GMR

Genetic diversity and differentiation in natural populations of *Arapaima gigas* from lower Amazon revealed by microsatellites

P.F. Fazzi-Gomes[1], N. Melo[2], G. Palheta[2], S. Guerreiro[2], M. Amador[1], A.K. Ribeiro-dos-Santos[1], S. Santos[1] and I. Hamoy[2]

[1]Laboratório de Genética Humana e Médica, Centro de Ciências Biológicas,
Universidade Federal do Pará, Belém, PA, Brasil
[2]Laboratório de Genética Aplicada,
Instituto Socioambiental e dos Recursos Hidricos,
Universidade Federal Rural da Amazônia, Belém, PA, Brasil

Corresponding author: I. Hamoy
E-mail: igor.hamoy@ufra.edu.br

ABSTRACT. Genetic variability is one of the important criteria for species conservation decisions. This study aimed to analyze the genetic diversity and the population differentiation of two natural populations of *Arapaima gigas*, a species with a long history of being commercially exploited. We collected 87 samples of *A. gigas* from Grande Curuai Lake and Paru Lake, located in the Lower Amazon region of Amazônia, Brazil, and genotyped these samples using a multiplex panel of microsatellite markers. Our results showed that the populations of *A. gigas* analyzed had high levels of genetic variability, which were similar to those described in previous studies. These two populations had a significant population differentiation supported by the estimates

of F_{ST} and R_{ST} (0.06), by Bayesian analysis (K = 2), and by population assignment tests, which revealed a moderate genetic distance.

INTRODUCTION

Arapaima gigas (Cuvier, 1829) is an endemic species of the Amazon Basin that inhabits lakes, floodplains, and flooded forests. It is the largest freshwater scale fish in the world and can weigh as much as 200 kg in body mass and can be 3 m in body length (Pereira-Filho and Roubach, 2010).

The activities of fishing and the trade of this species have great social and economic importance in the Amazon region, the records of the first commercial fisheries date back to the early eighteenth century (Veríssimo, 1895). However, the high commercial exploitation by the fishing industry over the years has led to a drastic reduction in the natural stocks of *A. gigas*. As a result, this species has become protected under the regulations of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (Hrbek et al., 2005). *A. gigas* is on the International Union for Conservation of Nature and Natural Resources (IUCN) Red List of Threatened Species, under the category of "Data Deficient" due to the lack of knowledge of this species (Arantes et al., 2011).

Genetic variability is one of the most important criteria for biodiversity conservation recommended by the IUCN (McNeely et al., 1990). To study the genetic variability of *A. gigas*, this study analyzed the genetic diversity and differentiation in two natural populations of *A. gigas* in the Lower Amazon using a multiplex panel of eight microsatellite markers.

MATERIAL AND METHODS

A total of 87 samples of *A. gigas* were collected, 41 of which from the Grande Curuai Lake (GCL), located on the south bank of the Amazon River. The other 46 were from Paru Lake (PAL), located on the north bank of the Amazon River; both lakes are located in the Lower Amazon, Pará State, Brazil (Figure 1A). A 10-g sample of muscle tissue was collected from each individual, preserved in 95% ethanol, and then stored at 4°C. Total genomic DNA was extracted from each tissue sample that had been digested in a proteinase K/sodium dodecyl sulfate solution, and DNA purification was performed using the standard phenol/chloroform method, followed by precipitation with isopropanol (Sambrook and Russell, 2001). The concentrations of the DNA samples were measured with a NanoDrop™ ND-1000 spectrophotometer (Thermo Scientific).

The samples were genotyped according to the protocol developed by Hamoy et al. (2008). In brief, the DNA products were visualized using the Applied Biosystems 3130 Genetic Analyzer and genotyped using the GeneMapper 3.7 software (Applied Biosystems). To check for possible genotyping errors and null alleles, the Micro-Checker program was used (van Oosterhout et al., 2004).

