UNIVERSIDADE ESTADUAL PAULISTA (UNESP)
CENTRO DE AQUICULTURA DA UNESP (CAUNESP)

Seleção genômica para resistência/resiliência a acantocefalose em tambaqui (*Colossoma macropomum*

John Fredy Gómez Agudelo

Jaboticabal
2024

UNIVERSIDADE ESTADUAL PAULISTA (UNESP)
CENTRO DE AQUICULTURA DA UNESP CAUNESP

Seleção genômica para resistência/resiliência a acantocefalose em tambaqui (*Colossoma macropomum)*

John Fredy Gómez Agudelo

Orientador: Prof. Dr. Diogo Teruo Hashimoto

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

Jaboticabal
2024

A282s

Agudelo, John Fredy Gomez
Seleção genômica para resistência/resiliência a acantocefalose em tambaqui (Colossoma macropomum / John Fredy Gomez Agudelo. -- Jaboticabal, 2024
116 p. : il., tabs., fotos

Tese (doutorado) - Universidade Estadual Paulista (UNESP), Centro de Aquicultura da Unesp, Jaboticabal
Orientadora: Diogo Hashimoto

1. Melhoramento Genético. 2. Acantocefalose. 3. Tambaqui. 4. SNP array. 5. Resistência/Resiliência. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Universidade Estadual Paulista (UNESP), Centro de Aquicultura da Unesp, Jaboticabal. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

UNIVERSIDADE ESTADUAL PAULISTA

Unidade Complementar - Jaboticabal

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: Seleção genômica para resistência/resiliência a acantocefalose em tambaqui *(Colossoma macropomum)*

AUTOR: JOHN FREDY GOMEZ AGUDELO

ORIENTADOR: DIOGO TERUO HASHIMOTO

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

Documento assinado digitalmente
gov.br DIOGO TERUO HASHIMOTO
Data: 03/05/2024 16:37:14-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. DIOGO TERUO HASHIMOTO (Participaçao Presencial)
Laboratorio de Genetica / Centro de Aquicultura da UNESP, CAUNESP, Jaboticabal-SP

Documento assinado digitalmente
gov.br GUSTAVO MORAES RAMOS VALLADAO
Data: 06/05/2024 10:46:23-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. GUSTAVO MORAES RAMOS VALLADÃO (Participaçao Virtual)
Departamento de Aquicultura / Universidade Nilton Lins, Manaus/AM

Documento assinado digitalmente
gov.br RAFAEL VILHENA REIS NETO
Data: 06/05/2024 11:27:44-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. RAFAEL VILHENA REIS NETO (Participaçao Virtual)
Departamento de Recursos Pesqueiros e Aquicultura / FCAVR/Unesp, Registro-SP

Profa. Dra. FABIANA PILARSKI (Participaçao Virtual)
Microbiologia e Parasitologia de Organismos Aquáticos / Centro de Aquicultura da Unesp, CAUNESP, Jaboticabal-SP

Documento assinado digitalmente
gov.br FABIANA PILARSKI
Data: 03/05/2024 17:18:52-0300
Verifique em https://validar.iti.gov.br

Dra. LUARA AFONSO DE FREITAS (Participaçao Virtual)
. / University of Wisconsin-Madison, EUA

Documento assinado digitalmente
gov.br LUARA AFONSO DE FREITAS JANUARIO
Data: 06/05/2024 10:39:42-0300
Verifique em https://validar.iti.gov.br

Jaboticabal, 30 de abril de 2024

Centro de Aquicultura - Unidade Complementar - Jaboticabal -
Via de Acesso Prof. Paulo Donato Castellani, s/nº, 14870000, Jaboticabal - São Paulo
http://www.caunesp.unesp.br/#!/pos-graduacao/

Apoio Financeiro

Agradeço ao CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico, pelo auxílio concebido (312250/2021-5)

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Agradeço a FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo. (2020/07959-5)

Agradeço a FAPEAM - Fundação de Amparo a Pesquisa do Estado do Amazonas (2020/07959-5)

Agradecimentos

Esta tese de Doutorado está dedicada à minha família, agradecimentos especiais para meu orientador e todos os companheiros do laboratório em genética em aquicultura e conservação (LaGeAc), e a todas as pessoas que me ajudaram de alguma forma na consecução deste objetivo.

Sumario

1. Introdução Geral 15
1.1 Produção aquícola e caracterização do tambaqui *Colossoma macropomum* 15
1.2 Descrição do parasita *Neoechinorhynchus buttnerae* e histórico 18
1.3 Diagnóstico e Tratamento 20
1.4 Melhoramento genético 22
1.5 Genômica aplicada ao melhoramento genético de peixes 24
1.6 Referênças 28
2. Objetivos 39
2.1 Objetivo geral 39
2.2 Objetivos específicos 39
CAPÍTULO I 40
Abstract 42
1. Introduction 43
2. Material and Methods 45
2.1. Ethics Statement 45
2.2. Experimental population and SNP genotyping 45
2.3. Selection of low-density subsets of SNPs for imputation accuracy 47
2.4. Accuracy of genotype imputation on low-density SNPs 48
2.5. Development of the low-density SNP array 50
3. Results 52
3.1 Selection of low-density subsets of SNPs for imputation accuracy 52
3.2 Accuracy of genotype imputation on low-density SNPs 53
3.4 Validation of the low-density SNP array 58
4. Discussion 63
5. Conclusions 65
6. References 67
CAPÍTULO II 71
Abstract 73

1. Introduction 74
2. Material and Methods 78
2.1 Ethics Statement 78
2.2 Experimental population 78
2.3 Experimental challenge 80
2.4 SNP genotyping 83
2.5 Statistical models 84
3. Results 86
4. Discussion 90
5. References 95
CAPÍTULO III 102
Abstract 103
1. Introduction 104
2. Material and methods 108
2.1. Ethics statement 108
2.2. Establishment of families, disease resistance challenge and genotyping 109
2.3. Genotype imputation 110
2.4. Genomic prediction analyzes 112
3. Results 115
3.1. Genotyping and Imputation 115
3.2. Genetic parameters and genomic prediction accuracy 116
4. Discussion 119
5. Conclusion 123
6. References 124
Considerações finais 131

Resumo geral

O tambaqui *Colossoma macropomum* é o principal peixe nativo produzido pela aquicultura nacional. Contudo, a cadeia produtiva desse peixe vem sendo afetada significativamente pela acantocefalose, doença causada pelo endoparasito *Neoechinorhynchus buttnerae*. A seleção de genótipos superiores para resistência e resiliência à acantocefalose é essencial para aprimorar a produção, promover a sustentabilidade ambiental e garantir a segurança alimentar. A resistência é a capacidade de um organismo de enfrentar e neutralizar os efeitos adversos de agentes externos, como doenças, toxinas ou condições ambientais desfavoráveis. Enquanto isso, a resiliência diz respeito à habilidade de um organismo se adaptar e se recuperar de desafios ambientais ou estresses, como doenças, mudanças climáticas ou condições adversas de cultivo. O objetivo deste projeto é subsidiar um programa de seleção genética para melhorar a cadeia produtiva do tambaqui, resistente e resiliente ao acantocéfalo *N. buttnerae*, por meio de estudos seleção genômica. Para isso, 70 famílias de tambaqui foram formadas, sendo 35 famílias pertencentes ao núcleo de melhoramento genético do Centro de Aquicultura da Unesp (CAUNESP), e 35 famílias pertencentes à uma piscicultura comercial no estado do Amazonas. Os fenótipos foram obtidos por meio de um desafio experimental onde um pool das famílias foi distribuído em três viveiros (réplicas) contendo zooplâncton hospedando a fase infectante do parasito. As características estimadas foram: Resistência (RT_{CT}) determinada pela contagem dos parasitas no intestino dos animais, densidade de parasitas (RT_{DP}), Resiliência final (RS_F), e resiliência total (RS_T). Este estudo foi dividido em três etapas. Inicialmente, conduzimos análises para estimar a acurácia da imputação de genótipos a partir de diferentes densidades de SNP (9K, 3K, 2K, 1K e 0,5K) em um *array* de média densidade (30K Affymetrix® Axiom® SerraSNP). Observamos que a densidade de 1K proporcionou o melhor equilíbrio entre o número mínimo de SNPs e a precisão da imputação de genótipos, alcançando uma precisão de 0,93 tanto no nível individual quanto no nível de SNP. Em seguida, utilizamos o subconjunto de 1K para desenvolver um *array* comercial por meio da genotipagem por sequenciamento direcionado Agriseq™ (Thermofisher). Na segunda etapa, procedemos à avaliação de parâmetros genéticos para as características RT_{CT}, RT_{DP}, RS_F e RS_T, nas quais observamos resultados de herdabilidade variando de 0,17 ± 0,04 a 0,20 ± 0,04. Esses números indicam a presença de níveis de herdabilidade baixos a moderados para a resposta à infecção por N. buttnerae em tambaqui. As correlações genéticas entre as características de resistência e resiliência mostraram-se consistentemente negativas (favoráveis) e estatisticamente significativas, variando de -0,512 ± 0,134 a -0,252 ± 0,166. Esses achados sugerem a viabilidade de selecionar animais simultaneamente para resistência e resiliência. A significativa variação genética encontrada para a resistência ao *N. buttnerae* e as correlações genéticas favoráveis com resiliência em tambaqui indicam que a seleção de genótipos superiores é uma estratégia promissora para reduzir

o impacto da acantocefalose na produção de tambaqui. Por último, desenvolvemos uma análise de predição genômica, onde notamos um aumento de até 70% na precisão da predição genômica para a resiliência total (RS_T) ao utilizar informações genômicas, apontando para uma vantagem esperada no ganho genético desses dados. Além disso, nossos resultados sugerem que a imputação é uma ferramenta eficaz para a predição genômica, podendo potencialmente aumentar em até 19% as características de resistência e até 48% a resiliência, comparado ao método tradicional baseado em pedigree. O *array* de 1K mostrou desempenho satisfatório na geração de estimativas confiáveis de valores genéticos para seleção genômica, especialmente em termos de precisão. Em conclusão, a seleção genômica, com imputação de genótipos, emerge como uma abordagem viável e econômica para melhorar a resistência à infecção por *N. buttnerae* em tambaqui, reduzindo os custos de genotipagem sem comprometer a precisão da predição genômica. O potencial impacto deste trabalho reside em viabilizar um pacote tecnológico para impulsionar a cadeia produtiva do tambaqui na indústria da aquicultura, garantindo segurança alimentar, alta qualidade genética e valor agregado.

General Abstract

The tambaqui *Colossoma macropomum* is the main native fish produced by national aquaculture. However, the productive chain of this fish has been significantly affected by acanthocephalosis, a disease caused by the endoparasite *Neoechinorhynchus buttnerae*. Selecting superior genotypes for resistance and resilience to acanthocephalosis is essential to improve production, promote environmental sustainability, and ensure food security. Resistance is the ability of an organism to confront and neutralize the adverse effects of external agents, such as diseases, toxins, or unfavorable environmental conditions. Meanwhile, resilience refers to an organism's ability to adapt and recover from environmental challenges or stresses, such as diseases, climate change, or adverse farming conditions. The aim of this project is to support a genetic selection program to improve the productive chain of tambaqui, resistant and resilient to the acanthocephalan *N. buttnerae*, through genomic selection studies. For this purpose, 70 tambaqui families were formed, with 35 families belonging to the genetic improvement nucleus of the Aquaculture Center of Unesp (CAUNESP), and 35 families belonging to a commercial fish farm in the state of Amazonas. Phenotypes were obtained through an experimental challenge where a pool of families was distributed in three ponds (replicas) containing zooplankton hosting the infective phase of the parasite. The estimated traits were: Resistance (RT_{CT}) determined by counting parasites in the animals' intestines, parasite density (RT_{DP}), Final resilience (RS_F), and total resilience (RS_T). This study was divided into three stages. Initially, we conducted analyses to estimate the accuracy of genotype imputation from different SNP densities (9K, 3K, 2K, 1K, and 0.5K) on a medium-density array (30K Affymetrix® Axiom® SerraSNP). We observed that the density of 1K provided the best balance between the minimum number of SNPs and the accuracy of genotype imputation, achieving an accuracy of 0.93 at both the individual and SNP levels. Next, we used the subset of 1K to develop a commercial *array* through targeted genotyping by Agriseq™ sequencing (Thermofisher). In the second stage, we proceeded to evaluate genetic parameters for the traits RT_{CT}, RT_{DP}, RS_F, and RS_T, where we observed heritability results ranging from 0.17 ± 0.04 to 0.20 ± 0.04. These numbers indicate the presence of low to moderate heritability levels for the response to *N. buttnerae* infection in tambaqui. Genetic correlations between resistance and resilience traits were consistently negative (favorable) and statistically significant, ranging from -0.512 ± 0.134 to -0.252 ± 0.166. These findings suggest the feasibility of selecting animals simultaneously for resistance and resilience. The significant genetic variation found for resistance to *N. buttnerae* and the favorable genetic correlations with resilience in tambaqui indicate that selecting superior genotypes is a promising strategy to reduce the impact of acanthocephalosis on tambaqui production. Finally, we developed a genomic prediction analysis, where we noticed an increase of up to 70% in the accuracy of genomic prediction for total resilience (RS_T) when using genomic information,

indicating an expected advantage in the genetic gain of these data. Additionally, our results suggest that imputation is an effective tool for genomic prediction, potentially increasing resistance traits by up to 19% and resilience by up to 48%, compared to the traditional pedigree-based method. The 1K array showed satisfactory performance in generating reliable estimates of genetic values for genomic selection, especially in terms of accuracy. In conclusion, genomic selection, with genotype imputation, emerges as a viable and cost-effective approach to improve resistance to *N. buttnerae* infection in tambaqui, reducing genotyping costs without compromising the accuracy of genomic prediction. The potential impact of this work lies in enabling a technological package to boost the productive chain of tambaqui in the aquaculture industry, ensuring food security, high genetic quality, and added value.

1. Introdução Geral

1.1 Produção aquícola e caracterização do tambaqui *Colossoma macropomum*

Um dos principais desafios da humanidade está relacionado em encontrar maneiras de suprir a alta demanda por alimentos exigida pela crescente população mundial. Projeções apontam que até 2050, a população mundial ultrapassará os 9 bilhões de habitantes (Troell et al., 2023). Nesse contexto, a produção de organismos aquáticos surge como uma alternativa saudável e sustentável de garantia de subsistência e geração de renda para uma população em constante expansão (Tavares-Dias & Mariano, 2015), mesmo diante de impactos econômicos e sociais como os decorrentes da pandemia de COVID-19. Sendo assim, a intensificação da produção animal através da aquicultura tem sido notável (FAO, 2022).

Com a estagnação da pesca e a crescente preferência por alimentos saudáveis, a aquicultura tem sido considerada uma das cadeias de produção de proteína animal mais promissoras atualmente (Souza & Viana, 2020). No entanto, a indústria ainda tem enfrentado desafios significativos, como as mudanças climáticas, o esgotamento dos recursos naturais e a necessidade de garantir a produção de alimentos de alta qualidade (Bacher, 2016).

Em 2020, a produção global de pescado proveniente da pesca e da aquicultura atingiu um total de 214 milhões de toneladas, sendo 178 milhões de toneladas de animais aquáticos e 36 milhões de toneladas de algas. Este leve aumento de 3% em relação ao recorde anterior

de 2018 (213 milhões de toneladas) foi principalmente limitado por uma queda de 4,4% na pesca de captura e pelos impactos da pandemia de COVID-19 em 2020. No entanto, esse declínio foi compensado pelo crescimento contínuo na aquicultura, embora a uma taxa anual mais lenta do que nos últimos dois anos. Este aumento é especialmente notável em regiões onde a aquicultura é impulsionada por investimentos em tecnologias avançadas, como é o caso da China, Chile e Noruega.

Em 2022, a produção aquícola no Brasil alcançou 617,3 mil toneladas, registrando um crescimento de 6,4% em relação ao ano anterior (IBGE, 2022). Esse aumento é atribuído principalmente ao crescimento da produção de tilápia *Oreochromis niloticus*, que totalizou 408,4 mil toneladas, representando 66,1% da produção total. Já as espécies nativas, como tambaqui, pacu, pirapitinga e seus híbridos, contribuem com 26,5% da produção total, totalizando 163,7 mil toneladas. O tambaqui, com 109,9 mil toneladas, ou seja, 17,8% do total de produção, é a segunda espécie mais produzida no Brasil.

O tambaqui *Colossoma macropomum* (Cuvier, 1818) pertence à família Serrasalmidae (Teleostei, Characiformes) e possui uma ampla distribuição na bacia do rio Amazonas e Orinoco (Jégu, 2003). É comum encontrá-lo em toda a região hidrográfica da Amazônia, que é composta por sete países, Brasil, Colômbia, Bolívia, Equador, Guiana, Peru e Venezuela (MMA, 2006). É um peixe reofílico neotropical, comumente conhecido como tambaqui no Brasil, gamitana no Peru (Fazzi-Gomes et al., 2017) e cachama negra na Colômbia (Santis et al., 2004). Sua produção tem sido realizada tanto em países da América

do Sul e Central, como em alguns países do Caribe e da Ásia, como a China, Indonésia, Malásia, Mianmar e Vietnã. (FAO, 2019).

O tambaqui apresenta diversas características que o tornam ideal para a aquicultura pois tem demonstrado rusticidade no manejo, tolerância a baixos níveis de oxigênio, pH e a altos níveis de amônia. Como espécie omnívora, sua dieta inclui sementes, frutos e insetos, além de possuir a habilidade de filtrar zooplâncton dos viveiros. O tambaqui também se adapta facilmente à ração comercial e demonstra um rápido crescimento em condições adequadas (Maciel-Honda et al., 2022). A qualidade de sua carne torna o tambaqui altamente procurado nos mercados locais e regionais, especialmente na região Norte, onde se concentra a maior parte da sua produção (IBGE, 2022).

Apesar das características favoráveis, a produção de tambaqui ainda enfrenta desafios principalmente relacionados a doenças causadas por parasitas e bactérias (Jerônimo et al., 2017; Lourenço et al., 2018; Pereira & Morey, 2018; Valladão et al., 2018). As doenças mais comuns, causadas por bactérias, incluem aquelas provocadas por *Flavobacterium columnare* e *Aeromonas hydrophila*. Além disso, diversos grupos de parasitas têm sido identificados nesta espécie, como Monogéneos, Tricodinídeos protozoários, mixosporídeos, crustáceos e helmintos. Em condições favoráveis, esses parasitas podem proliferar e resultar na perda de lotes inteiros (Fujimoto et al., 2015; Jerônimo et al., 2017; Valladão et al., 2018). Dentre esses parasitas, o *Neoechinorhynchus buttnerae* (Golvan, 1956) se destaca devido aos frequentes casos de infecções causados em peixes na região Norte do Brasil, onde foram

relatadas importantes perdas econômicas ((Chagas et al., 2019); Oliveira et al., 2015; Jerônimo et al., 2015, 2017; Pereira & Morey, 2018; Lourenço et al., 2017; Costa et al., 2020).

1.2 Descrição do parasita *Neoechinorhynchus buttnerae* e histórico

A doença da acantocefalose é causada pelo *Neoechinorhynchus buttnerae*, um endoparasita obrigatório cujo hospedeiro final pode ser o tambaqui e seus híbridos tambacu (*Piaractus mesopotamicus* × *Colossoma macropomum*) e tambatinga (*Colossoma macropomum* × *Piaractus Brachypomus*) (Castro et al., 2020a). O ciclo de vida do acantocéfalo se inicia com a liberação de ovos no ambiente aquático através das fezes do hospedeiro final, o tambaqui, seguido pela ingestão desses ovos pelo hospedeiro intermediário, um artrópode conhecido como *Cypridopsis vidua* (Sebastião et al., 2021). Dentro do hospedeiro intermediário, os ovos se desenvolvem na fase infecciosa chamada cistacanto (Lourenço et al., 2018). Quando o tambaqui se alimenta do zooplâncton presente nos tanques, ele ingere o hospedeiro intermediário. No intestino do peixe, as larvas se transformam em acantos e o parasita penetra na parede intestinal do hospedeiro usando seus "ganchos", até amadurecerem sexualmente para iniciar um novo ciclo (Nunez & Drago, 2017).

O parasita *N. buttnerae* é dotado de uma probóscide, uma estrutura retrátil na parte anterior, que contém uma série de espinhos ou "ganchos". É através dessa probóscide que o

parasita se fixa nas paredes do intestino do hospedeiro. A espécie de acantocéfalo pode ser determinada pelo número de espinhos presentes nessa estrutura. Além disso, o acantocéfalo não possui tubos digestivos e se alimenta dos nutrientes provenientes da alimentação do hospedeiro final diretamente através da mucosa intestinal (Valladão et al., 2020). Devido à penetração profunda nas paredes intestinais, a mucosa pode ser danificada, resultando em inflamação, granulomas, espessamento da parede intestinal e aumento da produção de muco (da Cunha et al., 2020). Nos casos de penetração profunda no tecido muscular é comum encontrar metaplasia, macrófagos, linfócitos e fibroblastos, isto como mecanismo de resposta à infecção parasitária (Lourenço et al., 2017; Nunes et al., 2023). Como consequência do processo infeccioso, a absorção de nutrientes é reduzida, e dependendo da intensidade parasitária, pode ocorrer obstrução intestinal, deformidades, bem como alteração do crescimento, perda de peso, caquexia, formação de estoques heterogêneos e até mortalidade (Jerônimo et al., 2017; Sebastião et al., 2021; Castro et al., 2020b).

Nos últimos anos, fortes infestações por *N. buttnerae* foram relatadas em pisciculturas nos estados do Amazonas, Roraima, Rondônia, Maranhão e Tocantins (Aguiar et al., 2018; Valladão et al., 2020; Maciel-Honda et al., 2023). O principal problema associado a essa parasitose é a grande perda econômica causada, sendo estimada em aproximadamente US$7.000 por hectare em tanques infectados por acantocéfalos (Silva-Gomes et al., 2017)

De acordo com estudos filogenéticos e taxonômicos, mais de 1.200 espécies, 157 gêneros e 26 famílias de acantocéfalos foram identificadas (Aguiar et al., 2018). Esses

parasitas têm a capacidade de infectar uma ampla variedade de vertebrados terrestres e aquáticos. Embora mais da metade das espécies de acantocéfalos sejam encontradas em peixes selvagens ou de cultivo, eles também podem parasitar répteis, aves, anfíbios, suínos, roedores e até mesmo seres humanos (Valladão et al., 2020).

