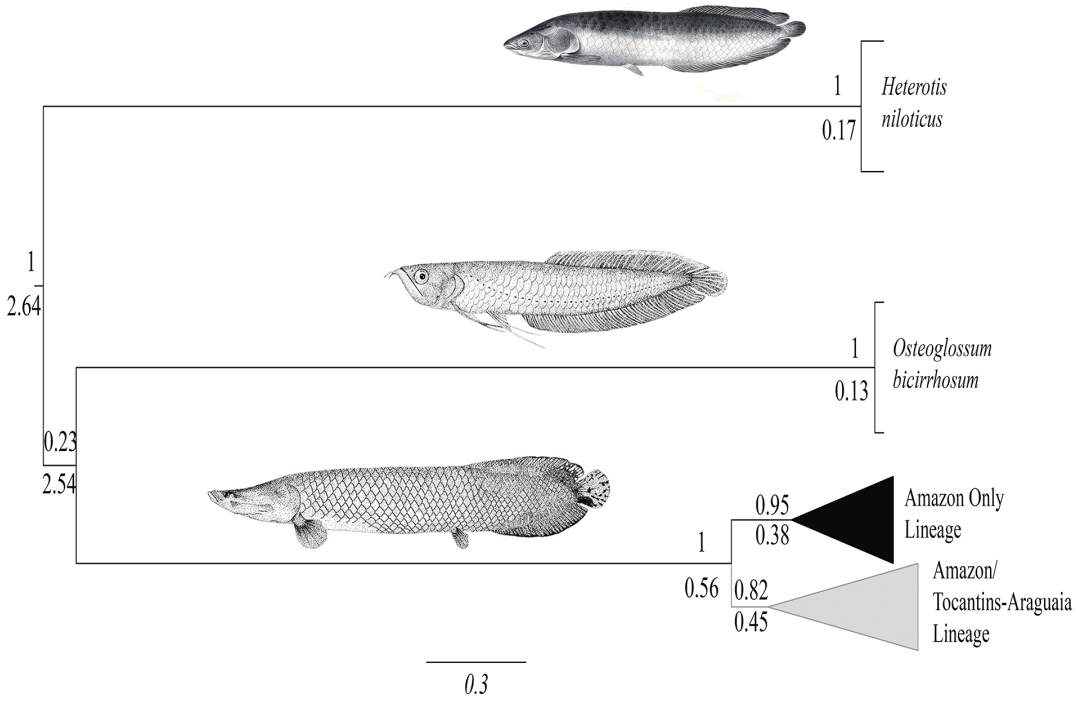

Figure 3. Topology showing the evolutionary relationships among Osteoglossidae taxa, based on the analysis of the Cytb gene, by Bayesian Inference. The values above the nodes show the statistical support, while those below the nodes indicate the mean coalescence time (Kyr). Illustrative picture of the Boulenger (1907) and MUSA (2017).

DISCUSSION

The genetic analyses of the natural arapaima populations of the Amazon and Tocantins-Araguaia basins revealed a clear pattern of differentiation between the basins, which did not share haplotypes. Two mitochondrial lineages were identified within the studied area, which diverged during the Pleistocene. This population structure is likely the result of a historical process of divergence, which is reflected in the occurrence of sympatry between the two lineages within the Amazon basin in the present day.

The genetic analysis of arapaima based on hypervariable mitochondrial markers allowed for the identification of different levels of genetic diversity within its geographic range. Significantly higher levels of genetic variability were observed in the Amazon basin in comparison with the Tocantins-Araguaia basin. The coexistence of the two mitochondrial lineages in the Amazon basin, in addition to factors such as the much larger populations size found in this basin, and its greater environmental heterogeneity, may be related to the much greater diversity of this aquatic system. Population surveys in areas of managed fishery in the Amazon basin indicated the existence of areas with high densities of arapaima, such as Mamirauá, where at least 50,000 individuals may be found within an area of approximately 1000 km^2 (Arantes et al. 2006), whereas in the Tocantins-Araguaia basin, the data on fishery catches indicate the existence of very much smaller populations (Mérona et al. 2010).

