SPECIAL ISSUE ORIGINAL ARTICLE

Genomic selection signatures in farmed *Colossoma macropomum* from tropical and subtropical regions in South America

John Fredy Gómez Agudelo[1] | Vito Antonio Mastrochirico-Filho[1] |
Carolina Heloisa de Souza Borges[1] | Raquel Belini Ariede[1] |
Lieschen Valeria Guerra Lira[1] | Rubens Ricardo de Oliveira Neto[1] |
Milena Vieira de Freitas[1] | Gustavo Adolfo Lenis Sucerquia[2] | Manuel Vera[3] |
Milthon Honorio Muñoz Berrocal[4] | Diogo Teruo Hashimoto[1]

[1]Aquaculture Center of UNESP, São Paulo State University (UNESP), Jaboticabal, Brazil

[2]Facultad de Ciencias Agrarias, Universidad de Antioquia, Medellín, Colômbia

[3]Facultad de Veterinaria, Universidad de Santiago de Compostela (USC), Lugo, Spain

[4]Facultad de Zootecnia, Universidad Nacional Agraria de la Selva, Tingo Maria, Perú

Correspondence
Milthon Honorio Muñoz Berrocal, Universidad Nacional Agraria de la Selva, Carretera Central km 1.21, Tingo María, Peru.
Email: milthon_munoz@yahoo.es

Funding information
Conselho Nacional de Desenvolvimento Científico e Tecnológico, Grant/Award Number: 140740/2016–3, 311559/2018–2 and 422670/2018-9; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior; Fundação de Amparo à Pesquisa do Estado de São Paulo, Grant/Award Number: 2016/18294–9, 2016/21011–9, 2017/19717-3, 2017/26900–9, 2018/08416–5, 2019/08972-8 and 2019/10662-7; Comisión Nacional de Investigación Científica y Tecnológica

Abstract

Tambaqui or cachama (*Colossoma macropomum*) is one of the most important neo-tropical freshwater fish used for aquaculture in South America, and its production is concentrated at low latitudes (close to the Equator, 0°), where the water temperature is warm. Therefore, understanding how selection shapes genetic variations and structure in farmed populations is of paramount importance in evolutionary biology. High-throughput sequencing to generate genome-wide data for fish species allows for elucidating the genomic basis of adaptation to local or farmed conditions and uncovering genes that control the phenotypes of interest. The present study aimed to detect genomic selection signatures and analyze the genetic variability in farmed populations of tambaqui in South America using single-nucleotide polymorphism (SNP) markers obtained with double-digest restriction site-associated DNA sequencing. Initially, 199 samples of tambaqui farmed populations from different locations (located in Brazil, Colombia, and Peru), a wild population (Amazon River, Brazil), and the base population of a breeding program (Aquaculture Center, CAUNESP, Jaboticabal, SP, Brazil) were genotyped. Observed and expected heterozygosity was 0.231–0.350 and 0.288–0.360, respectively. Significant genetic differentiation was observed using global F_{ST} analyses of SNP loci ($F_{ST} = 0.064$, $p < 0.050$). Farmed populations from Colombia and Peru that differentiated from the Brazilian populations formed distinct groups. Several regions, particularly those harboring the genes of significance to aquaculture, were identified to be under positive selection, suggesting local adaptation to stress under different farming conditions and management practices. Studies aimed at improving the knowledge of genomics of tambaqui farmed populations are

John Fredy Gómez Agudelo and Vito Antonio Mastrochirico-Filho contributed equally to the work.

essential for aquaculture to gain deeper insights into the evolutionary history of these fish and provide resources for the establishment of breeding programs.

1 | INTRODUCTION

Tambaqui or cachama (*Colossoma macropomum*) is a fish species that primarily inhabits the floodplains of the Amazon and Orinoco basins in South America (Araujo-Lima & Goulding, 1997). Tambaqui is a tropical fish, which typically prefers warm waters (Woynárovich & Van Anrooy, 2019), such as those in their natural habitats, where the temperature is above 26°C (e.g., Central Amazon; Affonso et al., 2015). Farming of this species began in the 1960s, and currently, it is considered the main native South American fish for aquaculture farmed under diverse environmental conditions. The leading producers include Brazil, Peru, and Colombia (Woynárovich & Van Anrooy, 2019). Production primarily occurs on small or medium fish farms using semi-intensive and intensive systems. In 2019, tambaqui accounted for nearly 19% (101,100 tons) of total Brazilian fish production (the major producer). Owing to its high rusticity, rapid growth, high productivity, and substantial commercial value in the international markets (IBGE, 2020; Valladão et al., 2018), tambaqui is one of the most important neotropical freshwater fish used for aquaculture in South America. Despite its smaller proportion, tambaqui production in Colombia and Peru has increased over the last few decades (Valladão et al., 2018). Nowadays, tambaqui is also being farmed in Asian countries, particularly China, Indonesia, Malaysia, Myanmar, and Vietnam (Valladão et al., 2018; Woynárovich & Van Anrooy, 2019).

Domestication is considered a prolonged process that allows for the adaptation of new species to captive environments. Selective breeding focusing on economic and adaptive traits may be applied to improve the productivity of farmed fish (Teletchea, 2018; Teletchea & Fontaine, 2014). However, even today, almost all tambaqui production from aquaculture in South America relies upon stocks that are not completely domesticated, and wild populations serve as a source to maintain farmed breeding populations (Fazzi-Gomes et al., 2017; Santos et al., 2007). Farming conditions, such as high stocking density, climatic fluctuations, and stressful management, negatively affect tambaqui production by suppressing immunity, enhancing disease susceptibility, and increasing mass mortality (Boijink et al., 2016). Apparently, tambaqui individuals exhibit a poorer growth performance in Brazilian subtropical regions (South Brazil), where temperatures are below 23°C. Moreover, fish die if the water temperature remains below 15°C for several days (Zaniboni-Filho & Meurer, 1997). These temperature limitations impede the natural range expansion and establishment of the species in waters where temperature drops below 15–18°C during

cold seasons (Woynárovich & Van Anrooy, 2019). Similarly, abrupt seasonal temperature variations in Brazilian subtropical regions, together with stresses triggered by intensive production, can result in the outbreak of diseases, such as those caused by the ectoparasite *Ichthyophthirius multifiliis* (Lira et al., 2020).

Genomic breeding programs for tambaqui were initiated in the first decade of the 21st century, with emphasis on improving the growth performance and disease resistance of these fish (Ariede et al., 2020; De Mello et al., 2015; Lira et al., 2020; Perazza et al., 2019). Although the assessment of genetic diversity is a fundamental step in the introduction of genomic breeding programs (Mastrochirico-Filho et al., 2019), there have been few large-scale population genetic studies of farmed populations (Mastrochirico-Filho et al., 2021; Nunes et al., 2017, 2020). Furthermore, no study has focus on the efficient breeding management of tambaqui, as opposed to a closely related serrasalmid species, pacu (*Piaractus mesopotamicus*) (Freitas et al., 2021). In addition, majority of the genetic studies aimed at understanding the degree of genetic variability and population structure were restricted to Brazilian stocks. Specifically, studies based on mitochondrial DNA and microsatellites have revealed moderate genetic differentiation among tambaqui broodstocks and loss of genetic variability in captivity, due perhaps to the small size of the founding populations and lack of appropriate breeding management (Aguiar et al., 2018; Ferreira et al., 2019; Gonçalves et al., 2018). As such, the formation of breeding units with unknown genetic variability in farmed populations may lead to the loss of genetic potential and an increase in inbreeding risk (Charlesworth & Willis, 2009; Melo et al., 2006). For instance, in rainbow trout (*Oncorhynchus mykiss*), different degrees of inbreeding depression in stocks decreased hatching, fecundity, and larval survival rates and generated individuals with morphological deformities (Yousefian & Nejati, 2008).

Genomic techniques, such as double-digest restriction site-associated DNA sequencing (ddRAD-Seq), can be applied for the discovery and genotyping of thousands of single nucleotide polymorphisms (SNPs) simultaneously. Moreover, ddRAD-Seq could be employed in genomic studies of several aquaculture species, even when limited genomic information was available (Houston et al., 2020; Robledo et al., 2018). The discovery and characterization of SNP markers have assumed relevance in genetic studies aimed at investigating the diversity, differentiation, and structure of wild and cultivated stocks for the expansion of aquaculture to several non-model South American species (Mastrochirico-Filho et al., 2019; Torati et al., 2019).

Furthermore, genomic approaches have been applied to detect selection signatures, which may be functionally linked to genetic variations in traits subject to selection (Lopez et al., 2015). In recent years, selection signatures have been intensively studied because of their potential association with genes that control the phenotypes of interest in farmed populations of fish (Gutierrez et al., 2016; López et al., 2019). However, contrary to the production of salmon, which has been intensively selected for economically important traits since the 1970s (Gjedrem, 2010), the domestication of tambaqui is relatively recent and involves farmed populations that have not undergone strong artificial selection as yet. Therefore, considering that the power of detection of SNP outliers depends primarily on the number of generations since domestication and the strength of selection (Gutierrez et al., 2016), novel insights into selection signatures focused on tambaqui production will expand our knowledge of gradual changes at the genomic level in populations adapted to farmed conditions, which may be associated with the traits of commercial interest. Additionally, the detection of novel selection signature may allow for the identification of genes formerly associated with farmed environments, which can serve as resources for increasing the productivity of this native South American fish species (Gutierrez et al., 2016; López et al., 2019; Vera et al., 2019).

To this end, the present study aimed to investigate the genetic diversity and population structure of farmed populations of tambaqui in commercial hatcheries from the leading producers of this species in South America (i.e., Brazil, Colombia, and Peru). Furthermore, the genomic selection signatures were investigated to identify the regions associated with the variations in economic traits and/or adaptation to captive environments in farmed tambaqui in tropical and subtropical regions of South America.

