

Imputation of genetic composition for missing pedigree data in Serrasalminae using morphometric data

Adriano Carvalho Costa^{1*}, Marcio Balestre², Hortência Aparecida Botelho³, Rilke Tadeu Fonseca de Freitas³, Richardson Cesar da Silva Gomes³, Sérgio Augusto de Sousa Campos⁴, Fábio Porto Foresti⁵, Diogo Teruo Hashimoto⁶, Diego Galetti Martins⁵, Fernanda Dotti do Prado⁵, Maria Andreia Corrêa Mendonça¹

¹Goiano Federal Institute – Dept. of Animal Science, Rod. Sul Goiana, km 01 – 75901-970 – Rio Verde, GO – Brazil.

²Federal University of Lavras – Dept. of Exact Science, C.P. 3037 – 37200-000 – Lavras, MG – Brazil.

³Federal University of Lavras – Dept. of Animal Science.

⁴Federal University of Lavras – Dept. of Food Science.

⁵São Paulo State University/School of Sciences – Biological Sciences Dept., Av. Eng. Luiz Edmundo Carrijo Coube, 1401 – 17033-360 – Bauru, SP – Brazil.

⁶São Paulo State University/FCAV – Biological Sciences Dept., Via de Acesso Prof. Paulo Donato Castellane, s/n – 14884-900 – Jaboticabal, SP – Brazil.

*Corresponding author <adriano.costa@ifgoiano.edu.br>

Edited by: Gerson Barreto Mourão

Received June 18, 2016

Accepted October 15, 2016

ABSTRACT: This study aimed to impute the genetic makeup of individual fishes of Serrasalminae family on the basis of body weight and morphometric measurements. Eighty-three juveniles, belonging to the genetic groups Pacu, Pirapitinga, Tambaqui, Tambacu, Tambatinga, Patinga, Paqui and Piraqui, were separated into 16 water tanks in a recirculation system, with two tanks per genetic group, where they remained until they reached 495 days of age. They were then weighed and analyzed according to the following morphometric parameters: Standard Length (SL), Head Length (HL), Body Height (BH), and Body Width (BW). The identity of each fish was confirmed with two SNPs and two mitochondrial markers. Two analyses were performed: one for the validating the imputation and another for imputing a genetic composition of animals considered to be advanced hybrids (post F1). In both analyses, we used linear mixed models with a mixture of normal distributions to impute the genetic makeup of the fish based on phenotype. We applied the mixed models method, whereby the environmental effects were estimated by the Empirical Best Linear Unbiased Estimator (EBLUE) and genetic effects are considered random, obtaining the Empirical Best Linear Unbiased Predictor (EBLUP) from the general (GCA) and the specific (SCA) combining ability effects. The results showed that validation of the genetic makeup imputation based on body weight can be used because of the strong correlation between the observed and imputed genotype. The fish classified as advanced hybrids had a genetic composition with a high probability of belonging to known genotypes and there was consistency in genotype imputation according to the different characteristics used.

Keywords: *Colossoma* spp., *Piaractus* spp., breeding, mixture model, round fish

Introduction

In Brazil, several breeders have been specializing in farming native species, as well as the production of hybrids to increase yield (Faustino et al., 2010; Hisano et al., 2013). Species of the family Serrasalminae, like Pacu (*Piaractus mesopotamicus*), Pirapitinga (*Piaractus brachypomus*) and Tambaqui (*Colossoma macropomum*), and their interspecific hybrids are the most important native fish in Brazilian aquaculture thanks to their high productivity (Ministério da Pesca e Aquicultura, 2012).

Studies of Serrasalminae fish are hampered by the classification of pure or hybrid by morphological evaluation (Gomes et al., 2012). Depending on the allelic interaction and the environment, hybrids can phenotypically resemble their parents, resulting in erroneous identification. Furthermore, effects of sex-linked genes (reciprocal crosses) and maternal and paternal effects can also influence the phenotypic variation in these fish (Griffing, 1956).

Molecular markers have been used with great precision to identify purebred animals and simple hybrids (Hashimoto et al., 2014). However, the low number of available markers is not enough for identifying more complex crosses, which can be identified only as post-F1 (i.e., information on the origin of the crosses is unknown).

Recently imputation through mixture distribution models has been widely used (Eirola et al., 2014). These models are a powerful tool for modeling in various areas. Finite mixtures have many applications. Direct applications emerge when each observation belongs to a subpopulation or group, although to which one is rarely known. In such a mixture, each subpopulation is described by its density, and the weightings of the mixtures are the probabilities of each observation belonging to a certain subpopulation (Luca and Zuccolotto, 2003). Indirect applications emerge when there is no division of data into subpopulations. In this case, the adjustment of the mixture is done by allowing great flexibility, such as multimodality (Abd-Elmaghd and Davis, 2006; Jang et al., 2006).