The Tocantins-Araguaia basin returned the lowest indices of diversity found within the study area and is characterized by much different morphodynamic features in comparison with the Amazon, including shorter periods of flooding (Latrubesse & Stevaux 2002, Ramalho et al. 2009). These features are likely to underpin fundamental differences in the ecology of arapaima, including the dispersal patterns of the offspring and adults. As the dispersal of juvenile and adult arapaima in floodplain environments is closely related to the annual flooding cycle (Castello 2008), changes in the hydrological dynamics of the river may impact migration patterns in this fish. The potential occurrence of population bottlenecks associated with genetic drift may also be responsible for the reduced genetic variability and the small number of haplotypes found in the Tocantins-Araguaia basin.

The genetic profile observed in the Tocantins-Araguaia basin reinforces the need for the adoption of effective management strategies in this region, where, in addition to the presence of exclusive haplotypes, there is a significant reduction in genetic variability. Vitorino et al. (2017) also recorded low levels of genetic heterozygosity in the arapaima populations of the Tocantins-Araguaia basin. While the results of the present demographic analysis indicated that the size of the population has remained relatively stable over time, this may reflect the type of marker analyzed, given that mitochondrial markers are appropriate for the recovery of historical events, with little potential resolution for the detection of recent processes, such as overfishing. Overall, these low indices may reflect a reduction in effective population size, followed by endogamous mating, which may erode the evolutionary potential of the species (Frankham et al. 2010).

The hypothesis that the genetic divergence between arapaimas is derived primarily from the geographic isolation of populations has been refuted progressively (Araripe et al. 2013, Fazzi-Gomes et al. 2017a, Vitorino et al. 2017). The factors most frequently associated with this structuring are historical reductions in stocks (Araripe et al. 2013, Vitorino et al. 2015, 2017), the sedentary behavior of the species (Araripe et al. 2013, Vitorino et al. 2015, 2017, Fazzi-Gomes et al. 2017a), the impact of fisheries on stocks (Hrbek et al. 2005, Vitorino et al. 2015, 2017), and the characteristics of each basin and its floodplain dynamics (Vitorino et al. 2015, 2017, Watson et al. 2016). It does seem likely that these processes may have influenced the structuring of the arapaima populations over time, molding their connectivity and restricting gene flow between the different hydrographic basins. The evidence of an historic process of differentiation, found in the present study, may have resulted in the current sympatry of the distinct mitochondrial lineages within the Amazon basin.

The current sympatric distribution of arapaima lineages and the analyses of nuclear markers (microsatellites) indicates that they have been subject to gene flow ever since the separation process. Given this, the varying patterns observed in the different studies are probably related primarily to the different markers analyzed (Toews & Brelsford 2012), given that the split which gave rise to the differentiation of the mitochondrial lineages was not long enough ago to ensure complete isolation and impede mixing events. One possible explanation for the pattern of genetic structure observed in arapaima is related to the evolution of the landscape of the Amazon region (Hoorn et al. 2010a, b, Shephard et al. 2010), given that a contact zone may have arisen between lineages, in particular on the lower Amazon. The analyses indicated that the

arapaima lineages separated approximately 564 Kyr ago, during the Pleistocene. Hrbek et al. (2005) point to a possible separation event in the region of Macapá, coinciding with this period, and the shifts in sea level that led to the eastward expansion of the Amazon River.

The divergence observed between the sympatric lineages from the Amazon basin may reflect the partial isolation of different parts of this basin during the Pleistocene (Harris & Mix 1999), followed by secondary contact. During the Pleistocene, different bodies of water may have been isolated by topographic formations, such as the Purus Arch, which may have isolated the westernmost populations from the rest of the Amazon basin. This same model has been used to account for the population structuring found in other groups of Neotropical fish, such as the Cichlidae (Farias & Hrbek 2008) and the Characidae (de Queiroz et al. 2017). Given the current frequencies of the different lineages within the basins, it seems likely that the lineage with the most ample current distribution was the first to colonize the two basins. The formation of geographic barriers during the glaciations of the Pleistocene (Harris & Mix 1999) may have contributed to the isolation of bodies of water in the middle and lower Amazon basin, leading to the differentiation of the Amazon Only lineage. The subsequent rise in sea level would have led to secondary contact and dispersal between the two lineages.