2 | MATERIALS AND METHODS

2.1 | Sampling

Samples were obtained from 361 tambaqui breeders across 12 commercial fish farms in Brazil (BR), Colombia (COL), and Peru (PER) as well as a wild population from the Amazon River (3°29′S, 60°66′W) (WILD). A base population used in breeding program research conducted by the Laboratory of Genetics in Aquaculture and Conservation (LaGeAC) at the Universidade Estadual Paulista (UNESP), Jaboticabal (São Paulo State, Brazil) (BRGEN), was also sampled. This population includes individuals from different fish farms and is studied to obtain genomic information, aimed at advancing aquaculture breeding programs for native fish species, with a particular focus on improving growth performance and disease resistance (Table 1).

The sampled farmed populations cannot be considered fully domesticated, as they are still based on genetically unselected fish. Farmed populations from Brazil (BR1 to BR6) correspond to samples from subtropical regions, which are characterized by wide variations in temperature (maximum temperatures during summer typically exceed 30°C and minimum temperatures during winter are below 15°C). Farmed populations from Colombia (COL1 to COL5) and Peru (PER) correspond to samples from the equatorial regions (tropical) of South America, which are characterized by constant and

TABLE 1 Parameters of genetic diversity in farmed populations of tambaqui *Colossoma macropomum* considering 1633 SNPs

Sampling site	N	N_a	H_o (SD)	H_e (SD)	F_{IS}	HWE	MAF > 0.1 (%)
BR1	14	1.968	0.350 (0.198)	0.351 (0.140)	0.060	1623	1401 (85.8)
BR2	19	1.975	0.314 (0.164)	0.349 (0.138)	0.131	1614	1409 (86.3)
BR3	20	1.966	0.231 (0.139)	0.345 (0.139)	0.357	1527	1408 (86.2)
BR4	11	1.983	0.311 (0.164)	0.358 (0.129)	0.193	1632	1452 (88.9)
BR5	15	1.950	0.256 (0.159)	0.342 (0.145)	0.287	1600	1394 (85.4)
BR6	18	1.971	0.290 (0.154)	0.345 (0.138)	0.190	1616	1.395 (85.4)
BR All	97	2.000	0.290 (0.109)	0.368 (0.112)	0.225	1195	1521 (93.1)
BRGEN	15	1.985	0.306 (0.160)	0.353 (0.131)	0.170	1616	1452 (88.9)
COL1	19	1.944	0.283 (0.179)	0.336 (0.152)	0.190	1592	1338 (81.9)
COL2	14	1.897	0.260 (0.183)	0.321 (0.164)	0.234	1609	1278 (78.3)
COL3	16	1.985	0.293 (0.157)	0.360 (0.130)	0.222	1603	1444 (88.4)
COL4	12	1.878	0.250 (0.183)	0.312 (0.168)	0.247	1617	1239 (75.9)
COL5	6	1.775	0.263 (0.236)	0.288 (0.185)	0.188	1633	1154 (70.6)
COL All	67	1994	0.274 (0.132)	0.351 (0.134)	0.230	1294	1410 (86.3)
PER	14	1.954	0.249 (0.157)	0.343 (0.144)	0.325	1601	1380 (84.5)
WILD	6	1.908	0.263 (0.195)	0.337 (0.152)	0.330	1633	1371 (84.0)

Note: Standard deviation values (SD) are between parentheses.

Abbreviations: F_{IS}, inbreeding coefficient; H_e, expected heterozygosity; H_o, observed heterozygosity; HWE, loci in Hardy–Weinberg equilibrium after FDR-BY correction (*p*-adjusted > 0.0063); MAF, minimum allelic frequency; *N*, sampling size; N_a, average number of alleles per loci.

high mean temperatures (~30°C) with very little annual variations. Unfortunately, environmental variables were not recorded during the collection period of farmed fish.

The fish farms in Brazil and Colombia are over 20 years old and have adopted intensive breeding systems in large facilities, with more intensified management to supply the increasing demand. The fish farms in Peru are relatively recent (10-year-old) and considered less developed than those in other main South American producers. Furthermore, Peruvian aquaculture is dominated by marine species, and tambaqui production remains small scale and at the subsistence level. The commercial identity and localization of the fish farms are confidential. The number of samples collected per population is presented in Table 1. The fish were individually tagged with pit tags (passive integrated transponder tags, FDX-B full-duplex model, 134.2 kHz) and maintained in a farm for subsequent analyses.

All procedures were performed with strict adherence to the recommendations of the National Council for Control of Animal Experimentation (CONCEA) (Brazilian Ministry for Science, Technology and Innovation), and the study approved by the Animal Use Ethics Committee (CEUA), number 019006/17. Fin samples were collected from each fish under anesthesia administered using benzocaine solution (200 mg·L^{-1}) (Sigma, St. Louis, USA), and all efforts were made to minimize suffering. Fin samples were stored in 95% ethanol at −20°C.

2.2 | ddRAD-Seq library construction

Prior to genomic analysis, a purity analysis was performed using species-specific multiplex polymerase chain reaction (PCR) (Hashimoto et al., 2011) to ensure that the samples originated from pure tambaqui specimens and not from interspecific hybrids. This preanalysis was applied because in Colombian and Brazilian hatcheries, hybrid individuals are commonly produced from one or two serrasalmid species, such as pirapitinga (*Piaractus brachypomus*) and pacu (*Piaractus mesopotamicus*).

The ddRAD-Seq libraries for SNP identification were constructed based on the protocol originally described by Peterson et al. (2012). Briefly, genomic DNA was extracted from 361 individuals using the PureLink Genomic DNA Kit. Each library was constructed using 45 individuals identified with barcodes (sequences of six base pairs). A total of eight libraries were constructed. The amount of genomic DNA was determined using Qubit® 3.0 (Thermo Scientific™) and standardized to the concentration of 10 ng·μL^{-1}. The purity of the nucleic acid samples was determined using spectrophotometry with NanoDrop™ One (Thermo Scientific™), and the absorbance ratio (A260/A280) was between 1.80 and 2.00. DNA integrity of the samples was visualized using 1% agarose gel electrophoresis. Genomic analyses of data from COL5 and WILD were restricted to a few individuals and affected by the difficulty in obtaining DNA from biological samples. For DNA digestion, the final reaction mixture contained 2.5 μl of DNA (totaling 25 ng) digested with 0.1 μl (10 U·ml^{-1}) of the restriction enzyme MluCI (5′-...AATT...-3′) and

0.1 μl (20 U·ml^{-1}) of the restriction enzyme SphI (5′-...GCATGC...-3′); 0.6 μl of the CutSmart 10× buffer; and 2.8 μl of water. The conditions for enzyme digestion were 37°C for 60 min. Based on the genome of tambaqui (GCA_904425465.1), computer simulations were performed to evaluate the cut frequency for each restriction enzyme using Nebcutter v2.0 (Vincze, 2003). The average distance between enzymatic sites of enzyme was 6500 bp for SphI and 250 bp for MluCI, which allowed for adjusting the amount of specific adapters for each enzyme, calculated as described by Peterson et al. (2012). Ligation was performed in a final volume of 10.0 μl, containing 0.2 μl of the T4 ligase enzyme, 1.0 μl of the 10× T4 DNA ligase reaction buffer, 1.0 μl of the P1 adapter (4.04 μM), and 1.0 μl of the P2 adapter (0.07 μM). For binding, the samples were incubated at 23°C for 90 min, followed by 65°C for 10 min. Short fragments resulting from DNA digestion and adapter ligation were removed using the Agencourt AMPure XP cleanup step Kit (Beckman Coulter, DE), following the manufacturer's protocol. This step allowed us to pool all samples for each library, and the libraries were then used to select specific fragments. Fragments of approximately 450 bp were selected using the E-Gel® Size Select 2% agarose gel (Invitrogen).

The enrichment step was then performed via PCR amplification using the Phusion® High-Fidelity PCR Kit. This step was necessary to incorporate the Illumina flow cell sequence and regions of the Illumina primer sequences. To increase the number of DNA fragments, 20 independent PCR assays were performed. All reactions were performed in a final volume of 20.0 μl, containing 9.5 μl of H$_2$O, 0.4 μl of 10-mM dNTPs, 4.0 μl of 5× Phusion HF Buffer, 2.5 μl of P1 5.0 μM Primer, 2.5 μl of P2 5.0 μM Primer, 0.12 μl of Phusion DNA Polymerase, and 1.0 μl of digested DNA. All reactions involved an initial stage of 68°C for 1 min, followed by 12 cycles at 95°C for 10 s, 60°C for 30 s, and 72°C for 30 s, and a final stage of 72°C for 7 min. The PCR samples were grouped in a final volume of 400 μl and were repurified using AMPure XP beads (Beckman Coulter, DE), according to the manufacturer's protocol. The pool of amplified sequences was analyzed by agarose gel electrophoresis. The concentration of each final library was verified by fluorometry using Qubit® 3.0 (Thermo Scientific™). The samples were sequenced on the Hiseq 4000 platform, generating 1.7 G (billions) of paired-end reads.

2.3 | SNP identification and filtering

Stacks v. 2.1 (Catchen et al., 2011, 2013) was used to analyze the raw sequences resulting from Illumina sequencing. Following the Stacks manual (https://catchenlab.life.illinois.edu/stacks/manual/#align), the analysis was initiated with the "process_radtags" module, in which the barcodes were assigned to individuals, and low-quality sequences and adapters were excluded. After demultiplexing and cleaning, the data were aligned against the tambaqui reference genome (GCF_904425465.1) using Burrows–Wheeler Aligner (BWA, v0.7.15) with the standard parameters (Li & Durbin, 2009). The mapped regions were used to identify genomic variants using Samtools (Li et al., 2009) and Vcftools (Petr et al., 2011), which filtered them

based on a general quality score of >20 and minimum allele count of ≥4. Samples with over 40% missing genotypes were removed (mind 0.4) using Plink v. 1.07 (Purcell et al., 2007). Quality control filters were also applied using Plink v. 1.07 (Purcell et al., 2007), excluding SNPs that did not pass in the following criteria: missing genotypes (geno) < 0.3 and minor allele frequency (MAF) < 0.05. Furthermore, the possible genotyping errors were discarded using Vcftools (Petr et al., 2011), removing SNPs deviating from the Hardy–Weinberg equilibrium considering HWE p-value $< 1 \times 10^{-6}$. Finally, the filtered SNPs were present in at least 70% individuals in a population (call rate > 0.70). The resulting variants were filtered to consider only one SNP per locus.