Mixture models are promising for the imputation of genetic makeup based on phenotype, especially in crosses with unknown parents. Therefore, this study aimed to impute the genetic composition of Serrasalminae fish on the basis of body weight and morphometric measurements when pedigree data were missing.

Materials and Methods

Biological material and experimental procedures

Data were collected in an experiment conducted in Lavras, Minas Gerais (21°14'43" S, 44°59'59" W, 919

m above sea level) for 465 days, divided between 180 days for adaptation and 285 experimental days. The experiment was carried out in a water recirculation laboratory, equipped with 16 water tanks each having a capacity of 500 liters. This system consisted of a bio-filter, pt100 probe, temperature controller (N540) and a pump (1/3 hp).

For this study, 96 Serrasalminae juveniles were acquired from two commercial fish farms in the state of São Paulo. These farmers reported there were 12 juveniles (30 days of age) from each of the following genetic groups (species and their hybrids): pacu, pirapitinga, tambaqui, patinga (♀pacu × ♂pirapitinga), paqui (♀pacu × ♂tambaqui), piraqui (♀pirapitinga × ♂tambaqui), tambacu (♀tambaqui × ♂pacu) and tambatinga (♀tambaqui × ♂pirapitinga). Upon arrival at the laboratory, the fish were placed in 16 indoor tanks (500 L) for a 5 month acclimatization period. During this adaptation period, 13 fish died, which reduced the number from 96 to 83. The stocking density was six fish per tank, with two tanks per genetic group. The tanks were controlled on the basis of three total renewals every hour and the temperature was kept at 28 °C throughout the study period. The fish were fed ad libitum with commercial food with 32 % crude protein, supplied three times a day (08h00 a.m. 12h00 a.m. and 04h00 p.m.) until reaching 100 g body weight, and then twice a day (08h00 a.m. and 02h00 p.m.) afterward. Dissolved oxygen was monitored daily and ammonia, nitrite, nitrate and pH every three days. At 180 days of age, the fish were counted and the values were used as fixed effects in the statistical model.

This study was approved by the Animal Experimentation Ethics Committee of Lavras University, Lavras, Brazil (Protocol 074/13 of 27 Feb 2014).

Morphometric evaluation

At the end of the experimental period, the fish were anaesthetized with benzocaine (60 mg L⁻¹), weighed and analyzed in terms of the following morphometric parameters: Standard Length (SL), between the tip of the snout to the end of the caudal fin; Head Length (HL), between the tip of the snout to the caudal border of the operculum; Body Height (BH), measured in front of the first dorsal fin ray; and Body Width (BW), measured in the region of the first dorsal fin ray.

Molecular analysis

To determine the group of each fish, fin samples were taken and fixed in absolute ethanol. Total DNA was extracted and purified with the Wizard Genomic DNA Purification Kit (Promega), following the manufacturer's recommendations. For identification of parentals and hybrids, two SNP (single nucleotide polymorphism) markers of nuclear genes and fragments of COI (cytochrome c oxidase subunit I) and CYTB (cytochrome B) mitochondrial genes were used for PC reactions. Both methods provide diagnostic electrophoretic fragments for each parental species and their interspecific hybrids (Table 1 and Figure 2). The sequences of the primers, restriction enzymes and reaction conditions were the same as those employed by Hashimoto et al. (2011). DNA fragment sizes were determined by electrophoresis on 2 % agarose gels stained with ethidium bromide (1 ng mL⁻¹) and visualized under UV illumination.

Data analysis

The analysis, including all 83 fish and their respective genetic composition assessed by the molecular approach, was performed to impute the genetic makeup of post F1 hybrids, based on all phenotypes measured (body weight and morphometric evaluation).

In a diallel context, we applied the method proposed by Griffing (1956), adapted to mixed model analysis, where the environmental effects were estimated through EBLUE and genetic effects were considered random, obtaining the EBLUP of the general (GCA) and specific (SCA) combining ability effects. However, since not all information about the parentals was available, the diallel analysis was performed by the mixed normal mixture model, which imputes the incidence matrices for the random effects (GCA and SCA) when an animal's genetic composition is unknown. To make inferences about the genetic composition of the sampled animals, we considered the imputed incidence matrix of GCA because it represents the additive genetic effect in the model. The reciprocal effect was not considered in the model. The mixed model presenting incomplete information for the pedigree can be given by:

$$y = X\beta + E[Z_1]\alpha_1 + E[Z_2]d + \varepsilon$$

where: y is the vector of phenotypic values (body weight and morphometric measurements), X is the fixed ef-

Table 1 – Method, gene and sizes of the polymerase chain reaction PCR products or restriction fragments.