A acantocefalose, devido às suas características, pode ser considerada uma doença silenciosa, muitas vezes passando despercebida pelo produtor, especialmente na ausência de um acompanhamento sanitário adequado (Valladão et al., 2020). Por outro lado, fatores que contribuem para a infestação por *N. buttnerae* em pisciculturas incluem a intensificação da produção, aliada a uma gestão inadequada e monitoramento insuficiente da qualidade da água (Biller & Chagas, 2022; de Alexandre Sebastião et al., 2022).

1.3 Diagnóstico e Tratamento

O diagnóstico convencional da acantocefalose é feito com eutanásia e posterior necropsia e dissecção do intestino para determinar a presença ou ausência do parasita *N. buttnerae*. Contudo, metodologias não invasivas como o uso de ferramentas moleculares como PCR e qPCR têm sido utilizadas para diagnóstico e controle da doença (da Cunha et al., 2020). No método abordado por da Cunha, et al. (2020) foi avaliada a presença de DNA do parasita *N. buttnerae* no sangue do hospedeiro por reação em cadeia da polimerase quantitativa (qPCR) e observaram que esta técnica apresentou 84% de eficiência, e 100% de especificidade, no entanto é necessário melhorar e aumentar a sensibilidade do método.

Os tratamentos para o controle da acantocefalose envolvem diversos produtos. O cloridrato de levamisol, por exemplo, ativa o sistema complementar e melhora a resposta ao estresse causado pela infestação parasitária (Feitoza da Silva et al., 2021). Outras opções incluem o uso de anti-helmínticos como o hidróxido de cálcio e o inseticida "terra de diatomáceas", que demonstraram eficácia de 48,2% e 36,4%, respectivamente. No entanto, a ivermectina e a piperazina apresentaram uma eficácia mais baixa, com 15,9% e 12,6%, respectivamente (Nunes et al., 2023). Ambos os estudos evidenciaram uma eficácia limitada nos tratamentos, destacando a necessidade de mais pesquisas para identificar medicamentos e produtos mais eficazes para o controle dessa doença (Nunes et al., 2023).

Além dos medicamentos convencionais, alguns produtos naturais comerciais à base de extratos vegetais têm sido empregados no controle de doenças parasitárias e bacterianas em espécies de peixes. Um estudo recente avaliou o potencial de dois desses produtos, BioFish® e Natumix®, no controle da acantocefalose, demonstrando resultados promissores em peixes infectados com *N. buttnerae* como monoterapia anti-acantocéfalo (Seixas et al., 2019). Embora o potencial de produtos à base de extratos de plantas naturais, óleos essenciais, ácidos orgânicos e óleo-resinas para o tratamento contra *N. buttnerae* tenha sido explorado em diversos estudos (Nunes et al., 2023), nenhum deles foi capaz de eliminar completamente o parasita.

Considerando a dificuldade de controlar a acantocefalose, especialmente devido à presença do zooplâncton infectado com ovos do parasita no ambiente aquático, medidas

preventivas tornam-se essenciais (Hilsdorf et al., 2022). Estas incluem o monitoramento regular dos tanques, observando sinais clínicos de infecção, bem como o controle da qualidade da água e a garantia de que a fonte de água não seja proveniente de outros cultivos. A amostragem de indivíduos para verificar a presença do parasita, juntamente com a secagem e desinfecção dos tanques antes da introdução de novos peixes, são práticas recomendadas. A utilização de sistemas indoor e bioflocos também tem se mostrado como uma alternativa eficaz (Valladão et al., 2020). Adicionalmente, é aconselhável solicitar registros sanitários antes de adquirir novos peixes (Hilsdorf et al., 2022).

1.4 Melhoramento genético

O melhoramento genético é essencial para o avanço de uma aquicultura mais eficiente e sustentável, além de contribuir para a redução dos custos de produção (Robledo et al., 2017; Lira et al., 2020). Com a crescente demanda por estratégias eficazes no controle de doenças, o melhoramento genético surge como uma alternativa viável, visando conferir maior resistência a patógenos, promover a saúde animal, aumentar a produtividade e aprimorar o bem-estar dos animais. Esse enfoque busca atingir eficiência e sustentabilidade na produção aquícola (Barría et al., 2018). Daí a importância dos programas de seleção em larga escala para a obtenção de organismos com um sistema imunológico mais eficiente, tornando-se uma ferramenta na aquicultura com potencial de substituir o uso indiscriminado de antibióticos (Yáñez et al., 2014; Gonen et al., 2015).

O melhoramento genético tradicional, baseado em pedigree, utiliza informações genealógicas sobre os indivíduos e dados fenotípicos para estimar seu valor genético (EBV) e orientar decisões de seleção (Gjedrem, 2010). Para características como resistência a doenças e rendimento de carcaça, difíceis de medir diretamente em animais vivos, a predição do valor genético (EBV) geralmente se fundamenta em dados de pedigree e fenotípicos de indivíduos relacionados, como irmãos completos e meios-irmãos. Essa abordagem, empregando o método tradicional de Melhor Predição Linear Não Viesada (BLUP), tem sido amplamente utilizada para tal propósito (Pablo Dufflocqa, 2019). No entanto, a seleção baseada em pedigree apresenta uma limitação importante, já que este método apenas aproveita a variação genética entre famílias. (Manso 2024). Com a crescente disponibilidade de tecnologias de genotipagem, o emprego de informações genômicas para avaliação genética está se tornando a metodologia padrão em muitas espécies de relevância comercial. (Yáñez et al., 2014)

Vários estudos foram realizados para o tambaqui, com o objetivo de: melhorar a resistência a doenças (Ariede et al., 2020;2022; Lira et al., 2020; 2023), performance de crescimento e rendimento de filé (Ariede et al., 2023), bem como estudos de variabilidade e estrutura genética populacional (Agudelo et al., 2022). É importante considerar os estudos genéticos como uma alternativa para controlar surtos nas produções, da mesma forma como já foi implementado em outras espécies, como a tilápia, os salmões e as carpas (Barría et al., 2018; Jeney et al., 2011; Wonmongkol et al., 2018). A busca por soluções para os desafios ocasionados por doenças no sistema de cultivo é de grande interesse para os produtores e

pesquisadores, uma vez que as perdas econômicas podem chegar a bilhões de dólares a cada ano (Ariede et al., 2022).

1.5 Genômica aplicada ao melhoramento genético de peixes

Nos últimos anos, houve um aumento significativo na disponibilidade e aplicação das informações genômicas em espécies aquícolas importantes. Isso inclui a montagem de genomas de referência em várias espécies, o desenvolvimento de SNP arrays, e as tecnologias de próxima geração (NGS) que incluem técnicas de genotipagem direcionada de alto rendimento, Embora essas ferramentas possam variar em termos de custos, tempo de processamento de amostras, análises de bioinformática e qualidade dos dados, sua aplicabilidade tem sido de grande importância para o avanço da aquicultura. ((Nascimento-Schulze et al., 2023; Yáñez et al., 2023; Zhang et al., 2023)

As ferramentas genômicas têm o potencial de revolucionar a aquicultura por meio de programas de seleção melhorando a acurácia e predição de valores genômicos (Prchal et al., 2023). Múltiplos estudos que utilizam marcadores moleculares associados a características produtivas foram extensivamente documentados em diversas espécies aquícolas (Palaiokostas e Houston, 2018).

O crescente uso de informações genômicas para impulsionar o avanço genético de características de difícil mensuração na aquicultura tem sido de suma importância nas últimas décadas (Yáñez et al., 2023). Uma estratégia para esse avanço é a seleção assistida por

marcadores (MAS), que se apoia na identificação de marcadores genéticos associados a loci de características quantitativas QTLs (*quantitative trait loci*). Essa abordagem tem sido adotada em diversas espécies de aquicultura como uma ferramenta para aprimorar a seleção (Barría et al., 2018; Houston et al., 2020). Além disso, outra abordagem amplamente utilizada para identificar marcadores moleculares associados a características produtivas são os estudos de associação genômica ampla (GWAS). Os GWAS auxiliam no entendimento da arquitetura genômica e os processos biológicos envolvidos na expressão de características de interesse produtivo na aquicultura (Ariede et al., 2022). Estudos utilizando GWAS foram reportados para identificar possíveis QTLs e genes associados a caraterísticas produtivas no tambaqui, como a resistência ao ectoparasito *Ichthyophthirius multifiliis* (Lira et al., 2023) e a bactéria *Aeromonas. hydrophila* (Ariede et al., 2022). Além disso, também foi descrito genes candidatos relacionados a ausência de espinhos intramusculares em tambaqui (Nunes et al., 2020).

A seleção genômica tem sido aplicada em espécies da aquicultura para acelerar o ganho genético em programas de melhoramento genético (Zenger et al., 2019). Esta ferramenta baseada na predição dos valores gnômicos (GEBV) dos marcadores presentes no genoma permite estimar características que são difíceis de medir por sua natureza poligênica como resistência a doenças, ou caraterísticas limitadas pelo sexo ou de interação fenótipo-ambiente. (Yang et al., 2024; Yáñez et al., 2023)

A predição genômica (GP) tem encontrado aplicação em diversos campos, como melhoramento de plantas, pecuária, aquicultura e genética humana. A GP combina dados fenotípicos individuais com informações genômicas de mapas de marcadores densos em todo o genoma (Kjetså et al., 2020). De maneira geral, a predição genômica tem se mostrado mais eficiente do que o método tradicional baseado em pedigree (EBV). Ao utilizar informações genômicas, essa técnica possibilita uma seleção mais precisa em programas de melhoramento genético (Ajasa et al., 2024). Estudos sobre resistência a doenças têm consistentemente demonstrado que a predição genômica produz valores de precisão mais elevados em comparação com métodos baseados apenas em pedigree (Aslam et al., 2020; Vallejo et al., 2024; Manso et al., 2024).

A metodologia de imputação de genótipos é uma alternativa para estudos genômicos, pois esta metodologia é capaz de inferir ou predizer genótipos ausentes ou que não foram encontrados durante o processo de genotipagem em indivíduos não genotipados ou genotipados em baixa densidade, tomando como referência uma população da mesma espécie genotipada em alta densidade (Kriaridou et al., 2023). A imputação pode ser usada para diminuir os custos de genotipagem que como foi relatado por (Yoshida et al., 2018a). A genotipagem de um painel de baixa densidade tem custo relativamente menor do que a genotipagem de um painel de alta densidade, por exemplo, um painel de 500 SNPs pode ter um custo de 5 a 15 US$ por amostra e um painel de alta densidade pode estar em torno de 45 a 75 US$ por amostra dependendo do número de amostras (Yoshida et al., 2018).

Diferentes estudos têm desenvolvido painéis de alta densidade em espécies aquícolas que podem ser utilizados futuramente em estudos de imputação. Painéis de alta densidade como de 665k para truta arco-íris (Bernard et al., 2022), salmão do Atlântico 50k (Yoshida et al., 2018b), *Crassostrea virginica* 200k (Xuereb et al., 2023), *Crassostrea ariakensis* 30 K (Zhang et al., 2023) são alguns exemplos. Para espécies nativas neotropicais, foi desenvolvido um painel multi-espécie de média densidade para pacu (*Piaractus mesopotamicus*) e tambaqui (*Colossoma macropomum*) com 30K marcadores (Mastrochirico-Filho et al., 2021). Do mesmo modo, tem sido desenvolvido painéis de baixa densidade para imputação de genótipos em outras espécies como: *Oncorhynchus mykiss*; *Salmo salar*; *Litopenaeus vannamei*; *Scophthalmus maximus*; *Cyprinus carpio; e Crassostrea gigas* (Fraslin et al., 2023; Tsairidou et al., 2020; Silva et al., 2022; Kriaridou et al., 2023).

A estratégia de combinação de painéis de alta, média e baixa densidade para imputação de genótipos é uma alternativa para a implementação da seleção genômica de características como resistência a doenças, ganho de peso, variabilidade e estrutura genética, bem como análise de parentesco. Em suma, o avanço das ferramentas genômicas na aquicultura promete impulsionar a seleção de características desejáveis, melhorar a produtividade e a resistência das espécies aquáticas, contribuindo assim para a sustentabilidade e eficiência do setor aquícola.

1.6 Referênças

Agudelo, J. F. G., Mastrochirico-Filho, V. A., Borges, C. H. de S., Ariede, R. B., Lira, L. V. G., Neto, R. R. de O., de Freitas, M. V., Sucerquia, G. A. L., Vera, M., Berrocal, M. H. M., & Hashimoto, D. T. (2022). Genomic selection signatures in farmed *Colossoma macropomum* from tropical and subtropical regions in South America. Evolutionary Applications, 15(4), 679–693. https://doi.org/10.1111/eva.13351

Aguiar, L. S., de Oliveira, M. I. B., de Matos, L. V., Gomes, A. L. S., da Costa, J. I., & da Silva, G. S. (2018). Distribution of the acanthocephalan *Neoechinorhynchus buttnerae* and semiquantitative analysis of histopathological damage in the intestine of tambaqui (*Colossoma macropomum*). Parasitology Research, 117(6), 1689–1698. https://doi.org/10.1007/s00436-018-5840-8

Ajasa, A. A., Boison, S. A., Gjøen, H. M., & Lillehammer, M. (2024). Genome-assisted prediction of amoebic gill disease resistance in different populations of Atlantic salmon during field outbreak. Aquaculture, 578. https://doi.org/10.1016/j.aquaculture.2023.740078

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

Ariede, R. B., Freitas, M. V., Agudelo, J. F., Borges, C. H., et al. (2020). Genetic (co) variation between resistance to *Aeromonas hydrophila* and growth in tambaqui (*Colossoma macropomum*). Aquaculture, 523, 735225.

Ariede, R. B., Freitas, M. V., Lira, L. V. G., Agudelo, J. F. G., Borges, C. H. S., Mastrochirico-Filho, V. A., Cáceres, P., Yáñez, J. M., & Hashimoto, D. T. (2022). Linkage map for chromosome-level genome anchoring and genome-wide association

study for resistance to *Aeromonas hydrophila* in *Colossoma macropomum*. Aquaculture, 560. https://doi.org/10.1016/j.aquaculture.2022.738462

Ariede, R. B., Lemos, C. G., Batista, F. M., Oliveira, R. R., Agudelo, J. F. G., Borges, C. H. S., Iope, R. L., Almeida, F. L. O., Brega, J. R. F., & Hashimoto, D. T. (2023). Computer vision system using deep learning to predict rib and loin yield in the fish *Colossoma macropomum*. Animal Genetics. https://doi.org/10.1111/age.13302

Aslam, M. L., Carraro, R., Sonesson, A. K., Meuwissen, T., et al. (2020). Genetic variation, GWAS and accuracy of prediction for host resistance to *Sparicotyle chrysophrii* in farmed gilthead sea bream (*Sparus aurata*). Frontiers in Genetics, 11, 594770.

Baca, L. C. De. El cultivo de la gamitana en latinoamérica. [s.l: s.n.].

Bacher, K. (2016). Perceptions and misconceptions of aquaculture: a global overview. https://doi.org/10.13140/RG.2.1.1399.3840

Barría, A., Christensen, K. A., Yoshida, G. M., Correa, K., 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.

Bernard, M., Dehaullon, A., Gao, G., Paul, K., Lagarde, H., Charles, M., Prchal, M., Danon, J., Jaffrelo, L., Poncet, C., Patrice, P., Haffray, P., Quillet, E., Dupont-Nivet, M., Palti, Y., Lallias, D., & Phocas, F. (2022). Development of a High-Density 665 K SNP Array for *Rainbow Trout* Genome-Wide Genotyping. Frontiers in Genetics, 13. https://doi.org/10.3389/fgene.2022.941340

Biller, J. D., & Chagas, E. C. (2022). Mechanisms of resistance and tolerance against parasites in fish: the impairments caused by *Neoechinorhynchus buttnerae* in *Colossoma macropomum*. Anais Da Academia Brasileira de Ciencias, 94(4). https://doi.org/10.1590/0001-3765202220210258

Castro, L. de A., Jerônimo, G. T., da Silva, R. M., Santos, M. J., Ramos, C. A., & Porto, S. M. de A. (2020a). Occurrence, pathogenicity, and control of acanthocephalosis caused by *Neoechinorhynchus buttnerae*: A review. In Revista Brasileira de

Parasitologia Veterinaria (Vol. 29, Issue 3, pp. 1–7). Brazilain Coll Veterinary Parasitology. https://doi.org/10.1590/s1984-29612020070

Chagas, E. C., Aquino-Pereira, S. L., Benavides, M. V., Brandão, F. R., Monteiro, P. C., & Maciel, P. O. (2019). *Neoechinorhynchus buttnerae* parasitic infection in tambaqui (*Colossoma macropomum*) on fish farms in the state of amazonas. Boletim Do Instituto de Pesca, 45(2). https://doi.org/10.20950/1678-2305.2019.45.2.499

Costa, C. M. de S., da Cruz, M. G., Lima, T. B. C., Ferreira, L. C., Ventura, A. S., Brandão, F. R., Chagas, E. C., Chaves, F. C. M., Martins, M. L., & Jerônimo, G. T. (2020). Efficacy of the essential oils of Mentha piperita, Lippia alba and Zingiber officinale to control the acanthocephalan *Neoechinorhynchus buttnerae* in *Colossoma macropomum*. Aquaculture Reports, 18. https://doi.org/10.1016/j.aqrep.2020.100414

da Cunha, F. P., Cardoso, A. C. de S., Merlano, J. A. R., Nornberg, B. F. da S., Marins, L. F., Jerônimo, G. T., & Almeida, D. V. (2020). Non-lethal molecular diagnostic for acanthocephalosis in *Colossoma macropomum*. Aquaculture, 519. https://doi.org/10.1016/j.aquaculture.2019.734860

de Alexandre Sebastião, F., Souza Rocha, M. J., Rodrigues Brandão, F., Braga de Oliveira, M. I., de Melo Souza, D. C., Carlos do Nascimento Barbosa, B., Castro Monteiro, P., Majolo, C., Crescêncio, R., Tavares-Dias, M., & Campos Chagas, E. (2022). Evaluation of the clinical safety and efficacy of fenbendazole and levamisole in the control of *Neoechinorhynchus buttnerae* in *Colossoma macropomum:* Acanthocephalosis treatments. Aquaculture International, 30(3), 1341–1351. https://doi.org/10.1007/s10499-022-00854-2

Dufflocq, P., Pérez-Enciso, M., Lhorente, J. P., & Yáñez, J. M. (2019). Accuracy of genomic predictions using different imputation error rates in aquaculture breeding programs: A simulation study. Aquaculture, 503, 225-230.

El estado mundial de la pesca y la acuicultura 2019. (2019). In El estado mundial de la pesca y la acuicultura 2019. FAO

El estado mundial de la pesca y la acuicultura 2022. (2022). In El estado mundial de la pesca y la acuicultura 2022. FAO. https://doi.org/10.4060/cc0461es

Fazzi-Gomes, P., Guerreiro, S., Palheta, G. D. A., de Melo, N. F. A. C., Santos, S., & Hamoy, I. (2017). High genetic diversity and connectivity in *Colossoma macropomum* in the Amazon basin revealed by microsatellite markers. Genetics and Molecular Biology, 40(1), 142–146. https://doi.org/10.1590/1678-4685-GMB-2015-0222

Feitoza da Silva, A., Pahor Filho, E., Gonçalves, L. U., Silva, T. F. A., & Urbinati, E. C. (2021). Levamisole modulates cortisol and innate immunity responses in tambaqui (*Colossoma macropomum*) parasitized with *Neoechinorhynchus buttnerae*. Aquaculture Research, 52(11), 5920–5924. https://doi.org/10.1111/are.15442

Fraslin, C., Robledo, D., Kause, A., & Houston, R. D. (2023). Potential of low-density genotype imputation for cost-efficient genomic selection for resistance to *Flavobacterium columnare* in rainbow trout (*Oncorhynchus mykiss*). Genetics Selection Evolution, 55(1), 59. https://doi.org/10.1186/s12711-023-00832-z

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

Gorjanc, G., Dumasy, J. F., Gonen, S., Gaynor, R. C., *et al.* (2017). Potential of low-coverage genotyping-by-sequencing and imputation for cost-effective genomic selection in biparental segregating populations. Crop Science, 57(3), 1404-1420.

Hilsdorf, A. W. S., Hallerman, E., Valladão, G. M. R., Zaminhan-Hassemer, M., Hashimoto, D. T., Dairiki, J. K., Takahashi, L. S., Albergaria, F. C., Gomes, M. E. de S., Venturieri, R. L. L., Moreira, R. G., & Cyrino, J. E. P. (2022). The farming and husbandry of *Colossoma macropomum*: From Amazonian waters to sustainable production. In Reviews in Aquaculture (Vol. 14, Issue 2, pp. 993–1027). John Wiley and Sons Inc. https://doi.org/10.1111/raq.12638

IBGE-Instituto Brasileiro de Geografia e Estatística. 2022. Produção da Pecuária Municipal Prod. Pec. munic., Rio de Janeiro, v. 50, p.1-12, 2022

Jégu, M (2003) Subfamily Serrasalminae (Pacus and piranhas). In: Check list of the freshwater fishes of South and Central America (eds Reis RE, Kullander SO, Ferraris CJ). Edipucrs, Porto Alegre

Jeney, G., Ardó, L., Rónyai, A., Bercsényi, M., & Jeney, Z. (2011). Resistance of genetically different common carp, *Cyprinus carpio* L., families against experimental bacterial challenge with *Aeromonas hydrophila*. Journal of Fish Diseases, 34(1), 65–70. https://doi.org/10.1111/j.1365-2761.2010.01211.x

Jerônimo, G. T., Pádua, S. B. de, Belo, M. A. de A., Chagas, E. C., Taboga, S. R., Maciel, P. O., & Martins, M. L. (2017). *Neoechinorhynchus buttnerae* (Acanthocephala) infection in farmed *Colossoma macropomum:* A pathological approach. Aquaculture, 469, 124–127. https://doi.org/10.1016/j.aquaculture.2016.11.027

Kjetså, M. H., Ødegård, J., & Meuwissen, T. H. E. (2020). Accuracy of genomic prediction of host resistance to salmon lice in Atlantic salmon (*Salmo salar)* using imputed high-density genotypes. Aquaculture, 526, 735415.

