

UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CENTRO DE AQUICULTURA DA UNESP

**ANÁLISES GENÔMICAS PARA ASSOCIAÇÃO DE
*SNPS COM RESISTÊNCIA À AEROMONAS
HYDROPHILA E RENDIMENTO DE FILÉ NO
TAMBAQUI (COLOSSOMA MACROPOMUM)***

Raquel Belini Ariede

Bióloga

Jaboticabal, São Paulo

2021

UNIVERSIDADE ESTADUAL PAULISTA - UNESP
CENTRO DE AQUICULTURA DA UNESP

**ANÁLISES GENÔMICAS PARA ASSOCIAÇÃO DE
HYDROPHILA E RENDIMENTO DE FILÉ NO
HYDROPHILA E RENDIMENTO DE FILÉ NO
TAMBAQUI (*COLOSSOMA MACROPOMUM*)**

Raquel Belini Ariede

Orientador: Dr. Diogo Teruo Hashimoto

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

Jaboticabal, São Paulo

2021

A698a Ariede, Raquel Belini
Análises genômicas para associação de SNPs com resistência à *Aeromonas hydrophila* e rendimento de filé no tambaqui (*Colossoma macropomum*) / Raquel Belini Ariede. -- Jaboticabal, 2021
iv, 129 p. : il. ; 29 cm

Tese (doutorado) - Universidade Estadual Paulista, Centro de Aquicultura, 2021

Orientador: Diogo Teruo Hashimoto

Banca examinadora: Fabiana Pilarski, Fabio Porto-Foresti, Gabriela Vanina Villanova, Gustavo Moraes Ramos Valladão

Bibliografia

1. Aquicultura. 2. Resistência à doenças. 3. Genética Animal. 4. Peixes.
I. Título. II. Jaboticabal-Centro de Aquicultura.

CDU 636.3.09

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



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Centro de Aqüicultura da Unesp - CAUNESP



CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: Análises genômicas para associação de SNPs com resistência à *Aeromonas hydrophila* e rendimento de filé no tambaqui (*Colossoma macropomum*)

AUTORA: RAQUEL BELINI ARIEDE

ORIENTADOR: DIOGO TERUO HASHIMOTO

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

Prof. Dr. DIOGO TERUO HASHIMOTO (Participação Virtual)
Laboratório de Genética de Peixes, Centro de Aqüicultura da Unesp, Caunesp -, Jaboticabal-SP

Prof. Dr. FABIO PORTO FORESTI (Participação Virtual)
Departamento de Ciências Biológicas / FC/UNESP, Bauru-SP

Profa. Dra. FABIANA PILARSKI (Participação Virtual)
Laboratório de Microbiologia e Parasitologia de Organismos Aquáticos / Centro de Aqüicultura da Unesp, Caunesp, Jaboticabal-SP

Prof. Dr. GUSTAVO MORAES RAMOS VALLADÃO (Participação Virtual)
Programa de Pós Graduação em Aqüicultura / Universidade Nilton Lins, Manaus-AM

Profa. Dra. GABRIELA VANINA VILLANOVA (Participação Virtual)
Consejo Nacional De Investigaciones Cientificas Y Técnicas, CONICET, Argentina

Jaboticabal, 20 de dezembro de 2021

Sumário	
Resumo	4
Abstract	5
1. Introdução Geral	6
1.1 Produção aquícola e caracterização do tambaqui <i>Colossoma macropomum</i>	6
1.2 Melhoramento genético em peixes	7
1.3 Características de crescimento e rendimento de filé	8
1.4 Resistência a doenças	9
1.5 Genômica aplicada ao melhoramento genético de peixes	11
1.6 Referências	12
2. Objetivos	21
2.1 Objetivo geral	21
2.2 Objetivos específicos	21
3. Atividades desenvolvidas	22
3.1 Capítulo I	24
3.2 Capítulo II	26
3.2.1 Material biológico	26
3.2.2 Construção das bibliotecas e seleção dos SNPs	26
3.2.3 Validação do Axiom array em tambaqui	30
Anexo I	35
Abstract	35
Introduction	36
Results	37
Discussion	39
Methods	41
Data Availability	44
References	44
Acknowledgements	48
Author contributions	48
Additional Information	48
Code Availability	48
Competing interests	49
Figures	49
Tables	49
3.3 Capítulo III	55
3.3.1 Introdução	57

3.3.2	Material e Métodos	59
3.3.3	Resultados	63
3.3.4	Discussão	70
3.3.5	Conclusão	74
3.3.6	Referências	75
3.4	Capítulo IV	85
3.4.1	Introdução	86
3.4.2	Material e Métodos	88
3.4.2.1	Comitê de Ética	88
3.4.2.2	População experimental	89
3.4.2.3	Experimento do desempenho de crescimento e rendimento de filé	89
3.4.2.4	Sistema de visão computacional e Deep Learning	92
3.4.2.5	Análises genômicas	94
3.4.2.6	Modelos estatísticos	96
3.4.2.7	Análise de associação genômica ampla (GWAS) para rendimento de filé	98
3.4.3	Resultados	100
3.4.3.1	Estatísticas descritivas	100
3.4.3.2	Parâmetros genéticos	106
3.4.3.3	Correlações fenotípicas e genéticas	108
3.4.3.4	Análises genômicas	111
3.4.4	Discussão	117
3.4.5	Referências	122

Dedicatória

*Ao meu avô Rodolfo Bellini,
minha maior saudade.*

Agradecimentos

Agradeço a Deus primeiramente por todo o cuidado comigo e com as pessoas que eu amo durante estes anos. Agradeço também cada pessoa que Ele colocou em meu caminho para tornar a caminhada mais leve.

Agradeço meus pais, irmãos, cunhados e sobrinhos que sempre foram meu porto seguro e minha alegria. Minha motivação vem deles.

Sou grata pelo meu namorado e sua família que me acolheram e foram minha segunda casa. Obrigada pelo cuidado e apoio sempre.

Ao prof. e orientador Diogo Hashimoto, pelo direcionamento e atenção desde o mestrado, foram seis anos de muito aprendizado e evolução. Obrigada por estar ao meu lado me ajudando.

A todos do laboratório LaGeAC e todos que passaram por ele durante todo este período, Milena, Rubens, John, Carol, Vito, Valéria, Celma, Marcelo e Shisley meus sinceros agradecimentos por toda ajuda, sempre acreditei que ninguém faz nada sozinho e se eu consegui chegar até aqui foi porque vocês estiveram comigo. Obrigada!

Minhas amigas Ariadne, Bruna e Yasmim, obrigada por cada conversa e momentos juntas, vocês me deram força e são mulheres incríveis que me inspiram a ser melhor a cada dia.

Obrigada a todos vocês!

Apoio Financeiro

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e a Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP (número do processo 2017/19717-3) pela Bolsa de Doutorado, e a FAPESP pela Bolsa Estágio de Pesquisa no Exterior - BEPE (número do processo 2019/10662-7).

Resumo

Tambaqui (*Colossoma macropomum*) é o peixe neotropical mais importante da aquicultura de água doce na América do Sul. Surto da doença causados pela bactéria *Aeromonas hydrophila* resultam em perdas significativas para a produção de tambaqui. Além disso, o aumento da produção dessa espécie em aquicultura é limitado pela falta de programas de melhoramento genético, especialmente para características de resistência a doenças e rendimento de filé. O objetivo desse estudo foi caracterizar o desempenho zootécnico de famílias de tambaqui *C. macropomum* para resistência à *Aeromonas hydrophila* e características de crescimento/rendimento de filé, por meio de genética quantitativa e associação genômica de SNPs (*single nucleotide polymorphisms*). Primeiramente, os resultados dos parâmetros genéticos para resistência a *A. hydrophila* e a correlação genética com o ganho de peso diário em juvenis de tambaqui demonstraram que é possível incluir essas características em programas de melhoramento seletivo. Em seguida, para avançar com as análises genômicas, foi desenvolvido um denso SNP *array* para o tambaqui, com a caracterização de 30 mil SNPs candidatos para a plataforma de genotipagem Axiom/ThermoFischer (30K SerraSNP *array*). Após essa etapa, foi investigado a arquitetura genética da resistência a *A. hydrophila* em tambaqui por um estudo de associação do genoma (GWAS) para identificar loci de características quantitativas (QTLs) e genes putativos associados a esta característica em tambaqui. Um denso mapa de ligação foi desenvolvido para o tambaqui usando os dados do genótipo; resultando em 17.374 SNPs distribuídos em 27 grupos de ligação. A análise GWAS identificou vários QTLs candidatos associados à resistência de *A. hydrophila*, distribuídos em seis grupos de ligação. Vários genes candidatos relacionados à resposta imune foram localizados próximos aos QTLs putativos. Em relação as características de crescimento/rendimento, as análises de genética quantitativa revelaram moderada-alta herdabilidades, o que demonstra que estão sob moderado controle genético e devem responder à seleção genética. Uma abordagem de *deep learning* e sistema de visão computacional (CVS) também foi validada para obter as medidas morfométricas do tambaqui em larga escala e de forma automatizada. As análises de GWAS para essas características confirmaram que estão sob controle poligênico, com a identificação de apenas alguns QTLs com significância estatística. Em conclusão, ambas as características avaliadas neste estudo, resistência a *A. hydrophila* e características de crescimento e rendimento corpóreo, demonstram estar sob controle genético e influenciadas por várias regiões genômicas, e podem ser inseridas em programas de melhoramento para tambaqui.

Palavras-Chave: tambaqui, SerraSNP array, desafio bacteriano, crescimento, SNPs associados.

Abstract

Tambaqui (*Colossoma macropomum*) is the most important neotropical fish in freshwater aquaculture in South America. Disease outbreaks caused by the bacteria *Aeromonas hydrophila* result in significant losses for tambaqui production. Furthermore, the increase in production of this species in aquaculture is limited by the lack of genetic improvement programs, especially for disease resistance traits and fillet yield. The objective of this study was to characterize the performance of tambaqui *C. macropomum* families for resistance to *Aeromonas hydrophila* and fillet growth/yield traits, through quantitative genetics and genomic association of SNPs (single nucleotide polymorphisms). First, the results of genetic parameters for resistance to *A. hydrophila* and the genetic correlation with daily weight gain in tambaqui juveniles demonstrated that it is possible to include these traits in selective breeding programs. Then, to proceed with genomic analyses, a dense SNP array for tambaqui was developed, with the characterization of 30,000 candidate SNPs for the Axiom/ThermoFischer genotyping platform (30K SerraSNP array). After this, the genetic architecture of resistance to *A. hydrophila* in tambaqui was investigated by a genome-wide association study (GWAS) to identify quantitative trait loci (QTLs) and putative genes associated with this trait in tambaqui. A dense linkage map was developed for tambaqui using genotype data; resulting in 17,374 SNPs distributed in 27 linkage groups. GWAS analysis identified several candidate QTLs associated with *A. hydrophila* resistance, distributed into six linkage groups. Several candidate genes related to the immune response were located close to the putative QTLs. Regarding growth/yield traits, quantitative genetic analysis revealed moderate-high heritability, which demonstrates that they are under moderate genetic control and must respond to genetic selection. A deep learning approach and computer vision system (CVS) was also validated to obtain the morphometric measurements of tambaqui on a large scale and in an automated way. GWAS analyzes for these traits confirmed that they are under polygenic control, with the identification of only a few QTLs with statistical significance. In conclusion, both traits evaluated in this study, resistance to *A. hydrophila* and growth and body yield traits, demonstrate to be under genetic control and influenced by several genomic regions, and can be included in improvement programs for tambaqui.

Keyword: tambaqui, SerraSNP array, bacterial challenge, growth, associated SNPs.

1. Introdução Geral

1.1 Produção aquícola e caracterização do tambaqui *Colossoma macropomum*

Segundo dados da *Food and Agriculture Organization of the United Nations* (FAO), em 2018, a produção total da aquíicultura foi de 82 milhões de toneladas (FAO, 2020). De forma geral, a Ásia foi a que mais contribuiu com a produção de peixes cultivados para consumo humano, e o Brasil encontra-se em 13º lugar dos países que mais contribuíram com a aquíicultura mundial (FAO, 2020). Além disso, em termos de aquíicultura de água doce, o Brasil ocupa a oitava posição no ranking global, com valores mais baixos apenas em relação à China, Índia, Indonésia, Vietnã, Bangladesh, Egito e Mianmar (FAO, 2020). A aquíicultura brasileira apresentou expressivo avanço nos últimos anos e, em 2018, segundo dados do IBGE (Instituto Brasileiro de Geografia e Estatística), atingiu um valor de produção de R\$ 3,3 bilhões. A produção total de peixes foi de 519,3 mil toneladas em 2018, representando um aumento de 3,4% em relação ao ano anterior (IBGE, 2019).

O tambaqui *Colossoma macropomum* (Cuvier, 1818) é um peixe neotropical e corresponde ao maior peixe da ordem Characiformes da bacia Amazônica. Esta espécie pode medir até um metro de comprimento, pesar 30 kg, e se alimenta principalmente de frutos, sementes e zooplâncton (Goulding and Carvalho, 1982). Esta espécie é amplamente distribuída nos rios Amazonas e Orinoco (Jégu, 2003), com ocorrência natural no Brasil, Venezuela, Colômbia, Peru e Bolívia (Araujo-Lima e Goulding, 1997).

Em termos de produção na aquíicultura, o tambaqui é a espécie nativa da América do Sul mais cultivada nas pisciculturas. No Brasil, está em segundo lugar no ranking das espécies mais produzidas, atrás somente da tilápia (IBGE, 2019). Em 2018, a produção nacional alcançou o valor de 102 mil toneladas. Além disso, o tambaqui é utilizado para cruzamentos com outras espécies de Serrasalminidae (*Piaractus mesopotamicus* e

Piaractus brachypomus), resultando em híbridos interespecíficos, que são muito utilizados nas regiões Centro-Oeste e Sudeste (Hashimoto *et al.*, 2012).

Além da aceitação de mercado, características produtivas 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 genético, que já tem sido feito por universidades, centros de pesquisa e empresas, especialmente no Brasil.

1.2 Melhoramento genético em peixes

A seleção genética de plantas e animais terrestres tem sido praticada desde que a agricultura começou (Brander, 2007). Em animais terrestres, quase nenhuma produção baseia-se em populações selvagens ou não melhoradas geneticamente (Gjedrem *et al.*, 2012). Em peixes, a seleção iniciou-se recentemente, sendo que a maioria da produção aquícola (mais de 90%) ainda é baseada em unidades geneticamente não tratadas (Gjedrem *et al.*, 2012). Se toda a produção da aquicultura fosse resultante de produtos melhorados geneticamente, estima-se que o volume de produção dobraria em seis gerações de seleção, levando em consideração apenas o valor de ganho genético. Portanto, o aumento do uso de produtos melhorados também permite melhor utilização dos recursos, tais como ração, laboratório, água e área de produção (Gjedrem *et al.*, 2012). Ainda existem vários entraves para que o melhoramento baseado na seleção seja amplamente difundido na aquicultura, como insuficiente conhecimento sobre a biologia de espécies, programas de melhoramento para poucas espécies, além da captura regular de espécimes selvagens para reposição de matrizes reprodutoras (Gjedrem, 2005).

O melhoramento genético é um ponto chave para o desenvolvimento de uma aquicultura mais eficiente e sustentável (Robledo *et al.*, 2017), não só como objetivo de aumentar o desempenho produtivo (ex., peso e rendimento de filé), mas também melhorar os organismos em relação à resistência a doenças (Gjedrem, 1997). Atualmente, já foram realizados vários experimentos de seleção em larga escala para aumentar a taxa de crescimento, resultando em ganhos de 10-20% (média de 12,5%) por geração (Gjedrem *et al.*, 2012), bem como seleção dos organismos com melhor eficiência do sistema imune (Yáñez *et al.*, 2014; Gonen *et al.*, 2015).

1.3 Características de crescimento e rendimento de filé

Em tambaqui, alguns esforços iniciais foram direcionados para avaliar parâmetros genéticos para características relacionadas ao crescimento, incluindo ganho de peso, comprimento padrão, tamanho da cabeça, altura e largura (Marcos *et al.*, 2016; Mello *et al.*, 2016; Ariede *et al.*, 2018; Perazza *et al.*, 2019). Estudos prévios demonstraram variação genética aditiva significativa para ganho de peso diário durante a fase juvenil em tambaqui, com valores de herdabilidade moderado a alto variando de 0.37 a 0.49 para essa característica em diferentes idades (Mello *et al.*, 2016; Perazza *et al.*, 2019).

Na tilápia GIFT, estudos demonstraram herdabilidade de magnitude moderada (0,20 a 0,33) para características de crescimento e rendimento de carcaça, mas com baixa correlação (0,35 a 0,44) entre crescimento e rendimento de filé (Nguyen *et al.*, 2010). Rendimento de filé é uma das características mais difíceis de ser avaliada em programas de melhoramento de peixes, pelo fato de não poder ser medida diretamente sob os candidatos de seleção, ou seja, implica no sacrifício do animal e descarte de um potencial reprodutor, o que limita o progresso genético nos programas tradicionais de seleção (Kause *et al.*, 2007). Até o momento, poucos estudos genéticos de rendimento corpóreo

foram desenvolvidos em tambaqui. Recentemente, por meio de predições da área de lombo e espaço entre as costelas em tambaqui usando ultrassonografia, a herdabilidade para essas características foram estimadas como moderada-alta (Perazza *et al.* 2019).

Estudos utilizando imagem de ultrassom tem sido realizado em outras espécies com o objetivo de substituir métodos de seleção que impliquem no sacrifício dos indivíduos ou de seus parentes, pela seleção direta de candidatos com base no uso de preditores como critérios indiretos (Haffray *et al.* 2013). Em truta arco-íris foi possível a predição dos rendimentos de partes corporais por meio de modelos de regressão linear compreendendo morfologia externa e medições de ultrassom. Para a tilápia do Nilo (*Oreochromis niloticus*), foi realizada uma abordagem semelhante, com medidas morfológicas, para prever o peso do filé e o rendimento do filé (Rutten *et al.* 2004). Tendo em vista os valores de herdabilidade estimados em diferentes espécies de peixe para rendimento, é possível aumentar o rendimento de filé por meio de acasalamentos seletivos, pois os valores variam de baixa a moderada magnitude indicando que esta característica está sob controle genético e pode ser selecionada (Gjerde *et al.*, 2012, Haffray *et al.*, 2013, Haffray *et al.*, 2012, Navarro *et al.*, 2009, Nguyen *et al.*, 2010, Powell *et al.*, 2008, Rutten *et al.*, 2005a, Vandeputte *et al.*, 2014).

1.4 Resistência a doenças

No Brasil, as perdas econômicas por doenças são estimadas em cerca de US\$ 84 milhões (Tavares-Dias e Martins, 2017). Devido à alta densidade de estocagem e intensificação na produção de tambaqui, vários patógenos têm potencial para causar perdas significativas de produção (Valladão *et al.*, 2018). As doenças parasitárias mais comuns são causadas por monogeneas, tricodinídeos, mixosporídeos (Franceschini *et al.* 2013) e pelos cosmopolitas *Piscinoodinium pillulare* (Martins *et al.* 2015) e

Ichthyophthirius multifiliis (Silva *et al.* 2013). Para doenças bacterianas, os principais agentes são *Aeromonas hydrophila* (Belém-Costa e Cyrino 2006) e *Flavobacterium columnare* (Sebastião *et al.* 2010).

A infecção por *A. hydrophila* é uma causa comum de altas taxas de mortalidade de peixes em todo o mundo (Harikrishnan e Balasundaram, 2005), como em carpa (Yin *et al.*, 2009), tilápia (Ardó *et al.*, 2008), truta (Nya e Austin, 2009), catfish (Zhang *et al.*, 2016) e pacu (Farias *et al.*, 2016). Este microorganismo é frequentemente associado à septicemia hemorrágica, erosão da nadadeira e distensão da cavidade abdominal, entre outras patologias (Paniagua *et al.*, 1990), e causam mortalidade em massa de várias espécies (Abdel-Tawwab *et al.*, 2008; Wang *et al.*, 2015; Sirimanapong *et al.*, 2014).

A seleção genética para resistência a doenças tem como objetivo aumentar o ganho genético relacionado ao sistema imunológico de peixes cultivados (Gjedrem e Baranski, 2009), sendo assim, uma alternativa ao uso descontrolado de antibióticos e à falta de vacinas. Em peixes neotropicais, programas de melhoramento por meio de abordagens tradicionais baseadas no pedigree foram iniciados recentemente em pacu e tambaqui (Mastrochirico-Filho *et al.*, 2019; Lira *et al.*, 2020). Foram observadas variação fenotípica significativa relacionada à resistência a infecção por *Aeromonas hydrophila* em pacu (Mastrochirico-Filho *et al.*, 2019) e *Ichthyophthirius multifiliis* em tambaqui (Lira *et al.*, 2020).

Tendo em vista os danos causados por *A. hydrophila*, pesquisas recentes demonstraram que o desenvolvimento de populações geneticamente resistentes a *A. hydrophila* é viável em várias espécies, inclusive peixes neotropicais, representando, portanto, uma alternativa importante para o controle da doença (Mahapatra *et al.*, 2008; Sahoo *et al.*, 2008; Ødegård *et al.*, 2010; Xiong *et al.*, 2017; Mastrochirico-Filho *et al.*, 2019; Srisapoome *et al.*, 2019).

1.5 Genômica aplicada ao melhoramento genético de peixes

O melhoramento genético em peixes tem sido aprimorado devido as inovações biotecnológicas desenvolvidas na aquicultura mundial, permitindo um aumento da produtividade e qualidade. Estudos utilizando marcadores moleculares associados a características produtivas foram documentados em várias espécies de aquicultura (revisado em Palaiokostas e Houston, 2018), no entanto, pouco tem sido aplicado em peixes nativos da América do Sul.

Uma abordagem eficiente para identificação de QTLs (*locus* de características quantitativas) é por meio do estudo de associação genômica ampla (GWAS) (Kong *et al.*, 2019). Estudos utilizando GWAS foi realizado para diversas características em diferentes espécies de peixes, tais como, resistência a doenças (Correa *et al.*, 2015; 2017; Tsai *et al.*, 2016; Robledo *et al.*, 2018; Barría *et al.* 2018; Kong *et al.*, 2019; Wang *et al.*, 2019; Mastrochirico-Filho *et al.*, 2020), características de crescimento e rendimento de filé (Gutierrez *et al.*, 2015, Tsai *et al.*, 2015; Gonzalez-Pena *et al.*, 2016) e determinação sexual (Gabián *et al.*, 2019; Cáceres *et al.*, 2019).

No Brasil, os estudos da arquitetura genética de características e a descoberta de QTLs em peixes nativos eram limitados pelo número relativamente pequeno de marcadores genéticos disponíveis (Mastrochirico-Filho *et al.*, 2016; Ariede *et al.*, 2018; Jorge *et al.*, 2018). Até o momento, estudos recentes de GWAS tem demonstrado a identificação de várias regiões genômicas ligadas a um fenótipo raro e peculiar no tambaqui, que é a falta de espinhos intramusculares (Nunes *et al.*, 2020). Além disso, para o pacu, uma espécie relativamente próxima ao tambaqui, foram gerados milhares de SNPs que foram usados para identificar a arquitetura genética da resistência à *A. hydrophila* por meio de GWAS. Esse estudo identificou 22 QTLs putativos associados à resistência a *A.*

hydrophila e vários genes candidatos relacionados à resposta imune foram localizados próximos aos QTLs, sugerindo que esta característica está sob controle poligênico no pacu (Mastrochirico-Filho *et al.*, 2020).

Entretanto, uma das principais limitações para o avanço dos estudos genômicos em peixes nativos neotropicais continua a ser a falta de uma plataforma de genotipagem de alto rendimento. *Arrays* de SNPs já foram desenvolvidas para as principais espécies de peixes cultivados em todo o mundo, como salmão do Atlântico, truta arco-íris, bagre, tilápia e carpa (Houston *et al.* 2014; Palti *et al.* 2015; Liu *et al.* 2014; Xu *et al.* 2014; Yáñez *et al.* 2020). Tem sido demonstrado que *arrays* com média-alta densidade de SNP facilitam estudos em escala de genoma, permitindo a avaliação simultânea de milhares de SNPs em espécies de peixes comercialmente importantes (Houston *et al.* 2014; Yáñez *et al.* 2016; Palti *et al.* 2015; Yáñez *et al.* 2020).

1.6 Referências

- Abdel-Tawwab, M., Abdel-Rahman, A. M., and Ismael, N. E. M. (2008). Evaluation of commercial live baker's yeast, *Saccharomyces cerevisiae*, as a growth and immunity promoter for Fry Nile tilapia, *Oreochromis niloticus* challenged in situ with *Aeromonas hydrophila*. *Aquaculture* 280, 185–189.
- Araujo-Lima, C., & Goulding, M., (1997). So Fruitful a Fish: Ecology, Conservation, and Aquaculture of the Amazon's Tambaqui. Columbia University Press.
- Ardó, L., Yin, G., Xu, P., Váradi, L., Szigeti, G., *et al.* (2008). Chinese herbs (*Astragalus membranaceus* and *Lonicera japonica*) and boron enhance the non-specific immune response of Nile tilapia (*Oreochromis niloticus*) and resistance against *Aeromonas hydrophila*. *Aquaculture* 275, 26–33.
- Ariede, R. B., Freitas, M. V., Hata, M. E., *et al.* (2018). Microsatellites associated with growth performance and analysis of resistance to *Aeromonas hydrophila* in tambaqui *Colossoma macropomum*. *Frontiers in genetics*, 9, 3.

- Barría, A., Christensen, K. A., Yoshida, G. M., *et al.* (2018). Genomic predictions and genome-wide association study of resistance against *Piscirickettsia salmonis* in coho salmon (*Oncorhynchus kisutch*) using ddRAD sequencing. *G3: Genes, Genomes, Genetics*, 8(4), 1183-1194.
- Belém-Costa A., Cyrino, J.E.P. (2006) Antibiotic resistance of *Aeromonas hydrophila* isolated from *Piaractus mesopotamicus* (Holmberg, 1887) and *Oreochromis niloticus* (Linnaeus, 1758). *Scientia Agricola* 63: 281–284.
- Brander, K.M. (2007). Global fish production and climate change. *Proc. Natl.Acad. Sci. U.S. A.* 104, 19709–19714.
- Cáceres, G., López, M. E., Cádiz, M. I., Yoshida, G. M., Jedlicki, A., *et al.* (2019). Fine mapping using whole-genome sequencing confirms anti-Müllerian hormone as a major gene for sex determination in farmed Nile tilapia (*Oreochromis niloticus* L.). *G3: Genes, Genomes, Genetics*, 9(10), 3213-3223.
- Calcagnotto, D., Schaefer, S.A, DeSalle, R., (2005). Relationships among characiform fishes inferred from analysis of nuclear and mitochondrial gene sequences. *Mol. Phylogenet.Evol.* 36 (1), 135–153.
- Correa, K., Lhorente, J.P., López, M.E., Bassini, L., *et al.* (2015). Genome-wide association analysis reveals loci associated with resistance against *Piscirickettsia salmonis* in two Atlantic salmon (*Salmo salar* L.) chromosomes. *BMC genomics*, 16(1), 854.
- Correa, K., Bangera, R., Figueroa, R., Lhorente, J. P., & Yáñez, J. M. (2017). The use of genomic information increases the accuracy of breeding value predictions for sea louse (*Caligus rogercresseyi*) resistance in Atlantic salmon (*Salmo salar*). *Genetics Selection Evolution*, 49(1), 1-5.
- FAO. (2020). The State of World Fisheries and Aquaculture 2020. Sustainability in Action. Food and Agriculture Organization of the United Nations.
- Farias, T.H.V., Levy-pereira, N., Alves, L.O., Dias, D.C., Tachibana, L., *et al.* (2016). Probiotic feeding improves the immunity of pacus, *Piaractus mesopotamicus*, during *Aeromonas hydrophila* infection. *Anim. Feed Sci. Technol.* 211, 137–144.
- Ferraz, R. B., Machado, A. M., Navarro, J. C., Cunha, I., Ozório, R., Salaro, A. L., *et al.* (2020). The fatty acid elongation genes *elovl4a* and *elovl4b* are present and functional in the genome of tambaqui (*Colossoma macropomum*). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 245, 110447.

- Franceschini, L., Zago, A.C., Schalch, S.H.C., Garcia, F., Romera, D.M., Silva, R.J.D. (2013). Parasitic infections of *Piaractus mesopotamicus* and hybrid (*P. mesopotamicus* x *Piaractus brachypomus*) cultured in Brazil. *Revista Brasileira de Parasitologia Veterinária* 22: 407–414.
- Gabián, M., Morán, P., Fernández, A. I., Villanueva, B., Chtioui, A., *et al.* (2019). Identification of genomic regions regulating sex determination in Atlantic salmon using high density SNP data. *BMC genomics*, 20(1), 764.
- Gjedrem, T., and Baranski, M. (2009). *Selective Breeding in Aquaculture: Na Introduction*. Springer Science & Business Media: 221 pp.
- Gjedrem, T., Robinson, N., and Rye, M. (2012). The importance of selective breeding in aquaculture to meet future demands for animal protein: A review. *Aquaculture* 350–353: 117–129.
- Gjedrem, T. (1997). Breeding to raise resistance. In *Furunculosis* (pp. 405-418). Academic Press.
- Gjedrem T, Olesen I (2005). Basic statistical parameters. In: Gjedrem T (ed.) *Selection and Breeding Programs in Aquaculture*, pp. 45–72. Springer, Dordrecht, The Netherlands.
- Gjerde, B., Mengistu, S. B., Ødegård, J., Johansen, H., & Altamirano, D. S. (2012). Quantitative genetics of body weight, fillet weight and fillet yield in Nile tilapia (*Oreochromis niloticus*). *Aquaculture*, 342, 117-124.
- Gomes, L.C., Simões, L.N., Araujo-Lima, C.A.R.M. (2010). Tambaqui (*Colossoma macropomum*). In: Baldisserotto, B., Gomes, L.C. (Eds.), *Espécies Nativas Para Piscicultura No Brasil*. EditoraUFSM, Santa Maria, pp. 175–204.
- Gonen, S., Bishop, S.C., Houston, R.D. (2015). Exploring the utility of cross-laboratory RAD-sequencing datasets for phylogenetic analysis. *BMC Research Notes* 8:299.
- Gonzalez-Pena, D., Nixon, S. E., O'Connor, J. C., *et al.* (2016). Microglia Transcriptome Changes in a Model of Depressive Behavior after Immune Challenge. *PLoS ONE* 11:e0150858.
- Goulding, M., & Carvalho, M. L. (1982). Life history and management of the tambaqui (*Colossoma macropomum*, Characidae): an important Amazonian food fish. *Rev. Bras. Zool.* 1(2), 107-133.
- Gutierrez, A. P., Yáñez, J. M., Fukui, S., Swift, B., & Davidson, W. S. (2015). Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic salmon (*Salmo salar*). *PLoS One*, 10(3), e0119730.

- Haffray, P., Bugeon, J., Pincent, C., Chapuis, H., Mazeiraud, E., Rossignol, M.-N., Chatain, B., Vandeputte, M., Dupont-Nivet, M., 2012. Negative genetic correlations between production traits and head or bony tissues in large all-female rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*:368–369 145–152. 10.1016/j.aquaculture.2012.09.023.
- Haffray, P., Bugeon, J., Rivard, Q., Quittet, B., Puyo, S., Allamelou, J.M., Vandeputte, M., Dupont-Nivet, M. (2013). Genetic parameters of in-vivo prediction of carcass, head and fillet yields by internal ultrasound and 2D external imagery in large rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*:410–411 236–244. 10.1016/j.aquaculture.2013.06.016.
- Harikrishnan, R., & Balasundaram, C. (2005). Modern trends in *Aeromonas hydrophila* disease management with fish. *Reviews in Fisheries Science*, 13(4), 281-320.
- Hashimoto, D. T., Senhorini, J. A., Foresti, F., & Porto-Foresti, F. (2012). Interspecific fish hybrids in Brazil: management of genetic resources for sustainable use. *Reviews in Aquaculture*, 4(2), 108-118.
- Houston, R. D., Taggart, J. B., Cézard, T., Bekaert, M., Lowe, N. R., *et al.* (2014). Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). *BMC genomics*, 15(1), 1-13.
- IBGE 2019. Pesquisa Pecuária Municipal – 2018. Disponível em: <https://sidra.ibge.gov.br/pesquisa/ppm/quadros/brasil/2018>
- Jégu, M. (2003). “Subfamily Serrasalminae (Pacus and piranhas),” in *Check list of the freshwater fishes of South and Central America*, eds. R. E. Reis, S. O. Kullander, C. J. Ferraris (Porto Alegre: Edipucrs), 182-196.
- Jorge, P. H., Mastrochirico-Filho, V. A., Hata, M. E., Mendes, N. J., *et al.* (2018). Genetic characterization of the fish *Piaractus brachipomus* by microsatellites derived from transcriptome sequencing. *Frontiers in Genetics*, 9, 46.
- Kause, A., Paananen, T., Ritola, O., *et al.* (2007). Direct and indirect selection of visceral lipid weight, fillet weight, and fillet percentage in a rainbow trout breeding program. *J. Anim. Sci.* 85, 3218–3227.
- Kong, S., Ke, Q., Chen, L., Zhou, Z., Pu, F., *et al.* (2019). Constructing a high-density genetic linkage map for large yellow croaker (*Larimichthys crocea*) and mapping resistance trait against ciliate parasite *Cryptocaryon irritans*. *Marine Biotechnology*, 21(2), 262-275.

- Lira, L.V., Ariede, R.B., Freitas, M.V., Mastrochirico-Filho, V.A., *et al.* (2020). Quantitative genetic variation for resistance to the parasite *Ichthyophthirius multifiliis* in the Neotropical fish tambaqui (*Colossoma macropomum*). *Aquaculture Reports*, 17, 100338.
- Liu, S., Sun, L., Li, Y., *et al.* (2014). Development of the catfish 250 K SNP array for genome-wide association studies. *BMC Research Notes* 7:135–47.
- Mahapatra, K.D., Gjerde, B., Sahoo, P.K., *et al.*, 2008. Genetic variations in survival of rohu carp (*Labeo rohita*, Hamilton) after *Aeromonas hydrophila* infection in challenge tests. *Aquaculture* 279, 29–34.
- Marcos, R., Povh, J. A., Fornari, D. C., *et al.* (2016). Weight gain and morphometric growth of genetically improved tambaqui (*Colossoma macropomum*). *Semina: Ciências Agrárias* 37(4), 2521-2528.
- Martins M.L., Cardoso L, Marchiori N., Padua S.B. (2015). Protozoan infections in farmed fish from Brazil: diagnosis and pathogenesis. *Revista Brasileira de Parasitologia Veterinária* 24:1–20.
- Mastrochirico-Filho, V.A., Hata, M.E., Sato, L.S., Jorge, P.H., *et al.* (2016). SNP discovery from liver transcriptome in the fish *Piaractus mesopotamicus*. *Conservation genetics resources*, 8(2), 109-114.
- Mastrochirico-Filho, V.A., Ariede, R.B., Freitas, M.V., Lira, L.V., *et al.* (2019). Genetic parameters for resistance to *Aeromonas hydrophila* in the Neotropical fish pacu (*Piaractus mesopotamicus*). *Aquaculture*, 513, 734442.
- Mastrochirico-Filho, V. A., Borges, C. H., Freitas, M. V., Ariede, R. B., Pilarski, F., Utsunomia, R., *et al.* (2020). Development of a SNP linkage map and genome-wide association study for resistance to *Aeromonas hydrophila* in pacu (*Piaractus mesopotamicus*). *BMC genomics*, 21(1), 1-13. <https://doi.org/10.1186/s12864-020-07090-z>
- Mello, F. D., Oliveira, C. A. L., Streit, Jr. D., *et al.* (2016). Estimation of Genetic Parameters for Body Weight and Morphometric Traits to Tambaqui *Colossoma macropomum*. *J. FisheriesSciences* 10(2), 096-100.
- Navarro, A., Zamorano, M.J., Hildebrandt, S., Ginés, R., Aguilera, C., Afonso, J.M., 2009. Estimates of heritabilities and genetic correlations for growth and carcass traits in gilthead seabream (*Sparus auratus* L.), under industrial conditions. *Aquaculture* 289:225–230. <http://dx.doi.org/10.1016/j.aquaculture.2008.12.024>.

