

UNIVERSIDADE ESTADUAL PAULISTA – UNESP
CENTRO DE AQUICULTURA DA UNESP

**Estudos de variabilidade genética em lotes
cultivados de tambaqui (*Colossoma
macropomum*) na América do Sul**

John Fredy Gomez Agudelo

Jaboticabal, São Paulo

2020

UNIVERSIDADE ESTADUAL PAULISTA – UNESP
CENTRO DE AQUICULTURA DA UNESP

**Estudos de variabilidade genética em lotes
cultivados de tambaqui (*Colossoma
macropomum*) na América do Sul**

John Fredy Gomez Agudelo

Orientador: Prof. Dr. Diogo Teruo Hashimoto

Dissertação apresentada ao Programa de Pós-graduação em Aquicultura do Centro de Aquicultura da UNESP-CAUNESP, como parte dos requisitos para obtenção do título de Mestre em Aquicultura.

Jaboticabal, São Paulo
2020

G633e Gómez Agudelo, John Fredy
Estudos de variabilidade genética em lotes cultivados de tabaqui
(Colossoma macropomum) / John Fredy Gómez Agudelo. -- Jaboticabal,
2020
ix, 61 p. : il. ; 29 cm

Dissertação (mestrado) - Universidade Estadual Paulista, Centro de
Aquicultura, 2020
Orientador: Diogo Hashimoto
Banca examinadora: Márcio Leite de Oliveira, Milthon H. Muñoz Berrocal
Bibliografia

1. Melhoramento genético. 2. Variabilidade genética. 3. Gargalo
genético. 4. Marcadores moleculares. I. Título. II. Jaboticabal-Centro de
Aquicultura.

CDU 639.3:636.082

Ficha Catalográfica elaborada pela STATI - Biblioteca da UNESP
Campus de Jaboticabal/SP - Karina Gimenes Fernandes - CRB 8/7418

John Fredy Gomez Agudelo

Estudos de variabilidade genética em lotes cultivados de tabaqui
(*Colossoma macropomum*) na América do Sul.

Dissertação apresentada ao programa de Pós-Graduação em Aquicultura,
Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de
Jaboticabal, como parte dos requisitos para obtenção do título de Mestre.

Orientador: Prof. Dr. Diogo Teruo Hashimoto

Comissão Examinadora:

Prof. Dr. Diogo Teruo Hashimoto - UNESP

Prof. Dr. Márcio Leite de Oliveira- NUPECCE

Prof. Dr. Milthon Honorio Muñoz Berrocal –
Universidad Nacional Agraria de la Selva, Tingo Maria, Peru

Jaboticabal, _____ de _____ de _____.

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: Estudos de variabilidade genética em lotes cultivados de tambaqui (*Colossoma macropomum*) na América do Sul

AUTOR: JOHN FREDY GOMEZ AGUDELO

ORIENTADOR: DIOGO TERUO HASHIMOTO

Aprovado como parte das exigências para obtenção do Título de Mestre em AQUICULTURA, pela Comissão Examinadora:


Prof. Dr. DIOGO TERUO HASHIMOTO
Laboratório de Genética de Peixes / Centro de Aquicultura - CAUNESP, Jaboticabal-SP


Prof. Dr. MARCIO LEITE DE OLIVEIRA
Departamento de Zootecnia / UNESP/Câmpus de Jaboticabal


Prof. Dr. MILTHON HONORIO MUÑOZ BERROCAL
. / Universidad Nacional Agraria de la Selva, Tingo María, Peru

Jaboticabal, 02 de março de 2020

AGRADECIMENTOS

Agradeço às circunstâncias da vida por ter tido a oportunidade de aprender e trabalhar em uma área tão empolgante quanto a genética, de conhecer e compartilhar com pessoas cujo conhecimento é inspirador, além de ter a experiência de estudar e fazer parte de um centro de pesquisa e universidade tão importantes.

A minha mãe e irmão que sempre me ajudaram e me apoiaram. Muito obrigado por tudo.

A minha namorada e meu filho que se esforçaram de longe para que eu pudesse fazer meus estudos.

Ao meu orientador, Prof. Dr. Diogo Teruo Hashimoto, por ter confiado em mim, por toda a ajuda e apoio durante o desenvolvimento do trabalho. Muito obrigado por tudo.

A todos os meus amigos e colegas do LaGeAC (Carol, Milena, Natalia, Raquel, Rubens, Valéria e Vito) muito obrigado pela ajuda e apoio prestado,

Ao Prof. Dr. Gustavo Adolfo Lenis Sucerquia por ajudar a vir para Brasil a fazer o estágio, e ajudar na coleta de amostras na Colômbia, para o projeto de mestrado.

Ao Prof. Dr. Milthon Muñoz Berrocal muito obrigado por ajudar nas coletas no Peru, e fazer parte da comissão examinadora, com apoio financeiro da FAPESP Processo 19/08972-8.

Ao CNPq pelo apoio financeiro para desenvolver o projeto, sem essa ajuda, seria impossível realizar este estudo.

As pisciculturas da Colômbia e Peru que colaboraram com as informações e a coleta de amostras. Muito obrigado pela colaboração.

Muito obrigado a todos

RESUMO

John F.G. Agudelo. **Estudos de variabilidade genética em lotes cultivados de tambaqui (*Colossoma macropomum*) na América do Sul.** 2020. Dissertação de mestrado – Centro de Aquicultura da Universidade Estadual Paulista, CAUNESP, Jaboticabal, 2020.

Atualmente, os estoques de peixes têm sofrido altos níveis de sobreexploração pela pesca e estão ameaçados de extinção devido às alterações dos habitats naturais. Portanto, a produção de pescado oriunda de uma aquicultura sustentável certamente poderá auxiliar a suprir a demanda de proteína animal. Programas de melhoramento genético podem garantir a maximização do desempenho produtivo nos sistemas de cultivo, pois são capazes de otimizar o rendimento da produção da espécie-alvo. Entretanto, é fundamental que se realize primeiramente estudos de variabilidade genética por meio de marcadores moleculares, de maneira a evitar o gargalo genético e depressão endogâmica. Para esta finalidade, SNPs (*Single Nucleotide Polymorphisms*) são considerados os marcadores moleculares mais apropriados, uma vez que constituem o polimorfismo mais abundante em qualquer organismo. O tambaqui (*Colossoma macropomum*) é um dos peixes de maior valor comercial e de potencial para a piscicultura da América do Sul, portanto, uma espécie-alvo para programas de melhoramento genético. O presente trabalho teve como objetivo analisar a variabilidade genética de estoques de tambaqui na América do Sul, utilizando marcadores SNPs obtidos por sequenciamento de nova geração. Inicialmente, foram sequenciados cerca de 350 reprodutores de tambaqui, sendo oriundos de pisciculturas de diferentes localidades: Sul e Norte do Brasil, Colômbia, Peru; e uma população selvagem (rio Amazonas) e a população base do núcleo de melhoramento do CAUNESP (Jaboticabal, SP). A estratégia de sequenciamento/genotipagem dos SNPs foi por meio da técnica ddRAD-seq (*Double-digest RAD-sequencing*). Após os filtros de qualidade, foram selecionados cerca de 1.440 SNPs de 220 amostras de tambaqui para os estudos de genômica populacional. Os valores de H_o variaram de 0.081 a 0.248; enquanto de H_e foram de 0.206 a 0.270. Os resultados da análise hierárquica de variância molecular (AMOVA) mostraram que a maior proporção da variância esteve presente dentro das populações (75.2%). Uma baixa diferenciação genética entre todas as populações foi detectada por análises de F_{st} global sobre os loci de SNPs ($F_{st} = 0,05$). A análise de estruturação genética identificou dois agrupamentos genéticos nas populações comerciais de tambaqui da América do Sul. Estudos genéticos enfocando um maior conhecimento do genoma dos estoques cultivados de tambaqui são considerados essenciais para aquicultura, principalmente no sentido de fornecer subsídios para o desenvolvimento de programas de manejo e melhoramento desta espécie.

ABSTRACT

John F.G. Agudelo. Genetic variability studies in cultivated tambaqui (*Colossoma macropomum*) lots in South America. 2020. Master's dissertation - Aquaculture Center at Universidade Estadual Paulista, CAUNESP, Jaboticabal, 2020.

Currently, fish stocks have suffered from high levels of overexploitation by fishing and are threatened with extinction due to changes in natural habitats. Therefore, the production of fish from sustainable aquaculture can certainly help to meet the demand for animal protein. Genetic improvement programs can guarantee the maximization of productive performance in cultivation systems, because they are able to optimize the production yield of the target species. However, it is essential that genetic variability studies be carried out first through molecular markers, in order to avoid the genetic bottleneck and inbreeding depression. For this purpose, SNPs (Single Nucleotide Polymorphisms) are considered the most appropriate molecular markers, since they constitute the most abundant polymorphism in any organism. Tambaqui (*Colossoma macropomum*) is one of the fish with the greatest commercial value and potential for fish farming in South America, therefore, a target species for breeding programs. The present work aimed to analyze the genetic variability of tambaqui stocks in South America, using SNPs markers obtained by new generation sequencing. Initially, about 350 tambaqui breeders were sequenced, coming from fish farms in different locations: South and North of Brazil, Colombia, Peru; and a wild population (Amazon River) and the base population of the CAUNESP improvement center (Jaboticabal, SP). The SNP sequencing / genotyping strategy was through the ddRAD-seq (Double-digest RAD-sequencing) technique. After the quality filters, about 1,440 SNPs from 220 tambaqui samples were selected for population genomics studies. H_o values ranged from 0.081 to 0.248; while from H_e they were 0.206 to 0.270. The results of the hierarchical analysis of molecular variance (AMOVA) showed that the highest proportion of variance was present within populations (75.2%). Low genetic differentiation among all populations was detected by global F_{st} analyzes on SNP loci ($F_{st} = 0.05$). The analysis of genetic structure identified two genetic groups in the commercial populations of tambaqui in South America. Genetic studies focusing on greater knowledge of the genome of cultivated stocks of tambaqui are considered essential for aquaculture, mainly in the sense of providing subsidies for the development of management and improvement programs for this species.

LISTA DE ILUSTRAÇÕES

Figura 1 – Mapa Localidade e número de amostras de tambaqui <i>Colossoma macropomum</i> coletadas para o estudo de variabilidade genética.	30
Figura 2 – Distribuição dos valores de parentesco (r_{xy}).	40
Figura 3 – Gráfico Delta K.....	43
Figura 4 – Análise de estruturação genética dos estoques de tambaqui <i>Colossoma macropomum</i>	43

LISTA DE TABELAS

Tabela 1 – Localidade e número de amostras de tambaqui <i>Colossoma macropomum</i> coletadas para o estudo de variabilidade genética.....	24
Tabela 2 – Dados dos SNPs e loci após os filtros de qualidade, resultantes do sequenciamento de nova geração das oito bibliotecas de ddRADseq dos estoques de tambaqui <i>Colossoma macropomum</i>	32
Tabela 3 – Dados dos SNPs e loci após os filtros de qualidade do plink, resultantes do sequenciamento de nova geração das oito bibliotecas de ddRADseq dos estoques de tambaqui <i>Colossoma macropomum</i>	33
Tabela 4 – Dados das sequencias iniciais e após os filtros de qualidade, resultantes do sequenciamento de nova geração das oito bibliotecas de ddRADseq dos estoques de tambaqui <i>Colossoma macropomum</i>	34
Tabela 5 – Média dos parâmetros de diversidade genética dos estoques de tambaqui <i>Colossoma macropomum</i> . n: número amostral analisado; Ho: heterozigidade observada; He: heterozigidade esperada; MAF: frequência alélica mínima; e Fis: coeficiente de endogamia.....	36
Tabela 6 – Tamanho efetivo da população (Ne) parâmetro baseado no excesso de heterozigidade, coeficiente de endogamia ($\Delta F = 1/2Ne$).....	37
Tabela 7 – Análise de parentesco para os 16 locais em populações de tambaqui (<i>Colossoma macropomun</i>) de acordo com o coeficiente rxy de Wang. Os valores são representados em porcentagem.....	38
Tabela 8 – Análise de variância molecular (AMOVA).....	39
Tabela 9 – Pairwise Fst dos estoques de tambaqui <i>Colossoma macropomum</i>	40

SUMÁRIO

1. Introdução geral	12
1.2 Aspectos gerais.....	12
1.2 Caracterização da espécie <i>Colossoma macropomum</i>	17
1.3 Problemas decorrentes da endogamia	21
1.4 Marcadores SNPs para análise de variabilidade genética.....	23
2. Objetivos	29
2.1 Geral	29
2.2 Específicos	29
3. Material e Métodos.....	29
3.1 Material.....	29
3.2 Caracterização dos locais de coleta.....	31
Colombia	31
Perú	32
Brasil	33
3.3 Construção de bibliotecas ddRADseq	34
3.4 Identificação e filtro dos SNPs.....	37
3.5 Análises de genética populacional	39
4. Resultados	40
4.1 Análises de bioinformática para seleção dos SNPs	40
4.2 Análises genético-populacionais	41
5. Discussão	48
6. Conclusão	53
7. Recomendações	53
8. Anexos	56
9. Referências.....	59

1. Introdução geral

1.2 Aspectos gerais

A demanda global por proteínas derivadas de pescado tem aumentado ao longo das últimas décadas, principalmente devido ao crescimento da população e preferência por alimentos saudáveis (Bacher, 2015). De acordo com o relatório da Organização das Nações Unidas para Alimentação e Agricultura (FAO, 2018), a produção pesqueira mundial atingiu aproximadamente 170,9 milhões de toneladas em 2016; 151,2 milhões de toneladas correspondem ao consumo humano e 19,7 milhões de toneladas representam usos não alimentares, como farinhas e óleos (FAO, 2018). Globalmente, em 2016, a produção aquícola a nível continental foi de 51,4 milhões de toneladas (30%) da produção total, e a atividade de pesca e captura de espécies continentais alcançou 11,6 milhões de toneladas (6,8%) da produção mundial. (FAO, 2018),

Em termos de consumo *per capita* de peixe mundialmente, houve um aumento de 9,0 kg (1961) para 20,2 kg em 2015, uma taxa média de crescimento de 1,5% ao ano. Os aumentos no consumo não se devem apenas ao aumento da produção, mas também a fatores relacionados à redução de resíduos e os desperdícios (FAO, 2018).

Enquanto isso, as principais populações de peixes estão sendo sobreexploradas pela pesca. As Nações Unidas através de objetivos de desenvolvimento sustentável (ODS), prospectam uma meta de 14.4% para regulamentar a exploração pesqueira e, desta forma, diminuir a pesca excessiva e por conseguinte reestabelecer as populações de peixes, para conseguir níveis que possam chegar ao máximo rendimento sustentável (MRS). Porém, a situação parece ser difícil para que as produções pesqueiras possam chegar a reestabelecer em um curto prazo o 33.1% das populações que estão sendo sobreexploradas na atualidade, devido ao fato de que para o reestabelecimento se necessita de tempo, podendo levar a vários ciclos de vida do animal (FAO, 2018).