Method/Molecular markers	Fragment size (bp)		
	<i>Piaractus mesopotamicus</i>	<i>Colossoma macropomum</i>	<i>Piaractus brachipomus</i>
Multiplex-PCR/SNP 1 (nuclear genes)	400	280	600
Multiplex-PCR/SNP2 (nuclear genes)	500	150	100
Multiplex-PCR/COI (mitochondrial gene)	307	435	307 and 610
RFLP-PCR/CYTB (mitochondrial gene)	152 and 513	261 and 405	665

SNP = single nucleotide polymorphism; COI = cytochrome c oxidase subunit I gene; RFLP-PCR = polymerase chain reaction-restriction fragment length polymorphism technique; CYTB = cytochrome B gene.

fects, which include the number of fish in each tank; Z_1 and Z_2 are the GCA and SCA matrices; β is the vector of fixed effects; and α_1 and d are the random effect related to GCA and SCA respectively. The random effects (α_1 and d) were assumed to be $\alpha_1 \sim MVN(0, I\sigma_a^2)$, and $d_1 \sim MVN(0, I\sigma_d^2)$ and the random residuals ε were assumed to be normally distributed, uncorrelated and have homogeneous variance. If all pedigrees are known, then Z_1 and Z_2 carry complete information, but since some genitors were missing, we had to impute Z_1 and Z_2 based on the genotype expectation.

The full likelihood function can be described as follows:

$$L(\theta|Y) = \prod_{j=1}^N p(y_j|\theta)$$

where: $\theta = (Xb + E[Z_1]\alpha_1 + E[Z_2]d, \sigma_e^2)$ and is the vector of model parameters.

The mixed normal mixture models considered g as denoting genetic group ($g = 1, 2, \dots, 6$), π_g as the maximization of proportions (initial presumption, a given animal belongs to a certain group), μ_g as the mean vector and σ_g^2 as the error variance. The full likelihood for the data can be represented by the following equation:

$$L(\theta|y) = \underbrace{\prod_{j=1}^k p(y_j|\theta)}_{\text{Full model}} \underbrace{\prod_{i=k+1}^n \left(\sum_{g=1}^l \pi_g(y_j|\mu_g, \sigma^2) \right)}_{\text{missing information}}$$

where: $\pi_g (\pi_1, \dots, \pi_6) = \frac{1}{6}$, that is, the initial probability of an individual with unknown genetic composition belonging to any of the crosses is 1/6, and $\mu_g (\mu_1, \dots, \mu_6)$ are the genotypic values related to the g -th crossing.

The observed values were grouped into products of sequences, in which the counter varied from j to k . The missing data on genetic composition were grouped into the counter ranging from i to n , where $i = k + 1$. The expectation of the j -th observation belonging to group g is given by the conditional probability of y_j given the group and its parameters, normalized by the sum of the probabilities of y_j . Therefore, it is a multinomial variable:

$$\hat{Z}_{2jg} = E[Z_{gj}] = \frac{\pi_g(y_j|\mu_g, \sigma^2)}{\sum_{i=1}^k \pi_i(y_j|\mu_i, \sigma^2)}$$

The mean vector of each cluster (g), i.e., the mean of the data for all observations, weighted by the probabilities of the group, was estimated by:

$$\mu_g = E[Z_1]\alpha_1 + E[Z_2]d$$

This part related to the mixed model refers to the expectation of the genotypic value under the g -th cross. For example, if the j -th individual presents a probability of $\hat{Z}_{2j1} = 0.9$ to be allocated in the first group (pure genotypic background), then, $E[Z_{1j1}] = 2 \cdot 0.9$, where 2 is the number of alleles obtained from a pure genotype.

Therefore, the empirical BLUP for GCA and SCA are given respectively by:

$$\hat{\alpha}_1 = \left(E[Z_1]'E[Z_1] + \frac{\sigma_e^2}{\sigma_a^2} \right)^{-1} E[Z_1]'(y - X\beta - \hat{Z}_2d)$$

$$d = \left(\hat{Z}_2'\hat{Z}_2 + \frac{\sigma_e^2}{\sigma_d^2} \right)^{-1} \hat{Z}_2'(y - X\beta - E[Z_1]\hat{\alpha}_1)$$

The program was formulated using PROC IML from SAS v. 8.0. The variance components were estimated by Restricted Maximum Likelihood (REML) using a convergence criterion equal to 1×10^{-5} (Patterson and Thompson, 1971).