- Nguyen, N.H, Ponzoni, R.W., Abu-Bakar, K.R., *et al.* (2010). Correlated response in fillet weight and yield to selection for increased harvest weight in genetically improved farmed tilapia (GIFT strain), *Oreochromis niloticus*. *Aquaculture* 305 1–5.
- Nunes, J. D. R., Liu, S., Pértille, F., Perazza, C. A., Villela, P. M. S., *et al.* (2017). Large-scale SNP discovery and construction of a high-density genetic map of *Colossoma macropomum* through genotyping-by-sequencing. *Scientific Reports*, 7, 46112.
- Nya, E. J., & Austin, B. (2009). Use of garlic, *Allium sativum*, to control *Aeromonas hydrophila* infection in rainbow trout, *Oncorhynchus mykiss* (Walbaum). *Journal of fish diseases*, 32(11), 963-970.
- Ødegard, J., Sommer, A., and Praebel, A. (2010). Heritability of resistance to viral nervous necrosis in Atlantic cod (*Gadus morhua*). *Aquaculture*. 300 (1-4), 59-64.
- Oliveira, E. G., Pinheiro, A. B., de Oliveira, V. Q., da Silva Júnior, A. R. M., de Moraes, M. G., *et al.* (2012). Effects of stocking density on the performance of juvenile pirarucu (*Arapaima gigas*) in cages. *Aquaculture*, 370, 96-101.
- Palaiokostas, C., & Houston, R. D. (2018). Genome-wide approaches to understanding and improving complex traits in aquaculture species. *Perspect. Agric. Vet. Sci. Nutr. Nat. Resour*, 12, 1-10.
- Palti, Y., Gao, G., Liu, S., Kent, M.P., *et al.* (2015). The development and characterization of a 57 K single nucleotide polymorphism array for rainbow trout. *Molecular ecology resources*, 15(3), 662-672.
- Paniagua, C., Rivero, O., Anguita, J., and Naharro, G. (1990). Pathogenicity factors and virulence for rainbow trout (*Salmo gairdneri*) of motile *Aeromonas* spp. isolated from a river. *J. Clin. Microbiol.* 28, 350–355.
- Peng, W., Xu, J., Zhang, Y., Feng, J., Dong, C., *et al.* (2016). An ultra-high density linkage map and QTL mapping for sex and growth-related traits of common carp (*Cyprinus carpio*). *Scientific reports*, 6(1), 1-16.
- Perazza, C. A., Ferraz, J. B. S., Almeida-Val, V. M. F., & Hilsdorf, A. W. S. (2019). Genetic parameters for loin eye area and other body traits of an importante Neotropical aquaculture species, *Colossoma macropomum* (Cuvier, 1818). *Aquaculture Research*, 50(10), 2907-2916.
- Powell, J., White, I., Guy, D., Brotherstone, S., 2008. Genetic parameters of production traits in Atlantic salmon (*Salmo salar*). *Aquaculture* 274 (2–4), 225–231.
- Robledo, D., Palaiokostas, C., Bargelloni, L., *et al.* (2017). Applications of genotyping by sequencing in aquaculture breeding and genetics. *Reviews in Aquaculture*, 1–13.

- Robledo, D., Matika, O., Hamilton, A., & Houston, R. D. (2018). Genome-wide association and genomic selection for resistance to amoebic gill disease in Atlantic salmon. *G3: Genes, Genomes, Genetics*, 8(4), 1195-1203.
- Rutten, M.J.M., Bovenhuis, H., Komen, H., 2004. Modeling fillet traits based on body measurements in three Nile tilapia strains (*Oreochromis niloticus* L.). *Aquaculture* 231:113–122. <http://dx.doi.org/10.1016/j.aquaculture.2003.11.002>.
- Rutten, M.J.M., Bovenhuis, H., Komen, H., 2005. Genetic parameters for fillet traits and body measurements in Nile tilapia (*Oreochromis niloticus* L.). *Aquaculture* 246: 125–32. <http://dx.doi.org/10.1016/j.aquaculture.2005.01.006>.
- Sahoo, P. K., Mahapatra, K. D., Saha, J. N., Barat, A., Sahoo, M., *et al.* (2008). Family association between immune parameters and resistance to *Aeromonas hydrophila* infection in the Indian major carp, *Labeo rohita*. *Fish & Shellfish Immunology*, 25(1-2), 163-169.
- Sebastião, F.A., Pilarski, F., Lemos, M.V.F. (2010). Isolation and molecular characterization of *Flavobacterium columnare* strains from fish in Brazil. *Journal of Bacteriology Research* 3:22–29.
- Silva, R.M., Tavares-Dias, M., Dias, M.W.R., Dias, M.K.R., Marinho, R.D.G.B. (2013). Parasitic fauna in hybrid tambacu from fish farms. *Pesquisa Agropecuária Brasileira* 48: 1049–1057.
- Sirimanapong, W., Thompson, K. D., Kledmanee, K., *et al.* (2014). Optimisation and standardisation of functional immune assays for striped catfish (*Pangasianodon hypophthalmus*) to compare their immune response to live and heat killed *Aeromonas hydrophila* as models of infection and vaccination. *Fish & shellfish immunology*, 40(2), 374-383.
- Srisapoome, P., Chatchaiphan, S., Bunnoy, A., Koonawootrittriron, S., & Na-Nakorn, U. (2019). Heritability of immunity traits and disease resistance of bighead catfish, *Clarias macrocephalus* Günther, 1864. *Fish & shellfish immunology*, 92, 209-215.
- Tavares-Dias, M., & Martins, M. L. (2017). An overall estimation of losses caused by diseases in the Brazilian fish farms. *Journal of Parasitic Diseases*, 41(4), 913-918.
- Tsai, H. Y., Hamilton, A., Guy, D. R., *et al.* (2016). Verification of SNPs Associated with Growth Traits in Two Populations of Farmed Atlantic Salmon. *Int. J. Mol. Sci.* 17:5.

- Tsai, H.-Y., Hamilton, A., Tinch, A. E., *et al.* (2015). Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics* 16:969.
- Valladão, G.M.R., Gallani, S.U., and Pilarski, F. (2018). South American fish for continental aquaculture. *Rev Aquacult.* doi:10.1111/raq.12164.
- Vandeputte, M., Garouste, R., Dupont-Nivet, M., Haffray, P., Vergnet, A., Chavanne, H., Laureau, S., Ron, T.B., Pagelson, G., Mazorra, C., Ricoux, R., Marques, P., Gameiro, M., Chatain, B. (2014). Multi-site evaluation of the rearing performances of 5 wild populations of European sea bass (*Dicentrarchus labrax*). *Aquaculture* 424–425 239–248. <http://dx.doi.org/10.1016/j.aquaculture.2014.01.005>.
- Wang, L., Wan, Z. Y., Bai, B., *et al.* (2015). Construction of a high-density linkage map and fine mapping of QTL for growth in Asian seabass. *Scientific reports*, 5.
- Wang, W., Tan, S., Luo, J., Shi, H., Zhou, T., *et al.* (2019). GWAS analysis indicated importance of NF- κ B signaling pathway in host resistance against motile *Aeromonas* septicemia disease in catfish. *Marine Biotechnology*, 21(3), 335-347.
- Xiong, X. M., Chen, Y. L., Liu, L. F., *et al.* (2017). Estimation of genetic parameters for resistance to *Aeromonas hydrophila* in blunt snout bream (*Megalobrama amblycephala*). *Aquaculture*, 479, 768-773.
- Xu, J., Zhao, Z., Zhang, X., Zheng, X., Li, J., *et al.* (2014). Development and evaluation of the first high-throughput SNP array for common carp (*Cyprinus carpio*). *BMC genomics*, 15(1), 307.
- Yáñez, J. M., Houston, R. D., and Newman, S. (2014). Genetics and genomics of disease resistance in salmonid species. *Front. Genet.* 5:415.
- Yáñez, J.M., Naswa, S., López, M.E., Bassini, L., *et al.* (2016). Genomewide single nucleotide polymorphism discovery in Atlantic salmon (*Salmo salar*): validation in wild and farmed American and European populations. *Molecular ecology resources*, 16(4), 1002-1011.
- Yáñez, J.M., Yoshida, G., Barria, A., Palma-Véjares, R., *et al.* (2020). High-Throughput Single Nucleotide Polymorphism (SNP) Discovery and Validation Through Whole-Genome Resequencing in Nile Tilapia (*Oreochromis niloticus*). *Marine Biotechnology*, 22(1), 109-117.
- Yin, G., Ardó, L. Á. S. Z. L. Ó., Thompson, K. D., Adams, A., Jeney, Z., & Jeney, G. (2009). Chinese herbs (*Astragalus radix* and *Ganoderma lucidum*) enhance immune

response of carp, *Cyprinus carpio*, and protection against *Aeromonas hydrophila*. *Fish & Shellfish Immunology*, 26(1), 140-145.

Zhang, D., Xu, D. H., & Shoemaker, C. (2016). Experimental induction of motile *Aeromonas* septicemia in channel catfish (*Ictalurus punctatus*) by waterborne challenge with virulent *Aeromonas hydrophila*. *Aquaculture Reports*, 3, 18-23.

2. Objetivos

2.1 Objetivo geral

- ✓ Analisar o desempenho de famílias de tabaqui *C. macropomum* para resistência à *Aeromonas hydrophila* e características de crescimento, por meio de genética quantitativa e associação genômica de SNPs (*Single Nucleotide Polymorphisms*).

2.2 Objetivos específicos

- ✓ Verificar parâmetros de genética quantitativa das famílias de tabaqui em relação à resistência à *Aeromonas hydrophila*, por meio de um experimento de desafio bacteriano;
- ✓ Avaliar desempenho genético das famílias de tabaqui em relação às características de crescimento e rendimento corpóreo;
- ✓ Caracterizar QTLs (*Quantitative Trait Locus*) no tabaqui, por meio de GWAS para resistência à *Aeromonas hydrophila* e rendimento corpóreo.

3. Atividades desenvolvidas

As seções de Material e Métodos, Resultados e Discussão desta Tese estão apresentados na forma de quatro Capítulos, conforme detalhamento abaixo.

O Capítulo I, intitulado “Genetic (co)variation between resistance to *Aeromonas hydrophila* and growth in tambaqui (*Colossoma macropomum*)”, está em formato de artigo científico e foi publicado na revista *Aquaculture*. Este trabalho avaliou estimativas de parâmetros genéticos para características de resistência a bactéria *Aeromonas hydrophila* e crescimento na fase juvenil em tambaqui.

O Capítulo II, intitulado “Development of a multi-species SNP array for Serrasalmid fish *Colossoma macropomum* and *Piaractus mesopotamicus*”, não se encontra na forma de manuscrito científico, pois descreve as contribuições do presente trabalho para a elaboração de um estudo mais amplo, que também envolve análises genômicas de outras Teses e Dissertações. Portanto, este Capítulo reporta as atividades referentes à descoberta e validação de milhares de SNPs por sequenciamento de nova geração em tambaqui, usando a técnica ddRADseq, que foram incorporados na plataforma Axiom. Ao final, em anexo, os resultados e discussão desse Capítulo podem ser encontrados, juntamente com a somatória dos outros estudos, no formato de artigo científico publicado na revista *Scientific Reports*.

O capítulo III, intitulado “A dense linkage map and genome-wide association study for host resistance against *Aeromonas hydrophila* in tambaqui (*Colossoma macropomum*) using a 30K SNP array” está em formato de artigo científico, o qual foi submetido a revista *Genomics*. Este capítulo se refere a um estudo de associação genômica ampla (GWAS) para resistência a *A. hydrophila* em tambaqui, a fim de conhecer a arquitetura genética desta característica e compreender os genes responsáveis

pelos mecanismos de defesa a este patógeno. Além disso, contém um denso mapa de ligação utilizando marcadores SNPs no tabaqui.

O capítulo IV, intitulado “Avaliação de parâmetros genéticos para desempenho de crescimento e rendimento corpóreo em tabaqui *Colossoma macropomum*”, está estruturado na forma de artigo científico, no entanto, não foi submetido a revista científica. Este Capítulo reporta os cálculos de parâmetros genéticos para desempenho de crescimento (peso e medidas morfométricas) e rendimento corpóreo em uma população experimental de tabaqui. Além disso, reporta resultados de GWAS para tais características.

3.1 Capítulo I

Este manuscrito foi publicado na Revista Aquaculture (ISSN: 0044-8486). Por esta razão, não pode ter o texto completo disponibilizado no Repositório da Unesp, devido às restrições de direitos autorais da revista científica. Segue o endereço para o acesso completo da publicação.: Ariede, R. B., Freitas, M. V., Agudelo, J. F., Borges, C. H., Lira, L. V., Yoshida, G. M., *et al.* (2020). Genetic (co) variation between resistance to *Aeromonas hydrophila* and growth in tambaqui (*Colossoma macropomum*). *Aquaculture*, 523, 735225.

<https://doi.org/10.1016/j.aquaculture.2020.735225>.

Genetic (co)variation between resistance to *Aeromonas hydrophila* and growth in tambaqui (*Colossoma macropomum*)

Raquel B. Ariede^a, Milena V. Freitas^a, John F. G. Agudelo^a, Carolina H. S. Borges^a, Lieschen V.G. Lira^a, Grazyella M. Yoshida^b, Fabiana Pilarski^a, José M. Yáñez^b, Diogo T. Hashimoto^{a*}

^a São Paulo State University (Unesp), Aquaculture Center of Unesp, 14884-900 Jaboticabal, SP, Brazil

^b Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, Santiago, Chile

* Corresponding author: Aquaculture Center of Unesp, Via de Acesso Prof. Paulo Donato Castellane, s/n, 14884-900 Jaboticabal/SP. Phone number: +55 16 3209-7477. E-mail: diogo.hashimoto@unesp.br

Abstract

Tambaqui (*Colossoma macropomum*) is the main native fish farmed in South American continental aquaculture. One of the major pathogens affecting world fresh water aquaculture is the bacterium *Aeromonas hydrophila*, which generates several disease outbreaks and production losses in farmed tambaqui stocks. Up to date, there are no studies focusing on understanding the quantitative basis for the genetic improvement for disease resistance in tambaqui. Thus, the objective of this study was to determine the genetic parameters for resistance to *A. hydrophila* and the genetic correlation with average daily gain in juveniles of tambaqui, to determine whether these traits can be included in selective breeding programs. Estimation of genetic parameters was performed using data from an experimental challenge performed in 18 full-sib families, using a total of 576 individuals. Before bacterial challenge, all animals were evaluated for average daily gain (ADG) in the juvenile phase during 30 days. The challenge spanned 120 hours (5 days) and disease resistance traits were defined as: i) binary survival (BS) and, ii) time of death (TD) of fish presenting clinical signs of *A. hydrophila* infection. The mean ADG was 0.49 g/day (SD=0.21) in the population and heritability for this trait was moderate ($h^2 = 0.37 \pm 0.13$). BS and TD varied considerably among families (26% to 89% and 10.7 h to 69.2 h, respectively), which indicated a significant genetic variation related to resistance to *A. hydrophila* infection. Low to moderate values for heritability were found for BS and TD (0.17 ± 0.06 and 0.23 ± 0.09 , respectively). The genetic correlations between resistance to *A. hydrophila* and ADG in juveniles of tambaqui were not significantly different from zero. The significant genetic variation found for *A. hydrophila* resistance in tambaqui indicates that selecting superior genotypes is a viable approach to reducing the impact of diseases outbreaks in aquaculture.

KEY-WORDS: heritability; genetic parameters; disease resistance; aeromoniosis.

3.2 Capítulo II

Descoberta de SNPs para o desenvolvimento de um denso SNP array no tambaqui *Colossoma macropomum*

3.2.1 Material biológico

A população de peixes utilizada para a descoberta de SNPs é resultante do núcleo de melhoramento genético do Centro de Aquicultura da UNESP, CAUNESP, Jaboticabal, SP. O material biológico para construção das bibliotecas de ddRADseq (*double digest restriction-site associated DNA*) consiste de duas fontes diferentes: 1) conjunto de parentais e progênie de 17 famílias de irmãos completos (n=494) e 2) uma população de reprodutores (n=18). As 17 famílias foram produzidas por um esquema de acasalamento hierárquico utilizando 8 fêmeas e 13 machos (aprox. 1 fêmea para cada 2 machos). No total, 21 parentais e 473 indivíduos F1 foram selecionados das 17 famílias (25 peixes por família). O segundo grupo é formado por reprodutores do núcleo de melhoramento genético, oriundos de distintas pisciculturas comerciais do Brasil.

3.2.2 Construção das bibliotecas e seleção dos SNPs

Amostras de sangue foram utilizadas para extração de DNA com o kit comercial *Wizard Genomic* da Promega, seguindo as instruções do fabricante. A quantificação do DNA extraído (ng/μl) foi medida pelo detector de fluorescência Qubit, usando o kit Qubit dsDNA HS Assay (Invitrogen, EUA). O protocolo para a construção das bibliotecas foi de acordo com Peterson *et al.* (2012), com modificações. Resumidamente, 25 ng de DNA genômico de cada indivíduo foram digeridos (8 U/reação) usando duas opções de

combinação de enzimas, NlaIII e MluCI (usado para as 17 famílias) e SphI e MluCI (usado para o estoque de reprodutores) (New England) e ligados a adaptadores específicos de cada enzima (P1 e P2, 0,25 μ M) utilizando T4 DNA ligase, a 23°C por 1h 30 min e 65°C por 10 min ou 37 °C durante 60 min, respectivamente, para cada conjunto de enzimas. Os adaptadores P1 possuem 5 nucleotídeos adicionais que funcionam como marcadores individuais (barcode). A seleção dos fragmentos digeridos foi realizada pelo equipamento E-Gel® SizeSelect™ II (Thermo) com fragmentos de 250pb. Posteriormente, foram realizados ensaios de PCR para incorporar a identificação de cada biblioteca. Cada biblioteca foi formada com aproximadamente 47 amostras. A PCR foi realizada de acordo com as condições da enzima Phusion® High-Fidelity. As reações foram purificadas com o kit AMPure XP Beads e a concentração foi verificada novamente por fluorometria no equipamento Qubit® 3.0 (Thermo Scientific™). Por fim, as bibliotecas foram enviadas para serem sequenciadas no Illumina HiSeq4000.

Como resultado do sequenciamento, um total de 1.466.250.912 *reads* foram obtidas, resultando em aproximadamente 300 bilhões de bases (Gb) (aproximadamente 32 Gb por biblioteca). Uma média de 10% das *reads* obtidas foram removidas de cada biblioteca, e 51 indivíduos também foram excluídos da análise devido ao baixo número de *reads*. O número de *reads* por indivíduo variou de 2,24 M (milhões) a 3,10 M, com uma média de 2,6 M. Os dados brutos do sequenciamento e quantidade de *reads* obtidas por biblioteca estão descritos na Tabela 1.

Tabela 1. Resultado de sequenciamento ddRADseq em *C. macropomum* para cada biblioteca e número de reads excluídas após filtros de qualidade.

Biblioteca	Total de Sequências	Reads com sequência do adaptador	Barcode não encontrado	Baixa qualidade	Sem o sítio de restrição	Reads retidas	Média de reads por indivíduo (milhões)
B1	137.820.548	2.815.107	6.741.102	132.869	2.360.639	125.770.831	2,73
B2	141.028.848	3.113.368	13.915.128	147.051	2.189.333	121.663.968	2,76
B3	136.451.030	2.893.884	5.982.954	143.002	2.390.367	125.040.823	2,71
B4	129.325.884	3.134.904	10.561.196	123.389	2.300.633	113.205.762	2,51
B5	126.345.426	2.509.719	6.818.818	116.082	2.183.722	114.717.085	2,49
B6	125.608.630	2.892.544	7.327.596	124.361	2.680.636	112.583.493	2,44
B7	111.205.174	1.760.515	8.832.422	182.746	2.148.245	98.281.246	2,47
B8	157.562.820	3.369.837	8.576.446	263.202	2.686.233	142.667.102	3,10
B9	137.987.898	2.770.362	7.001.170	363.118	2.235.703	125.617.545	2,73
B10	149.275.700	3.018.739	7.386.006	271.130	2.450.138	136.149.687	2,95
B11	113.638.954	2.508.043	5.880.742	176.325	1.828.925	103.244.919	2,24
B12	129.330.488	3.432.343	6.629.494	198.436	2.462.236	116.607.979	2,53

Os dados obtidos do sequenciamento foram analisados no *software* Stacks v. 2.41 para a identificação de SNPs (Catchen *et al.* 2013). Os módulos do *sotware* Stacks foram executados sequencialmente, sendo estes, *process_radtags*, *ustacks*, *cstacks*, *sstacks* e *populations*. Inicialmente, as *reads* foram separadas pelos barcodes e filtradas usando *process_radtags*, com parâmetros para remoção de *reads* de baixa qualidade. Em seguida, os *loci* foram construídos pela metodologia *de novo* devido à ausência de genoma de referência para a espécie *Colossoma macropomum*, utilizando uma profundidade de leitura mínima de três ($m = 3$). Um catálogo de *loci* foi construído em *cstacks* usando apenas sequências dos indivíduos parentais para o grupo das famílias e todos os indivíduos do estoque de reprodutores. O número máximo de incompatibilidades entre os *loci* de cada amostra ao construir o catálogo foi definido como três ($n = 3$). O programa *sstacks* comparou *loci* putativos individuais (construídos por *ustacks*) com o catálogo. O módulo *populations* foi executado para gerar dados de genótipos para as amostras de

população. Além disso, o *software* Plink 1.9 (Purcell, 2007) foi utilizado para excluir falsos SNPs usando a taxa de erro Mendeliana ($me = 0,1$) e equilíbrio de Hardy-Weinberg.

Para o conjunto de dados oriundo das 17 famílias, um catálogo contendo 1.154.132 *loci* foi gerado, dos quais 10.583 *loci* apresentava 10.955 SNPs. Para selecionar os SNPs, foram considerados apenas *loci* presentes em pelo menos 40% dos indivíduos. Além disso, para diferenciar potenciais SNPs de erros de sequenciamento, foram excluídos SNPs com frequências alélicas menores ($MAF < 0,1$) e SNPs com mais de 10% de taxa de erro mendeliana. Com isso, 2.612 SNPs deste conjunto de dados passaram no controle de qualidade.

O conjunto de dados originado de uma população de reprodutores gerou um catálogo com 2.436.687 *loci*, dos quais 57.713 *loci* apresentava 27.053 SNPs. Após os filtros de seleção de SNPs, tais como, *call rate* $> 0,70$, *MAF* (0,10) e equilíbrio de Hardy-Weinberg (0,05), 23.825 SNPs passaram nos filtros de qualidade.

Ao total, 26.437 SNPs foram descobertos nesse estudo para serem utilizados na construção de um denso SNP array, por meio da plataforma Axiom (Affymetrix). Além disso, SNPs adicionais de sequências derivadas de outros estudos com ddRADseq foram incluídas (**Anexo I**). Estas sequências são de estudos em andamento com o objetivo de explorar a diversidade genômica de populações cultivadas em tambaqui. Bancos de dados já publicados usando diferentes técnicas, como GBS e RNAseq, também foram usados como fonte para aumentar o poder de descoberta de SNPs, espalhados por todo o genoma (**Anexo I**). Desta forma, foi possível a obtenção inicial de 81.848 SNPs candidatos para desenvolver um denso array no tambaqui.

Então, uma lista de *loci* dos 81.848 SNPs candidatos foram fornecidos à Affymetrix, para caracterização *in silico* das melhores sondas a serem incluídas no array. A lista de SNPs estava configurada em sequências nucleotídicas de 71-mer, com os alelos

do SNP alvo destacados na posição 36. O valor *p-convert* (representando a probabilidade de um dado SNP resultar em um ensaio de SNP confiável no sistema *Axiom array*) foi calculado pela Affymetrix para cada sequência de SNP submetida. As sondas foram avaliadas para cada SNP na direção *forward* e *reverse*, e designados como recomendado, neutro ou não recomendado, evitando regiões com mais de um SNP (*wobble*), e considerando uma região adjacente de 25 pb monomórfica ao SNP. Além disso, uma segunda etapa de filtragem foi realizada para obter os melhores SNPs. Basicamente, os dados foram filtrados para atingir apenas um SNP por *locus*, sendo que a sonda com maior *p-convert* foi selecionada. A predição *in silico* para selecionar as melhores sondas de acordo com os critérios do *Axiom array* resultou em 42.851 marcadores recomendados para tabaqui (Tabela 2), dos quais 29.575 SNPs foram incorporados no *Axiom array*, denominado como *SerraSNP array*.

Tabela 2. Classificação dos SNPs de acordo com os critérios de controle de qualidade *Axiom*.

SNPs	
Categoria	tabaqui
recommended	42,851
neutral	11,857
not_recommended	27,140
total	81,848

3.2.3 Validação do *Axiom array* em tabaqui

No total, 59 amostras de DNA genômico foram enviadas à Thermo Fisher Scientific (EUA, Califórnia) para genotipagem utilizando o *SerraSNP array*, com os SNPs

selecionados previamente. Os dados gerados servem para testar e validar o desempenho do array e quantificar o número de SNPs segregantes nas várias populações amostradas.

O SerraSNP array foi testado em 5 populações diferentes de tabaqui (3 do Brasil, uma da Colômbia e uma do Peru), das quais três do Brasil são as mesmas populações utilizadas para a descoberta de SNP (Tabela 3).

Tabela 3. Estimativas genéticas descritivas de população incluídas na validação do array.

Espécie	População	n	SNPs MAF > 0.01	Média MAF	Ho	He
Tabaqui	Brasil (total)	28	20.380	0,266	0,362	0,356
	Piscicultura 1	9	19.572	0,267	0,39	0,357
	Piscicultura 2	9	17.172	0,268	0,426	0,352
	Piscicultura 3	10	19.078	0,274	0,379	0,365
	Colômbia - Piscicultura 4	15	15.649	0,247	0,351	0,331
	Peru - Piscicultura 5	15	15.075	0,248	0,353	0,329
	TOTAL	58	21.072	0,247	0,306	0,338

Os dados brutos contendo os resultados dos cálculos de intensidade (arquivos CEL) foram importados para o Axiom Analysis Suite (v2.0.035, Affymetrix) para análise de controle de qualidade e leitura do genótipo. As amostras com um valor de *dish control quality* (DQC) > 0,82 e *QC call rate* > 0,97 (seguindo o “Best Practices Workflow” recomendado pela Affymetrix), foram consideradas aptas na avaliação de controle de qualidade.

A análise de controle de qualidade classifica os SNPs em categorias de acordo com seu desempenho de agrupamento em relação aos critérios de controle de qualidade Axiom: (i) "alta resolução polimórfica", onde o SNP passa em todos os QC, (ii) "alta resolução monomórfica", onde o SNP passa em todos QC, mas apenas um genótipo é detectado, (iii) "taxa de leitura abaixo do limite", onde a taxa de leitura do genótipo é <97%, (iv) "nenhum homocigoto menor", onde o SNP passa em todo o QC, mas apenas

dois clusters são observados, (v) “Variante fora do alvo”, onde as propriedades atípicas do cluster surgem de variantes na região flanqueadora do SNP, e (vi) “outro”, onde o SNP não se enquadra em nenhuma das categorias anteriores. Para análises posteriores, apenas SNPs das categorias (i) e (iv) foram incluídos e classificados como polimórficos, uma vez que são mais prováveis de serem SNPs confiáveis e informativos.

Quase todas as amostras passaram no DQC e QC call rate > 97%, exceto uma. A classificação dos SNPs de acordo com sua qualidade mostrou que 71,25% (n=21,072) foram classificados como polimórficos (“PolyHighResolution” e “NoMinorHom”) (Tabela 4). O MAF médio desses SNPs polimórficos nas amostras totais de tabaqui foi 0,247 (Tabela 3), com a maioria dos SNPs distribuídos em MAF acima de 0,1 (Figura 1).

Tabela 4. Resumo da classificação dos SNP de acordo com seu desempenho no array Axiom SerraSNP.

Categoria	Tabaqui	
	Número de marcadores	%
<i>PolyHighResolution</i>	19.661	66,48
<i>NoMinorHom</i>	1.411	4,77
<i>MonoHighResolution</i>	1.468	4,96
<i>CallRateBelowThreshold</i>	1.443	4,88
<i>OffTargetVariant</i>	550	1,86
<i>Other</i>	5.042	17,05
Total	29.575	100.00

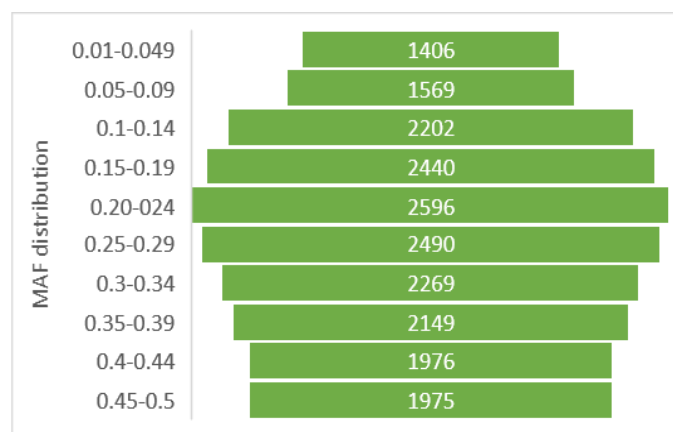


Figura 1. Distribuição dos SNPs de acordo com os valores do MAF.

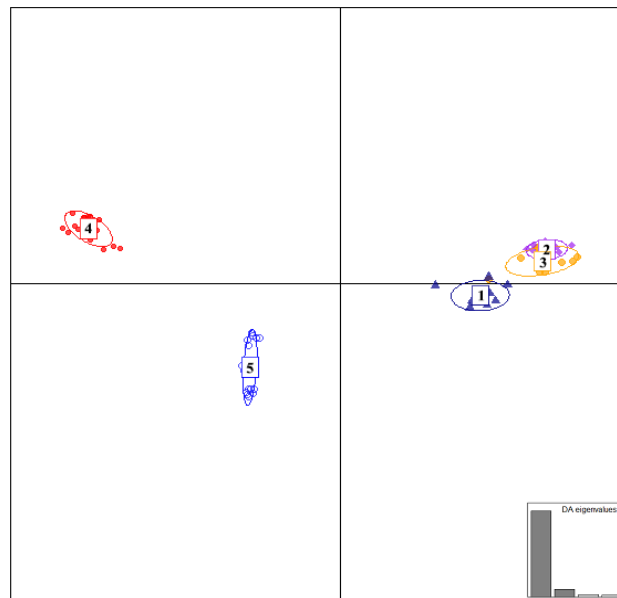


Figura 2. Análise DAPC de populações de tabaqui para demonstração da estrutura genética utilizando SNPs.

Parâmetros de genética populacional também foram calculados entre as diferentes pisciculturas. A maioria dos SNPs está segregando dentro de cada uma das populações, que variaram de 15.075 (Piscicultura 5) a 20.380 (Brasil- total) (Tabela 3). Os resultados indicam que as populações possuem altos níveis de diversidade genética baseados nos parâmetros de MAF e heterozigidade. As populações com maior MAF também revelaram os maiores valores de heterozigidade (piscicultura 3); e, conseqüentemente, essa população pode ser considerada com a maior variabilidade genética entre todas as analisadas no presente estudo. No geral, nenhuma evidência forte de deficiência de heterozigidade foi detectada, e a maioria das populações apresentaram níveis mais elevados de heterozigidade observada do que o esperado (de acordo com HWE).

Em relação à estrutura genética dos diferentes estoques de reprodutores, há evidências de agrupamento de amostras de acordo com a origem do estoque reprodutor (Figura 2). As pisciculturas 4 (Colômbia) e 5 (Peru) são claramente diferentes geneticamente em relação às populações do Brasil, e demonstraram baixa estrutura genética entre si, principalmente entre piscicultura 2 e piscicultura 3.

Devido à ausência de genoma de referência no tabaqui, a cobertura do genoma dos SNPs polimórficos não pode ser avaliada. Portanto, o grau do desequilíbrio de ligação (LD) foi avaliada por SNP *pruning* usando o programa plink para as diferentes populações de tabaqui (Figura 3). *Pairwise* r^2 usando diferentes limiares foi calculado com os SNPs polimórficos ($MAF > 0,01$). Em geral, diferentes perfis de LD podem ser detectados nas populações de tabaqui. Utilizando $r^2 > 0,8$, que é considerado um limiar alto de LD, os SNPs informativos variam de cerca de 11K (piscicultura 2) a 16K (piscicultura 3).

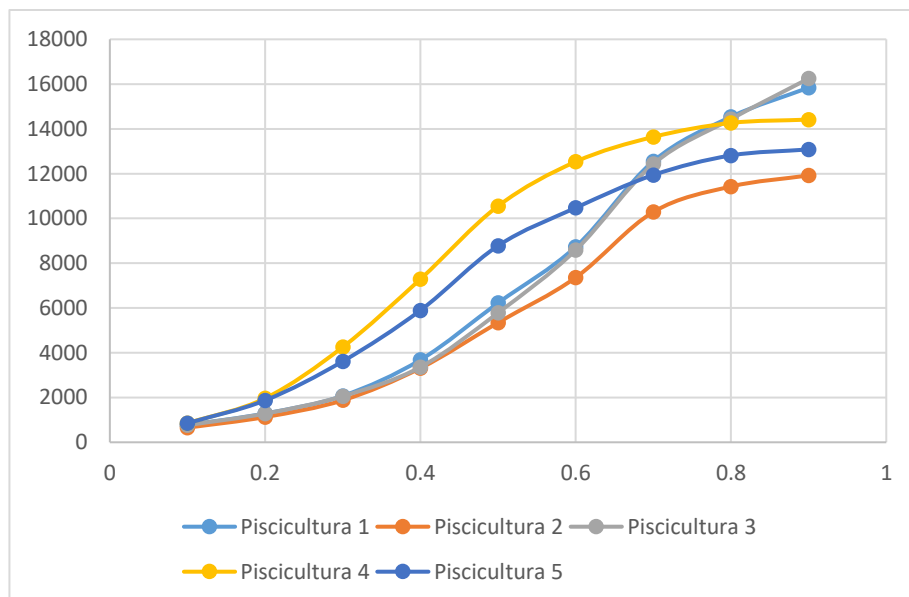


Figura 3. Desequilíbrio de ligação (LD) por SNP *pruning* entre marcadores de todas as populações de tabaqui amostradas.