Kriaridou, C., Tsairidou, S., Fraslin, C., Gorjanc, G., et al. (2023). Evaluation of low-density SNP panels and imputation for cost-effective genomic selection in four aquaculture species. Frontiers in genetics, 14, 1194266.

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

Lira, L. V. G., Mastrochirico-Filho, V. A., Mendes, N. J., Ariede, R. B., Yáñez, J. M., & Hashimoto, D. T. (2023). Genome-wide association study of host resistance to the ectoparasite *Ichthyophthirius multifiliis* in the Amazon fish *Colossoma macropomum*. Molecular Biology Reports, 50(1), 599–607. https://doi.org/10.1007/s11033-022-08062-0

Lourenço, F. D. S., Morey, G. A. M., & Malta, J. C. D. O. (2018). The development of *Neoechinorhynchus buttnerae* (Eoacanthocephala: *Neoechinorhynchidae*) in its intermediate host *Cypridopsis vidua* in Brazil. Acta Parasitologica, 63(2), 354–359. https://doi.org/10.1515/ap-2018-0040

Lourenço, F. de S., Morey, G. A. M., Pereira, J. N., & Malta, J. C. de O. (2017). Ocorrência de *Neoechinorhynchu*s (*Neoechinorhynchus*) *buttnerae* Golvan, 1956 (ACANTOCEPHALA: NEOCHINORHYNCHIDAE) em *Colossoma macropomum* (Cuvier, 1818) (CHARACIFORMES: SERRASALMIDAE) provenientes de uma piscicultura da amazônia brasileira. Folia Amazónica, 26(1), 1–8. https://doi.org/10.24841/fa.v26i1.414

Maciel-Honda, P. O., Neto, E. M. de S., Costa-Fernandes, T. de O., de Jesus, F. H. R., Chagas, E. C., & Tavares-Dias, M. (2023). First record of *Neoechinorhynchus buttnerae* and *Piscinoodinium pillulare* infection in *Colossoma macropomum* in the state of Tocantins, Brazil. Revista Brasileira de Parasitologia Veterinaria, 32(1). https://doi.org/10.1590/S1984-29612023001

Manso, S. C. S., Garcia, B. F., Mastrochirico-Filho, V. A., Porto-Foresti, F., Yáñez, J. M., & Hashimoto, D. T. (2024). Genomic predictions for resistance to *Aeromonas hydrophila* in pacu (*Piaractus mesopotamicus*). Aquaculture, 582. https://doi.org/10.1016/j.aquaculture.2024.740553

Mastrochirico-Filho, V. A., Ariede, R. B., Freitas, M. V., Borges, C. H., et al. (2021). Development of a multi-species SNP array for serrasalmid fish *Colossoma macropomum* and *Piaractus mesopotamicus*. Scientific Reports, 11(1), 19289.

MMA. Caderno da Região Hidrográfica Amazônica. [s.l:s.n.]. Disponível em: http://www.mma.gov.br/estruturas/161/_publicacao/161_publicacao03032011024106.pdf>

Nascimento-Schulze, J. C., Bean, T. P., Peñaloza, C., Paris, J. R., Whiting, J. R., Simon, A., Fraser, B. A., Houston, R. D., Bierne, N., & Ellis, R. P. (2023). SNP discovery and genetic structure in blue mussel species using low coverage sequencing and a medium

density 60 K SNP-array. Evolutionary Applications. https://doi.org/10.1111/eva.13552

Nunes, B. R. C., de Andrade Lopes, Y. V., de Souza, R. H. B., Pazdiora, R. D., Ushizima, T. T., Medeiros, S. P., Corrêa, L. T., Cardoso, I. I., Andrade, M. V. V., Maciel-Honda, P. O., Chagas, E. C., & de Alexandre Sebastião, F. (2023). Treatments for the control of *Neoechinorhynchus buttnerae* (Acanthocephala) in tambaqui *Colossoma macropomum.* Aquaculture International, 31(3), 1821–1835. https://doi.org/10.1007/s10499-023-01057-z

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

Oliveira ACB, Miranda E, Correa R (2012) Exigências nutricionais e alimentação do tambaqui. In: Nutriaqua: nutrição e alimentação de espécies de interesse para a aquicultura brasileira (eds Fracalossi DM, Cyrino JEP). Sociedade Brasileira de Aquiculura e Biologia Aquática, Florianópolis.

Palaiokostas, C., Cariou, S., Bestin, A., Bruant, J. S., et al. (2018). Genome-wide association and genomic prediction of resistance to viral nervous necrosis in European sea bass (Dicentrarchus labrax) using RAD sequencing. Genetics Selection Evolution, 50(1), 1-11.

Pereira, J. N., & Morey, G. A. M. (2018). Primeiro registro de *Neoechinorhynchus buttnerae* (Eoacantocephala, Neochinorhynchidae) em *Colossoma macropomum* (Characidae) em uma fazenda de peixes em Roraima, Brasil. Acta Amazonica, 48(1), 42–45. https://doi.org/10.1590/1809-4392201702411

Prchal, M., Palaiokostas, C., Gela, D., Piačková, V., Reschová, S., & Kocour, M. (2023). Genetic parameters and genomic prediction of resistance to koi herpesvirus disease

using a low-density SNP panel on two Amur mirror carp populations. Aquaculture Reports, 30. https://doi.org/10.1016/j.aqrep.2023.101582

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.

Santis, H. P., Restrepo, L. F., & Angel, M. O. (n.d.). Comparación morfométrica entre machos y hembras de Cachama Negra (Colossoma macropomum, Cuvier 1818) mantenidos en estanque.

Sebastião, F. de A., Braga de Oliveira, M. I., Rocha, M. J. S., Souza, D. C. de M., Ribeiro, P., Majolo, C., Crescêncio, R., & Chagas, E. C. (2021). Effect of a food additive in the control of the acanthocephalan *Neoechinorhynchus buttnerae* in *Colossoma macropomum*. Aquaculture Research, 52(2), 635–642. https://doi.org/10.1111/are.14920

Seixas, A. T., Gallani, S. U., Fernandes, I. M., Fernandes, I. E. C., Jerônimo, G. T., Pilarski, F., & Valladão, G. M. R. (2019). Effect of nutraceuticals on acanthocephalan *Neoechinorhynchus buttnerae* and its toxicity to the host tambaqui *Colossoma macropomum.* Journal of Helminthology, 94, e102. https://doi.org/10.1017/S0022149X19000956

Silva, N. M. L., Ianella, P., Yamagishi, M. E. B., Rocha, J. L., Teixeira, A. K., Farias, F. G., Guerrelhas, A. C., & Caetano, A. R. (2022). Development and validation of a low-density SNP panel for paternity and kinship analysis and evaluation of genetic variability and structure of commercial Pacific white shrimp (*Litopenaeus vannamei*) populations from Brazil. Aquaculture, 560. https://doi.org/10.1016/j.aquaculture.2022.738540

Silva-Gomes, A. L., Gomes Coelho-Filho, J., Viana-Silva, W., Braga-Oliveira, M. I., et al. (2017). The impact of *Neoechinorhynchus buttnerae* (Golvan, 1956) (Eoacanthocephala: Neochinorhynchidae) outbreaks on productive and economic

performance of the tambaqui *Colossoma macropomum* (Cuvier, 1818), reared in ponds. Latin american journal of aquatic research, 45(2), 496-500.

Souza, A. C. F., & Viana, D. C. (2020). Status atual da aquicultura no mundo: primeiros impactos da COVID-19. Research, Society and Development, 9(8), e462985798. https://doi.org/10.33448/rsd-v9i8.5798

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, 913-918.

Troell, M., Costa-Pierce, B., Stead, S., Cottrell, R. S., Brugere, C., Farmery, A. K., Little, D. C., Strand, Å., Pullin, R., Soto, D., Beveridge, M., Salie, K., Dresdner, J., Moraes-Valenti, P., Blanchard, J., James, P., Yossa, R., Allison, E., Devaney, C., & Barg, U. (2023). Perspectives on aquaculture's contribution to the Sustainable Development Goals for improved human and planetary health. Journal of the World Aquaculture Society, 54(2), 251–342. https://doi.org/10.1111/jwas.12946

Tsairidou, S., Hamilton, A., Robledo, D., Bron, J. E., & Houston, R. D. (2020). Optimizing low-cost genotyping and imputation strategies for genomic selection in atlantic salmon. G3: Genes, Genomes, Genetics, 10(2), 581–590. https://doi.org/10.1534/g3.119.400800

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

Valladão, G. M. R., Gallani, S. U., Jerônimo, G. T., & Seixas, A. T. de. (2020). Challenges in the control of acanthocephalosis in aquaculture: special emphasis on *Neoechinorhynchus buttnerae*. In Reviews in Aquaculture (Vol. 12, Issue 3, pp. 1360–1372). Wiley-Blackwell. https://doi.org/10.1111/raq.12386

Vallejo, R. L., Pietrak, M. R., Milligan, M. M., Gao, G., Tsuruta, S., Fragomeni, B. O., Long, R. L., Peterson, B. C., & Palti, Y. (2024). Genetic architecture and accuracy of predicted genomic breeding values for sea lice resistance in the St John River

aquaculture strain of North American Atlantic salmon. Aquaculture, 740819. https://doi.org/10.1016/j.aquaculture.2024.740819

Xuereb, A., Nahuelpi, R. M., Normandeau, E., Babin, C., Laporte, M., Mallet, A., Yáñez, J. M., Mallet, M., & Bernatchez, L. (2023). Design and validation of a high-density SNP array for the Eastern oyster (*Crassostrea virginica*). G3 Genes|Genomes|Genetics. https://doi.org/10.1093/g3journal/jkad071

Yáñez, J. M., Barría, A., López, M. E., Moen, T., Garcia, B. F., Yoshida, G. M., & Xu, P. (2023). Genome-wide association and genomic selection in aquaculture. In Reviews in Aquaculture (Vol. 15, Issue 2, pp. 645–675). John Wiley and Sons Inc. https://doi.org/10.1111/raq.12750

Yáñez, J. M., Houston, R. D., & Newman, S. (2014). Genetics and genomics of disease resistance in salmonid species. Frontiers in genetics, 5, 415.

Yang, B., Zhi, C., Li, P., Xu, C., Li, Q., & Liu, S. (2024). Genomic selection accelerates genetic improvement of resistance to Vibriosis in the Pacific oyster, *Crassostrea gigas*. Aquaculture, 584. https://doi.org/10.1016/j.aquaculture.2024.740679

Yoshida, G. M., Bangera, R., Carvalheiro, R., Correa, K., et al. (2018b). Genomic prediction accuracy for resistance against *Piscirickettsia salmonis* in farmed rainbow trout. G3: Genes, Genomes, Genetics, 8(2), 719-726.

Yoshida, G. M., Carvalheiro, R., Lhorente, J. P., Correa, K., Figueroa, R., Houston, R. D., & Yáñez, J. M. (2018a). Accuracy of genotype imputation and genomic predictions in a two-generation farmed Atlantic salmon population using high-density and low-density SNP panels. Aquaculture, 491, 147–154. https://doi.org/10.1016/j.aquaculture.2018.03.004

Zenger, K. R., Khatkar, M. S., Jones, D. B., Khalilisamani, N., et al. (2019). Genomic selection in aquaculture: application, limitations and opportunities with special reference to marine shrimp and pearl oysters. Frontiers in genetics, 9, 411091.

Zhang, K., Li, A., Qi, H., Yang, Q., Du, M., Wang, X., Zhang, Z., Wang, C., Wang, W., Zhang, G., & Li, L. (2023). The development of a 30 K SNP genotyping tool targeting

genomic regions of temperature and salinity adaptation in estuarine oyster. Aquaculture, 566. https://doi.org/10.1016/j.aquaculture.2022.739168

2. Objetivos

2.1 Objetivo geral

- Subsidiar um programa de seleção genética visando aprimorar a cadeia produtiva do tambaqui (*Colossoma macropomum*), tornando-o mais resistente e resiliente ao acantocéfalo *Neoechinorhynchus buttnerae*, por meio de estudos de associação genômica ampla e seleção genômica.

2.2 Objetivos específicos

- Desenvolver e validar um *array* SNP 1K para tambaqui, que pode ser aplicado em conjunto com técnicas de imputação, para uso em programas de melhoramento.

- Realizar um desafio experimental contra o endoparasita *N. buttnerae* em tambaqui (*Colossoma macropomum*).

- Estimar os parâmetros de genética quantitativa associados à resistência e resiliência ao parasita acantocéfalo em famílias de tambaqui,

- Analisar os parâmetros genéticos das características relacionadas à doença causada por *N. buttnerae* em uma população experimental de tambaqui, comparando métodos baseados em pedigree e informações genômicas.

- Avaliar a acurácia da predição genômica para as características de resistência (RTct) e resiliência (RSf e RSt) ao endoparasita em tambaqui. utilizando dados genômicos, empregando um painel de SNPs de baixa densidade.

CAPÍTULO I

Accuracy of genotype imputation of a low-density SNP array for the Amazon fish *Colossoma macropomum*

Accuracy of genotype imputation of a low-density SNP array for the Amazon fish *Colossoma macropomum*

Running title: Genotype imputation accuracy for the Tambaqui fish

John F. G. Agudelo[a], Vito A. Mastrochirico-Filho[a,b], Baltasar F. G. Neto, Raquel B. Ariede[a], José M. Yáñez[c], Gustavo M. R. Valladão[d], Diogo T. Hashimoto[a]*

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

[b] São Paulo State University (Unesp), School of Sciences, 17033-360 Bauru, SP, Brazil

[c] Faculty of Veterinary and Livestock Sciences, University of Chile, Santiago, Chile

[d] Nilton Lins University, Manaus, AM, Brazil

* 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

Acknowledgments: The authors thank CNPq (312250/2021-5), CAPES (Code 001), FAPESP (2020/07959-5) and FAPEAM (2020/07959-5) by funding this study.

Abstract

Tambaqui *Colossoma macropomum* is the main Amazon native species target for aquaculture in South America, however breeding programs for this species are still incipient. High-density SNP arrays are valuable resources for aquaculture breeding initiatives for selecting superior genotypes. However, genotyping expenses can be prohibitive, particularly for non- or semi-industrial species. A cost-effective strategy for genotyping native fish involves creating low-density SNP arrays followed by genotype imputation to higher density. In the present study, a 1K SNP array was developed and validated for tambaqui, which can be applied together with imputation techniques for the use in breeding programs. The imputation accuracy from different SNP densities (9K, 3K, 2K, 1K and 0.5K) to a medium-density SNP array (30K Affymetrix® Axiom® SerraSNP) was assessed. The 1K density offered the best trade-off regarding the minimal number of SNPs without compromising the accuracy of genotype imputation (accuracy at individual and SNP level of 0.93). Therefore, the 1K subset was then used to create a commercial array using Agriseq™ targeted genotyping-by-sequencing (Thermofisher). The 1K array was validated in 192 DNA samples, revealing the high quality of the low-density SNP array for genotyping tambaqui individuals. Through genome-wide coverage and high polymorphism levels, this low-density panel proved to be effective for exploring genetic variation within diverse populations. Population analyses using the 1K panel yielded an efficient and useful tool in terms of genetic characterization of sampled broodstocks. The low-density SNP panel presented here is a valuable resource for genetic improvement programs designed for this Amazon native species.

Keywords: Tambaqui; SNP Array; Genotyping; Breeding Programs; Genetic Improvement

1. Introduction

The tambaqui fish (*Colossoma macropomum*) is classified as an omnivorous species, with preference for frugivorous behavior, in contrast to its closely related predatory counterparts, the piranhas of the Serrasalmidae family (Woynárovich and Van Anrooy, 2019). This unique species exhibits remarkable resilience, rapid growth, artificial feed acceptance, substantial productivity, and holds significant commercial value in international markets (Valladão et al., 2018). Tambaqui stands out as the principal Amazon fish species cultivated in Brazil, particularly concentrated in the northern region of the country. As evidenced by statistics from the Brazilian Institute of Geography and Statistics (IBGE, 2022), aquaculture production of tambaqui attained 94.6 thousand tons in 2021. Moreover, this remarkable fish species possesses desirable characteristics that make it advantageous to aquaculture development in Latin America. As a result, tambaqui aquaculture has experienced successful expansion beyond Brazil, extending to countries such as Bolivia, Colombia, Peru, Ecuador, and Venezuela (Valladão et al., 2018).

In Brazil, most of the native fish aquaculture rely on unselected stocks, in which breeding programs are still in the beginning stage of development. Genetic selection programs have been initiated in tambaqui targeting efficient growth and disease resistance (Agudelo et al., 2022). As research and breeding techniques continue to advance, the application of genomic selection for complex traits is poised to revolutionize tambaqui aquaculture and contribute significantly to the industry's growth. This modern approach allows fish breeders to make more precise breeding decisions, thereby accelerating the

genetic progress of tambaqui populations. One of the primary challenges in implementing genomic selection-based breeding programs for native species is associated with the high cost of investment. In addition, a significant portion of native fish production is carried out by small to medium-scale producers, who generally lack the financial solvency and specialized workforce to invest in genetic improvement programs.

Recently, more comprehensive genomic studies have been conducted on tambaqui, including the sequencing of a reference genome (Hilsdorf et al., 2021), population genetic studies (Agudelo et al., 2022), linkage mapping, identification of Quantitative Trait Loci (QTLs) (Ariede et al., 2020; 2022), and the development of a dense array of Single Nucleotide Polymorphisms (SNPs) (Mastrochirico-Filho et al., 2021). However, the high cost of genotyping thousands of SNPs for accurate genomic prediction in a species primarily utilized in small-scale aquaculture, remains a limiting factor in the routine application of genomic selection programs.

An alternative approach to mitigate costs involves the adoption of lower density marker panels and the imputation of unobserved genotypes using data from higher density SNP arrays of related animals, thereby enhancing the genomic prediction accuracy (Phocas et al., 2022). Recent research has assessed the imputation accuracy from low- to high density panels in different fish species (Tsai et al., 2017; Yoshida et al., 2018; Tsairidou et al., 2020; Garcia et al., 2022; Kriaridou et al., 2023). Nevertheless, a fully validated low-density SNP array suitable for genotype imputation at a commercial scale is yet to be made available in tambaqui.

A low-density SNP panel combined with imputation techniques may be used to decrease costs of genotyping without compromising the reliability of genomic predictions. This study evaluated the feasibility of genotype imputation from low- (0.5, 1, 2, 3 and 9K) to the medium-density the Affymetrix® Axiom® SerraSNP array (30K), as a reference (Mastrochirico-Filho et al., 2021), in an experimental breeding population from Brazil.

2. Material and Methods

2.1. 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 number 1038/21) of Faculdade de Ciências Agrárias e Veterinárias, UNESP, Campus Jaboticabal, SP, Brazil.

2.2. Experimental population and SNP genotyping

The population under investigation in this study constitutes the breeding population maintained at the Aquaculture Center, São Paulo State University (UNESP), located in Jaboticabal city, São Paulo State, Brazil. The animals used for SNP genotyping were previously described in an analysis of GWAS for resistance against *Aeromonas hydrophila* in tambaqui by Ariede et al. (2020, 2022), which consisted of 18 half and full-sib families.

Breeding involved 9 females and 14 males, with four males each mating with two females (resulting in four half-sibling groups, equivalent to eight full-sibling families), three males each mating with one female (forming one half-sibling group, equivalent to three full-sibling families), two males each mating with one female (yielding one half-sibling group, equivalent to two full-sibling families), and five males and five females paired individually (resulting in five full-sibling families). To obtain these families, induced spawning was conducted using carp pituitary extract (Danubio piscicultura LTDA), administered in two doses with a 12-hour interval. The first and second dosages were 0.5 and 5.5 mg/kg, respectively, and a single dosage of 2.5 mg/kg of carp pituitary extract was used for males, simultaneously with the second dosage for females (Pinheiro and Silva, 1988).

Genomic DNA was extracted from blood samples obtained from 280 fish, consisting of 23 parents and 257 offspring, using the commercial Wizard Genomic Kit from Promega, following the manufacturer's instructions. The quality of the DNA was evaluated through two methods. Firstly, a NanoDrop spectrophotometer® from Thermo Fisher Scientific was utilized, which allowed for the assessment of DNA concentration and purity. Secondly, agarose gel electrophoresis was performed to further verify the integrity and quality of the DNA samples. To quantify the extracted DNA accurately, the Qubit fluorescence detector, in conjunction with the Qubit dsDNA BR Assay kit® from Invitrogen, was employed. This approach enabled precise measurements of DNA concentration in units of nanograms per microliter (ng/µl). The combination of these techniques ensured a comprehensive assessment of the DNA's quality, facilitating reliable subsequent analyses and experiments.

SNP genotyping of the fish was conducted using the multispecies SerraSNP Affymetrix® Axiom® array, a custom-designed array developed by our research group and commercially available for pacu *Piaractus mesopotamicus* and tambaqui (Mastrochirico-Filho et al., 2021). The SerraSNP Affymetrix® Axiom® array comprises 29,575 SNPs specific for tambaqui. The genomic DNA samples were sent to Thermo Fisher Scientific, located in California, USA, for the genotyping process. The raw data, containing the intensity calculation results (CEL files), was imported into the Axiom Analysis Suite (v2.0.035, Affymetrix) for quality control analysis and genotype calling, utilizing default parameters. To pass the initial quality control assessment, samples with a dish quality control (DQC) value exceeding 0.82 and a QC (quality control) call rate exceeding 0.97 (following the "Best Practices Workflow" recommended by Affymetrix) were considered satisfactory. Furthermore, additional quality control procedures were conducted using the software PLINK 1.9 (Purcell et al., 2007), as follows: 1) Minor allele frequency (MAF) filtering was performed with a threshold of < 0.05, which was applied to the entire population, rather than within each family, to enhance the stringency in excluding potential genotyping errors; 2) Mendelian error analysis was conducted for each family, using the parameters –me 0.05 0.1, to identify and eliminate individuals and/or SNPs based on the Mendelian error rate.