A estagnação da pesca por captura começou a ser observada no final da década de 1980. Neste momento, houve um aumento significativo na aquicultura, e um aumento contínuo na oferta de pescado para o consumo humano. De acordo com as últimas estatísticas disponíveis, em 2016, a produção de peixes pela aquicultura continental e marinha apresentou acelerado crescimento e já representa 47% da produção total de pescado (FAO, 2018).

Neste contexto da estagnação da pesca, a aquicultura terá que suprir a maior parte do aumento na demanda por pescado, de maneira correta e sustentável. Esta importância da aquicultura ocorre em um momento em que o mundo se tornou mais consciente sobre as questões ambientais, e consumidores passaram a exigir produtos mais seguros (Bacher, 2015).

No âmbito da produção aquícola regional na América do Sul, a atividade da aquicultura é liderada pelo Chile, seguido pelo Brasil e Equador, que juntos representaram 90% da produção da América Latina em 2015 e atingiram uma produção de 2,1 milhões toneladas. O Peru, com 4% da produção, ocupa o quarto lugar e o restante dos países juntos produz 6% (Regalado, 2017). Segundo estatísticas da FAO, a produção no Brasil em 2017 foi de 595.000 toneladas, das quais 514.070 toneladas são de espécies de água doce e 80.930 toneladas são de espécies marinhas (FAO Fish Stat, 2018). Em 2018, com um crescimento de 4,5%, 722.560 toneladas de peixe foram produzidas (Peixebr, 2019). No Peru, no ano 2017 a produção atingiu 100.453 toneladas, 61.028 toneladas de espécies de água doce e 39.425 de espécies marinhas (FAO Fish Stat, 2018). Por outro lado, a produção na Colômbia atingiu 100.000 toneladas, sendo 97.600 toneladas em espécies de água doce e 2.400 toneladas de espécies marinhas (FAO Fish Stat, 2018). A produção da aquicultura na América Latina

e no Caribe deve passar de 2,7 milhões de toneladas (2016) a 4 milhões de toneladas em 2030, com uma meta de crescimento prevista de 49,2% (FAO, 2018).

Existem basicamente duas alternativas para aquicultores melhorarem o desempenho produtivo nos sistemas de cultivo. O primeiro é por meio de uma ampliação em suas instalações, o que frequentemente é inviável, pois nem sempre existem áreas ou volume de água disponíveis, além da ampliação quase sempre resultar em elevação nos custos de produção. O segundo é maximizar sua produtividade, ou seja, melhorar o rendimento da produção por peixe produzido por unidade de área cultivada, tanto através de melhorias ambientais realizadas na propriedade, tais como a prática de um manejo adequado, com equilibrada taxa de arraçamento e controle da qualidade da água, quanto por meio de um programa de melhoramento genético com o objetivo de aumentar o ganho genético nos peixes que estão sendo cultivados (Tave, 1995).

Programas de melhoramento genético são cruciais para o desenvolvimento da piscicultura, não só com o objetivo de aumentar as taxas de crescimento e o desempenho produtivo das espécies cultivadas, mas também para reduzir os custos de produção, melhorar a resistência dos organismos cultivados a doenças, melhorar a qualidade dos produtos e, assim, garantir a utilização mais eficiente de alimento, água, e áreas de cultivo disponíveis (Gjedren, 1997).

Entre as técnicas utilizadas no Brasil para obtenção de material genético mais produtivo, a hibridação interespecífica é responsável pela produção de milhares de toneladas de peixes ao ano (Ibge, 2016). O tambaqui (*Colossoma macropomum*) junto com duas espécies de peixes amplamente utilizadas na aquicultura, que também pertencem à família Serrasalminidae, a pirapitinga (*Piaractus brachypomus* Cuvier, 1818) e o pacu (*Piaractus mesopotamicus* Holmberg, 1887), três espécies muito importantes

comercialmente na aquicultura.(FAO, 2019). Porém, apesar do aparente bom resultado zootécnico da hibridação como método de melhoramento entre espécies nativas, existe uma séria preocupação com a possibilidade de escape e introgressão genética desses híbridos sobre as populações selvagens (Hashimoto *et al.*, 2012, 2014), além do risco da implementação enganosa de híbridos não férteis no plantel de reprodutores, pela dificuldade de se diferenciar híbridos de espécies puras, principalmente nos estágios iniciais de vida (Hashimoto *et al.*, 2011). Portanto, para evitar possíveis impactos, a aplicação de técnicas como a seleção e cruzamento dirigido entre indivíduos de uma mesma espécie com maior variabilidade genética se faz necessário, usando a variabilidade genética presente nestas populações é uma alternativa mais sustentável (Hilsdorf e Orfão, 2011).

Vários fatores são importantes para que os programas de melhoramento por seleção conduzam a ganhos genéticos expressivos e duradouros. Dentre eles, é fundamental realizar um adequado programa de pré-melhoramento, principalmente por meio da avaliação da variabilidade genética inicial, que irá evitar os problemas decorrentes do estreitamento da base genética de certas espécies ou estoques (Ponzoni, 2006). Estudos de variabilidade genética por meio de marcadores moleculares efetivos são necessários para a organização de uma estrutura de geração e análise de dados genéticos em longo prazo. A avaliação genética torna-se ainda mais importante diante das alterações nas frequências gênicas, e possibilidade de redução da variabilidade genética no processo de seleção de organismos geneticamente superiores, o que pode resultar no aumento da probabilidade de acasalamento de indivíduos aparentados e surgimento de genes deletérios, dizimando o potencial de linhagem da produção piscícola (Oliveira *et al.*, 2011).

A piscicultura de água doce no Brasil é dividida em dois setores: a produção de alevinos e o cultivo de peixes engorda (Suplicy, 2007). Na década de 1980, a indústria do comércio de alevinos foi fortemente focada para fornecer peixes a empreendimentos de lazer ou repovoamento de represas. A produção de alevinos para a piscicultura comercial visando o abastecimento de restaurantes, supermercados e fábricas de processamento e filetagem só foram estabelecidos na década de 1990, o que trouxe consideráveis mudanças na demanda do consumidor (Suplicy, 2007). *C. macropomun* é um peixe com características muito particulares que a tornam adequada para a aquicultura, os avanços obtidos em sua reprodução, a facilidade de manejo, o rápido crescimento e a qualidade de sua carne, fizeram desta espécie uma das mais demandadas pelos mercados locais e regionais (BACA, 2015). Devido à sobreexploração do tambaqui, a população selvagem diminuiu consideravelmente entre 1976 e 1996. Essa foi uma das razões para aumentar os esforços e desenvolver técnicas confiáveis de propagação dessa espécie para a produção de alevinos tanto para repovoamento quanto aquicultura (FAO, 2019).

O tambaqui foi introduzido em locais fora de sua distribuição nativa, o que levou a um aumento na produção. Atualmente, o tambaqui e outras espécies de Serrasalminidae foram estendidas para a maioria dos países da América do Sul e Central e em alguns países do Caribe e vários países da Ásia, particularmente China, Indonésia, Malásia, Mianmar e Vietnã. (FAO, 2019). No Brasil há uma crescente exigência por peixes de melhor qualidade, o que mostra à necessidade do estabelecimento de estratégias de manejo genético para as pisciculturas nacionais, principalmente para os produtores de alevinos, que é considerado a base deste setor.

1.2 Caracterização da espécie *Colossoma macropomum*

O tambaqui *Colossoma macropomum* (Cuvier, 1818) está incluído dentro da família Serrasalminidae (Teleostei, Characiformes) e possui uma ampla distribuição na bacia do rio Amazonas e Orinoco (Jégu, 2003). Por isso, é comum encontrá-lo em toda a região hidrográfica da Amazônia, que é composta por 8 países, 63,88% do território está nos limites do Brasil, 16,14% está na Colômbia 15,61% Bolívia; 2,31%; Equador; 1,35% na Guiana; 0,60% no Peru e 0,11% na Venezuela. (MMA, 2006)

Na Bacia Amazônica, foram relatadas cerca de 2.500 espécies de peixes. Entre eles, o tambaqui (*Colossoma macropomum*), considerado o maior caracídeo da Amazônia, é a segunda maior espécie escamada, depois de *Arapaima gigas* (Ostoglossidae), (Baca, 2015). O tambaqui é uma espécie importante do ponto de vista cultural e econômico da pesca artesanal e, além da produção aquícola, é uma das espécies mais importantes de peixes escamados da bacia amazônica (FAO, 2019).

A espécie *Colossoma macropomun* (Cuvier, 1818) é um peixe reofílico neotropical, comumente conhecido como tambaqui no Brasil, Gamitana no Peru (Fazzi-Gomes et al., 2017) e Cachama Negra na Colômbia (Pineda Santis; Restrepo Betancur; Olivera Ángel, 2004). O formato do corpo é semelhante ao das piranhas (peixe em formato redondo) podendo medir um metro de comprimento e pesar 30 kg. Os hábitos alimentares são contrários à piranha, portanto, se alimenta principalmente de frutos, sementes e zooplâncton. É considerado um peixe semi-migratório, realizando migrações sazonais dos lagos de várzea para a reprodução e alimentação (Araújo-Lima e Ruffino, 2003, Fazzi-Gomes et al., 2017, Ariede et al., 2018). quando o nível dos rios aumenta devido às chuvas, e atingem a temperatura média de 27 ° C os indivíduos sexualmente maduros migram rio acima para desovar. Os locais onde geralmente se reproduzem são áreas das margens dos rios, próximas a leitos naturais ou zonas de inundação conectadas

aos rios, a desova pode ocorrer em lugares com áreas de floresta ao longo das margens com vegetação flutuante ou abaixo delas. Em seu ambiente natural, a maturação sexual ocorre entre 4 a 5 anos e 5 a 7 kg de peso, sob condições de cativeiro em clima tropical, esse evento ocorre aos 3 ou 4 anos e peso de 4 a 5 kg (FAO, 2019).

Em relação à sua biologia, o tambaqui prefere águas quentes, ricas em nutrientes encontrados em seu habitat natural, e são intolerantes a grandes variações de temperatura (Valladão et al., 2018). Serralmidos do gênero *Colossoma* têm respiração braquial forçada, em concentrações de oxigênio inferiores a 0,5 mg/L o tambaqui têm a capacidade de captar oxigênio da camada superficial da água para respirar, essa característica é denominada respiração aquática da superfície, essa adaptação eco-mórfica gera a extensão do derma da mandíbula inferior (Baca, 2015).

Devido aos aspectos migratórios, análises moleculares têm indicado que esta espécie pode formar uma população panmítica na bacia Amazônica (Santos et al., 1997), mesmo no Rio Madeira, onde cachoeiras não atuam efetivamente como barreiras geográficas para isolamento de populações (Farias et al., 2010). Por isso, as populações de tambaqui poderão ser afetadas negativamente pela instalação e construção de hidrelétricas na bacia Amazônica.

O tambaqui é um dos principais recursos explorados pela pesca extrativa na região Norte. A produção de pescado desta espécie pela pesca extrativista tem alcançado mais de quatro mil toneladas por ano (MPA, 2012). Entretanto, devido principalmente a esta alta taxa de pesca predatória, os estoques desse peixe já são considerados sobre-explorados ou ameaçados de sobre-exploração no Brasil, de acordo a Instrução Normativa do Ministério do Meio Ambiente (MMA, nº 05/2004). Assim, o cultivo desta espécie é uma alternativa para suprir a demanda do mercado consumidor e uma estratégia para auxiliar na redução da pressão de pesca sobre as populações naturais.

Em termos de produção na aquicultura, o tambaqui é a espécie nativa da América do Sul mais cultivada nas pisciculturas. No período de 2003-2009, a produção de tambaqui cresceu 123% (MPA, 2010). Segundo estimativas da FAO em 2017, a produção de *Colossoma macropomun* no Brasil atingiu 105.000 toneladas, na Colômbia atingiu 1.800 toneladas e o Peru teve uma produção de 1.047 toneladas (FAO –Fisheries 2019).

Além disso, o tambaqui é utilizado para cruzamentos com outras espécies de Serrasalmidae (*Piaractus mesopotamicus* e *Piaractus brachypomus*), resultando em híbridos interespecíficos, que são muito utilizados nas regiões Centro-Oeste e Sudeste (Hashimoto *et al.*, 2012). Em 2015, os híbridos tambatinga (*C. macropomum* x *P. brachypomus*) e tambacu (*C. macropomum* x *P. mesopotamicus*) tiveram uma produção de 37 mil t, o que representou o segundo maior valor para peixes nativos brasileiros (IBGE, 2016). Portanto, o tambaqui e seus híbridos têm contribuído com mais de 35% da produção da aquicultura nacional e, além disso, tem recebido destaque em outros países da América do Sul, como Equador, Bolívia, Colômbia, Peru e Venezuela (Valladão *et al.*, 2018).

Além da qualidade nutricional da carne, aceitação de mercado, características zootécnicas e adaptações favoráveis aos sistemas de cultivo, o destaque do tambaqui se deve especialmente aos avanços em pesquisa e estudos científicos (Gomes *et al.*, 2010; Oliveira *et al.*, 2012). Por isso, é considerado o principal peixe nativo da aquicultura sul-americana e, desta forma, uma espécie alvo para programas de melhoramento por seleção clássica, que já tem sido feito por universidades, centros de pesquisa e empresas públicas e privadas, especialmente no Brasil. Todas estas informações justificam a aplicação de biotecnologias modernas com intuito de aumentar a produção do tambaqui, em especial o uso de marcadores moleculares para análise de variabilidade genética.

Do ponto de vista genético, alguns estudos em pisciculturas do Norte e Nordeste do Brasil têm demonstrado uma baixa diversidade genética em estoques cultivados de tambaqui. Por meio de análises do DNA mitocondrial, foram observados apenas dois haplótipos entre 93 exemplares de tambaqui analisados de 10 pisciculturas (Gomes *et al.*, 2012), o que demonstra que estes lotes podem estar sendo afetados por alto grau de endogamia. Além disso, com o uso de marcadores microssatélites, uma baixa diversidade genética em estoques de reprodutores de três pisciculturas também foi relatada, como reduzido número de alelos e heterozigosidade em relação a uma população selvagem (Santos *et al.*, 2016).