Vale salientar que o SerraSNP array foi desenvolvido em uma plataforma multi-espécie, em conjunto com 30K SNPs de pacu *Piaractus mesopotamicus*. Portanto, a publicação dos dados é conjunta e o manuscrito está publicado na revista *Scientific Reports*, cujos detalhes podem ser encontrados no **Anexo I**.

Anexo I

Este manuscrito foi publicado na Revista Scientific Reports (ISSN: 2045-2322), que é Open Access e seu conteúdo pode ser disponibilizado. Segue o endereço para o acesso completo da publicação: Mastrochirico-Filho, V.A., Ariede, R.B., Freitas, M.V. *et al.* Development of a multi-species SNP array for serrasalmid fish *Colossoma macropomum* and *Piaractus mesopotamicus*. *Sci Rep* **11**, 19289 (2021). <https://doi.org/10.1038/s41598-021-98885-x>

Development of a multi-species SNP array for Serrasalmid fish *Colossoma macropomum* and *Piaractus mesopotamicus*

Authors

Vito A. Mastrochirico-Filho^{1,+}, Raquel B. Ariede^{1,+}, Milena V. Freitas¹, Carolina H. S. Borges¹, Lieschen V. G. Lira¹, Natália J. Mendes¹, John F. G. Agudelo¹, Pablo Cáceres², Milthon H. M. Berrocal³, Gustavo A. L. Sucerquia⁴, Fabio Porto-Foresti¹, José M. Yáñez², Diogo T. Hashimoto^{1*}

+ These authors contributed equally to the work

Affiliations

1. São Paulo State University (Unesp), Aquaculture Center of Unesp, 14884-900 Jaboticabal, SP, Brazil
2. Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, Santiago, Chile
3. Facultad de Zootecnia, Universidad Nacional Agraria de la Selva, Tingo Maria, Perú
4. Facultad de Ciencias Agrarias, Universidad de Antioquia, Medellín, Colômbia

*corresponding author(s): Diogo T. Hashimoto (diogo.hashimoto@unesp.br)

Abstract

Scarce genomic resources have limited the development of breeding programs for serrasalmid fish *Colossoma macropomum* (tambaqui) and *Piaractus mesopotamicus* (pacu), the key native freshwater fish species produced in South America. The main objectives of this study were to design a dense SNP array for this fish group and to validate its performance on farmed populations

from several locations in South America. Using multiple approaches based on different populations of tambaqui and pacu, a final list of 29,575 and 29,612 putative SNPs was selected, respectively, to print an Axiom AFFYMETRIX (THERMOFISHER) SerraSNP array. After validation, 74.17% (n = 21,963) and 71.25% (n = 21,072) of SNPs were classified as polymorphic variants in pacu and tambaqui, respectively. Most of the SNPs segregated within each population ranging from 14,199 to 19,856 in pacu; and from 15,075 to 20,380 in tambaqui. Our results indicate high levels of genetic diversity and clustered samples according to their hatchery origin. The developed SerraSNP array represents a valuable genomic tool approaching in-depth genetic studies for these species.

Introduction

Tambaquis, pacus and piranhas are included in the family Serrasalminae (Ostariophysi: Characiformes)¹, and they are broadly distributed throughout the major river systems of South America². Serrasalminae (round-shaped fish) are a diverse group comprising 101 valid species, represented by a variety of feeding strategies and associated morphological adaptations³. While piranhas are popularly known for voracious and carnivorous behavior, tambaquis and pacus are omnivorous, possessing special teeth to eat fruit and seeds, and have a great ability to harness natural food sources (zooplankton). Beyond their feeding habitats, these species have excellent traits of economic interest for aquaculture including high growth performance, ease of reproduction and market appreciation^{4,5,6}.

Tambaqui (*Colossoma macropomum*), pacu (*Piaractus mesopotamicus*) and pirapitinga (*Piaractus brachipomus*) are the major serrasalminid fish produced by the aquaculture industry in South America⁴. Tambaqui and pirapitinga are mainly farmed in the Amazon region (north region of Brazil, Colombia and Peru), while pacu is produced at higher latitudes in South America (southern Brazil and northern Argentina). These species have also been introduced for farming purposes in several countries in Asia, including China, Indonesia, Malaysia, Myanmar and Viet Nam^{5,7}. About 142 thousand tonnes of tambaqui was produced in 2016, of which 96.4 % was produced by Brazil (approximately 137 thousand tonnes)⁸. Pacu is also primarily produced in Brazil, representing the second largest native fish species production in the country, with 12 thousand tonnes produced in 2019⁹.

Despite the potential for serrasalminid species in aquaculture, their production remains entirely based on genetically unimproved stocks, and there is enormous untapped potential for selective breeding in these species. Traditional pedigree-based breeding programs have only recently been initiated for these species^{10,11,12}. Currently, there are no reports of the incorporation of genomic information to breeding programs for serrasalminid species, which can be explained by the lack of genomic resources for this fish group. For example, only a few studies have described the discovery of novel molecular markers for these species^{13,14,15}. However, the recent availability of genetic maps constructed using GBS (Genotyping by Sequencing) for tambaqui¹⁶ and RADseq (Restriction site-associated DNA sequencing) for pacu¹⁷, create new opportunities for the application of molecular markers to uncover the genetic basis of economically and ecologically important traits.

Genome wide association studies (GWAS) are considered useful tools for dissecting complex traits in both natural and farmed populations^{18,19}. In addition, the use of genomic selection has the potential to increase genetic gain^{20,21} and, consequently, boost emerging breeding programs of new species. A few GWAS have been performed for serrasalminid species. For example, suggestive Quantitative Trait Loci (QTLs) and genes have been associated with resistance against an important bacterial pathogen in pacu¹⁷, and to a lack of intermuscular bones in tambaqui²². One of the major limitations for application of GWAS and genomic selection in

serrasalmid fish is the lack of a standard high-throughput and robust genotyping assay. Dense single nucleotide polymorphism (SNP) arrays have been developed for popular farmed fish species, such as Atlantic salmon^{23,24}, rainbow trout²⁵, coho salmon²⁶, catfish²⁷, tilapia²⁸ and carp²⁹. These arrays have been broadly used to study traits of economic importance and to apply genomic selection in aquaculture breeding programs^{30,31}.

In the present study, the main objective was to design and evaluate a multi-species SNP array for two key serrasalmid fish of economic importance, *C. macropomum* and *P. mesopotamicus*, using SNPs from multiple sources including different populations from commercial broodstocks and breeding programs. The SerraSNP array was developed using AFFYMETRIX Axiom (THERMOFISHER) technology and validated in samples from various populations of pacu and tambaqui. This dense SNP array was also tested in a closely related serrasalmid (*P. brachypomus*) to assess its potential use in other commercial species.

Results

The sequence data used to generate the SNP dataset from ddRADseq and RADseq is part of a larger ongoing study aimed at exploring the genomic diversity of farmed tambaqui and pacu. Therefore, the descriptive data sequence of ddRADseq and RADseq will be prepared and presented in separate studies. After the *de novo* formation of the loci catalog and consequent read alignment, thousands of SNPs were identified in both species. Following quality-control filtering of putative SNPs, an initial selection of 130,403 and 81,848 filtered SNPs from pacu and tambaqui, respectively were identified. The *in silico* prediction to select the best probes according to the Axiom array criteria resulted in 99,682 and 42,851 recommended markers for pacu and tambaqui, respectively (Table 1). Further filtering step was applied to choose 30K SNPs for each species, selecting the final list of 29,612 putative SNPs for pacu and 29,575 putative SNPs for tambaqui to be included at the SerraSNP Axiom array. Most of the SNPs incorporated in the SerraSNP array were obtained by RADseq (17,932) in pacu and by ddRADseq (25,929) in tambaqui (Table 2). 2,200 validated SNPs positioned in the tambaqui linkage map described by Nunes-Silva *et al.* (2017)¹⁶ passed the SNP filtering steps and were included in the final array design (Table 2).

The SerraSNP array was used to genotype 94 and 58 samples of pacu and tambaqui, respectively. A list containing the putative polymorphic SNPs and the flanking sequences for tambaqui and pacu were provided respectively, including: the source of sequencing and the classification based on *p-convert* values and wobble criteria (Supplementary Table S1 and Supplementary Table S2). Almost all samples passed the QC and genotype call rate > 97 % threshold, except two samples of pacu and one of tambaqui. The classification of SNPs according to their quality showed that 74.17 % (n = 21,963) and 71.25 % (n = 21,072) were classified as polymorphic (either “Poly High Resolution” and “No Minor Hom”) in pacu and tambaqui, respectively (Table 3). The average MAF of these polymorphic SNPs in the combined total samples of pacu and tambaqui were 0.203 and 0.247, respectively (Table 4). In general, the lower MAF value in pacu is due to the larger number of SNPs distributed between MAF 0.01 and 0.049, which is mostly the result of SNPs belonging to the RADseq and ddRADseq dataset (Fig. 1). The different techniques for SNP discovery showed low interference on population segregation of SNPs (Table 4; Supplementary Table S3). Most of SNPs were segregating within each population with MAF > 0.01, ranging from 14,199 (Hatchery2) to 19,856 (Hatchery4) in pacu; and from 15,075 (Hatchery9) to 20,380 (Hatchery1) in tambaqui (Table 4).

In relation to SNPs annotation, 14,033 (66.9 %) and 13,046 SNPs (62.7 %) were located into transcribed regions in pacu and tambaqui, respectively (Supplementary Table S4). Regarding the RNA-Seq derived SNPs in pacu 2,047 SNPs (48.3 %) were in intergenic regions; 1,486

(35.1%) in untranslated regions (5' and 3' UTR); and 705 (16.6%) in coding sequences (cds), including 490 SNPs as synonymous and 211 as missense variants (Table 5). In relation to RNA-Seq-derived SNPs in tambaqui, 208 SNPs (69.8 %) were identified in intergenic regions, 31 (10.4 %) in 5' and 3' UTR; and 59 (19.8%) were considered as synonymous SNPs (Table 5).

To assess the utility of the SerraSNP array on a closely related species, which is also relevant for aquaculture, a set of DNA samples of pirapitinga (*Piaractus brachypomus*) was also tested. The samples passed the QC call rate when using a 94 % threshold in the 30 K SNP dataset of pacu, which demonstrates a high SNP conversion rate across this species. However, the polymorphic rate was much lower compared to pacu, resulting in 3,042 SNPs (about 10 % for all 30K SNPs of pacu on the SerraSNP array) (Table 4). The transferability of tambaqui markers into pacu or pirapitinga DNA resulted in failed SNP calling, even using a lower QC call rate (< 70 %). The same occurred when investigating the pacu markers using tambaqui DNA samples.

Parameters of population genetics were calculated between different hatcheries of both species. The expected heterozygosity ranged between 0.29 to 0.31 in pacu and between 0.33 to 0.37 in tambaqui. The expected heterozygosity for the pirapitinga samples was 0.35. Our results indicated that populations of both species revealed high levels of genetic diversity using the parameters of MAF and heterozygosity (Table 4). Overall, no strong evidence of heterozygous deficiency was detected, and most populations had higher levels of observed heterozygosity than expected (according to HWE). In both species, the populations with highest MAF also revealed the highest heterozygosity values (Hatchery3 for pacu, and Hatchery1 for tambaqui); and, consequently, these populations have the highest genetic variability values in the present study (Table 4).

According to IBS analysis for pacu (Supplementary Fig. S1) and tambaqui (Supplementary Fig. S2), and DAPC results (Fig. 2), there is evidence of the genetic structure of pacu and tambaqui samples clustering according to their hatchery origin. In pacu, hatchery2 and hatchery3 tended to cluster separately from the group formed by hatchery1 and hatchery4, which shows high genetic similarity (Supplementary Fig. S1). In tambaqui, hatchery8 (Colombia) and hatchery9 (Peru) show clear evidence of genetic differentiation in relation to populations from Brazil, which demonstrated low genetic structure to each other, particularly between hatchery6 and hatchery7 (Supplementary Fig. S2).

Due to the absence of a reference genome at chromosome level (at least not published yet) for pacu and tambaqui, the genome coverage of the polymorphic SNPs cannot be evaluated in the present study. However, correlation analysis between previously mapped SNPs included on the SerraSNP array and the corresponding linkage map for tambaqui (SNPs from GBS technique¹⁶) and pacu (RADseq¹⁷) revealed high and positive values for the relative amount (0.90 and 0.99, respectively) and density (0.98 and 0.99, respectively) (Supplementary Table S5), suggesting a wide coverage of these markers in the genome. Moreover, the extent of Linkage Disequilibrium (LD) between markers was assessed by SNP pruning for the different populations of pacu and tambaqui (Fig. 3). Pairwise r^2 was calculated among the polymorphic SNPs (MAF > 0.01) and a range of r^2 pruning thresholds were applied (from 0.1 to 0.9) to determine the number of markers remaining after each filtering step in each population of both species. Thus, we have a picture of the number of SNPs with different levels of redundant genotypic information. Generally, different profiles of increasing numbers of markers, at increasing levels of LD pruning, can be detected in populations of pacu. For instance, the lowest and the highest number of pruned SNPs from pacu hatchery1 and hatchery3 were at different levels of LD. Pacu hatchery2 and hatchery4 showed similar numbers of SNPs pruned at different LD levels. Assuming a $r^2 > 0.8$, which is considered a strong LD threshold, the number of informative SNPs showing non-highly redundant genotypic information ranges from about 9K (hatchery3) to 19K (hatchery1), depending on the population of pacu analysed. In contrast, the number of pruned SNPs at different

levels of LD from tambaqui populations, showed a similar trend between the different populations analysed. Assuming a $r^2 > 0.8$, the number of informative SNPs showing non-highly redundant genotypic information ranges from about 11 K (hatchery6) to 15 K (hatchery5), depending on the population of tambaqui analysed.

Discussion

Dense SNP arrays have been shown to facilitate genome-scale studies by allowing the simultaneous evaluation of thousands of SNPs in commercially important fish species, such as Atlantic salmon^{23,24}, rainbow trout²⁵ and tilapia²⁸. These markers have facilitated the analysis of GWAS for important commercial traits (growth rate, sex determination and disease resistance) and also the implementation of genomic selection in fish species^{32,33,34,35}. Although pacu and tambaqui represent the main native fish of continental aquaculture in South America, studies aimed at the incorporation of genome-wide SNP information to boost the breeding programs of these species are scarce. Therefore, the broad utility (multi-species) and open access format of the array presented here will allow the advancement of genomic studies in this fish group and support ongoing and emerging breeding programs in South America.

After assessment of the cluster properties of each of the SNPs in the SerraSNP array, most of the markers on the Axiom platform were designated as high quality and polymorphic, with validation rates of 74.17% and 71.25% in pacu and tambaqui, respectively. This conversion rate is similar to previous SNP arrays developed for other aquaculture species^{36,37}, demonstrating the efficacy of our multi-strategy design and stringent filtering steps for robust SNP discovery. A balanced conversion rate was observed across the four main SNP discovery techniques, except the RNA-Seq-derived SNPs with lower conversion rate (33%) in tambaqui, which could be linked to the low coverage sequencing (454 Roche technology) and the small sampling used to generate the RNA-Seq database (*i.e.*, the previously published dataset was not sequenced with primarily target of SNP discovery), or even to the absence of a reference genome resulting in false positive markers²³. Otherwise, the previous published GBS-derived SNPs used in tambaqui registered the highest conversion rate (85%), as result of a particular strategy in SNP discovery and validation by using linkage mapping experiments¹⁶. In pacu, the larger dataset of candidate SNPs provided to AFFYMETRIX were derived from RNA-Seq, but only 10% were incorporated on the Axiom array. This high discrepancy compared to RADseq and ddRADseq (40-50%) is due to the larger size of the RNA contigs, which resulted in multiple SNPs per contig; therefore, as only one SNP per contig was selected, high drop out of SNPs derived from the RNA-Seq dataset was observed.

Overall, although a high conversion rate of QC-filtered SNPs was reported, it is likely that a low proportion of false positive SNPs discovered in these sequencing experiments would remain, particularly in the ddRAD and RADseq dataset. This is due to unique potential sources of error and bias in the library preparation protocol, especially related to PCR duplicates (clonal DNA fragments originated during PCR steps) that can lead to downstream genotyping errors (for review, see Andrews *et al.*, 2016)³⁸. Moreover, false positive or failed SNPs may also occur in RNA-seq-derived SNPs because Mendelian errors in pedigreed samples were not possible to be detected by this method and/or due to limitations in the genotyping technology (SNPs located close to exon-intron boundaries).

It is also worth noting that the rate of monomorphic SNPs was low, with values of 5.98% and 4.96% in pacu and tambaqui, respectively, especially when compared to the array developed for Atlantic Salmon using similar techniques (12-53%)²³. The high values of monomorphic SNPs (large number of false positive SNPs) in the latter study were mainly attributed to the RNA-seq technique (53%), which is particularly susceptible to false positive marker discovery²³. Moreover, the conversion rate in the Atlantic Salmon array (46%) was relatively low comparing to the

SerraSNP array (71-74%), mostly due to the duplicated genome nature in salmonids, in which was necessary to apply special strategies to avoid false positive SNPs, such as effective removal of putative paralogous variants by RR-seq (reduced-representation sequencing) of a haploid fish (conversion rate of 74%).

In the present study, a high fraction of polymorphic SNPs was identified in putative transcribed regions of the genomes of pacu and tambaqui (66.9 and 62.7%, respectively) (Supplementary Table S4). Moreover, particularly for the RNA-Seq-derived SNPs, several markers were also identified in coding sequences or regulatory regions of transcripts; therefore, they are likely to be functional and could be linked to phenotypes of economic interest. Thus, the enrichment of associated SNPs to transcribed genes in the SerraSNP array will be useful for the determining the genetic architecture of target traits in aquaculture and, consequently, for the inclusion of genomic selection in breeding programs of these species.

Population genetic analysis showed that the SNP discovery strategy used here allowed us to identify and develop a high-quality 60K genotyping array which can be reliably used to genotype different populations of farmed pacu and tambaqui. Low to moderate genetic diversity values and genetic differentiation between farmed populations has already been found for both species probably due to the lack of controlled management of broodstocks, although genetic similarity was found in between some farms most likely due to physical proximity favoring an exchange of fish between them^{39,40}.

Results from the assessment of the SNP segregation between different populations indicate that the SNP panel developed in the present study would be useful for genetic applications across different populations of pacu and tambaqui, including high-resolution population genomics, genome wide-association studies and genomic selection. Nevertheless, the performance of this SNP array is expected to have a slight decrease in some populations particularly due to genetic composition of the target populations used for SNP discovery (instead ascertainment bias of the techniques). For instance, most of the samples used for SNP discovery in tambaqui were from Brazil, creating bias for Colombia and Peru hatcheries. Otherwise, a similar rate of SNPs segregation at population level was detected across the different techniques of SNP discovery, even when using limited sampling source (only 3 families by ddRAD in pacu).

The conversion rate for using SNP markers between species (pacu into tambaqui, and *vice-versa*) or using an additional species (pirapitinga) is low. This situation is comparable to the 250K SNP array for catfish, in which the transferability for different species also resulted in a low number of polymorphic SNPs²⁷. Only 10% of the 30K SNPs for pacu on the SerraSNP array were polymorphic to pirapitinga; thus, the probes designed from pacu sequences could hybridize to the genomic DNA of pirapitinga, but the level of polymorphic SNPs in the latter species was limited. However, as there is no SNP marker data available for pirapitinga to date, the 3K polymorphic SNPs identified are a valuable resource for the application of genomics to population genetic and evolutionary studies, and for selective breeding in in this species; especially given the high values of genetic diversity identified ($MAF = 0.28$; $H_o = 0.45$ and $H_e = 0.35$).

Although a larger sample size for the population of each species for LD analysis and haplotypes construction is still necessary, the preliminary results of LD pruning demonstrate that the SNP panel of pacu and tambaqui presented here will be useful for the design of low-density SNP panels providing little redundant genotypic information. As observed previously in other fish species⁴¹, lower density SNP panels can be designed and applied for genomic selection and breeding, with fewer tag markers selected on interesting traits. In addition, correlation values between the mapped SNPs included in the SerraSNP array and linkage maps, previously published for tambaqui¹⁶ and pacu¹⁷, showed a high coverage and representation of the SerraSNP

array throughout the genome, which reflects to its reliability and usefulness for genetic studies aiming the development of breeding programs for both species.

This study describes the development and analysis of a dense SNP array for two *Serrasalminidae* species. A large database of SNP markers using multiple approaches was developed for pacu and tambaqui, both native species widely farmed in South America. Following stringent filtering criteria, SNP assays for these two fish species were combined on the 60K SNP array, to generate a 30K high quality SNP panel for each species. Testing of the array in diverse populations revealed a high number of informative SNPs that are shared between species populations. Also, the array can be used to assess genetic diversity and population structure between populations. The SerraSNP array has open access, which will facilitate the study of important economic and ecological/evolutionary traits for these two *Serrasalminidae* species, including applications such as genomic selection, QTL mapping, phylogenetic analyses and conservation genetic programs.

Methods

Ethics Statement

This study was conducted in strict accordance with the recommendations of the National Council for Control of Animal Experimentation (CONCEA) (Brazilian Ministry for Science, Technology and Innovation) and was approved by the Ethics Committee on Animal Use (CEUA numbers 19.005/17 and 19.006/17) of Faculdade de Ciências Agrárias e Veterinárias, UNESP, Campus Jaboticabal, SP, Brazil.

Sample Information

The tambaqui samples used for SNP discovery were obtained from broodstocks of different commercial hatcheries in South America and from the breeding nucleus of the Aquaculture Center (CAUNESP) of São Paulo State University (UNESP), Brazil (Table 6).

The commercial hatcheries represented ten different populations, of which six were from Brazil, two from Peru and two from Colombia. The samples from the Caunesp breeding nucleus were composed of 29 selected breeders, plus the parents and offspring from 20 full-sib families generated with a hierarchical mating scheme using 6 dams and 17 sires. In total, the 20 full-sib family dataset consisted of 23 parents and 577 F1 individuals (\cong 30 fish per family). The SNP array was validated with 5 different populations of tambaqui (three from Brazil, one from Colombia and one from Peru). The three from Brazil were the same populations used for SNP discovery (Table 4).

The individuals of pacu used for SNP discovery were obtained from the breeding nucleus of CAUNESP and represented 17 full-sib families (Table 1), which were generated with a hierarchical mating scheme using 8 dams and 15 sires. In total, 23 parents and 477 F1 individuals were selected from the 17 families (\cong 25 fish per family). The SNP array was tested on four different populations of pacu collected from broodstocks of different commercial hatcheries in Brazil, which were not the same populations used for SNP discovery (Table 4).

SNP Discovery and Filtering

DNA extraction was performed using the DNeasy Blood & Tissue kit (QIAGEN). Purified DNA was quantified using the Qubit dsDNA BR Assay kit (INVITROGEN). There is no reference genome available for tambaqui and pacu, therefore, the ddRADseq approaches were used for SNP discovery because they are appropriate for genome reduction and *de novo* assembly^{42,43}. Previous published databases using different techniques, such as GBS, RADseq and RNA-Seq (Table 6), were also used to increase the power of SNP discovery throughout the genome.

ddRADseq library construction was completed as previously described by Peterson *et al.* (2012)⁴³. Briefly, 25 ng of genomic DNA from each individual was double-digested (8 U/reaction) using one of two enzyme combinations, I (NlaIII and MluCI) or II (SphI and MluCI) (Table 1), and ligated to specific adapters for each enzyme (P1 and P2, 0.25 μ M) using T4 DNA ligase, at 23 °C for 1h 30 min and 65 °C for 10 min. P1 adapters had an additional 5 nucleotides that served as individual markers (barcode). Size selection of the digested fragments was performed with E-Gel SizeSelect II (THERMOFISHER) equipment. Subsequently, PCR assays were performed to incorporate the indexes identification of each library (about 48 samples/library). Phusion enzyme conditions were used to perform the PCR assay. Reactions were purified with the AMPure XP Beads Kit and analysed by Agilent Bioanalyzer and Qubit. ddRADseq libraries were sequenced by NOVOGENE (Sacramento, USA) on ILLUMINA HiSeq4000 (PE 150 bp).

After sequencing, data were analyzed using the software package Stacks for *de novo* SNP identification (STACKS v. 2.41⁴⁴). The subprograms of Stacks were implemented sequentially (*process_radtags*, *ustacks*, *cstacks*, *sstacks*). The *stacks populations* module was used to generate genotype output data for the population samples. Initially, reads were demultiplexed and filtered using *process_radtags*, with parameters that removed reads with uncalled bases, discard reads with low quality scores and excluded sequences in which barcodes and RAD cutsites were not found. Subsequently, the loci were constructed by *de novo* methodology due to the absence of reference genomes for both species, using a minimum read depth of three ($m=3$). A loci catalog was constructed in *cstacks* using a subset of individuals for those derived from broodstocks or using the parental individuals for those resulting from family structure. The maximum number of mismatches between sample loci when building the catalog was set to three ($n=3$). The *sstacks* program was used in order to match individual putative loci (constructed by *ustacks*) against the catalog, followed by the module *populations* for SNP discovery, with the filter parameters of minor allele frequency (MAF) > 0.05 and genotyping call rate > 0.7. To differentiate putative SNPs from sequencing errors, PLINK 1.9 software⁴⁵ was used to exclude SNPs using Hardy-Weinberg equilibrium (assessed on each individual broodstock) and Mendelian error rate (*me* 0.1 for family structure). Moreover, individuals with high missing genotype rates (*mind* 0.3) were also discarded.

The raw sequences from previous RNA-Seq databases were downloaded from the NCBI database (see Data Availability section) and the following steps were performed. The RNA-Seq database of pacu was originated from 36 individuals challenged against the bacteria *Aeromonas hydrophila* (for details, see Mastrochirico-Filho *et al.*, 2020⁴⁶). The RNA-Seq raw data of tambaqui was obtained from 10 liver samples and 8 muscle and skin samples, all of them collected from random individuals, as described in detail by Ariede *et al.* (2018)¹⁴ and Gomes *et al.* (2019)⁴⁷, respectively. High quality reads were used to assemble a *de novo* transcriptome using TRINITY v.2.9.1, adopting standard parameters⁴⁸, such as a kmer value of 25 and contigs longer than 200 bp. The read mapping was performed by BOWTIE2 v.2.3.4.3⁴⁹. Redundancy was reduced using the CD-HIT-EST v.4.6.8 software, clustering sequences with 95% identity⁵⁰. After mapping,

the SAMTOOLS program⁵¹ was used to manipulate the BAM files for SNP discovery. The filtering was performed with a quality equal to or higher than 20 (Phred score) and variants with a minimum sequencing depth of 10 reads. Abundant and repetitive SNPs in small areas were excluded. The VCFtools program⁵² was used to eliminate SNPs with MAF values less than 0.05.

SNPs from GBS were the same used for linkage mapping in tambaqui described in detail by Nunes *et al.*, (2017)¹⁶.

SNP selection for Axiom array design

A list consisting of 130,403 and 81,848 putative SNP markers for pacu and tambaqui was provided to THERMOFISHER (AFFYMETRIX) as 71-mer nucleotide sequences, with both alleles at the target SNP highlighted at position 36. A *p-convert value* (probability of a putative SNP locus converting to a reliable assay on the Axiom array) was produced by THERMOFISHER team for each submitted SNP adjacent sequence. Probes were analysed for each SNP side (forward and reverse), and then classified as recommended, neutral, or not recommended, based on *p-convert* values and wobble criteria (*i.e.*, nearby interfering polymorphisms or another SNP at the flanking sequence of the target marker).

The initial analysis performed by THERMOFISHER resulted in a superior number of recommended markers than the total capacity of the Axiom MyDesign custom array. Therefore, a second step of filtering was carried out to obtain the best 30K SNPs for each species. Basically, the data was filtered to achieve only one SNP per RAD/ddRAD locus or RNA contig. When one RAD/ddRAD locus or RNA contig was identified with multiple recommended SNPs, the SNP marker with highest *p-convert* was selected.

SNP array validation

In total, 152 genomic DNA samples (94 pacu and 58 tambaqui individuals) were sent to THERMOFISHER (USA, California) for genotyping using the SerraSNP array. The results were used to validate the performance of the array and to quantify the number of segregating SNPs in the sampled populations.

Raw data consisting of intensity calculations (CEL files) was imported into the Axiom Analysis Suite (v2.0.035, AFFYMETRIX) for genotype calling and quality control. Samples presenting a QC call rate > 0.97 and quality control (DQC) > 0.82 passed the quality control assessment (following the “Best Practices Workflow” recommended by AFFYMETRIX). Quality control analysis then classified the SNPs into categories according to their clustering performance in relation to Axiom quality control criteria: (i) “polymorphic high resolution” where the SNP passes all QC, (ii) “monomorphic high resolution” where the SNP passes all QC, but only one genotype is detected, (iii) “call rate below threshold” where genotype call rate is < 97 %, (iv) “no minor homozygote” where the SNP passes all QC but only two clusters are observed, (v) “off-target variant” where atypical cluster properties arise from variants in the SNP flanking region, and (vi) “other” where the SNP does not fall into any of the previous categories. Only SNPs from categories (i) and (iv) were included and identified as polymorphic for further analyses, as they are most probably reliable and informative SNPs.

Descriptive statistics of minor allele frequencies (MAF), expected (H_e) and observed (H_o) heterozygosity; and IBS (Identity by State) analysis followed by multi-dimensional scaling (MDS), discriminant analysis of principal components (DAPC) and linkage disequilibrium (LD) were performed using PLINK 1.9 software⁴⁵. A predefined tambaqui genome (not published) was

adopted as reference to identify SNPs into transcribed regions using BEDOPS v2.4.40⁵³. To evaluate SNPs from RNA-Seq with potential functional effects, the list of polymorphic SNPs selected for the array was annotated with SNPEff v 5.0⁵⁴ using Uniprot proteins database as reference.

At the moment, there are still no reference genomes at chromosome level available for both species. However, linkage maps were previously developed for tambaqui¹⁶ and pacu¹⁷ using 7,192 and 17,453 SNPs, respectively. Of these total, 25.6 % and 42.6 % of mapped SNPs were included respectively in the SerraSNP array, and were correlated to the linkage maps to evaluate the level of coverage and genetic representativeness of the SerraSNP array for both species.

Data Availability

The raw fastq files obtained by RADseq methods for SNP discovery in tambaqui⁵⁵ and pacu⁵⁶ are available in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA). RNA sequencing information is encompassed for pacu⁵⁷ and tambaqui⁵⁸ by NCBI BioProject ID PRJNA632934 and PRJNA358254, respectively.

References

1. Fricke, R., Eschmeyer, W. N. & Van der Laan, R. (eds). Eschmeyer's catalog of fishes: genera, species, references <http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp> (2020).
2. Jégu, M. Subfamily Serrasalminae (Pacus and piranhas) in *Check List of the Freshwater Fishes of South and Central America* (Eds. Reis, R. E., Kullander, S.O. & Ferraris, C. J.) 182-196 (Edipucrs, 2003).
3. Mateussi, N. T. B., Melo, B. F., Ota, R. P., Roxo, F. F., Ochoa, L. E., Foresti, F. & Oliveira, C. Phylogenomics of the Neotropical fish Family Serrasalminidae with a novel intrafamilial classification (Teleostei: Characiformes). *Mol. Phylogenet. Evol.* **153**, 106945; <https://doi.org/10.1016/j.ympev.2020.106945> (2020)
4. Valladão, G. M. R., Gallani, S. U. & Pilarski, F. South American fish for continental aquaculture. *Rev. Aquacult.* **10**, 351-369; <https://doi.org/10.1111/raq.12164> (2018)
5. Woynárovich, A. & Van Anrooy, R. *Field guide to the culture of tambaqui (Colossoma macropomum, Cuvier, 1816)*. FAO Fisheries and Aquaculture Technical Paper 624 (FAO Food and Agriculture Organization of the United Nations, 2019).
6. Freitas, M. V., Lira, L. V. G., Ariede, R. B., Agudelo, J. F. G., Neto, R. R. O., Borges, C. H. S., Mastrochirico-Filho, V. A., Neto, B. F. G., Carvalheiro, R. & Hashimoto, D. T. Genotype by environment interaction and genetic parameters for growth traits in the Neotropical fish pacu (*Piaractus mesopotamicus*). *Aquaculture* **530**, 735933; <https://doi.org/10.1016/j.aquaculture.2020.735933> (2020)
7. Honglang, H. Freshwater fish seed resources in China in *Assessment of freshwater fish seed resources for sustainable aquaculture. FAO Fisheries Technical Paper. No. 501* (Eds. Bondad-Reantaso, M. G.) 185-199 (FAO Food and Agriculture Organization of the United Nations, 2007).
8. IBGE (Instituto Brasileiro de Geografia e Estatística). Produção da Pecuária Municipal 2016 <https://biblioteca.ibge.gov.br/index.php/biblioteca-atálogo?view=detalhes&id=784> (2016)
9. IBGE (Instituto Brasileiro de Geografia e Estatística). Produção da Pecuária Municipal 2019 <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=784> (2019).