As described above, we used the GCA incidence matrix (additive part of the model), which was obtained using the imputation of body weight of observations, and grouped the animals with segregation similarities. The likelihood ratio test was carried out to check for differences between expected and observed frequencies, according to the type of cross observed, as pure animals, single-cross hybrids or three-way cross hybrids, among others. The analysis of concordance between the genetic composition of the animals imputed by the mixed normal mixture model for each trait, and between them and the genetic composition reported by the farmers was formulated by Pearson correlation using the Mantel test, with the Vegan package 2.0-10 and the R software version 3.1.0. Also, cross-validation was performed using the leave-one-out process to impute each fish's genetic composition based on body weight and other morphometric traits analyzed. Then, we later sought an association between the true genotype and the genotype imputed by Pearson correlation using the Mantel test, with the Vegan package 2.0-10 and the R software version 3.1.0. For cross-validation, only genotypes presenting known pedigree were used and the full likelihood was given by:

$$L(\theta|y) = \underbrace{\prod_{j=1}^{k-1} p(y_j|\theta)}_{\text{Full model}} \underbrace{\left(\sum_{g=1}^l \pi_g(y_k|\mu_g, \sigma^2) \right)}_{\text{missing information for k-th genotype}}$$

Results

During the adaptation period, 13 fish died, which reduced the number from 96 to 83. After this period, there was no mortality and the water quality parameters were within the suitable range for production, according to the limits recommended by Resolution 357/2005 from Brazil's National Environmental Council - CONAMA (Ministério do Meio Ambiente, 2005).

The molecular analysis (Figures 1A, B, C and D) indicated a high error rate (53 %) of the farmers in identifying the genetic composition of the fish. Pacu, Pirapitinga and the hybrid Paqui were correctly classified by the farmers (Table 2). Tambaqui and the hybrids Patinga, Piraqui, Tambatinga and Tambacu were erroneously classified. According to molecular analysis, among the 12 animals misclassified as Patinga, seven were identified as

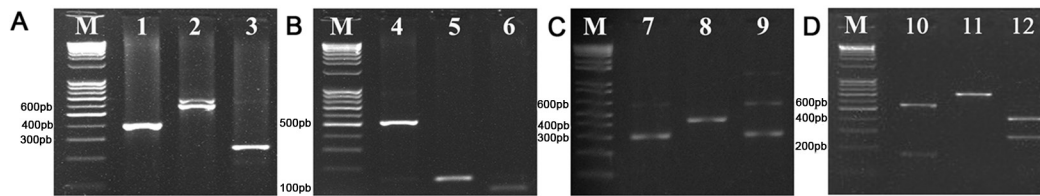


Figure 1 – Electrophoresis analysis of *Piaractus mesopotamicus* (Pacu), *Piaractus brachypomus* (Pirapitinga) and *Colossoma macropomum* (Tambaqui) using multiplex-PCR species-specific primers for the nuclear gene SNP 1 (single nucleotide polymorphism) (A), nuclear gene SNP 2 (B), COI mitochondrial gene (cytochrome c oxidase subunit I) (C) and CYTB (cytochrome B) mitochondrial gene (D). Lanes: 1-Pacu; 2-Pirapitinga; 3-Tambaqui; 4-Pacu; 5-Tambaqui; 6-Pirapitinga; 7-Pacu; 8-Tambaqui; 9-Pirapitinga; 10-Pacu; 11-Pirapitinga; 12-Tambaqui; M = Molecular weight marker: 1 kb.