In turn, the differentiation of the mitochondrial lineage that is exclusive to the Amazon basin may have been influenced by a series of processes that occurred during the formation of the Amazon River, which are also relevant to the evolution of many of the region's plants and animals (Aleixo et al. 2007, Hoorn et al. 2010a, b, Hovikoski et al. 2010). The physical, geomorphological and chemical characteristics of the local aquatic environments vary considerably within the Amazon basin, and these conditions probably varied considerably over time, throughout the formation of the basin (Figueiredo et al. 2009, Junk et al. 2011). Arapaima appears to be capable of adapting to these varying environments, and was already widely distributed within this basin. At the present time, the arapaima is found at higher densities in floodplains and is less abundant in rivers and other environments with relatively high temperatures and low pH (Castello 2008). Further research, that include an in-depth analysis of the factors that determine the phylogeographic profile of arapaima is needed. Within the Amazon basin, the frequency of occurrence of the representatives of each mitochondrial lineage indicated a predominance of specimens from the Amazon Only Lineage in the lower (eastern) Amazon basin, while specimens belonging to the Amazon/Tocantins-Araguaia Lineage are more common in the upper (western) Amazon basin.

The structuring observed in the Tocantins-Araguaia basin, which has only a single lineage, and the high F*st* values recorded in the extremes of this basin are consistent with previous studies in this region, which indicate low levels of gene flow and structuring in the resident populations (Vitorino et al. 2015, 2017). Historical population bottlenecks, followed by genetic drift may have further reinforced this structuring. These bottlenecks may have been a consequence of the reduced flood pulse of the Tocantins-Araguaia river system, which results in limited connectivity between isolated environments (floodplain lakes) within the basin (Latrubesse & Stevaux 2002, Hatanaka et al. 2006, Aquino & Latrubesse 2008). The accentuated genetic structuring observed in the specimens from Quatro Bocas, in the southern extreme of the species' range, is related to the geographic isolation of this population, and possibly also a reduction in the local population.

The level of divergence observed between the lineages in the *COI* and *ND2* genes reinforces the conclusion that the specimens analyzed represent only a single species. In pairs of congeneric osteoglossid fish species, Mu et al. (2012) recorded a divergence of 8.6% in the *ND2* sequences between *O. bicirrhosum* and *O. ferreirai*, for example, and 23.7% between *S. formosus* and *S. leichardti*, values dozens of times higher than the 0.3% difference recorded here in the arapaima.

Similarly, the divergence of 0.4% estimated between the arapaima lineages in the *COI* region is well below the 2% threshold for valid species applied in the DNA barcoding approach (Herbert et al. 2003). In fact, the divergence in the *COI* region found between other congeneric osteoglossids ranges from 16% between *S. formosus* and *S. leichardti* (40 times higher than that recorded between the arapaima lineages) and 25% between *S. formosus* and *S. jardinii* (more than 60 times higher than that found in the arapaima). In a recent study of *Triportheus albus*, de Queiroz et al. (2017) used these same criteria to confirm that different genetic lineages belonged to the same taxon. In this case, the intraspecific divergence in the *COI* region of *T. albus* ranged from 0.3% to 0.7%, in comparison with a divergence of 17.5% between *T. brachipomus* and *T. guentheri*. While new *Arapaima* species have been proposed in recent years, such as *A. agassizii* Valenciennes, 1847 and *A. leptossoma* Stewart, 2013 (Stewart 2013a, b), these forms are described as being very rare, with no recent records from the Amazon region. However, the holotype of *A. agassizii* is known only from illustrations, while the holotype of *A. leptossoma* is an immature specimen (SL = 77.6 cm) deposited in the ichthyological collection of the National Amazonian Research Institute (INPA) in Manaus (INPA-16847).