10. Mastrochirico-Filho, V. A., Ariede, R. B., Freitas, M. V., Lira, L. V., Agudelo, J. F. G., Pilarski, F., Neto, R. V. R., Yáñez, J. M. & Hashimoto, D. T. Genetic parameters for resistance to *Aeromonas hydrophila* in the Neotropical fish pacu (*Piaractus mesopotamicus*). *Aquaculture* **513**, 734442; <https://doi.org/10.1016/j.aquaculture.2019.734442> (2019)
11. Ariede, R. B., Freitas, M. V., Agudelo, J. F. G., Borges, C. H. S., Lira, L. V. G., Yoshida, G. M., Pilarski, F., Yáñez, J. M. & Hashimoto, D. T. Genetic (co) variation between resistance to *Aeromonas hydrophila* and growth in tambaqui (*Colossoma macropomum*). *Aquaculture* **523**, 735225; <https://doi.org/10.1016/j.aquaculture.2020.735225> (2020)
12. Lira, L. V., Ariede, R. B., Freitas, M. V., Mastrochirico-Filho, V. A., Agudelo, J. F. G., Barría, A., Yáñez, J. M. & Hashimoto, D. T. Quantitative genetic variation for resistance to the parasite *Ichthyophthirius multifiliis* in the Neotropical fish tambaqui (*Colossoma macropomum*). *Aquac. Rep.* **17**, 100338; <https://doi.org/10.1016/j.aqrep.2020.100338> (2020)
13. Mastrochirico-Filho, V. A., Hata, M. E., Sato, L. S., Jorge, P. H., Foresti, F., Rodriguez, M. V., Martínez, P., Porto-Foresti, F. & Hashimoto, D. T. SNP discovery from liver transcriptome in the fish *Piaractus mesopotamicus*. *Conserv. Genet. Resour.* **8**, 109-114; <https://doi.org/10.1007/s12686-016-0521-3> (2016)
14. Ariede, R. B., Freitas, M. V., Hata, M. E., Mastrochirico-Filho, V. A., Utsunomia, R., Mendonça, F. F., Foresti, F., Porto-Foresti, F. & Hashimoto, D. T. Development of microsatellite markers using next-generation sequencing for the fish *Colossoma macropomum*. *Mol. Biol. Rep.* **45**, 9-18; <https://doi.org/10.1007/s11033-017-4134-z> (2018)
15. Jorge, P. H., Mastrochirico-Filho, V. A., Hata, M. E., Mendes, N. J., Ariede, R. B., Freitas, M. V., Vera, M., Porto-Foresti, F. & Hashimoto, D. T. Genetic characterization of the fish *Piaractus brachypomus* by microsatellites derived from transcriptome sequencing. *Front. Genet.* **9**, 46; <https://dx.doi.org/10.3389%2Ffgene.2018.00046> (2018)
16. Nunes, J. R. S., Liu, S., Pértille, F., Perazza, C. A., Villela, P. M. S., Almeida-Val, V. M. F., Hilsdorf, A. W. S., Liu, Z. & Coutinho, L. L. Large-scale SNP discovery and construction of a high-density genetic map of *Colossoma macropomum* through genotyping-by-sequencing. *Sci. Rep.* **7**, 46112; <https://dx.doi.org/10.1038/srep46112> (2017)
17. Mastrochirico-Filho, V. A., Borges, C. H. S., Freitas, M. V., Ariede, R. B., Pilarski, F., Utsunomia, R., Carneiro, R., Gutierrez, A. P., Peñaloza, C., Yáñez, J. M., Houston, R. D. & Hashimoto, D. T. Development of a SNP linkage map and genome-wide association study for resistance to *Aeromonas hydrophila* in pacu (*Piaractus mesopotamicus*). *BMC Genomics* **21**, 672; <https://doi.org/10.1186/s12864-020-07090-z> (2020)
18. Yue, G. H. Recent advances of genome mapping and marker-assisted selection in aquaculture. *Fish Fish.* **15**, 376-396; <https://doi.org/10.1111/faf.12020> (2014)
19. Yue, G. H. & Wang, L. Current status of genome sequencing and its applications in aquaculture. *Aquaculture* **468**, 337-347; <https://doi.org/10.1016/j.aquaculture.2016.10.036> (2017)
20. Meuwissen, T. H., Hayes B. J. & Goddard M. E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819-1829 (2001).
21. Sonesson, A. K. & Meuwissen, T. H. Testing strategies for genomic selection in aquaculture breeding programs. *Genet. Sel. Evol.* **41**, 37; <https://dx.doi.org/10.1186%2F1297-9686-41-37> (2009)
22. Nunes, J. R. S., Pértille, F., Andrade, S. C. S., Perazza, C. A., Villela, P. M. S., Almeida-Val, V. M. F., Gao, Z. -X., Coutinho, L. L. & Hilsdorf, A. W. S. Genome-wide association study reveals genes associated with the absence of intermuscular bones in tambaqui (*Colossoma macropomum*). *Anim. Genet.*; <https://doi.org/10.1111/age.13001> (2020)
23. Houston, R. D., Taggart, J. B., Cézard, T., Bekaert, M., Lowe, N. R., Downing, A., Talbot, R., Bishop, S. C., Archibald, A. L., Bron, J. E., Penman, D. J., Davassi, A., Brew, F., Tinch, A. E., Gharbi, K. & Hamilton, A. Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). *BMC genomics* **15**, 1-13 (2014).

24. Yáñez, J. M., Naswa, S., López, M. E., Bassini, L., Correa, K., Gilbey, J., Bernatchez, L., Norris, A., Neira, R., Lhorente, J. P., Schnable, P. S., Newman, S., Mileham, A., Deeb, N., Di Genova, A. & Maass, A. Genomewide single nucleotide polymorphism discovery in Atlantic salmon (*Salmo salar*): validation in wild and farmed American and European populations. *Mol. Ecol. Resour.* **16**, 1002-1011; <https://doi.org/10.1111/1755-0998.12503> (2016)
25. Palti, Y., Gao, G., Liu, S., Kent, M. P., Lien, S., Miller, M. R., Rexroad 3rd, C. E. & Moen, T. The development and characterization of a 57 K single nucleotide polymorphism array for rainbow trout. *Mol. Ecol. Resour.* **15**, 662-672; <https://doi.org/10.1111/1755-0998.12337> (2015)
26. Barría, A., Christensen, K. A., Yoshida, G., Jedlicki, A., Leong, J. S., Rondeau, E. B., Lhorente, J.P., Koop, B., Davidson, W.S. & Yáñez, J. M. Whole genome linkage disequilibrium and effective population size in a coho salmon (*Oncorhynchus kisutch*) breeding population using a high-density SNP array. *Front. Genet.* **10**, 498; <https://doi.org/10.3389/fgene.2019.00498> (2019)
27. Liu, S., Sun, L., Li, Y., Sun, F., Jiang, Y., Zhang, Y., Zhang, J., Feng, J., Kaltenboeck, L., Kucuktas, H. & Liu, Z. Development of the catfish 250K SNP array for genome-wide association studies. *BMC Res. Notes* **7**, 135 (2014).
28. Yáñez, J. M., Yoshida, G., Barria, A., Palma-Véjares, R., Dante, T., Díaz, D., Cáceres, G., Cádiz, M. I., López, M. E., Lhorente, J. P., Jedlicki, A., Soto, J., Salas, D. & Maass, A. High-Throughput Single Nucleotide Polymorphism (SNP) Discovery and Validation Through Whole-Genome Resequencing in Nile Tilapia (*Oreochromis niloticus*). *Mar. Biotechnol.* **22**, 109-117; <https://doi.org/10.1007/s10126-019-09935-5> (2020)
29. Xu, J., Zhao, Z., Zhang, X., Zheng, X., Li, J., Jiang, Y., Kuang, Y., Zhang, Y., Feng, J., Li, C., Yu, J., Li, Q., Zhu, Y., Liu, Y., Xu, P. & Sun, X. Development and evaluation of the first high-throughput SNP array for common carp (*Cyprinus carpio*). *BMC Genomics* **15**, 307; <https://doi.org/10.1186/1471-2164-15-307> (2014)
30. Yáñez, J. M., Newman, S. & Houston, R. D. Genomics in aquaculture to better understand species biology and accelerate genetic progress. *Front. Genet.* **6**, 128; <https://doi.org/10.3389/fgene.2015.00128> (2015)
31. You, X., Shan, X. & Shi, Q. Research advances in the genomics and applications for molecular breeding of aquaculture animals. *Aquaculture* **526**, 735357; <https://doi.org/10.1016/j.aquaculture.2020.735357> (2020)
32. Correa, K., Lhorente, J. P., López, M. E., Bassini, L., Naswa, S., Deeb, N., Di Genova, A., Maass, A., Davidson, W. S. & Yáñez, J. M. Genome-wide association analysis reveals loci associated with resistance against *Piscirickettsia salmonis* in two Atlantic salmon (*Salmo salar* L.) chromosomes. *BMC Genomics* **16**, 854; <https://doi.org/10.1186/s12864-015-2038-7> (2015)
33. Palaiokostas, C., Ferraresso, S., Franch, R., Houston, R. D. & Bargelloni, L. Genomic prediction of resistance to pasteurellosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. *G3-Genes Genom. Genet.* **6**, 3693-3700; <https://doi.org/10.1534/g3.116.035220> (2016)
34. Bangera, R., Correa, K., Lhorente, J. P., Figueroa, R. & Yáñez, J. M. Genomic predictions can accelerate selection for resistance against *Piscirickettsia salmonis* in Atlantic salmon (*Salmo salar*). *BMC Genomics* **18**, 121; <https://doi.org/10.1186/s12864-017-3487-y> (2017)
35. Vallejo, R. L., Silva, R. M. O., Evenhuis, J. P., Gao, G., Liu, S., Parsons, J. E., Martin, K. E., Wiens, G. D., Lourenco, D. A. L., Leeds, T. D. & Palti, Y. Accurate genomic predictions for BCWD resistance in rainbow trout are achieved using low-density SNP panels: Evidence that long-range LD is a major contributing factor. *J. Anim. Breed. Genet.* **135**, 263-274; <https://doi.org/10.1111/jbg.12335> (2018)
36. Zeng, Q., Fu, Q., Li, Y., Waldbieser, G., Bosworth, B., Liu, S., Yang, Y., Bao, L., Yuan, Z., Li, N. & Liu, Z. Development of a 690 K SNP array in catfish and its application for genetic mapping and validation of the reference genome sequence. *Sci. Rep.* **7**, 1-14; <https://doi.org/10.1038/srep40347> (2017)

37. Nugent, C. M., Leong, J. S., Christensen, K. A., Rondeau, E. B., Brachmann, M. K., Easton, A. A., Ouellet-Fagg, C. L., Crown, M. T. T., Davidson, W. S., Koop, B. F., Danzmann, R. G. & Ferguson, M. M. Design and characterization of an 87k SNP genotyping array for Arctic charr (*Salvelinus alpinus*). *PLoS One* **14**, e0215008; <https://doi.org/10.1371/journal.pone.0215008> (2019)
38. Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.*, **17**, 81–92. <https://doi.org/10.1038/nrg.2015.28> (2016)
39. Aguiar, J. P., Gomes, P. F. F., Hamoy, I. G., dos Santos, E. B., Schneider, H. & Sampaio, I. 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. *Aquac. Res.* **49**, 1914-1925; <https://doi.org/10.1111/are.13647> (2018)
40. 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. Assessing genetic diversity for a pre-breeding program in *Piaractus mesopotamicus* by SNPs and SSRs. *Genes* **10**, 668; <https://dx.doi.org/10.3390%2Fgenes10090668> (2019)
41. Tsairidou, S., Hamilton, A., Robledo, D., Bron, J.E. & Houston, R. D. Optimizing low-cost genotyping and imputation strategies for genomic selection in Atlantic salmon. *G3-Genes Genom. Genet.* **10**, 581-590; <https://doi.org/10.1534/g3.119.400800> (2020)
42. Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A. & Johnson, E. A. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3**, e3376; <https://doi.org/10.1371/journal.pone.0003376> (2008)
43. Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7**, e37135; <https://doi.org/10.1371/journal.pone.0037135> (2012)
44. Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. Stacks: building and genotyping loci de novo from short-read sequences. *G3-Genes Genom. Genet.* **1**, 171-182; <https://doi.org/10.1534/g3.111.000240> (2011)
45. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J. & Sham, P. C. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559-575; <https://dx.doi.org/10.1086%2F519795> (2007)
46. Mastrochirico-Filho, V. A., Hata, M. E., Kuradomi, R. Y., Freitas, M. V., Ariede, R. B., Pinheiro, D. G., Robledo, D., Houston, R. & Hashimoto, D. T. Transcriptome Profiling of Pacu (*Piaractus mesopotamicus*) Challenged with Pathogenic *Aeromonas hydrophila*: Inference on Immune Gene Response. *Front. Genet.* **11**, 604; <https://doi.org/10.3389/fgene.2020.00604> (2020)
47. Gomes, F., Watanabe, L., Vianez, J., Nunes, M., Cardoso, J., Lima, C., Schneider, H. & Sampaio, I. Comparative analysis of the transcriptome of the Amazonian fish species *Colossoma macropomum* (tambaqui) and hybrid tambacu by next generation sequencing. *PLoS One* **14**, e0212755; <https://doi.org/10.1371/journal.pone.0212755> (2019)
48. Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. & Regev, A. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **29**, 644; <https://dx.doi.org/10.1038%2Fnbt.1883> (2011)
49. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357; <http://www.nature.com/doi/10.1038/nmeth.1923> (2012)
50. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659; <https://doi.org/10.1093/bioinformatics/btl158> (2006)

51. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & 1000 Genome Project Data Processing Subgroup. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**, 2078-2079; <https://doi.org/10.1093/bioinformatics/btp352> (2009)
52. Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R. & 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158; <https://doi.org/10.1093/bioinformatics/btr330> (2011)
53. Neph, S., Kuehn, M. S., Reynolds, A. P., Haugen, E., Thurman, R. E., Johnson, A. K., Rynes, E., Maurano, M. T., Vierstra, J., Thomas, S., Sandstrom, R., Humbert, R. & Stamatoyannopoulos, J. A. BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919-1920; <https://doi.org/10.1093/bioinformatics/bts277> (2012)
54. Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80-92. <https://doi.org/10.4161/fly.19695> (2012)
55. *NCBI Sequence Read Archive* <http://identifiers.org/ncbi/insdc.sra:SRP293925> (2020)
56. *NCBI Sequence Read Archive* <http://identifiers.org/ncbi/insdc.sra:SRP293156> (2020)
57. *NCBI Sequence Read Archive* <http://identifiers.org/ncbi/insdc.sra:SRP261699> (2020)
58. *NCBI Sequence Read Archive* <http://identifiers.org/ncbi/insdc.sra:SRP095431> (2020)

Acknowledgements

This work was supported by São Paulo Research Foundation (FAPESP grant 2016/21011-9, 2016/18294-9, 2017/26900-9, 2017/19717-3, 2018/08416-5, 2019/10662-7 and 2019/08972-8) providing a PhD fellowship for VAMF and funds for the sequencing service; Comisión Nacional de Investigación Científica y Tecnológica (International Call FAPESP-CONICYT) providing funds for the experiment costs; National Council for Scientific and Technological Development (CNPq grant 311559/2018-2, 422670/2018-9 and 140740/2016-3) providing a PhD scholarship and financing for the study design; and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES - Finance Code 001 and CAPES/PRINT) providing funds for project costs.

Author contributions

D.T.H. and J.M.Y. conceptualized, acquire funding and supervised this project; V.A.M.F. performed bioinformatics analysis and analysed the data with R.B.A. and C.H.S.B.; P.C collaborated with the construction of the SNP array; D.T.H., M.H.M.B. and G.A.L.S. coordinated the sample collection; R.B.A., M.V.F., V.A.M.F., L.V.G.L., N.J.M and J.F.G.A. performed DNA extraction and library construction; D.T.H. wrote the manuscript with V.A.M.F and R.B.A. collaboration; J.M.Y and F.P reviewed, approved and contributed with the language improvement of the manuscript. The authors declare that there is no conflict of interest.

Additional Information

Code Availability

Parameters of software involved in the analysis are described in the following paragraph.

FastQC: v0.11.8, standard parameters

RADseq methods:

1. *Stacks*: v. 2.0, $m=3$, $n=3$, $r=0.70$

2. *Plink*: v. 1.9, $--hwe 10^{-6} -me 0.1 0.9 --mind 0.3$

RNA-Seq method:

1. *Trimmomatic*: v.0.36, *LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:36*

2. *Trinity*: v.2.9.1, standard parameters *kmer-length* value of 25 and contiguous sequences (contigs) longer than 200 bp.

3. *BWA (Burrows-Wheeler Aligner)*: v. 0.7.17, *mem* algorithm

4. *SAMtools* programs: version 1.9, default parameters, *vcfutils.pl varFilter -d 10 -a 3*

Competing interests

The authors declare no competing financial interests

Figures

Figure 1. SNP distribution according to the MAF values of each species. 21,963 validated SNPs (74.17 %) in pacu (blue) and 21,072 validated SNPs (71.25 %) in tambaqui were considered. The validation of SNPs was carried out in 94 pacu and in 58 tambaqui individuals, respectively

Figure 2. DAPC analysis from hatcheries of pacu (Hatchery 1-4) (a) and tambaqui (Hatchery 5-9) (b) to demonstrate the genetic structure using the SNPs. DAPC was performed adopting an optimum number of principal components (PC = 10) calculated using the α -score function of the Aegenet software

Figure 3. Linkage disequilibrium (LD) by SNP pruning between markers among all the sampled pacu and tambaqui populations (Hatchery 1-9). 21,963 validated SNPs (74.17 %) in pacu (blue) and 21,072 validated SNPs (71.25 %) in tambaqui were considered

Tables

Table 1. SNPs classification according the Axiom quality control criteria (p-convert).

Category	Number of SNPs

	Pacu	Tambaqui
Recommended	99,682	42,851
Neutral	18,308	11,857
Not_recommended	12,413	27,140
Total	130,403	81,848

Table 2. SNP performance in the Axiom array, according to the method of SNP discovery in each species.

Species	Method	Putative SNPs	Selected SNPs	Polymorphic SNPs	Conversion rate	MAF
Pacu	RNA-Seq	80,004	8,052	5,842	0.72	0.254
	RADseq	43,355	17,932	13,345	0.74	0.175
	ddRADseq	7,044	3,628	2,776	0.76	0.160
	Total	130,403	29,612	21,963	0.74	0.203
Tambaqui	RNA-Seq	18,282	1,446	480	0.33	0.234
	GBS	6,803	2,200	1,888	0.85	0.237
	ddRADseq	56,763	25,929	18,704	0.72	0.260
	Total	81,848	29,575	21,072	0.71	0.247

Table 3. Summary of the SNP classification according to their performance in the Axiom SerraSNP array of each species.

Category	Pacu		Tambaqui	
	Number of markers	%	Number of markers	%
<i>PolyHighResolution</i>	16,780	56.67	19,661	66.48
<i>NoMinorHom</i>	5,183	17.50	1,411	4.77
<i>MonoHighResolution</i>	1,771	5.98	1,468	4.96
<i>CallRateBelowThreshold</i>	1,166	3.94	1,443	4.88
<i>OffTargetVariant</i>	570	1.92	550	1.86

<i>Other</i>	4,142	13.99	5,042	17.05
Total	29,612	100.00	29,575	100.00

Table 4. Descriptive population genetic estimates for the sampled pacu and tambaqui populations included in the validation of the array.

Species	Population	n	SNPs MAF > 0.01	Average MAF	H _o	H _e
Pacu	Hatchery 1	42	19,608	0.221	0.306	0.307
	Hatchery 2	21	18,461	0.214	0.307	0.294
	Hatchery 3	13	14,199	0.235	0.402	0.313
	Hatchery 4	18	19,856	0.210	0.285	0.292
	All populations	94	21,963	0.203	0.276	0.283
Tambaqui	Brazil - combined	28	20,380	0.266	0.362	0.356
	Hatchery 5	9	19,572	0.267	0.39	0.357
	Hatchery 6	9	17,172	0.268	0.426	0.352
	Hatchery 7	10	19,078	0.274	0.379	0.365
	Colombia - Hatchery 8	15	15,649	0.247	0.351	0.331
	Peru - Hatchery 9	15	15,075	0.248	0.353	0.329
	All populations	58	21,072	0.247	0.306	0.338
Pirapitinga	Wild	10	3,042	0.278	0.454	0.348

Table 5. Classification of SNPs from RNA-Seq included in the array, considering the annotation of 4,093 SNPs from pacu and 294 SNPs from tambaqui. Percentage values are between parenthesis

Category	Number of SNPs	
	Pacu (%)	Tambaqui (%)
Intergenic	2,047 (48.3)	208 (69.8)
UTR 3' prime	1,202 (28.4)	17 (5.7)
UTR 5' prime	284 (6.7)	14 (4.7)
Exon		
Synonymous SNP	490 (11.6)	59 (19.8)
Missense SNP	215 (5.0)	-
Total of effects	4,238 (100)	298 (100)

Table 6. Details about the sampling and methods used for SNP discovery for each species. *enzyme combination I (SphI and MluCI). **3 families used the enzyme combination I (SphI and MluCI) and 17 families used the combination II (NlaIII and MluCI).

Species	Source	Method	Samples	Origin
Tambaqui	Hatchery 1 - broodstock	ddRADseq*	11	South Brazil
	Hatchery 2 - broodstock	ddRADseq*	17	South Brazil
	Hatchery 3 - broodstock	ddRADseq*	17	South Brazil
	Hatchery 4 - broodstock	ddRADseq*	15	North Brazil
	Hatchery 5 - broodstock	ddRADseq*	18	North Brazil
	Hatchery 6 - broodstock	ddRADseq*	21	North Brazil
	Hatchery 7 - broodstock	ddRADseq*	15	Colombia
	Hatchery 8 - broodstock	ddRADseq*	16	Colombia
	Hatchery 9 - broodstock	ddRADseq*	31	Peru
	Hatchery 10 - broodstock	ddRADseq*	8	Peru
	Breeding nucleus (Caunesp) - broodstock	ddRADseq*	29	South Brazil
	Breeding nucleus (Caunesp) – 20 full-sib families	ddRADseq**	600	South Brazil
	Nunes <i>et al.</i> (2017) ¹⁶ - 1 full-sib family	GBS	124	North Brazil
	Ariede <i>et al.</i> (2018) ¹⁴ – pool of 10 random individuals	RNA-Seq	10	South Brazil
	Gomes <i>et al.</i> (2019) ⁴⁷ – pool of 8 random individuals	RNA-Seq	8	North Brazil
Pacu	Mastrochirico-Filho <i>et al.</i> (2020) ¹⁷ – 14 full-sib families	RADseq	400	South Brazil
	Breeding nucleus (Caunesp) – 3 full-sib families	ddRADseq*	100	South Brazil
	Mastrochirico-Filho <i>et al.</i> (2020) ⁴⁶ - 3 full-sib families	RNA-Seq	36	South Brazil

3.3 Capítulo III

Mapa de ligação para ancoragem do genoma em escala cromossômica e estudo de associação genômica ampla para resistência contra *Aeromonas hydrophila* em *Colossoma macropomum*

Raquel B. Ariede^a, Milena V. Freitas^a, Lieschen V.G. Lira^a, John F.G. Agudelo^a, Carolina H.S. Borges, Pablo Cáceres^b, José M. Yáñez^b, Diogo T. Hashimoto^{a*}

^a Universidade Estadual Paulista (Unesp), Centro de Aquicultura da Unesp, 14884-900 Jaboticabal, SP, Brasil

^b Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, Santiago, Chile

* Autor correspondente: Centro de Aquicultura da Unesp, Via de Acesso Prof. Paulo Donato Castellane, s/n , 14884-900 Jaboticabal/SP. Telefone: +55 16 3209-7477.

E-mail: diogo.hashimoto@unesp.br

Destaques do manuscrito:

- Desenvolvimento de um mapa de ligação denso para realizar análises de nível genômico em tambaqui.
- GWAS revelou que a arquitetura genética da resistência de *Aeromonas hydrophila* é poligênica e identificou vários QTLs putativos.
- Famílias de genes semelhantes do sistema imunológico relatadas em estudos anteriores sugerem sobreposição de sistemas biológicos para resistência a *Aeromonas hydrophila*.

Resumo

Tambaqui (*Colossoma macropomum*) é o peixe neotropical mais importante na aquicultura de água doce na América do Sul. Surto de doenças causadas pela bactéria *Aeromonas hydrophila* resultaram em perdas significativas para a produção de tambaqui. A reprodução seletiva usando informações genômicas é uma estratégia poderosa para melhorar a resistência a doenças em espécies de peixes. O primeiro passo para incorporar informações genômicas em programas de melhoramento é determinar o número de *loci* ou genes que podem afetar as características alvo. Neste estudo, investigamos a arquitetura genética de resistência a *A. hydrophila* em tambaqui por meio de um estudo de associação genômica ampla (GWAS) para identificar locos de características quantitativas (QTLs) e genes putativos associados a essa característica em tambaqui cultivado. Os genótipos de polimorfismo de nucleotídeo único (SNP) de 275 animais pertencentes a 18 famílias de irmãos completos foram obtidos usando a nova plataforma de genotipagem Axiom 30K SerraSNP array para tambaqui. A análise genômica revelou valores de herdabilidade baixos a moderados para ambas as definições de características de resistência: $h^2 = 0.15 (\pm 0.07)$ para teste de sobrevivência e $h^2 = 0.27 (\pm 0.10)$ para tempo de morte. Um mapa de ligação denso foi desenvolvido para tambaqui usando os dados do genótipo; resultando em 17.374 SNPs distribuídos em 27 grupos de ligação. O comprimento dos grupos de ligação variou de 124 (LG1) a 50 (LG14) cM, com um comprimento total de mapa integrado de 2.298,91 cM. A análise GWAS identificou vários QTLs putativos associados à resistência de *A. hydrophila*, distribuídos em seis grupos de ligação. Vários genes candidatos relacionados à resposta imune foram localizados próximos aos QTLs putativos, como *irf4a*, *ripk11*, *il20ra*, *c8a*, *c8b*, *ube3c* e *ccr9a*. Este estudo representa a primeira aplicação de uma matriz SNP de 30K para construir um mapa de ligação denso e identificar variantes genéticas associadas à resistência de *A. hydrophila* em tambaqui.

Palavras-chave: Motile *Aeromonas* septicemia, GWAS, QTL, mapa de ligação, herdabilidade

3.3.1 Introdução

Tambaqui (*Colossoma macropomum*) é considerado um peixe onívoro, com preferência por comportamento frugívoro; diferente das conhecidas piranhas, uma espécie de Serrasalminidae relacionada com comportamento predatório [1]. Esta espécie apresenta alta rusticidade, crescimento rápido, aceitação de rações artificiais, alta produtividade e valor comercial para mercados internacionais [2]. Tambaqui é o principal peixe nativo cultivado no Brasil, principalmente na região Norte do país. A produção aquícola de tambaqui atingiu 136 mil toneladas em 2015 [3]. Tambaqui tem características desejáveis para o desenvolvimento da aquicultura na América Latina e também é cultivado em países como Bolívia, Brasil, Colômbia, Peru, Equador e Venezuela [2].

Os surtos de doenças são responsáveis por perdas de bilhões de dólares para a aquicultura todos os anos, e as infecções bacterianas podem ser uma grande ameaça à sustentabilidade da produção de peixes cultivados [4, 5]. Devido à intensificação na produção de tambaqui e altas densidades de estocagem nas condições de cultivo, importantes patógenos têm sido identificados nestes peixes, que têm o potencial de causar perdas significativas de produção, por exemplo, surtos de doenças causadas por *Aeromonas hydrophila* [2]. A septicemia por *Aeromonas* é uma doença comum na aquicultura causada por *Aeromonas hydrophila*, um patógeno bacteriano gram-negativo, que geralmente está associado a nadadeiras avermelhadas ou corroídas, septicemia externa/interna e hemorragia [6].

Embora praticamente toda a produção de peixes neotropicais ainda seja baseada em estoques não selecionados, abordagens genéticas para melhorar a resistência a doenças infecciosas têm destacado a possibilidade de controlar surtos em espécies de aquicultura, como carpa [7, 8, 9], tilápia [10] e salmonídeos [11,12,13]. O melhoramento

seletivo para resistência a doenças é um desafio porque essa característica é difícil de medir em candidatos à seleção [14, 15]. Além disso, embora os principais QTL que afetam a resistência a doenças virais tenham sido identificados [16, 17], a resistência bacteriana normalmente exibe uma arquitetura poligênica ou oligogênica [18, 19].

Estudos de associação genômica ampla (GWAS) investigaram a arquitetura genética da resistência contra *A. hydrophila* em algumas espécies de peixes, incluindo a carpa *Labeo rohita* [20], pacu *Piaractus mesopotamicus* [18] e híbridos de bagres entre espécies de *Ictalurus* [19]. Nesses estudos, vários loci de características quantitativas (QTLs) foram identificados como significativamente associados à resistência contra *A. hydrophila*, e estão próximos de genes candidatos relacionados à resposta imune.

O desafio experimental contra *A. hydrophila* em tambaqui, mostrou anteriormente que há variação genética significativa para resistência a esta bactéria; demonstrando o potencial de melhoramento genético da resistência à aeromoniose nesta espécie [21, 22]. No entanto, a base genômica da resistência a *A. hydrophila* ainda não foi identificada em tambaqui. Painéis de SNPs densos contendo dezenas de milhares de SNPs são necessários para realizar GWAS para uma melhor compreensão dos loci ou genes que afetam as características de resistência [23, 24]. Recentemente, um array de genotipagem Axiom SNP multi-espécies foi desenvolvido para as espécies de serrasalmídeos pacu e tambaqui (array SerraSnp 60K), contendo 30K marcadores SNP para cada espécie. Esta matriz SNP está agora disponível para dissecar características complexas de importância econômica e aplicar a seleção genômica em programas de melhoramento aquícola de ambas as espécies [59].

O presente estudo teve como objetivo i) desenvolver um mapa de ligação denso em tambaqui usando os SNPs de 30K da matriz de genotipagem SerraSnp Axiom, e ii) realizar um GWAS para resistência contra *A. hydrophila* em 18 famílias de irmãos

completos de tambaqui. O objetivo foi desvendar a arquitetura genética da resistência de *A. hydrophila* e identificar genes candidatos responsáveis pela variação dos mecanismos de defesa a esse patógeno. Os resultados do estudo são fundamentais para acelerar o melhoramento genético para resistência de *A. hydrophila* através da seleção genômica em tambaqui.

3.3.2 Material e Métodos

3.3.2.1 Comitê de Ética

Este estudo foi conduzido em estrita conformidade com as recomendações do Conselho Nacional de Controle de Experimentação Animal (CONCEA) (Ministério da Ciência, Tecnologia e Inovação) e foi aprovado pelo Comitê de Ética no Uso de Animais (CEUA número 019006/17) da Faculdade de Ciências Agrárias e Veterinárias, UNESP, Campus Jaboticabal, SP, Brasil.

3.3.2.2 Material Biológico

A população utilizada neste estudo é a população reprodutora mantida no Centro de Aquicultura da Universidade Estadual Paulista (UNESP), município de Jaboticabal (Estado de São Paulo, Brasil). Testes experimentais foram conduzidos no Laboratório de Genética em Aquicultura e Conservação (LaGeAC, UNESP, Brasil). A população foi composta por 576 peixes pertencentes a 18 famílias de irmãos completos gerados a partir de 9 fêmeas e 14 machos. Mais detalhes sobre a descrição das condições de cultivo e manejo populacional podem ser encontrados em Ariede *et al.* [22].

3.3.2.3 Desafio experimental

Um protocolo passo a passo do desafio experimental foi descrito em Ariede *et al.* [22]. Resumidamente, os peixes foram distribuídos em três tanques de 0,5 m³ para o teste de desafio. Cada tanque foi mantido com um sistema de recirculação de água independente. No total, 576 peixes foram usados no experimento de desafio bacteriano. Aproximadamente 190 peixes foram colocados em cada tanque, com aproximadamente dez indivíduos de cada família por tanque. Os peixes foram inoculados por injeção intraperitoneal de uma dose letal 50 predefinida (DL50) de células vivas de *A. hydrophila* ($3,4 \times 10^5$ UFC/g de peso corporal). A mortalidade dos peixes foi observada durante todo o dia (24 h) nos três dias iniciais do desafio; e em intervalos de 8 h nos demais dias de desafio. Indivíduos mortos apresentando sinais clínicos de *A. hydrophila* foram registrados e retirados imediatamente dos tanques. Os dados fenotípicos de resistência a *A. hydrophila* e amostras de nadadeira foram coletadas de todos os peixes e armazenadas em etanol 98% a -20°C até a extração do DNA genômico.

3.3.2.4 Genotipagem dos SNPs

A estratégia de genotipagem teve como objetivo maximizar a variância fenotípica dentro da população, mantendo uma representação equilibrada de peixes por família. A genotipagem não foi totalmente aleatória, mas focada em selecionar a mesma quantidade de peixes suscetíveis e resistentes das diferentes famílias; tentando manter um número semelhante de animais por família. Cada família foi representada com uma média de 13 (DP=2) peixes/família.

O DNA genômico foi extraído de amostras de sangue de 280 peixes usando o kit comercial Wizard Genomic, Promega, seguindo as instruções do fabricante. Os peixes

foram genotipados usando o array SerraSNP Affymetrix® Axiom®, projetado por nosso grupo de pesquisa e disponível comercialmente [59]. As amostras de DNA genômico foram enviadas para a Thermo Fisher Scientific (EUA, Califórnia) para genotipagem.

Os genótipos passaram por controle de qualidade usando o *Software AXIOM Analysis Suite da Affymetrix* usando as configurações padrão. Os dados brutos contendo os resultados dos cálculos de intensidade (arquivos CEL) foram importados para o *Axiom Analysis Suite* (v2.0.035, Affymetrix) para análise de controle de qualidade e identificação de genótipos. Amostras com um valor de controle de qualidade (DQC) > 0,82 e um call rate QC > 0,97 (segundo o “*Best Practices Workflow*” recomendado pela Affymetrix), foram consideradas aprovadas na avaliação inicial de controle de qualidade. Etapas adicionais de controle de qualidade foram realizadas filtrando SNPs por frequência de alelos menores (MAF) inferior a 0,05 e a taxa de erro mendeliana de 0,01 no programa Plink [25].

3.3.2.5 Mapa de ligação e Análise de sintenia

Um mapa de ligação foi criado usando Lep-MAP3 [26]. O módulo *ParentCall2* foi usado para imputar possíveis genótipos ausentes ou para corrigir genótipos parentais errôneos com base em dados de progênie. O módulo *Filtering2* foi usado para remover marcadores com distorção de segregação significativa (tolerância de dados = 0,001). Marcadores foram atribuídos a grupos de ligação (GL) por *SeparateChromosomes2*, testando vários lodLimit e ajustados para 27 grupos de ligação (correspondentes aos 27 pares de cromossomos do tabaqui) [27]. O valor de LOD foi selecionado aumentando o valor de LOD de um em um, sem parâmetro de marcador mínimo até que o número de GL se estabilizasse. Posteriormente, marcadores órfãos foram atribuídos a grupos de ligação existentes (pontuação LOD menor que o *SeparateChromosomes2*) usando

JoinSingles2 e ordenados dentro de grupos de ligação usando o módulo *OrderMarkers2*. Devido à pequena variação estocástica nas distâncias dos marcadores entre as corridas, o módulo *OrderMarkers2* foi repetido 10 vezes em cada grupo de ligação para produzir o mapa de ligação. O mapa de ligação gerado foi desenhado usando *LinkageMapView* [28].

A confiabilidade da atribuição de loci SNP no GL e a respectiva ordenação de loci dentro dos GLs foi verificada por análise de sintonia entre os cromossomos de *Astyanax mexicanus* e os GLs de tambaqui, usando o software Circa (<http://omgenomics.com/circa/>). O genoma de *A. mexicanus* foi utilizado como referência por ser a espécie filogeneticamente mais próxima com um genoma de referência disponível à nível cromossômico.

3.3.2.6 Métodos PBLUP e ssGBLUP

As características são definidas como em Ariede *et al.* [22]. Resumidamente, a resistência foi avaliada como sobrevivência ao teste de desafio usando Teste de Sobrevivência (TS, sobrevivência binária) e Tempo de Morte (TD, avaliado em horas). Os componentes de variância e valores genéticos estimados (EBVs) foram inicialmente avaliados usando o método convencional baseado em pedigree (PBLUP), de acordo com Ariede *et al.* [22].

Os componentes de variância e valores genéticos genômicos (GEBV) foram estimados usando o método de avaliação genética de etapa única (ssGBLUP) [29]. A matriz de parentesco \mathbf{H} foi utilizada [30], na qual os dados de genótipo e pedigree são combinados. A matriz \mathbf{H} inversa é:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

Onde \mathbf{A}^{-1} é o inverso do numerador da matriz de relacionamento para todos os indivíduos incluídos no pedigree, \mathbf{A}_{22}^{-1} é o inverso da matriz de relacionamento baseada em pedigree para indivíduos genotipados, e \mathbf{G}^{-1} é o inverso da matriz de relacionamento genômico.

O TD foi analisado como um modelo linear usando AIREMLF90 da família de programas BLUPF90. O TS foi analisado inicialmente como um modelo de limiar usando THRGIBBS1F90 da família de programas BLUPF90 [31], e então, usando um modelo linear usando AIREMLF90 de BLUPF90. A significância estatística dos SNPs foi avaliada ajustando o valor de p com a correção de Bonferroni. Os SNPs foram considerados significativos no genoma quando excederam o limite de Bonferroni para testes múltiplos ($\alpha = 0,05$) de $0,05/tg$, onde tg é o número total de SNPs usados no GWAS, e significativo em todo o cromossomo quando o limite de Bonferroni para testes múltiplos ultrapassou ($\alpha = 0,05$) $0,05/tc$, onde tc é o número médio de SNPs por cromossomo. Os gráficos de Manhattan foram desenhados com base em p -values de SNP individuais distribuídos em diferentes grupos de ligação. Cada locus contendo SNPs associados foi alinhado contra o genoma em scaffolds do *Colossoma macropomum* (acesso ao conjunto NCBI RefSeq: GCF_904425465.1) usando a ferramenta BLASTn para investigar genes candidatos (dentro da região de 500Kb) em torno de SNPs associados à resistência de *A. hydrophila*.