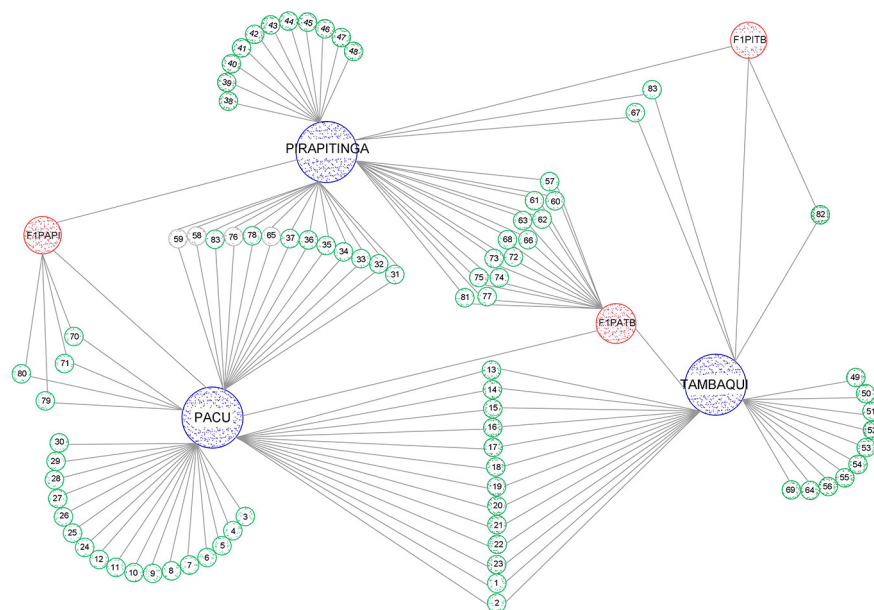


Figure 2 – Relationship of animal with parental species; F1PAPI = simple hybrid of Pacu with Pirapitinga; F1PATB = simple hybrid of Pacu with Tambaqui; F1PITB = simple hybrid of Pirapitinga with Tambaqui; Animals in green have a probability of above 90 % of belonging to mating/cross in question. Those in white have a probability of below 90 %.

Table 2 – Classification and number of Serrasalmidæ fish according to the identification of fish farmers and molecular analysis, and level of correct classification by farmers (%).

Classification		Correct classification by farmers
Fish farmers (number of fish)	Molecular analysis (number of fish)	
Pacu (10)	Pacu (10)	100.00
Patinga (12)	Pacu (7) and PF1H (5)	0.00
Paqui (11)	Paqui (11)	100.00
Pirapitinga (11)	Pirapitinga (11)	100.00
Piraqui (12)	Paqui (1) and PF1H (11)	0.00
Tambaqui (8)	Tambaqui (7) and PF1H(1)	87.50
Tambatinga (9)	Paqui (1) and PF1H (8)	0.00
Tambacu (10)	Piracu (7) and PF1H (3)	0.00

PF1H = post-F1 hybrids.

Pacu and the other five were classified as post-F1 hybrids (PF1H). It was not possible to identify the genetic composition of these post-F1 hybrids via molecular markers. The same occurred in 12 hybrids of Piraqui, classified as one Paqui and 11 PF1H. From eight Tambaqui, seven were classified as Tambaqui and one as PF1H. The Tambacu were identified as seven Piracu (♀pirapitinga × ♂pacu) and three PF1H. Tambatinga were also identified as one Paqui and eight advanced hybrids.

In the leave-one-out process to impute the genetic composition of animals with known genetic composition, based on body weight and morphometric measurements, the correlation ranged from 0.54 to 0.56 between the true genotypes and the imputed genotypes (Table 3).

With regard to the information provided by fish farmers about the genetic composition and that imputed

by the mixture model according to the evaluated traits, the genetic composition reported by the farmers was significantly correlated with the genetic composition imputed to head length only. This result evidenced a low level of consistency between the genetic composition imputed for the traits and that reported by farmers (Table 4). There was also a significant and positive correlation between genetic compositions obtained by imputation, with 40 % moderate correlations ($0.3 \leq r \leq 0.7$) and 60 % high correlations ($r > 0.7$), indicating concordance in the imputation of the genetic composition of fish for the different variables used.

Figure 2 shows the relationship of the animals used in this experiment with their parental species. Fish were identified by number from 1 to 83, and up to number 55, molecular analysis was considered to infer the relationship of animals with their parents, while above number 55, the information imputed by the mixed normal mixture models was used.

Most fish that had their genetic composition imputed showed a high probability of belonging to known genotypes (Figure 2 and Table 5). Three animals with imputed genetic composition were highly likely to be Tambaqui (56, 64 and 69), one animal (78) with imputed genetic composition was highly likely to be a simple hybrid between Pacu and Pirapitinga, and one fish was considered highly likely to have the genetic composition of $\frac{1}{4}$ Pirapitinga and $\frac{3}{4}$ Tambaqui (82) (Table 5). Animals identified as numbers 58, 59, 65 and 76, which were white in color, were assigned with a low probability of belonging to the cross between Pacu and Pirapitinga, since they had probabilities below 90 % (Figure 2). Thus, it cannot be affirmed that these animals belonged to this cross.