The microsatellite markers analyzed in the present study also confirmed that the specimens represent a single taxon, given that these highly variable markers did not indicate any structuring between the lineages, but rather between the stocks that contribute differentially to the genetic composition of the arapaima in the two basins (Supplementary Material – Figure S2). Escobar et al. (2015) also used mitochondrial sequences and microsatellites to confirm the existence of a single species of *Piaractus brachypomus* in the Amazon and Orinoco basins, but with distinct Evolutionarily Significant Units (ESUs).

Based on the results of the present study, together with those of previous studies that have demonstrated the effectiveness of managed fisheries for the recovery of arapaima populations (Amaral & Almeida 2013, Araripe 2013, Arantes & Castello 2013), we would recommend the more widespread adoption of more effective measures for the conservation of this resource, not only in the Amazon region, but also in the Tocantins-Araguaia basin, adapted to the scenario found in each region. Up to now, commercial arapaima fisheries in Brazil have been regulated through normative instructions at both the national (federal) and state levels. The initial federal norm established a minimum body length of 150 cm for the harvesting of the species, and was followed by the definition of closed seasons, which were different for the Amazon and Tocantins-Araguaia basins, due to their distinct flood pulses, which determine the reproductive cycle. Subsequently, a minimum size was established for the marketing of processed sides of arapaima, which must be at least 120 cm long when fresh or 110 cm, when salted. Unfortunately, however, the vast area inhabited by the species has limited the effectiveness of these measures, and illegal harvesting is a major threat to arapaima stocks, especially in

areas with reduced genetic diversity, such as the Tocantins-Araguaia basin.

Incentives for the implementation of these measures will be fundamental to the conservation of the arapaima, especially considering the genetic diversity found within the distribution of the species, and its varying ecological and behavioral characteristics in different environments. Further ecological and genetic studies are also needed, especially in key areas, such as the Amazon estuary, the Negro, Purus, and Madeira rivers, in Brazil, and the Essequibo River in Guyana. The assessment of the genetic diversity of other areas of fishery management will also be important to ensure the development and implementation of effective conservation strategies, adapted to the local characteristics of the target populations.

ACKNOWLEDGMENTS

The authors are grateful to the reviewers for their important contributions to improving the manuscript. We would like to thank, Issakar Lima Sousa, Ruiter Braga and IDSM for support in obtaining samples. We are also grateful to Pró-Reitoria de Pesquisa e Pós-Graduação PROPESP/UFPA.

REFERENCES

AKAIKE H. 1974. A new look at the statistical model identification. IEEE Trans on Automatic Control 19: 716-723.

ALEIXO A & DE FÁTIMA-ROSSETTI D. 2007. Avian gene trees, landscape evolution, and geology: towards a modern synthesis of Amazonian historical biogeography? J Ornithol 148: 443-453.

AMARAL E & ALMEIDA O. 2013. Produtividade e eficiência econômica da pesca de pirarcucu (*Arapaima gigas*) nas áreas de manejo das Reservas Amanã e Mamirauá. In: Amaral E (Eds), BIOLOGIA, CONSERVAÇÃO E MANEJO PARTICIPATIVO DE PIRARUCUS NA PAN-AMAZÔNIA. Tefé-AM: Instituto de Desenvolvimento Mamirauá, p. 151-162.

AQUINO S & LATRUBESSE EM. 2008. Relações entre o regime hidrológico e os ecossistemas aquáticos da planície aluvial do rio Araguaia. Acta Sci Bio Sci 30.

ARANTES CC & CASTELLO L. 2013. Implicações da biologia, ecologia e contagens para o manejo do pirarucu. In: Amaral E (Eds), Biologia, Conservação e Manejo Participativo de Pirarucus na Pan-Amazônia. Tefé-AM: Instituto de Desenvolvimento Mamirauá, p. 33-42.