3.3.3 Resultados

3.3.3.1 Desafio experimental

A mortalidade total foi de 43,5%, variando de 26 a 89% da família mais resistente para a menos resistente. O tempo médio de mortes durante a infecção por *A. hydrophila* foi de 32,2 h (DP = 23,2), variando de 10,7 h a 69,2 h entre as famílias.

3.3.3.2 Genotipagem dos SNPs e controle de qualidade

Após a análise de CQ recomendada pela Affymetrix (*Best Practices Workflow*), um total de 22.540 SNPs e 275 indivíduos foram usados para o mapa de ligação. A taxa de erro mendeliana excluiu 486 SNPs e 3 indivíduos. Finalmente, a filtragem MAF eliminou 3.229 SNPs, o que resultou em um total de 18.825 SNPs usados para GWAS.

3.3.3.3 Mapa de ligação e Análise de sintonia

Após a análise de possíveis genótipos parentais ausentes ou errôneos no módulo *ParentalCall*, um total de 16.590 SNPs foram atribuídos a grupos de ligação (GL) de acordo com as características cariotípicas da espécie (tamanho relativo do cromossomo e número diplóide) em LOD 19. Os marcadores restantes foram atribuídos ao GL existente no LOD 16, que recuperou 784 marcadores. 1.451 marcadores foram descartados porque nenhuma associação com o mapa de ligação foi detectada. A ordem dos marcadores com a melhor probabilidade para cada GL foi combinada para produzir o mapa de ligação final. Um total de 17.374 SNPs foram atribuídos a 27 grupos de ligação (Fig.1), correspondendo aos 27 pares de cromossomos ($2n = 54$) descritos em tambaqui [32].

O número de SNP no GL variou de 1.106 (GL1) a 444 (GL27). O comprimento dos GLs do mapa sex-average variou de 124,93 (GL1) a 50,29 (GL27) cM, com um comprimento de 2.298,91 cM. Uma alta correlação foi encontrada entre o comprimento e o número de marcadores por GL (correlação de Pearson, $r^2=0,83$). A distância média entre os marcadores foi de 0,13 cM (DP = 0,017). A maior e a menor densidade de marcadores foi encontrada em GL3 e GL25, com média de 0,10 e 0,17 cM, respectivamente (Tabela 1).

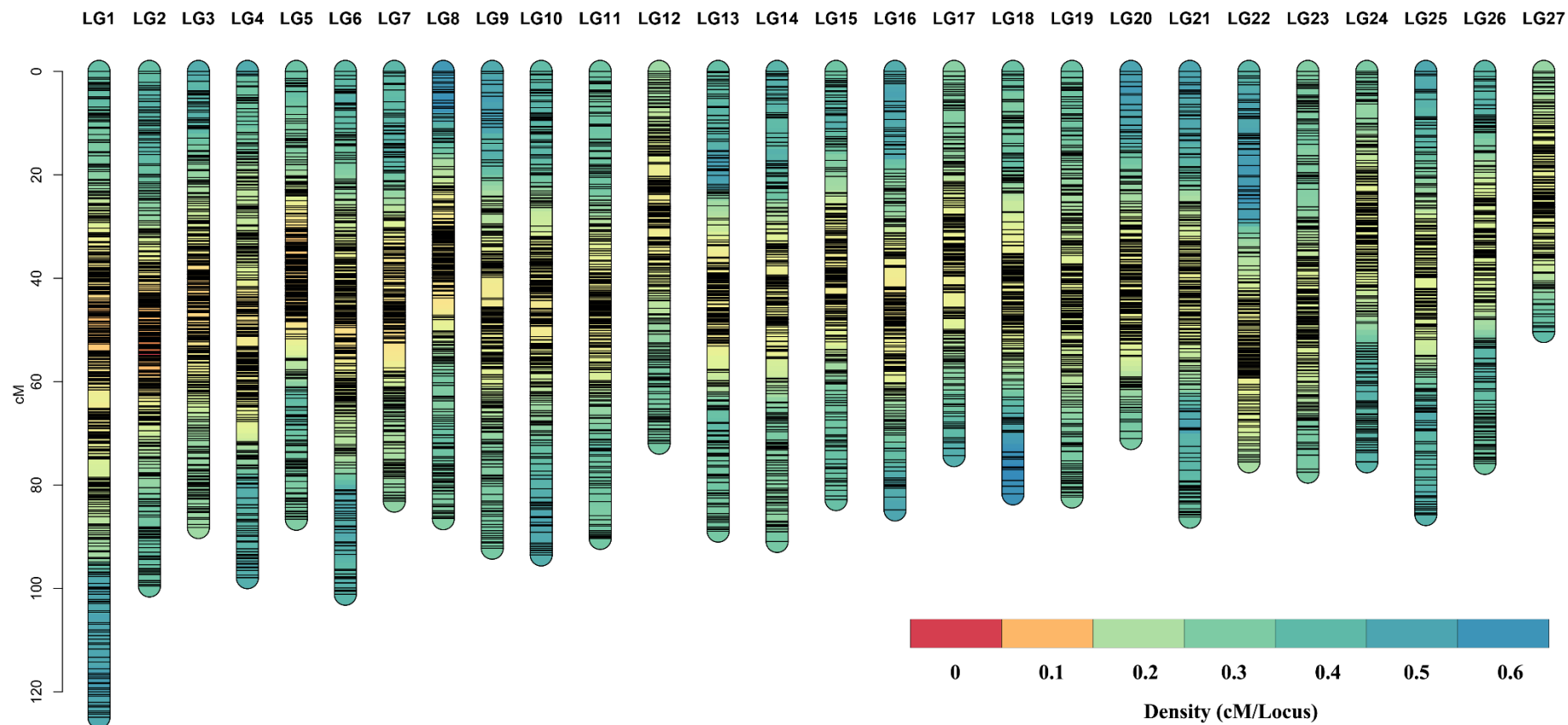


Fig 1. Mapa de ligação sex-average de tambaqui *Colossoma macropomum* mostrando os 27 grupos de ligação construídos usando 17.374 SNPs. As cores representam uma escala de densidade de marcadores que variam de regiões de baixa densidade (azul) a regiões de alta densidade (vermelho).

Tabela 1. Resumo do número de marcadores e comprimento do mapa sex-average e densidade dos 27 grupos de ligação em tabaqui.

	Número de marcadores	sex- average	
		Comprimento (cM)	Densidade (SNP/cM)
GL1	1106	124,932	0.112958
GL2	877	99,5	0.113455
GL3	811	88,165	0.108711
GL4	799	97,877	0.122499
GL5	772	86,566	0.112132
GL6	772	101,143	0.131014
GL7	737	83,052	0.112689
GL8	684	86,53	0.126506
GL9	675	92,153	0.136523
GL10	650	93,523	0.143882
GL11	647	90,336	0.139623
GL12	629	71,91	0.114324
GL13	629	88,892	0.141323
GL14	631	90,916	0.144082
GL15	611	82,802	0.135519
GL16	623	84,83	0.136164
GL17	582	74,284	0.127636
GL18	555	81,741	0.147281
GL19	560	82,271	0.146913
GL20	530	71,014	0.133989
GL21	532	86,175	0.161983
GL22	531	75,539	0.142258
GL23	517	77,459	0.149824
GL24	513	75,474	0.147123
GL25	501	85,692	0.171042
GL26	456	75,843	0.166322
GL27	444	50,296	0.113279
Total	17,374	2,298.915	0.132319

Sintenia genômica conservada entre os genomas do tabaqui e do peixe-cego (*Astyanax mexicanus*) foi detectada usando os 17.374 loci dos SNPs mapeados. Alinhamentos significativos do BLASTn contra o genoma do peixe-cego foram comparados apenas a 5.799 loci mapeados (33,4%), onde 5.080 foram atribuídos aos 25 cromossomos e 799 aos scaffolds não montados nos cromossomos de *A. mexicanus*. Apesar da pequena porcentagem de alinhamentos ao genoma de peixe-cego, um alto nível

de sintenia genômica foi observado entre o tambaqui e esta espécie de referência. O número de segmentos alinhados para cada cromossomo de *A. mexicanus* variou de 463 (Am3) a 50 (Am11), o que foi altamente correlacionado com os tamanhos dos cromossomos (correlação de Pearson, $r^2=0,94$). Uma relação de 1:1 foi detectada em pelo menos 20 GLs (77,7%). GL1 e GL7 foram separados em dois cromossomos, Am7/Am8 e Am10/Am11, respectivamente. Dois GLs (GL4 e GL20) foram fundidos em um único cromossomo (Am3). A sintenia do genoma não foi conclusiva para três GLs (GL24, GL26 e GL27), mostrando uma distribuição dispersa para vários cromossomos de *A. mexicanus* (Fig. 2).

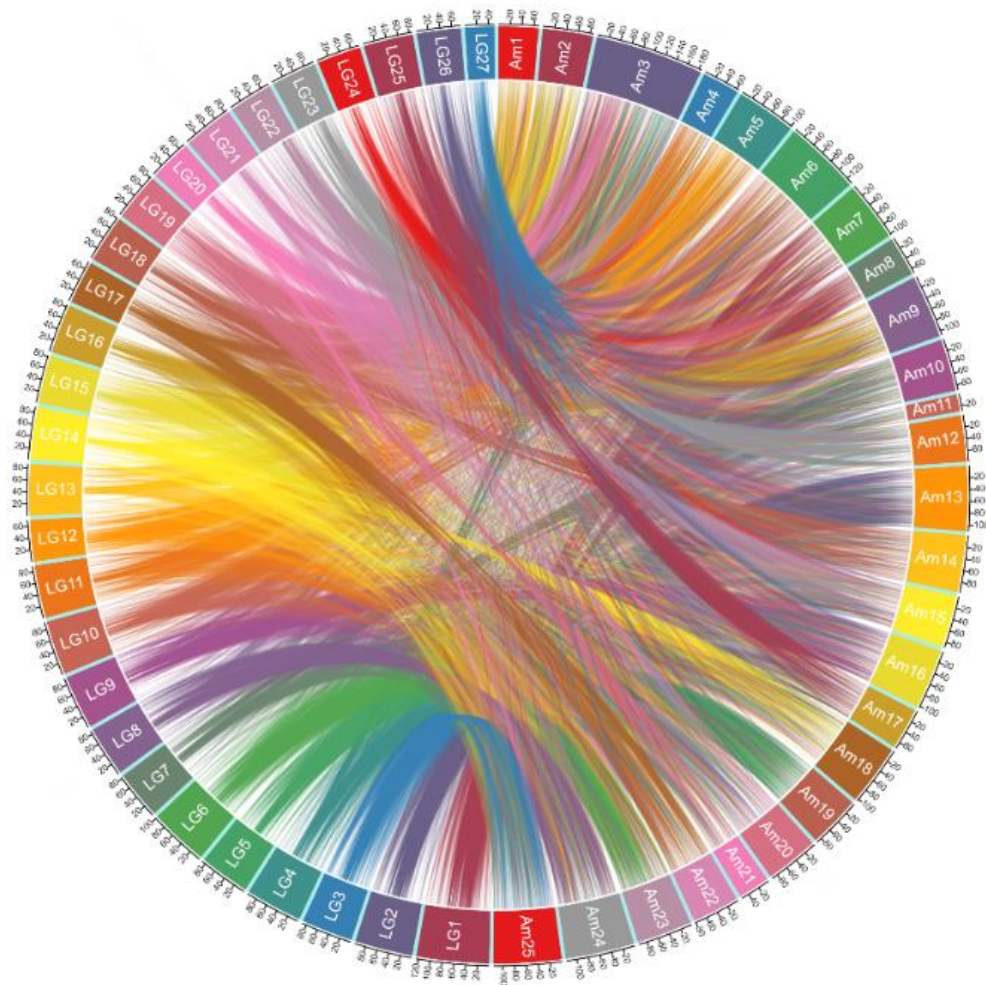


Fig 2. Sintenia genômica entre os grupos de ligação do tambaqui *Colossoma macropomum* (GL) e os cromossomos da espécie de peixe-cego (*Astyanax mexicanus*) (Am).

3.3.3.4 Herdabilidade e Análise de associação genômica ampla (GWAS)

No total, 19.404 SNPs e 275 animais genotipados passaram pelo controle de qualidade (call rate > 97%; MAF > 0.05), e 301 indivíduos com apenas fenótipos foram analisados no GWAS. As herdabilidades estimadas com o método ssGBLUP foram semelhantes (TD = 0.27 ± 0.10 e TS = 0.15 ± 0.07) aos obtidos com PBLUP e resultados previamente relatados em Ariede *et al.* [22] (TD = 0.23 ± 0.09 e TS = 0.17 ± 0.06) (Tabela 2). Um total de 9 SNPs com associação significativa em nível genômico à resistência de *A. hydrophila* foram detectados para TD em 2 GLs (GLs 4 e 9) e um scaffold não atribuído a nenhum GL. Foram identificados 14 SNPs com associação significativa em nível cromossômico, dos quais apenas 2 estavam associados a TS (GLs 8 e 9), apenas 10 associados a TD (GLs 4, 5, 9 e 13) e 2 estavam associados a ambas as características (GL10) (Tabela 3, Fig. 3). Os QTLs significativos tanto em nível cromossômico quanto genômico explicaram uma baixa proporção da variância para TD e TS, com valores de aproximadamente 5% e 1%, respectivamente. A análise de genes candidatos usando os SNPs mais associados contra o genoma de *Colossoma macropomum* revelou vários genes do sistema imunológico localizados próximos aos QTLs identificados (Tabela 3).

Tabela 2. Estimativas genômicas de variância genética aditiva (σ_a^2), variância residual (σ_e^2), e herdabilidade (h^2) para resistência a *A. hydrophila* em tambaqui *Colossoma macropomum*, mensuradas em tempo de morte (TD) e teste de sobrevivência (TS). Erro padrão em parênteses.

Característica	σ_a^2	σ_e^2	h^2
TD	329.01 (140.88)	872.21 (91.456)	0.27390 (0.10448)
TS	0.35744E-01 (0.18605E-01)	0.19976 (0.16070E-01)	0.15178 (0.75099E-01)

Tabela 3. Marcadores SNP associados à resistência a *Aeromonas hydrophila* medidos como tempo de morte (TD) e teste de sobrevivência (TS) em tambaqui. A localização e posição são em relação ao genoma de referência do tambaqui *Colossoma macropomum*. Os genes candidatos do sistema imunológico estão localizados dentro da janela de 1 Mb.

SNP	Posição (pb)	Variância genética %	-Log ₁₀ (P-value)	Nível de Significância	Característica	GL	Localização (scaffold)	Genes
AX-508803570	7,771,015	0.42	7.32	Genome	TD	9	NW_023494805.1	phlpp1, ripk11, irf4a, dusp22b, igfbp1b, cldn1
AX-508812066	7,697,873	0.26	6.54	Genome				
AX-508790544	7,744,214	0.07	5.71	Genome				
AX-508807585	6,690,346	0.29	6.72	Genome	TD	9	NW_023494805.1	c8a, c8b, ube3c
AX-508791788	8,376,130	0.18	6.62	Genome	TD	9	NW_023494805.1	ripk11, irf4a, dusp22b, wrnip1
AX-509092322	218,735	0.19	5.79	Genome	TD	4	NW_023495946.1	il20ra
AX-509110675	219,872	0.15	4.26	Chromosome				
AX-508831435	73,422	0.09	5.74	Genome	TD	n/a	NW_023495411.1	shrprbck1r
AX-508798776	4,863,528	0.37	5.71	Genome	TD	9	NW_023494805.1	dnajb6b
AX-509111803	4,896,188	0.04	4.43	Chromosome				
AX-509123997	5,285,665	0.11	6.19	Genome				
AX-508808864	12,902,387	0.41/0.46	5.31	Chromosome	TD/TS	10	NW_023494793.1	klhl22, cabin1, ido1
AX-509117671	17,850,572	0.27	5.02	Chromosome	TD	9	NW_023494805.1	bag1, yes1, colec12
AX-508786584	40,748,603	0.19	5.00	Chromosome	TD	5	NW_023494784.1	ptpn4a, dnajc3a
AX-508811089	8,958,752	0.42	4.56	Chromosome	TD	9	NW_023494805.1	wrnip1
AX-509145981	8,849,858	0.12	4.34	Chromosome				
AX-509133856	12,739,948	0.17	4.43	Chromosome	TD	4	NW_023494795.1	kita, exoc1, clocka, pdgfra
AX-508789842	9,972,154	0.29/0.35	4.37	Chromosome	TD/TS	10	NW_023494793.1	dusp2, prf1.5
AX-509109805	20,475,802	0.32	4.27	Chromosome	TD	9	NW_023494805.1	klhl20, eif2b3
AX-508786632	6,407,628	0.22	4.19	Chromosome	TD	13	NW_023494786.1	-
AX-508804703	19,193,829	0.35	4.16	Chromosome	TD	9	NW_023494805.1	abcf2a, klhl18, ccr9a
AX-509091650	23,050,823	0.04	4.44	Chromosome	TS	9	NW_023494805.1	cldn15a, casp23, plaa, tkfc, dnajc3b, slc7a7, rbpjb
AX-508827488	19,734,527	0.24	4.17	Chromosome	TS	8	NW_023494788.1	ppid, etfdh, rxfp1

* n/a significa não atribuído a um grupo de ligação.

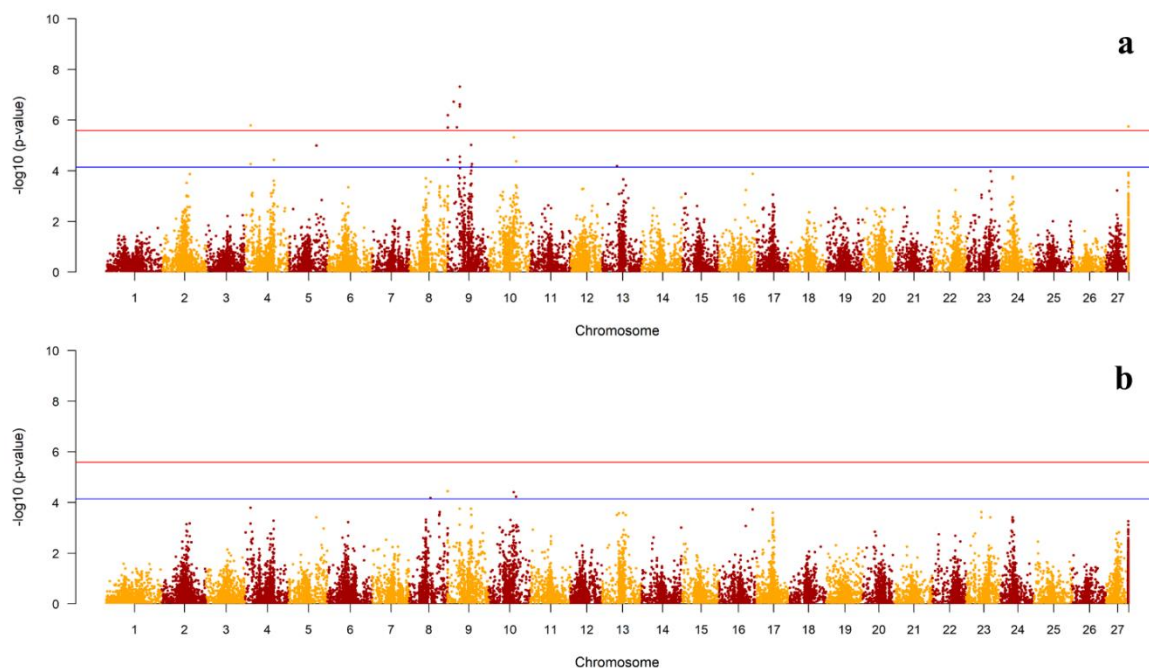


Fig 3. Estudo de associação genômica ampla (GWAS) para tempo de morte (TD) e teste de sobrevivência (TS) contra a infecção por *A. hydrophila* em tambaqui por ssGBLUP. Gráfico de Manhattan com $-\log_{10}(\text{p-values})$ distribuídos em diferentes grupos de ligação. Marcadores que ultrapassaram a significância do genoma (linha vermelha) e/ou do cromossomo (linha azul) após a correção de Bonferroni foram considerados em associação com a resistência. Os marcadores não atribuídos a um grupo de ligação foram agrupados no lado direito do gráfico.

3.3.4 Discussão

3.3.4.1 Mapa de ligação

Neste estudo desenvolvemos o mapa genético de maior densidade disponível até agora em tambaqui, que foi composto por 17.374 loci e um intervalo médio de marcadores de 0,13 cM. Este mapa teve maior resolução do que o primeiro mapa de ligação produzido em tambaqui, que utilizou informações de 7.734 SNPs em uma família de irmãos completos para obter um comprimento de mapa de 2.811 cM, com intervalo de marcador médio de 0,39 cM [33]. Aqui, utilizamos o genoma do peixe-cego (*Astyanax mexicanus*) como genoma de referência para realizar a análise de sintenia em vez do peixe-zebra, pois o primeiro é mais filogeneticamente relacionado ao tambaqui do que o segundo [34, 35]. O resultado foi uma homologia considerável entre os GLs e cromossomos de ambas as espécies. Uma sintenia conservada na maioria dos GLs foi identificada entre os genomas

de tabaqui e *A. mexicanus* (Fig. 2), e relações sintênicas fracas foram encontradas em 3 GLs. Portanto, o mapa de ligação resultante com sintonia conservada ao genoma de *A. mexicanus* oferecerá uma estrutura para o mapeamento de genes candidatos responsáveis por características econômicas a serem incluídas em programas de melhoramento desta espécie.

3.3.4.2 Herdabilidade e Análise de associação genômica ampla (GWAS)

Em geral, vários estudos para diferentes características demonstraram que as avaliações genômicas de uma única etapa fornecem uma previsão mais precisa do mérito genético do que o método clássico PBLUP em espécies de aquicultura [24, 36, 37]. Os valores baixos/moderados de herdabilidade em tabaqui, estimados pelo método ssGBLUP, concordam com resultados anteriores realizados na mesma população usando estimativas baseadas em pedigree para resistência contra *A. hydrophila* [22]. A pequena diferença nos valores de herdabilidade ao comparar PBLUP com ssGBLUP pode ter sido influenciada pela baixa quantidade de animais genotipados ($n = 275$). Em geral, nossos resultados ficaram dentro do intervalo daqueles relatados anteriormente para resistência de *A. hydrophila* em outras espécies de peixes (0,12 a 0,39), como *Clarias macrocephalus*, *Megalobrama amblycephala*, *Labeo rohita* e *Piaractus mesopotamicus* [21, 38, 39, 40].

GWAS revelou um padrão de arquitetura poligênica para resistência contra *A. hydrophila* em tabaqui, que foi semelhante ao relatado em outras espécies de peixes desafiadas contra *A. hydrophila* [18, 20] e para outras doenças bacterianas em outros peixes [41, 42]. A seleção assistida por marcadores foi proposta como um método viável para o cultivo de bagres devido ao baixo número de QTLs explicando uma grande variação fenotípica (~20%) envolvida na resistência de *A. hydrophila* (3 QTLs apresentados em 3 grupos de ligação) [19]. No entanto, esta não é uma estratégia eficaz

para melhorar geneticamente a resistência contra *A. hydrophila* em tambaqui devido à ausência de QTLs de grande efeito. Este padrão é semelhante a um estudo prévio de mapeamento de QTL que associou vários SNPs explicando uma baixa variância genética para resistência de *A. hydrophila* em pacu [18]. Portanto, a aplicação da seleção genômica [43] pode ser a estratégia mais adequada para incorporar informações genômicas para acelerar a reprodução seletiva para resistência de *A. hydrophila* em tambaqui, como tem sido demonstrado em espécies de salmonídeos [11, 13, 24, 44, 45]. A acurácia dos valores genéticos de avaliações genéticas baseadas em genômica deve ser comparada à seleção tradicional baseada em pedigree em estudos futuros.

A avaliação de genes candidatos próximos aos principais SNPs revelou vários genes relacionados à resposta imune, particularmente aqueles de significância genômica, que provavelmente podem estar envolvidos na resistência à infecção por *A. hydrophila* (Tabela 3). Algumas dessas famílias de genes (*dusp*, sistema complemento, *casp*, quimiocinas e perforina) também foram relatadas em outros estudos que exploraram os mecanismos do sistema imunológico de resistência de peixes contra *A. hydrophila* [18, 19]. Esses resultados sugerem que mecanismos em comum estão envolvidos na defesa bacteriana de diferentes espécies de peixes, como também relatado em espécies de salmonídeos [46]. Por exemplo, dual-specificity phosphatases (*dusp2* e *dusp22*, próximas ao AX-508789842, AX-508803570, AX-508791788, AX-508812066, e AX-508790544) são um subconjunto de proteínas tirosina fosfatases que pertencem à família MAPK que regulam a expressão gênica envolvida na resposta imune a alguns patógenos [47]. A via MAPK protege o hospedeiro contra a infecção por *A. hydrophila* em *Pseudosciaena crocea* [48]. Além disso, próximo ao AX-508789842 está a proteína formadora de poros (*prf1.5*), que é uma molécula envolvida na citólise mediada por células T e natural killer, inserindo-se na membrana alvo formando um poro que permite que as proteínas citolíticas

entrem na célula e desencadeiem a autodestruição. A expressão diferencial do gene da perforina tem sido sugerida como um papel importante na defesa imunológica e resistência contra a infecção por *A. hydrophila* em *L. rohita* [20].

O sistema imunológico inato é um mecanismo de defesa fundamental em peixes. Essa resposta é dividida em barreiras físicas e resposta imune celular e humoral, e é modulada por várias moléculas como quimiocinas, interleucinas, interferons e fatores do complemento [49]. Algumas dessas moléculas foram detectadas na análise GWAS para resistência a *A. hydrophila* em tambaqui. Quimiocinas e receptores de quimiocinas (como o *ccr9a*, próximo a AX-508804703) estão envolvidos na estimulação do recrutamento, ativação e adesão de células a locais de infecção ou lesão [50]. As interleucinas (*il20ra*, próximas a AX-509092322 e AX-509110675) e interferons (*irf4a*, próximas a AX-508803570, AX-508791788, AX-508812066, AX-508790544) pertencem a uma grande classe de proteínas conhecidas como citocinas, que desencadeiam as defesas protetoras do sistema imunológico para ajudar a erradicar patógenos [51]. A ativação das cascatas do complemento resulta na via citolítica envolvendo o complexo de ataque à membrana (MAC), que é montado pelos componentes do complemento C5b, C6, C7, C8 (detectado próximo ao SNP AX-508807585) e múltiplas moléculas C9 [52].

Outros genes candidatos envolvidos na resposta imune do hospedeiro a *A. hydrophila*, incluem *ube3c* (próximo ao AX-508807585) que atua no processo de ubiquitinação, que serve como um sinal que desencadeia a ativação de NF- κ B por indução de inflamação para inibir a proliferação bacteriana [53]. O NF- κ B é um fator de transcrição que desempenha papéis importantes em muitos processos biológicos, incluindo imunidade inata e adaptativa [54, 55]. Vários genes relacionados a esta via foram encontrados no presente estudo (*ripk11*, *dusp22b*, *rbck1r*, *clocka* e *casp23*), semelhantemente identificado em pesquisa realizada para resistência de *A. hydrophila* em

híbridos de bagres por GWAS [19]. Vários genes da via de sinalização NF- κ B também foram regulados positivamente em *Pseudosciaena crocea* e *Megalobrama amblycephala* infectados com *A. hydrophila* [48, 56], sugerindo uma contribuição para as respostas de defesa do hospedeiro para diferentes espécies de peixes.

As claudinas são importantes componentes funcionais e estruturais das junções oclusivas, que são estruturas de contato célula-célula que formam barreiras teciduais e poros para evitar a passagem descontrolada de moléculas e íons através do espaço intercelular [57]. A expressão gênica das claudinas foi regulada significativamente na resposta inflamatória intestinal e do baço contra a infecção por *A. hydrophila* em carpa comum [49, 58]. No presente estudo, o gene *cldn1* (próximo ao SNP AX-508812066, AX-508790544, AX-509091650) apresentou associação significativa com a resistência do hospedeiro contra *A. hydrophila* em tambaqui; sugerindo que a família de genes claudina está envolvida em respostas imunes a infecções bacterianas em diferentes espécies de peixes.

3.3.5 Conclusão

A fim de melhorar a seleção para resistência a doenças em tambaqui, geramos e aplicamos ferramentas genômicas para entender a arquitetura genômica de resistência contra *A. hydrophila* através da construção de um mapa de ligação e análise GWAS. Nossos resultados indicam que a resistência de *A. hydrophila* está sob controle poligênico em tambaqui e mais estudos são necessários para validar se os genes putativos encontrados aqui desempenham um papel na variação genética para resistência. A implementação da seleção assistida por marcadores provavelmente não é a abordagem mais eficiente para melhorar a resistência contra *A. hydrophila* em tambaqui, ao contrário do que foi sugerido para bagres em um estudo GWAS semelhante [19]. A seleção

genômica pode ser testada em novos estudos para melhorar a resistência a uma das doenças mais problemáticas que afeta a aquicultura sul-americana.

Financiamento

Este trabalho foi financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (bolsa FAPESP 2017/19717-3, 2019/10662-7 e 2018/08416-5), Conselho Nacional de Desenvolvimento Científico e Tecnológico (bolsa CNPq 311559/2018-2 e 422670/2018 -9), e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES - Código Financeiro 001). Nenhum potencial conflito de interesse foi relatado pelos autores.

3.3.6 Referências

- [1] Woynárovich, A., Van Anrooy, R., 2019. Field guide to the culture of tambaqui (*Colossoma macropomum*, Cuvier, 1816). FAO Fisheries and Aquaculture Technical Paper, (624), I-121.
- [2] Valladão, G.M.R., Gallani, S.U., Pilarski, F., 2018. South American fish for continental aquaculture. *Reviews in Aquaculture*, 10(2), 351-369. <https://doi.org/10.1111/raq.12164>
- [3] IBGE 2019. Pesquisa Pecuária Municipal – 2018. Disponível em: <https://sidra.ibge.gov.br/pesquisa/ppm/quadros/brasil/2018>.
- [4] Subasinghe, R. P., Bondad-Reantaso, M. G., McGladdery, S. E., 2001. Aquaculture development, health and wealth.
- [5] Huang, K., and Nitin, N., 2019. Edible bacteriophage based antimicrobial coating on fish feed for enhanced treatment of bacterial infections in aquaculture industry. *Aquaculture*, 502, 18-25. <https://doi.org/10.1016/j.aquaculture.2018.12.026>

- [6] Zhang, D., Xu, D. H., Shoemaker, C., 2016. Experimental induction of motile *Aeromonas septicemia* in channel catfish (*Ictalurus punctatus*) by waterborne challenge with virulent *Aeromonas hydrophila*. *Aquaculture Reports*, 3, 18-23. <https://doi.org/10.1016/j.aqrep.2015.11.003>
- [7] Jia, Z., Chen, L., Ge, Y., Li, S., Peng, W., Li, C., Zhang, Y., Hu, X., Zhou, Z., Shi, L., Xu, P., 2020. Genetic mapping of *Koi herpesvirus resistance* (KHVR) in Mirror carp (*Cyprinus carpio*) revealed genes and molecular mechanisms of disease resistance. *Aquaculture*, 519, 734850. <https://doi.org/10.1016/j.aquaculture.2019.734850>
- [8] Palaiokostas, C., Vesely, T., Kocour, M., Prchal, M., Pokorova, D., Piackova, V., Pokedal, L., Houston, R.D., 2019. Optimizing genomic prediction of host resistance to koi herpesvirus disease in carp. *Frontiers in genetics*, 10, 543. <https://doi.org/10.3389/fgene.2019.00543>
- [9] Tadmor-Levi, R., Hulata, G., David, L., 2019. Multiple interacting QTLs affect disease challenge survival in common carp (*Cyprinus carpio*). *Heredity*, 123(5), 565-578. <https://doi.org/10.1038/s41437-019-0224-0>
- [10] El-Magd, M. A., El-Said, K. S., El-Semlawy, A. A., Tanekhy, M., Afifi, M., Mohamed, T. M., 2019. Association of MHC IIA polymorphisms with disease resistance in *Aeromonas hydrophila*-challenged Nile tilapia. *Developmental & Comparative Immunology*, 96, 126-134. <https://doi.org/10.1016/j.dci.2019.03.002>
- [11] Bangera, R., Correa, K., Lhorente, J. P., Figueroa, R., Yáñez, J. M. 2017. Genomic predictions can accelerate selection for resistance against *Piscirickettsia salmonis* in Atlantic salmon (*Salmo salar*). *BMC genomics*, 18(1), 121. <https://doi.org/10.1186/s12864-017-3487-y>
- [12] Robledo, D., Matika, O., Hamilton, A., & Houston, R.D., 2018. Genome-wide association and genomic selection for resistance to amoebic gill disease in Atlantic

- salmon. G3: Genes, Genomes, Genetics, 8(4), 1195-1203.
<https://doi.org/10.1534/g3.118.200075>
- [13] Yoshida, G.M., Bangera, R., Carvalheiro, R., Correa, K., Figueroa, R., Lhorente, J.P., Yáñez, J.M., 2018. Genomic prediction accuracy for resistance against *Piscirickettsia salmonis* in farmed rainbow trout. G3 Genes, Genomes, Genet. 8, 719–726. <https://doi.org/10.1534/g3.117.300499>
- [14] Sonesson, A. K., Meuwissen, T. H., 2009. Testing strategies for genomic selection in aquaculture breeding programs. Genetics Selection Evolution, 41(1), 37. <https://doi.org/10.1186/1297-9686-41-37>
- [15] Yáñez, J. M., Lhorente, J. P., Bassini, L. N., Oyarzún, M., Neira, R., Newman, S., 2014a. Genetic co-variation between resistance against both *Caligus rogercresseyi* and *Piscirickettsia salmonis*, and body weight in Atlantic salmon (*Salmo salar*). Aquaculture, 433, 295-298. <https://doi.org/10.1016/j.aquaculture.2014.06.026>
- [16] Houston, R. D., Haley, C. S., Hamilton, A., Guy, D. R., Tinch, A. E., Taggart, J.B., McAndrew, B.J., Bishop, S.C., 2008. Major quantitative trait *loci* affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). Genetics, 178(2), 1109-1115. <https://doi.org/10.1534/genetics.107.082974>
- [17] Moen, T., Baranski, M., Sonesson, A. K., Kjøglum, S., 2009. Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): population-level associations between markers and trait. BMC genomics, 10(1), 368. <https://doi.org/10.1186/1471-2164-10-368>
- [18] Mastrochirico-Filho, V. A., Borges, C. H., Freitas, M. V., Ariede, R. B., Pilarski, F., Utsunomia, R., Carvalheiro, R., Gutierrez, A.P., Peñaloza, C., Yáñez, J.M., Houston, R.D., Hashimoto, D.T., 2020. Development of a SNP linkage map and genome-wide