2.4 | Genetic diversity analyses

The genetic diversity parameters were estimated considering the geographic regions (Brazil, Colombia, and Peru) and sampled fish farms. All remaining SNPs were pruned for linkage disequilibrium (LD) using the indep-pairwise option (--indep-pairwise 50 10 0.1) in Plink v. 1.07 (Purcell et al., 2007) to avoid the effects of linked SNPs. The observed (H_o) and expected (H_e) heterozygosity, average number of alleles per locus (N_a), and MAF were estimated using Plink v. 1.07 (Purcell et al., 2007). To assess heterozygote deficiency in a fish farm, deviations from the Hardy–Weinberg equilibrium (HWE p-value < 0.05) were predicted with the exact test in Plink v. 1.07 (Purcell et al., 2007), using the FDR-BY correction method described to Narum (2006). The inbreeding coefficient (F_{IS}) for each fish farm was estimated using the R package hierfstat (Goudet & Jombart, 2015). The contemporary effective population size (N_e) for each fish farm was calculated using the linkage disequilibrium method in NeEstimator V2.01 (p-value < 0.05) (Do et al., 2014). For this analysis, rare alleles were avoided while retaining SNPs with the highest polymorphic content (PIC) and MAF > 0.1 (Marandel et al., 2020). Identity by state (IBS) distances by multidimensional scaling (MDS) analysis were performed using Plink v. 1.07 (Purcell et al., 2007) to estimate the genomic relatedness and similarity between individuals from different geographic regions.

The genetic structure of fish farms from different countries was determined based on the global and pairwise estimates of FST using hierfstat (Goudet & Jombart, 2015) and Arlequin 3.5 (Excoffier & Lischer, 2010). The genetic distances between stocks were measured as described by Nei (1972). Mixing levels between stocks were inferred by estimating the most probable number of population units (K) using Structure V2.3.4 (Falush et al., 2007), without prior information of the population. First, the distribution of ΔK was determined, and an ad hoc statistic based on the rate of change in the logarithmic probability of data between successive K values was used. The cluster interval (K) was predefined from 1 to 14. The analysis was performed with 70 replicates (i.e., five repetitions for each K value) and 50,000 iterations after 100,000 burn-in steps. The most probable value

of K to explain the population structure was the modal value of ΔK, as described by Evanno et al. (2005). The results of structure analysis were viewed using Structure Harvester (Earl & Von Holdt, 2012). Discriminant analysis of principal components (DAPCs) (Jombart et al., 2010) was also used to examine the population structure with the R package adegenet (Jombart, 2008), describing the genetic clusters considering the geographic location of the fish farms as predefined groups. The find.clusters function was applied to determine the number of clusters, and the xvalDAPC and optim.a.score functions were used to calculate the optimal number of principal components to be retained for the analysis. Analysis of molecular variance (AMOVA) was performed in Arlequin 3.5 (Excoffier & Lischer, 2010) (10,000 permutations to test significance) considering the clusters obtained in structure analysis and DAPC.

2.5 | Identification of SNP outliers

Three methods were used for the identification of SNP outliers in the populations, considering groups compounding all commercial fish farms from Brazil (group BR), Colombia (group COL), and Peru (group PER). BRGEN and WILD were excluded from the analysis. Subsequently, data from the adopted fish farms were grouped into different clusters according to the results of structure analysis and DAPC. The first and second methods were performed using Arlequin v 3.5 (Excoffier & Lischer, 2010). The first method was based on the hierarchical island model (H) (Excoffier et al., 2009), which can reduce the number of false-positive outliers. The significance of outliers was assessed using 10,000 simulations, 100 demes, and 5 groups. The second method was the simulation of the finite island model (nH) using the same specifications as the hierarchical island method (H). For both methods, loci presenting p-values ≤0.01 were considered as the putative outliers. The third method was performed using Bayescan v. 2.1 (Foll & Gaggiotti, 2008), with 20 pilot runs, 100,000 iterations, and 50,000 burn-in steps. The program calculates the posterior odds (PO) based on the probability of a particular locus being under selection using a value of prior odds equivalent to 100:1 and the proportion of loci with a strong increase in FST relative to the other loci. Loci presenting q-values ≤0.05 were considered the putative outliers. All resulting SNPs were considered in this analysis. To evaluate the effect of SNP outliers on the differentiation of fish farms, a second cluster analysis was performed using IBS and DAPC considering only the putative SNP outliers.

The ddRAD-tags containing the putative outliers were annotated by BLASTN search against the tambaqui genome in GenBank (GCA_904425465.1). The significance threshold for homologies (e-value) was $<e^{-4}$. The search for genes was performed when the loci were anchored in the same region within a range of <250 kb on both sides. The genetic variant annotation for each candidate SNP to be under selection was performed using SnpEff (Cingolani et al., 2012).

3 | RESULTS

3.1 | SNP identification

Approximately 300 Gb (billions of nucleotide bases) reads were sequenced through the libraries, resulting in an average of approximately 220 M reads per library. After the application of the quality filters on sequences, approximately 180 M reads per library were retained for SNP characterization. On average, approximately 4 M reads per individual were considered, and a total of 207 M reads were excluded due to ambiguous barcodes. For genomic analysis, 199 samples were considered (after individual missingness filtering, mind = 0.4). A total of 6323 SNPs were obtained after aligning the sequences against the tambaqui reference genome (GCA_904425465.1), excluding SNPs with a quality score of <20, minimum allele count of <4 reads, and HWE p-value of $<1 \times 10^{-6}$. Subsequently, 1276 SNPs presenting MAF < 0.05 and proportion of missing genotypes >30% were discarded, leaving 5047 SNPs for downstream analyses. All remaining SNPs attained a minimum call rate of 70% individuals in a population (call rate > 0.70).

3.2 | Genetic diversity

The descriptive statistics of genetic diversity for each population are presented in Table 1 and are based on 1633 SNPs (3414 SNPs were eliminated after LD pruning).

Overall, there were no large differences in diversity parameters among the fish farms and between the farmed and WILD populations. Most farmed populations from Colombia (COL1 to COL5) and Peru (PER) presented lower N_a, H_o, and H_e values than the farmed populations from Brazil (BR). The diversity estimates of farmed populations from Colombia and Peru were similar to those of the wild population (WILD). Markers with rare alleles and lower MAF values were more abundant in farmed populations from Colombia (e.g., COL5 and COL4), except COL3, whereas farmed populations from Brazil presented higher MAF frequencies. All inbreeding coefficients (F_{IS}) were positive, ranging between 0.060 (BR1) and 0.357 (BR3), and the SNP loci were in Hardy–Weinberg disequilibrium in most populations (except COL5 and WILD) (Table 1). The F_{IS} value over all loci was 0.223, and the effective population size (N_e) was low and variable among the farmed populations, ranging from 9.8 in BRGEN to 75.6 in COL5. N_e could not be estimated for WILD because the confidence interval extremely large (infinite), making the values unreliable. N_e estimates for COL5 should be interpreted cautiously because of the small number of individuals analyzed. The N_e estimates for inbreeding rate (ΔF) per generation ranged from 0.9% in BR6 to 5.1% in BRGEN (Table 2).

3.3 | Population genetic differentiation and structure

A low but significant genetic differentiation among all populations was detected based on global FST values using SNP loci ($F_{ST} = 0.064$, $p < 0.05$). In FST pairwise analysis (Table 3; Figure 1; Table S1a), low-to-moderate genetic differentiation was observed between all population pairs. All values were significantly greater than zero, except for some comparisons involving geographically close populations (shown in bold letters in Table 3); these values ranged from 0.002 for the genetic differentiation between BRGEN and WILD, to 0.109, the highest genetic differentiation detected between BR6 and COL5 (Table 3). The Nei genetic distances showed a similar pattern of little differentiation among farmed populations in Brazil. Meanwhile, the highest genetic distance was noted between the farmed populations in Brazil and Colombia (COL1 to COL5) (Table 3).

TABLE 2 Effective population size (Ne) parameter based on linkage disequilibrium, considering 1633 SNPs

Sampling site	N	Ne	CI 95%	Jackknife CI	ΔF
BR1	14	12.1	11.8–12.2	6.6–26.5	0.041
BR2	19	17.8	17.5–18.1	10.5–37.2	0.028
BR3	20	19.9	19.6–20.4	10.8–55.3	0.025
BR4	11	54.7	53.5–62.1	11.1–Inf.	0.009
BR5	15	15.1	15.1–15.8	8.2–40.5	0.033
BR6	18	26.5	25.8–27.1	13.7–95.2	0.018
BRGEN	15	9.8	9.7–10.0	3.6–28.3	0.051
COL1	19	14.0	13.6–14.0	9.5–21.6	0.036
COL2	14	34.8	32.6–35.6	15.6–640.2	0.014
COL3	16	46.9	45.2–49.0	21.0–Inf.	0.011
COL4	12	17.2	16.7–17.8	7.1–188.1	0.029
COL5	6	75.6	42.6–62.6	7.3–Inf.	0.006
PER	14	25.8	24.7–26.4	10.8–2164.4	0.019
WILD	6	Inf.	Inf.–Inf.	8.3–Inf.	Infinite

Abbreviations: CI, confidence interval 95%; inbreeding coefficient (ΔF = 1 / 2Ne); Inf.: infinite; N, number of breeders for each fish farm.