The observed (H_O) and expected (H_E) heterozygosities and their possible deviations from Hardy-Weinberg equilibrium (HWE) were calculated in the Arlequin 3.5.1.3 program (Excoffier and Lischer, 2010).
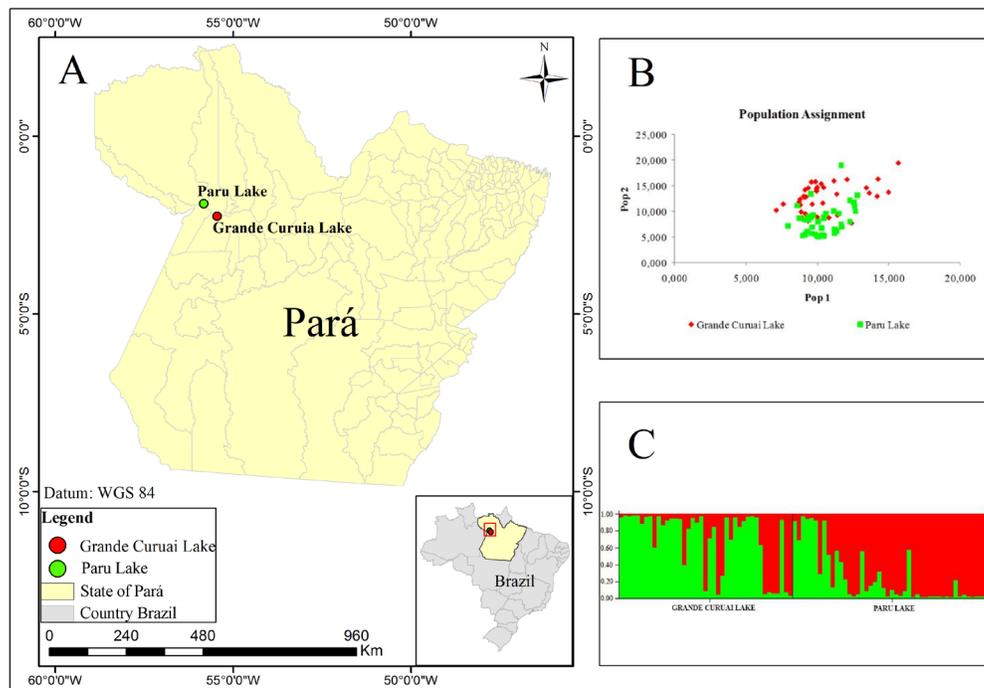
Figure 1. A. Map location of lakes where collections of native *Arapaima gigas* samples were performed. **B.** Representation of the genetic signatures of native populations of *Arapaima gigas*. **C.** Standard population structure supported by Bayesian analysis indicating the existence of two groups (K = 2).

Then, we determined the Bonferroni correction (Rice, 1989) P values. The number of alleles per locus (N_A) and allelic richness (A_R) were estimated using the program Fstat version 2.9.3.2 (Goudet, 2002). The polymorphic information content (PIC) (Hearne et al., 1992) was calculated using the Cervus 3.0 program (Kalinowski et al., 2007). The interpopulation genetic differentiation was observed in both the R_{ST} and the F_{ST} using the Arlequin 3.5.1.3 program (Excoffier and Lischer, 2010).

To infer the number of genetically homogenous populations (K) that most likely occurred during the study, we created a database in the Structure 2.2 program (Pritchard et al., 2000). In this Bayesian analysis, we ran 106 simulations with the K value ranging from 1 to 4. For the burn-in test, we ran 20,000 simulations. The most likely value of K was determined by the ΔK method described by Evanno et al. (2005) using the Structure Harvester 0.6.94 program (Earl and vonHoldt, 2012). We tested the population assignment using the program GenAlEx 6.5 (Peakall and Smouse, 2012) to determine the likelihood of inclusion of the individuals in each population.

RESULTS AND DISCUSSION

A total of 25 alleles were identified for the eight loci that were analyzed in the 87 *A. gigas* specimens. The N_A, A_R, H_O, H_E, and PIC values are shown in Table 1. The results of the Micro-Checker program did not indicate the presence of null alleles or stutter bands.

Table 1. Genetic diversity index of native populations of *Arapaima gigas* from the Grande Curuai Lake and Paru Lake.

Loci	Grande Curuai Lake (N = 41)					Paru Lake (N = 46)				
	N_A	A_R	H_O	H_E	PIC	N_A	A_R	H_O	H_E	PIC
AgCAm02	12	11.951	0.878	0.822	0.79	10	9.591	0.826	0.711*	0.67
AgCAm13	13	12.878	0.854	0.833	0.80	8	7.736	0.435	0.475	0.55
AgCAm15	6	5.999	0.707	0.610	0.53	6	5.723	0.783	0.633	0.57
AgCAm16	9	8.951	0.732	0.802	0.76	9	8.593	1.000	0.730*	0.68
AgCAm18	5	5.000	0.829	0.702	0.64	5	4.868	0.848	0.686	0.61
AgCTm03	6	6.000	0.829	0.778	0.73	7	6.852	0.891	0.739*	0.69
AgCTm05	7	6.975	0.756	0.769	0.72	6	5.852	0.652	0.649	0.58
AgCTm07	11	11.000	0.625	0.833*	0.82	7	6.854	0.609	0.753	0.71
Mean	8.62	8.594	0.776	0.768		7.25	7.008	0.755	0.672	

N_A = number of alleles, A_R = allelic richness, H_O = observed heterozygosity, H_E = expected heterozygosity. *Marker out of the Hardy-Weinberg equilibrium, after Bonferroni correction (adjusted P < 0.005).