ARANTES CC, GARCEZ DS & CASTELLO L. 2006. Densidades de pirarucu (*Arapaima gigas*, Teleostei, Osteoglossidae) em lagos das Reservas de Desenvolvimento Sustentável Mamirauá e Amanã, Amazonas, Brasil. UAKARI 2: 37-43.

ARARIPE J, RÊGO PS, QUEIROZ H, SAMPAIO I & SCHNEIDER H. 2013. Dispersal capacity and genetic structure of *Arapaima gigas* on different geographic scales using microsatellite markers. PloS One 8: e54470.

ARARIPE J. 2013. Avaliação genética do manejo do pirarucu (*Arapaima gigas*) na Reserva Mamirauá. In: Amaral E (Eds), Biologia, Conservação e Manejo Participativo de Pirarucus na Pan-Amazônia. Tefé-AM: Instituto de Desenvolvimento Mamirauá, p. 69-86.

AYRES JM. 2006. As Matas de Várzeas do Mamirauá, 3ª ed., Belém: Sociedade Civil Mamirauá, 123p.

BOSSUYT F & MILINKOVITCH MC. 2000. Convergent adaptive radiations in Madagascan and Asian ranid frogs reveal covariation between larval and adult traits. Proc Natl Acad Sci USA 97: 6585-6590.

BOULENGER GA. 1907. The fishes of the Nile. London (UK): Egyptian Government, 558 p.

BRITO MA, SCHENEIDER H, SAMPAIO I & SANTOS S. 2015. DNA barcoding reveals high substitution rate and mislabeling in croaker fillets (Sciaenidae) marketed in Brazil: The case of "pescada branca" (*Cynoscion leiarchus* and *Plagioscion squamosissimus*). Food Res Int 70: 40-46.

CASTELLO L. 2008. Lateral migration of *Arapaima gigas* in floodplains of the Amazon. Ecol Freshw Fish 17: 38-46.

CITES. 2017. Convention on International Trade in Endangered Species of Wild Fauna and Flora. 2017. [accessed 2018 Mar 23]. http://www.cites.org/esp/app/index.php.

DARRIBA D, TABOADA GL, DOALLO R & POSADA D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9: 772-772.

DE QUEIROZ LJ, TORRENTE-VILARA G, QUILODRAN C, DA COSTA DORIA CR & MONTOYA-BURGOS JI. 2017. Multifactorial genetic divergence processes drive the onset of speciation in an Amazonian fish. PloS One12: e0189349.

DRUMMOND AJ, SUCHARD M A, XIE D & RAMBAUT A. 2012. Bayesian phylogenetics with BEAUTI and the BEAST 1.7. Mol Biol Evol 29: 1969-1973.

EARL DA & VONHOLDT BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4: 359-361.

ESCOBAR LMD, ANDRADE-LÓPEZ J, FARIAS IP & HRBEK T. 2015. Delimiting Evolutionarily Significant Units of the Fish, *Piaractus brachypomus* (Characiformes: Serrasalmidae), from the Orinoco and Amazon River Basins with Insight on Routes of Historical Connectivity. J Herd 106: 428-438.

EVANNO G, REGNAUT S & GOUDET J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14: 2611-2620.

EXCOFFIER L & LISCHER HE. 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.

FARIAS IP & HRBEK T. 2008. Patterns of diversification in the discus fishes (*Symphysodon* spp. Cichlidae) of the Amazon basin. Mol Phylogenet Evol. 49: 32-43.

FAZZI-GOMES PF, MELO N, PALHETA G, GUERREIRO S, AMADOR M, RIBEIRO-DOS-SANTOS AK & HAMOY I. 2017a. Genetic diversity and differentiation in natural populations of *Arapaima gigas* from lower Amazon revealed by microsatellites. Genet Mol Res 16.

FAZZI-GOMES P, GUERREIRO S, PALHETA GDA, MELO NFAC, SANTOS S & HAMOY IG. 2017b. High genetic diversity and connectivity in *Colossoma macropomum* in the Amazon basin revealed by microsatellite markers. Genet Mol Biol 40: 142-146.