TABLE 3 Values of population differentiation based on F_{ST} (below the diagonal) and Ney genetic distance (above the diagonal)

	WILD	PER	COL5	COL4	COL3	COL2	COL1	BRGEN	BR6	BR5	BR4	BR3	BR2	BR1
BR1	0.025	0.059	0.099	0.095	0.047	0.089	0.084	0.042	0.037	0.041	**0.006**	0.037	0.044	0
BR2	0.028	0.058	0.092	0.086	0.046	0.082	0.080	0.046	**0.006**	0.040	0.014	0.035	0	0.048
BR3	0.015	0.042	0.078	0.076	0.033	0.074	0.067	0.035	0.036	**0.005**	**0.008**	0	0.043	0.039
BR4	**0.007**	0.030	0.058	0.059	0.018	0.053	0.051	**0.014**	**0.012**	0.016	0	**0.011**	0.021	**0.005**
BR5	0.024	0.044	0.086	0.080	0.036	0.077	0.070	0.041	0.039	0	0.020	**0.004**	0.048	0.043
BR6	0.026	0.056	0.092	0.085	0.044	0.083	0.080	0.042	0	0.044	**0.013**	0.041	**0.007**	0.037
BRGEN	**0.007**	0.032	0.054	0.048	0.023	0.044	0.051	0	0.050	0.050	0.020	0.044	0.058	0.049
COL1	0.040	**0.008**	0.051	0.049	0.009	0.040	0	0.057	0.091	0.083	0.065	0.077	0.092	0.097
COL2	0.046	0.033	**0.006**	**0.007**	0.019	0	0.042	0.048	0.094	0.088	0.067	0.079	0.091	0.095
COL3	**0.008**	**0.001**	0.023	0.023	0	0.022	0.016	0.032	0.056	0.048	0.028	0.040	0.058	0.055
COL4	0.042	0.039	**0.0001**	0	0.028	0	0.057	0.055	0.098	0.091	0.071	0.084	0.099	0.102
COL5	0.048	0.036	0	**0.003**	0.029	**0.004**	0.050	0.057	0.109	0.092	0.072	0.086	0.106	0.107
PER	**0.009**	0	0.039	0.046	**0.004**	**0.006**	**0.010**	0.039	0.066	0.058	0.039	0.049	0.065	0.072
WILD	0	0.022	0.057	0.047	**0.020**	0.048	0.050	**0.002**	0.033	0.036	**0.004**	**0.029**	0.039	0.029

Note: All values were significant (*p*-value < 0.05) except those shown in bold letters.

The genetic distance between these populations was evaluated based on clusters by calculating the average measures of IBS of the SNP markers and then summarizing them using MDS to determine population structures in different countries (Figure 2). The results were posteriorly confirmed by DAPC, retaining 25 PCs through a-score optimization (Figure 3). Samples from Brazil and Colombia formed separate clusters, confirming the genetic differentiation between the farmed populations. Furthermore, genetic similarity was observed between farmed populations from Colombia and Peru and between WILD and BRGEN (Figures 2 and 3).

To assess the level of mixing between samples, cluster analyses using the Bayesian model were performed based on the distribution of ΔK (Figure 4). The results indicated that K = 2 was the most adequate value for explaining the population structure of the farmed populations of tambaqui in South America. Overall, all genetic structure analyses identified the presence of two main clusters: (1) BR1, BR2, BR3, BR4, BR5, and BR6 (orange group) and (2) COL1, COL2, COL4, and COL5 (blue group). COL3, PER, WILD, and BRGEN populations were considered a mixture of both genetic clusters, although the cluster found in their respective geographical regions was the most abundant (i.e., cluster 1 for WILD and BRGEN and cluster 2 for COL3 and PER) (Figure 4).

Based on the above population clusters, two AMOVAs were performed according to the structure hypothesis (Model I: group 1: BR1, BR2, BR3, BR4, BR5, BR6, and BRGEN and group 2: WILD, COL1, COL2, COL3, COL4, COL5, and PER) and DAPC results (Model II: group 1: BR1, BR2, BR3, BR4, BR5, and BR6; group 2: BRGEN and WILD; and group 3: COL1, COL2, COL3, COL4, COL5, and PER). There were no significant differences between the two models. The highest proportion of genetic variance was present within the populations (F_{ST}) (~93%), whereas the genetic variance among groups (F_{CT} = ~3%) and among populations within groups (F_{SC} = ~3%) was significantly low (Table 4).

3.4 | Regions with SNP outliers and their putative genes

The outliers were detected using two groups based on the results of structure and DAPC analyses; the first group was composed of farmed populations from Brazil (BR1, BR2, BR3, BR4, BR5, and BR6), and the second group was composed of farmed populations from Colombia (COL1, COL2, COL3, COL4, and COL5) and Peru (PER). The analysis was performed using two different software and three different methods. A graphical representation of the outliers identified by the Arlequin (nH, H) and Bayescan methods is provided in Figure S1. A total of 82 SNPs were identified as outliers, considering all three approximations. Of these, two outliers (2.4%) were detected by all methods, 12 outliers (14.6%) by the Arlequin methods (nH, H), and only 1 (1.2%) identified by the Bayescan method alone. The corresponding outliers were characterized and associated with functional genes, as shown in Table 5. Most outliers were identified exclusively by the hierarchical island model (nH) (67; 81.8%)

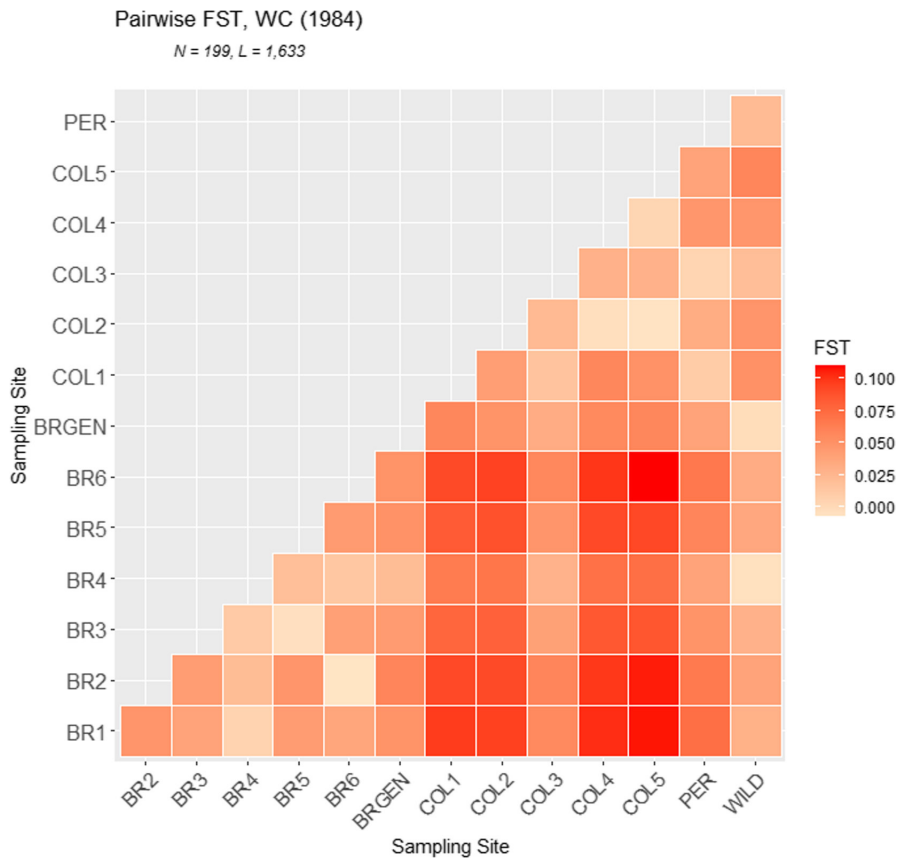**FIGURE 1** Heatmap showing differentiation between tambaqui stocks from South America based on F_{ST}
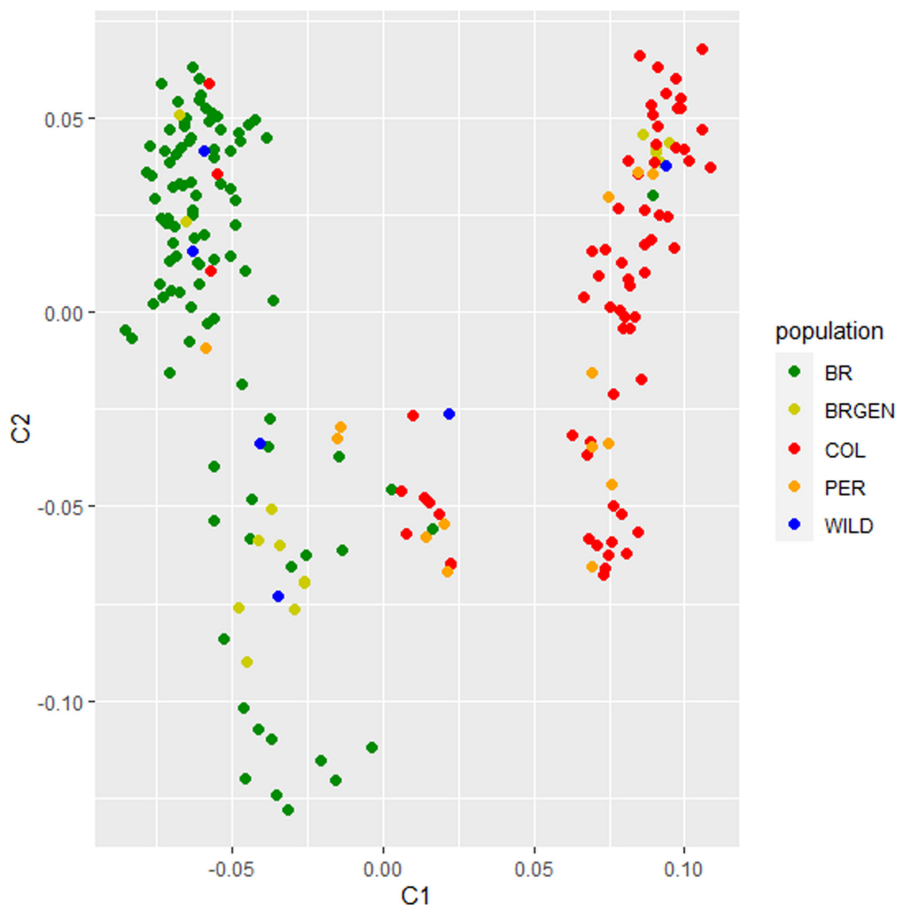

FIGURE 2 Population genetic analysis of tambaqui stocks from South America. Multidimensional scaling (MDS) analysis resulted from 199 individuals. Individuals were plotted according to their coordinates on the first two components (C1 and C2)