2.3.Selection of low-density subsets of SNPs for imputation accuracy

Different low-density SNP subsets were selected to evaluate the accuracy of genotype imputation. For this step, we used the genotypes of 81 tambaqui individuals from 5 different

broodstock populations, considering 2 populations from North and 3 populations from Southeast regions of Brazil. All samples were genotyped by the multispecies SerraSNP Affymetrix® Axiom® array, according to the topic 2.2. The quality control (QC) filter was carried out by the PLINK 1.9 software, using Hardy-Weinberg equilibrium (HWE) (p-value < 0.05 after Bonferroni correction), call rate of 90%, and MAF (< 0.01). In addition, the extent of linkage disequilibrium (LD) between SNPs was assessed within each broodstock population using the LD-based variant pruner by PLINK. A range of r^2 pruning thresholds, varying from 0.9 to 0.6, was then applied to eliminate redundant combinations of SNPs. This process resulted in the selection of subsets of common SNPs among the broodstock populations, ensuring the retention of non-highly redundant genotypic information for further analyses.

2.4. Accuracy of genotype imputation on low-density SNPs

After extracting the low-density subsets of SNPs, a cross-validation system consisting of 10 replicates was employed to evaluate the imputation accuracy in the population comprising the 34 parents and 238 offspring. In each replicate, a random subset of animals was assigned using pipelines implemented in the R software (R Core Team, 2018). For this setup, 10% of the individuals were designated for validation purposes, while the remaining 90% served as a reference for imputation. As a result, each replicate consisted of 27 to 29 individuals for validation, leading to a total of 272 individuals used in the analysis. To simulate the imputation scenario, individuals in the validation category had their SNPs

masked, leaving only the low-density subsets of SNPs representing the low-density panel. In contrast, individuals classified in the reference category remained unmodified in their genotype data. For the validation category, genotypes were imputed using the FIMPUTE3 software (Sargolzaei et al., 2014). This process was conducted across all 10 replicates to mitigate the effects of stochastic sampling. Subsequently, the imputed genotypes obtained from the validation replicates were compared to the true genotypes to determine the accuracy of the imputation. The Pearson correlation coefficient (R) was employed as a measure to assess the accuracy, and the "R" value was estimated using the following formula:

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(g_i - \bar{g})}{\sqrt{\sum_{i'1}^{n}(x_i - \bar{x})^2 \sum_{i'1}^{n}(g_i - \bar{g})^2}}$$

where for each SNP in which the true genotypes were masked, the possible genotypes were coded as 0, 1, or 2, representing the potential number of the smallest allele for that SNP in the target population; x_i represents the imputed genotype for individual *i*, and $\bar{x}$ represents the mean value of the imputed genotypes across all individuals. Likewise, g_i and $\bar{g}$ represent the variables corresponding to the observed genotypes, and *n* denotes the number of individuals in the validation group. The calculation of R, used as a measure of imputation accuracy, was performed at both the SNP marker and individual levels. SNPs that were imputed accurately demonstrated an accuracy of over 80% (R > 0.8) across all animals. The evaluation of imputation accuracy was carried out within the chromosomes of tambaqui, which consist of 27 pairs. Considering that the ends of the chromosomes are known to have

low imputation accuracy, one hundred markers were evaluated at both ends of each chromosome in the lower density subsets (1K, and 0.5K densities).

2.5. Development of the low-density SNP array

The development of the low-density SNP array was centered on a subset of SNPs that ensured both a reduced number of SNPs and maintained a high imputation accuracy. For constructing the array, the AgriSeq® Targeted Genotyping by Sequencing (Agriseq tGBS) genotyping platform was chosen. The SNPs were required to be at least 400 base pairs (bp) long, with the target SNP positioned at position 201. All the SNPs used in the low-density array were derived from the Affymetrix® Axiom® SerraSNP array (Mastrochirico-Filho et al., 2021), where the SNPs are represented by 71-mer nucleotide sequences. To align these 71-mer sequences against the tambaqui genome (NCBI RefSeq assembly accession: GCA_904425465.1, assembled at the chromosome level by Ariede et al., 2022), the BLASTn pipeline was employed. The goal was to obtain the necessary total 400 bp loci for each SNP to implement the Agriseq tGBS strategy. Strict criteria were applied during the alignment process, mandating a high percentage of sequence similarity and full coverage of the query sequence concerning the genomic region of the tambaqui genome. The parameters applied were -perc_identity ranging from 97% to 99% and -qcov_hsp_perc 100%. The resulting 400 bp flanking sequences containing the SNPs were extracted using bioinformatics pipelines and subsequently submitted to Thermo Fisher Scientific for quality control analysis and primers design.

The validation of the Agriseq tGBS low-density array involved genotyping a total of 192 DNA samples from tambaqui individuals, considering 96 individuals from fish farms located in the Northern region of Brazil (Amazon region), and 96 individuals produced in the Southeast region of Brazil. DNA extraction and quality parameters were performed according to the topic 2.2. The genomic DNA samples were sent to Thermo Fisher Scientific in Austin, Texas, USA, for SNP genotyping. The SNP data were subjected to a QC check using the PLINK, applying the parameters MAF > 0.01, call-rate for samples and SNPs > 0.15, Hardy-Weinberg equilibrium (p = 1e-06). The performance of the developed low-density SNP array was evaluated applying the PLINK 1.9 software (Chang et al., 2015) for estimation of the MAF frequencies and mean of heterozygosity per genotyped individual. To evaluate the genetic structure among the analyzed populations, the presence of genotypic clusters among the populations from the North and Southeast regions was investigated. This was achieved by performing a two-dimensional multidimensional scaling analysis (MDS) on the genomic identity-by-state (IBS) matrix. Furthermore, population structuring analysis was performed using Structure 2.3.4 software (Pritchard et al., 2000) to estimate the optimum number of population clusters (K). The range of K values was predefined from 1 to 3.. The analysis was performed in 5 replicated runs for each K value using 50,000 iterations after a burn-in period of 5,000 iterations. The determination of the number of clusters in the dataset was performed using Structure Harvester software (Evanno et al., 2005).

3. Results

3.1 Selection of low-density subsets of SNPs for imputation accuracy

The QC processes to select high-quality SNPs present in each broodstock population (Broodstock 1-5), which composed the low-density subsets for the evaluation of genotype imputation accuracy, were described in Table 1 and Figure 1. The high-quality SNPs ranged from 14,985 SNPs in Broodstock 4 to 16,792 SNPs in Broodstock 5. The Broodstock 5 had a higher proportion of high-quality SNPs, with higher MAF values and genotyped markers present in most of its individuals (call rate > 0.90). On the other hand, Broodstock 4 had a higher proportion of its markers removed (approximately 13.7%) (Table 1).

Table 1. QC process of the SNPs used to compose the low-density subsets for evaluating the accuracy of genotype imputation in tambaqui. The high-quality SNPs were filtered from five tambaqui broodstocks of fish farms.

QC process	Broodstock 1 (N = 20)	Broodstock 2 (N = 33)	Broodstock 3 (N = 9)	Broodstock 4 (N = 9)	Broodstock 5 (N = 10)
Before QC[1]	15,821	15,821	17,372	17,372	17,372
HWE[2]	15,814	15,793	17,372	17,372	17,372
Call rate[3]	15,814	15,793	16,682	16,71	17,296
MAF[4]	15,668	15,626	16,398	14,985	16,792

[1] SNPs from the validation of the *Affymetrix® Axiom® SerraSNP* array (Mastrochirico-Filho et al., 2021).

[2] SNPs in Hardy-Weinberg Equilibrium (HWE) ($p > 0.05$, after Bonferroni correction).

[3] SNPs with call rate > 90%

[4] SNPs with minor allele frequency (MAF) > 0.01

In relation to the filtering of linkage disequilibrium aimed at removing redundant information among the broodstock, the squared correlation (r^2) between pairs of SNPs varied from 1 (no pruning) to 0.6. This process resulted in the creation of distinct low-density subsets: 9,471 SNPs (9K), 3,541 SNPs (4K), 2,169 SNPs (2K), 1,022 SNPs (1K), and 500 SNPs (0.5K), as illustrated in Figure 1. The imputation accuracy of each subset was then evaluated.

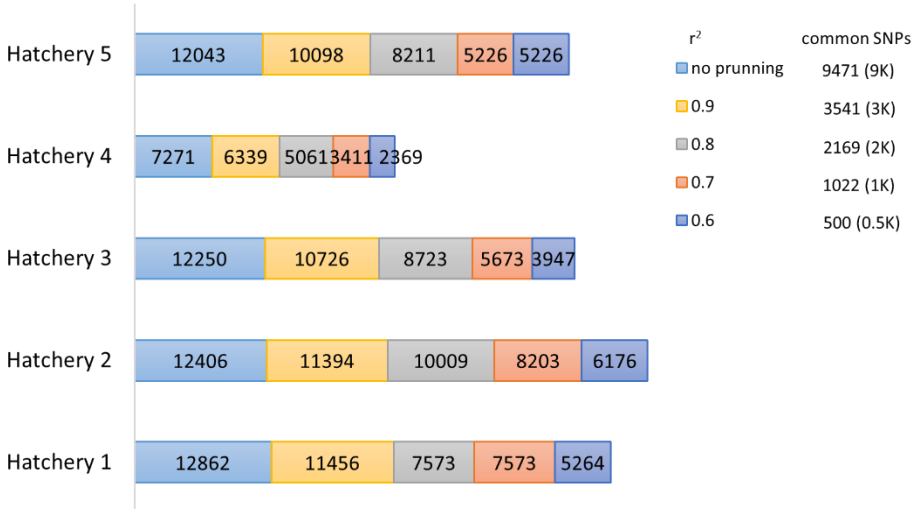


Figure 1. Linkage disequilibrium filtering (LD) applied to remove redundant SNP combinations. The common SNPs among the populations were used to form the low-density subsets that will be evaluated for imputation accuracy.

3.2 Accuracy of genotype imputation on low-density SNPs

After QC analysis, 280 genotyped samples and 17,368 SNPs were used as final dataset for assessing the accuracy of imputation. The genotype imputation accuracies, both at the SNP and individual levels, for each density is illustrated in Figure 2 and Table 2. In brief,

consistently high to moderate accuracy values were maintained across densities ranging from 9K to 1K. However, a notable decrease in accuracy was observed when the density reduced to 0.5K. Specifically, at the SNP level, mean accuracies ranged from 0.98 (SD 0.04) to 0.93 (SD 0.10) for densities 9K to 1K, decreasing to 0.81 (SD 0.17) at the 0.5K density. At the individual level, the mean accuracy reached 0.83 (SD 0.05) for the 0.5K density, whereas it exceeded 0.93 for the other densities. The proportion of SNPs with accuracy greater than 0.8 was notably lower at the 0.5K density, ranging from 88.4% when assessing accuracy solely based on SNPs to 76.8% when considering accuracy based on animals. Conversely, for the other densities at the SNP level, more than 90% of the SNPs exhibited accuracies exceeding 0.8. accuracy, enhancing the imputation performance of the 1K subset along the chromosome regions. Notably, a minority of SNPs with accuracies below 80% were more prevalent, particularly for LG22.

In contrast, the 0.5K SNP density resulted in significantly lower accuracy at the chromosome extremities, with average values ranging from 0.34 (SD 0.22) to 0.91 (SD 0.08). This indicated a higher prevalence of low and variable imputation accuracies compared to the other densities (see Figure S1). Considering the accuracy for each low-density subset, 1K SNPs were identified as the optimal density for constructing the low-density SNP array for tambaqui.

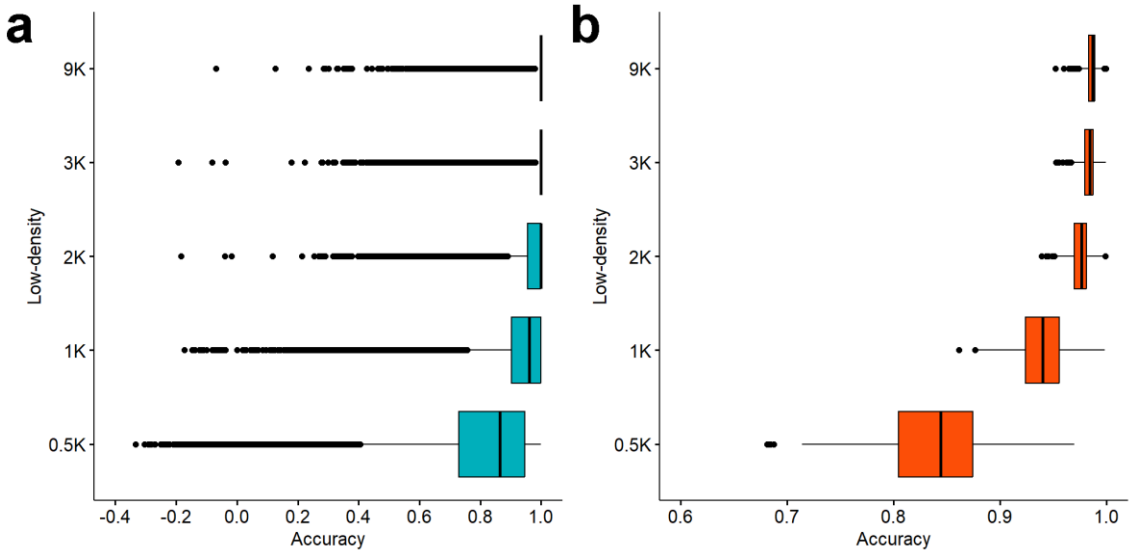


Figure 2. Comparing the Pearson correlation coefficient (r) between imputed and observed genotypes as a measure of imputation accuracy (R) for each low-density subsets of SNPs. The accuracy was assessed for imputed SNPs (a) and at the individual level (b).

Table 2. The summary of the accuracy of genotype imputation using different low-density (LD) subsets of SNPs. The accuracy was described for imputed SNPs and at the individual level. The mean (SD), minimum and maximum accuracy, as well as percentage (%) of SNPs with accuracy greater than 0.8 in each scenario, were registered (% > 0.8).

		Accuracy at SNP level		Accuracy at individual level			
LD subsets	Imputed SNPs	Mean (SD)	% > 0.8	Mean (SD)	Min.	Max.	% > 0.8
9K	7,897	0.98 (0.04)	99.0	0.98 (0.01)	0.95	0.99	100
3K	13,816	0.98 (0.04)	98.9	0.98(0.01)	0.95	0.99	100
2K	14,882	0.97 (0.05)	97.9	0.98 (0.01)	0.94	0.99	100
1K	14,822	0.93 (0.10)	90.8	0.94 (0.02)	0.86	0.99	100
0.5K	16,836	0.81 (0.17)	64.0	0.84 (0.05)	0.68	0.97	76.8

The analysis of accuracy means was also conducted concerning the MAF values, as shown in Figure 3. The outcomes revealed that imputation accuracies displayed only a marginal increment in their values with the increase in MAF values. Notably, rare SNPs were absent from the analysis due to their removal during the QC steps. Interestingly, SNPs with low MAF values (0.01-0.049) exhibited high to moderate mean accuracy, surpassing even SNPs with more frequent alleles when lower densities were taken into consideration.

The accuracy of imputed SNPs across the 27 chromosomes for each low-density subset is demonstrated in Figures 4a-e. The findings revealed that imputed SNPs situated at both the beginning and end of each chromosome displayed lower imputation accuracy compared to SNPs located in other regions of the chromosomes. This decrease in accuracy at both ends was particularly pronounced when evaluating the 1K and 0.5K subsets (Figure S1). Nevertheless, for the 1K subset, the mean accuracy fluctuated between 0.67 (SD 0.20) and 0.96 (SD 0.07) in these regions, suggesting no significant decline in imputation accuracy (refer to Figure S1). Consequently, a substantial proportion of imputed SNPs exhibited high, enhancing the imputation performance of the 1K subset along the chromosome regions. Notably, a minority of SNPs with accuracies below 80% were more prevalent, particularly for LG22.

In contrast, the 0.5K SNP density resulted in significantly lower accuracy at the chromosome extremities, with average values ranging from 0.34 (SD 0.22) to 0.91 (SD 0.08). This indicated a higher prevalence of low and variable imputation accuracies compared to the other densities (see Figure S1). Considering the accuracy for each low-density subset, 1K

SNPs were identified as the optimal density for constructing the low-density SNP array for tambaqui.

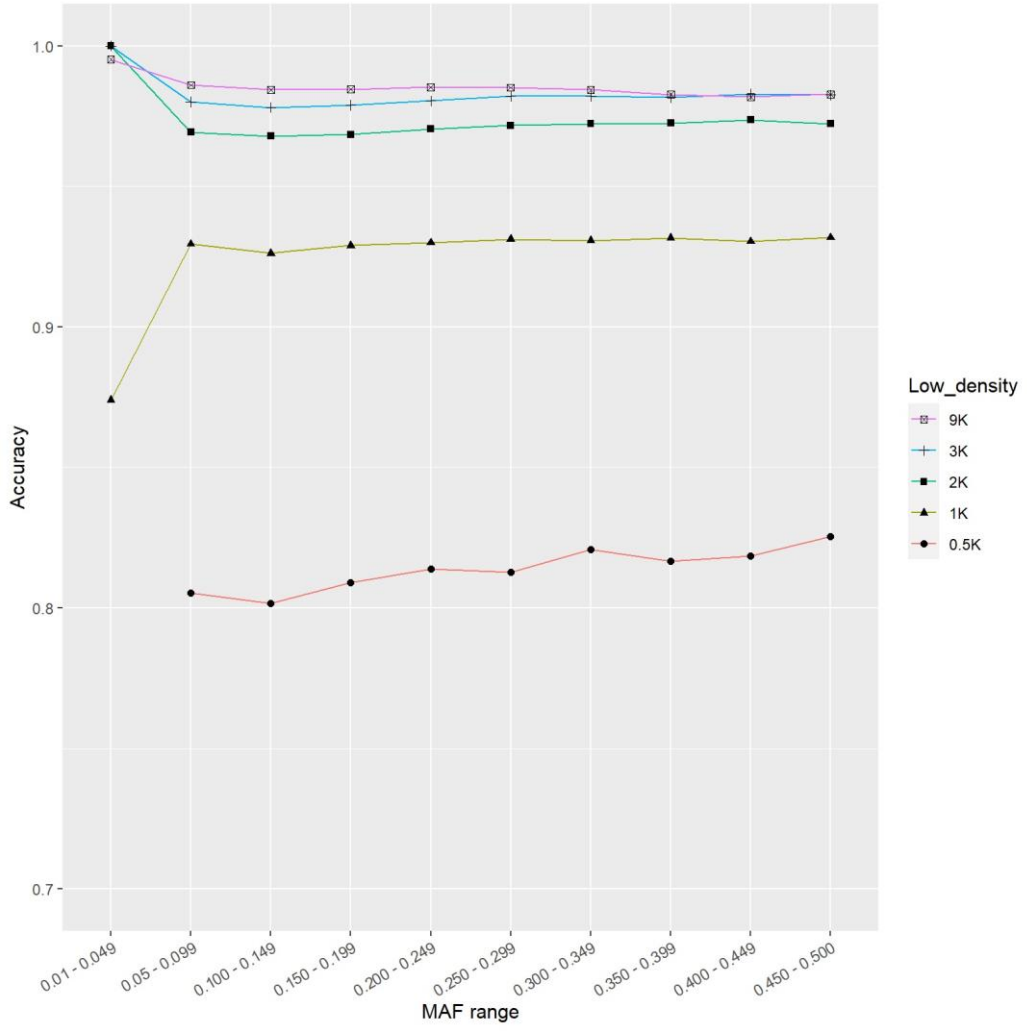


Figure 3. Correlation between accuracy (R^2) and minor allele frequency (MAF) range examined for different low-density subsets: 9,471 SNPs (9K), 3,541 SNPs (4K), 2,169 SNPs (2K), 1,022 SNPs (1K), and 500 SNPs (0.5K).

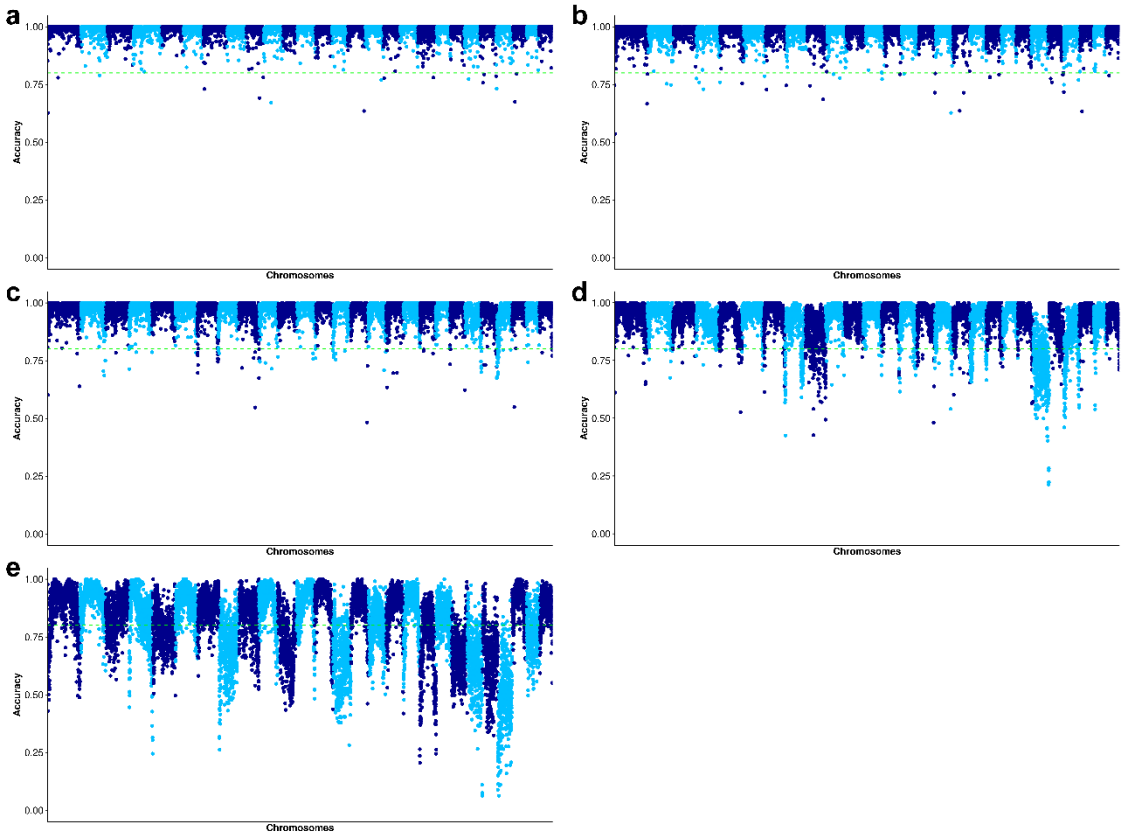


Figure 4. Squared Pearson correlation measure of imputation accuracy (R^2) across the 27 chromosomes of tambaqui (*Colossoma macropomum*) after imputation using low-density subsets, included a) 9K, b) 3K, c) 2K, d) 1K, e) 0.5K SNPs.