The likelihood ratio test of the values obtained from the imputation showed the following genotypes: two simple hybrids of Pirapitinga and Tambaqui (67 and

Table 5 – Likelihood ratio test (LRT) for imputed genotypes.

Number	Genotype			LRT
	Pacu	Pirapitinga	Tambaqui	p-value
3	-	-	100	-
1	50	50	-	-
2	-	50	50	0.55
13	25	50	25	0.28
4	75	25	-	0.60
1	-	25	75	-

83), thirteen three-way cross hybrids ($\frac{1}{4}$ Pacu + $\frac{1}{4}$ Tambaqui and $\frac{1}{2}$ Pirapitinga) (57, 60, 61, 62, 63, 66, 68, 72, 73, 74, 75, 77, 81) and four hybrids of Pacu with Pirapitinga, back-crossed with Pacu ($\frac{3}{4}$ Pacu + $\frac{1}{4}$ Pirapitinga) (numbers 70, 71, 79 and 80) (Figure 2 and Table 5).

Discussion

The high misclassification rate by the fish farmers exhibited in the present work can be ascribed to the difficulties in making this classification based on morphological visualization only, which is the most widely used technique for assignment of species of fish. Depending on the type of gene action, maternal and paternal effects, cultivation environment, and the interaction of all these factors, the hybrids may phenotypically resemble their parents, resulting in erroneous identification. Misclassification of Serrasalminidae purchased from commercial fish farms in Brazil was also reported by Gomes et al. (2012) and Hashimoto et al. (2014). Hashimoto et al. (2014) used multiplex-PCR and PCR-RFLP, and reported that juvenile Pacu, Tambaqui and Tambatinga were wrongly marketed as Tambacu in the states of Minas Gerais, São Paulo and Sergipe. Tambaqui is also being marketed as Tambacu in the states of Pará and Piauí, and also as Tambatinga in Pará (Gomes et al., 2012).

Likewise, Gomes et al. (2012) stated that fish farms in the states of Pará and Piauí are selling Tambaqui erroneously. In fact, they are Tambacu and Tambatinga. The same authors also found that Tambaqui had been sold as Tambacu in the state of Pará. Thus, hybrid animals are traded in Brazil as pure species or other types of interspecific hybrids. When using the molecular markers of Pintado (*Pseudoplatystoma coruscans*) and Cachara (*Pseudoplatystoma reticulatum*), it has also been observed that morphological evaluation is not efficient: in fish farms, post-F1 hybrid fish are used as breeders (Hashimoto et al., 2013). Misconception in the classification of exotic fish species, like tilapia and carp, has also promoted the use of hybrid animals as breeders (Mair, 2007; Mia et al., 2005). Thus, the identification of the genetic composition of fish by species based on morphological visualization only, routinely done by fish farmers, has not proven to be effective.

Recent studies using molecular markers suggest that a minimum of four nuclear markers are necessary to correctly identify hybrid animals when introgressive

Table 3 – Pearson correlation using Mantel test between true genotype and genotype imputed for each trait.

Traits	Genotype imputed				
	Weight	SL	HL	BH	BW
True genotype	0.56*	0.55*	0.54*	0.55*	0.54

SL = standard length; HL = head length; BH = body height; BW = body width; *significant at 5 % probability.

Table 4 – Pearson correlation using Mantel test between the genetic composition matrix reported by farmers and matrices imputed for each trait.

Traits	SL	HL	BH	BW	FAR
Weight	0.88*	0.69*	0.86*	0.76*	0.02 ^{ns}
SL	-	0.57*	0.75*	0.81*	0.05 ^{ns}
HL	-	-	0.73*	0.50*	0.34*
BH	-	-	-	0.64*	0.08 ^{ns}
BW	-	-	-	-	0.11 ^{ns}

SL = standard length; HL = head length; BH = body height; BW = body width; FAR = farmers; *significant at 5 % probability; ns = not significant.

hybridization occurs (Boecklen and Howard, 1997; Sanz et al., 2008; Hashimoto et al., 2014). Tests using different numbers of markers have demonstrated that the variation in these numbers is directly related to the differentiation between loci among parental species (Sanz et al., 2006; Bohling et al., 2013); that is, the greater the difference, the smaller the number of markers required in the analysis. This was confirmed in this study, where the markers were 100 % fixed between species and were useful for detecting even introgressive hybridization.

It was possible to discern five genetic groups (three pure species and two hybrids) as well as post-F1 hybrids (Table 2). The mixed normal mixture models proved to be a complementary technique applicable to molecular analysis, aiming to infer the genetic composition of animals when the molecular approach was not sufficiently informative, as observed by the correlations obtained (0.54 to 0.56) between the true and the imputed genotypes.