FIGURE 3 Plot of discriminant analysis of principal components (DAPC) showing relationship of 199 individual fish color coded by sample site, representing the structure between tambaqui stocks from South America. Twenty five PCs were retained using a-score optimization
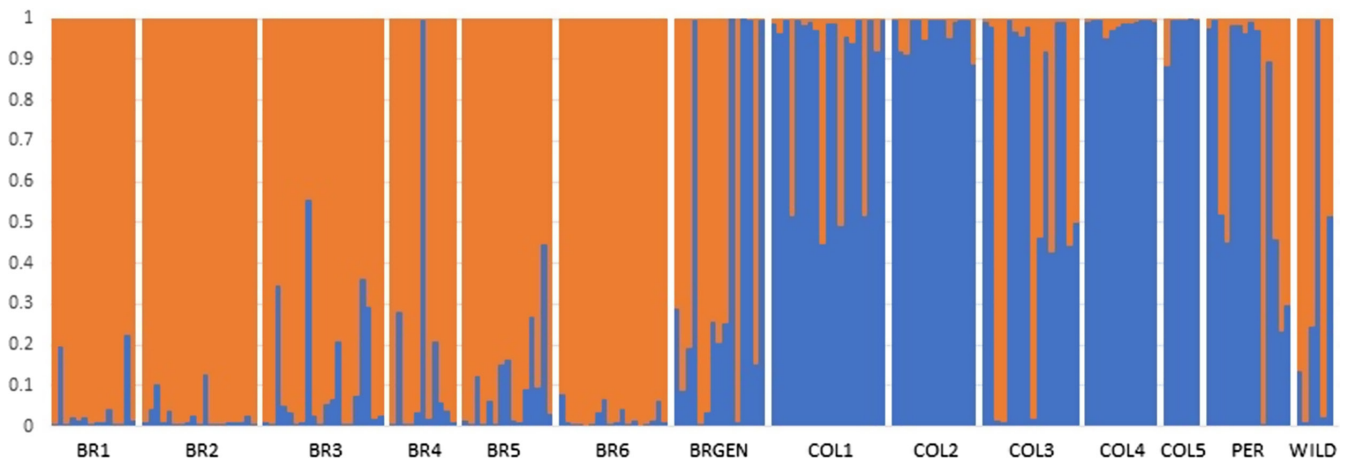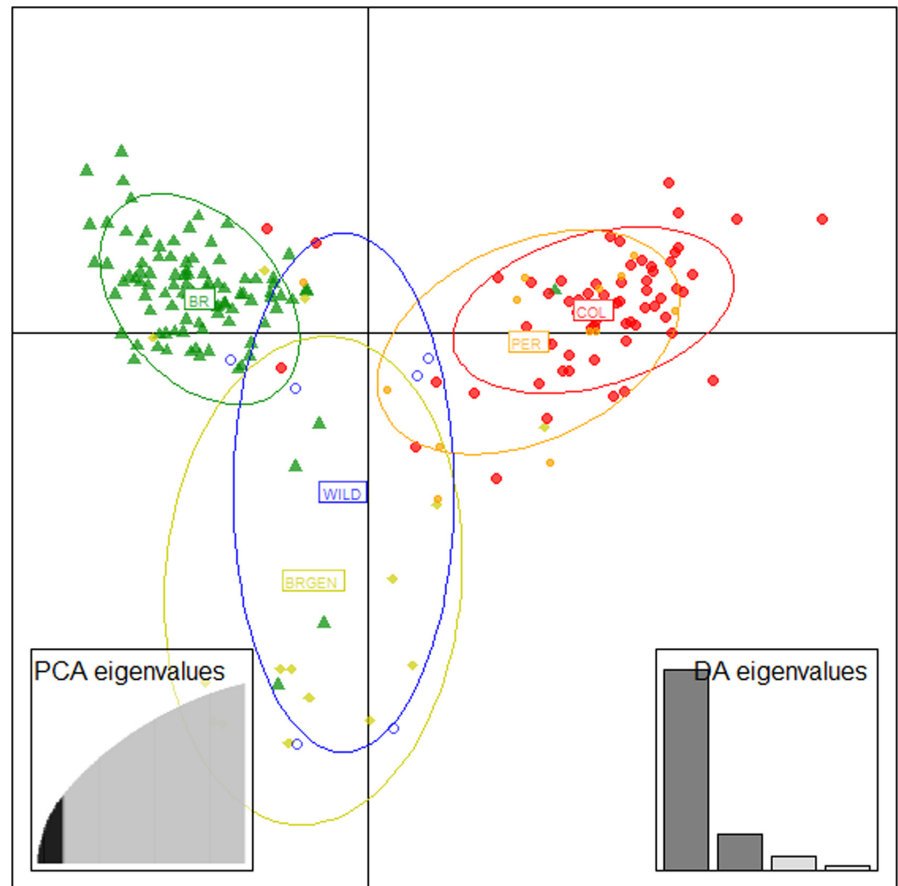

FIGURE 4 Analysis of genetic structure of tambaqui stocks from South America. Genetic structure was analyzed approaching K = 2 according to Delta K statistics. Each vertical bar represents an individual. Populations are separated by vertical white bars. The color proportions of each bar correspond to the estimated fractions of association of the individuals in each of the clusters

(Figure S1) and were characterized and associated with functional genes (see Table S1b). The Bayescan analysis revealed moderate evidence of selection considering the three detected outliers [log10 (PO) = 1.000–1.833; q-value < 0.05]. The Arlequin analysis revealed outliers with apparent evidence of selection, presenting high observed FST values for both the nH (obs F_{ST} = 0.227–0.411) and H (obs F_{ST} = 0.335–0.430) (p-value < 0.01) models. The visualization of IBS and DAPC (5 PCs retained) revealed two distinct clusters: one

formed by the COL and PER samples and the other formed by the BR samples (Figure S2a and b).

Regarding annotation of the putative SNP outliers, 36.6% SNPs were found in transcript regions, 35.1% were classified as intron variants, 9.3% were classified as upstream gene variants, and 9.0% were located in intergenic regions, and the closest gene was identified. Regarding the effects of SNP outliers, we imputed the overall impact of all variants as largely modifying (97.6%) (see Table S1c).

TABLE 4 Analysis of molecular variance (AMOVA) of tambaqui stocks *Colossoma macropomum*

	F-statistic	Variance component	% Variation
Model I—Structure (K = 2)			
Among groups (F_{CT})	0.033*	10.33	3.31
Among populations within groups (F_{SC})	0.035*	10.60	3.40
Within populations (F_{ST})	0.067*	290.96	93.29
Model II—DAPC analysis			
Among groups (F_{CT})	0.035*	10.93	3.51
Among populations within groups (F_{SC})	0.030*	9.14	2.94
Within populations (F_{ST})	0.064*	290.96	93.55

Note: For Model I, the populations were grouped following population units (K) described by structure analyses (K = 2), while for model II, the populations were grouped by DAPC results. *p-value > 0.01.

4 | DISCUSSION

4.1 | Levels of genetic diversity and population structure

The present study used an SNP dataset (1633 SNPs) to unveil the genetic variability and population structure of farmed tambaqui populations in South America. Considering the massive amount of genetic data, the genetic diversity (H_e and MAF values) in the farmed populations was moderate or high, contrary to the pattern of low genetic variability in farmed tambaqui populations demonstrated in a previous analysis based on mitochondrial DNA and microsatellite data (Aguiar et al., 2018; Ferreira et al., 2019; Gonçalves et al., 2018). Diversity estimates were higher for the farmed populations from Brazil than for the farmed populations from Colombia and Peru. The F_{IS} values were positive and high for all populations, and the loci were in Hardy–Weinberg disequilibrium in most populations, revealing homozygous excess. This pattern can be explained by common breeding practices in aquaculture, in which random mating is typically not performed

TABLE 5 Summary of outliers identified by the three utilized methods

	BayeScan			Arlequin (H)		Arlequin (nH)		Blastn	
SNP ID	Log10(PO)	*q*-value	F_{ST}	Obs F_{ST}	*p*-value	Obs F_{ST}	*p*-value	Gene	Annotations
cmj_57835:205	1.833	0.014	0.255	0.361	0.008	0.291	0.005	ATF6	I,ST
cmj_49626:177	1.000	0.052	0.211	0.380	0.011	0.301	0.002	ENO1A	ST
cmj_139241:10	1.269	0.033	0.240	–	–	–	–	IRF4	I
cmj_40107:214	–	–	–	0.413	0.001	0.335	0.001	OPN3	ST
cmj_115873:85	–	–	–	0.430	0.003	0.354	0.002	SBK1	ST
cmj_13492:21	–	–	–	0.427	0.005	0.301	0.002	ULK1	ST
cmj_158265:204				0.395	0.006	0.347	0.0003	PLXNA2	ST
cmj_3210:95				0.371	0.006	0.296	0.004	WRNA	S
cmj_185105:197	–	–	–	0.397	0.006	0.304	0.004	INO80C	ST
cmj_47461:171				0.383	0.007	0.312	0.0007	SPRED1	I
cmj_34560:96	–	–	–	0.361	0.007	0.283	0.006	RBP2B	S
cmj_44265:124	–	–	–	0.389	0.008	0.311	0.004	ERCC3	I, ST
cmj_97663:203	–	–	–	0.391	0.008	0.311	0.004	FGF4	I
cmj_85247:50				0.352	0.009	0.280	0.008	NMS	ST
cmj_919:262	–	–	–	0.335	0.010	0.281	0.007	ZFAND5A	I

Abbreviations: H, hierarchical island model. Blastn annotations: I, genes related to defense response against pathogens; nH, finite model; S, genes related to sensory system; ST, genes related to stress responses.

Mining of the ddRAD-tags containing the outliers found genes with important functions surrounding these markers (Table 5; Table S1B). Among the 14 outliers detected by different methods and identified exclusively by the Bayescan method (Table 5), nine outliers (60.0%) were surrounded by genes related to stress response (ST) and six outliers (40.0%) were related to defense response against pathogens under stressful conditions (I) (Table 5).

or by broodstock exchange between farmed populations without proper management, which may increase the proportion of homozygous genotypes for different alleles that might have been originally fixed in the farm of origin, due to inbreeding or genetic drift. Moreover, the lack of control of homozygous excess was reflected in the low N_e estimates for most farmed populations, being lower than a minimum of 50 animals per generation to

—WILEY┘

control inbreeding, as recommended by the Food and Agriculture Organization of the United Nations (FAO, 1998).