- association study for resistance to *Aeromonas hydrophila* in pacu (*Piaractus mesopotamicus*). <https://doi.org/10.21203/rs.3.rs-26811/v1>
- [19] Wang, W., Tan, S., Luo, J., Shi, H., Zhou, T., Yang, Y., Jin, Y., Wang, X., Niu, D., Yuan, Z., Gao, D., Dunham, R., Liu, Z., 2019. GWAS Analysis Indicated Importance of NF- κ B Signaling Pathway in Host Resistance Against Motile *Aeromonas* Septicemia Disease in Catfish 2009. <https://doi.org/10.1007/s10126-019-09883-0>
- [20] Robinson, N., Baranski, M., Mahapatra, K. Das, Saha, J.N., Das, S., Mishra, J., Das, P., Kent, M., Arnyasi, M., Sahoo, P.K., 2014. A linkage map of transcribed single nucleotide polymorphisms in rohu (*Labeo rohita*) and QTL associated with resistance to *Aeromonas hydrophila* 1–23. <https://doi.org/10.1186/1471-2164-15-541>
- [21] Mastrochirico-Filho, V. A., Ariede, R. B., Freitas, M. V., Lira, L. V., Agudelo, J. F., Pilarski, F., Reis Neto, R.V., Yáñez, J.M., Hashimoto, D.T., 2019. Genetic parameters for resistance to *Aeromonas hydrophila* in the Neotropical fish pacu (*Piaractus mesopotamicus*). *Aquaculture*, 513, 734442. <https://doi.org/10.1016/j.aquaculture.2019.734442>
- [22] Ariede, R.B., Freitas, M.V., Agudelo, J. F., Borges, C. H., Lira, L. V., Yoshida, G.M., Pilarski, F., Yáñez, J.M., Hashimoto, D.T., 2020. Genetic (co) variation between resistance to *Aeromonas hydrophila* and growth in tambaqui (*Colossoma macropomum*). *Aquaculture*, 523, 735225. <https://doi.org/10.1016/j.aquaculture.2020.735225>
- [23] Yáñez, J. M., Houston, R. D., Newman, S., 2014b. Genetics and genomics of disease resistance in salmonid species. *Frontiers in genetics*, 5, 415. <https://doi.org/10.3389/fgene.2014.00415>
- [24] Barría, A., Christensen, K. A., Yoshida, G. M., Correa, K., Jedlicki, A., Lhorente, J.P., Davidson, W.S., Yáñez, J.M., 2018. Genomic predictions and genome-wide

- association study of resistance against *Piscirickettsia salmonis* in coho salmon (*Oncorhynchus kisutch*) using ddRAD sequencing. *G3: Genes, Genomes, Genetics*, 8(4), 1183-1194. <https://doi.org/10.1534/g3.118.200053>
- [25] Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., Bakker, P.I.W, Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, 81(3), 559-575. <https://doi.org/10.1086/519795>
- [26] Rastas, P., 2017. Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics*, 33(23), 3726-3732. <https://doi.org/10.1093/bioinformatics/btx494>
- [27] Ribeiro, L. B., Matoso, D. A., Feldberg, E., 2014. Chromosome mapping of repetitive sequences in four Serrasalminae species (Characiformes). *Genetics and Molecular Biology*, 37(1), 46-53. <https://doi.org/10.1590/S1415-47572014000100009>
- [28] Ouellette, L.A., Reid, R.W., Blanchard, S.G., Brouwer, C.R., 2018. LinkageMapView - rendering high-resolution linkage and QTL maps. *Bioinformatics*, 34(2), 306-307. <https://doi.org/10.1093/bioinformatics/btx576>
- [29] Wang, H., Misztal, I., Aguilar, I., Legarra, A., Fernando, R. L., Vitezica, Z., Okimoto, R., Wing, T., Hawken, R., Muir, W.M., 2014. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. *Frontiers in Genetics*, 5, 134. <https://doi.org/10.3389/fgene.2014.00134>
- [30] Aguilar, I., Misztal, I., Johnson, D.L., Legarra, A., Tsuruta, S., Lawlor, T.J., 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic

- information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93, 743–752.
<https://doi.org/10.3168/jds.2009-2730>
- [31] Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., Lee, D. H., 2002. BLUPF90 and related programs (BGF90). In Proceedings of the 7th world congress on genetics applied to livestock production (Vol. 33, pp. 743-744).
<https://doi.org/9782738010520>
- [32] Sato, L. S., Jorge, P. H., Mastrochirico-Filho, V. A., 2020. Triploidy in tambaqui *Colossoma macropomum* identified by chromosomes of fish larvae. *J Aquac Mar Biol.* 9(3), 65-69.
- [33] Nunes, J., Liu, S., Pértille, F., Perazza, C.A., Villela, P.M.S., Almeida-Val, V.M.F., Hilsdorf, A.W.S., Liu, Z., Coutinho, L.L., 2017. Large-scale SNP discovery and construction of a high-density genetic map of *Colossoma macropomum* through genotyping-by-sequencing. *Sci Rep* 7, 46112. <https://doi.org/10.1038/srep46112>
- [34] Lima FCT, Malabarba LR, Buckup PA, da Silva JFP, Vari RP, Harold A, Benine R, Oyakawa OT, Pavanelli CS, Menezes NA, Lucena CAS, Malabarba MCSL, Lucena ZMS, Reis RE, Langeani F, Cassati L, Bertaco VA, Moreira C, Lucinda PHF: Genera *incertae sedis* in Characidae. Check List of the Freshwater Fishes of South and Central America. Edited by: Reis RE, Kullander SE, Ferraris Jr CJ. 2003, Porto Alegre: Edipucrs, 106-169.
- [35] Oliveira, C., Avelino, G.S., Abe, K.T., Mariguela, T.C., Benine, R.C., Ortí, G., Vari, R.P., Castro, R.M.C., 2011. Phylogenetic relationships within the speciose family Characidae (Teleostei: Ostariophysii: Characiformes) based on multilocus analysis and extensive in group sampling. *BMC Evol Biol* 11, 275.
<https://doi.org/10.1186/1471-2148-11-275>

- [36] Vallejo, R. L., Leeds, T. D., Gao, G., Parsons, J. E., Martin, K. E., Evenhuis, J. P., Fragomeni, B. O., Wiens, G.D., Palti, Y., 2017. Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. *Genetics Selection Evolution*, 49(1), 1-13.
<https://doi.org/10.1186/s12711-017-0293-6>
- [37] Lu, S., Liu, Y., Yu, X., Li, Y., Yang, Y., Wei, M., Zhou, Q., Wang, J., Zhang, Y., Zheng, W., Chen, S., 2020. Prediction of genomic breeding values based on pre-selected SNPs using ssGBLUP, WssGBLUP and BayesB for Edwardsiellosis resistance in Japanese flounder. *Genetics Selection Evolution*, 52(1), 1-10.
<https://doi.org/10.1186/s12711-020-00566-2>
- [38] Mahapatra, K. D., Gjerde, B., Sahoo, P. K., Saha, J. N., Barat, A., Sahoo, M., Mohanty, B.R., Ødegård, J., Rye, M., Salte, R., 2008. Genetic variations in survival of rohu carp (*Labeo rohita*, Hamilton) after *Aeromonas hydrophila* infection in challenge tests. *Aquaculture*, 279(1-4), 29-34.
<https://doi.org/10.1016/j.aquaculture.2008.03.054>
- [39] Xiong, X., Chen, Y., Liu, L., Wang, W., Andrew, N., Gao, Z., 2017. Estimation of genetic parameters for resistance to *Aeromonas hydrophila* in blunt snout bream (*Megalobrama amblycephala*). *Aquaculture* 479, 768–773.
<https://doi.org/10.1016/j.aquaculture.2017.07.011>
- [40] Srisapoome, P., Chatchaiphan, S., Bunnoy, A., Koonawootrittriron, S., Na-Nakorn, U., 2019. Heritability of immunity traits and disease resistance of bighead catfish, *Clarias macrocephalus* Günther, 1864. *Fish & shellfish immunology*, 92, 209-215.
<https://doi.org/10.1016/j.fsi.2019.05.060>

- [41] Palaiokostas, C., Ferraresso, S., Franch, R., Houston, R. D., Bargelloni, L., 2016. Genomic prediction of resistance to pasteurellosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. *G3: Genes, Genomes, Genetics*, 6(11), 3693-3700. <https://doi.org/10.1534/g3.116.035220>
- [42] Correa, K., Lhorente, J. P., López, M. E., Bassini, L., Naswa, S., Deeb, N., Genova, A.D., Maass, A., Davidson, W.S., Yáñez, J.M., 2015. Genome-wide association analysis reveals *loci* associated with resistance against *Piscirickettsia salmonis* in two Atlantic salmon (*Salmo salar* L.) chromosomes. *BMC genomics*, 16(1), 854. <https://doi.org/10.1186/s12864-015-2038-7>
- [43] Meuwissen, T.H.E., Hayes, B. J., Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4), 1819-1829.
- [44] Tsai, H.Y., Robledo, D., Lowe, N.R., Bekaert, M., Taggart, J.B., Bron, J.E., Houston, R.D., 2016. Construction and annotation of a high density SNP linkage map of the Atlantic salmon (*Salmo salar*) genome. *G3 Genes, Genomes, Genet.* 6, 2173–2179. <https://doi.org/10.1534/g3.116.029009>
- [45] Correa, K., Bangera, R., Figueroa, R., Lhorente, J. P., Yáñez, J. M., 2017. The use of genomic information increases the accuracy of breeding value predictions for sea louse (*Caligus rogercresseyi*) resistance in Atlantic salmon (*Salmo salar*). *Genetics Selection Evolution*, 49(1), 1-5. <https://doi.org/10.1186/s12711-017-0291-8>
- [46] Yáñez, J.M., Yoshida, G.M., Parra, A., Correa, K., Barria, A., Bassini, L.N., Christensen, K., Lopez Dinamarca, M.E., Carvalheiro, R., Lhorente, J.P., Pulgar, R.E., 2019. Comparative genomic analysis of three salmonid species identifies functional candidate genes involved in resistance to the intracellular bacterium *Piscirickettsia salmonis*. *Front. Genet.* 10, 1–13. <https://doi.org/10.3389/fgene.2019.00665>

- [47] Tárrega, C., Pulido, R., 2009. A one-step method to identify MAP kinase residues involved in inactivation by tyrosine- and dual-specificity protein phosphatases. *Anal. Biochem.* 394, 81–86. <https://doi.org/10.1016/j.ab.2009.07.006>
- [48] Mu, Y., Ding, F., Cui, P., Ao, J., Hu, S., Chen, X., 2010. Transcriptome and expression profiling analysis revealed changes of multiple signaling pathways involved in immunity in the large yellow croaker during *Aeromonas hydrophila* infection. *BMC Genomics* 11, 506. <https://doi.org/10.1186/1471-2164-11-506>
- [49] Jiang, Y., Feng, S., Zhang, S., Liu, H., Feng, J., Mu, X., Sun, X., Xu, P., 2016. Transcriptome signatures in common carp spleen in response to *Aeromonas hydrophila* infection. *Fish Shellfish Immunol.* 57, 41–48. <https://doi.org/10.1016/j.fsi.2016.08.013>
- [50] Laing, K.J., Secombes, C.J., 2004. Trout CC chemokines: comparison of their sequences and expression patterns. *Molecular immunology*, 41(8), 793-808. <https://doi.org/10.1016/j.molimm.2004.03.038>
- [51] Parkin, J., Cohen, B., 2001. An overview of the immune system. *The Lancet* 357(9270), 1777-1789. [https://doi.org/10.1016/S0140-6736\(00\)04904-7](https://doi.org/10.1016/S0140-6736(00)04904-7)
- [52] Rebl, A., Goldammer, T., 2018. Fish and Shell fish Immunology Under control: The innate immunity of fish from the inhibitors' perspective. *Fish Shellfish Immunol.* 77, 328–349. <https://doi.org/10.1016/j.fsi.2018.04.016>
- [53] Bionda, T., Behrends, C., 2016. Unveiling ubiquitinome rearrangements induced by *Salmonella* infection. *Autophagy* 12(9), 1683-1684. <https://doi.org/10.1080/15548627.2016.1203490>
- [54] Silverman, N., Maniatis, T., 2001. NF- κ B signaling pathways in mammalian and insect innate immunity. *Genes & development*, 15(18), 2321-2342. <https://doi.org/10.1101/gad.909001>

- [55] Li, Q., Verma, I. M., 2002. NF- κ B regulation in the immune system. *Nature Reviews Immunology* 2 (10), 725-734. <https://doi.org/10.1038/nri910>
- [56] Tran, N.T., Gao, Z.X., Zhao, H.H., Yi, S.K., Chen, B.X., Zhao, Y.H., Lin, L., Liu, X.Q., Wang, W.M., 2015. Transcriptome analysis and microsatellite discovery in the blunt snout bream (*Megalobrama amblycephala*) after challenge with *Aeromonas hydrophila*. *Fish Shellfish Immunol.* 45, 72–82. <https://doi.org/10.1016/j.fsi.2015.01.034>
- [57] Krause, G., Winkler, L., Piehl, C., Blasig, I., Piontek, J., Müller, S. L., 2009. Structure and function of extracellular claudin domains. *Annals of the New York Academy of Sciences* 1165(1), 34.
- [58] Syakuri, H., Jung-Schroers, V., Adamek, M., Brogden, G., Irnazarow, I., Steinhagen, D., 2014. Beta-glucan feeding differentiated the regulation of mRNA expression of claudin genes and prevented an intestinal inflammatory response post *Aeromonas hydrophila* intubation in common carp, *Cyprinus carpio* L. *Journal of fish diseases* 37(2), 149-156. <https://doi.org/10.1111/jfd.12121>
- [59] Mastrochirico-Filho, V. A., Ariede, R. B., Freitas, M. V., Borges, C. H., Lira, L. V., Mendes, N. J., *et al.* (2021). Development of a multi-species SNP array for serrasalmid fish *Collossoma macropomum* and *Piaractus mesopotamicus*. *Scientific Reports*, 11(1), 1-11.

3.4 Capítulo IV

Avaliação de parâmetros genéticos para desempenho de crescimento e rendimento corpóreo em tambaqui *Collossoma macropomum*

Resumo

O objetivo deste estudo foi estimar componentes de variância e herdabilidade para desempenho de crescimento e rendimento corpóreo em tambaqui, assim como suas correlações genéticas/fenotípicas, em uma população experimental na região Sudeste do Brasil. Uma abordagem de *deep learning* e sistema de visão computacional (CVS) também foi validada para obter as medidas morfométricas do tambaqui de forma automatizada. Estimativas de parâmetros genéticos foram realizadas utilizando dados de 11 famílias de irmãos completos, com um total de 365 indivíduos, avaliados por 15 meses. Ao final, oito características de crescimento foram mensuradas. Após a biometria, cada peixe foi processado nas seguintes regiões corporais: cabeça, costela, lombo, C+L (costela + lombo) e carcaça (correspondendo ao esqueleto vertebral, nadadeiras e vísceras). As estimativas de herdabilidade para características de crescimento e rendimento corpóreo (peso e proporção - %) variaram de baixo a alto; com exceção da costela. Os valores de herdabilidade dos resíduos da regressão das partes corporais foram mais baixos. As correlações genéticas entre % das partes corporais e características morfométricas foram favoráveis e altamente correlacionados, com exceção de % Cabeça cujas correlações foram desfavoráveis. O CVS validado nesse conjunto de imagens se mostrou resiliente e pode ser utilizado para fenotipagem em larga escala e automatizada do tambaqui. Em conclusão, as características de crescimento e rendimento corpóreo provavelmente estão sob moderado controle genético e devem responder à seleção genética. Algumas medidas morfométricas podem ser utilizadas como critérios de seleção eficientes para rendimento corpóreo nessa população de tambaqui.

3.4.1 Introdução

O tambaqui *Colossoma macropomum* é o peixe amazônico mais importante da aquicultura continental da América do Sul, especialmente no Brasil, cuja produção total foi de 102 mil toneladas em 2018 (IBGE 2019). O tambaqui e seus híbridos tambatinga (*C. macropomum* x *P. brachypomus*) e tambacu (*C. macropomum* x *P. mesopotamicus*) têm contribuído com aproximadamente 30% da produção da aquicultura nacional, abaixo apenas da tilápia (cerca de 60% da produção brasileira). Além disso, sua produção também ocorre em vários países da Ásia, como China, Indonésia, Malásia, Mianmar e Vietnã (Woynárovich e Van Anrooy, 2019). Por isso, essa espécie é fundamental para ser explorada em programas de seleção genética que visem uma produção eficiente e sustentável.

Em geral, peso corporal é o principal critério de seleção em programas de melhoramento genético de peixes com objetivo de melhor desempenho de crescimento (Gjedrem, 2012; Fraslin *et al.*, 2018). Em tambaqui, estudos prévios já foram desenvolvidos para avaliar parâmetros genéticos de características relacionadas ao crescimento, incluindo ganho de peso, comprimento padrão, tamanho da cabeça, altura e largura (Marcos *et al.*, 2016; Mello *et al.*, 2016; Perazza *et al.*, 2019; Ariede *et al.*, 2018, 2020). Os resultados desses estudos revelaram herdabilidades em diferentes populações que variaram de baixa a alta magnitude (0,10 a 0,46), e com alta correlação entre elas, principalmente de peso corporal com comprimento padrão (Mello *et al.*, 2016; Marcos *et al.*, 2016; Perazza *et al.*, 2019).

Recentemente, a área do lombo e espaço entre as costelas foram avaliadas indiretamente por ultrassom no tambaqui por serem considerados cortes nobres nessa espécie, de forma similar aos cortes comercializados na indústria de suínos (Perazza *et al.*, 2019). Ambas as características foram caracterizadas com alta herdabilidade (0,39 e

0,50), e as correlações da área do lombo com diferentes medidas corporais variaram de 0,16 a 0,94. No entanto, espaço entre as costelas apresentou baixa variabilidade fenotípica, o que dificultaria a inclusão dessa característica em um programa de seleção (Perazza *et al.*, 2019).

Nenhum estudo ainda foi realizado para avaliar a variação genética do rendimento de partes corporais de tambaqui. Na literatura, a herdabilidade para rendimento de filé em peixes varia de baixa a moderada (0,03 a 0,35) (Rutten *et al.*, 2005; Kause *et al.*, 2007; Powell *et al.*, 2008; Navarro *et al.*, 2009; Nguyen *et al.*, 2010; Gjerde *et al.*, 2012; Haffray *et al.*, 2012, 2013; Vandeputte *et al.*, 2014). Contudo, os experimentos para rendimento de filé em peixes relatam ganhos genéticos eficientes e moderados (Fraslin *et al.*, 2018).

O rendimento de filé é raramente usado como objetivo de seleção em programas de melhoramento de peixes porque é uma característica difícil de ser avaliada (filé, cabeça ou rendimento de carcaça), particularmente porque não pode ser medida diretamente sob os candidatos de seleção, ou seja, implica no sacrifício do animal e descarte de um potencial reprodutor (Kause *et al.*, 2007). O rendimento de filé é geralmente realizado indiretamente usando o desempenho de irmãos sem o fenótipo alvo (ou seja, peixes candidatos com fenótipos são utilizados para coleta de dados), o que reduz drasticamente o progresso genético. Esta característica também pode ser selecionada indiretamente por meio de características correlacionadas (por exemplo, peso corporal), mas pode levar a ganhos genéticos preditivos não confiáveis (Rutten *et al.*, 2005; Navarro *et al.*, 2009). Portanto, a estimativa de parâmetros genéticos é necessária para entender o nível de variação genética e a relação entre as características analisadas, principalmente o rendimento de filé.

A fenotipagem em larga escala (fenômica) é uma abordagem emergente que provavelmente irá impulsionar a melhor compreensão de alguns fenômenos biológicos,

particularmente pela automação da medição de dados de todas as variáveis importantes (Houle *et al.*, 2010). As tecnologias para fenotipagem de alto rendimento estão agora se tornando disponíveis; especialmente aqueles que usam inteligência artificial e *deep learning*, o que criará oportunidades para incorporar a piscicultura inteligente na indústria da aquicultura (Yang *et al.*, 2021). Recentemente, um sistema de visão computacional foi desenvolvido com sucesso por processamento de imagem e segmentação em pacu *Piaractus mesopotamicus*, uma espécie muito próxima filogeneticamente do tambaqui, resultando em uma abordagem automatizada rápida e não invasiva para medição de características biométricas, que pode ser usada para predição de formato corpóreo (Freitas *et al.*, em preparação).

O objetivo deste estudo foi estimar componentes de variância e herdabilidade para desempenho de crescimento e rendimento de partes corporais em tambaqui, assim como suas correlações genéticas/fenotípicas, e a validação de um sistema de visão computacional para fenotipagem em larga escala em uma população experimental na região Sudeste do Brasil.

3.4.2 Material e Métodos

3.4.2.1 Comitê de Ética

Todos os procedimentos com os peixes deste projeto foram analisados e aprovados pelo Comitê de Ética em Uso Animal (CEUA nº 019006/17) da Faculdade de Ciências Agrárias e Veterinárias, UNESP, Campus Jaboticabal, SP, Brasil. Os peixes foram anestesiados durante os experimentos com benzocaína (40 mg / L) e todos os esforços foram feitos para garantir o bem-estar e minimizar o sofrimento animal.

3.4.2.2 População experimental

A população analisada nesse estudo foi composta por 365 indivíduos de *C. macropomum*, pertencentes a 11 famílias de irmãos completos, gerados por um esquema de acasalamento hierárquico (6 fêmeas e 10 machos). Estas mesmas famílias também foram avaliadas em um estudo de resistência a *Aeromonas hydrophila* (Ariede *et al.*, 2020), que incluiu a análise de outras sete famílias adicionais (18 famílias ao total). No presente estudo, a inclusão de todas as 18 famílias não foi possível devido ao baixo número amostral, pois somente 11 famílias apresentaram quantidade de indivíduos suficientes para avaliação das características de crescimento e rendimento. Informações sobre a reprodução, formação das famílias e manejo reprodutivo estão descritos em Ariede *et al.* (2020).

3.4.2.3 Experimento do desempenho de crescimento e rendimento de filé

Ao total, aproximadamente 30 animais foram utilizados de cada uma das 11 famílias para o delineamento experimental (inteiramente ao acaso) de avaliação das características de crescimento e rendimento de filé. Cada família de irmãos completos foi mantida separada até o momento da marcação com as pit-tags (aproximadamente 8 meses de idade) em tanques de 1m³, conectados em sistema de recirculação de água, com filtros mecânicos e biológicos, sistema de aeração externa e temperatura controlada por controlador térmico conectado a aquecedor (2x500 w). Temperatura, oxigênio dissolvido e pH foram mensurados com a sonda Multiparameter Water Quality Checker U-50 (Horiba, Kyoto, Japão). A temperatura da água foi mantida a 30 °C ($\pm 0,5$), com oxigênio dissolvido a 5,0 mg / L ($\pm 1,0$) e pH a 7,2 ($\pm 1,0$).

Após a marcação, os juvenis das diferentes famílias foram pesados (peso inicial) e distribuídos em três tanques escavados de 200 m², sendo aprox. 10 indivíduos de cada família por tanque. Então, cada tanque foi composto por aproximadamente 110 animais, que compartilharam este ambiente comum durante o período de avaliação. O experimento de desempenho de crescimento foi realizado por cerca de 15 meses. Os valores médios dos parâmetros da água mensurados foram: temperatura (28,3 °C), pH (6,6) e oxigênio (5,6 mg / l). Os tanques experimentais pertencem às instalações do Centro de Aquicultura da Unesp (Jaboticabal, Brasil), onde o núcleo de melhoramento é mantido. Durante o experimento, os peixes foram alimentados com rações peletizadas de 2 a 6 mm (36% de proteína bruta, ração comercial Nutripiscis Presence). A quantidade de ração foi estabelecida de acordo com a biomassa de cada tanque (5% do peso total de peixes), com base em uma subamostra pesada a cada trimestre.

Oito características de crescimento foram avaliadas (Figura 1), obtendo-se as seguintes medidas: peso total (**We**); comprimento padrão (**SL**), compreendido entre a extremidade anterior da cabeça e o menor perímetro do pedúnculo caudal (região da inserção da nadadeira caudal); comprimento da cabeça (**HL**), compreendido entre a extremidade anterior da cabeça e a borda caudal do opérculo; altura da cabeça (**HH**); altura 1 (**H1**) do corpo, medida na região do 1º raio da nadadeira dorsal ao 1º raio da nadadeira pélvica; altura 2 (**H2**), entre a base posterior da nadadeira dorsal e base anterior da nadadeira anal; comprimento da pelve (**PL**), compreendido entre a base anterior da nadadeira pélvica e a base anterior da nadadeira anal; largura (**Wi**). Após a biometria, cada peixe foi processado nas seguintes regiões corporais: cabeça, costela, lombo, C+L (costela + lombo) (Figura 1) e carcaça (correspondendo ao esqueleto vertebral, nadadeiras e vísceras). As regiões corporais foram pesadas individualmente com balança analítica. A filetagem foi realizada por apenas uma pessoa.

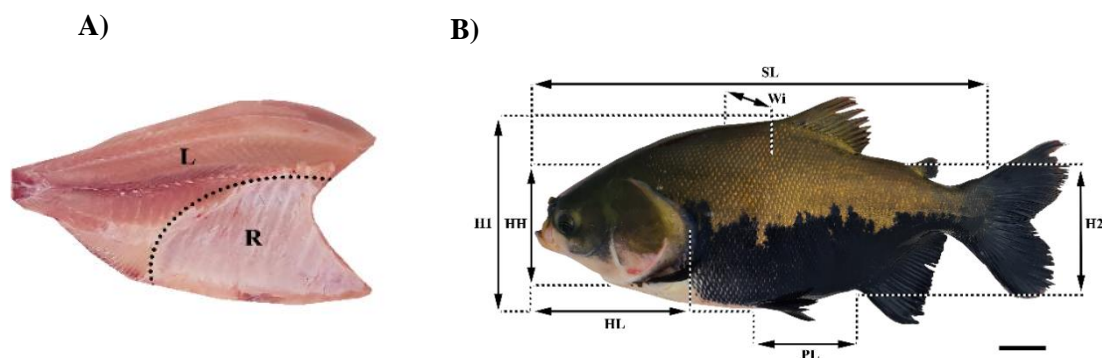


Figura 1. A) cortes utilizados para avaliação de rendimento de filé, L= lombo e R = costela. B) Características morfométricas mensuradas em tambaqui (*Colossoma macropomum*).

Os rendimentos foram calculados dividindo o peso das diferentes regiões corporais pelo peso corporal total (Tabela 1): rendimento de cabeça (% de cabeça), rendimento de costela (% costela), rendimento de lombo (% lombo) e rendimento de C+L (% C+L). Devido o cálculo do rendimento ser uma proporção, as respostas esperadas à seleção sobre os rendimentos podem ser tendenciosas, imprevisíveis e não preditivas do ganho genético real (Gunsett, 1987). Isso ocorre porque os componentes das razões podem ser altamente correlacionados (ver Pearson, 1897 e Sutherland, 1965). Por esta razão, outros índices e critérios de seleção podem ser utilizados para avaliar o rendimento de filé (Vandeputte *et al.*, 2020). Conforme sugerido por Koch *et al.* (1963) e outros estudos prévios (Haffray *et al.*, 2013; Haffray *et al.*, 2012; Prchal *et al.*, 2018; Vandeputte *et al.*, 2017; Vandeputte *et al.*, 2014), o uso do resíduo da regressão linear entre o peso do filé e o peso corporal transformados em log pode ser utilizado como alternativa para o rendimento do filé. Essa abordagem se mostrou mais eficiente do que a predição do rendimento do filé em porcentagem (Vandeputte *et al.*, 2020). Portanto, obteve-se novas variáveis (resíduos) da regressão linear dos pesos das regiões corporais e W_e transformados em log (Logr_Cabeça, Logr_Costela, Logr_Lombo, Logr_C+L).

O sexo dos animais do presente estudo foi identificado por inspeção visual das gônadas e confirmado por histologia. As gônadas foram fixadas em formalina tamponada a 10% e após 24 h foram transferidas para etanol 70%, acondicionadas em cassetes plásticas. Por serem muito pequenas, as mesmas foram protegidas com papel de filtro para evitar o risco de perda durante o processamento. As amostras foram submetidas ao método tradicional: desidratação, diafanização, parafinização e emblocamento em parafina histológica. Os cortes foram realizados em micrótomo rotativo Leica 2125, com 4 micrômetros de espessura. As amostras foram coradas por Hematoxilina e Eosina. As lâminas foram montadas com Entellan.

3.4.2.4 Sistema de visão computacional e Deep Learning

As medidas morfométricas também foram estimadas de forma automatizada por meio de um sistema de visão computacional (CVS) desenvolvida para o tambaqui. Entretanto, devido ao baixo número amostral de imagens de tambaqui para o treinamento do modelo por *deep learning* (n=302), um conjunto de imagens de aproximadamente 1.354 animais de pacu também foram utilizados no desenvolvimento do CVS (Freitas *et al.*, em preparação). Resumidamente, o CVS foi implementado por meio de *deep learning*, adotando uma família de métodos conhecida como *Region Proposal Convolutional Neural Networks* (R-CNN), adaptado do método de *Mask RCNN* (He *et al.*, 2018). O CVS foi treinado para realizar a segmentação automática da área da cabeça, corpo e nadadeiras, por meio da classificação dos pixels. Em seguida, o valor de cada pixel nas imagens foi transformado linearmente em centímetro, pela calibração em uma régua afixada na base das imagens, dos quais é possível obter medidas morfométricas e das áreas corporais. Esta técnica foi testada e validada nas imagens capturadas de cada

indivíduo de tabaqui do presente estudo, por meio do cálculo da correlação linear entre os valores reais e preditos pelo CVS.

3.4.2.4.1 Aquisição de dados de imagem e conjuntos de dados

Imagens de todos os animais foram coletadas no final do período de experimento de crescimento. Para capturar as fotos, foi utilizada uma câmera digital de 12 megapixels configurada para melhor resolução e foco automático. Para manter a padronização da distância das imagens, a câmera foi fixada em uma plataforma fixa a 0,8 cm de distância. Uma régua com medidas conhecidas foi colocada em cada imagem para futura calibração entre os milímetros e os pixels gerados pela imagem. O software Image J foi utilizado para a aquisição manual das medidas morfométricas e da área de cada animal, com base no comprimento dos pixels do sistema métrico (mm). A calibração foi realizada por meio de medidas feitas diretamente no comprimento da régua.

3.4.2.4.2 Segmentação e análise de imagens

Cada imagem do banco de dados foi usada para segmentação manual da área da cabeça (A); corpo (B) e nadadeiras (C), usando o software Labelbox (Figura 3). As imagens segmentadas foram agrupadas para construir um conjunto de dados rotulado, onde cada pixel da imagem original correspondia a uma das classes previamente definidas (cabeça, corpo ou nadadeiras). Para evitar viés de padronização e erros individuais, a segmentação manual das imagens foi realizada por apenas um profissional e todas as fotos foram posteriormente selecionadas e aprovadas por um avaliador treinado. Caso uma medida não fosse aceita, a mesma imagem era enviada de volta ao rotulador para ser reavaliada.

A partir das imagens segmentadas, o valor de cada pixel nas imagens foi transformado linearmente em milímetros medindo a distância correspondente em pixels das pontas da régua afixada à base das imagens. As medidas morfométricas da altura da cabeça (HH), comprimento da cabeça (HL), comprimento padrão (SL) e altura (H1) foram então estimados usando esta transformação linear.

O conjunto de dados utilizado para o treinamento do algoritmo foi composto por 302 imagens de tambaqui e 1.354 imagens de pacu. O conjunto de dados foi dividido em três conjuntos separados: conjunto de treinamento, desenvolvimento e validação. Para validação, consideramos 10% do conjunto de dados original (proporcional ao número do conjunto de dados de cada espécie). Das imagens restantes, 5% foi dividido como um conjunto de desenvolvimento, que usamos para ajustar os hiperparâmetros e para realizar a experimentação, evitando assim usar diretamente o conjunto de validação e possivelmente overfitting nele. A métrica sendo otimizada para as medições foi a intersecção sobre a união (IoU). No final de cada época, calculamos a IoU proporcional ao conjunto de desenvolvimento. A IoU foi calculada usando a área A prevista e a área B esperada (anotada) com a seguinte fórmula $A \cap B / A \cup B$. Quanto mais próximo de 1,0 a IoU, mais precisas as previsões eram.

Para verificar a validade da predição das medidas morfométricas, foi realizada a correlação de Pearson entre as medidas estimadas manualmente e pelo sistema computacional.

3.4.2.5 Análises genômicas

Amostras de sangue foram utilizadas para extração de DNA com o kit comercial *Wizard Genomic* da Promega, seguindo as instruções do fabricante. A quantificação do DNA extraído (ng/ μ l) foi medida pelo detector de fluorescência Qubit, usando o kit Qubit

dsDNA HS Assay (Invitrogen, EUA). O protocolo para a construção das bibliotecas foi de acordo com Peterson *et al.* (2012), com modificações. Resumidamente, 450 ng de DNA genômico de cada indivíduo foram digeridos (8 U/reação) usando a combinação de duas enzimas de restrição, SphI e MluCI (New England) e ligados a adaptadores específicos (P1 e P2, 0,25 μ M) utilizando a enzima T4 DNA ligase, a 23°C por 4 horas e 65°C por 10 minutos. Os adaptadores P1 possuem 5 nucleotídeos adicionais que funcionam como marcadores individuais (barcode). A seleção dos fragmentos digeridos foi realizada pelo Purelink™ Quick Gel Extraction Kit (Thermo Fisher Scientific) com fragmentos de aproximadamente 280pb. Posteriormente, foram realizados ensaios de PCR para incorporar a identificação de cada biblioteca. Cada biblioteca foi formada com aproximadamente 47 amostras. A PCR foi realizada de acordo com as condições da enzima Platinum SuperFi DNA Polymerase™. As reações foram purificadas com o kit ProNex® Size-Selective Purification System (Promega) e a concentração foi verificada novamente por fluorometria no equipamento Qubit™ 3.0 (Thermo Fisher Scientific). Por fim, as bibliotecas foram enviadas para serem sequenciadas no DNBseq 150PE (BGI).

Os dados obtidos do sequenciamento foram analisados usando o pacote de software Stacks v. 2.41 para identificação dos SNPs (Catchen *et al.* 2011). As sequências foram demultiplexadas e filtradas utilizando o **process_radtags**. As sequências que passaram nos filtros de qualidade anterior foram alinhadas ao genoma de referência do *Colossoma macropomum* (Hilsdorf *et al.*, 2021), mas que foi refinado por uma etapa prévia de ancoragem em nível cromossômico, permitindo a aplicação futura de GWAS. A obtenção do genoma do tambaqui em nível cromossômico foi realizado por meio do software Chromonomer (Catchen *et al.*, 2020), que requer uma descrição da montagem do genoma, que consiste em um arquivo AGP (A Golden Path) com a estrutura dos *scaffolds* (GenBank: GCA_904425465.1), um arquivo contendo a informação genética

do mapa de ligação (gerado no Cap. III), incluindo o grupo de ligação e posição centiMorgan (cM) de cada SNP, e um arquivo SAM ou BAM descrevendo as posições de alinhamento dos *loci* dos SNPs na montagem física do genoma de referência (GenBank: GCA_904425465.1). Este último arquivo foi gerado usando o software BWA v.0.7.12 (Li e Durbin 2009), a partir da indexação do genoma de referência (usando a função **index**) e os alinhamentos dos *loci* do mapa genético usando o algoritmo **mem**; todos os parâmetros foram definidos como *default*. O *output* do Chromonomer é um arquivo fasta (*chromonome.fa*), que contém o genoma em nível cromossômico, de acordo com a orientação do mapa genético.