Por outro lado, um distinto padrão também foi descrito para uma piscicultura de tambaqui no Norte do Brasil, tendo em vista que um lote cultivado foi caracterizado com uma expressiva representação da diversidade genética típica de estoques selvagens (Aguiar *et al.*, 2013). Desta forma, os estudos indicam distintos cenários para a produção do tambaqui no Norte e Nordeste do Brasil, com alguns produtores de alevinos fornecendo peixes com alto risco de endogamia (baixa variabilidade genética), enquanto outros conseguem manter a diversidade genética característica de populações de *Colossoma* (Aguiar *et al.*, 2013). Estes dados demonstram a necessidade de monitorar geneticamente como estão os lotes de tambaqui em outras regiões da América do Sul, como forma de auxiliar na redução de endogamia e subsidiar práticas corretas de manejo genético.

1.3 Problemas decorrentes da endogamia

Infelizmente, existem poucos estudos de endogamia com peixes, o que faz ser o aspecto menos estudado na genética de peixes. A maioria dos estudos foi conduzida com salmonídeos, particularmente com truta arco-íris. Enquanto a endogamia é uma importante forma de manipulação genética que pode ser usada para melhorar um estoque quando bem planejado e dirigido, processos de endogamia sem planejamento e descontrolado podem arruinar um estoque com a depressão por endocruzamento, o que causa uma diminuição no crescimento e na viabilidade, acoplada com um aumento de anormalidades (Tave, 1999).

Com aumento da consanguinidade em programas de piscicultura, aumenta-se a chance de ocorrer a depressão por endocruzamento, que resulta do fato de indivíduos biologicamente aparentados terem maior chance de possuírem genes recessivos deletérios. Assim, a probabilidade de anomalias biológicas e zootécnicas aumenta quando os pais são aparentados e, quanto maior for o nível de consanguinidade, maiores serão os danos biológicos, tais como o crescimento e a sobrevivência, causando também aumento na frequência de deformidades de coluna vertebral e opérculos (Su *et al.*, 1996; Imsland *et al.*, 2001).

Desta forma, os níveis de consanguinidade devem ser mantidos abaixo dos limites críticos, sob aspecto da biologia e genética (Toledo-Filho *et al.*, 1998). Entretanto, especialmente na piscicultura brasileira, há uma tendência dos plantéis de matrizes serem formados por peixes que possuem alto grau de parentesco (Jorge, 2016; Mastrochirico-Filho, 2016). Isto se deve principalmente aos seguintes fatores: formação de lotes de reprodutores a partir de estoques já endogâmicos de outras pisciculturas; falta de controle da reprodução e desconhecimento da origem dos animais; falta de recursos para obtenção de matrizes de boa qualidade genética; utilização de baixo

número de reprodutores; carência de profissionais especializados na piscicultura; e ausência de programas de monitoramento de endogamia (Hashimoto *et al.*, 2012).

Devido aos prejuízos que podem acarretar para os produtores de peixes, há uma imediata necessidade para o monitoramento da variabilidade genética dos plantéis de matrizes. Com a informação do perfil genético, em conjunto com práticas adequadas de manejo genético, pode-se reduzir e até evitar os principais problemas resultantes do uso inadequado de animais em programas de produção (Jorge, 2016; Mastrochirico-Filho, 2016). Mesmo quando os estoques cultivados são formados por peixes selvagens, o processo de endogamia deve ser monitorado e evitado, pois geralmente uma baixa amostragem de peixes coletados não representa a variabilidade genética natural e os acasalamentos são realizados sem qualquer critério de seleção (Hashimoto *et al.*, 2012).

O outro motivo pela qual se torna imprescindível o monitoramento genético em pisciculturas é o uso das espécies cultivadas como potenciais bancos genéticos *ex-situ* para repovoamentos de populações selvagens. O objetivo destes programas de repovoamento tem sido contribuir para a conservação do potencial biológico de populações selvagens de peixes, cujo hábitat tenha sido alterado e que corram, portanto, risco de redução ou até mesmo de extinção (Toledo-Filho *et al.*, 1992; Koljonen *et al.*, 2002; Machado-Schiaffino *et al.*, 2007).

Existem três métodos mais usados para estimar os níveis de consanguinidade em piscicultura: método da genealogia, método do número de parentais e método dos marcadores genéticos. Dentre estes, os marcadores genéticos mostram-se mais vantajosos por possibilitarem cálculos relativos de probabilidade de parentesco (irmãos ou meio-irmãos) (Jorge, 2016; Mastrochirico-Filho, 2016), sem que haja necessidade de conhecimento prévio da genealogia ou do número de reprodutores efetivamente usados por geração (Toledo-Filho *et al.*, 1998).

A aplicação de tecnologias moleculares para monitoramento da endogamia requer metodologias polimórficas e com alto poder de resolução para revelar as variações genéticas entre os indivíduos (Liu e Cordes, 2004). Os SNPs têm se mostrado eficiente para esta finalidade e representam uma das principais classes de marcadores genéticos atualmente aplicados em peixes comerciais (Vera *et al.*, 2013; Mastrochirico-Filho, 2016).

1.4 Marcadores SNPs para análise de variabilidade genética

Atualmente, o uso de marcadores moleculares é uma área de pesquisa de grande importância para o setor de aquicultura, sendo aplicada em diversos programas de melhoramento genético para caracterização genética de reprodutores, identificação de híbridos, determinação de taxas de cruzamento, estimativas de diversidade e construção de mapas genéticos (De Queiroz *et al.*, 2016). Os marcadores moleculares são ferramentas poderosas capazes de detectar a singularidade de indivíduos, populações ou espécies, através da detecção de polimorfismos existentes em pequenas regiões de seus materiais genéticos. Uma variedade de tipos de marcadores tem sido utilizada em análises genéticas aplicados à aquicultura. (Liu e Cordes, 2004).

Essas técnicas moleculares são de grande importância e têm sido um instrumento eficiente utilizado na análise da diversidade genética em peixes, devido ao seu alto potencial de detecção de polimorfismos e por estar baseado no PCR (Polymerase Chain Reaction), uma técnica simples, rápida e eficiente (Jacometo *et al.*, 2010). Com o advento da tecnologia de PCR, as análises de DNA se tornaram mais acessíveis, o que causou uma revolução na biologia e a forma de estudar os seres vivos. A PCR é considerada uma ferramenta poderosa para estudos genéticos e moleculares envolvendo regiões específicas do DNA, além de permitir a análise de um elevado número de

indivíduos de qualquer organismo. Na aquicultura, o interesse pela utilização de técnicas moleculares PCR dependentes tem disponibilizado uma gama considerável de marcadores (Liu e Cordes, 2004).

Os marcadores moleculares para mapeamento de loci de características quantitativas (QTL) foram aplicados em várias espécies de peixes, resultando em uma ferramenta fundamental para programas de melhoramento por meio da implementação da seleção assistida por marcadores (MAS), este é um método que busca encontrar regiões genômicas relacionadas a variações genéticas de fenótipos importantes do ponto de vista zootécnico, ou seja, marcadores moleculares ligados a uma característica de interesse (Ariede et al., 2018).

Os mais utilizados são do tipo PCR-RFLP (*Restriction Fragment Length Polymorphism*), SPAR (*Single Primer Amplification Reaction*), AFLP (*Amplified Fragment Length Polymorphism*), microssatélites e SNP (*Single Nucleotide Polymorphism*). Dentre estes, os SNPs são considerados os marcadores mais promissores da atualidade, cujo uso tem aumentado exponencialmente nos diferentes organismos estudados, incluindo espécies não modelos (Renaut *et al.*, 2010; Seeb *et al.*, 2011).

Os marcadores SNPs são polimorfismos causados por mutações pontuais, que geram diferentes alelos de determinado nucleotídeo pertencente a um locus específico. Estas diferenças nas sequências nucleotídicas são geradas por substituições de bases e já foram bem caracterizadas desde o início do sequenciamento de DNA (Liu e Cordes, 2004). Os SNPs são uma ferramenta preponderante por características como ampla distribuição, quantidade, precisão e relativa facilidade.

Atualmente, SNP é o principal foco para o desenvolvimento de marcadores moleculares, uma vez que constitui o polimorfismo mais abundante em qualquer

organismo, adaptável a automação de genotipagem, e revelam polimorfismo oculto não detectado com outros marcadores e métodos (Liu e Cordes, 2004).

Teoricamente, um SNP de um determinado locus pode apresentar até quatro alelos (A, T, C e G). Na prática, no entanto, a maioria dos SNPs é normalmente limitada a dois alelos (frequentemente duas pirimidinas C/T ou duas purinas A/G) (Liu e Cordes, 2004; Deagle *et al.*, 2013). Uma vantagem dos marcadores SNPs, da mesma forma que os microssatélites, é o fato de apresentarem a herança do tipo co-dominância. Obviamente, o nível de polimorfismo não é tão elevado como nos marcadores microssatélites (multi alelos), mas esta desvantagem é equilibrada pela sua abundância no genoma (Liu e Cordes, 2004; Vera *et al.*, 2013). Os SNPs podem ser encontrados a cada 300 - 1000 pares de bases (pb) em mamíferos (Brouillette *et al.*, 2000), e são aparentemente ainda mais frequentes (1 SNP/100 pb) em outros vertebrados, como nos peixes (Vera *et al.*, 2013). Para ser considerado um SNP, é necessário que o alelo menos frequente apresente uma frequência de 1% ou superior (Vignal *et al.*, 2002).

Estas características demonstram que este marcador é ideal para diversos estudos biológicos, pois possibilitam análises genômicas complexas, com alto rendimento e elevada cobertura, o que tem causado uma revolução na forma de estudar os seres vivos. Os marcadores SNPs já foram utilizados para estudos comparativos de genômica evolutiva, genômica populacional, identificação de híbridos interespecíficos, identificação de sequências relacionadas com o sexo, seleção genômica, mapeamento de genes por meio de mapas de ligação e detecção de alelos associados com características economicamente importantes (Sarropoulou *et al.*, 2008; Kamiya *et al.*, 2012; Yue *et al.*, 2013; Recknagel *et al.*, 2013).

Os SNPs são os marcadores genéticos que tiveram maior relevância em diferentes áreas da biologia, embora existam diferentes tipos de marcadores

moleculares, uma de suas características e que o tornem impressionante eles podem estar diretamente relacionados a genes candidatos de características de interesse, isso difere, por exemplo, da maioria dos marcadores do tipo AFLP. Diferenças entre marcadores AFLP e microssatélites também podem ser evidenciadas com os dados gerados pelo SNP, que podem ser padronizados mais facilmente em laboratórios e diferentemente dos microssatélites, estes podem ter padrões mais complexos de mutações e os custos da automação de genotipagem pode ser mais acessível. Um dos problemas com a amplificação do locus do SNP pode ser o polimorfismo de baixa frequência ou a duplicação de genes, tornando a identificação de marcadores um processo incomum e geralmente mais trabalhoso. Além do número e automação de genótipos, o desenvolvimento de SNPs pode envolver vários processos de validação.(Renaut; Nolte; Bernatchez, 2010)

No passado, uma das maiores dificuldades para o uso rotineiro de SNPs era a caracterização e descoberta destes marcadores. Historicamente, numerosas abordagens para a descoberta de SNPs foram descritas, principalmente a partir da comparação das sequências de lócus específicos. A realização de sequenciamento direto (Sanger) de genes candidatos era considerada a estratégia mais simples, apesar de dispendiosa, para busca de SNPs. Em uma escala maior, a melhor alternativa para a caracterização de SNPs se fundamentava na comparação de sequências de fragmentos clonados, particularmente de projetos de EST (*Expressed Sequence Tags*) utilizando diferentes tipos de tecidos (Vignal *et al.*, 2002). Entretanto, além dos custos elevados, é necessário muito esforço laboratorial, tempo e expertise para este tipo de análise.

Atualmente, as tecnologias de sequenciamento de nova geração (NGS) suprimiram esta demanda e revolucionaram o cenário da Genética e Biologia Molecular, pois possibilitaram à descoberta de milhares de SNPs em qualquer organismo, inclusive

espécies não modelos (Seeb *et al.*, 2011). Para identificar um maior número de SNPs associados a genes, é necessário um maior rendimento das leituras de sequências. O Illumina/HiSeq2500, por exemplo, pode gerar até 180 bilhões de pares de bases (pb), em um rendimento de 600 milhões de sequências com cerca de 300 pb. Por esta razão, as tecnologias NGS tornaram-se muito interessante para sequenciamento *de novo* (sem referência) de genomas eucarióticos (Zhou *et al.*, 2010; Mutz *et al.*, 2013).

Uma forma de acessar o genoma para rápida prospecção e genotipagem de marcadores SNPs é por meio de estratégias de redução genômica para adquirir conjuntos de *reads* redundantes, como o caso do RAD-seq (*restriction site associated DNA sequencing*). Esta técnica foi descrita inicialmente por Baird *et al.* (2008) e permite realizar o sequenciamento e genotipagem de forma simultânea (Seeb *et al.*, 2011). O RAD-seq consiste em digerir o DNA genômico com enzima de restrição, seguido por fragmentação mecânica para reduzir os fragmentos a tamanhos adequados para o sequenciamento. Os fragmentos digeridos são então ligados aos adaptadores com *barcodes* únicos para cada indivíduo, de tal forma que podem ser *multiplexados* em um *pool*. Em seguida, as regiões adjacentes aos sítios de restrição de múltiplos indivíduos são sequenciadas simultaneamente em uma única corrida de sequenciamento. Por isso, o RAD-seq é considerado uma tecnologia de baixo custo e com fácil aplicação para espécies não-modelos (Andrews *et al.*, 2016).

Diferentes variações para o RAD-seq têm sido descritas, tais como o *Genotyping by sequencing* (GBS) (Elshire *et al.*, 2011), 2bRAD (Wang *et al.*, 2012), *Sequence-based genotyping* (SBG) (Truong *et al.*, 2012), *Reduced representation libraries* (RRLs) (Van Tassell *et al.*, 2008), *Multiplexed shotgun genotyping* (MSG) (Andolfatto *et al.*, 2012), ezRAD (Toonen *et al.*, 2013) e *Double-digest RAD* (ddRAD) (Peterson *et al.*, 2012). Em síntese, as diferenças estão relacionadas na quantidade de enzimas de

restrição, tipos de corte das enzimas, métodos de fragmentação do DNA, forma de seleção dos fragmentos digeridos, construção da biblioteca ou a plataforma de sequenciamento NGS (*next-generation sequencing*) (para revisão, Andrews *et al.*, 2016).