3.4 Validation of the low-density SNP array

The SNPs amount of the 1K low-density array in the genetic map of tambaqui (Ariede et al., 2022) was succinctly outlined in Table 3. There are between 19 and 74 markers distributed across the specified chromosomal regions. This implies that approximately 3.58% to 10.33% of all SNPs covering the chromosomal regions of the genetic map were incorporated into the low-density panel.

Table 3. Characterization of the SNPs from the 1K SNP array into the genetic map of tambaqui.

Chromosomes	Genetic Map (nº SNPs)[1]	1K array (nº SNPs)[2]	% SNPs[3]
1	1,106	54	4.88
2	877	74	8.44
3	811	67	8.26
4	799	71	8.89
5	772	40	5.18
6	772	72	9.33
7	737	58	7.87
8	684	35	5.12
9	675	30	4.44
10	650	56	8.62
11	647	48	7.42
12	629	58	9.22
13	629	65	10.33
14	631	50	7.92
15	611	52	8.51
16	623	32	5.14
17	582	38	6.53
18	555	33	5.94
19	560	55	9.82
20	530	40	7.55
21	532	43	8.08
22	531	19	3.58
23	517	39	7.54
24	513	29	5.65
25	501	44	8.78
26	456	33	7.24
27	444	27	6.08
total	17,374	1,262	7.26

[1] The total number of SNPs distributed in the genetic map of tambaqui (Ariede et al., 2022).
[2] The distribution of the SNPs from the 1K SNP array into the genetic map.
[3] Proportion (%) of the SNPs from the 1K SNP array in relation to the total of SNPs per chromosome of the genetic map.

For the validation of the 1K SNP array, the QC analysis of the genotyping retained 843 SNPs (70% of the total SNPs) and 169 individuals (86% of the total individuals). Low frequencies of the mutant alleles (MAF values from 0.01 to 0.05) represented only 1.6% of the total SNPs. Highly polymorphic SNPs with MAF values ranging from 0.45 to 0.50 were more abundant, representing 21.7% of the total SNPs (Figure 5a). Differences in mean heterozygosity per individual were evident, highlighting genetic distinctions between the populations in the Southeast and North regions of Brazil (Figure 5b). The Southeast population exhibited a higher frequency of heterozygous loci (74%) compared to individuals from the North population, underscoring greater genetic diversity in the Southeast.

The genetic structure analysis supported the hypothesis of the onset of genetic differentiation between the tambaqui populations in the Southeast and North regions (Figure 6a and 6b). This validation underscores the efficacy of the low-density SNP panel in capturing genetic diversity within both populations.

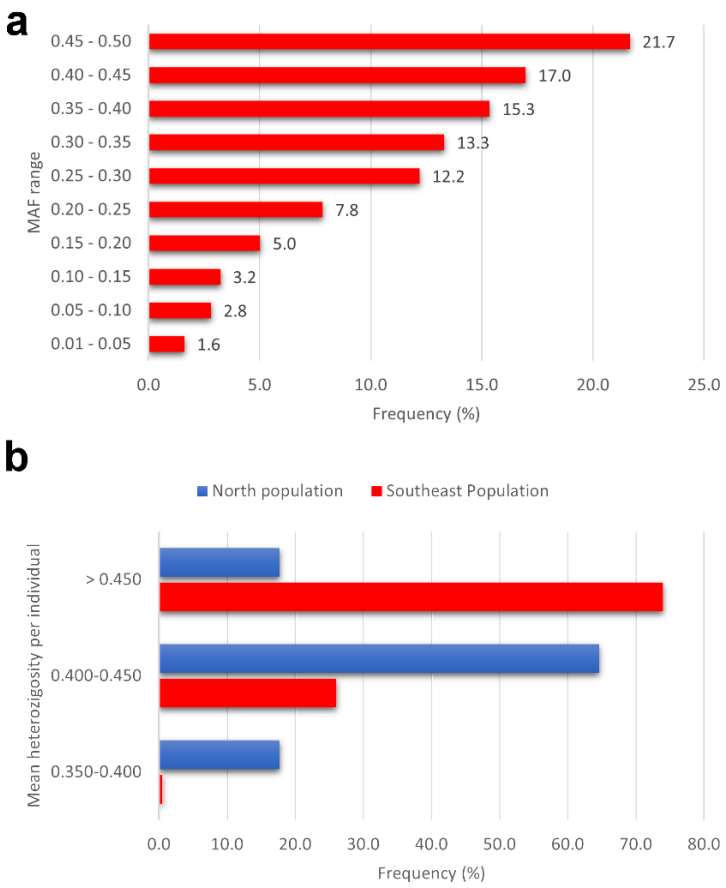


Figure 5. Frequency of MAF values of tambaqui populations from North and Southeast regions of Brazil (a). Mean of heterozygosity per individual from North (blue) and Southeast (red) populations (b).

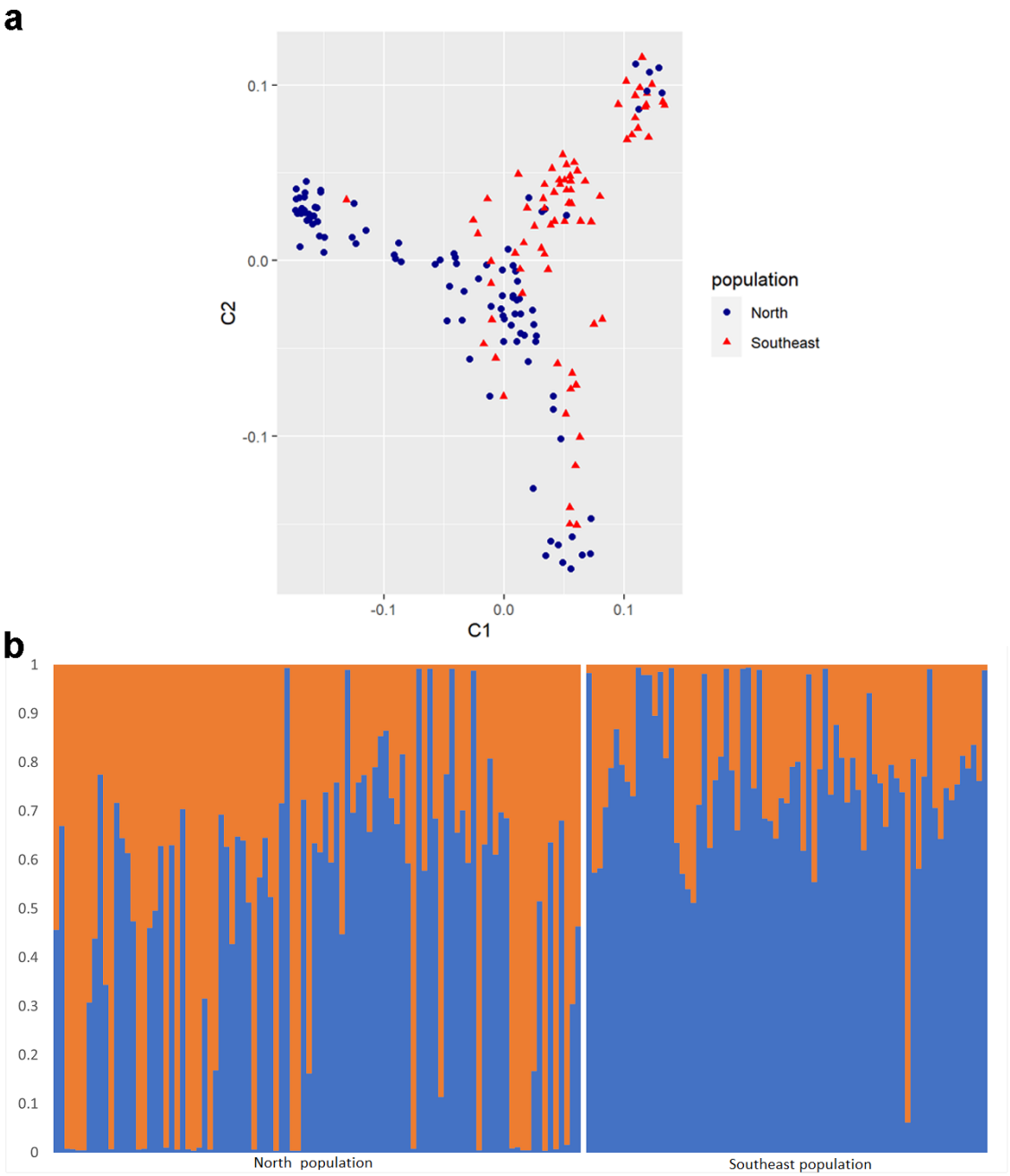


Figure 6. Analysis of the genetic structure of tambaqui populations from North and Southeast regions of Brazil. The figure (a) represents the multidimensional scaling (MDS) plot of the individual IBS distances for the North (blue) and Southeast (red) populations. The figure (b) represents the STRUCTURE analysis as a function of the number of putative genetic clusters (K). The genetic structure was analyzed approaching K = 2 according to the Delta K statistic. Each vertical bar represents an individual. Populations are separated by vertical white bars. The color proportions of each bar correspond to the estimated fractions of association of the individuals in each of the clusters.

4. Discussion

The developed SNP array stands as one of the pioneering low-density panels optimized through genotype imputation for the Neotropical aquaculture. This molecular tool allows substantial reduction in genotyping costs (in a scenario of genotyping 1,000 fish), from approximately $50 per sample when using the medium-density array (Axiom technology), to $10 per sample (Agriseq platform), representing a reduction of 80.0% in genotyping expenses. The low-cost genotyping strategy holds significant relevance for target species predominantly cultivated by small-scale farmers, herein exemplified by the Amazon fish tambaqui.

Our results indicated that high imputation accuracy values (above 0.80) slightly decreased with decreasing SNP density in the validation process, a trend observed in prior studies on low-density SNP arrays for Atlantic salmon (Tsairidou et al., 2020) and Nile tilapia (Yoshida et al., 2019). Approximately 99.0% to 90.8% of genotypes achieved an accuracy above 0.80 when marker densities ranged from 9K to 1K SNPs, respectively. However, imputation accuracy sharply declined when the density decreased to 0.5K SNPs. Similar trends were noted at the genotyped animal level, with accurate imputations being more prevalent between 9K and 1K densities (100%), decreasing to 76.8% at the 0.5K density. Based on these findings, the 1K low-density genotyping array, followed by genotype imputation using as reference the 30K SerraSNP Affymetrix® Axiom® array, could be tested for further cost-effective genomic selection programs in tambaqui.

Previous studies have suggested a positive correlation between high MAF values and accurate imputation, in which SNPs presenting low MAF are challenging for imputation (Brøndum et al., 2012; Ma et al., 2013). Although our results revealed a subtle increase in accuracy with rising MAF values, we did not observe a high correlation between these parameters. Thus, our findings suggests that MAF values are closely linked to inaccurately imputed genotypes, and the imputation method effectively mitigated any impact of MAF on the results. Consequently, the imputation accuracy method remains preferable in imputation evaluations, as highlighted in prior studies (Calus et al., 2014; Garcia et al., 2022).

SNPs located at chromosome ends also present challenges for imputation, displaying lower accuracy across all tested densities. Accuracy notably declined, particularly with very low SNP densities, such as the 0.5K density in this study, consistent with previous research (Ventura et al., 2016; Yoshida et al., 2018). Chromosomal ends typically exhibit lower imputation accuracy. If traits of interest are influenced by this region, genomic association studies may introduce bias and compromise the investigation of imputed genotypes. To address this issue, including more SNPs at chromosome ends, either by increasing SNP array density or utilizing more comprehensive SNP arrays in future analyses, could serve as a potential strategy to be tested.

The genotyping technology employed for constructing the SNP array was targeted genotyping-by-sequencing (*Agriseq tGBS*), which was carefully selected for its high accuracy in genotyping heterozygous loci and lower rates of missing data (Ott et al., 2017), which was corroborated in the present study in the validation process. Moreover, the genetic

diversity and structure analysis corroborated the distinct patterns previously observed between the North and South populations of farmed tambaqui in Brazil (Mastrochirico-Filho et al., 2021; Agudelo et al. 2022), and reveals the efficiency of using the 1K low-density array in population analysis.

The development of tambaqui production is currently in its early domestication stages and lacks technologies that could facilitate the establishment of effective breeding programs. The susceptibility of tambaqui stocks to diseases (bacterial and parasites), exacerbated by the intensification of production, has led to significant economic losses (Ariede et al., 2020; Valladão et al., 2020; Hilsdorf et al., 2022). The genetic control of disease resistance is known to be distributed among several Quantitative Trait Loci (QTLs), each elucidating only a small portion of the genetic variation. In this case, the prediction of genetic merit of animals, based on genomic estimated breeding values (GEBV, referred to as Genomic Selection), has been shown to enhance the genetic gain in aquaculture breeding (Yáñez et al., 2023). Hence, the validated strategy presented here, involving a low-density array followed by genotype imputation, can serve as a framework for efficiently incorporate disease resistance into genomic selection programs for the Amazon fish tambaqui.

5. Conclusions

We proposed a cost-effective solution involving a 1K SNP array with imputation techniques, offering precise genotyping without exorbitant expenses. The 1K density, validated with high

accuracy (0.93), proves to be an optimal trade-off in terms of accuracy of imputation. In addition, this genomic tool may be employed to population genomics studies in order to evaluate genetic variability and differentiation of populations.

Conflict of interest statement

The authors have no conflict of interest to declare.

Funding information

CNPq (312250/2021-5), CAPES (Code 001), FAPESP (2020/07959-5) and FAPEAM (2020/07959-5)

Author contributions

Conceptualization: JFGA, VAMF and DTH; Formal analysis: JFGA, VAMF, BFGN, RBA and DTH; Funding acquisition: JFGA, GMRV, and DTH; Investigation: JFGA, VAMF, BFGN, RBA, GMRV and DTH; Methodology JFGA, VAMF, BFGN, RBA, GMRV and DTH; Project administration: DTH; Supervision: DTH; Roles/Writing – original draft: JFGA, VAMF, BFGN, RBA and DTH; Writing – review & editing JFGA, VAMF, BFGN, RBA, JMY GMRV and DTH.

Supplementary material

The following online material is available for this article: Figure S1 – Mean accuracy (R2)

6. References

Agudelo, J. F. G., Mastrochirico-Filho, V. A., Borges, C. H. de S., Ariede, R. B., Lira, L. V. G., Neto, R. R. de O. et al. (2022) Genomic selection signatures in farmed *Colossoma macropomum* from tropical and subtropical regions in South America. Evolutionary Applications, 15(4), 679–693. Available from: https://doi.org/10.1111/eva.13351

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.

Ariede, R. B., Freitas, M. v., Lira, L. V. G., Agudelo, J. F. G., Borges, C. H. S., Mastrochirico-Filho, V. A. et al. (2022) Linkage map for chromosome-level genome anchoring and genome-wide association study for resistance to *Aeromonas hydrophila* in *Colossoma macropomum*. Aquaculture, 560. Available from: https://doi.org/10.1016/j.aquaculture.2022.738462

Brøndum, R.F., Ma, P., Lund, M.S., Su, G. (2012) Short communication: genotype imputation within and across Nordic cattle breeds. Journal of Dairy Science, 95, 6795–6800. Available from: https://doi.org/10.3168/jds.2012-5585

Calus, M.P.L., Bouwman, A.C., Hickey, J.M., Veerkamp, R.F., Mulder, H.A. (2014) Evaluation of measures of correctness of genotype imputation in the context of genomic prediction: a review of livestock applications Animal, 8(11), 1743–1753. doi:10.1017/s1751731114001803

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., Lee, J. J. (2015) Second-generation PLINK: Rising to the challenge of larger and richer datasets. GigaScience, 4(1). Available from: https://doi.org/10.1186/s13742-015-0047-8

Evanno, G., Regnaut, S., Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol., 14, 2611–2620.

Garcia, B.F., Yoshida, G.M., Carvalheiro, R., Yáñez, J.M. (2022) Accuracy of genotype imputation to whole genome sequencing level using different populations of Nile tilapia. Aquaculture, 551. Available from: https://doi.org/10.1016/j.aquaculture.2022.737947

Hilsdorf, A.W.S., Uliano-Silva, M., 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. GigaByte gigabyte29, 1-14. doi: 10.46471/gigabyte.29.

Hilsdorf, A. W. S., Hallerman, E., Valladao, G. M. R., Zaminhan-Hassemer, M., Hashimoto, D. T., Dairiki, J. K., & Cyrino, J. E. P. (2022) The farming and husbandry of Colossoma

macropomum: From Amazonian waters to sustainable production. Reviews in Aquaculture, 14(2), 993-1027.

IBGE-Instituto Brasileiro de Geografia e Estatística. (2022) Produção da Pecuária Municipal 2022, Available from: https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/9107-producao-da-pecuaria-municipal.html?=&t=resultados, Accessed on November 20, 2023

Kriaridou, C., Tsairidou, S., Fraslin, C., Gorjanc, G., Looseley, M.E., Johnston, I.A. et al. (2023) Evaluation of low-density SNP panels and imputation for cost-effective genomic selection in four aquaculture species. Frontiers in Genetics, 14, 1194266. doi: 10.3389/fgene.2023.1194266

Ma, P., Brøndum, R.F., Zhang, Q., Lund, M.S., Su, G. (2013) Comparison of different methods for imputing genome-wide marker genotypes in Swedish and Finnish Red Cattle. Journal of Dairy Science, 96, 4666–4677. Available from: https://doi.org/10.3168/jds.2012-6316

Mastrochirico-Filho, V.A., Ariede, R.B., Freitas, M.V. *et al.* (2021) Development of a multi-species SNP array for serrasalmid fish *Colossoma macropomum* and *Piaractus mesopotamicus*, Scientific Reports, 11**,** 19289. https://doi.org/10.1038/s41598-021-98885-x

Ott, A., Liu, S., Schnable, J.C., Yeh, C.E., Wang, K., Schnable, P.S. (2017) tGBS® genotyping-by-sequencing enables reliable genotyping of heterozygous loci. Nucleic Acids Research, 45(21), e178. Available from: https://doi.org/10.1093/nar/gkx853

Phocas, F. (2022) Genotyping, the Usefulness of Imputation to Increase SNP Density, and Imputation Methods and Tools. In: Ahmadi, N., Bartholomé, J. (eds) Genomic Prediction of Complex Traits, Methods in Molecular Biology, vol 2467. Humana, New York, NY. Available from: https://doi.org/10.1007/978-1-0716-2205-6_4

Pinheiro-Lima, J.L.P., Silva, M.C.N., (1988) Tambaqui (*Colossoma macropomum*–Cuvier, 1818): ampliação do período de desova. CODEVASF,29, Brasília.

Pritchard, J., Stephens, M., Donnelly, P. (2000) Inference of population structure using multilocus genotype data. Genetics, 155, 945–959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D. et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics, 81(3), 559-575.

Sargolzaei, M., Chesnais, J.P., Schenkel, F.S. (2014) A new approach for efficient genotype imputation using information from relatives. BMC Genomics, 15, 478. Available from: https://doi.org/10.1186/1471-2164-15-478

Tsai, H.Y., Hamilton, A., Guy, D.R., et al. (2017) Genome wide association study identifies SNPs associated with early growth performance in Atlantic Salmon (*S. Salar*). Aquaculture, 472, 165-165. Available from: https://doi.org/10.1534/g3.117.040717

Tsairidou, S., Hamilton, A., Robledo, D., Bron, J.E., Houston, R.D. (2020) Optimizing Low-Cost Genotyping and Imputation Strategies for Genomic Selection in Atlantic Salmon. G3 Genes|Genomes|Genetics, 10(2), 581–590. Available from: https://doi.org/10.1534/g3.119.400800

Valladao, G. M. R., Gallani, S. U., Jerônimo, G. T., & Seixas, A. T. D. (2020) Challenges in the control of acanthocephalosis in aquaculture: special emphasis on *Neoechinorhynchus buttnerae*. Reviews in Aquaculture, 12(3), 1360-1372.

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

Ventura, R.V., Miller, S.P., Dodds, K.G. et al. (2016) Assessing accuracy of imputation using different SNP panel densities in a multi-breed sheep population. Genetics Selection Evolution, 48, 71. Available from: https://doi.org/10.1186/s12711-016-0244-7

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.

Yanez, J. M., Barria, A., Lopez, M. E., Moen, T., Garcia, B. F., Yoshida, G. M., & Xu, P. (2023) Genome-wide association and genomic selection in aquaculture. Reviews in Aquaculture, 15(2), 645-675.

Yoshida, G.M., Carvalheiro, R., Lhorente, J.P., et al. (2018) Accuracy of genotype imputation and genomic predictions in a two-generation farmed Atlantic salmon population using high-density and low-density SNP panels. Aquaculture, 491: 147-154.