Por fim, o arquivo *chromonome.fa* foi utilizado como referência para alinhar as sequências individuais de ddRAD-seq (resultantes do *process_radtags* - Stacks), por meio do software BWA, após a etapa de indexação. Em seguida, prosseguiu-se com o programa **gstacks** que criou *loci* incorporando as leituras de ddRAD que foram alinhadas ao genoma de referência. O programa **populations** foi utilizado para gerar dados de genótipos para as amostras das famílias em estudo. A fim de diferenciar SNPs putativos de erros de sequenciamento, o software Plink 1.9 (Purcell, 2007) foi usado para excluir SNPs utilizando o *geno* com valor de 0.05, ou seja, SNPs com mais de 5% de taxa de erro de genotipagem foram excluídos, frequências de alelos menores (MAF) inferiores a 0.05, e desequilíbrio de Hardy-Weinberg com correção de Bonferroni. Em relação ao filtro de indivíduos, as amostras que apresentavam mais de 15% (*mind* 0.15) de genótipos ausentes foram excluídas.

3.4.2.6 Modelos estatísticos

Todas as características de crescimento foram analisadas por meio de um modelo animal univariado, conforme definido abaixo:

$$y_{ij} = \mu + t_i + s_g + a_j + e_{igj}$$

Onde, y_{ij} foi o fenótipo para o peixe j (peso, medidas morfométricas, variáveis de rendimento), no tanque i ; μ foi o efeito fixo da média geral; t_i foi o efeito fixo do tanque i , s_g foi o efeito fixo do sexo g e covariável peso inicial; a_j foi o efeito aleatório genético animal do indivíduo j ; e e_{igj} foi o resíduo aleatório para o peixe j .

O modelo foi ajustado no *software* BLUPF90 (Misztal *et al.*, 2009). No modelo, o efeito aleatório genético animal foi assumido como sendo $\sim N(0, A\sigma_a^2)$, onde A é a matriz de parentesco genético aditiva baseada no pedigree entre todos os animais incluídos na população e σ_a^2 é a variância genética aditiva. Os resíduos foram assumidos como $\sim N(0, I\sigma_e^2)$, onde I é uma matriz de identidade e σ_e^2 é a variância residual.

A herdabilidade foi calculada como:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$$

Onde, σ_a^2 foi a variância genética aditiva e σ_e^2 a variância residual.

As correlações genéticas $r_g(m, n)$ entre características de crescimento e rendimento foram calculadas a partir dos componentes de covariância estimados utilizando análises bivariadas independentes, por meio de ajustes no modelo animal conforme definido acima, conforme descrito por Falconer e Mackay, 1996:

$$r_{g(m,n)} = \frac{Cov_g(m, n)}{\sqrt{Var_g m \times Var_g n}}$$

Onde, $Cov_g(m, n)$ foi a covariância genética entre as duas características, $Var_g m$ e $Var_g n$ foi a variância genética aditiva da característica m e característica n , respectivamente. As correlações foram classificadas de acordo com Navarro *et al.* (2009), com os valores de 0 - 0,40 considerados baixos, 0,45 - 0,55 moderados e 0,60 - 1 altos.

A correlação fenotípica $r_p(m,n)$ foi calculada como:

$$r_{p(m,n)} = \frac{Cov_p(m,n)}{\sqrt{Var_p m \times Var_p n}}$$

Onde, $Cov_p(m,n)$ foi a covariância fenotípica entre as duas características, $Var_p m$ e $Var_p n$ foi a variância fenotípica da característica m e característica n , respectivamente.

3.4.2.7 Análise de associação genômica ampla (GWAS) para rendimento de filé

Os componentes de variância e o valor genômico genômico (GEBV) foram estimados usando o método ssGBLUP (Wang *et al.*, 2014). A matriz de parentesco utilizada foi a matriz H (Aguilar *et al.*, 2010), na qual os dados de genótipo e pedigree foram combinados. O inverso da matriz H foi:

$$H^{-1} = A^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & G^{-1} - A_{22}^{-1} \end{bmatrix}$$

onde, A^{-1} é a matriz de relação do numerador inverso para todos os animais, A_{22}^{-1} é o inverso de uma matriz de relacionamento baseada em linhagem para animais genotipados (todos os peixes neste caso), e G^{-1} é a matriz de relacionamento genômico inverso. Na primeira iteração, os pesos para cada SNP foram atribuídos como 1, o que significa que todos os SNPs tiveram o mesmo peso.

As características de crescimento e rendimento de filé foram avaliadas como linear utilizando AIREMLF90 e BLUPF90 (Misztal *et al.*, 2009). A significância estatística dos SNPs foi avaliada ajustando o valor de p com a correção de Bonferroni. Os SNPs foram considerados significativos em nível genômico quando excederam o limite de Bonferroni para testes múltiplos (alfa = 0,05) de $0,05 / tg$, onde tg é o número total de SNPs usados no GWAS, e significativo em nível cromossômico quando foi maior que

0,05 / t_c , onde t_c é o número médio de SNPs por cromossomo. Os gráficos de Manhattan foram desenhados com base em valores de p dos SNP individuais, distribuídos em diferentes grupos de ligação. Cada locus contendo SNPs associados foi alinhado contra os scaffolds do genoma de *Colossoma macropomum* (NCBI RefSeq assembly access: GCF_904425465.1) usando a ferramenta BLASTn para investigar genes candidatos (janela de 500Kb) em torno dos SNPs associados a características de crescimento e rendimento de filé.

3.4.3 Resultados

3.4.3.1 Estatísticas descritivas

A média geral de cada característica de crescimento e rendimento corpóreo avaliado neste experimento está descrita na Tabela 1. A média geral de peso corporal final (BW) foi 899,3 g ($\pm 462,0$). A família 8 apresentou a maior média de BW (1186,04 \pm 406,8 g) e a família 6 a menor média de BW (595,3 \pm 318,3 g) (Tabela 2). A média geral de SL foi de 298,1 mm ($\pm 58,9$), H1 de 137,6 mm ($\pm 27,3$) e Wi de 52,6 mm ($\pm 11,4$). Em geral, as famílias 5 e 8 registraram os maiores valores médios para as medidas morfométricas, enquanto a família 3 revelou os menores índices (Tabela 2).

Em relação ao rendimento das partes corporais, as famílias 1 e 8 registraram menor média de % Cabeça (24,7), enquanto a família 11 demonstrou a maior média de % Cabeça (27,9); as famílias 6 e 12 com a menor e maior média de % Costela, com valores de 19,5 e 24,5, respectivamente; as famílias 11 e 12 com a menor e maior média de % Lombo, cujos valores foram 30,7 e 38,4, respectivamente. Destaca-se a família 11 em relação a proporção das medidas corpóreas, pois em média registrou maior valor de % Cabeça (27,9), mas com menor valor de % Lombo (30,7) e % Costela (20,2).

Tabela 1. Estatísticas descritivas tais como média, desvio padrão, coeficiente de variação, mínimo e máximo em machos e fêmeas das características avaliadas para desempenho de crescimento e rendimento corpóreo.

Variável	Média	SD ¹	CV ²	Macho	Fêmea	Min ³	Max ⁴
SL	298,1	58,9	5,06	268,09	326,99	146,06	419,97
HL	98,1	18	5,45	89,22	106,71	51,06	134,32
H1	137,6	27,3	5,04	123,91	150,88	66,96	202,93
H2	98,9	19,7	5,02	89,03	108,48	45,85	148,56
HH	89	15,8	5,63	82,09	95,91	47,66	133,02
PL	64,9	15	4,32	57,89	71,94	26,35	101,35
Wi	52,6	11,4	4,61	47,41	57,64	20	80
% Cabeça	25,5	2,3	11,08	26,13	24,97	20,8	33,8
% Costela	20,4	2,1	9,71	19,94	20,87	13,2	27,7
% Lombo	32,9	2,6	12,65	32,49	33,41	24,6	42,5
% C+L	53,2	3,4	15,64	52,44	54,28	37,8	70,2
BW	899,3	462	1,94	658,14	1143,49	92,29	2191,95
Peso cabeça	224,1	107,2	2,09	167,46	280,43	24,5	506,63
Peso costela	188,5	105,1	1,79	135,13	243,10	15,15	517,51
Peso lombo	299,5	160,2	1,86	217,02	385,20	28,18	799,48
Peso C+L	488,1	263,1	1,85	352,16	628,31	43,3	1223,2
Logr_Cabeça	-7,91482E-16	0,032	0	-0,001	-0,000	-0,145	0,093
Logr_Costela	7,18959E-16	0,039	0	0,002	-0,002	-0,187	0,129
Logr_Lombo	6,67029E-16	0,033	0	0,001	0,001	-0,096	0,110
Logr_C+L	1,7683E-15	0,023	0	0,001	8,79E-06	-0,095	0,116

¹Desvio padrão; ²Coeficiente de variação; ³Mínimo; ⁴Máximo.

Tabela 2. Média e desvio padrão para as características analisadas em tabaqui. O peso está em gramas (g) e as medidas morfométricas em milímetro (mm).

Família	SL ± SD	HL ± SD	H1 ± SD	H2 ± SD	HH ± SD	PL ± SD	Wi ± SD	BW ± SD	%Cabeça	%Costela	%Lombo	%C+L
1	279,8 ± 60,1	91,5 ± 18,4	127,8 ± 27,1	92,6 ± 20,8	84,7 ± 16,6	59,7 ± 16,1	48,7 ± 11,7	759 ± 389	24,7 ± 2,6	19,7 ± 2,4	33 ± 2,3	52,8 ± 3,3
3	251,1 ± 67,3	86 ± 20,2	117 ± 31,8	83,8 ± 22,4	79,9 ± 21,5	51,4 ± 16,1	43,1 ± 10,9	612,1 ± 484,2	26,6 ± 2,0	19,7 ± 2,2	32,4 ± 1,9	52,2 ± 3,0
5	334,2 ± 35	109,6 ± 11,5	152,2 ± 14,5	109,6 ± 11,1	96,1 ± 8,5	71,9 ± 9,4	58,4 ± 7,7	1151,2 ± 330,7	25,7 ± 1,9	20,7 ± 1,8	33,5 ± 2,3	54,3 ± 1,9
6	261,4 ± 42,7	90 ± 14,5	120,8 ± 19,6	87,5 ± 14,3	81,2 ± 12,3	56,4 ± 11,3	46,8 ± 9	595,3 ± 318,3	27 ± 2,1	19,5 ± 1,8	32,2 ± 1,9	51,7 ± 2,5
8	332,8 ± 49,1	108,8 ± 15,4	153,3 ± 21,5	109,8 ± 15,5	96,9 ± 13,3	73,1 ± 11,8	58,8 ± 9,7	1186,04 ± 406,8	24,7 ± 1,6	20,8 ± 2,1	33,8 ± 2,5	54,6 ± 3,7
9	321,7 ± 46,9	105,7 ± 12,4	151,7 ± 21,6	106,9 ± 15,9	93,2 ± 12,9	71,2 ± 13	57,7 ± 9,6	1104,1 ± 426,2	24,8 ± 1,7	21 ± 2,1	33,7 ± 2,1	54,8 ± 2,8
10	320,1 ± 54,4	105,3 ± 15,8	151,8 ± 26,2	109,7 ± 19,1	94,8 ± 13,4	68,8 ± 14,3	56,6 ± 11,2	1139,5 ± 559,3	25,2 ± 2,5	21,3 ± 2,4	32,1 ± 2,9	53,4 ± 3,7
11	272,4 ± 58,8	91,9 ± 19,1	126,4 ± 26,3	91,2 ± 19,5	83,8 ± 15,3	58,7 ± 14,4	49,5 ± 10,5	697,2 ± 398,1	27,9 ± 2,3	20,2 ± 1,8	30,7 ± 2,2	50,9 ± 3,0
12	292,1 ± 57,1	92,2 ± 17,3	131,6 ± 24,8	92,4 ± 16	81,4 ± 12,9	61,3 ± 15,1	48,4 ± 10,8	767,3 ± 323,6	26,5 ± 2,1	24,5 ± 2,6	38,4 ± 3,7	58,5 ± 5,1
15	275,3 ± 53,5	89,5 ± 15,9	126,8 ± 24,9	90,4 ± 18,1	84,5 ± 17	61,5 ± 14,7	47,9 ± 10,4	706,6 ± 374,6	25,2 ± 2,1	20,8 ± 2,1	33 ± 2,8	53,8 ± 3,7
18	312,2 ± 46,7	100,4 ± 14,5	141,3 ± 24,9	103,3 ± 16,4	93,3 ± 14,3	69,2 ± 12,2	55 ± 10,3	974,1 ± 399,7	25,2 ± 1,7	20,1 ± 1,8	33,8 ± 2,3	53,9 ± 1,9

Ao final do período de avaliação do crescimento, foi possível observar que as fêmeas apresentaram maior média de peso corporal do que os machos, sendo de 1.143,49 g (\pm 464,9) e 658,14 g (\pm 386,7), respectivamente. Para todas as características morfométricas, as fêmeas também apresentaram valores médios maiores que os machos, no entanto, para as características de rendimento corpóreo os valores foram similares entre os sexos. Os resultados das análises histológicas revelaram que os machos ainda estavam imaturos, com as gônadas macroscopicamente transparentes e pequenas. Poucos indivíduos machos estavam no início da espermatogênese (fase mitótica) (Fig. 2), caracterizado por cistos crescentes de espermatogônias do tipo B e alguns cistos de espermatócitos (Almeida *et al.*, 2016). As fêmeas também estavam todas imaturas, com os ovários macroscopicamente transparentes, pequenos e achatados. Não foram detectados oócitos vitelogênicos ou pré-ovulatórios, sendo a maioria das fêmeas classificadas com ovário em estágio de crescimento primário e oócitos perinucleolares (Figura 2) (Almeida *et al.*, 2016).

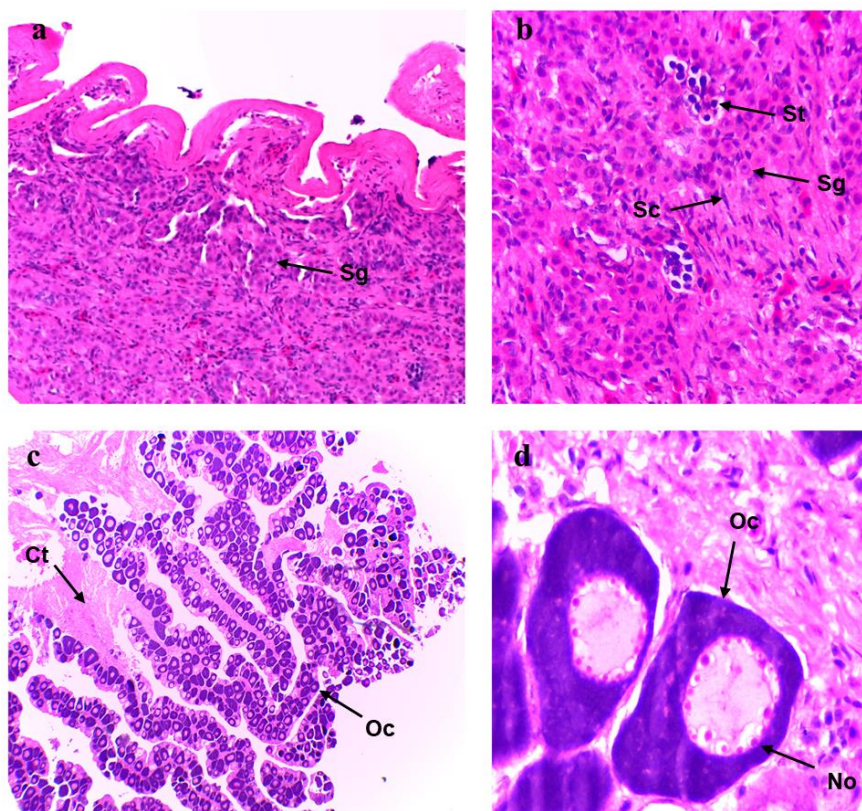


Figura 2. Testículo (a e b) e ovário (c e d) imaturos de tambaqui *Colossoma macropomum*. Sg:espermatogônia; St: espermatócito; Sc: célula de Sertoli; Ct: tecido conectivo; Oc: oócito crescente; No: nucléolo periférico.

Em relação fenotipagem por CVS, após a transformação linear de pixel para cm, estimaram-se as medidas morfométricas e áreas para cada animal de tambaqui. O resultado revelou que o CVS integrado com imagens de pacu pode ser implementado com confiança para estimar medidas morfométricas do tambaqui, pois o sistema mostrou-se resiliente mesmo em diferentes condições de fundo de imagem e espécie em questão, conforme demonstrado no resultado de segmentação automática (Figura 3). Os valores de correlação entre as medidas reais e preditos pelo CVS foram altos e positivos para as medidas morfométricas (Tabela 3), indicando que esse sistema pode ser utilizado para fenotipagem em larga escala e automatizada do tambaqui.

Tabela 3. Correlação fenotípica entre medidas corporais e áreas preditas por CVS e reais.

	SL	H1	HL	HH	Area C+L	Area Cabeça
Correlação	0.98	0.98	0.95	0.92	0.98	0.97

Em relação ao desempenho de classificação entre cabeça e nadadeiras versus corpo, foram detectadas diferenças visíveis em apenas cerca de 3% das imagens, conforme demonstrado na Figura 3. Isso ocorre principalmente porque as classes de cabeça e nadadeiras apresentavam o menor número de pixels em relação ao corpo, e devido à dificuldade de estabelecer os limites entre estas categorias ou com o fundo da imagem, mesmo para avaliadores humanos. Pela inspeção visual das classificações de pixel das diferentes classes (corpo, nadadeira e cabeça), foi possível observar as seguintes falhas (Figura 3): 1) os pixels preditos do corpo podem ser erroneamente atribuídos como parte da cabeça (A2); 2) os pixels preditos podem não preencher todo o espaço correspondente da parte da cabeça (B2); 3) os pixels preditos podem não preencher todo o espaço correspondente da respectiva parte corpórea (C2). A figura 3 (D2) é um exemplo de atribuição correta de todas as áreas (cabeça, nadadeiras e corpo) preditas pelo programa. Entretanto, não foram detectadas atribuições erradas entre as classes das partes do peixe com o fundo da imagem, ou seja, nenhum pixel do fundo da imagem foi reconhecido erroneamente como peixe.

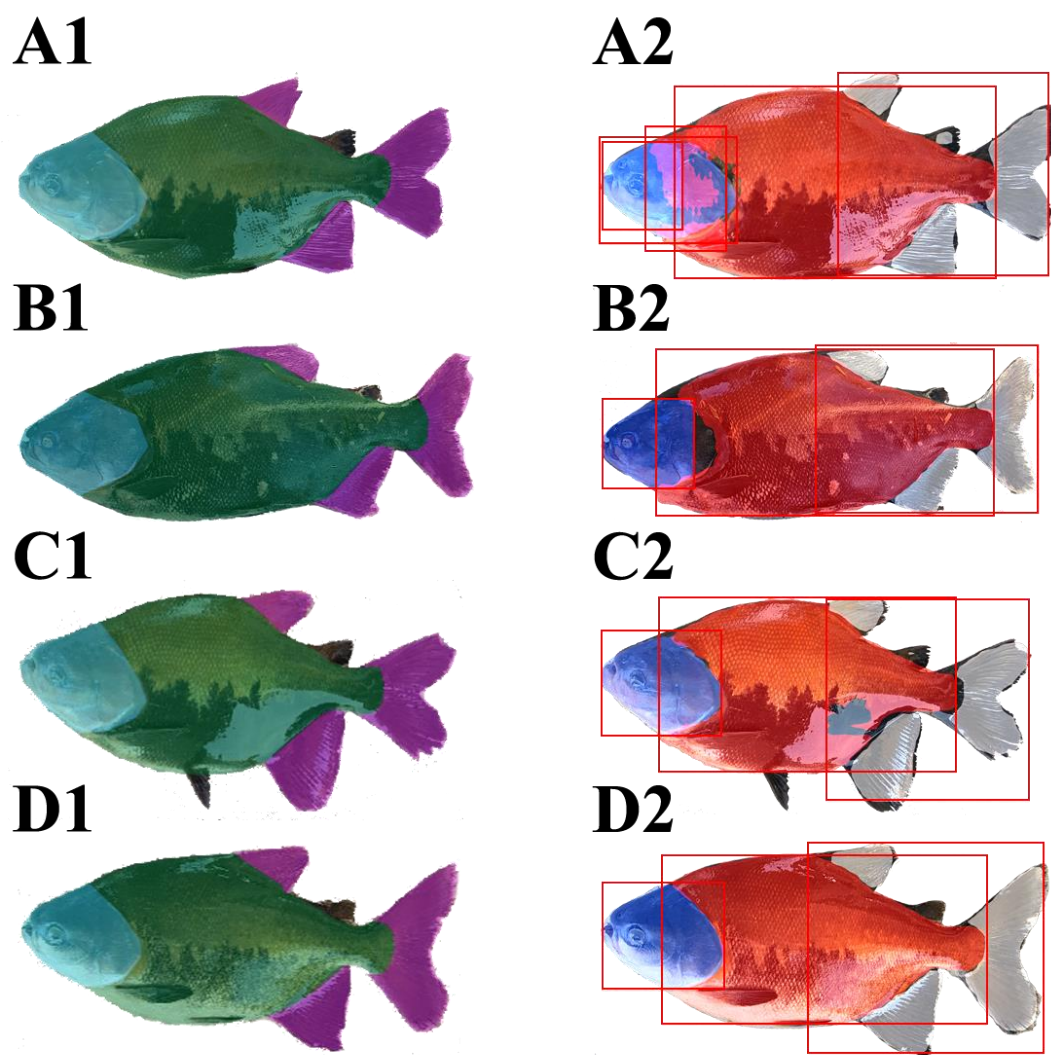


Figura 3. Comparação entre a segmentação manual (esperada) (A1, B1, C1 e D1) à esquerda; e segmentação automática do peixe usando CVS e deep learning (predita) nos peixes avaliados a direita (A2, B2, C2 e D2). Os quadrados vermelhos são gerados pelo sistema para calcular as áreas preditas e estabelecer limites entre as partes corpóreas avaliadas.

3.4.3.2 Parâmetros genéticos

Os valores estimados de herdabilidade e variância para cada característica estão descritos na Tabela 4. Em geral, as herdabilidades das características de crescimento (morfometrias e peso corporal) variaram de moderado a alto, com **W1** apresentando o menor valor ($h^2 = 0,24 \pm 0,11$) e **H2** o maior valor ($h^2 = 0,40 \pm 0,12$).

Em relação as estimativas de herdabilidade de rendimento (peso e % das partes corporais), os valores variaram de baixo a moderado ($h^2 = 0,10$ a $0,32$), com exceção de

% Costela que o valor não foi significativo. Os valores de herdabilidade dos resíduos da regressão das partes corporais foram não significativos, com exceção do Logr_Cabeça, que apresentou uma baixa herdabilidade ($h^2 = 0,15$).

Tabela 4. Estimativas de herdabilidade genômica, componentes de variância genética aditiva e variância residual para características de desempenho de crescimento utilizando modelo linear animal.

Trait	h^2 (SD)	σ_a^2 (SD)	σ_e^2 (SD)
SL (mm)	0,32 (0,12)	933,05 (399,93)	1927,20 (259,36)
HL (mm)	0,35 (0,13)	100,25 (45,43)	179,16 (26,49)
H1 (mm)	0,37 (0,12)	233,61 (93,32)	393,69 (56,92)
H2 (mm)	0,40 (0,12)	126,29 (48,46)	187,42 (28,51)
HH (mm)	0,27 (0,11)	56,70 (26,64)	146,58 (18,49)
PL (mm)	0,34 (0,12)	63,48 (27,11)	122,29 (17,10)
Wi (mm)	0,24 (0,11)	33,08 (17,44)	99,88 (12,40)
Área Cabeça (mm)	0,32 (0,12)	0.10539E+07 (0.46182E+06)	0.2253E+07 (0.30015E+06)
Área C+L (mm)	0,35 (0,12)	0.20651E+08 (0.84005E+07)	0.37664E+08 (0.52586E+07)
% Cabeça	0,27 (0,12)	1,43 (0,71)	3,70 (0,48)
% Costela	0,02 (0,04)	0,10 (0,18)	3,73 (0,32)
% Lombo	0,10 (0,07)	0,72 (0,52)	6,16 (0,58)
% C+L	0,17 (0,09)	1,72 (1,05)	8,37 (0,90)
BW (g)	0,32 (0,12)	62738 (27139)	131570 (17571)
Peso cabeça (g)	0,31 (0,12)	3274,80 (1459,60)	7094,80 (948,89)
Peso costela (g)	0,32 (0,12)	3174,40 (1389,40)	6743,20 (901,07)
Peso lombo (g)	0,32 (0,12)	7767,20 (3353,40)	15948 (2153,20)
Peso C+L (g)	0,32 (0,12)	20725 (8899,50)	42378 (5714,20)
Logr_Cabeça	0,15 (0,11)	0,159E-03 (0,119E-03)	0,849E-03 (0,974E-04)
Logr_Costela	0,01 (0,03)	0,147E-04 (0,465E-04)	0,139E-02 (0,116E-03)
Logr_Lombo	0,03 (0,04)	0,329E-04 (0,478E-04)	0,105E-02 (0,905E-04)
Logr_C+L	0,01 (0,03)	0,649E-05 (0,171E-04)	0,445E-03 (0,377E-04)

3.4.3.3 Correlações fenotípicas e genéticas

As correlações genéticas e fenotípicas para as características analisadas estão apresentadas na Tabela 5. Todas as correlações genéticas entre as medidas de morfometria e peso das partes corporais foram altas e favoráveis, com exceção do valor moderado entre Peso Cabeça~H1 (0,492). No geral, as correlações entre os resíduos da regressão das partes corporais e características morfométricas foram não significativas (dados não demonstrados).

As correlações genéticas entre % das partes corporais e características morfométricas foram favoráveis e altamente correlacionados, com exceção de % cabeça cujas correlações foram desfavoráveis (Tabela 6). A medida PL está mais correlacionada com a maioria das % das partes corporais (Cabeça, Lombo e C+L). Em relação % Costela, a medida Area Cabeça está mais correlacionada e com menor desvio padrão.

As correlações genéticas entre % e resíduos da regressão das respectivas partes corporais foram positivas e altamente correlacionados (Tabela 7).

Em relação as correlações fenotípicas, os valores encontrados entre as características de morfometria e peso das partes corporais foram altos e favoráveis, variando de 0,839 a 0,990.

Tabela 5. Correlação genética \pm desvio padrão (acima da diagonal) e correlações fenotípicas (abaixo da diagonal) para características de crescimento e rendimento em tambaqui (*Collossoma macropomum*).

	BW	SL	HL	H1	H2	HH	PL	Wi	Área C+L	Área cabeça	Peso C+L	Peso Lombo	Peso Costela	Peso Cabeça
BW	*	0,996 \pm 0,003	0,995 \pm 0,00	1 \pm 0,00	1 \pm 0,00	1 \pm 0,00	1 \pm 0,00	0,992 \pm 0,00	0,999 \pm 0,00	0,998 \pm 0,00	0,999 \pm 0,00	0,999 \pm 0,00	1 \pm 0,00	0,997 \pm 0,01
SL	0,953	*	0,992 \pm 0,00	1 \pm 0,00	0,997 \pm 0,00	1 \pm 0,00	0,999 \pm 0,00	0,992 \pm 0,00	0,775 \pm 0,00	0,993 \pm 0,00	0,996 \pm 0,00	0,999 \pm 0,00	0,988 \pm 0,01	0,995 \pm 0,00
HL	0,920	0,960	*	0,998 \pm 0,00	0,999 \pm 0,00	0,995 \pm 0,00	0,992 \pm 0,00	0,832 \pm 0,00	0,995 \pm 0,00	1 \pm 0,00	1 \pm 0,00	0,999 \pm 0,00	0,993 \pm 0,00	1 \pm 0,00
H1	0,940	0,974	0,945	*	1 \pm 0,00	0,996 \pm 0,00	0,995 \pm 0,00	0,999 \pm 0,00	1 \pm 0,00	1 \pm 0,00	1 \pm 0,00	1 \pm 0,00	0,967 \pm 0,18	0,492 \pm 0,212
H2	0,925	0,949	0,920	0,955	*	0,996 \pm 0,00	0,993 \pm 0,00	1 \pm 0,00	0,998 \pm 0,00	1 \pm 0,00	0,999 \pm 0,00	1 \pm 0,00	0,997 \pm 0,00	0,998 \pm 0,15
HH	0,881	0,911	0,892	0,898	0,947	*	0,940 \pm 0,01	0,999 \pm 0,00	0,992 \pm 0,00	0,995 \pm 0,00	0,999 \pm 0,00	1 \pm 0,00	0,995 \pm 0,00	0,997 \pm 0,00
PL	0,897	0,919	0,867	0,90	0,930	0,908	*	0,997 \pm 0,00	0,988 \pm 0,00	1 \pm 0,00	0,992 \pm 0,00	0,988 \pm 0,00	0,989 \pm 0,00	0,990 \pm 0,00
Wi	0,929	0,931	0,897	0,927	0,908	0,864	0,873	*	0,987 \pm 0,00	0,997 \pm 0,00	0,989 \pm 0,00	0,976 \pm 0,00	0,993 \pm 0,00	1 \pm 0,00
Area C+L	0,980	0,963	0,912	0,951	0,931	0,878	0,911	0,860	*	0,997 \pm 0,00	0,999 \pm 0,00	0,999 \pm 0,00	1 \pm 0,00	1 \pm 0,00
Area Cabeça	0,958	0,942	0,934	0,922	0,907	0,863	0,873	0,852	0,962	*	0,997 \pm 0,00	0,996 \pm 0,00	0,997 \pm 0,00	0,999 \pm 0,00
Peso C+L	0,995	0,936	0,896	0,936	0,916	0,870	0,894	0,870	0,974	0,944	*	0,999 \pm 0,00	0,999 \pm 0,00	0,988 \pm 0,00
Peso Lombo	0,984	0,923	0,887	0,920	0,901	0,856	0,882	0,869	0,960	0,933	0,990	*	0,999 \pm 0,00	0,987 \pm 0,00
Peso Costela	0,975	0,922	0,876	0,927	0,905	0,859	0,880	0,839	0,958	0,925	0,977	0,939	*	0,991 \pm 0,00
Peso Cabeça	0,982	0,943	0,922	0,929	0,905	0,864	0,878	0,865	0,965	0,967	0,970	0,965	0,941	*

Tabela 6. Correlações genéticas e desvio padrão entre características morfométricas e características de rendimento expressa em porcentagem.

	% CABEÇA	% COSTELA	% LOMBO	% C+L
BW	-0,838 ± 0,12	0,999 ± 0,26	0,907 ± 0,19	0,952 ± 0,05
SL	-0,827 ± 0,14	0,990 ± 0,00	0,924 ± 0,18	0,954 ± 0,04
HL	-0,767 ± 0,20	0,993 ± 0,00	0,876 ± 0,21	0,928 ± 0,06
H1	-0,837 ± 0,13	0,995 ± 0,00	0,895 ± 0,16	0,947 ± 0,05
H2	-0,819 ± 0,17	0,988 ± 0,00	0,891 ± 0,20	0,937 ± 0,05
HH	-0,824 ± 0,16	0,999 ± 0,01	0,950 ± 0,22	0,964 ± 0,04
PL	-0,868 ± 0,11	0,950 ± 0,32	0,976 ± 0,21	0,987 ± 0,02
WI	-0,842 ± 0,17	0,997 ± 0,36	0,919 ± 0,30	0,948 ± 0,05
AREA CABEÇA	-0,775 ± 0,26	0,999 ± 0,00	0,854 ± 0,27	0,925 ± 0,01
AREA C+L	-0,835 ± 0,12	0,988 ± 0,00	0,898 ± 0,36	0,945 ± 0,01

Tabela 7. Correlações genéticas e desvio padrão entre características de rendimento expressa como porcentagem e log-residual.

Traits	Genetic correlation	SD
Logr_Cabeça ~ % Cabeça	0,980	0,00
Logr_Costela ~ % Costela	0,955	0,00
Logr_Lombo ~ % Lombo	0,986	0,04
Logr_C+L ~ % C+L	0,970	0,00

3.4.3.4 Análises genômicas

O genoma do tabaqui disponível no NCBI tem um comprimento total de 1,22 Gb, fragmentado em 1.269 Scaffolds (Hilsdorf *et al.*, 2021). O mapa de ligação utilizado para atribuir o genoma em nível cromossômico contém 17,374 SNPs (Capítulo III). Após o alinhamento dos *loci* dos SNPs do mapa de ligação no genoma de referência por BWA, o resultado do programa Chromonomer registrou apenas 3,772 marcadores no mapa final, agrupando 172 Scaffolds. Entretanto, esse resultado permitiu construir um genoma com cobertura de 95% da informação inicial, derivando em um genoma em nível cromossômico com comprimento total de 1,17 Gb, sendo que somente 0,05 Gb não foram ancorados em pseudomoléculas (cromossomos) (Tabela 8).

Tabela 8. Comprimento dos cromossomos após atribuição de SNPs nos 27 grupos de ligação do tabaqui utilizando o software Chromonomer.

Chromosome	lengths
LG1	67,263,696bp
LG2	53,233,441bp
LG3	49,677,613bp
LG4	47,201,105bp
LG5	46,638,033bp
LG6	56,555,689bp
LG7	47,628,250bp
LG8	43,745,175bp
LG9	45,465,291bp
LG10	45,648,328bp
LG11	40,542,391bp
LG12	40,668,426bp
LG13	40,889,137bp
LG14	43,985,697bp
LG15	40,261,227bp
LG16	43,836,131bp
LG17	36,041,427bp
LG18	38,372,523bp
LG19	36,813,950bp
LG20	33,938,749bp
LG21	39,081,045bp
LG22	42,743,206bp
LG23	36,663,456bp
LG24	35,261,499bp
LG25	35,884,927bp
LG26	34,444,640bp
LG27	33,262,948bp
Total length	1,155,748,000bp

Em relação aos dados de ddRADseq, 1.918.598.642 *reads* brutas foram sequenciadas, resultando em aproximadamente 300 Giga bases (Gb) (aproximadamente 30 Gb por biblioteca). Após os parâmetros de filtro (remoção de sequências de baixa qualidade e *reads* com *barcodes* ausentes ou ambíguos), uma média de 22% das *reads* foram removidas de cada biblioteca, ou seja, 78% das leituras foram retidas para análise. Além disso, 46 indivíduos foram excluídos devido ao baixo número de *reads* (<2M). O

número de *reads* por indivíduo variou de 3.2 M a 5.0 M, com uma média de 4.3 M. Dados brutos do sequenciamento e *reads* filtradas de cada biblioteca são mostrados na Tabela 9.

Após o mapeamento das ddRAD *reads* no genoma em nível cromossômico, as análises do programa Stacks resultaram em 94.232 SNPs (84.407 loci) para 296 indivíduos, que passaram nos controles de qualidade. No software Plink, após o filtro mind e geno, foram excluídos 19 indivíduos e 28.053 SNPs. Os filtros de MAF e HWE excluíram 111 e 5.398 SNPs, respectivamente. Sendo assim, 277 indivíduos e 60.670 SNPs passaram em todos os controles de qualidade e foram utilizados para as análises de associação genômica ampla (GWAS).