A estratégia RAD-seq já tem sido aplicada para várias finalidades, como filogenia e filogeografia, identificação de híbridos, genética de populações e análises de variabilidade genética (Narum *et al.*, 2013; Andrews *et al.*, 2016). A genotipagem por meio do sequenciamento RAD (DNA associado aos locais de restrição) e o ddRAD (digestão dupla RAD) são uma ferramenta rápida e acessível com custos relativamente baratos para descobrir e genotipar ao mesmo tempo uma quantidade abundante de SNPs de várias amostras e espécies. No entanto, essas metodologias devem ser otimizadas de acordo com as condições ou características, pois o número de locais encontrados depende do tamanho do genoma, estrutura e conteúdo base das espécies estudadas, bem como a extensão do polimorfismo do SNP pode variar. A metodologia ddRADseq avançou significativamente nos rendimentos de sequenciamento, bem como na análise de genótipo baseada em dados curtos de sequência de leitura, permitindo a descoberta simultânea de alto rendimento e a genotipagem de polimorfismos de sequência com ou sem genoma de referência existente. Fazendo uma comparação da metodologia RADseq com o ddRADseq, esta última nos permite maior flexibilidade e robustez na recuperação da região e menor custo do material genômico necessário nas amostras, além do tempo de trabalho exigido pelo pesquisador (Peterson *et al.*, 2012). O ddRADseq tornou-se de relevância pois esta metodologia tem sido usada com eficiência em diversos estudos genômicos de populações mostrando ser um método confiável de grande alcance em análises para determinar variabilidade genética dentro e entre as populações, estrutura, e níveis de diferenciação da população a pequena e grande

escala entre outras informações de interesse genético.

2. Objetivos

2.1 Geral

- Analisar a variabilidade genética e o grau de similaridade de estoques cultivados de tambaqui (*Colossoma macropomum*) na América do Sul.

2.2 Específicos

- Realizar a descoberta, caracterização e validação de marcadores SNPs de *C. macropomum* obtidos por sequenciamento ddRAD-seq;
- Estimar os parâmetros de diversidade genética de estoques cultivados de *C. macropomum* para programas de manejo e pré-melhoramento genético;
- Realizar análise de estrutura genética de lotes cultivados de tambaqui em diferentes regiões da América do Sul, especialmente onde há poucos estudos genéticos.

3. Material e Métodos

3.1 Material

Para realizar a análise da variabilidade genética de estoques de tambaqui na América do Sul, foram coletados 356 indivíduos de três diferentes países (Brasil, Peru e Colômbia) (Figura 1), que estão mantidos no banco de material biológico e DNA do Laboratório LaGeAC (Laboratório de Genética em Aquicultura e Conservação), Centro de Aquicultura (CAUNESP, Jaboticabal - SP) (Tabela 1). As amostras foram obtidas de 15 lotes de reprodutores de pisciculturas comerciais, além de uma população selvagem (rio Amazonas) e a população base do núcleo de melhoramento genético do CAUNESP. Para manter a integridade destas pisciculturas, uma vez que o objetivo do estudo é a

aplicação de marcadores para estudos de variabilidade, e não uma análise comercial ou prestação de serviços, os nomes e informações das pisciculturas foram preservados. Os animais foram individualmente marcados com pit-tags (*passive integrated transponder tag*, modelo *full-duplex* FDX-B, 134.2-kHz) e mantidos vivos nas instalações das pisciculturas para que seja realizado o posterior manejo genético dos reprodutores, por meio da análise de parentesco.

As coletas foram feitas com animais anestesiados (benzocaína 60 mg/l, Sigma). As amostras selvagens de tambaqui foram as mesmas utilizadas no estudo de Ariede *et al.* (2018). As amostras dos animais provenientes da Colômbia e Peru foram obtidas por meio de parcerias internacionais, com o Dr. Gustavo Adolfo Lenis Sucerquia (Facultad de Ciencias Agrarias, Universidad de Antioquia, Colômbia) e Dr. Milthon Muñoz Berrocal (Facultad de Zootecnia de la Universidad Nacional Agraria de la Selva, Peru).

Tabela 1. Localidade e número de amostras de tambaqui *Colossoma macropomum* coletadas para o estudo de variabilidade genética.

Origem	Identificação da Piscicultura	Número amostral incluído na biblioteca de ddRADseq
Brasil	BRS1	19
	BRS2	20
	BRS3	22
	BRN1	36
	BRN2	23
	BRN3	27
Peru	PER1	34
	PER2	23
	PER3	25
	PER4	8
Coômbia	COL1	20
	COL2	20
	COL3	24
	COL4	15
	COL5	12
População selvagem	WILD	8
Núcleo de melhoramento genético do CAUNESP	GEN	20

3.2 Caracterização dos locais de coleta

Para a coleta de amostras, um questionário com informações relacionadas à produção foi disponibilizado, com o objetivo de estabelecer um contexto produtivo desses locais. As informações fornecidas são tratadas de maneira geral e serão referenciadas a cada local como pontos de coleta ou grupos amostrais. Os questionários podem ser vistos com mais detalhes nos anexos um e dois, Peru e Colômbia, respectivamente.

Colombia

Na Colômbia, foram obtidos 96 amostras de 5 diferentes pontos de amostragem caracterizados por ser locais de produção de alevinos de *Colossoma macropomum*. Eles são distribuídos a partir da região amazônica colombiana, região andina e região do Caribe. Os pontos de coleta foram chamados COL1; COL2; COL3; COL4; e COL5.

Nestes 5 pontos de coleta foram obtidas informações desses locais de produção e verificou-se que elas têm entre 25 e 50 anos de operação, essas instalações são produtoras de alevinos de *C macropomum*, seus híbridos com *Piaractus brachyomus* além da produção de outras espécies nativas, como *Prochilodus magdalenae* e algumas espécies de bargres. Entre 120.000 e 10 milhões de alevinos são vendidos com distribuição local, regional e nacional, com diferentes destinos, como: engorda, pesque e pague, e programas de extensão que promovem a aquicultura. De acordo com as informações oferecidas pelos funcionários das instalações de coleta, o aparecimento de deformidades é mínimo entre 1% e 3%, geralmente apresentando nas nadadeiras, e indivíduos que não se tornam viáveis e morrem em estágios iniciais. Por outro lado, são

relatadas doenças causadas por parasitas externos como *Ichthyophthirius multifiliis*, e *Oodinium* que ocorrem em alguns locais devido à queda de temperatura ou pós-manejo, sendo descrita pouca mortalidade por essas causas.

Os lotes de reprodutores são formados por entre 20 e 100 reprodutores, e uma renovação entre 3 e 10 anos, cuja origem ocorre na maioria dos locais de produções comerciais, embora também seja indicado o uso de indivíduos silvestres para a introdução de novos indivíduos. A idade atual das matrizes varia de 5 a 20 anos e sua reprodução ocorre nos meses chuvosos, nesta estação varia em diferentes regiões e pode ser entre fevereiro a maio, junho a julho ou agosto a novembro. Para o acasalamento é utilizado a partir de uma proporção de um para um, dois para um, macho ou fêmea ou pool de reprodutores, tudo isso de acordo com as características reprodutivas dos indivíduos. Dos 5 pontos, apenas um realiza um manejo reprodutivo com lotes separados e locais diferentes para os reprodutores; o manejo genético não é realizado como tal em nenhuma das instalações.

Perú

No Peru, foram feitas coletas em 4 locais diferentes distribuídos na geografia peruana, 3 desses locais sobre a influência na região amazônica peruana e um no centro do país, estes foram nomeados como PER1; PER2; PER3; e PER4. Os pontos de coleta são produções comerciais dedicados à produção de alevinos de *Colossoma macropomum* e outras espécies nativas como *P. brachypomus* e *Arapaima gigas*. Um único local foi dedicado à produção exclusiva de Tambaqui. A operação desses locais é recente, entre 3 a 10 anos. Nestes locais produzem cerca de 500 mil alevinos por ano, e as larvas e pós-larvas são comercializadas no nível nacional, regional e local, os animais são destinados a produções de engorda e repovoamento, a taxa de deformidade é mínima e geralmente 1% é relatado nas nadadeiras; a maior mortalidade ocorre nas

larvas devido às baixas temperaturas; a mortalidade pós-manejo e a ocorrência de doenças e o uso de tratamentos não são informados.

O número efetivo de reprodutores com quem o trabalho começou nesses locais varia entre 25 e 130 reprodutores; a introdução de novos indivíduos é realizada com indivíduos selvagens, outras pisciculturas e indivíduos F1 x F1; a renovação dos reprodutores ocorre entre 2 e 9 anos, a origem das matrizes para a reposição é de indivíduos selvagens e outras pisciculturas, a idade dos reprodutores é de 4 a 9 anos. A reprodução ocorre nos meses chuvosos entre setembro e março, a proporção de acasalamentos é de 1 macho para 1 fêmea ou vários machos e várias fêmeas. Nenhum controle reprodutivo ou manejo genético foi relatado.

Brasil

No Brasil, foram obtidas amostras de 6 pontos de coleta de locais dedicados à produção de alevinos de *Colossoma macropomum*, amostras coletadas em duas regiões: 3 locais no sudeste (estado de São Paulo) BRS1, BRS2 e BRS3 e três na região norte (no estado do Tocantins) BRN1, BRN2 e BRN3. Instalações com mais de 20 anos, e além do tambaqui também produzem híbridos de *C. macropomum* com *P. brachypomus* e *P. mesopotamicus*, além de produzir outras espécies nativas. O número efetivo de reprodutores com quem o trabalho começou nesses locais varia entre 25 e 50 reprodutores; a introdução de novos indivíduos é realizada com indivíduos de outras pisciculturas e indivíduos F1 x F1.

O grupo de melhoramento genético do CAUNESP GEN foi formado por indivíduos de diferentes pisciculturas e foram formadas famílias que foram posteriormente analisadas e usadas para obter informações genéticas e desenvolver estudos de melhoramento genético.

A distribuição dos grupos de coleta pode ser observada em detalhe no mapa de coleta com seus respectivos nomes e o número amostral de cada ponto coletado (Fig. 1).

Figura 1. Mapa localidade e número de amostras de tambaqui *Colossoma macropomum* coletadas para o estudo de variabilidade genética.



3.3 Construção de bibliotecas ddRADseq

Com o objetivo de garantir que os indivíduos analisados fossem animais puros de *C macropomum*, foi realizada uma análise de pureza utilizando a técnica de PCR multiplex espécie específico, nesta análise, é realizada uma extração de material genético e, posteriormente, uma PCR com primers específicos para cada uma dessas espécies, permitindo identificar, por meio da amplificação em gel agarosa, um padrão de bandas para espécimes puros e híbridos e determinar se o material coletado pertence a esta espécie e sua pureza. Essa pré-análise foi determinada porque, na maioria das

pisiculturas na Colômbia e no Brasil, são produzidos híbridos com uma ou duas espécies da família *Serrasalminidae* pirapitinga e pacu. Nesta análise, amostras de diferentes locais foram selecionadas aleatoriamente e comparadas com controles de amostras pertencentes às outras duas espécies e híbridos.

Para prospectar e caracterizar os SNPs, foi realizada a construção das bibliotecas do ddRADseq baseado no protocolo originalmente descrito por Peterson *et al.* (2012). Inicialmente, foi extraído o DNA genômico de 356 indivíduos, utilizando o Kit *PureLink™ Genomic DNA*. Todos os indivíduos receberam um *barcode* (sequência de seis pares de bases), para que estes fossem identificados no *pool*, totalizando aproximadamente 45 marcações individuais para cada biblioteca. Ao total, foram construídas 8 bibliotecas.

As quantidades de DNA genômico foram determinadas no equipamento *Qubit® 3.0 (Thermo Scientific™)* e padronizadas para 10ng/ul. A análise de pureza das amostras de ácido nucleico foi obtida por espectrofotometria no equipamento *NanoDrop™ One (Thermo Scientific™)*, onde foi obtido um valor de razão de absorvância (A) A260/A280 entre de 1.80 e 2.00; a integridade do DNA das amostras também foi visualizada por meio do gel de agarose 1%.

Para essa etapa de digestão do DNA, foi obtido um volume final de reação de 6.0 µL, em que foram utilizados 2.5 µL de DNA (totalizando 25 ng), digeridos com 0.1 µL (10U/mL) da enzima de restrição *MluCI* (5´...AATT...3´) e 0.1 µL (20U/ml) da enzima *SphI* (5´...GCATGC...3´), 0.6 µL do *CutSmart 10X Buffer* e 2.8 µL de água. As condições estabelecidas para o tempo de digestão foram de 37°C durante 60 min.

Com base em um genoma de peixe já sequenciado, a espécie piranha *Pygocentrus nattereri*, simulações computacionais foram realizadas para avaliar a frequência do corte de cada enzima de restrição, por meio do programa *Nebcutter v2.0*

(Vincze *et al.*, 2003). Verificou-se que a distância média entre sítios enzimáticos da enzima *SphI* é de 6.500 pb, enquanto para *MluCI* é de 250 pb, o que permitiu ajustar a quantidade de adaptador específico de cada enzima, de acordo com o cálculo fornecido por Peterson *et al.* (2012). A reação de ligação foi realizada em um volume final de 10.0 µL, com 0.2 µL da enzima T4 ligase, 1.0 µL do tampão 10X *T4 DNA ligase Reaction Buffer*, 1.0 µL do adaptador P1 (4.04 µM), e 1.0 µL do adaptador P2 (0.07 µM), que tem um *barcode* específico para cada indivíduo. Este processo de ligação ocorreu com as amostras incubadas a 23°C durante 90 min e a 65°C por 10 min.

Na etapa seguinte, foi realizada uma PCR para verificar a efetividade da digestão e da ligação dos adaptadores. A PCR foi realizada em um volume final de 12.50 µL contendo 7.9 µL de H₂O, 1.25 µL de *10X PCR Buffer*, 0.5 µL do *primer* P1 (10.0 µM), 0.5 µL do *primer* P2 (10.0 µM), 0.5 µL dNTPs (10 mM), 0.75 µL MgCl₂ (25 mM), 0.1 µL *Platinum® Taq DNA Polymerase* e 1.0 µL de DNA digerido. As condições estabelecidas para a PCR foram: ciclo inicial a 94°C por 2 min, seguido por 20 ciclos de 94°C por 30 s, 60°C por 30 s e 72°C por 45 s. As amplificações foram checadas por eletroforese em gel de agarose.

Para remover os fragmentos curtos resultante da digestão do DNA genômico e da ligação dos adaptadores foi usado o *Kit Agencourt AMPure XP cleanup step* (Beckman Coulter, DE), seguindo o protocolo proposto pelo fabricante. Esta etapa também permitiu juntar todas as amostras em um *pool* de cada biblioteca, que foi utilizado para seleção dos fragmentos específicos. A seleção dos fragmentos foi de aproximadamente 450 pb por meio do *E-Gel® Size Select 2% Agarose Gel (Invitrogen)*.

Em seguida, foi realizada a etapa de enriquecimento via amplificação por PCR com a *Phusion® High-Fidelity PCR Kit*. Esta etapa é necessária para incorporar a sequência do *flowcell* Illumina e as regiões das sequências dos *primers* da Illumina.