Although the genetic diversity in farmed populations from Brazil may be favored due to the exchange of stocks among breeders, broodstocks in all analyzed populations were not appropriately renewed by individuals originating, for instance, from future progenies or wild environments. Furthermore, the reduced number of breeders allied to the lack of directed mating of nonparental individuals can decrease genetic diversity over the generations of farmed populations from South America (Brown et al., 2005; Mastrochirico-Filho et al., 2019).

The loss of genetic variability in farmed tambaqui has been described in Brazilian hatcheries, suggesting the lack of genetic monitoring and adequate breeding management in farmed populations of the species (Aguiar et al., 2018; Ferreira et al., 2019; Gonçalves et al., 2018). Therefore, our results can be attributed to the combination of a limited number of founders in farmed populations without proper genetic management, leading to random genetic drift and inbreeding caused by domestication processes, which may have compromised the genetic diversity of the farmed populations. To minimize the impacts of the low number of individuals used in the hatcheries, additional studies on the genomics and management of South American stocks are fundamental for the proper functioning of initial breeding programs.

Analysis of genetic differentiation between the farmed populations based on pairwise F_{ST} values and Nei distances revealed higher genetic differentiation of Brazilian stocks from Colombian and Peruvian stocks. Meanwhile, genetic similarity was observed between farmed populations from Colombia and Peru as well as between WILD and BRGEN (Figures 2 and 3). However, the observed genetic differentiation values were not expressive, highlighted by the overall F_{ST} estimate of 0.064 (p-value < 0.05). Moreover, AMOVA revealed the highest significant differentiation within populations (93.29%, p < 0.01) and not between groups (3.31%, p-value < 0.01). Considering the panmictic nature of the wild populations from the Amazon Basin (Santos et al., 2007), the low-to-moderate genetic differentiation between farmed populations from the northern and central regions of South America was significant, exhibiting unique genetic structures. Primarily, the genetic differences observed between the studied farmed populations, and confirmed by structure and DAPC analyses, may be related to the geographical isolation of tambaqui production from the central region of Brazil, where the access to farmed populations from northern region is limited, restricting the gene flow due to the geographical barrier and the exchange of broodstocks among the hatcheries. Simultaneously, stock foundation from captured wild individuals did not likely favor the high genetic differentiation between fish farms, as evidenced by the low of F_{ST} values between the WILD and farmed populations. Consistently, the panmictic nature of tambaqui populations in the Amazon River has been documented (Santos et al., 2007). Interestingly, farmed populations from Colombia and Peru, despite being located the nearest to the WILD populations, showed a higher genetic differentiation from the WILD populations, emphasizing the

lack of renewal of broodstocks by new individuals originating from the wild and poor management practices.

4.2 | SNP outliers and their putative genes

The application of large-scale SNP datasets has become increasingly popular for the detection of selection signatures due to increased marker density and, consequently, increased genomic coverage (Gutierrez et al., 2016). In the present study, 82 SNPs (1.62%) were identified as outliers using a large SNP dataset (~5 K). Previous studies analyzing selection signatures in aquaculture species have detected few regions potentially under selection. For instance, Gutierrez et al. (2016) identified moderate evidence of 44 putative selection signatures in an Atlantic salmon population using a 6.5K SNP array. Furthermore, Mäkinen et al. (2014) analyzed three populations of Atlantic salmon and found little evidence of signatures of selection using the Bayescan software.

In the present study, although significant evidence of selection was detected in 82 genomic regions, only three outliers were detected by the Bayescan method, which is considered a more conservative approach than the Arlequin methods, and only two outliers were detected by all selected methods. Although previous studies have reported a high rate of false-positive outliers using the Arlequin and Bayescan methods (Narum & Hess, 2011), both approaches have been used to detect selection signatures in aquaculture species (Gutierrez et al., 2016; Liu et al., 2015; Vera et al., 2019). Given the differences between the methods used to identify the evidence of selection and caution with which these loci under selection should be evaluated due to the high false-positive rate, we considered the most consistent outliers supported by both Arlequin methods, in addition to those identified exclusively by the Bayescan method. The detected outliers may be related to domestication processes in many parts of South America. Tambaqui production has become more intensive over the years and involves diverse farm management practices, including handling, confinement, fertilization, and other operations. Nevertheless, intensive management procedures may produce some level of disturbance, leading to stress responses and, consequently, poor fish performance. Additionally, tambaqui is considered a susceptible species to extreme environmental conditions, and although its production is possible in a wide range of South American territories, fish farmers are discouraged from cultivating this species during the winter, specifically in the southern and southeastern regions (subtropical area of Brazil), when there are long periods of low temperature (Fernandes et al., 2018). Temperature fluctuations affect the physiological processes of all fish, particularly under stressful conditions (Alfonso et al., 2020). These changes in the physiological processes may increase susceptibility to diseases, such as those caused by parasites and bacteria, as already described in cases of disease outbreaks during tambaqui production (Ariede et al., 2020; Lira et al., 2020).

Of the three outliers detected by the Bayescan method (q-value < 0.05) and the 12 outliers detected by both Arlequin methods

(*p*-value < 0.01), nine were related to stress response and six were related to immunity (Table 5). However, we were particularly interested in the six SNPs that were identified as outliers in this analysis (cmj_57835:205, cmj_49626:177, cmj_139241:10, cmj_115873:85, cmj_13492:21, cmj_158265:204, and cmj_85247:50). These outliers were found to be associated with genes that integrate the genetic components of stress responses to body temperature, confinement, and management practices, such as handling and feeding deprivation. The loci harboring the outliers cmj_57835:205 and cmj_158265:204 have been implicated in thermal stress response, and they were, respectively, associated with cyclic AMP-dependent transcription factor ATF-6 alpha (*ATF6*) (Wang et al., 2016) and plexin A2 (*PLXNA2*), which were involved in the muscle development and growth of threespine stickleback (*Gasterosteus aculeatus*) under different temperatures (Metzger & Schulte, 2018); Outliers cmj_49626:177 and cmj_115873:85 showed homology with proteins involved in stress response to handling and confinement. Meanwhile, cmj_49626:177 was associated to enolase 1a (*ENO1a*), which is involved in stress response to repeated handling in Senegalese sole; suboptimal rearing conditions in Gilthead seabream (Raposo de Magalhães et al., 2021); and hypoxia/thermal stresses in zebrafish larvae (Long et al., 2015), pikeperch (Swirplies et al., 2019), and tambaqui (Prado-Lima & Val, 2016). Moreover, cmj_115873:85 was associated with a serine/threonine protein kinase (*SBK1*), which has been mapped to differentially expressed transcripts related to handling and confinement stress response in rainbow trout (Gonzalez-Pena et al., 2016). Stress response to food deprivation has been widely discussed in fish farming procedures aimed at evaluating the metabolic regulation and reallocation of energy under these critical periods (Lutfi et al., 2018). Two outliers (cmj_13492:21 and cmj_85247:50) were surrounded by genes encoding the serine/threonine protein kinase *ULK1*, which is involved in autophagy in response to starvation and nutrition deficiency in farmed species (Fan et al., 2019; Wu et al., 2020), and neuromedin S (*NMS*), which is involved in the neuroendocrine regulation of feeding and, possibly, in the reallocation of energy under stress response (Li et al., 2015).

In tambaqui production, domestication is a very recent process, and it has not yet been fully implemented, hampering the presence of a large number of fixed outliers over the generations. The lack of genetic monitoring and inadequate breeding management may compel fish farmers to repeatedly acquire new broodstock from the wild environment, interrupting the fixation of selection signatures and loci involved in adaption to local conditions. Nevertheless, the present study is the first to aim at identifying the loci under selection in tambaqui. Although these findings must be carefully evaluated, they will serve as a basis for expanding our understanding of the genetics of adaptability of fish to farmed systems.

To the best of our knowledge, the present study is the first to investigate the genetic composition of tambaqui stocks from the major producers in South America using genome-wide SNP markers. Therefore, our results provide fundamental knowledge of the genetic profiles of tambaqui stocks from different parts of South America, considering the importance of generating subsidies for the progress of its production. The genetic diversity and structure of stocks highlight the risk of inbreeding and emphasize the necessity of directed mating of stocks to maintain genetic variability and avoid inbreeding. Evidence of selection signatures in tambaqui stocks associated with different farming environments was detected. Even though the number of outliers was low, most were associated with stress tolerance and immunity. Therefore, further studies are warranted to validate these predicted genes as important candidates to improve tambaqui production in many parts of South America.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The raw fastq files obtained by ddRAD-Seq methods for SNP discovery in tambaqui are available in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) (PRJNA680381). The data that support the results of this study (list of SNPs and SNPs under selection) are openly available in the DRYAD Digital Repository at https://doi.org/10.5061/dryad.mgqnk991h.

ORCID

John Fredy Gómez Agudelo https://orcid.org/0000-0001-9823-8751

Vito Antonio Mastrochirico-Filho https://orcid.org/0000-0002-3550-2671

Carolina Heloisa de Souza Borges https://orcid.org/0000-0002-6950-3082

Raquel Belini Ariede https://orcid.org/0000-0002-0641-164X

Lieschen Valeria Guerra Lira https://orcid.org/0000-0003-0971-001X

Rubens Ricardo de Oliveira Neto https://orcid.org/0000-0003-3222-7894

Milena Vieira de Freitas https://orcid.org/0000-0002-6434-2590

Manuel Vera https://orcid.org/0000-0003-1584-6140

Milthon Honorio Muñoz Berrocal https://orcid.org/0000-0002-1704-4603

Diogo Teruo Hashimoto https://orcid.org/0000-0002-8808-2498

REFERENCES

Affonso, A. G., Queiroz, H. L., & Novo, E. M. L. M. (2015). Abiotic variability among different aquatic systems of the central Amazon floodplain during drought and flood events. *Brazilian Journal of Biology*, 75(4 suppl 1), 60–69. https://doi.org/10.1590/1519-6984.04214

Aguiar, J. P., Gomes, P. F. F., Hamoy, I. G., dos Santos, S. E. B., Schneider, H., & Sampaio, I. (2018). Loss of genetic variability in the captive stocks of tambaqui, *Colossoma macropomum* (Cuvier, 1818), at breeding centres in Brazil, and their divergence from wild populations. *Aquaculture Research*, 49(5), 1914–1925. https://doi.org/10.1111/are.13647

Alfonso, S., Gesto, M., & Sadoul, B. (2020). Temperature increase and its effects on fish stress physiology in the context of global warming. *Journal of Fish Biology*, 98(6), 1496–1508. https://doi.org/10.1111/jfb.14599

Araujo-Lima, C. R. M., & Goulding, M. (1997). *So fruitful fish: Ecology, conservation, and aquaculture of the Amazon's tambaqui*. Columbia University Press.