FIGUEIREDO J, HOORN C, VAN DER VEN P & SOARES E. 2009. Late Miocene on set of the Amazon River and the Amazon deep-sea fan: Evidence from the Foz do Amazonas Basin Geology. Geol 37: 619-612.

FRANKHAM R, BALLOU JD & BRISCOE DA. 2010. Introduction to conservation genetics. New York: Cambridge University Press, 618 p.

GUINDON S, DUFAYARD JF, LEFORT V, ANISIMOVA M, HORDIJK W & GASCUEL O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59: 307-321.

HALL TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Sympo 41: 95-98.

HARRIS SE & MIX AC. 1999. Pleistocene precipitation balance in the Amazon Basin recorded in deep sea sediments. Quat Res. 51: 14-26.

HATANAKA T, HENRIQUE-SILVA F & GALETTI JRPM. 2006. Population substructuring in a migratory freshwater fish *Prochilodus argenteus* (Characiformes, Prochilodontidae) from the São Francisco River. Genetica 126: 153-159.

HEBERT PD, CYWINSKA A & BALL SL. 2003. Biological identifications through DNA barcodes. Biol Sciences 270: 313-321.

HOORN C, WESSELINGH, FP, HOVIKOSKI J & GUERRERO J. 2010A. The development of the amazonian mega-wetland (Miocene; Brazil, Colombia, Peru, Bolivia). In: Hoorn C & Wesselingh FP (Eds), AMAZONIA LANDSCAPE AND SPECIES EVOLUTION: A LOOK INTO THE PAST. A John wiley & Sons, p. 123-142.

HOORN C ET AL. 2010B. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. Science 330: 927-931.

HOVIKOSKI J, WESSELINGH FP, RÄSÄNEN M, GINGRAS M & VONHOF HB. 2010. Marine influence in Amazonia: evidence from the geological record. In: Hoorn C & Wesselingh FP (Eds), Amazonia Landscape and Species Evolution: A Look Into The Past. John Wiley & Sons, p. 143-160.

HRBEK T, CROSSA M & FARIAS IP. 2007. Conservation strategies for *Arapaima gigas* (Schinz 1822) and the Amazonian várzea ecosystem. Bras J Biol 67: 909-917.

HRBEK T, FARIAS IP, CROSSA M, SAMPAIO I, PORTO JI & MEYER A. 2005. Population genetic analysis of *Arapaima gigas* one of the largest freshwater fishes of the Amazon basin: implications for its conservation. Anim Conserv 8: 297-308.

IUCN. 2017. IUCN Red List of Threatened Species. 2017. [accessed 2018 Mar 23]. Available at http://www.iucnredlist.org/details/1991/0.

JAKOBSSON M & ROSENBERG NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. BMC Bioinformatics 23: 1801-1806.

JUNK WJ, PIEDADE MTF, SCHÖNGART J, COHN-HAFT M, ADENEY JM & WITTMANN F. 2011. A classification of major naturally-occurring Amazonian lowland wetlands. Wetlands 31: 623-640.

LATRUBESSE EM & STEVAUX JC. 2002. Geomorphology and environmental aspects of the Araguaia fluvial basin, Brazil. Zeitschrift fur Geomorphologie Supplementband, 109-127.

LIBRADO P & ROZAS J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. BMC Bioinformatics 25: 1451-1452.

MÉRONA BD, JURAS AA, MENDES DOS SANTOS G & CINTRA IHA. 2010. Os peixes e a pesca no baixo rio Tocantins: vinte anos depois da UHE Tucuruí. http://www.eletronorte.gov.br/opencms/export/sites/eletronorte/publicacoes/publicacoes/Os_Peixes_e_a_Pesca_no_Baixo_Rio_Tocantins.pdf. Accessed May 15, 2017.

MU XD, WANG XJ, SONG HM, YANG YX, LUO D, GU DE, XU M, LIU C, LUO JR & HU YC. 2012. Mitochondrial DNA as effective molecular markers for the genetic variation and phylogeny of the family Osteoglossidae. Gene 511: 320-325.