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, Volume 9, Issue 8, 2597–2607, https://doi.org/10.1534/g3.119.400116

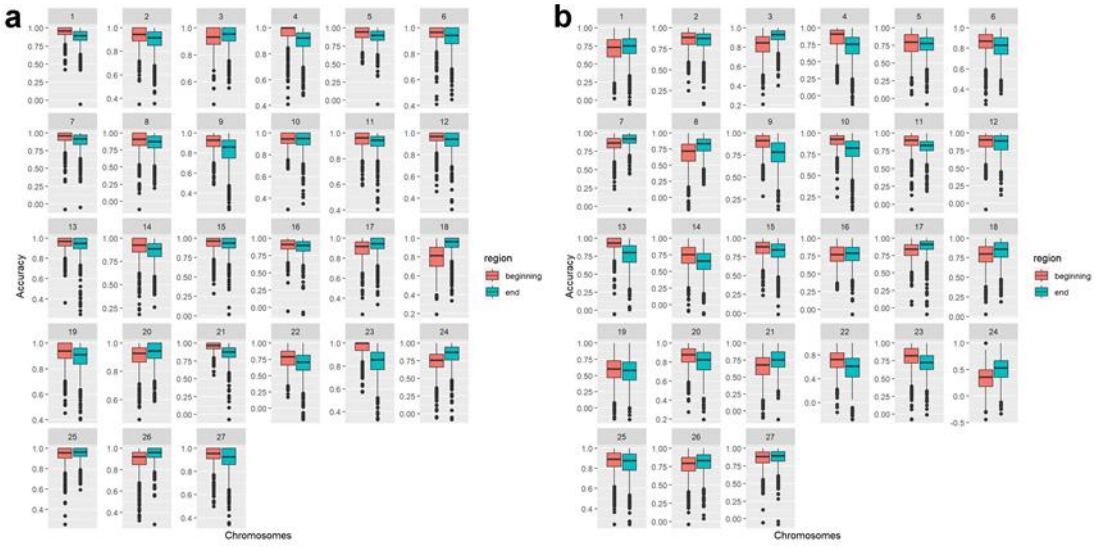


Figure S1. Mean accuracy (R^2) of imputed extremities (beginning in red and end in blue) of the 27 chromosomes of tambaqui using the 1K (a) and 0.5K (b) low-density subsets. Each chromosome extremity was analyzed by adopting 100 SNP markers.

CAPÍTULO II

Genetic parameters for host-response to acanthocephaliasis caused by the endoparasite *Neoechinorhynchus buttnerae* in the Amazon fish *Colossoma macropomum*

Genetic parameters for host-response to acanthocephaliasis caused by the endoparasite *Neoechinorhynchus buttnerae* in the Amazon fish *Colossoma macropomum*

John F. G. Agudelo[a], Baltasar F. G. Neto[a], Vito A. Mastrochirico-Filho[a], Raquel B. Ariede[a], Luis F. S. Moreira[d], Leilane da Silva Noronha[d], José M. Yáñez[b], Alexandre Honczaryk[c], Gustavo M. R. Valladão[d], 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

[c] National Institute for Amazonian Research (INPA) Manaus/AM – Brazil

[d] Nilton Lins University, Manaus/AM – Brazil.

* 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

Highlights of the manuscript:

- An experimental challenge against *Neoechinorhynchus buttnerae* was performed in tambaqui.

- Low-Moderated values for heritability were estimated for resistance and resilience against the endoparasite.
- Genetic correlations between resistance and resilience were low-moderated and negative.

Abstract

Tambaqui (*Colossoma macropomum*) is the main native fish produced by Amazon aquaculture. However, the production chain of this fish has been significantly affected by acanthocephaliasis, caused by the endoparasite *Neoechinorhynchus buttnerae*. The selection of superior genotypes for resistance and resilience to acanthocephaliasis is a fundamental approach to improve production in the context of environmental sustainability and food security. Thus, the objectives of this study were to 1) conduct an experimental challenge against the endoparasite *N. buttnerae* in tambaqui, and 2) determine the genetic parameters of disease traits caused by *N. buttnerae* in an experimental breeding population of tambaqui. Estimation of genetic parameters was performed using a genomic relationship matrix (SNPs data) obtained from an experimental challenge performed in 70 full-sib families. The experimental challenge was successfully performed in three replicate tanks, using the parasite eggs in the ichthyoplankton as source for a natural route of infection, which in turn was induced through oral filtration the plankton. The total gain in body weight of all individuals was measured during the infestation period, as a measure of Resilience (RS_F) and averaged 42 g (SD = 21.8). The average parasite abundance per animal at the end of the challenge period, indicating Resistance (RT_{CT}), was about 205.2 parasites (SD = 178.6). The heritability estimates for these traits ranged from 0.17 ± 0.04 to 0.20 ± 0.04, indicating the presence of low to moderate heritability levels for host response against *N. buttnerae* in tambaqui. Genetic correlations between resistance and resilience traits were consistently negative (favorable) and statistically significant, ranging from -0.512 ± 0.134 to -0.252 ± 0.166. These results suggest the possibility of selecting animals for both resistance and

resilience simultaneously. The significant genetic variation found for *N. buttnerae* resistance and favorable genetic correlations with resilience in tambaqui indicated that selecting superior genotypes is a viable approach to reduce the impact of acanthocephaliasis disease in aquaculture.

Keywords: heritability; disease resistance; aquaculture

1. Introduction

The tambaqui (*Colossoma macropomum*) is a Neotropical fish species which belong to the order Characiformes and the family Serrasalmidae (Calcagnotto et al., 2005). The tambaqui possesses a unique set of teeth and filtering apparatus and distinctive omnivorous feeding behavior, consuming a wide variety of natural foods such as fruits/seeds and zooplankton (Woynárovich and Anrooy, 2019). This species is widely distributed in the Amazon and Orinoco rivers, naturally occurring in countries including Brazil, Venezuela, Colombia, Peru, and Bolivia (Araujo-Lima and Goulding, 1997). Tambaqui is a prominent freshwater species in South American aquaculture and has also been successfully introduced in various Asian countries including China, Indonesia, Malaysia, Myanmar, and Vietnam (Woynárovich and Anrooy, 2019). The global production of tambaqui in aquaculture reached approximately 142 thousand tons in 2016, with Brazil accounting for 96.4% of it (i.e., around 137 thousand tons) (IBGE, 2017).

Economic losses of aquaculture attributed to diseases in Brazil alone are estimated to be around US$ 84 million (Tavares-Dias and Martins, 2017). The high stocking density and intensification in tambaqui production have led to the identification of significant pathogens affecting these fish in farming conditions, which in turn represents a potential threat to sustainable tambaqui aquaculture. The acanthocephalan *Neoechinorhynchus buttnerae* is the causative agent of acanthocephaliasis in tambaqui cultivated in different Brazilian states, such as Roraima, Amazonas, Rondônia, Pará, Amapá, and Maranhão (Valladão et al., 2020), significantly affecting a vast area of the Legal Amazon Region. The main clinical manifestation in tambaqui infected with *N. buttnerae* is a reduction in growth rate. Typically, this parasite neither cause fish mortality nor generates external clinical signs, which hinders the rapid diagnosis of acanthocephaliasis. As a result, the losses are associated with the high cost on feed, which is intended for fish that fail to achieve the expected weight gain. Once present in the fish's intestine, the parasite disrupts the integrity of the intestinal mucosa by attaching itself to the host and competes for nutrients in the lumen (Valladão et al., 2020). The diseases pathological features are caused by the parasitic attachment structure within the host, known as the proboscis, which contains spines. The acanthocephalan can affect the mucosal, submucosal, and muscular layers, as well as cause laceration of the intestinal mucosa, leading to inflammation (Jerônimo et al., 2017).

The life cycle of *N. buttnerae* is indirect, involving the presence of an invertebrate intermediate host (zooplankton) and a vertebrate definitive host, which is the tambaqui. The adult parasite presents in the fish's intestine releases fertilized eggs into the environment,

which are ingested by the intermediate host (Lourenço et al., 2018). Simultaneously, the intermediate hosts are ingested by the tambaqui, which possesses a filtering feeding structure for zooplankton. Subsequently, the developed larvae with proboscis attach to the intestinal wall of the host, mature sexually, and initiate a new cycle (Lourenço et al., 2018). Due to the parasite's cycle, acanthocephaliasis becomes a problem with uncontrolled spread in the aquatic environment and among fish farms, as any source of water containing zooplankton infested with the parasite's eggs can lead to new infections.

For the control and prevention of acanthocephaliasis, it is necessary to treat the fish with antiparasitic agents and manipulate the environment (controlling the water source and zooplankton) (Valladão et al., 2020). However, both forms of control have some drawbacks that compromise their effectiveness and safety, such as environmental pollution, drug resistance and the consequent recurrence of infections. Regarding zooplankton, there is currently no effective way (chemical or physical) to control them in a farming system, with the only alternative being to cultivate tambaqui in inland recirculation systems (Valladão et al., 2020), which is unfeasible for most developing regions in Brazil due to high energy and infrastructure costs. Due to the lack of effective treatments against acanthocephaliasis, genetic selection represents a potential strategy to reduce the negative impact of the disease in tambaqui production. Although there are currently no reports of selective breeding programs specifically focused on enhancing the host response to endoparasites in fish, the development of genetically improved populations against resistance to bacteria, virus, and ectoparasites have been demonstrated to be feasible in several aquaculture species, therefore,

representing an important alternative for controlling diseases (Kube et al., 2012; Vallejo et al., 2017; Griot et al., 2021).

Previous studies have estimated the presence of significant genetic variation for resistance to different parasites in various aquaculture species when resistance is defined either as parasite load or density (*e.g.*, Correa et al., 2017, Bassini et al., 2019). In addition, a recent study demonstrated the presence of quantitative genetic variation for resistance to the ectoparasite *Ichthyophthirius multifiliis* in tambaqui (Lira et al., 2020). These results suggest the feasibility to improve resistance to endoparasite infestation in tambaqui by utilizing selective breeding.

Growth rate is the main breeding goal in aquaculture species (Gjedrem, 2000; Gjedrem and Rye, 2018). Genetic variation for growth rate has been widely reported for different aquaculture species, including tambaqui with heritabilities ranging from 0.23 to 0.81 and expected genetic gains up to 31% (Ariede et al., 2020; Campos et al., 2020). However, there is limited knowledge on how parasite presence and burden can modify growth rate of affected fish up to date. In fact, growth performance of an animal in the face of infection can be considered a completely different trait (Doeschl-Wilson et al., 2012). According to Bishop (2012), the net impact on performance of a given level of infection and the productivity of an animal in the face of infection, can be defined as tolerance and resilience, respectively. These two traits must be considered as different to resistance (Bishop and Woolliams, 2014), which in this case can be just defined as the parasite load. To date,

there are no studies aiming at determining additive genetic variation neither for resistance, tolerance nor resilience against acanthocephaliasis in tambaqui.

In order to incorporate resistance and resilience traits against the endoparasite *N. buttnerae* into breeding programs for tambaqui, it is crucial to estimate genetic parameters to assess the level of genetic variation and the genetic relationship between these traits. Hence, the aim of this study was to estimate (co)variance components and heritability for resistance and resilience against *N. buttnerae* in tambaqui. This was achieved by utilizing data from an experimental challenge conducted in a pedigreed population from Brazil.

2. Material and Methods

2.1 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 number 1038/21) of Faculdade de Ciências Agrárias e Veterinárias, UNESP, Campus Jaboticabal, SP, Brazil.

2.2 Experimental population

The fish population analyzed in the present study was composed of 70 families, originated from two breeding groups: 35 families originated from the genetic improvement

nucleus of the Aquaculture Center of Unesp, Jaboticabal, São Paulo state, Brazil, generated by partial factorial mating involving 9 females and 11 males; and the remaining 35 families were generated in a commercial fish farm in the Amazonas state by partial factorial mating involving 16 females and 17 males.

Reproduction was induced using carp pituitary extract (Danubio LTDA) dissolved in saline solution (0.9% NaCl) and administered in two doses, 12 hours apart (first and second doses were of 0.5 and 5.5 mg/kg, respectively). For the males, a single dose equivalent to 2.5 mg/kg of carp pituitary extract was administered simultaneously with the second dose given to the females. Each female's spawn was divided for fertilization with different males, and *vice versa* (*i.e.*, sperm from males used to fertilize multiple females). The following proportion was adopted for fertilization: 5 ml of oocytes (approximately 5,000 oocytes) fertilized with 50 µl of semen in 500 ml of water, that were then placed in 20-liter experimental fiberglass incubators.

After 12 to 20 hours post-fertilization, the larvae hatched from each family were individually maintained in 20 liters conical fiberglass incubators for one week. Subsequently, approximately 200 larvae from each family were sampled and combined to create a "balanced pool," which was then transferred to a 200 m² tank. In the fry and rearing stage, 1.2 mm feed containing 40% crude protein was provided twice a day, using commercial Nutripiscis Presence feed. In the end, a pool of fry from the state of São Paulo and another from the state of Amazonas was formed, each theoretically composed of approximately 35 families.

The strategy of forming larval pools from the families was adopted due to the lack of physical infrastructure to maintain each family separately in tanks until the tagging process with *pit-tags*. One of the advantages of pooling is to avoid common environmental effects (generally called as hapa effect), which can inflate heritability estimates since the families are raised separately until pit-tagging (Vandeputte and Haffray, 2014).

After reaching a size ranging from 3-5 cm (at 5 months of age), the animals from the state of São Paulo were transported by air to the experimental facilities of the National Institute of Amazonian Research (INPA, Manaus, Brazil), free from acanthocephalan, where the fish were reared until the beginning of the experiment. The transport of the animals was carried out in plastic bags with 75% oxygen and 25% water (with 0.5% salt for prophylaxis). At the same time, the animals from the commercial fish farm in Amazonas were transported by land in 500-liter transport container supplied with oxygen and a salinity of 0.5%. The fish were kept in 5,000-liter tanks supplied with artesian well water to avoid contact with the acanthocephalan parasite until they reached an average weight of 25 g. Subsequently, the fish were marked with pit-tags (FDX-B) and weighed, referred to as the initial tagging weight - BW_0.

2.3 Experimental challenge

As for the experimental design of infection, the pool of fish from the 70 families was equally distributed across three replicates (120 m² excavated ponds) among the São Paulo and Amazon groups, following the pit-tagging process. Prior to releasing the animals, the

three ponds were previously populated with 20 parasitized tambaquis by acanthocephalans (*i.e.* trojans or disease transmitters) for 90 days to ensure the presence of parasite eggs in the infected zooplankton (Figure 1). Additionally, one month before the start of the experiment, pumps were installed to enable water exchange between the three ponds, ensuring that all replicates received a similar zooplankton input, resulting in similar fish infection levels across different ponds (experimental units). Finally, one week before releasing the pool of fish to be challenged against the acanthocephalan, the trojans were removed from the experimental environment. The ponds belong to the infrastructure of the INPA experimental fish farm, where acanthocephaliasis occurs frequently. Approximately 650 fish were allocated in each pond, representing an expected number of 10 individuals/family per replicate.

After 30 days in the ponds with parasite eggs, 5 individuals from each pond were sampled and euthanized in a benzocaine solution (300 mg/L). The intestines were collected to assess the development of infection, revealing that endoparasitism had already begun in the population, with fish infested with 2 to 157 parasites. After 60 days, all fish from the three ponds were collected for the first data sampling and to monitor the development of acanthocephaliasis. The fish were weighed (referred to as initial infection weight - BW_1) and immediately returned to their respective ponds for the continuation of the experiment. This period was chosen based on pilot projects conducted by our research group to ensure that the fish had sufficient time to be parasitized, considering that the parasite's life cycle is completed between 30 and 60 days. Therefore, a random subsample of 30 animals from each pond was

first collected and euthanized to verify if 100% of the animals were infected. The results demonstrated that all fish in the subsample were infected at BW_1, with a range of 1 to 244 parasites per fish, ensuring that the overall population (or almost all animals) had acanthocephaliasis within the 60-day period after exposure to parasite eggs. The average BW_1 was 78.27±17.78 g, with an average weight gain relative to BW_0 of 41.72±19.53 g.

After BW_1 process, the experiment continued for an additional 60 days to evaluate the weight gain or loss of each fish. At the end of this period, all animals were collected, euthanized in a benzocaine solution (300 mg/L), weighed (referred to as final infection weight - BW_2), and necropsied to collect the intestine in 70% alcohol. Then, the exact quantification of the number of parasites per animal was conducted in laboratory.

Temperature, dissolved oxygen and pH were measured with a Multiparameter Water Quality Checker U-50 (Horiba, Kyoto, Japan). Water temperature was maintained at 28°C ± 0.5, with dissolved oxygen at 5.0 mg/L ± 1.0 and pH at 7.7 ± 1.1 during the challenge period.

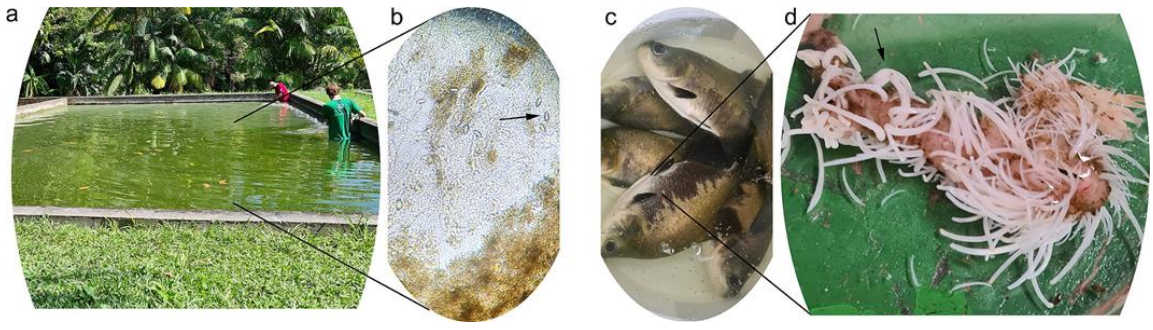


Fig. 1. (a) experimental tanks prepared for the parasite challenge experiment; (b) zooplankton infected with parasite eggs (arrow); (c) tambaqui juveniles of the experiment; (d) Intestine of

2.4 SNP genotyping

We randomly selected 1,290 animals from the challenged population to carry out single-nucleotide polymorphism (SNP) genotyping for further estimation of genetic parameters, using the genomic relationship matrix. Genomic DNA extraction was performed using the Dneasy Blood & Tissue DNA extraction kit® (Qiagen) from either blood or fin tissue samples. The DNA quality was evaluated using a NanoDrop spectrophotometer® (Thermo Fisher Scientific) and agarose gel electrophoresis. The concentration of extracted DNA (ng/µl) was measured using the Qubit fluorescence detector and the Qubit dsDNA HS Assay kit® (Invitrogen). Subsequently, the genomic DNA samples were sent to Thermo Fisher Scientific in Austin, Texas, USA, for genotyping. The AgriSeq® Targeted Genotyping by Sequencing (Agriseq tGBS) genotyping platform was used to genotype 1,246 SNPs, which were originated as subset of the dense array from Affymetrix® Axiom® SerraSNP 30K SNPs (Mastrochirico-Filho et al., 2021). The 1,246 SNPs were selected based on genomic information of 81 tambaquis from five different populations (two from North and three from South Brazil) genotyped using the SerraSNP Affymetrix® Axiom® multispecies array (29,575 SNPs of tambaqui). After a genotype quality control (QC) step (Hardy-Weinberg equilibrium with p-value < 0.05 after Bonferroni correction, call-rate of 90% and a minor allele frequency (MAF) threshold of 0.01), the linkage disequilibrium (LD) extent between SNPs was evaluated within each breeding population. A r^2 pruning threshold of 0.7 was applied to eliminate redundant combinations of SNPs. This process led to the

identification of common SNPs across all populations, ensuring the retainment of high informative SNPs. After genotyping of the challenged animals using the Agriseq tGBS, a QC was applied before downstream analysis: MAF (minor allele frequency) > 0.01, call-rate for SNPs >0.7 and call-rate for samples > 0.8. All QC filters were applied using the PLINK 1.9 software (Chang et al., 2015).

2.5 Statistical models

Resistance (RT_{CT}) to the acanthocephalan was assessed by the burden of parasites within the animals, representing the degree of infestation for all fish at the end of the experiment (120 days from the start of the challenge test). The parasitic burden was determined by directly counting the parasites in the intestines of the animals (Figure 1). Density of parasites (RT_{DP}) was also assessed by dividing the quantity of parasites (RT_{CT}) by the final weight of each fish (BW_2). Final Resilience (RS_F) to the acanthocephalan was evaluated by the body weight gain/loss in final period of evaluation, i.e., from the initial establishment of parasitism in all animals (BW_1) until the end of the experiment (BW_2), defined as $RS_F = BW_2 - BW_1$. Total Resilience (RS_T) was similarly calculated subtracting the final weight from the initial weight (prior exposure to parasites), defined as $RS_T = BW_2 - BW_0$.

All traits were analyzed using a linear mixed model to adjust the continuous variables, according to the univariate animal model below:

$$y = X\beta + Z\alpha + e$$

where y is a vector of phenotypes (RT_{CT}, RT_{DP}, RS_F and RS_T), β is a vector of fixed effects (tank and origin for all phenotypes and BW_2 as co-variate for RT_{CT} only), α is a vector of random additive polygenic genetic effects that follows a normal distribution ~ N $(0, G\sigma_\alpha^2)$, where G is the genomic relationship matrix (VanRaden, 2008) constructed using the genotyped SNPs post-QC and σ_α^2 is the additive genetic variance, X and Z are incidence matrices, $e$is the random residual error with a distribution ~ N$(0, I\sigma_e^2)$, where I is an identity matrix and σ_e^2 is the residual variance. Several models were tested using Akaike Information Criterion (AIC) comparisons with different hypotheses regarding the fixed effects to define the best model for each phenotype. The São Paulo and Amazon populations were not produced at the same time (*i.e.* they had different ages). However, all animals have the same age within each population, *i.e.*, age and origin had the same binary distribution. For this reason, fitting age and origin effects at same time would cause overparameterization of the models. The genetic and residual variances were used to estimate the heritability (h^2) of each trait defined as:

$$h^2 = \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_e^2)$$

The genetic correlations (r_g) between all traits were calculated from the covariance components estimated by using independent bivariate linear animal models as defined above, as follows (Falconer and Mackay, 1996):

$$r_g = \frac{\sigma_{a_{x,y}}}{\sqrt{\sigma^2_{a_x} * \sigma^2_{a_y}}}$$

where $\sigma_{a_{x,y}}$, is the genetic covariance between two traits (x and y), $\sigma^2_{a_x}$ and $\sigma^2_{a_y}$ are the additive genetic variance of trait x and trait y, respectively. The phenotypic correlations (r_p) were similarly calculated, but rather than genetic, phenotypic variances and co-variances between traits were used. All (co)variance parameters were estimated using the AIREMLF90 module from the BLUPF90 package (Misztal, 2002).