Ariede, R. B., Freitas, M. V., Agudelo, J. F. G., Borges, C. H. S., Lira, L. V. G., Yoshida, G. M., Pilarski, F., Yáñez, J. M., & Hashimoto, D. T. (2020). Genetic (co)variation between resistance to *Aeromonas hydrophila* and growth in tambaqui (*Colossoma macropomum*). *Aquaculture*, 523, 735225. https://doi.org/10.1016/j.aquaculture.2020.735225

Boijink, C. L., Queiroz, C. A., Chagas, E. C., Chaves, F. C. M., & Inoue, L. A. K. A. (2016). Anesthetic and anthelminthic effects of clove basil (*Ocimum gratissimum*) essential oil for tambaqui (*Colossoma macropomum*). *Aquaculture*, 457, 24–28. https://doi.org/10.1016/j.aquaculture.2016.02.010

Brown, R. C., Woollians, J. A., & McAndrew, B. J. (2005). Factors influencing effective population size in commercial populations of gilthead seabream, Sparus aurata. *Aquaculture*, 247(1-4), 219–225. https://doi.org/10.1016/j.aquaculture.2005.02.002

Catchen, J. M., Amores, A., Hohenlohe, P. H., Willian, C., & John, H. P. (2011). Stacks: Building and genotyping loci de novo from short-read sequences. *G3-Genes Genomes Genetics*, 1(3), 171–182. https://doi.org/10.1534/g3.111.000240

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124–3140. https://doi.org/10.1111/mec.12354

Charlesworth, D., & Willis, J. H. (2009). The genetics of inbreeding depression. *Nature Reviews Genetics*, 10(11), 783–796. https://doi.org/10.1038/nrg2664

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118. *Fly*, 6(2), 80–92. https://doi.org/10.4161/fly.19695

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

De Mello, F. D., Oliveira, C. A. L., Ribeiro, R. P., Resende, E. K., Povh, J. A., Fornari, D. C., Barreto, R. V., Mcmanus, C., & Streit, D. Jr (2015). Growth curve by Gompertz nonlinear regression model in female and males in tambaqui (*Colossoma macropomum*). *Anais Da Academia Brasileira De Ciências*, 87(4), 2309–2315. https://doi.org/10.1590/0001-3765201520140315

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). Ne Estimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, 14(1), 209–214. https://doi.org/10.1111/1755-0998.12157

Earl, D. A., & VonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, 14(8), 2611–2620. https://doi.org/10.1111/j.1365-294x.2005.02553.x

Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity*, 103(4), 285–298. https://doi.org/10.1038/hdy.2009.74

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Falush, D., Stephens, M., & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Molecular Ecology Notes*, 7(4), 574–578. https://doi.org/10.1111/j.1471-8286.2007.01758.x

Fan, X., Hou, T., Sun, T., Zhu, L., Zhang, S., Tang, K., & Wang, Z. (2019). Starvation stress affects the maternal development and larval fitness in zebrafish (*Danio rerio*). *Science of the Total Environment*, 695, 133897. https://doi.org/10.1016/j.scitotenv.2019.133897

FAO. (1998). *Secondary guidelines for development of national farm animal genetic resources management plans. Management of small populations at risk. Initiative for domestic animal diversity*. FAO.

Fazzi-Gomes, P. F., Melo, N. F., Palheta, G., Aguiar, J., Sampaio, I., Santos, S., Moreira, F., Ribeiro-dos-Santos, Â. K., & Hamoy, I. (2017). Characterization of the genetic resources of farmed Tambaqui in Northern Brazil. *Journal of Agricultural Science*, 9(10), 76. https://doi.org/10.5539/jas.v9n10p76

Fernandes, E. M., de Almeida, L. C. F., Hashimoto, D. T., Lattanzi, G. R., Gervaz, W. R., Leonardo, A. F., & Neto, R. V. R. (2018). Survival of purebred and hybrid *Serrasalmidae* under low water temperature conditions. *Aquaculture*, 497, 97–102. https://doi.org/10.1016/j.aquaculture.2018.07.030

Ferreira, L. D. A., Fazzi-Gomes, P. F., Guerreiro, S., Rodrigues, M. D. N., Ribeiro-dos-Santos, Â. K., Santos, S., & Hamoy, I. (2019). Genetic variability of tambaqui broodstocks in the Brazilian state of Pará. *Revista Brasileira De Zootecnia*, 48, e20180106. https://doi.org/10.1590/rbz4820180106

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180(2), 977–993. https://doi.org/10.1534/genetics.108.092221

Freitas, M. V., Ariede, R. B., Hata, M. E., Mastrochirico-Filho, V. A., Pazo, F. D., Villanova, G. V., Mendonça, F. F., Porto-Foresti, F., & Hashimoto, D. T. (2021). Haplotypes traceability and genetic variability of the breeding population of pacu (*Piaractus mesopotamicus*) revealed by mitochondrial DNA. *Genetics and Molecular Biology*, 44(1), e20200249. https://doi.org/10.1590/1678-4685-gmb-2020-0249

Gjedrem, T. (2010). The first family-based breeding program in aquaculture. *Reviews in Aquaculture*, 2(1), 2–15. https://doi.org/10.1111/j.1753-5131.2010.01011.x

Gonçalves, R. A., dos Santos, C. H. A., de Sá Leitão, C. S., de Souza, É. M. S., & de Almeida-Val, A. M. F. (2018). Genetic basis of *Colossoma macropomum* broodstock: Perspectives for an improvement program. *Journal of the World Aquaculture Society*, 50(3), 633–644. https://doi.org/10.1111/jwas.12564

Gonzalez-Pena, D., Gao, G., Baranski, M., Moen, T., Cleveland, B. M., Kenney, P. B., Vallejo, R. L., Palti, Y., & Leeds, T. D. (2016). Genome-wide association study for identifying loci that affect fillet yield, carcass, and body weight traits in rainbow trout (*Oncorhynchus mykiss*). *Frontiers in Genetics*, 7, 203. https://doi.org/10.3389/fgene.2016.00203

Goudet, J., & Jombart, T. (2015). *hierfstat: Estimation and tests of hierarchical F-statistics (Version 0.04-22)*. https://cran.rproject.org/web/packages/hierfstat/index.html

Gutierrez, A. P., Yáñez, J. M., & Davidson, W. S. (2016). Evidence of recent signatures of selection during domestication in an Atlantic salmon population. *Marine Genomics*, *26*, 41–50. https://doi.org/10.1016/j.margen.2015.12.007

Hashimoto, D. T., Mendonça, F. F., Senhorini, J. A., de 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*(1-2), 49–53. https://doi.org/10.1016/j.aquaculture.2011.08.018

Houston, R. D., Bean, T. P., Macqueen, D. J., Gundappa, M. K., Jin, Y. H., Jenkins, T. L., Selly, S. L. C., Martin, S. A. M., Stevens, J. R., Santos, E. M., Davie, A., & Robledo, D. (2020). Harnessing genomics to fast-track genetic improvement in aquaculture. *Nature Reviews Genetics*, *21*(7), 389–409. https://doi.org/10.1038/s41576-020-0227-y

IBGE-Instituto Brasileiro de Geografia e Estatística. (2020). *Produção da Pecuária Municipal 2019*. Rio de Janeiro, v. 47, p.1–8. https://biblioteca.ibge.gov.br/visualizacao/periodicos/84/ppm_2019_v47_br_informativo.pdf

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 94. https://doi.org/10.1186/1471-2156-11-94

Li, H., Bob, H., Alec, W., Tim, F., Jue, R., Nils, H., Gabor, M., Goncalo, A., Richard, D., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, S., Guo, X., Lu, L. F., Lu, X. B., Wu, N., & Zhang, Y. A. (2015). Regulation pattern of fish irf4 (the gene encoding IFN regulatory factor 4) by STAT6, c-Rel and IRF4. *Developmental & Comparative Immunology*, *51*(1), 65–73. https://doi.org/10.1016/j.dci.2015.02.018

Lira, L. V. G., Ariede, R. B., Freitas, M. V., Mastrochirico-Filho, V. A., Agudelo, J. F. G., Barría, A., Yáñez, J. M., & Hashimoto, D. T. (2020). Quantitative genetic variation for resistance to the parasite *Ichthyophthirius multifiliis* in the Neotropical fish tambaqui (*Colossoma macropomum*). *Aquaculture Reports*, *17*, 100338. https://doi.org/10.1016/j.aqrep.2020.100338

Liu, L., Ang, K. P., Elliott, J. A. K., Kent, M. P., Lien, S., MacDonald, D., & Boulding, E. G. (2015). A genome scan for selection signatures comparing farmed Atlantic salmon with two wild populations: Testing colocalization among outlier markers, candidate genes, and quantitative trait loci for production traits. *Evolutionary Applications*, *10*(3), 276–296. https://doi.org/10.1111/eva.12450

Long, Y., Yan, J., Song, G., Li, X., Li, X., Li, Q., & Cui, Z. (2015). Transcriptional events co-regulated by hypoxia and cold stresses in Zebrafish larvae. *BMC Genomics*, *16*(1), 385. https://doi.org/10.1186/s12864-015-1560-y