Para aumentar o número dos fragmentos de DNA, foram realizadas 20 reações independentes de PCR. A PCR foi realizada em volume final de 20.0 µL, sendo 13.6 µL de H₂O, 0.4 µL dNTPs 10 mM, 4.0 µL 5X *Phusion HF Buffer*, 2.5 µL de primer P1 5.0 µM, 2.5 µL de *Primer P2* 5.0 µM, 0.12 µL *Phusion DNA Polymerase* e 1.0 µL de DNA digerido. As reações ocorreram em uma etapa inicial de 68°C por 1 min; seguido por 12 ciclos a 95°C por 10 s, 60°C por 30 s, 72°C por 30 s; e uma etapa final de 72°C por 7 min. As amostras de PCR foram agrupadas totalizando um volume final de 400 µL, e foram novamente purificadas por *AMPure XP beads* (Beckman Coulter, DE), de acordo com o protocolo proposto pelo fabricante. O *pool* das sequências amplificadas foi analisado em eletroforese de gel de agarose.

A concentração de cada biblioteca final foi verificada por fluorometria no equipamento *Qubit® 3.0 (Thermo Scientific™)*. As amostras foram enviadas para serem sequenciadas na plataforma Hiseq 4000, gerando 1.7 G (bilhões) de *reads paired-end*.

3.4 Identificação e filtro dos SNPs

O *software* utilizado para analisar as leituras de sequências brutas obtidas do sequenciamento *Illumina* foi o *Stacks* v. 2.1 (Catchen *et al.*, 2011; Catchen *et al.*, 2013). Este programa foi desenvolvido para analisar *reads* geradas por sequenciamento de nova geração (NGS), com dados de leitura curta para organismos modelos e não-modelos (Davey *et al.*, 2013).

Dentro do programa *Stacks* existem diversos subprogramas que devem ser executados para a análise dos dados. A análise foi iniciada com o programa *process_radtags (Stacks)*, no qual os *barcodes* foram atribuídos aos indivíduos, e foram excluídos os adaptadores e sequências de baixa qualidade. Em seguida, foi usado o *ustacks* para construir os *loci* de cada indivíduo, com uma profundidade mínima de *rad-*

tags para construir um alelo putativo (-m 6), e máxima incompatibilidade de nucleotídeos permitidas entre os alelos de uma amostra (-M 1). A montagem *de novo* de um catálogo com uma subamostragem foi realizada por meio do programa *cstacks*, com um número máximo de incompatibilidades permitidas entre as *rad-tags* dos *loci* de um (-n 1). Em seguida, através do programa *sstacks*, os *loci* putativos criados no programa *ustacks* foram mapeadas ao catálogo.

O programa *tsv2bam* transpõem os dados que estão orientados por amostra, para que sejam orientados por *loci*. O programa *gstacks* foi executado para identificar variantes e, dessa forma, genotipar cada amostra. E para finalizar a execução do *Stacks*, o programa *populations* foi utilizado, com parâmetros de *call rate* variando de 0.3 a 0.75. (Tabela 2). Além disso, o comando *write-random-snp* foi usado para gerar um SNP por *loci*, gerando mais confiança na genotipagem.

Tabela 2. Dados dos SNPs e *loci* após os filtros de qualidade, resultantes d sequenciamento de nova geração das oito bibliotecas de ddRADseq dos estoques de tambaqui *Collossoma macropomum*.

Min. <i>Call rate</i>	Número de SNPs	Número de <i>loci</i> mantido	Número de <i>loci</i> removido
0.3	42.169	65.493	494.319
0.5	14.282	21.459	538.353
0.6	5.109	7.802	552.011
0.65	2.205	3.482	556.330
0.7	364	730	559.082
0.75	30	167	559.645

O *Software Plink* foi utilizado para realizar o controle de qualidade nos SNPs identificados no *Stacks*. Foram utilizados filtros para exclusão de indivíduos que apresentaram mais do que 30% de genótipos faltantes (*mind*); e SNPs que apresentaram frequência alélica mínima (MAF) inferior a 5%. Para isso, vários valores foram testados com os parâmetros *geno*, *mind*, *min-maf*. (Tabela 4) para escolher os valores adequados à qualidade e quantidade de SNP e indivíduos filtrados.

Tabela 3. Dados dos SNPs e *loci* após os filtros de qualidade do plink, resultantes do sequenciamento de nova geração das oito bibliotecas de ddRADseq dos estoques de tambaqui *Colossoma macropomum*.

geno	mind	min-maf	SNPs	N indivíduos filtrados
0.3	0.3	0.05	3271	176
0.3	0.4	0.05	2676	200
0.3	0.5	0.05	2186	221
0.25	0.5	0.05	1440	220

3.5 Análises de genética populacional

A heterozigosidade observada (H_o) e esperada (H_e) foi estimada usando o programa CERVUS 3.0. As frequências alélicas mínimas (MAF) foram obtidas no programa GENEPOP 4.0.11, e o coeficiente de endogamia (F_{is}) foi estimado usando o *software* FSTAT 2.9.3.2. Para estimar a diferenciação genética entre os estoques, os valores globais e *pairwise* de F_{st} foram calculados usando o FSTAT versão 2.9.3.2. A significância desses valores foi estimada com 10.000 permutações.

Os níveis de mistura entre os estoques foram inferidos estimando-se o número ideal de grupos populacionais (K) usando o STRUCTURE versão 2.3.4, sem informações prévias sobre a população. Primeiramente, foi determinada a distribuição de ΔK , uma estatística *ad hoc* com base na taxa de variação na probabilidade logarítmica de dados entre valores sucessivos de K. O intervalo de clusters (K) foi predefinido de 1 a 16. A análise foi realizada em 80 execuções replicadas (isto é, 5 repetições para cada valor de K) usando 50.000 iterações após um período de queima de 20.000 execuções. O valor mais provável de K para explicar a estrutura da população foi o valor modal desse ΔK . Os resultados da análise do STRUCTURE foram visualizados através do programa STRUCTURE HARVESTER. Os resultados das execuções independentes do STRUCTURE foram resumidos e corrigidos para o melhor K usando o *software* CLUMPP versão 1.1.2.

A diferenciação genética entre as populações distribuídas em diferentes regiões geográficas foram avaliadas por análise de variância molecular (AMOVA) no *software* ARLEQUIN versão 3.5.2.2, usando 10.000 permutações.

4. Resultados

4.1 Análises de bioinformática para seleção dos SNPs

No total, foram sequenciadas cerca de 300 Gb (bilhões de bases nucleotídicas) para as oito bibliotecas construídas, resultando em média cerca de 220 M (milhões) de *reads* por biblioteca (Tabela 5). Após os filtros de qualidade das sequências, aproximadamente 180 M de *reads* por biblioteca foram mantidas para as análises de construção dos *loci* e caracterização dos SNPs. Em média, analisou cerca de 4 M de *reads* por indivíduo. A maioria das *reads* excluídas foi devido a falta de um *barcode*, resultando na exclusão de um total de 207 M *reads* (média de 25 M por biblioteca).

Tabela 4. Dados das sequencias iniciais e após os filtros de qualidade, resultantes do sequenciamento de nova geração das oito bibliotecas de ddRADseq dos estoques de tabaqui *Colossoma macropomum*.

Biblioteca	Total de sequencias	<i>Reads</i> contendo sequencias de adptador	<i>Barcode</i> não encontrado	Baixa qualidade	RAD <i>Cutsite</i> não encontrado	<i>Reads</i> retidas para análises de SNPs	Média de <i>reads</i> por indivíduo
Cm1	241.392.290	4.545.161	33.740.098	197.595	7.713.341	195.196.095	4.337.691
Cm2	244.917.924	5.131.081	21.528.426	215.399	7.913.372	210.129.646	4.568.036
Cm3	214.601.018	1.899.968	44.372.884	165.688	7.425.169	160.737.309	3.494.289
Cm4	250.460.088	4.429.306	8.256.138	237.040	8.384.081	229.153.523	4.981.598
Cm5	195.285.654	7.732.084	21.515.074	156.536	7.667.737	158.214.223	3.439.440
Cm6	242.239.138	8.171.579	19.767.432	199.826	9.905.197	204.195.104	4.439.024
Cm7	218.507.310	7.933.702	23.893.582	209.617	18.244.639	168.225.770	3.657.082
Cm8	169.623.186	3.021.253	34.633.550	122.612	14.198.143	117.647.628	2.941.191

A análise de montagem *de novo* e construção dos *loci* resultou em um catálogo referência composto de 578.766 *loci*. O mapeamento individual e posterior seleção dos

SNPs resultaram em diferentes valores de SNPs e *loci* analisados dependendo da aplicação do parâmetro de *call rate* (Tabela 3). Para as análises genômicas, foi selecionado o *call rate* de 0.65, o que resultou em 2.205 SNPs, pertencentes a 3.482 *loci* analisados. Foram excluídos 138 indivíduos por apresentarem mais do que 50% de genótipos faltantes, portanto, foram analisadas geneticamente 218 amostras ao total.

4.2 Análises genético-populacionais

Os resultados descritivos dos valores de diversidade genética por população estão reportados na (Tabela 6). De uma forma geral, os valores de heterozigosidade e MAF dos estoques comerciais de reprodutores foram similares com o da população selvagem, revelando que os índices de diversidade genética ainda não sugerem evidências de endogamia para os estoques cultivados.

Os valores de H_o variaram de 0.081 (PER1) a 0.248 (GEN); enquanto de H_e foram de 0.206 (PER1) a 0.270 (GEN) (Tabela 4). A população base do núcleo de melhoramento genético do CAUNESP (GEN) apresentou o maior índice de diversidade por ter sido composto a partir de animais oriundos de diferentes origens do Brasil, que foi uma estratégia adotada para obter um alto valor de variabilidade genética para formação das famílias. Em relação MAF, o menor valor observado (0.193) foi reportado em COL3, enquanto que o maior valor (0.267) foi detectado em PER1. Em relação ao índice genético avaliado em cada população por meio dos valores de F_{is} (coeficiente de endogamia), foram observados valores positivos em 15 dos lotes avaliados, indicando deficiência de heterozigotos nos genótipos, no grupo amostral BRN1 foi observado um valor de -0.017 indicando estar mais próximo do equilíbrio.

A estimativa de N_e pode ser observada em valores que variam de 4,2 no GEN a 126,7 no BRS1, é observado um tamanho populacional reduzido em vários locais dos diferentes grupos amostrais em comparação com o BRS1, os valores de PR1, PR2 e a

população natural WILD não puderam ser estimados, pois tendem ao infinito. A taxa de consanguinidade (ΔF) foi calculada levando em consideração os valores gerados de N_e e como resultado, podem ser observados valores mais baixos para BRS1 com 0,008 e alta para GEN com 0,238. (Tabela 7).

Tabela 5. Média dos parâmetros de diversidade genética dos estoques de tambaqui *Colossoma macropomum*. n: número amostral analisado; H_o : heterozigosidade observada; H_e : heterozigosidade esperada; MAF: frequência alélica mínima; e F_{is} : coeficiente de endogamia.

Local de amostragem	N	H_{obs}	H_{exp}	MAF	FIS
BRS1	11	0.233±0.192	0.263±0.175	0.211±0.129	0.125
BRS2	17	0.194±0.167	0.249±0.180	0.213±0.132	0.229
BRS3	17	0.199±0.185	0.236±0.186	0.219±0.135	0.176
BRN1	15	0.231±0.230	0.248±0.189	0.221±0.138	-0.017
BRN2	18	0.223±0.192	0.249±0.183	0.219±0.140	0.102
BRN3	21	0.177±0.154	0.251±0.179	0.215±0.140	0.302
PER1	8	0.081±0.149	0.206±0.226	0.267±0.145	0.642
PER2	8	0.123±0.163	0.228±0.208	0.238±0.127	0.489
PER3	13	0.163±0.152	0.242±0.181	0.209±0.135	0.335
COL1	17	0.199±0.185	0.236±0.186	0.215±0.137	0.159
COL2	16	0.200±0.195	0.236±0.195	0.236±0.132	0.157
COL3	16	0.213±0.160	0.261±0.163	0.193±0.131	0.190
COL4	12	0.203±0.205	0.236±0.202	0.246±0.133	0.146
COL5	8	0.211±0.195	0.239±0.195	0.263±0.153	0.128
WILD	7	0.187±0.181	0.262±0.189	0.223±0.125	0.313
GEN	14	0.248±0.187	0.270±0.167	0.203±0.136	0.088

Table 6. Tamanho efetivo da população (N_e) parâmetro baseado no excesso de heterozigosidade coeficiente de endogamia ($\Delta F = 1/2N_e$).

Local de amostragem	N_e	ΔF
BRS1	126.7 (96.2-184.2)	0.008
BRS2	9.6 (9.3-9.6)	0.104
BRS3	20.9 (20-21.8)	0.048
BRN1	6.8 (6.6-7)	0.147
BRN2	10.1 (9.8-10.3)	0.099
BRN3	13.7 (13.4-14.1)	0.073
PER1	Infinite	
PER2	Infinite	
PER3	19.9 (18.8-21.1)	0.050
COL1	7.6 (7.4-7.9)	0.132
COL2	18.9 (18.1-19.9)	0.053
COL3	28.7 (27.4-30.1)	0.035
COL4	10.4 (9.9-10.8)	0.096
COL5	64.0 (48.8-91.9)	0.016
WILD	Infinite	
GEN	4.2 (3.9-4.3)	0.238

Intervalo de confiança usado na análise = 95%.

Na análise de parentesco, mostra-se que todas os locais da amostragem apresentam indivíduos relacionados (irmãos completos ou meios-irmãos) em diferentes proporções, o PER2 não apresenta indivíduos relacionados e o BRS2 apresenta uma porcentagem notável de 24% de indivíduos irmãos completos. (Tabela 8); (Figura1).

Tabela 7. Análise de parentesco para os 16 locais em populações de tambaqui (*Colossoma macropomun*) de acordo com o coeficiente rxy de Wang. Os valores são representados em porcentagem.