We found most loci in HWE. However, there were four loci identified in HWE, three due to heterozygous excess and one due to heterozygous deficiency (Table 1).

The values of F_{ST} and R_{ST} between the populations of GCL and PAL were both equal to 0.06 (P < 0.5). The genetic signature test showed that 82% of the individuals were assigned to their local populations, whereas the other 18% came from another population (Figure 1B). The results of the Bayesian analysis supported the existence of two clusters (K = 2) in the database analyzed, with the GCL and the PAL populations each forming a distinct cluster (Figure 1C).

We adopted the multiplex panel method from Hamoy et al. (2008) to quantify the genetic variability of the populations of *A. gigas* that were investigated. Compared with other microsatellite analysis approaches, this approach provides a quick and inexpensive method of evaluation.

The average N_A values in the GCL and PAL populations of *A. gigas* were as high as and similar to those found by Hamoy et al. (2008), who found an average N_A = 8.0 for a population of *A. gigas* native to the Sauaçu Lake in the Lower Amazon region. In comparison, some studies using other microsatellite markers, such as the ones by Araripe et al. (2013) and Farias et al. (2003), found mean values of N_A of approximately 6.5 in populations of the Lower Amazon and southeast of Pará.

The average A_R values found in the GCL and PAL populations of *A. gigas* were higher than those reported by Araripe et al. (2013), in which the average A_R values for the four large populations analyzed were all greater than 5.5.

The average values of H_O in the GCL and PAL populations were similar to those found by Hamoy et al. (2008), who reported an H_O average of 0.69. The PIC markers in this study were all above 0.5 and are thus considered as informative and of good quality (Botstein et al., 1980).

We found that the GCL population had only one locus (AgCTm07) out of HWE, which was caused by heterozygous deficiency, whereas in the LAP population, the imbalance occurred in three loci. These loci were AgCAm02, AgCAm16 and AgCTm03, all of which were caused by heterozygous excess. The latter could potentially be explained by a sampling bias, as these loci showed no significant deviations in the GCL population. Moreover, Hamoy et al. (2008) used the same multiplex panel but found no locus with significant deviation from HWE, either.

Taken together, these findings suggest that the *A. gigas* populations of the Lower Amazon region analyzed in this study have high levels of genetic variability. This high-level variability was also verified by a study by Hrbek et al. (2005) on two segments of mtDNA, which showed that the expression of these segments of mtDNA were higher in samples from

the areas far from large urban centers of the Amazon, where the commercial exploitation of this fish has been reduced. The high-level genetic variability that we found in this study possibly reflects the effectiveness of the past and current regulations on the commercial exploitation of this species.

The populations of *A. gigas* that we investigated showed a moderate population differentiation based on their estimated F_{ST} and R_{ST}, according to the standards proposed by Hartl and Clark (2010). This differentiation is also supported by Bayesian analysis. The results of this study are consistent with the findings of Hrbek et al. (2005), who found a significant differentiation among various populations, and suggested that this differentiation was an artifact of anthropogenic actions caused by genetic drift, not by isolation by distance. Hrbek et al. (2007) studied various populations along the Amazon River and also found significant population differentiation, but their results were associated with isolation by distance. In addition, Araripe et al. (2013) reported moderate to high genetic differentiation in their population studies at the meso and large scales as well.

The genetic signature test showed that most individuals are classified as belonging to their local populations, which can be explained by the fact that *A. gigas* is considered as a sedentary species, which only performs small lateral migrations for a short distance (Castello, 2008). This result is different from another Amazonian species as reported by Fazzi-Gomes et al. (2017), who performed the same analysis in the native populations of *Colossoma macropomum* in the Amazon basin and found high levels of mixtures between populations, resulting in a panmictic population.

The moderate genetic differentiation and the high genetic variability that we found in the two populations of *A. gigas* of the Lower Amazon region are consistent with previous literature and are important for the development of the regulation plans on this species. Future studies on the characterization of population and the monitoring of genetic variability of other natural and farm-raised populations of *A. gigas* are needed for the understanding and protection of this species.