López, M. E., Linderoth, T., Norris, A., Lhorente, J. P., Neira, R., & Yáñez, J. M. (2019). Multiple selection signatures in farmed Atlantic salmon adapted to different environments across hemispheres. *Frontiers in Genetics*, *10*, 901. https://doi.org/10.3389/fgene.2019.00901

Lopez, M. E., Neira, R., & Yáñez, J. M. (2015). Applications in the search for genomic selection signatures in fish. *Frontiers in Genetics*, *5*, 458. https://doi.org/10.3389/fgene.2014.00458

Lutfi, E., Ningping, G., Johansson, M., Sánchez-Moya, A., Björnsson, B. T., Gutiérrez, J., Navarro, I., & Capilla, E. (2018). Breeding selection of rainbow trout for high or low muscle adiposity differentially affects lipogenic capacity and lipid mobilization strategies to cope with food deprivation. *Aquaculture*, *495*, 161–171. https://doi.org/10.1016/j.aquaculture.2018.05.039

Mäkinen, H., Vasemägi, A., McGinnity, P., Cross, T. F., & Primmer, C. R. (2014). Population genomic analyses of early-phase Atlantic Salmon (*Salmo salar*) domestication/captive breeding. *Evolutionary Applications*, *8*(1), 93–107. https://doi.org/10.1111/eva.12230

Marandel, F., Charrier, G., Lamy, J. B., Le Cam, S., Lorance, P., & Trenkel, V. M. (2020). Estimating effective population size using RADseq: Effects of SNP selection and sample size. *Ecology and Evolution*, *10*(4), 1929–1937. https://doi.org/10.1002/ece3.6016

Mastrochirico-Filho, V. A., Ariede, R. B., Freitas, M. V., Borges, C. H. S., Lira, L. V. G., Mendes, N. J., Agudelo, J. F. G., Cáceres, P., Berrocal, M. H. M., Sucerquia, G. A. L., Porto-Foresti, F., Yáñez, J. M., & Hashimoto, D. T. (2021). Development of a multi-species SNP array for Serrasalmid fish *Colossoma macropomum* and *Piaractus mesopotamicus*. *Scientific Reports*, *11*(1), 19289. https://doi.org/10.1038/s41598-021-98885-x

Mastrochirico-Filho, V. A., del Pazo, F., Hata, M. E., Villanova, G. V., Foresti, F., Vera, M., Martínez, P., Porto-Foresti, F., & Hashimoto, D. T. (2019). Assessing genetic diversity for a pre-breeding program in *Piaractus mesopotamicus* by SNPs and SSRs. *Genes*, *10*(9), 668. https://doi.org/10.3390/genes10090668

Melo, D. C., Oliveira, D. A. A., Ribeiro, L. P., Teixeira, C. S., Sousa, A. B., Coelho, E. G. A., Crepaldi, D. V., & Teixeira, E. A. (2006). Caracterização genética de seis plantéis comerciais de tilápia (*Oreochromis*) utilizando marcadores microssatélites. *Arquivo Brasileiro De Medicina Veterinária E Zootecnia*, *58*(1), 87–93. https://doi.org/10.1590/s0102-09352006000100013

Metzger, D. C. H., & Schulte, P. M. (2018). Similarities in temperature-dependent gene expression plasticity across timescales in threespine stickleback (*Gasterosteus aculeatus*). *Molecular Ecology*, *27*(10), 2381–2396. https://doi.org/10.1111/mec.14591

Narum, S. (2006). Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conservation Genetics*, *7*(5), 783–787. https://doi.org/10.1007/s10592-005-9056-y

Narum, S. R., & Hess, J. E. (2011). Comparison of FST outlier tests for SNP loci under selection. *Molecular Ecology Resources*, *11*(Suppl 1), 184–194. https://doi.org/10.1111/j.1755-0998.2011.02987.x

Nei, M. (1972). Genetic distance between populations. *The American Naturalist*, *106*(949), 283–292. https://doi.org/10.1086/282771

Nunes, J., Liu, S., Pértille, F., Perazza, C. A., Villela, P. M. S., Almeida-Val, V. M. F., Hilsdorf, A. W. S., Liu, Z., & Coutinho, L. L. (2017). Large-scale SNP discovery and construction of a high-density genetic map of Colossoma macropomum through genotyping-by-sequencing. *Scientific Reports*, *7*(1), 46112. https://doi.org/10.1038/srep46112

Nunes, J. R. S., Pértille, F., Andrade, S. C. S., Perazza, C. A., Villela, P. M. S., Almeida-Val, V. M. F., Gao, Z. X., Coutinho, L. L., & Hilsdorf, A. W. S. (2020). Genome-wide association study reveals genes associated with the absence of intermuscular bones in tambaqui (*Colossoma macropomum*). *Animal Genetics*, *51*(6), 899–909. https://doi.org/10.1111/age.13001

Perazza, C. A., Ferraz, J. B. S., Almeida-Val, V. M. F., & Hilsdorf, A. W. S. (2019). Genetic parameters for loin eye area and other body traits of an important Neotropical aquaculture species, *Colossoma macropomum* (Cuvier, 1818). *Aquaculture Research*, *50*(10), 2907–2916. https://doi.org/10.1111/are.14245

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, *7*(5), e37135. https://doi.org/10.1371/journal.pone.0037135

Prado-Lima, M., & Val, A. L. (2016). Transcriptomic characterization of Tambaqui (*Colossoma macropomum*, Cuvier, 1818) exposed to three climate change scenarios. *PLoS One*, *11*(3), e0152366. https://doi.org/10.1371/journal.pone.0152366

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Raposo de Magalhães, C., Schrama, D., Nakharuthai, C., Boonanuntanasarn, S., Revets, D., Planchon, S., Kuehn, A., Cerqueira, M., Carrilho, R., Farinha, A. P., & Rodrigues, P. M. (2021). Metabolic plasticity of gilthead seabream under different stressors: Analysis of the stress responsive hepatic proteome and gene expression. *Frontiers in Marine Science*, *8*. https://doi.org/10.3389/fmars.2021.676189

Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P., & Houston, R. (2018). Applications of genotyping by sequencing in aquaculture breeding and genetics. *Reviews in Aquaculture*, *10*(3), 670–682. https://doi.org/10.1111/raq.12193

Santos, M. C. F., Ruffino, M. L., & Farias, I. P. (2007). High levels of genetic variability and panmixia of the tambaqui *Colossoma macropomum* (Cuvier, 1816) in the main channel of the Amazon River. *Journal of Fish Biology*, *71*, 33–44. https://doi.org/10.1111/j.1095-8649.2007.01514.x

Swirplies, F., Wuertz, S., Baßmann, B., Orban, A., Schäfer, N., Brunner, R. M., Hadlich, F., Goldammer, T., & Rebl, A. (2019). Identification of molecular stress indicators in pikeperch *Sander lucioperca* correlating with rising water temperatures. *Aquaculture*, *501*, 260–271. https://doi.org/10.1016/j.aquaculture.2018.11.043

Teletchea, F. (2018). *Fish domestication: An overview*. Animal domestication. IntechOpen. https://doi.org/10.5772/intechopen.79628

Teletchea, F., & Fontaine, P. (2014). Levels of domestication in fish: implications for the sustainable future of aquaculture. *Fish and Fisheries*, *15*(2), 181–195. https://doi.org/10.1111/faf.12006

Torati, L. S., Taggart, J. B., Varela, E. S., Araripe, J., Wehner, S., & Migaud, H. (2019). Genetic diversity and structure in *Arapaima gigas* populations from Amazon and Araguaia-Tocantins river basins. *BMC Genetics*, *20*(1), 13. https://doi.org/10.1186/s12863-018-0711-y

Valladão, G. M. R., Gallani, S. U., & Pilarski, F. (2018). South American fish for continental aquaculture. *Reviews in Aquaculture*, *10*(2), 351–369. https://doi.org/10.1111/raq.12164

Vera, M., Pardo, B. G., Cao, A., Vilas, R., Fernández, C., Blanco, A., Gutierrez, A. P., Bean, T. P., Houston, R. D., Villalba, A., & Martínez, P. (2019). Signatures of selection for bonamiosis resistance in European flat oyster (*Ostrea edulis*): New genomic tools for breeding programs and management of natural resources. *Evolutionary Applications*, *12*(9), 1781–1796. https://doi.org/10.1111/eva.12832

Vincze, T. (2003). NEBcutter: A program to cleave DNA with restriction enzymes. *Nucleic Acids Research*, *31*(13), 3688–3691. https://doi.org/10.1093/nar/gkg526

Wang, X., Zhang, T., Mao, H., Mi, Y., Zhong, B., Wei, L., Liu, X., & Hu, C. (2016). Grass carp (*Ctenopharyngodon idella*) ATF6 (activating transcription factor 6) modulates the transcriptional level of GRP78 and GRP94 in CIK cells. *Fish & Shellfish Immunology*, *52*, 65–73. https://doi.org/10.1016/j.fsi.2016.03.028

Woynárovich, A., & Van Anrooy, R. (2019). *Field guide to the culture of tambaqui* (Colossoma macropomum, *Cuvier, 1816*). FAO Fisheries and Aquaculture Technical Paper No. 624. Rome, FAO. (pp. 132).

Wu, P., Wang, A., Cheng, J., Chen, L., Pan, Y., Li, H., Zhang, Q., Zhang, J., Chu, W., & Zhang, J. (2020). Effects of starvation on antioxidant-related signaling molecules, oxidative stress, and autophagy in Juvenile Chinese Perch Skeletal Muscle. *Marine Biotechnology*, *22*(1), 81–93. https://doi.org/10.1007/s10126-019-09933-7

Yousefian, M., & Nejati, A. (2008). Inbreeding Depression by Family Matching in Rainbow Trout (*Oncorhynchus mykiss*). *Journal of Fisheries and Aquatic Science*, *3*(6), 384–391. https://doi.org/10.3923/jfas.2008.384.391

Zaniboni-Filho, E., & Meurer, S. (1997). Limitações e potencialidades do cultivo de tambaqui (Colossoma macropomum Cuvier, 1818) na região subtropical brasileira. *Boletim Instituto Pesca*, *24*, 169–172.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.