Local de amostragem	Não relacionado	Meio irmão	Irmão completo
BRS1	96.364	3.636	0
BRS2	63.704	11.852	24.444
BRS3	89.706	8.088	2.206
BRN1	73.333	6.667	20
BRN2	84.722	8.333	6.944
BRN3	94,286	4.286	1.429
PER1	96.429	3.571	0
PER2	100	0	0
PER3	96.154	2.564	1.282
COL1	80.882	6.618	12.5
COL2	84.167	13.333	2.5
COL3	94.167	5.833	0
COL4	87.692	10.769	1.538
COL5	84.211	15.789	0
GEN	84.615	12.088	3.297

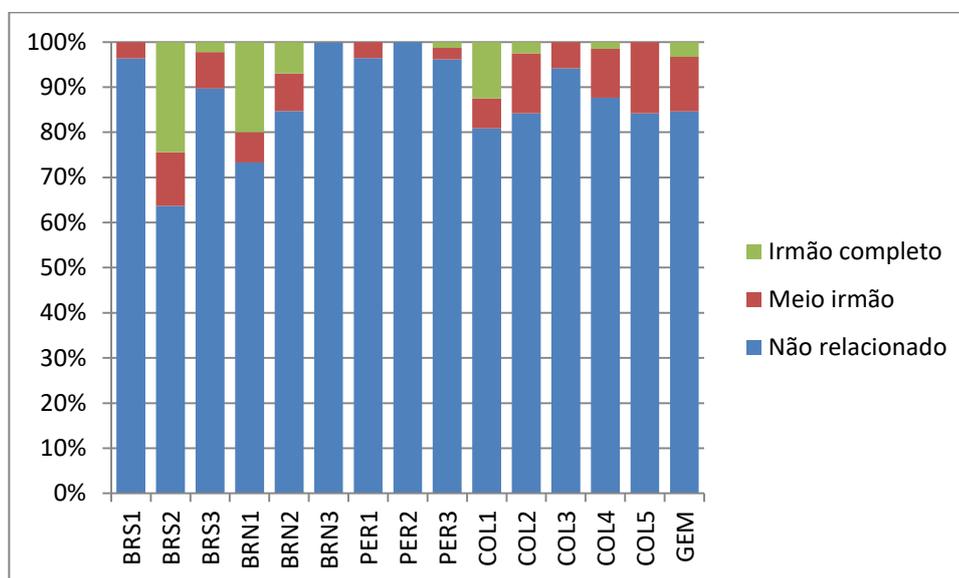


Figura 2. Distribuição dos valores de parentesco (rxy). Valores-limite para análise de parentesco: indivíduos não relacionados ($r < 0,125$), meio irmão ($0,125 \leq r \leq 0,375$) e irmãos completos ($r > 0,375$).

Os resultados da análise hierárquica de variância molecular (AMOVA) mostrou que a maior proporção de variância esteve presente dentro das populações (75,217%) (Tabela 9). Na análise de AMOVA, foi testada a hipótese de agrupamento das populações de acordo com sua procedência (grupo 1: Sul do Brasil; grupo 2: Norte; grupo 3: Peru; grupo 4: Colômbia; grupo 5: Selvagem; grupo 6: População do melhoramento), o que mostrou uma variação de somente 3,458% entre as populações dentro dos grupos e de 4,622% entre os grupos, e entre indivíduos dentro das populações 16.704%.

Tabela 8. Análise de variância molecular. (AMOVA)

Fonte de variação	Quadrados Soma dos componentes	Varição da variação	Porcentagem
Entre grupos	419.115	0.909	4.622
Entre populações dentro de grupos	362.006	0.680	3.458
Entre indivíduos dentro de populações	3363.091	3.284	16.704
Dentro de indivíduos	2566.500	14.788	75.217
Total	6710.712	19.661	

Uma significativa e baixa diferenciação genética entre todas as populações foi detectada por análises de F_{st} global sobre os *loci* de SNPs ($F_{st} = 0,05$), corroborando com dados prévios de tabaqui realizados em estoques cultivados (Araújo Ferreira et al., 2019; Gonçalves et al., 2019). Nas análises de F_{st} pairwise (Tabela 9), em geral, foi

observada diferenciação genética baixa a moderada entre todos os pares populacionais. Estes valores de F_{ST} oscilaram entre 0.001 e 0.18. A maior diferenciação genética, que ainda é considerada moderada, foi detectada entre os estoques BRS3, COL2 e COL5 ($F_{st} = 0,18$), enquanto que a menor estruturação foi evidenciada entre BRN1 e BRS1 ($F_{st} = 0,001$).

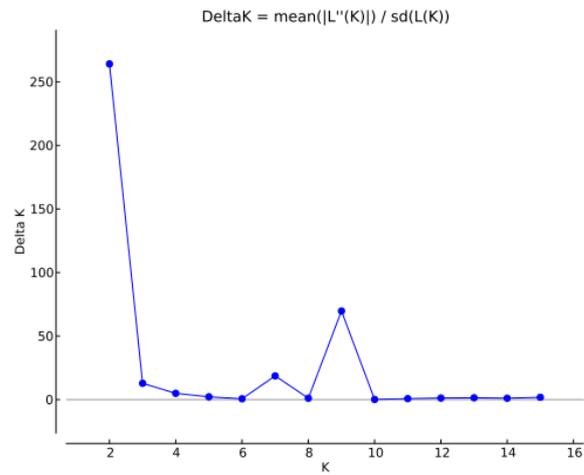
Para avaliar o nível de mistura entre as amostras, as análises de agrupamento baseadas no modelo bayesiano foram realizadas com base na distribuição de ΔK (Figura 2). Os dados indicaram que $K = 2$ é o mais adequado para explicar a estrutura populacional dos estoques cultivados de tabaqui da América do Sul, que foram avaliados no presente estudo (Figura 3). Portanto, as análises de estruturação genética identificaram dois *clusters*: 1) BRS1, BRS2, BRS3, BRN1, BRN2 e BRN3; e 2) COL1, COL2, COL4 e COL5. Os estoques PER1, PER2, PER3, COL3, WILD e a população base do programa de melhoramento genético do CAUNESP (GEN) podem ser considerados como uma mistura de ambos *clusters* genéticos.

Tabela 9. Pairwise Fst dos estoques de tabaqui *Collossoma macropomum*.

	BRS3	BRN2	BRN3	BRS2	BRN1	BRS1	GEN	WILD	COL3	COL4	COL2	COL5	COL1	PER3	PER1	PER2
BRS3	0															
BRN2	0.013	0														
BRN3	0.040	0.038	0													
BRS2	0.063	0.071	0.005	0												
BRN1	0.054	0.069	0.034	0.021	0											
BRS1	0.044	0.044	0.002	0.016	0.001	0										
GEN	0.082	0.082	0.013	0.023	0.050	0.028	0									
WILD	0.092	0.075	0.018	0.035	0.067	0.003	0.004	0								
COL3	0.079	0.089	0.022	0.033	0.034	0.012	0.026	0.041	0							
COL4	0.157	0.148	0.083	0.099	0.120	0.086	0.048	0.061	0.038	0						
COL2	0.182	0.171	0.102	0.126	0.148	0.117	0.057	0.071	0.077	0.024	0					
COL5	0.187	0.164	0.109	0.121	0.153	0.108	0.059	0.075	0.063	0.011	0.009	0				
COL1	0.136	0.145	0.053	0.073	0.096	0.064	0.043	0.068	0.013	0.030	0.048	0.066	0			
PER3	0.173	0.171	0.065	0.077	0.100	0.067	0.055	0.043	0.017	0.019	0.053	0.077	0.003	0		
PER1	0.120	0.139	0.046	0.051	0.104	0.055	0.038	0.065	0.032	0.081	0.093	0.088	0.030	0.035	0	
PER2	0.124	0.121	0.035	0.040	0.089	0.051	0.064	0.067	0.003	0.070	0.109	0.067	0.050	0.059	0.014	0

Valores significativos. Nível de significância de $P = 0.05$.

Figura3: Delta K



K

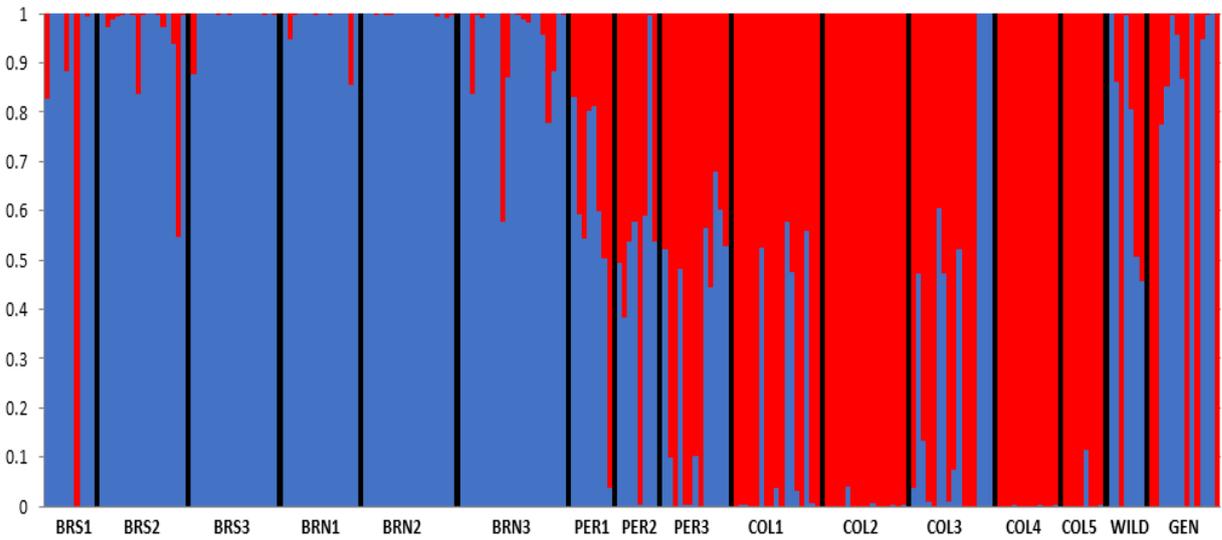


Figura 4. Análise de estruturação genética dos estoques de tambaqui *Colossoma macropomum*. Cada barra vertical representa um indivíduo. As pisciculturas são separadas por linhas brancas verticais. As proporções de cores de cada barra correspondem às frações estimadas de associação dos indivíduos de cada um dos clusters..

5. Discussão

A produção de peixes geneticamente melhorados é uma das estratégias para intensificar a piscicultura, embora atualmente menos de 10% dos peixes de criação sejam melhorados geneticamente. Esta situação é apresentada em um contexto em que o aumento do consumo exige a disponibilidade de estoques de peixes, para atender às demandas do mercado e às demandas dos consumidores (Gonçalves et al., 2019). Portanto, as pesquisas para avaliar a diversidade genética em pisciculturas são de relevância para o manejo reprodutivo e desenvolvimento de estratégias para a produção sustentável (Oliveira et al., 2019). Da mesma forma, é uma alternativa para reduzir a pressão da pesca sobre populações naturais que, como no caso da *C macropomum*, é classificada como uma espécie com sobrepesca ou risco de superexploração devido aos altos volumes de capturas. (Aguiar et al., 2013).

Em relação ao contexto produtivo, os efeitos negativos que implica a perda da variabilidade genética em populações cultivadas têm implicação em parâmetros economicamente importantes, como sobrevivência, resistência a doenças, crescimento e perda de potencial de adaptação. (Ferreira et al., 2019). Portanto, a caracterização de recursos genéticos nos programas de pré-melhoramento e melhoramento em tambaqui é uma questão prioritária a ser focada na aquicultura, pois é uma espécie de importância econômica para a piscicultura nacional e internacional.

Considerando os resultados dos parâmetros de diversidade genética estimados nos reprodutores de tambaqui, os valores de heterozigocidade (H_o e H_e) e MAF indicam baixa variabilidade genética para os grupos amostrais cultivados, os quais oscilam em H_o 0,08 a 0,23, H_e 0,20 a 0,26 e MAF 0,19 a 0,26. Esses valores não diferem muito para a população Selvagem (H_o 0,18, H_e 0,26 e MAF 0,22). Em relação ao grupo GEN (H_o 0,24, H_e 0,27 e MAF 0,20), este apresentou o maior índice de

diversidade por ter sido composto de animais de diferentes origens do Brasil. Esses valores de heteroziguidade que relacionam indivíduos cultivados e a população selvagem são semelhantes aos dados de SNPs para uma espécie do grupo de peixes da família (Serralsamidae), o pacu *Piaractus mesopotamicus* (Mastrochirico-Filho et al., 2019a). Em geral, estes valores obtidos para SNPs são menores quando comparados com marcadores do tipo microssatélites, que demonstram maiores valores de heteroziguidade devido à natureza multialélica, como demonstrado nos estudos realizados com estoques de reprodutores de tambaqui no Norte do Brasil (Araújo Ferreira, et al., 2019; Gonçalves et al., 2019).

Espera-se que as populações cultivadas tendam a revelar valores de variabilidade reduzida devido à seleção intencional e ao número reduzido de reprodutores na população base inicial (Mastrochirico-Filho et al., 2019a). Da mesma forma, fatores como domesticação, confinamento, grau de parentesco e efeitos da endogamia contribuem para reduzir a heteroziguidade e como resultado, tendem à perda da variabilidade genética. (Oliveira *et al.*, 2011).

A baixa variabilidade genética de indivíduos selvagens mostrada no presente trabalho contrasta com a alta variabilidade genética de populações selvagens em outros estudos comparativos com estoques cultivados e desenvolvidos com microssatélites. De acordo com o estudo de (Mastrochirico-Filho et al., 2019a), com a espécie *P. mesopotamicus*, foi registrado também valores semelhantes de diversidade genética entre indivíduos selvagens e cultivados por meio de SNPs. A baixa variabilidade no presente estudo mostra um cenário em que a população selvagem de tambaqui pode estar sujeita a uma redução em sua população (Santos; Ruffino; Farias, 2007). Portanto, surge a hipótese de que a espécie está sendo afetada pela sobrepesca ou por interferência humana. Entretanto, essa hipótese precisa ser corroborada com estudos mais amplos de

variabilidade na região amazônica, com maior número de indivíduos e utilizando a mesma metodologia para estabelecer dados comparativos sobre a diversidade nessa população com maiores critérios.

Os valores do F_{IS} auxiliam a entender melhor o padrão da variabilidade em algumas populações, por exemplo, PER1 e PER2 apresentaram menores valores de variabilidade e os maiores valores de F_{IS} (0,64 e 0,48, respectivamente). Esses valores indicam um déficit heterozigótico que pode estar afetando a variabilidade desses lotes e, portanto, mostrando sinais de endogamia nessas populações, ajustando-se aos dados de baixa variabilidade. Por outro lado, as populações GEN e BRN1 apresentaram maiores valores de H_o e um FIS de 0,08 e -0,017 com excesso de heterozigotos.

A baixa variabilidade e as sinais de endogamia podem resultar do efeito fundador, o que parece ser a principal causa de baixa variabilidade devido ao baixo número de reprodutores. Tomando como exemplo as populações do Peru e da Colômbia, onde estão disponíveis estimativas do número de reprodutores dessas instalações, e na maioria delas, havia estoques de matrizes que atingem até 120 indivíduos, da mesma forma, baixos valores de variabilidade e sinais de endogamia são evidenciados em alguns lotes. Isso pode ser apresentado porque não existem registros de origem dos animais, pedigree e falta de informação genética, além de manejo reprodutivo inadequado, situação que também ocorre em diferentes regiões do Brasil em produções de aquicultura.