Moreover, the fish classified as post-F1 hybrids by means of molecular analysis showed a high probability of belonging to known genotypes. This was because where molecular analysis classified fish as post-F1, it was possible to allocate these individuals with a high probability to known groups. Furthermore, the imputations were highly concordant for different types of traits (0.72 mean correlation). This high concordance between results for various traits evidences the effectiveness of the adopted imputation strategy, showing high repeatability in the allocation of animals into groups. Therefore, the application of mixed normal mixture models provides further reliability in the description of animals whose genetic composition is unknown, being more reliable than the use of morphological visualization only.

The lack of correlation between the genetic composition matrix reported by farmers and that imputed for most traits was expected, due to the high error rate in the classification of the genetic composition of the animals by farmers, i.e., the farmers' difficulty in correctly identifying hybrid individuals. The presence of significant moderate to high correlation between genetic compositions imputed between the evaluated traits can be a function of these characteristics as correlation is found in fish (Reis Neto et al., 2012; Melo et al., 2013).

As for the classification of animals that had their genetic composition imputed by the mixture models, a number of these animals were highly likely to be three-way cross hybrids or back-crossed hybrids. In some studies, it has also been found that Serrasalminae fish hybrids are fertile, such as Tambacu (♀ tambaqui × ♂ pacu), Tambatinga (♀ tambaqui × ♂ pirapitinga), and Patinga (♀ pacu × ♂ pirapitinga) (Hashimoto et al., 2014). The presence of fertile hybrids is worrisome from the biological point of view. It can pose a threat to pure species if they reach the natural stocks by escape or operating error, since they have higher genetic variability, which enhances their adaptation to different environments (Porto-Foresti et al., 2013; Hashimoto et al., 2014).

The imputation process indicated that three fish were likely to be Tambaqui. By means of molecular analysis, these animals were classified as advanced hybrids. Thus, these fish can be derived from an absorbent cross at the advanced stage of Tambaqui with one or more species. The imputation of a hybrid of Pacu with Pirapitinga and two simple hybrids of Pirapitinga with Tambaqui may be due to the mating of F₁ hybrids, causing them to show the same genetic composition as simple hybrids.

The results of the differentiation of parental species and their hybrids demonstrated that the combined use of different techniques can be an excellent alternative, and can provide data for characterizing other hybrids currently produced in Brazilian fish farming. Thus, the combination of these two strategies has great potential for identifying genetic composition when crosses are unknown. Mixed models of mixtures proved to be a useful technique in imputing the genetic composition of fish species of the family Serrasalminae based on the phenotypes assessed in this work. Thus, it can be concluded that the technique described here combined with molecular determination can be used to make inferences about genetic makeup in fish.

Acknowledgments

This study was supported by the Brazilian National Council for Scientific and Technological Development (CNPq).

References

- Abd-Elmaghd, W.; Davis, L. 2006. Density estimation using mixtures of mixtures of gaussians. *Lecture Notes in Computer Science* 3954: 410-422.
- Boecklen, W.J.; Howard, D.J. 1997. Genetic analysis of hybrid zones: numbers of markers and power of resolution. *Ecology* 78: 2611-2616.
- Bohling, J.H.; Adams, J.R.; Waits, L.P. 2013. Evaluating the ability of Bayesian clustering methods to detect hybridization and introgression using an empirical red wolf data set. *Molecular Ecology* 22: 74-86.
- Eirola, E.; Lendasse, A.; Vandewalle, V.; Biernacki, C. 2014. Mixture of Gaussians for distance estimation with missing data. *Neurocomputing* 131: 32-42.
- Faustino, F.; Nakaghi, L.S.; Marques, C.; Ganeco, L.N.; Makino, L.C. 2010. Structural and ultrastructural characterization of the embryonic development of *Pseudoplatystoma* spp. hybrids. *The International Journal of Developmental Biology* 54: 723-730.
- Gomes, F.; Schneider, H.; Barros, C.; Sampaio, D.; Hashimoto, D.; Porto-Foresti, F.; Sampaio, I. 2012. Innovative molecular approach to the identification of the tambaqui (*Colossomamacropomum*) and its hybrids: the tambacu and the tambatinga. *Anais da Academia Brasileira de Ciências* 84: 89-97.
- Griffing, B. 1956. Concept of general and specific combining ability in relation to diallel crossing systems. *Australian Journal of Biological Science* 9: 463-493.