ACKNOWLEDGMENTS

REFERENCES

Arantes CC, Castello L, Cetra L and Schilling A (2011). Environmental influences on the distribution of *Arapaima* in Amazon floodplains. *Environ. Biol. Fish.* DOI 10.1007/s10641-011-9917-9.

Araripe J, do Rêgo PS, Queiroz H, Sampaio I, et al. (2013). Dispersal capacity and genetic structure of *Arapaima gigas* on different geographic scales using microsatellite markers. *PLoS One* 8: e54470. http://dx.doi.org/10.1371/journal. pone.0054470

Botstein D, White RL, Skolnick M and Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32: 314-331.

Castello L (2008). Lateral migration of *Arapaima gigas* in floodplains of the Amazon. *Ecol. Freshwat. Fish* 17: 38-46. http://dx.doi.org/10.1111/j.1600-0633.2007.00255.x

Earl DA and vonHoldt BM (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4: 359-361. http://dx.doi.org/10.1007/s12686-011-9548-7

Evanno G, Regnaut S and Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611-2620. http://dx.doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier L and Lischer HEL (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10: 564-567. http://dx.doi.org/10.1111/j.1755-0998.2010.02847.x

Farias IP, Hrbek T, Brinkmann H, Sampaio I, et al. (2003). Characterization and isolation of DNA microsatellite primer for *Arapaima gigas*, an economically important but severely over-exploited fish species of the Amazon basin. *Mol. Ecol. Notes* 3: 128-130. http://dx.doi.org/10.1046/j.1471-8286.2003.00375.x

Fazzi-Gomes P, Guerreiro S, Palheta GDA, Hamoy IG, et al. (2017). High genetic diversity and connectivity in *Colossoma macropomum* in the Amazon basin revealed by microsatellite markers. *Genet. Mol. Biol.*, in press.

Goudet J (2002). FSTAT, a Program to Estimate and Test Gene Diversities and Fixation Indices (version 2.9.3.2). Available at [http://www2.unil.ch/popgen/softwares/fstat.htm].

Hamoy IG, Santos EJM and Santos SEB (2008). Rapid and inexpensive analysis of genetic variability in *Arapaima gigas* by PCR multiplex panel of eight microsatellites. *Genet. Mol. Res.* 7: 29-32. http://dx.doi.org/10.4238/vol7-1gmr394

Hartl DL and Clark AG (2010). Principles of population genetics. 4rd edn. Sinauer Associates, Sunderland.

Hearne CM, Ghosh S and Todd JA (1992). Microsatellites for linkage analysis of genetic traits. *Trends Genet.* 8: 288-294. http://dx.doi.org/10.1016/0168-9525(92)90256-4

Hrbek T, Farias IP, Crossa M, Sampaio I, et al. (2005). Population genetic structure of *Arapaima gigas*, world's biggest freshwater fish: implications for conservation. *Anim. Conserv.* 8: 297-308. http://dx.doi.org/10.1017/S1367943005002210

Hrbek T, Crossa M and Farias IP (2007). Conservation strategies for *Arapaima gigas* (Schinz, 1822) and the Amazonian várzea ecosystem. *Braz. J. Biol.* 67 (Suppl): 909-917. http://dx.doi.org/10.1590/S1519-69842007000500015

Kalinowski ST, Taper ML and Marshall TC (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 16: 1099-1106. http://dx.doi.org/10.1111/j.1365-294X.2007.03089.x

McNeely JA, Miller KR, Reid WV, Mittermeier RA, et al. (1990). Conserving the world's biological diversity. IUCN, Gland, Switzerland: WRI, CI, WWF- US, and the Work Bank, Washington, DC.

Peakall R and Smouse PE (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28: 2537-2539. http://dx.doi.org/10.1093/bioinformatics/bts460

Pereira-Filho M and Roubach R (2010). Pirarucu (*Arapaima gigas*). In: Espécies nativas para a piscicultura no Brasil (Baldisserotto B and Gomes LC, eds.). Universidade Federal de Santa Maria, Santa Maria.

Pritchard JK, Stephens M and Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

Rice WR (1989). Analyzing tables of statistical tests. i43: 223-225.

Sambrook J and Russell D (2001). Molecular Cloning: A Laboratory Manual. 3rd edn. Cold Spring Harbor Laboratory Press, New York.

van Oosterhout C, Hutchinson WF, Wills DPM and Shipley P (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* 4: 535-538. http://dx.doi.org/10.1111/j.1471-8286.2004.00684.x

Veríssimo J (1895). A Pesca na Amazônia. Livraria Clássica, Rio de Janeiro.