Segundo Aguiar et al. (2018), os estoques de reprodutores no Brasil sofreram uma perda drástica da variabilidade genética e esse fato é atribuído ao efeito fundador, à deficiência de programas para o manejo da reprodução e ao tamanho efetivo das populações. Isso traz cenários preocupantes com o risco de depressão por endogamia (Aguiar et al., 2018). Do mesmo modo, Oliveira *et al.* (2011) evidenciaram que várias

populações do norte do Brasil apresentaram uma diversidade genética reduzida em comparação com indivíduos silvestres e que as causas prováveis desse efeito são o manejo reprodutivo inadequado e a origem desconhecida dos reprodutores (Oliveira *et al.*, 2011).

As estratégias a serem implementadas para recuperar os estoques das matrizes são a introdução de novos animais da população natural para introduzir um novo *pool* genético e aumentar o fluxo genético, realizar programas de manejo para seleção reprodutiva e, além disso, o controle de registros reprodutivos.

Uma ferramenta importante para reduzir as taxas de endogamia é a análise de parentesco, que nos permite a possibilidade de direcionar os acasalamentos de indivíduos não relacionados (Mastrochirico-Filho *et al.*, 2019a). Os valores observados na análise de parentesco mostram que, exceto o local PER3, todos os estoques têm indivíduos com algum grau de parentesco (meios-irmãos ou irmãos completos), o que demonstra a utilidade dessas análises para as produções não usarem indivíduos com algum grau de consanguinidade, com os riscos associados à endogamia, afetando as características produtivas e a viabilidade da progênie. Nesse caso, pode-se observar como a população BRS2 que apresentou o maior percentual de indivíduos relacionados 36,2% também apresentou alto coeficiente de endogamia, e baixo valor efetivo da população N_e 9,6 indivíduos. Esta informação é relevante porque permite monitorar os indivíduos do lote para realizar a seleção dos acasalamentos. Este tipo de ferramenta nos permite realizar um manejo reprodutivo que garante níveis adequados de variabilidade genética do estoque de reprodutores (Mastrochirico-Filho *et al.*, 2019a).

Na análise molecular AMOVA estes resultados são semelhantes com estudos de estoques de reprodutores comerciais realizados em outras espécies do mesmo grupo

de peixe (Serralsaminae), o pacu *P. mesopotamicus* (Mastrochirico-Filho et al., 2019a) e a pirapitinga *P. brachypomus* (Jorge et al., 2018).

As análises de estruturação revelaram um cenário interessante porque leva em consideração a estruturação genética de estoques cultivados de diferentes regiões da América do Sul, de forma semelhante que foi realizada com sucesso para o programa de seleção genética do pacu *P. mesopotamicus* (Mastrochirico-Filho et al., 2019b). Os resultados indicaram que há baixa diferenciação entre as populações ($F_{st} = 0,05$). A diferenciação dos pares de populações foi dada com valores estimados de baixo ($<0,05$) a moderado ($0,05 - 0,25$).

De fato, observa-se que a diferenciação entre os estoques do Brasil em relação à Colômbia apresenta valores baixos a moderados, variando de 0,01 (menor valor de diferenciação baixa) até 0,18 (valor de diferenciação moderada). Embora a diferença não seja muito acentuada, apenas três valores de baixa diferenciação são observados nas pisciculturas dois países. Esse fato é evidenciado na Figura 4, pois o Brasil e a Colômbia formam duas estruturas diferentes. De fato, os valores de menor diferenciação apresentados pelo COL3 em relação ao BRN3, BRS21, BRN1 e BRS1, e são mostrados na Figura 4 como uma mistura das duas estruturas, situação que se apresenta em menor grau com o COL1.

A diferenciação das instalações do Peru em relação ao Brasil e Colômbia mostra valores mais baixos. Embora valores moderados de diferenciação sejam observados, há uma quantidade maior de valores de baixa diferenciação ($<0,05$). Na Figura 4, o Peru é observado como uma mistura dos dois agrupamentos genéticos (Brasil e Colômbia), como também observado com as populações WILD, GEN, COL1 e COL3.

6. Conclusão

A prospecção de recursos genéticos em programas de pré-melhoramento do tambaqui é um dos temas prioritários para ser focada na aquicultura nacional, pois esta espécie tem alta importância econômica para piscicultura nacional e mundial e, atualmente, há uma crescente demanda do consumidor por produtos de melhor qualidade. A identificação de SNPs por NGS são fundamentais para conhecimento da estrutura genética de populações naturais e cultivadas, e realização de análises de diversidade genética das espécies não modelo, fornecendo melhor controle sobre o direcionamento das famílias nos acasalamentos em programas de manejo genético.

De maneira geral, os resultados deste estudo permitem conhecer um contexto de como os recursos genéticos do tambaqui são encontrados em diferentes locais da América do Sul, com o objetivo de fornecer informações úteis para a realização da prática produtiva e torná-la mais eficiente e rentável a produção, além de ajudar a reduzir a pressão da pesca sofrida por esta espécie, ajudando sua conservação.

Em conclusão, será possível orientar um programa de melhoramento genético adequado para o tambaqui, como tem sido feito para outras espécies da piscicultura mundial, como salmão, catfish, carpa e tilápia.

7. Recomendações

O presente estudo conseguiu contextualizar os recursos genéticos em diferentes locais da América do Sul para a espécie *Colossoma macropomum*, o que é de grande importância, pois, além de ser a espécie nativa mais produzida na América do Sul, é também uma espécie ameaçada pela sobrepesca e pelos danos ambientais em seu habitat natural.

Levando em consideração a relevância da produção do Tambaqui na América do Sul, os resultados deste trabalho podem contribuir promovendo estratégias para o desenvolvimento de programas de pré-melhoramento e melhoramento genético nesta espécie, e direcionar processos adequados para a seleção de indivíduos que estejam em conformidade com as características desejadas.

A validação da variabilidade genética e dos dados sobre endogamia e pureza em estoques de reprodutores é essencial na inicialização de um programa de melhoramento por seleção, garantindo que os resultados positivos no ganho genético sejam evidentes e duradouros, evitando problemas com o estreitamento da base genética do núcleo reprodutor.

Neste estudo foi possível observar sinais de endogamia e núcleos de reprodutores com algum grau de parentesco nas pisciculturas analisadas. Os dados aqui visualizados são corroborados por informações obtidas junto aos funcionários das pisciculturas, dos quais não relataram nenhum manejo genético e/ou nenhum planejamento e controle nos programas de reprodução. Ao identificar este problema, estratégias podem ser utilizadas para a recuperação desses lotes, utilizando as informações genéticas para realizar um processo reprodutivo adequado, ajudando a manter a qualidade dos reprodutores e sua progênie.

Seguindo as recomendações apropriadas, a possibilidade do lote apresentar peixes com deformidades anatômicas ou problemas fisiológicos que possam afetar o desenvolvimento normal e as características de interesse produtivo é evitada, conseqüentemente, uma melhora econômica é observada no futuro. No caso da utilização de animais com o objetivo de repovoar ambientes naturais, o uso de informações genéticas deve ser mais rigoroso, a fim de realizar um processo adequado com indivíduos de um banco genético com alta variabilidade, garantindo a viabilidade

dos indivíduos e do fluxo gênico entre as populações selvagens, sem alterar a população natural, garantindo assim um processo de repovoamento adequado.

Em programas de melhoramento genético estas informações podem ser utilizadas para direcionar cruzamentos reprodutivos com a finalidade de formar grupos ou famílias que nos permitem ter um núcleo reprodutivo geneticamente capaz, garantindo a realização de programas de melhoramento sem a incidência de gargalo genético.

Desta forma, os resultados deste estudo são de grande importância para o desenvolvimento da aquicultura, uma vez que a implementação de programas de melhoramento genético e pré-melhoramento podem contribuir para aumentar o desempenho produtivo das pisciculturas, melhorando fatores como crescimento e resistência a doenças, por exemplo, tornando a produção mais eficiente em termos de uso de recursos como água e ração, além de melhorar a qualidade dos produtos, gerando um resultado positivo nos custos de produção.

Os dados e informações genéticas obtidas, podem ser eficazes na formação de famílias ou núcleos reprodutivos, o que permitirá a implementação de programas de rastreabilidade para os animais vendidos para a formação de estoques reprodutores ou para indivíduos destinados a outros objetivos, como engorda, por exemplo. A rastreabilidade nos permite monitorar toda a cadeia produtiva, contribuindo para o controle e a qualidade dos produtos, a fim de detectar problemas em qualquer ponto do processo, assim como é feito com grande sucesso em outras espécies de importância zootécnica.

Por outro lado, é essencial que os indivíduos do núcleo reprodutivo e os animais selecionados para os programas de melhoramento genético tenham técnicas de identificação adequadas, facilitando o controle de cada indivíduo, o gerenciamento do

manejo, e o mais importante, o uso eficiente de registros reprodutivos e produtivos, bem como registros de origem e formação genealógica. Outras recomendações importantes para que a produção seja eficiente e lucrativa e os programas de melhoramento genético e de pré-melhoramento sejam viáveis é ter pessoal especializado para trabalhar nas instalações, ter um número adequado de matrizes (50 -100), controle de programas reprodutivos e programas de monitoramento genético.

Finalmente com base nos resultados apresentados, é feito um apelo às autoridades competentes para que adotem ações com perspectivas para melhorar a aquicultura do tambaqui e, em geral, das espécies nativas, além de fortalecer pesquisas e estudos nas diferentes áreas em que podem contribuir para o desenvolvimento produtivo e conservação de as espécies. As informações contidas neste estudo são de grande importância para o Péru e Colômbia, pois poucos estudos são reportados sobre este assunto nesses locais. No Brasil estes dados contribuíram para a construção de um maior conhecimento sobre o tambaqui, permitindo o desenvolvimento de estratégias para que haja avanços tecnológicos na produção, progresso na aquicultura de espécies nativas, o que resultará em um crescimento econômico do setor aquícola e auxiliará na preservação de populações selvagens ameaçadas pela ação humana.

8. Anexos

Anexo 1. Questionário aos piscicultores de 4 instalações de produção de *Colossoma macropomum* no Peru. PER1; PER2; PER3; PER4.

Perguntas	PERU			
	PER1	PER2	PER3	PER4
1. Quando a piscicultura foi fundada	Sem informação	2017	2011	2003
2. Quais as espécies produzidas	Sem informação	Cachama Negra, (<i>Colossoma Macropomum</i>)	Cachama branca (<i>Piaractus Brachyopomum</i>); Cachama Negra, Tambaqui (<i>Colossoma Macropomum</i>);	Cachama Negra, (<i>Colossoma Macropomum</i>); Cachama branca (<i>Piaractus Brachyopomum</i>); Paiche (<i>Arapaima gigas</i>)
3. Existe a formação de híbridos? Se sim quais?	Sem informação	(X) Não	(X) Não	(X) Não
4. Qual o volume de alevinos vendidos?	Sem informação	200.000 Alevinos e 500.000 post-larva	500.000 Alevinos	120 milhões de alevinos
5. Forma de comercialização	Sem informação	(X) post-larva (X) Alevinos	(X) Alevinos e post-larva	(X) Alevinos
6. Destino dos alevinos:	Sem informação	(X) Local	(X) Nacional	(X) Regional
7. Finalidade da produção	Sem informação	(X) Engorda (X) Repovoamento	(X) Engorda	(X) Engorda
8. Utilizam algum canal de comercialização? Se sim qual?	Sem informação	(X) Não	(X) Não	(X) Não
9. Existem indivíduos com deformidades morfológicas? Se sim quais?	Sem informação	(X) Não	(X) Não	(X) Sim menos de 1%, falta de aletas
10. Existe mortalidade frequente após manejo?	Sem informação	(X) Não	(X) Não	(X) Não
11. Existe mortalidade frequente em baixas temperaturas	Sem informação	(X) Não	(X) Não	(X) Não
12. Quais as doenças mais observadas no lote? Existe a utilização de algum tratamento?	Sem informação	Ninguno	Sim larvas y pos-larvas	Não
13. Qual o número efetivo de indivíduos (pacu) utilizados no início da formação da piscicultura?	Sem informação	Não	Não	Não
14. Como é realizada a introdução de novos indivíduos:	Sem informação	25 indivíduos	130	60 (30 machos y 30 fêmeas)
15. A renovação do plantel ocorre de quanto em quanto tempo?	Sem informação	(X) Indivíduos selvagens (X) Outras pisciculturas	(X) F1 e pós F1	(X) Indivíduos Selvagens, (X) Outras piscicultura
16. Qual o período de reprodução?	Sem informação	entre 7 a 9 anos	Cada 2 años	7 anos
17. Qual a origem das matrizes?	Sem informação	Estação chuvosa Iquitos y San Pablo de Loreto (provincia de Mariscal Ramón Castilla), piscicultura y medio natural, respectivamente	Septiembre a Marzo	Septiembre a Marzo
18. Qual a idade dos reprodutores?	Sem informação	3 a 9 Anos de idade	4 Anos de idade	4 Anos de idade
19. Como ocorre a reprodução? Número de fêmeas e machos utilizados para a formação dos alevinos	Sem informação	(X) Uma fêmeas y um macho.	1 fêmea e 1 macho	(X) Várias fêmeas e vários machos
20. Utilizam melhoramentos genéticos na piscicultura?	Sem informação	Não	Não	Não