As herdabilidades genômicas estimadas com o método ssGBLUP foram similares aos valores obtidos por PBLUP (dados não mostrados). Um total of 3 SNPs com significância cromossômica foram associados a oito características de crescimento/rendimento (HL, SL, Peso corporal, Area da Cabeça, Área C+L, peso C+L, peso Cabeça e Peso costela), distribuídos em dois grupos de ligação (LG 11 e 19) (Figura 4), com a variância genética variando de 0.012 a 0.232% (Tabela 10). O restante das características não teve SNPs associados com significância estatística. A análise de genes candidatos usando os top SNPs contra o genoma de *Colossoma macropomum* revelou alguns genes associados ao crescimento localizados próximos aos QTLs identificados (Tabela 10), tais como *ebf3a*, *ptprea* e *adamts17*.

Tabela 9. Marcadores SNP associados ao crescimento e rendimento de filé em tambaqui. A localização e posição são em relação ao genoma de referência de *Colossoma macropomum*. Os genes candidatos estão localizados dentro da janela de 1 Mb.

Biblioteca	Sequências Totais	Reads com sequência do adaptador	Barcode Não encontrado	Baixa Qualidade	Sítio de corte não encontrado	Reads retidas	Média de reads por indivíduo (milhões)
B1	272.744.834	11.260.275	61.873.920	1.154.886	16.362.650	182.093.103	4.9
B2	301.052.914	10.109.525	66.710.494	1.242.188	26.268.814	196.721.893	5.0
B3	179.765.880	8.115.131	14.690.396	916.844	11.739.842	144.303.667	3.2
B4	206.981.034	7.330.031	35.351.580	1.389.827	6.693.769	156.215.827	3.9
B5	260.721.062	8.935.263	20.110.722	1.932.199	9.631.535	220.111.343	5.0
B6	208.633.548	8.336.530	8.723.618	1.621.837	7.503.100	182.448.463	3.9
B7	245.120.976	9.415.215	11.435.444	2.769.136	8.797.751	212.703.430	4.6
B8	243.578.394	14.208.417	10.385.430	2.752.302	9.233.638	206.998.607	4.5

Tabela 10. Resultado do sequenciamento ddRADseq para *C. macropomum* para cada biblioteca e número de *reads* excluídas após filtros de qualidade.

SNP	Posição (bp)	Variância Genética (%)	-Log ₁₀ (P-value)	Nível de Significância	Característica	GL	Localização (scaffold)	Genes
47254_96	30851409	0.232; 0.213	4.75; 4.61	Chromosome	HL, SL	11	NW_023494790.1	ebf3a; glrx3;
47098_63	30428771	0.225	4.69	Chromosome	Peso Cabeça	11	NW_023494790.1	ebf3a; ptprea; foxI1
143706_49	10282970	0.012; 0.019; 0.026; 0.013; 0.012	4.98; 4.77; 4.85; 5.21; 4.86	Chromosome	Peso; Área Cabeça; Área C+L; Peso Costela; Peso C+L	19	NW_023494796.1	kdm4c; adamts17; pax5; rnf38; pgm2; melk; zcchc7

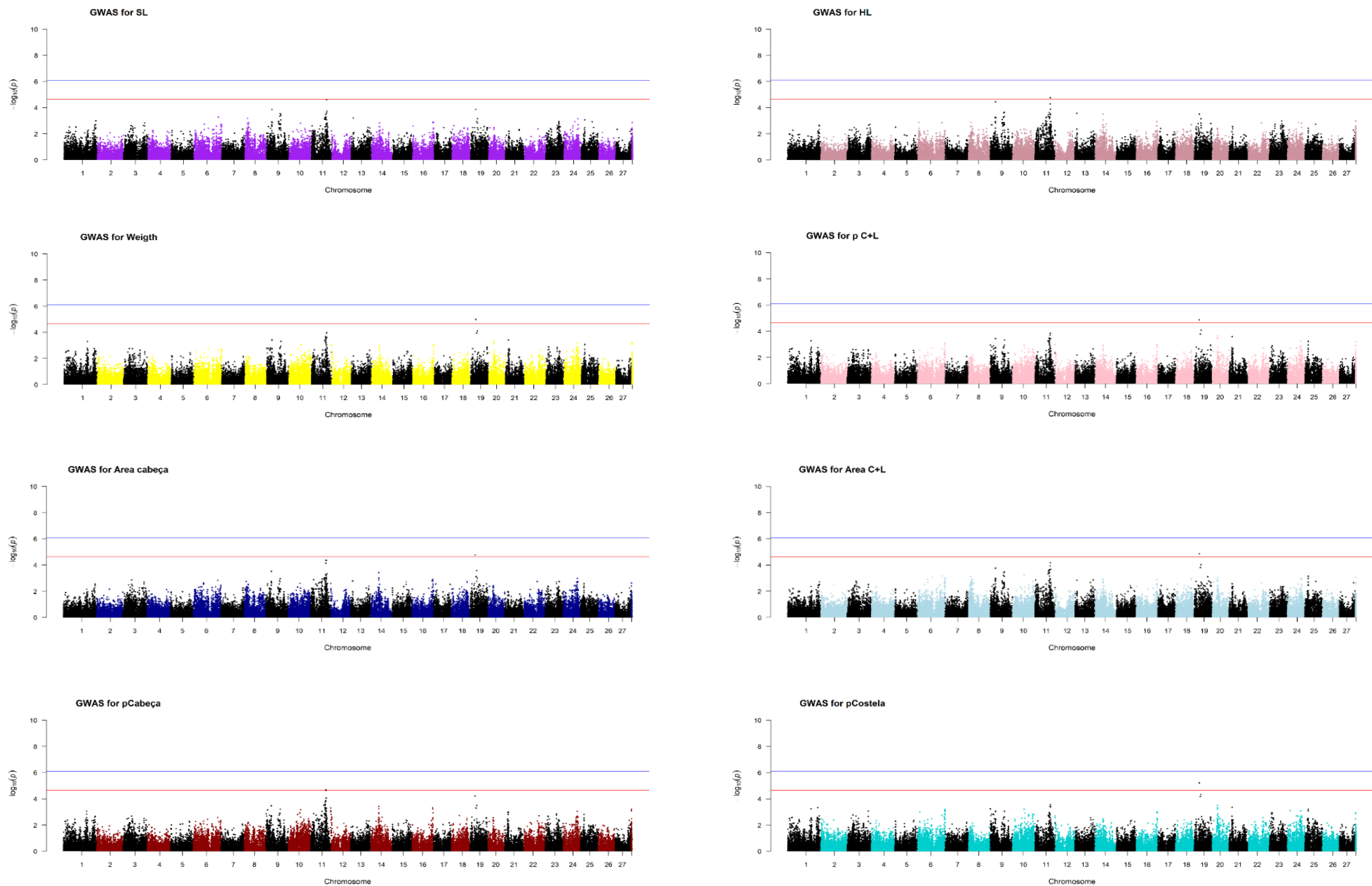


Figura 4. Estudo de associação genômica ampla (GWAS) para características de crescimento e rendimento de filé em tabaqui por ssGBLUP. Gráfico de Manhattan com $-\log_{10}$ (valores p) distribuídos em diferentes grupos de ligação. Os marcadores que ultrapassaram a significância do genoma (linha vermelha) e / ou do cromossomo (linha azul) após a correção de Bonferroni foram considerados em associação com o crescimento e rendimento. Os marcadores não atribuídos a um grupo de ligação foram agrupados no lado direito do gráfico, após o grupo de ligação 27.

3.4.4 Discussão

Em geral, as estimativas de herdabilidade para características de crescimento (morfometrias e peso corporal) variaram de moderada a alta para a população desse estudo ($h^2 = 0,24-0,40$), sugerindo a presença de componentes genéticos na herança dessas características e que provavelmente responderão ao processo de seleção genética. Os resultados corroboram com valores descritos para outras espécies de peixes, como seabass ($h^2 = 0,29-0,61$) (Dupont-Nivet *et al.*, 2010), tilápia do Nilo ($h^2 = 0,19-0,55$) (Gjerde *et al.*, 2012; Trong *et al.*, 2013), pacu ($h^2 = 0,27-0,74$) (Freitas *et al.*, 2020), e também em tambaqui ($h^2 = 0,22-0,49$) (Mello *et al.*, 2016). Todavia, valores ligeiramente menores foram reportados por Perazza *et al.* (2019) em tambaqui para medidas equivalentes, porém, deve-se levar em consideração que estes valores podem variar de acordo com a população analisada, dentre outras variáveis, como tamanho do tanque, densidade de animais por m³, idade, etc.

Além do peso corpóreo e morfometria, outra característica que merece destaque na aquicultura é o rendimento de partes corporais. Em geral, o rendimento do filé tem baixa variação fenotípica (coeficiente de variação na faixa de 2% a 6%) e o peso do filé é proporcional ao peso corporal (Vandeputte *et al.*, 2019). No presente estudo, o coeficiente de variação oscilou de 9,71 a 15,64%. A variação de % C+L foi alta (37,8-60,4%), sendo maiores do que os relatados para outras espécies como tilápia, bagre, carpa e robalo (Nguyen *et al.*, 2010; Sang *et al.*, 2009; Cibert *et al.*, 1999; Saillant *et al.*, 2009), e similar ao reportado para salmonídeos, como a truta arco-íris e salmão do Atlântico (Einen *et al.*, 1999; Kause, *et al.*, 2007; Powell *et al.*, 2008).

Neste estudo, as herdabilidades para rendimento corpóreo (%) variaram de baixo a moderado (0,10 a 0,27). Esses resultados indicam menores valores em relação a um

estudo recente que estimou alta herdabilidade para rendimento corpóreo de duas características no tambaqui, sendo a área do lombo ($h^2 = 0,39$) e espaço entre as costelas ($h^2 = 0,50$) (Perazza *et al.*, 2019), e com outros estudos relatados para carpa comum, tilápia do Nilo e seabass (0,20-0,35) (Kocour *et al.*, 2007; Saillant *et al.*, 2009; Nguyen *et al.*, 2010; Vandeputte *et al.*, 2017), o que indica que estas características estão sob controle genético.

As herdabilidades dos pesos das partes corpóreas foram semelhantes à do peso corporal, como geralmente visto neste tipo de estudos, pois há mais variação fenotípica no tamanho e peso corporal do que nas proporções das diferentes partes do corpo. Devido o cálculo do rendimento ser uma proporção e as respostas esperadas à seleção sobre os rendimentos serem tendenciosas (Gunsett, 1987), estudos recentes tem utilizado o resíduo da regressão (Logr) do peso do filé sobre o peso corporal como substituto para o rendimento do filé, o que tem demonstrado ser mais eficiente para predição do rendimento do filé (Haffray *et al.*, 2013; Haffray *et al.*, 2012; Prchal *et al.*, 2018; Vandeputte *et al.*, 2017; Vandeputte *et al.*, 2014). Entretanto, os resultados desse estudo demonstraram valores não significativos para herdabilidades de log residual, com exceção da cabeça, o que pode ser atribuído a baixa variação fenotípica devido ao baixo número amostral. Em estudos prévios, os valores são similares entre Logr e % de rendimento corpóreo para seabass (Vandeputte *et al.*, 2017), o que diferiu dos valores registrados para a característica de cabeça no presente estudo, que foi menor em log residual (0,15) do que em % (0,27).

No geral, as estimativas de herdabilidade para desempenho de crescimento (peso e características morfométricas) e rendimento corpóreo sugerem que essas características estão sob moderado-alto controle genético e devem responder à seleção. Além disso, algumas características morfométricas foram favoráveis e altamente correlacionados com

rendimento de partes corporais, implicando que a seleção simultânea pode ser alcançada entre essas características de crescimento.

Em relação às correlações genéticas, em geral, o peso foi altamente correlacionado com todas as características morfométricas, sugerindo que ganhos genéticos de peso podem ser obtidos por respostas correlacionadas dessas características. Altas correlações entre medidas corporais e peso também foi observado em outras espécies, como pacu (*Piaractus mesopotamicus*) (Freitas *et al.*, 2020), linguado (*Paralichthys olivaceus*) (Li *et al.*, 2018; Liu *et al.*, 2011; Tian *et al.*, 2011) e tilápia do Nilo (Rutten *et al.*, 2005). A correlação genética entre o peso das partes corporais e o peso total é próximo a unidade, similarmente a prévios estudos (Gjerde *et al.*, 2012, Haffray *et al.*, 2012, Navarro *et al.*, 2009, Nguyen *et al.*, 2010, Powell *et al.*, 2008, Rutten *et al.*, 2005).

A seleção para aumentar o rendimento do filé ou de partes corporais é um desafio porque essa característica não pode ser medida em animais vivos. Os métodos de seleção possíveis são a seleção de irmãos, em que os valores genéticos dos candidatos à seleção são estimados com base em medições feitas em irmãos abatidos, ou seleção indireta, em que características mensuráveis em candidatos de seleção vivos e geneticamente correlacionados ao rendimento de filé são usados como critérios de seleção indireta (Haffray *et al.*, 2013). Rendimento de Lombo, Costela e C+L estão fortemente correlacionados com peso corpóreo, mas outras medidas morfométricas demonstraram uma correlação genética mais alta, como por exemplo, PL~%Lombo, PL~%C+L e Area da Cabeça~%Costela, indicando serem critérios de seleção eficientes para rendimento corpóreo nessa população de tambaqui.

Uma das características desfavoráveis para a produção do tambaqui é o grande tamanho da cabeça, que não é atraente para o consumidor por diminuir o rendimento do filé (Mello *et al.*, 2016), como demonstrado no presente estudo em que pode alcançar até

33,8% do peso total. As análises genéticas demonstraram que o % Cabeça registrou valores altos e negativos de correlação genética com peso total e medidas morfométricas, como por exemplo PL~%Cabeça (-0,868), de forma similar com a tilápia (-0,94, Rutten *et al.*, 2005). Este resultado revela que é possível selecionar animais com menor proporção de cabeça usando características indiretas e eficientes como critério de seleção, o que é desejável para a indústria da aquicultura do tambaqui.

Um outro resultado interessante do presente estudo é a validação de um método de fenotipagem não invasivo para obter as medidas morfométricas do tambaqui. Com essa ferramenta, será possível realizar a predição do rendimento de filé sem sacrifício do animal, por meio de biometria automatizada, reduzindo o tempo de coleta de dados e evitando o estresse animal por manejo excessivo. Estas vantagens permitem uma fenotipagem em larga escala, o que traz avanços para acelerar o melhoramento genético dessa espécie, similarmente ao realizado para tilápia (Fernandes *et al.*, 2020) e pacu (Freitas *et al.*, em preparação).

A maioria das características economicamente importantes, como crescimento e rendimento, são características quantitativas e complexas afetadas por vários genes (Ali *et al.*, 2020; Gong *et al.*, 2021). Neste estudo, por meio das análises de GWAS, não foi descoberto regiões genômicas significativamente associadas em nível genômico. Apenas três loci revelaram associação significativa em nível cromossômico para algumas das características (8 de 22 características).

A investigação dos genes candidatos detectou que o SNP 47254_96 está próximo dos genes *ebf3a* (EBF transcription factor 3a) envolvido na regulação positiva da transcrição pela RNA polimerase II; e *glrx3* (glutaredoxin 3), envolvido na maturação da hemoglobina. O SNP 47098_63 está próximo dos genes *ptprea* (Protein tyrosine phosphatase receptor type Ea), envolvido na atividade da proteína tirosina fosfatase, além

disso é um regulador negativo do sinal do receptor de insulina no músculo esquelético (Bak *et al.* 2008), e seus parálogos (*ptpra*, *ptprd* e *ptprf*) são conhecidos como genes de susceptibilidade à obesidade (Ahmad *et al.* 1997); e gene *foxI1* (Forkhead box protein I1). O SNP 143706_49 está próximo dos genes *kdm4c* (lysine demethylase 4C) que regula a expressão gênica e a segregação cromossômica; *adamts17* (ADAM Metallopeptidase With Thrombospondin Type 1 Motif 7) que regula a migração das células do músculo liso vascular; *pax5* (Paired box protein Pax-5) que regula a adesão e migração celular e facilita a sinalização do receptor de células pré-B promovendo o desenvolvimento de células B maduras; *rnf38* (Ring Finger Protein 38) desempenha papel em diversos processos celulares, incluindo transdução de sinal e apoptose; *pgm2* (Phosphoglucomutase-2) associado ao processo metabólico da glicose; *melk* (Maternal Embryonic Leucine Zipper Kinase) regulação do ciclo celular, autorrenovação de células-tronco, apoptose e regulação de splicing; e por fim, o gene *zcchc7* (Zinc Finger CCHC-Type Containing 7) que atua na ligação de RNA, e ligação de íon zinco.

Dentre os genes identificados próximos aos SNPs, vale ressaltar o gene *ptpre*, que já foi identificado em um estudo de GWAS em garoupa (*Epinephelus lanceolatus*) demonstrando que pode ser um gene comum em atuar no crescimento de peixes (Wu *et al.*, 2019). Além disso, *ptpre* também provou ter funções de crescimento conhecidas ou estar associado a características de crescimento em humanos e outros animais (Li *et al.* 2018; Vaughan *et al.* 2015; Wang *et al.* 2016). Por exemplo, um estudo de GWAS para peso em suínos identificou *ptpre* como um gene candidato próximo ao SNP significativamente associado (Wang *et al.* 2016). Em peixe, foi identificado que este gene desempenha um papel crucial na resistência à insulina responsável pela obesidade no fígado (Wu *et al.*, 2019). Outro gene de destaque é o *glrx3*, associado a endotermia dos peixes, pois depende principalmente da atividade do músculo vermelho, que produz calor

como um subproduto durante a contração muscular (por exemplo, atum e tubarões-brancos) (Wang *et al.*, 2021). O gene *ebf3a* também foi identificado em um estudo de GWAS com tabaqui para identificar genes relacionados a falta de ossos intermusculares, demonstrando que esse gene pode influenciar diferentes características nesta espécie (Nunes *et al.*, 2020).

No presente estudo, poucos loci associados a características quantitativas importantes para a produção de filé e peso em tabaqui foram detectados. O pequeno efeito desses loci reforça as evidências da natureza poligênica dessas características. Nossos resultados corroboram resultados anteriores que mostraram a natureza poligênica do rendimento de filé e características relacionadas ao crescimento em diferentes espécies de aquicultura, sem evidência de genes de efeito importante ou regiões genômicas avaliadas por GWAS (Gutierrez *et al.* 2015; Tsai *et al.* 2015; Gonzalez-Pena *et al.* 2016; Yoshida *et al.* 2019).

3.4.5 Referências

- Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta S., Lawlor T.J. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science*, 93(2), 743-752.
- Ahmad, F., Azevedo, J.L., Cortright, R., Dohm, G.L., Goldstein, B.J. (1997). Alterations in skeletal muscle protein-tyrosine phosphatase activity and expression in insulin-resistant human obesity and diabetes. *J Clin Invest* 100:449–458.
- Ali, A., Al-Tobasei, R., Lourenco, D., Leeds, T., Kenney, B., Salem, M. (2020). Genome-wide identification of loci associated with growth in rainbow trout. *BMC genomics*, 21(1), 1-16.
- Almeida, F. D., Lopes, J. S., Crescencio, R., Izel, A. C. U., Chagas, E. C., Boijink, C. (2016). Early puberty of farmed tabaqui (*Colossoma macropomum*): Possible influence of male sexual maturation on harvest weight. *Aquaculture*, 452, 224-232.

- Ariede, R. B., Freitas, M. V., Agudelo, J. F., Borges, C. H., Lira, L. V., Yoshida, G. M., Pilarski, F., Yáñez, J.M., Hashimoto, D. T. (2020). Genetic (co) variation between resistance to *Aeromonas hydrophila* and growth in tambaqui (*Colossoma macropomum*). *Aquaculture*, 523, 735225. <https://doi.org/10.1016/j.aquaculture.2020.735225>
- 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*. *Front. Genet.* 9, 3. <https://doi.org/10.3389/fgene.2018.00003>.
- Bak, A., Jacob, A.I., Aga-Mizrachi, S., Brutman-Barazani, T., Sampson, S.R., Elson, A. (2008). Cytosolic protein tyrosine phosphatase- ϵ is a negative regulator of insulin signaling in skeletal muscle. *Endocrinology* 149:605–614.
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait J. H. (2011). Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes|genomes|genetics*, 1(3), 171-182.
- Catchen, J., Amores, A., & Bassham, S. (2020). Chromonomer: a tool set for repairing and enhancing assembled genomes through integration of genetic maps and conserved synteny. *G3: Genes, Genomes, Genetics*, 10(11), 4115-4128.
- Cibert, C., Fermon, Y., Vallod, D.J., Meunier, F. (1999). Morphological screening of carp *Cyprinus carpio*: relationship between morphology and fillet yield. *Aquatic Living Resources* 12, 1–10.
- Dupont-Nivet, M., Chevassus, B., Mauger, S., Haffray, P., Vandeputte, M. (2010). Side effects of sexual maturation on heritability estimates in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture Research* 41, e878–e880.
- Einen, O., Mørkøre, T., Rørå, A.M.B., Thomassen, M.K. (1999). Feed ration prior to slaughter - a potential tool for managing product quality of Atlantic salmon (*Salmo salar*). *Aquaculture* 178, 149–169.
- Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to quantitative genetics*. 1996. Harlow, Essex, UK: Longmans Green, 3.
- Fernandes, A. F., Turra, E. M., de Alvarenga, É. R., Passafaro, T. L., Lopes, F. B., Alves, G. F., Singhc, V., Rosa, G.J.M. (2020). Deep Learning image segmentation for extraction of fish body measurements and prediction of body weight and carcass traits in Nile tilapia. *Computers and electronics in agriculture*, 170, 105274.

- Fraslin, C., Dupont-Nivet, M., Haffray, P., Bestin, A., Vandeputte, M. (2018). How to genetically increase fillet yield in fish: new insights from simulations based on field data. *Aquaculture*, 486, 175-183.
- Freitas, M. V., Lira, L. V., Ariede, R. B., Agudelo, J. F., Borges, C. H., Mastrochirico-Filho, V. A., Garcia Neto, B. F., Carneiro, R., Hashimoto, D.T. (2020). Genotype by environment interaction and genetic parameters for growth traits in the Neotropical fish pacu (*Piaractus mesopotamicus*). *Aquaculture*, 735933. <https://doi.org/10.1016/j.aquaculture.2020.735933>
- Gjedrem, T. (2012). Genetic improvement for the development of efficient global aquaculture: a personal opinion review. *Aquaculture* 344–349, 12–22. <https://doi.org/10.1016/j.aquaculture.2012.03.003>.
- Gjerde, B., Mengistu, S. B., Ødegård, J., Johansen, H., Altamirano, D. S. (2012). Quantitative genetics of body weight, fillet weight and fillet yield in Nile tilapia (*Oreochromis niloticus*). *Aquaculture*, 342, 117-124.
- Gong, J., Zhao, J., Ke, Q., Li, B., Zhou, Z., Wang, J., Zhou, T., Zheng, W., Xu, P. (2021). First genomic prediction and genome-wide association for complex growth-related traits in Rock Bream (*Oplegnathus fasciatus*). *Evolutionary Applications*.
- Gonzalez-Pena, D., Gao, G., Baranski, M., Moen, T., Cleveland, B.M., Kenney, P.B., Vallejo, R.L., Palti, Y., Leeds, T.D. (2016). Genome-wide association study for identifying loci that affect fillet yield, carcass, and body weight traits in rainbow trout (*Oncorhynchus mykiss*). *Front Genet* 7:203.
- Gunsett, F.C. (1987). Merit of utilizing the heritability of a ratio to predict the genetic change of a ratio. *J. Anim. Sci.* 65, 936–942.
- Gutierrez, A. P., Yáñez, J. M., Fukui, S., Swift, B., Davidson, W. S. (2015). Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic salmon (*Salmo salar*). *PloS one*, 10(3), e0119730.
- Haffray, P., Bugeon, J., Pincet, C., Chapuis, H., Mazeiraud, E., Rossignol, M.-N., Chatain, B., Vandeputte, M., Dupont-Nivet, M., (2012). Negative genetic correlations between production traits and head or bony tissues in large all-female rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*:368–369 145–152. [10.1016/j.aquaculture.2012.09.023](https://doi.org/10.1016/j.aquaculture.2012.09.023).
- Haffray, P., Bugeon, J., Rivard, Q., Quittet, B., Puyo, S., Allamelou, J.M., Vandeputte, M., Dupont-Nivet, M. (2013). Genetic parameters of in-vivo prediction of carcass, head and fillet yields by internal ultrasound and 2D external imagery in large

- rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*:410–411 236–244. 10.1016/j.aquaculture.2013.06.016.
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).
- Hilsdorf, A. W. S., Silva, M. U., Coutinho, L. L., Montenegro, H., Almeida-Val, V. M. F., Pinhal, D. (2021). Genome assembly and annotation of the tambaqui (*Colossoma macropomum*): an emblematic fish of the Amazon River basin. *bioRxiv*.
- Houle, D., Govindaraju, D. R., & Omholt, S. (2010). Phenomics: the next challenge. *Nature reviews genetics*, 11(12), 855-866.
- IBGE. (2019). Pesquisa Pecuária Municipal – 2018. Disponível em: <https://sidra.ibge.gov.br/pesquisa/ppm/quadros/brasil/2018>
- Kause, A., Paananen, T., Ritola, O., Koskinen, H., (2007). Direct and indirect selection of visceral lipid weight, fillet weight, and fillet percentage in a rainbow trout breeding program. *Journal of Animal Science* 85, 3218–3227.
- Koch, R.M., Swiger, L.A., Chambers, D., Gregory, K.E. (1963). Efficiency of feed use in beef cattle. *Journal of Animal Science* 22, 486–494.
- Kocour, M., Mauger, S., Rodina, M., Gela, D., Linhart, O., Vandeputte, M. (2007). Heritability estimates for processing and quality traits in common carp (*Cyprinus carpio* L.) using a molecular pedigree. *Aquaculture* 270, 43–50. <https://doi.org/10.1016/j.aquaculture.2007.03.001>.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics*, 25(14), 1754-1760.
- Li, Y., Zhang, B., Lu, S., Tian, Y., Yang, Y., Chen, S. (2018). Genetic parameters estimates for growth performance traits at harvest in Japanese flounder (*Paralichthys olivaceus*). *Aquaculture*, 489, 56-61.
- Liu, Y. X., Wang, G. X., Wang, Y. F., Si, F., Sun, Z. H., Zhang, X. Y., Wang, J.D.; Liu, H. J. (2011). Estimation of genetic parameters for growth traits of Japanese flounder *Paralichthys olivaceus* using an animal model. *Fisheries Science*, 77(1), 87-93.
- Marcos, R., Povh, J. A., Fornari, D. C., Oliveira, C. A. L., Ribeiro, P., Lopera-Barrero, N. M., Filho, R. A. C. C., Abreu, J. S., Murari, P. J. F. (2016). Weight gain and morphometric growth of genetically improved tambaqui (*Colossoma macropomum*). *Semina: Ciências Agrárias* 37(4), 2521-2528. <https://doi.org/10.5433/1679-0359.2016v37n4Sup1p2521>.

- Mello, F. D., Oliveira, C. A. L., Streit, Jr. D., Resende, E. K., Oliveira, S. N., Fornari, D. C., Barreto, R. V., Povh, J. A., Ribeiro, R. P. (2016). Estimation of Genetic Parameters for Body Weight and Morphometric Traits to Tambaqui *Colossoma macropomum*. *J. Fisheries Sciences* 10(2), 096-100.
- Misztal, I., Legarra, A., & Aguilar, I. (2009). Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science*, 92(9), 4648-4655.
- Navarro, A., Zamorano, M.J., Hildebrandt, S., Ginés, R., Aguilera, C., Afonso, J.M., (2009). Estimates of heritabilities and genetic correlations for growth and carcass traits in gilthead seabream (*Sparus auratus* L.), under industrial conditions. *Aquaculture* 289, 225–230.
- Nguyen, N. H., Ponzoni, R. W., Abu-Bakar, K. R., Hamzah, A., Khaw, H. L., Yee, H. Y. (2010). Correlated response in fillet weight and yield to selection for increased harvest weight in genetically improved farmed tilapia (GIFT strain), *Oreochromis niloticus*. *Aquaculture*, 305(1-4), 1-5.
- Nunes, J. R. S., Pértille, F., Andrade, S. C. S., Perazza, C. A., Villela, P. M. S., Almeida-Val, V. M. F., Gao, Z.X., Coutinho, L.L., Hilsdorf, A. W. S. (2020). Genome-wide association study reveals genes associated with the absence of intermuscular bones in tambaqui (*Colossoma macropomum*). *Animal Genetics*, 51(6), 899-909.
- Pearson, K. (1897). Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organ. *Proceedings of the Royal Society of London* 60, 489–498.
- Perazza, C. A., Ferraz, J. B. S., Almeida-Val, V. M. F., Hilsdorf, A. W. S. (2019). Genetic parameters for loin eye area and other body traits of an important Neotropical aquaculture species, *Colossoma macropomum* (Cuvier, 1818). *Aquaculture Research*. <https://doi.org/10.1111/are.14245>.
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one*, 7(5), e37135.
- Powell, J., White, I., Guy, D., Brotherstone, S., (2008). Genetic parameters of production traits in Atlantic salmon (*Salmo salar*). *Aquaculture* 274 (2–4), 225–231.
- Prchal, M., Bugeon, J., Vandeputte, M., Kause, A., Vergnet, A., Zhao, J., Gela, D., Genestout, L., Bestin, A., Haffray, P., Kocour, M. (2018). Potential for genetic

- improvement of the main slaughter yields in common carp with in vivo morphological predictors. *Frontiers in genetics*, 9, 283.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, 81(3), 559-575.
- Rutten, M.J.M., Bovenhuis, H., Komen, H. (2005). Genetic parameters for fillet traits and body measurements in Nile tilapia (*Oreochromis niloticus* L.). *Aquaculture* 246, 101–113.
- Saillant, E., Dupont-Nivet, M., Sabourault, M., Haffray, P., Laureau, S., Vidal, M.-O.O., Chatain, B. (2009). Genetic variation for carcass quality traits in cultured sea bass (*Dicentrarchus labrax*). *Aquat. Living Resour.* 22, 105–112. <https://doi.org/10.1051/alr/2009010>.
- Sang, N., Thomassen, M., Klemetsdal, G., Gjøen, H. M. (2009). Prediction of fillet weight, fillet yield, and fillet fat for live river catfish (*Pangasianodon hypophthalmus*). *Aquaculture*, 288(3-4), 166-171.
- Sutherland, T.M. (1965). The correlation between feed efficiency and rate of gain, a ratio and its denominator. *Biometrics* 21, 739–749.
- Tian, Y., Xu, T., Liang, Y., Chen, S. (2011). Estimates of genetic and phenotypic parameters for weight and length in *Paralichthys olivaceus* (Temminck et Schlegel). *Acta Oceanologica Sinica*, 30(6), 58-64.
- Trọng, T. Q., Mulder, H. A., van Arendonk, J. A., Komen, H. (2013). Heritability and genotype by environment interaction estimates for harvest weight, growth rate, and shape of Nile tilapia (*Oreochromis niloticus*) grown in river cage and VAC in Vietnam. *Aquaculture*, 384, 119-127.
- Tsai, H. Y., Hamilton, A., Tinch, A. E., Guy, D. R., Gharbi, K., Stear, M. J., Matika, O., Bishop, S.C., Houston, R. D. (2015). Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC genomics*, 16(1), 1-9.
- Vandeputte, M., Bugeon, J., Bestin, A., Desgranges, A., Allamellou, J. M., Tyran, A. S., Allal, F., Dupont-Nivet, M., Haffray, P. (2019). First evidence of realized selection response on fillet yield in rainbow trout *Oncorhynchus mykiss*, using sib selection or based on correlated ultrasound measurements. *Frontiers in genetics*, 10, 1225.

- Vandeputte, M., Fraslin, C., Haffray, P., Bestin, A., Allal, F., Kocour, M., Prchal, M., Dupont-Nivet, M. (2020). How to genetically increase fillet yield in fish: Relevant genetic parameters and methods to predict genetic gain. *Aquaculture*, 519, 734877.
- Vandeputte, M., Garouste, R., Dupont-Nivet, M., Haffray, P., Vergnet, A., Chavanne, H., Laureau, S., Ron, T.B., Pagelson, G., Mazorra, C., Ricoux, R., Marques, P., Gameiro, M., Chatain, B. (2014). Multi-site evaluation of the rearing performances of 5 wild populations of European sea bass (*Dicentrarchus labrax*). *Aquaculture* 424–425 239–248. <http://dx.doi.org/10.1016/j.aquaculture.2014.01.005>.
- Vandeputte, M., Puleda, A., Tyran, A. S., Bestin, A., Coulombet, C., Bajek, A., Baldit, G., Vergnet, A., Allal, F., Bugeon, J., Haffray, P. (2017). Investigation of morphological predictors of fillet and carcass yield in European sea bass (*Dicentrarchus labrax*) for application in selective breeding. *Aquaculture*, 470, 40–49.
- Vaughan, L.K., Wiener, H.W., Aslibekyan, S., Allison, D.B., Havel, P.J., Stanhope, K.L., O'Brien, D.M., Hopkins, S.E., Lemas, D.J., Boyer, B.B. (2015). Linkage and association analysis of obesity traits reveals novel loci and interactions with dietary n-3 fatty acids in an Alaska Native (Yup'ik) population. *Metab Clin Exp* 64:689–697.
- Wang, H., Misztal, I., Aguilar, I., Legarra, A., Fernando, R. L., Vitezica, Z., Okimoto, R., Wing, T., Hawken, R., Muir, W.M. (2014). Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. *Frontiers in Genetics*, 5, 134.
- Wang, X., Liu, X., Deng, D., Mei, Y., Li, X. (2016). Genetic determinants of pig birth weight variability. *BMC Genet* 17:S15.
- Wang, X., Qu, M., Liu, Y., Schneider, R. F., Song, Y., Chen, Z., Zhang, H., Zhang, Y., Yu, H., Zhang, S., Li, D., Qin, G. *et al.*, (2021). Genomic basis of evolutionary adaptation in a warm-blooded fish. *The Innovation*, 100185.
- Wojnárovich, A., & Van Anrooy, R. (2019). Field guide to the culture of tambaqui (*Colossoma macropomum*, Cuvier, 1816). *FAO Fisheries and Aquaculture Technical Paper*, (624), I-121.
- Wu, L., Yang, Y., Li, B., Huang, W., Wang, X., Liu, X., Meng, Z., Xia, J. (2019). First genome-wide association analysis for growth traits in the largest coral reef-dwelling bony fishes, the Giant grouper (*Epinephelus lanceolatus*). *Marine Biotechnology*, 21(5), 707-717.

- Yang, X., Zhang, S., Liu, J., Gao, Q., Dong, S., Zhou, C. (2021). Deep learning for smart fish farming: applications, opportunities and challenges. *Reviews in Aquaculture*, 13(1), 66-90.
- Yoshida, G. M., Lhorente, J. P., Correa, K., Soto, J., Salas, D., Yáñez, J. M. (2019). Genome-wide association study and cost-efficient genomic predictions for growth and fillet yield in Nile tilapia (*Oreochromis niloticus*). *G3: Genes, Genomes, Genetics*, 9(8), 2597-2607.