MUSA. 2017. Amazonas (BR): Museu da Amazônia, [accessed 2017 Oct 22]. http://museudaamazonia.org.br/.

NELSON JS, GRANDE CM & WILSON VH. 2016. Fishes of the world. 5th ed., Hoboken, New Jersey: John Wiley & Sons, 752 p.

PAITHANKAR KR & PRASAD KS. 1991. Precipitation of DNA by polyethylene glycol and etanol. Nucleic Acids Res 19: 1346.

PRITCHARD JK, STEPHENS M, ROSENBERG NA & DONNELLY. 2000. Association mapping in structured populations. Am J Hum Genet 67(2000): 170-181.

R DEVELOPMENT CORE TEAM. 2017. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing: 1-2. [accessed 2017 May 31]. http://www.r-project.org/.

RAMALHO EE, MACEDO J, VIEIRA TM, VALSECCHI J, MARMONTEL M & QUEIROZ HL. 2009. Ciclo Hidrológico nos Ambientes de Várzea da Reserva de Desenvolvimento Sustentável Mamirauá–Médio Rio Solimões, Período De 1990 A 2008. UAKARI 51: 61-87.

ROGERS AR & HARPENDING H. 1992. Population growth makes waves in the distribution of pairwise genetic diferences. Mol Biol Evol 9: 552-569.

ROSENBERG NA. 2004. DISTRUCT: a program for the graphical display of population structure. Mol Ecol Resour 4: 137-138.

SALZBURGER W, EWING GB & VON HAESELER A. 2011. The performance of phylogenetic algorithms in estimating haplotype genealogies with migration. Mol Ecol 20: 1952-1963.

SAMBROOK J & RUSSELL DW. 2001. Molecular cloning: a laboratory manual. New York. Cold Spring Harbor Laboratory.

SHEPHARD GE, MÜLLER RD, LIU L & GURNIS M. 2010. Miocene drainage reversal of the Amazon River driven by plate-mantle interaction. Nature Geosc 3: 870-875.

STEWART DJ. 2013a. Re-description of *Arapaima agassizii* (Valenciennes) a rare fish from Brazil (Osteoglossomorpha: Osteoglossidae). Copeia. 2013: 38-51.

STEWART DJ. 2013b. A New Species of *Arapaima* (Osteoglossomorpha: Osteoglossidae) from the Solimões River Amazonas State Brazil. Copeia 2013: 470-476.

TAMURA K, STECHER G, PETERSON D, FILIPSKI A & KUMAR S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30: 2725-2729.

TOEWS DP & BRELSFORD A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol Ecol 21: 3907-3930.

VITORINO CA, NOGUEIRA F, SOUZA IL, ARARIPE J & VENERE PC. 2017. Low genetic diversity and structuring of the *Arapaima* (Osteoglossiformes, Arapaimidae) population of the Araguaia-Tocantins basin. Front Genet 8:1-10.

VITORINO CA, OLIVEIRA RCC, MARGARIDO VP & VENERE PC. 2015. Genetic diversity of Arapaima gigas (Schinz, 1822) (Osteoglossiformes: Arapaimidae) in the Araguaia-Tocantins basin estimated by ISSR marker. Neotrop Ichthyol 3: 557-568.

WARD RD, ZEMLAK TS, INNES BH, LAST PR & HEBERT PD. 2005. DNA barcoding Australia's fish species. Biol Sciences 360: 1847-1857.

WATSON LC, STEWART DJ & KRETZER AM. 2016. Genetic Diversity and Population Structure of the Threatened Giant *Arapaima* in Southwestern Guyana: Implications for Their Conservation. Copeia 104: 864-872.

ZARDOYA R & DOADRI I. 1999. Molecular Evidence on the Evolutionary and Biogeographical Patterns of European Cyprinids. J Mol Evol 49: 227-237.