- Hashimoto, D.T.; Mendonça, F.F.; Senhorini, J.A.; Oliveira, C.; Foresti, F.; Porto-Foresti, F. 2011. Molecular diagnostic methods for identifying Serrasalmid fish (Pacu, Pirapitinga, and Tambaqui) and their hybrids in the Brazilian aquaculture industry. *Aquaculture* 321: 49-53.
- Hashimoto, D.T.; Prado, F.D.; Senhorini, J.A.; Foresti, F.; Porto-Foresti, F. 2013. Detection of post-F1 fish hybrids in broodstock using molecular markers: approaches for genetic management in aquaculture. *Aquaculture Research* 44: 876-884.
- Hashimoto, D.T.; Senhorini, J.A.; Foresti, F.; Martínez, P.; Porto-Foresti, F. 2014. Genetic identification of F1 and Post-F1 Serrasalmid juvenile hybrids in Brazilian aquaculture. *PlosOne* 9: e89902.
- Hisano, H.; Fonseca, G.G.; Russo, M.R.; Della Flora, M.A.L.; Ishikawa, M.M.; Pádua, S.B. 2013. Hybrid sorubim viscera protein concentrate in the diets for barred sorubim. *Boletim do Instituto de Pesca* 39: 37-44.
- Jang, K.S.; Han, S.; Lee, I.; Woo, Y.W. 2006. Lip localization based on active shape model and gaussian mixture model. *Lecture Notes in Computer Science* 4319: 1049-1058.
- Luca, G.; Zuccolotto, P. 2003. Finite and infinite mixtures for financial durations. *International Journal of Statistics* 51: 431-455.
- Mair, G.C. 2007. Genetics and breeding in seed supply for inland aquaculture. p. 519-548. In: Bondad-Reantaso, M.G., ed. *Assessment of freshwater fish seed resources for sustainable aquaculture*. FAO, Rome, Italy.
- Melo, C.C.V.; Reis Neto, R.V.; Costa, A.C.; Freitas, R.T.F.; Freato, T.A.; Souza, U.N. 2013. Direct and indirect effects of measures and reasons morphometric on the body yield of Nile tilapia, *Oreochromis niloticus*. *Acta Scientiarum. Animal Sciences* 35: 357-363.
- Mia, M.Y.; Taggart, J.B.; Gilmour, A.E.; Gheyas, A.A.; Das, T.K.; Kohinoor, A.H.M.; Rahman, M.A.; Sattar, M.A.; Hussain, M.G.; Mazid, M.A.; Penman, D.J.; McAndrew, B.J. 2005. Detection of hybridization between Chinese carp species (*Hypophthalmichthys molitrix* and *Aristichthys nobilis*) in hatchery broodstock in Bangladesh, using DNA microsatellite loci. *Aquaculture* 247: 267-273.
- Ministério da Pesca e Aquicultura [MPA]. 2012. Statistical Bulletin of Fisheries and Aquaculture from Brazil = Boletim Estatístico da Pesca e Aquicultura do Brasil. MPA, Brasília, DF, Brazil (in Portuguese).
- Ministério do Meio Ambiente [MMA]. 2005. CONAMA Resolution n° 357, Mar 17th, 2005 = Resolução CONAMA n° 357, de 17 de março de 2005. CONAMA, Brasília, DF, Brazil. p.58-63 (in Portuguese).
- Patterson, H.D.; Thompson, R. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* 58: 545-554.
- Porto-Foresti, F.; Hashimoto, D.T.; Prado, F.D.; Senhorini, J.A.; Foresti, F. 2013. Genetic markers for the identification of hybrids among catfish species of the family Pimelodidae. *Journal of Applied Ichthyology* 29: 643-647.
- Reis Neto, R.V.; Freitas, R.T.F.; Serafini, M.A.; Costa, A.C.; Freato, T.A.; Rosa, P.V.; Allaman, I.B. 2012. Interrelationships between morphometric variables and rounded fish body yields evaluated by path analysis. *Revista Brasileira de Zootecnia* 41: 1576-1582.
- Sanz, N.; Cortey, M.; Pla, C.; García-Marín, J.L. 2006. Hatchery introgression blurs ancient hybridization between brown trout (*Salmotrutta*) lineages as indicated by complementary allozymes and mtDNA markers. *Biological Conservation*. 130: 278- 289.
- Sanz, N.; García-Marín, J.L.; Viñas, J.; Roldán, M.; Pla, C. 2008. Spawning groups of European anchovy: population structure and management implications: population structure of spawning groups of European anchovy. *ICES Journal of Marine Sciences* 65: 1635-1644.