3. Results

Descriptive statistics for each replicate tank and origin group of the fish used in the challenge test is presented in Table 1 and Figure 2. The average BW_2 was 119.8 ± 32.1 g, with a maximum of 296 g and a minimum of 38 g. Final Resilience (RS_F) of the animals at the end of the growth assessment period (BW_2- BW_1), was on average 42 g ± 21.8 g. Fish belonging to tank 1 obtained the highest RS_F (44.8 g, ± 21.0 g), while tank 2 showed the lowest RS_F (39.8 g ± 20.8 g) (Table 1 and Figure 2). Regarding the origin, animals from São Paulo demonstrated a slightly higher value of RS_F (43.1 g ± 23.0 g) in relation to those from Amazonas (40.4 g ± 20.0 g) (Table 1). Maximum RS_F reported for an individual was 166 g, while some animals demonstrated negative values of RS_F, such as -33 g.

Table 1. Descriptive statistics of traits evaluated in a pedigreed population of tambaqui challenged to the endoparasite *Neoechinorhynchus buttnerae.*

Trait		Tank			Origin		Overall Mean
		1	2	3	SP	AM	
RS_F	Mean	44.8	39.8	41.6	43.1	40.4	42
	SD	21	20.8	23.3	23	20	21.8
	Min	-33	-7	-7	-33	-3	-33
	Max	151	166	154	166	151	166
	CV	46.7	52.3	56.1	53.2	49.5	52
RS_T	Mean	85.5	80.9	74.8	93.5	61.3	80.3
	SD	33.3	31.2	33.5	30.9	25.8	33
	Min	-80	8	-18	22	-80	-80
	Max	202	241	249	249	179	249
	CV	38.9	38.5	44.8	33.1	42.2	41.1
RT_{CT}	Mean	335.1	156.5	108.9	218.9	185.4	205.2
	SD	214.7	100.1	94.8	197	146	178.6
	Min	31	9	8	8	8	8
	Max	1570	746	847	1570	1016	1570
	CV	64.1	64	87.1	90	78.8	87.1
RT_{DP}	Mean	2.6	1.2	0.8	1.6	1.3	1.5
	SD	1.8	0.9	1.0	1.6	1.2	1.5
	Min	0.0	0.0	0.0	0.0	0.0	0.0
	Max	12.5	5.2	12.0	12.5	7.7	12.5
	CV	69.2	69.4	123.4	97.4	92.8	97.0
Total samples		703	699	733	1,240	895	2,135

RS_F: Final Resilience**; RS_T:** Total Resilience; **RT_{CT}:** Resistance by parasite count; **RT_{DP}:** Resistance by parasite density; **Mean**: average values; **SD**: Standard deviation; **Min**: Minimum values; **Max**: Maximum values; **CV**: coefficient of variation; **SP**: São Paulo; **AM**: Amazonas.

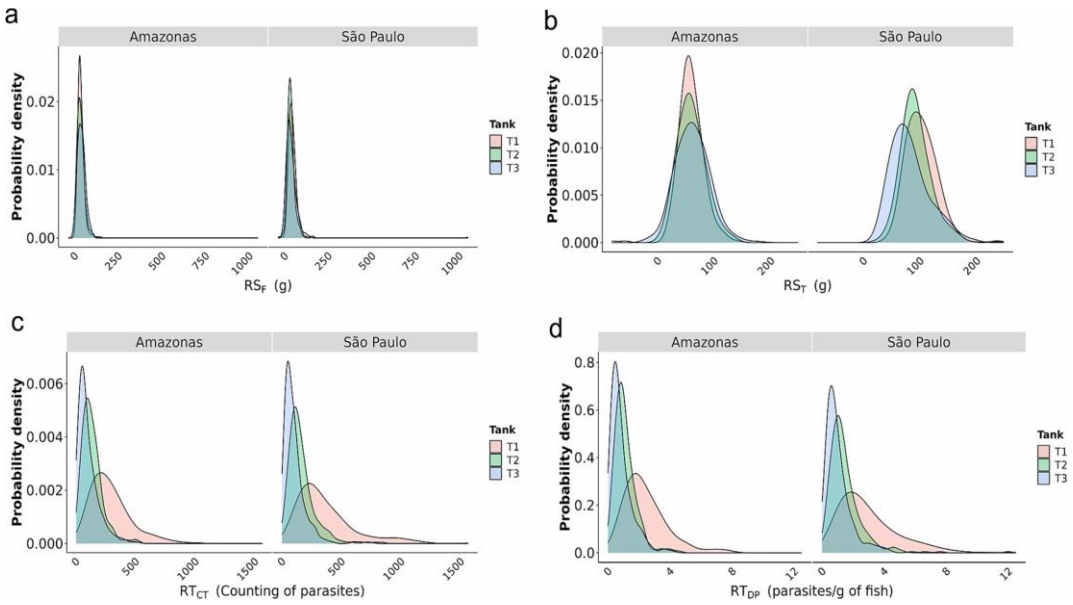


Fig. 2. Phenotype density distribution according to the origin (Amazonas: left and S˜ao Paulo: right) and tank (T1: red, T2: green, T3: blue). (a) RS_F: Final Resilience; (b) RS_T: Total Resilience; (c) RT_{CT}: Resistance by parasite count; (d) RT_{DP}: Resistance by parasite density. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The average abundance of parasites per animal at the end of the challenge period (RT_{CT}), which measures resistance, was 205.2 parasites ± 178.6. Among the experimental replicates (Table 1 and Figure 2), tank 1 had the highest average number of parasites (335.1 ± 214.7), while tank 3 exhibited the lowest abundance (108.9 ± 94.8). Regarding their origin, fish from São Paulo showed a higher average number of parasites (218.9 ± 197.0) compared to those from Amazonas (185.4 ± 146.0) (Table 1). The maximum reported parasite abundance for an individual was 1,570, and the minimum recorded was 8. No mortality was observed during the parasitic experiment period.

Significant additive-genetic variation was observed for all traits related to *N. buttnerae* resistance and resilience in tambaqui. Estimated heritabilities are presented in Table 2. The results showed low to moderate heritability values, which varied between 0.170 ± 0.040 and 0.203 ± 0.040 for RS_F and RS_T.

Table 2. Estimation of variance components and heritabilities for traits evaluated in tambaqui challenged to the endoparasite *Neoechinorhynchus buttnerae.*

Trait	σ_α^2 ± SE	σ_e^2 ± SE	h^2
RS_F	69.1 ± 17.9	332.2 ± 17.3	0.171 ± 0.040
RS_T	153.9 ± 35.8	597.7 ± 31.8	0.203 ± 0.040
RT_{CT}	4614.6 ± 1204.8	20022 ± 1077	0.185 ± 0.040
RT_{DP}	4877.9 ± 1253.2	19971 ± 1088.5	0.194 ± 0.045

RS_F: Final Resilience; **RS_T:** Total Resilience; **RT_{CT}:** Resistance by parasite count; **RT_{DP}:** Resistance by parasite density; σ_α^2: genetic variance; σ_e^2: residual variance; h^2: heritability.

The phenotypic correlations between the two trait definitions of Resistance (RT_{CT} and RT_{DP}) and Resilience (RS_F and RS_T) were low (Table 3), with values ranging from 0.210 to -0.080 for the correlation between RT_{CT} and RS_F, and RT_{DP} and RS_F, respectively. All the genetic correlations were negative (favorable) and significant (Table 3), which reveals the feasibility of selecting animals resilient, but also resistant. The genetic correlations were low to moderate and varied between -0.491 ± 0.138 for RT_{DP}/RS_T and -0.252 ± 0.166 for RT_{CT}/RS_T.

Table 3. Phenotypic and genetic correlations for traits evaluated in tambaqui challenged to the endoparasite *Neoechinorhynchus buttnerae.*

Traits	Phenotypic correlation		Genetic correlation	
	RT_{CT}	RT_{DP}	RT_{CT}	RT_{DP}
RS_F	0.150	-0.080	-0.304 (0.169)	-0.484(0.149)
RS_T	0.210	0.010	-0.252(0.166)	-0.491(0.138)

RS_F: Final Resilience**; RS_T:** Total Resilience; **RT_{CT}:** Resistance by parasite count; **RT_{DP}:** Resistance by parasite density.

4. Discussion

In the past, evaluating disease resistance was deemed a challenging task, especially when conducted under field conditions, owing to the multitude of environmental factors capable of impacting fish mortality. Currently, the genetic selection for disease resistance in fish predominantly relies upon experimental challenge testing, wherein animals are subjected to controlled pathogen exposure. This is often accomplished through either cohabitation with infected fish or direct pathogen inoculation, as demonstrated by Ødegård et al. (2011). In terms of parasite challenge, several studies have been conducted for ectoparasites, particularly for sea lice copepods in salmon, such as *Lepeophtheirus salmonis* in Europe and *Caligus rogercresseyi* in Chile (Tsai et al., 2020; Bassini et al., 2019). Other rare examples are the study of genetic parameters estimated against the trematode parasite *Uvulifer* sp. in tilapia under field infection condition (Charo-Karisa et al., 2021); and in tambaqui, where a cohabitation challenge test against the protozoan *Ichthyophthirius multifiliis* was recently developed by temperature fluctuations to assess heritability estimates (Lira et al., 2020). However, endoparasite controlled challenge has not been reported in fish, and our

experimental protocol allowed for the first time the evaluation of quantitative genetic parameters of resistance and resilience against *N. buttnerae* in fish.

In general, while utilizing cohabitation challenge could potentially provide a more accurate simulation of the immune response seen during natural infections, a majority of challenges concerning bacterial and viral infections in fish have predominantly employed intraperitoneal injection of the pathogen directly into the host (Gjedrem, 2000; Ødegård et al., 2011; Yáñez et al., 2014b; Mastrochirico-Filho, et al., 2019; Srisapoome et al., 2019). In the context of parasites, challenge trials have mainly been conducted through cohabitation (Xu et al., 2007; Taylor et al., 2009; Salte et al., 2010; Gjerde et al., 2011; Glover et al., 2017; Lira et al., 2020). Here, we have described a successful challenge using infection by the presence of parasite eggs in the infected zooplankton, reproducing the natural route of infection where tambaqui filters the plankton using the specific feeding structure. Under the conditions of the current experimental challenge, the fish exhibited endoparasites on their intestine by 30 days after the beginning of the test, with a substantial parasitic burden (all animals infected, except one free of parasites) by 60 days. These observations are consistent with the known life cycle of the *N. buttnerae*, characterized by a life cycle completed between 30 and 60 days (Lourenço et al., 2018).

A majority of the investigations focusing on the evaluation of genetic parameters associated with disease resistance have concentrated on bacterial and viral pathogens (Gjedrem, 2000; Ødegård et al., 2011; Yáñez et al., 2014b; Elaswad and Dunham, 2018; Gjedrem and Rye, 2018). Regarding parasites, there are limited available studies, and they

are primarily centered around three parasites that impact salmonid species, especially the Atlantic Salmon *Salmo salar*: salmon louse (*Lepeophtheirus salmonis*) (Glover et al., 2005; Kolstad et al., 2005; Gjerde et al., 2011; Rochus et al., 2018), sea lice (*Caligus rogercressegi*) (Lhorente et al., 2012; Tsai et al., 2016; Correa et al., 2017a; 2017b; Bassini et al., 2019), and amoebic gill disease (*Neoparamoeba perurans*) (Taylor et al., 2009; Kube et al., 2012; Boison et al., 2019; Lillehammer et al., 2019). As a result, the central outcome of this investigation was the assessment of genetic parameters for evaluating resistance and resilience against an endoparasite in tambaqui, which can serve as a framework for developing genetically resistant fish against endoparasites. Some initial efforts have been addressed to evaluate genetic parameters for disease traits in tambaqui, particularly resistance against the bacteria *A. hydrophila* (Ariede et al., 2020) and the ectoparasite *Ichthyophthirius multifiliis* (Lira et al., 2020), which revealed moderated-high heritabilities and, consequently, that resistance disease traits can be included into the breeding goal of this species.

In general, studies conducted to assess genetic parameters in resistance against parasites rely on quantifying parasite levels. These studies have reported moderate to high heritability estimates for this trait, ranging from 0.28 to 0.48 (Taylor et al., 2009; Gjerde et al., 2011; Kube et al., 2012; Lhorente et al., 2012; Yáñez et al., 2014a; Tsai et al., 2016; Boison et al., 2019). In the present study, the estimation of heritability against the endoparasite acanthocephalan demonstrated low to moderate and significant values for different phenotypes/traits, denominated as resistance and resilience. Animals can present different defense mechanisms to protect themselves from the negative effects of parasitic

infections (Klemme et al., 2020). One of these defense mechanisms refers to the term "resistance", which encompasses mechanisms that reduce infection load or parasite growth after the establishment of the disease, with both nonspecific and specific immune responses playing a crucial role (Klemme and Karvonen, 2017). Another natural defense mechanism pertains to the host's "tolerance", which is linked to the animal's capacity to diminish the adverse effects of infection by limiting and repairing the damages inflicted by parasites (Bishop, 2012; Klemme and Karvonen, 2017). Some authors have used the term "resilience" similarly to tolerance (Doeschl-Wilson et al., 2012), although there are controversies about the use of the term resilience, which generally can be attributed to animals' adaptability in various environmental circumstances (Bishop & Morris, 2007). The significant genetic variation found in this study indicates the viability of genetically improving resistance and resilience against the endoparasite in tambaqui through genetic selection. Furthermore, as all the genetic correlations exhibited both negative and significant values, it becomes viable to select animals that possess not only resistance but also a combination of resilience traits. Further research remains necessary to assess the extent and orientation of the genetic correlation or re-ranking of estimated breeding values (EBVs) for growth-related traits (resilience) across the juvenile and harvest phases within this specific breeding population.

To conclude, the evident substantial genetic diversity concerning resistance and resilience against the endoparasite *N. buttnerae* within the examined tambaqui breeding population suggests the potential for genetic improvement of these traits. Hence, this approach presents a viable solution to combat the disease within aquaculture systems,

particularly for the aquaculture of this native Amazonian fish, predominantly sustained by small subsistence producers.

Acknowledgment

We would like to thank all the members of the Laboratory of Genetics in Aquaculture and Conservation (LaGeAC) and the Laboratory of Parasitology and Pathology of Aquatic Organisms at Nilton Lins University, for all the support and assistance given during the management, collections and counting of parasites.

Author contributions

Conceptualization: JFGA, GMRV, and DTH; Formal analysis: JFGA, VAMF, BFGN; Funding acquisition: JFGA; GMRV, and DTH; Investigation: JFGA, LFSM, LSN, VAMF, BFGN, RBA, JMY; Methodology: JFGA, LFSM, LSN, AH, GMRV; Project administration: DTH; Supervision: DTH; Roles/Writing – original draft: JFGA and DTH; Writing – review & editing: JFGA, VAMF, BFGN, RBA and DTH.

Funding

This work was supported by Amazonas Research Foundation (FAPEAM grant 20/07959-5), National Council for Scientific and Technological Development (CNPq grant 312250/2021-5), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil

(CAPES - Finance Code 001); PROCAD/AMAZONIA 2018 (grant number 88881.200614/2018-01).

Data availability statement

The raw data of the experiments used in this study is available upon request to the corresponding author.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

5. References

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

Ariede, R. B., Freitas, M. V., Agudelo, J. F., Borges, C. H., Lira, L. V., Yoshida, G. M., & Hashimoto, D. T. 2020. Genetic (co) variation between resistance to *Aeromonas hydrophila* and growth in tambaqui (*Colossoma macropomum*). Aquaculture, 523, 735225.

Bassini, L. N., Lhorente, J. P., Oyarzún, M., Bangera, R., Yáñez, J. M., Neira, R. 2019. Genetic parameters for *Piscirickettsia salmonis* resistance, sea lice (*Caligus rogercresseyi*) susceptibility and harvest weight in rainbow trout (*Oncorhynchus*

mykiss). *Aquaculture*, *510*,276-282. https://doi.org/10.1016/j.aquaculture.2019.05.008.

Bishop, S. C. 2012. A consideration of resistance and tolerance for ruminant nematode infections. In Frontiers in Genetics (Vol. 3, Issue DEC). https://doi.org/10.3389/fgene.2012.00168

Bishop, S. C., & Morris, C. A. 2007. Genetics of disease resistance in sheep and goats. Small ruminant research, 70(1), 48-59.

Bishop, S. C., & Woolliams, J. A. 2014. Genomics and disease resistance studies in livestock. Livestock Science, 166(1), 190–198. https://doi.org/10.1016/j.livsci.2014.04.034

Boison, S.A., Gjerde, B., Hillestad, B., Makvandi-Nejad, S., Moghadam, H.K. 2019. Genomic and transcriptomic analysis of amoebic gill disease resistance in Atlantic salmon (*Salmo salar* L.). Front. Genet 10. https://doi.org/10.3389/fgene.2019.00068.

Campos, E. C., Oliveira, C. A. L., Araújo, F. C. T., Todesco, H., Souza, F. N., Rossi, R. M., et al. 2020. Genetic parameters and response to selection for growth in tambaqui. animal, 14(9), 1777-1785.

Calcagnotto, D., Schaefer, S. A., DeSalle, R. 2005. Relationships among characiform fishes inferred from analysis of nuclear and mitochondrial gene sequences. Molecular phylogenetics and evolution, 36(1), 135-153. https://doi.org/10.1016/j.ympev.2005.01.004.

Charo-Karisa, H., Ali, S. E., Marijani, E., Ibrahim, N. A., Trinh, T. Q., Chadag, M. V., & Benzie, J. A. 2021. Genetic parameters for black spot disease (diplopstomiasis) caused by Uvulifer sp. infection in Nile tilapia (*Oreochromis niloticus* L.). Aquaculture, 532, 736039.

Correa, K., Bangera, R., Figueroa, R., Lhorente, J.P., Yáñez, J.M. 2017b. The use of genomic information increases the accuracy of breeding value predictions for sea louse (*Caligus rogercresseyi*) resistance in Atlantic salmon (*Salmo salar*). Genet. Sel. Evol. 49, 1–5. https://doi.org/10.1186/s12711-017-0291-8.

Correa, K., Lhorente, J.P., Bassini, L., López, M.E., Di Genova, A., Maass, A., Davidson, W.S., Yáñez, J.M. 2017a. Genome wide association study for resistance to *Caligus rogercresseyi* in Atlantic salmon (*Salmo salar* L.) using a 50K SNP genotyping array. Aquaculture 472, 61–65. https://doi.org/10.1016/j.aquaculture.2016.04.008.

Doeschl-Wilson, A. B., & Kyriazakis, I. 2012. Should we aim for genetic improvement in host resistance or tolerance to infectious pathogens? Frontiers in genetics, 3, 272.

Elaswad, A., Dunham, R., 2018. Disease reduction in aquaculture with genetic and genomic technology: current and future approaches. Rev. Aquaculture. 10, 876–898.https://doi.org/10.1111/raq.12205.

Falconer, D. S., and Mackay, T. F. C. 1996. Introduction to quantitative genetics. 1996. Harlow, Essex, UK: Longmans Green, 3.

Gjedrem, T. 2000. Genetic improvement of cold-water fish species. Aquaculture research, 31(1), 25-33.

Gjedrem, T., & Rye, M. 2018. Selection response in fish and shellfish: a review. Reviews in Aquaculture, 10(1), 168-179.

Gjerde, B., Ødegård, J., Thorland, I., 2011. Estimates of genetic variation in the susceptibility of Atlantic salmon (*Salmo salar*) to the salmon louse *Lepeophtheirus salmonis*. Aquaculture 314, 66–72. https://doi.org/10.1016/j.aquaculture.2011.01.026.

Glover, K. A., Aasmundstad, T., Nilsen, F., Storset, A., & Skaala, Ø. 2005. Variation of Atlantic salmon families (*Salmo salar* L.) in susceptibility to the sea lice *Lepeophtheirus salmonis* and *Caligus elongatus*. Aquaculture, 245(1-4), 19-30. https://doi.org/10.1016/j.aquaculture.2004.11.047.

Glover, K.A., Hansen, T., Besnier, F., Solberg, M.F., Fjelldal, P.G., Eide, A.G.S., Dalvin, S., Nilsen, F. 2017. Cloned and outbred Atlantic salmon display equal parasite dispersion when infected with the salmon louse. Aquaculture 480, 83–88. https://doi.org/10.1016/j.aquaculture.2017.08.008.

Griot, R., Allal, F., Phocas, F., Brard-Fudulea, S., Morvezen, R., Bestin, A., & Vandeputte, M. 2021. Genome-wide association studies for resistance to viral nervous necrosis in

three populations of European sea bass (*Dicentrarchus labrax*) using a novel 57k SNP array DlabChip. Aquaculture, 530, 735930.

IBGE. 2017. Produção da Pecuária Municipal 2015. Rio de Janeiro, 41, 1-108.

Jerônimo, G.T., de Pádua, S.B., de Andrade Belo, M.A., Chagas, E. C., Taboga, S.R., Maciel, P.O., & Martins, M.L. 2017. *Neoechinorhynchus buttnerae* (Acanthocephala) infection in farmed *Colossoma macropomum*: a pathological approach. Aquaculture, 469, 124-127.

Klemme, I., & Karvonen, A. 2017. Vertebrate defense against parasites: Interactions between avoidance, resistance, and tolerance. Ecology and Evolution, 7(2), 561-571.

Klemme, I., Hyvärinen, P., & Karvonen, A. 2020. Negative associations between parasite avoidance, resistance and tolerance predict host health in salmonid fish populations. Proceedings of the Royal Society B, 287(1925), 20200388.

Kolstad, K., Heuch, P. A., Gjerde, B., Gjedrem, T., & Salte, R. 2005. Genetic variation in resistance of Atlantic salmon (*Salmo salar*) to the salmon *louse Lepeophtheirus salmonis*. Aquaculture, 247(1-4), 145-151. https://doi.org/10.1016/j.aquaculture.2005.02.009.

Kube, P.D., Taylor, R.S., Elliott, N.G. 2012. Genetic variation in parasite resistance of Atlantic salmon to amoebic gill disease over multiple infections. Aquaculture 364–365, 165–172. https://doi.org/10.1016/j.aquaculture.2012.08.026.

Lhorente, J.P., Gallardo, J.A., Villanueva, B., Araya, A.M., Torrealba, D.A., Toledo, X.E., Neira, R. 2012. Quantitative genetic basis for resistance to *Caligus rogercresseyi* sea lice in a breeding population of Atlantic salmon (*Salmo salar*). Aquaculture 324–325, 55–59. https://doi.org/10.1016/j.aquaculture.2011.10.046.