SUPPLEMENTARY MATERIAL

Table SI. Distribution of haplotypes based on the analysis of the concatenated data set (Cytb and Control Region). GenBank accession numbers for the Ctyb/Control Region. The Amazon Only Lineage is outlined in black, and the Amazon/Tocantins-Araguaia Lineage, in gray. The Amazon basin is delimited in red, and the Tocantins-Araguaia basin, in blue. IQ-Iquitos, LT-Letícia, MM-Mamirauá, JR-Jurutí, ST-Santarém, TC-Tucuruí, CS-Caseara, LC-Lagoa da Confusão, NS-Novo Santo Antônio and QB-Quatro Bocas.

Figure S1. Bayesian Skyline Plot (*Cytb* and Control Region) showing the estimated effective size of the

***Arapaima* population over time. The solid dark blue line represents the mean values and the blue shaded area represents the 95% confidence interval of the historic effective population size. The graphs show: A) Amazon Only Lineage, B) Amazon/Tocantins-Araguaia Lineage – populations of the Amazon basin and C) Amazon/Tocantins-Araguaia Lineage – populations of the Tocantins-Araguaia basin.**

Figure S2. Population structure in *Arapaima* based on a Bayesian analysis, without an *a priori* definition of the stocks. MM-Mamirauá, ST-Santarém and TC-Tucuruí.

How to cite
NOGUEIRA F, RÊGO PS, QUEIROZ H, VENERE P, VARELA ES, SAMPAIO I, SCHNEIDER H & ARARIPE J. 2020. Genetic diversity and structuring in the arapaima (Osteoglossiformes, Osteoglossidae) population reveal differences between the Amazon and the Tocantins-Araguaia basins. An Acad Bras Cienc 92: e20180496. DOI 10.1590/0001-3765202020180496.

Manuscript received on May 18, 2018; accepted for publication on November 13, 2018

FABRÍCIA NOGUEIRA[1]
https://orcid.org/0000-0002-8658-075X

PÉRICLES S. DO RÊGO[1]
https://orcid.org/0000-0002-4537-946X

HELDER QUEIROZ[2]
https://orcid.org/0000-0002-4425-3208

PAULO VENERE[3]
https://orcid.org/0000-0001-7236-8857

EDUARDO S. VARELA[4]
https://orcid.org/0000-0003-0424-2683

IRACILDA SAMPAIO[5]
https://orcid.org/0000-0002-2137-4656

HORACIO SCHNEIDER[5]
https://orcid.org/0000-0002-5987-6395

JULIANA ARARIPE[1]
https://orcid.org/0000-0002-8014-3081

[1]Laboratório de Genética e Conservação, Instituto de Estudos Costeiros, Universidade Federal do Pará, Alameda Leandro Ribeiro, S/N, Aldeia, 68610-000 Bragança, PA, Brazil

[2]Instituto de Desenvolvimento Sustentável Mamirauá, Estrada do Bexiga, 2584, Fonte Boa, 69470-000 Tefé, AM, Brazil

[3]Instituto de Biociências, Universidade Federal de Mato Grosso, Avenida Fernando Corrêa da Costa, 2367, Boa Esperanca, 78060-900 Cuibá, MT, Brazil

[4]Embrapa Pesca e Aquicultura, Palmas, Prolongamento da Avenida NS 10, cruzamento com a Avenida LO 18, Sentido Norte, Loteamento Água Fria, 77015-012 Santo Amaro, TO, Brazil

[5]Laboratório de Genética e Biologia Molecular, Instituto de Estudos Costeiros, Universidade Federal do Pará, Alameda Leandro Ribeiro, S/N, Aldeia, 68610-000 Bragança, PA, Brazil

Correspondence to **Juliana Araripe**
E-mail: araripe@ufpa.br

Author contributions

Collection and sample loaning: FN, HQ, PV, ESV and JA; DNA extraction, sequencing and genotyping: FN and JA; Genetic analysis: FN, PSR and JA; Interpretation of data and writing of the manuscript: FN, PSR, HQ, PV, ESV, IS, HS and JA; all authors approved the final version of the manuscript.