9. Referências

- Araújo-Lima CARM, Ruffino ML (2003) Migratory fishes of the Brazilian Amazon. In: Migratory fishes of South America (eds Carolsfeld J, Harvey B, Ross C, Baer A). World Fisheries Trust/The World Bank/International Development Research Centre, Ottawa.
- Aguiar, J. da P.; Gomes, P. F. F.; Hamoy, I. G.; Santos, S. E. B. Dos; 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*, [s. l.], v. 49, n. 5, p. 1914–1925,
- Aguiar J, Schneider H, Gomes F, *et al.* (2013) Genetic variation in native and farmed populations of Tambaqui (*Colossoma macropomum*) in the Brazilian Amazon: Regional discrepancies in farming systems. *An Acad Bras Ciênc.* **85**: 1439–1447.
- Andolfatto P, Davison D, Erezyilmaz D, *et al.* (2011) Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Res* **21**: 610–617.
- Andrews KR, Good JM, Miller MR, *et al.* (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* **17**: 81-92.
- Ariede, R. B.; Freitas, M. V.; Hata, M. E.; Mastrochirico-Filho, V. A.; Pilarski, F.; Batlouni, S. R.; Porto-Foresti, F.; Hashimoto, D. T. (2018); Microsatellites associated with growth performance and analysis of resistance to *Aeromonas hydrophila* in tambaqui *Colossoma macropomum*. *Frontiers in Genetics*, [s. l.], v. 9, n. JAN, p. 1–8, a.
- Ariede, R. B.; Freitas, M. V.; Hata, M. E.; Matrochirico-filho, V. A.; Utsunomia, R.; Mendonça, F. F.; ForestI, F.; Porto-foresti, F.; Hashimoto, D. T. (2018); Development of microsatellite markers using next-generation sequencing for the fish *Colossoma macropomum*. *Molecular Biology Reports*, [s. l.], v. 45, n. 1, p. 9–18, 2018. b. Disponível em: <http://dx.doi.org/10.1007/s11033-017-4134-z>
- Baca, L. C. De. El cultivo de la gamitana en latinoamérica. [s.l: s.n.].
- Bacher K (2015) Perceptions and Misconceptions of Aquaculture: A Global Overview. *Globefish Research Programme*, vol.120; FAO: Rome, Italy, 35 pp.
- Baird NA, Etter PD, Atwood TS, *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**: e3376.
- Brander, KM (2015) Global fish production and climate change.
- Calcagnotto, D.; De Almeida Toledo-Filho, S. (2000). Loss of genetic variability at the transferrin locus in five hatchery stocks of tambaqui (*Colossoma macropomum*). *Genetics and Molecular Biology*, [s. l.], v. 23, n. 1, p. 127–130,

- Caribe, A. L. Y. Potencial Acuicola General De America Latina. [s. l.], n. Sofia 2010, 2011.
- Catchen JM, Amores A, Hohenlohe P, *et al.* (2011) Stacks: building and genotyping loci *de novo* from short-read sequences. *G3* **1**: 171-182.
- De Queiroz, C. A.; Sousa, N. R.; DA Silva, G. F.; Inoue, L. A. K. A. (2016). Impacts of stocking on the genetic diversity of *Colossoma macropomum* in central Amazon, Brazil. *Genetics and Molecular Research*, [s. l.], v. 15, n. 2, p. 1–9,
- Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol. Bioinform.* **1**: 47-50.
- Estado Mundial, E. Fao 2018 De La Pesca Y La Acuicultura. [s.l: s.n.]. Organization of the United Nations – FAO Fisheries and Aquaculture Department, Roma.
- Evangelista-Barreto, N. S. (2019). Genetic diversity in natural populations of *Colossomamacropomum* in the Brazilian Amazon region and in populations farmed in Northeast Brazil based on ISSR markers. *Aquaculture International*, [s. l.], p. 1423–1434,
- FAO. Field guide to the culture of tambaqui (*Colossoma macropomum*. [s.l: s.n.].
- FAO - Fisheries and Aquaculture Information and Statistics Branch - 16/10/2019.
- Farias IP, Torrico JP, García-Dávila C, *et al.* (2010) Are rapids a barrier for floodplain fishes of the Amazon basin? A demographic study of the keystone floodplain species *Colossoma macropomum* (Teleostei: Characiformes). *Mol. Phylogenet. Evol.* **56**: 1129-1135.
- Fazzi-Gomes, P.; Guerreiro, S.; Palheta, G. D. A.; De Melo, N. F. A. C.; Santos, S.; Hamoy, I. High. (2017) Genetic diversity and connectivity in *Colossoma macropomum* in the Amazon basin revealed by microsatellite markers. *Genetics and Molecular Biology*, [s. l.], v. 40, n. 1, p. 142–146.
- Ferreira, L. de 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**, [s. l.], v. 48, n. 0
- Gjedren, T (1997) Selective breeding to improve aquaculture production. *World Aquaculture-Baton Rouge* **28**: 33-46.
- Gomes LC, Simões LN, Araujo-Lima CARM (2010) Tambaqui (*Colossoma macropomum*). In: *Espécies nativas para a piscicultura no Brasil* (eds Baldisserotto B, Gomes LC). UFSM, Santa Maria.
- Gomes, F, Schneider, H, Barros, C, *et al.* (2012) Innovative molecular approach to the identification of *Colossoma macropomum* and its hybrids. *An Acad Bras Ciênc.* **84(2)**: 517-526.
- Hardie DC, Hebert PDN (2003) The nucleotypic effects of cellular DNA content in cartilaginous and ray-finned fishes. *Genome* **46**: 683-706.
- Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* **2**: 618-620.

- Hashimoto DT, Senhorini JA, Foresti F, *et al.* (2012) Interspecific fish hybrids in Brazil: management of genetic resources for sustainable use. *Rev. Aquaculture* **4**: 108-118.
- Hashimoto DT, Alves AL, Varela ES, *et al.* (2012) Genética na Piscicultura: Importância da variabilidade genética, marcação e coleta para análise de DNA. 1. ed. Brasília: Embrapa, 32p.
- Hashimoto DT, Senhorini JA, Foresti F, *et al.* (2014) Genetic identification of F1 and post-F1 Serrasalmid juvenile hybrids in Brazilian aquaculture. *Plos One* **9**:e89902.
- Hiltsdorf, AWS, Orfão, LH (2011) Aspectos gerais do melhoramento genético no Brasil. *Revista Brasileira de Zootecnia*, v.40, p. 317-324.
- IBGE (2016) Produção da Pecuária Municipal 2015. Rio de Janeiro, 43, 1-47.
- Imslund AK, Foss A, Naevdal G, *et al.* (2001) Selection or adaptation: differences in growth performance of juvenile turbot (*Scophthalmus maximus* Rafinesque) from two close-by localities o Norway. *Sarsia* **86**: 43-51.
- Jacometo, C. B.; Barrero, N. M. L.; Rodriguez-Rodriguez, M. D. P.; Gomes, P. C.; Povh, J. A.; Junior, D. P. S.; Vargas, L.; De Resende, E. K.; Ribeiro, R. P. (2010). Variabilidade genética em tambaquis (Teleostei: Characidae) de diferentes regiões do Brasil. *Pesquisa Agropecuaria Brasileira*, [s. l.], v. 45, n. 5, p. 481–487,
- Jégu, M (2003) Subfamily Serrasalminae (Pacus and piranhas). In: Check list of the freshwater fishes of South and Central America (eds Reis RE, Kullander SO, Ferraris CJ). Edipucrs, Porto Alegre.
- Jorge PH (2016) Sequenciamento do transcriptoma e caracterização de microssatélites na pirapitinga *Piaractus brachypomus* para análises de variabilidade genética. Dissertação de Mestrado. Instituto de Biociências de Botucatu, Unesp.
- Koljonen ML, Tahtinen J, Saisa M, *et al.* (2002) Maintenance of genetic diversity of Atlantic salmon (*Salmo salar*) by captive breeding programmes and the geographic distribution of microsatellite variation. *Aquaculture* **212**: 69-92.
- Liu ZJ, Cordes JF (2004) DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* **238**: 1-37.
- Lopes, T. S.; Streit Júnior, D. P.; Ribeiro, R. P.; Povh, J. A.; Lopera-Barrero, N. M.; Vargas, L.; Pinto Filho, C.; Queiroz, J. R. (2009). Diversidade genética de estoques de reprodutores de. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia*, [s. l.], v. 61, n. 3, p. 728–735,
- Machado-Schiaffino G, Dopico E, Garcia-Vazquez E (2007) Genetic variation losses in Atlantic salmon stocks created for supportive breeding. *Aquaculture* **264**: 59-65.
- Mastrochirico-Filho, 2016 Análise de parentesco e variabilidade genética de pacu (*Piaractus mesopotamicus*) por meio de marcadores SNPs: subsídios para o melhoramento genético. Dissertação de Mestrado. Centro de Aquicultura da Unesp.

- 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, et al. (2019) Assessing Genetic Diversity for a Pre-Breeding Program in *Piaractus mesopotamicus* by SNPs and SSRs. *Genes*, [s. l.], v. 10, n. 9, p. 668.
- Mastrochirico-Filho, V.A; Ariede, R. B.; Freitas, M, V.; Lira, L. V.G.; Agudelo. J. F.G.; Pilarski, F.; Reis Neto, R. V.; Yañez, J, M.; Hashimoto, D. T. (2019) Genetic parameters for resistance to *Aeromonas hydrophila* in the Neotropical fish pacu (*Piaractus mesopotamicus*). **Aquaculture**, [s. l.], v. 513, n. June, p. 734442, Disponível em: <https://doi.org/10.1016/j.aquaculture.2019.734442>
- Merino, M.; Bonilla, S.; Bages, F. Diagnóstico del estado de la Acuicultura en Colombia. Autoridad Nacional de Pesca y Acuicultura – AUNAP. [s.l: s.n.].
- MMA. **Caderno da Região Hidrográfica Amazônica**. [s.l: s.n.]. Disponível em: <http://www.mma.gov.br/estruturas/161/_publicacao/161_publicacao03032011024106.pdf>
- MPA (2010) Produção pesqueira e aquícola – Estatística 2008 e 2009. Ministério da Pesca e Aquicultura.
- MPA (2012) Boletim estatístico da pesca e aquicultura – Brasil 2010. Ministério da Pesca e Aquicultura.
- Mutz KO, Heilkenbrinker A, Lönne M, et al. (2013) Transcriptome analysis using next-generation sequencing. *Curr. Opin. Biotechnol.* **24**: 22-30.
- Narum SR, Buerkle CA, Davey JW, et al. (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Mol Ecol* **22**: 2841-2847.
- Oliveira, C. de S. T.; Moreira, R. F. C.; Soares Filho, A. A.; Fonteles, S. B. A.; Evangelista-Barreto, N. S. (2019). Genetic diversity in natural populations of *Colossomacropomum* in the Brazilian Amazon region and in populations farmed in Northeast Brazil based on ISSR markers. *Aquaculture International*, [s. l.], p. 1423–1434,
- Oliveira ACB, Miranda E, Correa R (2012) Exigências nutricionais e alimentação do tambaqui. In: Nutriaqua: nutrição e alimentação de espécies de interesse para a aquicultura brasileira (eds Fracalossi DM, Cyrino JEP). Sociedade Brasileira de Aquicultura e Biologia Aquática, Florianópolis.
- PEIXEBR. Anunário PeixeBr da Piscicultura. (2019). **Associação Brasileira de Piscicultura**, [s. l.], p. 138.
- Peterson BK, Weber JN, Kay EH, et al. (2012) Double digest RADseq: an inexpensive method for de novo snp discovery and genotyping in model and non-model species. *PLoS One* **7**: e37135.
- Pineda Santis, H.; Restrepo Betancur, L.; Olivera Ángel, M. (2004) Comparación morfológica entre machos y hembras de Cachama Negra (*Colossoma macropomum*, Cuvier 1818) mantenidos en estanque. *Revista Colombiana de Ciencias Pecuarias*, [s. l.], v. 17, n. 4, p. 24–29.

- Ponzoni, RW (2006) Genetic improvement effective dissemination: Keys to prosperous and sustainable aquaculture industries. WorldFish Center, Penang.
- Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Heredity* **86**: 248-249.
- Recknagel H, Elmer KR, Meyer A (2013) A hybrid genetic linkage map of two ecologically and morphologically divergent midas cichlid fishes (*Amphilophus* spp.) obtained by massively parallel DNA sequencing (ddRADSeq). *G3* **3**: 65-74.
- Regalado, F. (2017) Potencial acuícola en el Perú. [s. l.], p. 34–39.
- Renaut S, Nolte AW, Bernatchez L (2010) Mining transcriptoma sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Mol. Ecol.* **19**: 115-131.
- Santos MCF, Ruffino ML, Farias IP (2007) High levels of genetic variability and panmixia of the tambaqui *Colossoma macropomum* (Cuvier, 1816) in the main channel of the Amazon River. *J. Fish Biol.* **71**: 33-44.
- Santos, CHA, Santana, GX, Sá Leitão, CS, *et al.* (2016) Loss of genetic diversity in farmed populations of *Colossoma macropomum* estimated by microsatellites. *Anim. Genet.* **47**: 373-376.
- Sarropoulou E, Nousdili D, Magoulas AGK (2008) Linking the genomes of nonmodel teleosts through comparative genomics. *Mar. Biotechnol.* **10**: 227-233.
- Seeb JE, Carvalho G, Hauser L, *et al.* (2011) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Mol. Ecol. Res.* **11**: 1-8.
- Su GS, Liljedahl LE, Gall GAE (1996) Effects of inbreeding on growth and reproductive traits in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* **142**: 139-148.
- Suplicy FM (2007) Freshwater fish seed resources in Brazil. FAO Fisheries Technical Paper. No. 501. FAO, Rome.
- Tave, D (1995) Selective breeding programmes or medium-sized fish farms. FAO Fisheries Technical Paper. N°352, FAO, Roma.
- Tave D (1999) Inbreeding and brood stock management. *Fisheries Technical Paper. n° 392*. Roma, FAO.
- Toledo-Filho SA, Almeida-Toledo LF, Foresti F, *et al.* (1998) Programas genéticos de seleção, hibridação e endocruzamento aplicados à piscicultura. Cadernos de Ictiogenética 4, CCS/USP, São Paulo. **Molecular Ecology Resources**, [s. l.], v. 11, n. SUPPL. 1, p. 1–8, 2011.
- Torati, L. S.; Taggart, J. B.; Varela, E. S.; Araripe, J.; Wehner, S.; Migaud, H. Genetic diversity and structure in *Arapaima gigas* populations from Amazon and Araguaia-Tocantins river basins. *BMC Genetics*, [s. l.], v. 20, n. 1, p. 1–13, 2019.

- Valladão, G. M. R.; Gallani, S. U.; Pilarski, F. (2018) South American fish for continental aquaculture. *Reviews in Aquaculture*, [s. l.], v. 10, n. 2, p. 351–369.
- Valladão GMR, Gallani SU, Pilarski F (2016) South American fish for continental aquaculture. *Rev Aquacult.* doi:10.1111/raq.12164
- Vera M, Alvarez-Dios JA, Fernandez C, *et al.* (2013) Development and validation of single nucleotide polymorphisms (SNPs) markers from two transcriptome 454-runs of turbot (*Scophthalmus maximus*) using high-throughput genotyping. *Int. J. Mol. Sci.* **14**: 5694-711.
- . The genetical structure of populations. *Ann. Hum. Genet.* 1951, 15(1), 323-354.
- Vignal A, Milan D, SanCristobal M, *et al.* (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genet. Sel. Evol.* **34**: 275-305.
- Yue GH (2013) Recent advances of genome mapping and marker-assisted selection in aquaculture. *Fish Fish.* doi: 10.1111/faf.12020
- Zhou X, Ren L, Meng Q, *et al.* (2010) The next-generation sequencing technology and application. *Protein Cell* 1: 520-536.