Lillehammer, M., Boison, S.A., Norris, A., Løvoll, M., Bakke, H., Gjerde, B. 2019. Genetic parameters of resistance to amoebic gill disease in two Norwegian Atlantic salmon populations. Aquaculture 508, 83–89. https://doi.org/10.1016/j.aquaculture.2019.04.072.

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

Lourenço, F. D. S., Morey, G. A. M., & Malta, J. C. D. O. 2018. The development of Neoechinorhynchus buttnerae (Eoacanthocephala: Neoechinorhynchidae) in its intermediate host Cypridopsis vidua in Brazil. Acta Parasitologica, 63(2), 354-359.

Mastrochirico-Filho, V. A., Ariede, R. B., Freitas, M. V., Lira, L. V., Agudelo, J. F., Pilarski, F., Neto, R. V. R., 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.

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. 28, No. 07, p. 743).

Ødegård, J., Baranski, M., Gjerde, B., & Gjedrem, T. 2011. Methodology for genetic evaluation of disease resistance in aquaculture species: challenges and future prospects. Aquaculture research, 42, 103-114.

Rochus, C.M., Holborn, M.K., Ang, K.P., Elliott, J.A.K., Glebe, B.D., Leadbeater, S., Tosh, J.J., Boulding, E.G. 2018. Genome-wide association analysis of salmon lice (*Lepeophtheirus salmonis*) resistance in a North American Atlantic salmon population. Aquac. Res. 49, 1329–1338. https://doi.org/10.1111/are.13592.

Salte, R., Bentsen, H.B., Moen, T., Tripathy, S., Bakke, T.A., Ødegård, J., Omholt, S., Hansen, L.P. 2010. Prospects for a genetic management strategy to control *Gyrodactylus salaris* infection in wild Atlantic salmon (*Salmo salar*) stocks. Can. J.Fish. Aquat. Sci. 67, 121–129. https://doi.org/10.1139/f09-168.

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 Immun. 92 (February), 209–215. https://doi.org/10.1016/j.fsi.2019.05.060.

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, 913-918.

Taylor, R.S., Kube, P.D., Muller, W.J., Elliott, N.G. 2009. Genetic variation of gross gill pathology and survival of Atlantic salmon (*Salmo salar* L.) during natural amoebic gill disease challenge. Aquaculture 294, 172–179. https://doi.org/10.1016/j.aquaculture.2009.06.007.

Tsai, H.Y., Hamilton, A., Tinch, A.E., Guy, D.R., Bron, J.E., Taggart, J.B., Gharbi, K., Stear, M., Matika, O., Pong-Wong, R., Bishop, S.C., Houston, R.D. 2016. Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. Genet. Sel. Evol. 48, 1–11. https://doi.org/10.1186/s12711-016-0226-9.

Vallejo, R. L., Leeds, T. D., Gao, G., Parsons, J. E., Martin, K. E., Evenhuis, J. P., & 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.

Valladão, G. M. R., Gallani, S. U., Jerônimo, G. T., & Seixas, A. T. de. 2020. Challenges in the control of acanthocephaliasis in aquaculture: special emphasis on *Neoechinorhynchus buttnerae*. In Reviews in Aquaculture, 12, 1360–1372. https://doi.org/10.1111/raq.12386

Vandeputte, M., & Haffray, P. 2014. Parentage assignment with genomic markers: a major advance for understanding and exploiting genetic variation of quantitative traits in farmed aquatic animals. Frontiers in genetics, 5, 432.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. Journal of dairy science, 91(11), 4414-4423.

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.

Xu, D.H., Klesius, P.H., Shoemaker, C.A., 2007. Evaluation of a cohabitation challenge model in immunization trials for channel catfish *Ictalurus punctatus* against *Ichthyophthirius multifiliis*. Dis. Aquat. Organ. 74, 49–55. https://doi.org/10.3354/dao074049.

Yáñez, J.M., Houston, R.D., Newman, S. 2014b. Genetics and genomics of disease resistance in salmonid species. Front. Genet. 5, 1–13. https://doi.org/10.3389/fgene.2014.00415.

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.

CAPÍTULO III

Genomic predictions for *Neoechinorhynchus buttnerae* resistance in the Amazonian Tambaqui fish (*Colossoma macropomum*)

Genomic predictions for *Neoechinorhynchus buttnerae* resistance in the Amazonian Tambaqui fish (*Colossoma macropomum*)

John F. G. Agudelo[a], Baltasar F. G. Neto[a], Raquel B. Ariede[a], Vito A. Mastrochirico-Filho[a, b], José M. Yáñez[c], Gustavo M. R. Valladão[d], Diogo T. Hashimoto[a]*

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

[b] São Paulo State University (Unesp), School of Sciences, 17033-360 Bauru, SP, Brazil

[c] Faculty of Veterinary and Livestock Sciences, University of Chile, Santiago, Chile

[d] Nilton Lins University, Manaus, AM, Brazil

Abstract

Infections caused by the endoparasite *Neoechinorhynchus buttnerae* have resulted in substantial economic losses in tambaqui (*Colossoma macropomum*) production. An alternative to enhance resistance to *N. buttnerae* infection is the adoption of genomic selection programs, which can promote sustainable development in Brazilian aquaculture. Additionally, the use of low-density SNP panels, optimized through genotype imputation, emerges as a cost-effective tool for implementing genomic selection on a commercial scale. The objectives of this study were: 1) to estimate genetic parameters for resistance (RTct) and resilience (RSf and RSt) to the endoparasite *Neoechinorhynchus buttnerae* in tambaqui, and 2) to assess the accuracy of genomic prediction for these traits using genomic data, employing

a low-density SNP array (1K), and 12K imputed genotypes; then compare the accuracy of genomic prediction with that of the traditional pedigree-based method. Overall, there was an observed increase of up to 70% in the accuracy of genomic prediction for total resilience (RSt) with the utilization of genomic information, indicating an expected advantage in genetic gain from this data. Additionally, the findings suggest that imputation serves as an effective tool for genomic prediction, potentially yielding gains of up to 19% in resistance characteristics and up to 48% in resilience compared to the traditional pedigree-based method. In conclusion, the 1K array demonstrated sufficient performance for generating reliable estimations of genomic breeding values, particularly in terms of accuracy. These results provide valuable insights for designing genotyping strategies geared towards cost reduction in the practical application of genomic selection for *N. buttnerae* resistance.

Keywords: Genotyping; Breeding Programs; Genetic Improvement; Acanthocephalan

1. Introduction

Acanthocephalosis disease is caused by *Neoechinorhynchus buttnerae*, an obligate endoparasite (metazoan) whose host is the tambaqui (*Colossoma macropomum*) (Castro et al., 2020a). Strong infestations by *N. buttnerae* have been reported in the North, Northeast, and Central-West regions of Brazil (Aguiar et al., 2018; Valladão et al., 2020; Maciel-Honda et al., 2023). The main problem associated with this parasitosis is the significant economic loss in tambaqui production, estimated at approximately US$7,000 per hectare in tanks infected by acanthocephalans (Silva-Gomes et al., 2017). These losses are primarily due to the continuous use of feed that does not translate into weight gain because *N. buttnerae* absorbs nutrients directly from fish food through its integument (Valladão et al., 2020). As a

consequence of the infectious process, nutrient absorption is reduced, and depending on the parasite intensity, it can cause deformities, altered growth, weight loss, cachexia, and heterogeneous Stokes (Jerônimo et al., 2017; Sebastião et al., 2021) (Castro et al., 2020b).

Different challenges associated with the control of acanthocephalosis in fish farms can be highlighted, including the indiscriminate use of antiparasitic agents, environmental pollution, and drug resistance. Hence, genetic selection emerges as a promising strategy to mitigate the negative impact of the disease on tambaqui production, as demonstrated by Agudelo et al. (2024), where significant genetic variation was found for resistance to *N. buttnerae* and favorable genetic correlations with resilience in tambaqui, indicating that the selection of superior genotypes is a viable approach to reduce the impact of acanthocephalosis disease on tambaqui production.

Genetic selection and breeding programs for most neotropical species are rare, resulting in insufficient enhancement of crucial traits such as disease resistance. Consequently, there is an urgent call for genetic studies and the integration of genomic tools to fortify these traits, offering a powerful solution for controlling outbreaks and mitigating the spread of pathogens. (Ariede et al., 2022) Similar to its successful implementation in aquaculture species like Atlantic salmon, rainbow trout, coho salmon, common carp, channel catfish, and Pacific oyster (Barria et al., 2018; Garcia et al., 2018; Gutierrez et al., 2018; Palaiokostas et al., 2018; Vallejo et al., 2018; Yoshida et al., 2018a; Bangera et al., 2017; Correa et al., 2017; Ødegård et al., 2014; Tsai et al., 2016). Continuously enhancing genetic

practices and leveraging genomic tools are essential for optimizing desirable traits, thereby safeguarding the health and productivity of aquatic species.

In recent years, genetic studies have been carried out to select individuals resistant to diseases caused by bacteria and parasites in neotropical species, specifically pacu (*Piaractus mesopotamicus*) and tambaqui (*Colossoma macropomum*). (Ariede et al., 2018; Mastrochirico-Filho et al., 2019; Ariede et al., 2020; Lira et al., 2020,2023; Agudelo et al., 2024; Manso et al., 2024).

Genomic selection (GS) has been widely applied to aquaculture species to accelerate genetic gain in selection and genetic improvement programs. This approach has proven to be highly effective in achieving significant improvements in breeding programs. (Zenger *et al.*, 2019). This tool, based on the prediction of genomic values (GEBV) of markers present in the genome, enables the estimation of traits that are difficult to measure, such as disease resistance, traits limited by sex, or those influenced by phenotype-environment interactions. (Yang et al., 2024; Yáñez et al., 2023)

In general, genomic prediction (GEBV) has proven to be more efficient than the traditional pedigree-based method (EBV). This technique utilizes genomic information to estimate the genetic value of individuals, thereby enabling more precise selection in genetic improvement programs (Ajasa et al., 2024). As observed in various studies on disease resistance, genomic prediction has consistently yielded higher accuracy values compared to pedigree-based methods. (Aslam et al(Vallejo et al., 2024); Manso et al., 2024)

One challenge of genomic selection lies in the accessibility of genotyping technologies, which are typically associated with high costs, thus, limiting their routine use in production. In this sense, the development of economically viable genotyping strategies is crucial for the aquaculture industry (Kriaridou et al., 2023). One strategy to optimize genomic selection (GS) involves using low-density SNP panels (1,000 - 6,000 SNPs), which can yield similar prediction accuracies as medium/high-density panels. Even panels with as few as 200 to 3,000 SNPs outperform pedigree-based approaches in selection accuracy (Prchal et al., 2023). In this regard, various studies have shown that employing lower densities in SNP panels is an effective strategy. Finding a cost-benefit balance between SNP panel density and prediction accuracies can enhance the feasibility of genomic selection (Qu et al., 2024;Dufflocq et al., 2019; Prchal et al., 2023;Kriaridou et al., 2020; Yoshida et al., 2018).

The tambaqui (*Colossoma macropomum*) is the most significant native species in aquaculture in South America. Nevertheless, high-density conditions and climate fluctuations increase susceptibility to diseases, leading to economic losses in production (Agudelo et al 2022). However, over the past three years, genomic studies have been conducted for this species, serving as the foundation and reference for further research. As examples of these studies, we can mention: genome assemblying of *C. macropomum* (Hilsdorf et al., 2021), and the design and development of a multispecies SNP array, featuring 30k SNPs specific to tambaqui, (Mastrochirico-Filho et al., 2021). The mapping of quantitative trait loci (QTL) associated to resistance to *I. multifiliis* in tambaqui using genome-wide association study (GWAS) is also noteworthy (Lira et al., 2023). Additionally,

a linkage map was also constructed for genome-wide anchoring at the chromosomal level, along with a genome-wide association study for resistance to *Aeromonas hydrophila* in *C. macropomum* (Ariede et al., 2022). Consequently, the application and advancement of new genomic tools will aid in selecting economically important traits and enhancing the genetic improvement of the species.

The objectives of this study were: 1) to estimate genetic and genomic parameters for resistance (RTct) and resilience (RSf and RSt) to the endoparasite *Neoechinorhynchus buttnerae* in Tambaqui, and 2) to assess the accuracy of genomic prediction for these traits using genomic data, employing a low-density SNP array (1K), and 12K imputed genotypes; then compare the accuracy of genomic prediction with that of the traditional pedigree-based method.

2. Material and methods

2.1. Ethics statement

This study was conducted in strict accordance with the recommendations of the National Council for control Animal Experimentation (CONCEA) (Brazilian Ministry for Science, Technology and Innovation) and was approved by the Ethics Committee on Animal Use (CEUA number 1038/21) of Faculdade de Ciências Agrárias e Veterinárias, UNESP, Campus Jaboticabal, SP, Brazil.

2.2.Establishment of families, disease resistance challenge and genotyping

The fish population analyzed in the present study was composed of 70 families, originated from two breeding groups: 35 families originated from the genetic improvement nucleus of the Aquaculture Center of Unesp, Jaboticabal, São Paulo state, Brazil, generated by partial factorial mating involving 9 females and 11 males; and the remaining 35 families were generated in a commercial fish farm in the Amazonas state by partial factorial mating involving 16 females and 17 males. Further details regarding family production, the process for establishing the experimental population, as well as the experimental challenge and acquisition of phenotypic data, are available for review (Agudelo et al., 2024).

We randomly selected 1,290 animals from the challenged population to carry out single-nucleotide polymorphism (SNP) genotyping for further estimation of genetic parameters and genomic prediction. Genomic DNA extraction was performed using the Dneasy Blood & Tissue DNA extraction kit® (Qiagen) from either blood or fin tissue samples. The DNA quality was evaluated using a NanoDrop spectrophotometer® (Thermo Fisher Scientific) and agarose gel electrophoresis. The concentration of extracted DNA (ng/µl) was measured using the Qubit fluorescence detector and the Qubit dsDNA HS Assay kit® (Invitrogen). Subsequently, the genomic DNA samples were sent to Thermo Fisher Scientific in Austin, Texas, USA, for genotyping. The AgriSeq® Targeted Genotyping by Sequencing (Agriseq tGBS) genotyping platform was used to genotype 1,246 SNPs, which were originated as subset from the medium dense SNP array from Affymetrix® Axiom®

(30K SerraSNP array) (Mastrochirico-Filho et al., 2021). After genotyping, the data was subjected to a quality control (QC) check using PLINK 1.9 software (Chang et al., 2015), using the parameters: MAF (minor allele frequency) > 0.01, call-rate for SNPs >0.7 and call-rate for samples > 0.8.

2.3.Genotype imputation

Genotype imputation was performed using a dataset consisting of 1290 challenged animals, genotyped with 1K SNPs from the AgriSeq® tGBS platform. The genomic data of 717 animals, originating from various tambaqui populations and genotyped with 15K SNPs from the Axiom Serra SNP® matrix, served as a reference for imputation. Notably, the reference population shares some relationship with the challenged population, as it includes some parents (n = 53) from the reference group.

The assessment of imputation accuracy was designed based on a cross-validation scheme consisting of 10 replicates. In each replicate, a random subset of animals was designated using pipelines implemented in the R software (R Core Team, 2018). In this setup, 10% of the individuals were designated for validation purposes, while the remaining 90% served as reference for imputation. As a result, each replication consisted of approximately 43 individuals in the validation groups, totaling 717 individuals used in the analysis. To simulate the imputation scenario, the SNPs of individuals in the validation category were masked, leaving only the low-density SNP subsets representing the low-density panel (1K). In contrast, individuals classified in the reference category remained unchanged in their

genotypic data. For the validation category, genotypes were imputed using the FIMPUTE3 software (Sargolzaei et al., 2014). This process was conducted across all 10 replications to mitigate the effects of stochastic sampling. Subsequently, the imputed genotypes obtained from the validation replicates were compared to the actual genotypes to determine the accuracy of imputation. The Pearson correlation coefficient (R) was employed as a measure to assess accuracy, and the value of "R" was estimated using the following formula:

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(g_i - \bar{g})}{\sqrt{\sum_{i'1}^{n}(x_i - \bar{x})^2 \sum_{i'1}^{n}(g_i - \bar{g})^2}}$$

where for each SNP in which the true genotypes were masked, the possible genotypes were coded as 0, 1, or 2, representing the potential number of the smallest allele for that SNP in the target population; x_i represents the imputed genotype for individual i, and $\bar{x}$ represents the mean value of the imputed genotypes across all individuals. Likewise, g_i and $\bar{g}$ represent the variables corresponding to the observed genotypes, and n denotes the number of individuals in the validation group. The calculation of R, used as a measure of imputation accuracy, was performed at both the SNP marker and individual levels. SNPs that were imputed accurately demonstrated an accuracy of over 80% ($R^2 > 0.8$) across all animals. The evaluation of imputation accuracy was carried out within the chromosomes of tambaqui, which consist of 27 pairs. After validation, 12,107 SNPs were imputed with high imputation accuracy ($R^2 > 0.8$).

2.4.Genomic prediction analyzes

Before performing genomic prediction, the pedigree from challenged animals was reconstructed using the low-density SNPs. We used the likelihood probability method via the CERVUS program (Kalinowski et al., 2007). The individuals were retained based on the parameter assigned to both parents called "trio confidence". This parameter indicates the probability or certainty that the three individuals in question are related in a specific way, such as parent-child, full siblings, or half-siblings, among other possible relationships. In this way, from the 1,290 animals a total of 717 animals were maintained for downstream analyses. Genomic prediction analyses were conducted on two pieces of dataset: 1) genotypes from a pre-built low-density SNP array of 1K (1,246 SNPs), based on the AgriSeq® tGBS platform; and 2) imputed 1K genotypes to 12K. Both strategies were tested to evaluate the balance between low-density SNP arrays (lower genotyping cost) and prediction accuracy.

The resistance to the acanthocephalan was evaluated by the parasite load counting (RT_{CT}) within the animals, representing the degree of infestation of all fish at the end of the experiment (120 days from the beginning of the challenge test). The parasite load was determined by directly counting the parasites in the intestines of the animals. The final resilience (RS_F) to the acanthocephalan was assessed by the weight gain or loss in the final evaluation period, i.e., from the initial establishment of parasitism in all animals (BW1) to the end of the experiment (BW2), defined as RS_F = BW2 - BW1. The total resilience (RS_T) was calculated similarly, subtracting the final weight from the initial weight (before exposure to parasites, BW0), defined as RS_T = BW2 - BW0.

Two different methods were used for genomic prediction: pedigree-based linear prediction (PBLUP) and genomic-based linear prediction (GBLUP). The PBLUP method involved estimating genetic values using only the reconstructed pedigree information (without genotypes), aiming to compare and assess the benefits of including genomic information in genetic evaluations. All traits were analyzed using a mixed linear model to adjust continuous variables, according to the univariate animal model below:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z\alpha} + \boldsymbol{e}$$

where y is a vector of phenotypes (RT_{CT}, RS_F and RS_T), β is a vector of fixed effects (tank and origin for all phenotypes and BW_2 as co-variate for RT_{CT} only), α is a vector of random additive polygenic genetic effects that follows a normal distribution ~ N $(0, G\sigma_\alpha^2)$, where G is the genomic relationship matrix (VanRaden, 2008) constructed using the genotyped SNPs post-QC. However, when pedigree data is used, the G matrix is replaced by the pedigree relationship matrix (A) and σ_α^2 is the additive genetic variance, X and Z are incidence matrices, e is the random residual error with a distribution ~ N$(0, I\sigma_e^2)$, where I is an identity matrix and σ_e^2 is the residual variance. Several models were tested using Akaike Information Criterion (AIC) comparisons with different hypotheses regarding the fixed effects to define the best model for each phenotype. The populations from São Paulo and Amazonas were not produced at the same time (i.e., they had different ages). However, all animals have the same age within each population, meaning age and origin have the same binary distribution.

Therefore, adjusting the effects of age and origin simultaneously would cause overparameterization in the models.

The BLUPF90 software (Misztal, 2018) was used to solve the mixed linear models and calculate estimated breeding values (EBVs) for both phenotypic EBVs (PBLUP) and genomic EBVs (GBLUP). PBLUP, GBLUP1K and GBLUP_12K_imp were evaluated for all traits (RT_{CT}, RS_F and RS_T).

The accuracy of genomic prediction was compared using a cross-validation scheme through 5 replicates. Firstly, all animals were randomly divided into five validation sets. As a result, each replicate consisted of approximately 143 individuals in the validation groups, totaling 717 individuals used in the analysis. For each set, predictions were made by masking the phenotypes of the animals (20%) and using the remaining fish (80%) as the training group to estimate the genetic values of the unphenotyped individuals. Prediction accuracies to assess the performance of each model were estimated using the following formula (Legarra & Reverter, 2018):

$$\boldsymbol{\rho}^2_{\mathbf{cov(w,p)}} = \frac{\mathbf{cov}\,(\hat{\mathbf{u}}\mathbf{w},\hat{\mathbf{u}}\mathbf{p})}{(\mathbf{1}+\bar{\boldsymbol{F}}-\mathbf{2}\bar{\boldsymbol{f}})\boldsymbol{\sigma}^2\mathbf{u},\infty}$$

where, $\rho^2_{cov(w,p)}$ represents the accuracy of genomic prediction; cov (ûw,ûp) is the covariance between the genetic values or genomic genetic values using "complete" data (ûw, without missing phenotypes) and "partial" data (ûp, missing phenotypes from the validation sets); $\bar{F}$ is the average coefficient of inbreeding for the validation animals; $\bar{f}$ is the average coefficient

of relatedness; and σ^2u,∞ is the same as that of the pedigree model. These inbreeding components were considered to be zero, as the population analysed was recently established. Finally, the accuracy was calculated for each genomic prediction model for all traits of resistance and resilience to *N. buttnerae* and compared with those obtained by the PBLUP model.

3. Results

3.1.Genotyping and Imputation

Table 1 presents a summary of each genotyping technology and platform (Axiom Serra 30 K SNP® and Agriseq tGBS), along with the number of genetic markers (SNPs) post quality control filters. For the 1k scenario, the minor allele frequency (MAF) and levels of observed and expected heterozygosity (H_O and H_E) were as follows: MAF= 0.333 ± 0.114, H_O=0.435 ± 0.130, and H_E=0.418 ± 0.097. These metrics were slightly lower for the imputed 12K scenario: MAF= 0.311± 0.135, H_O= 0.397 ± 0.164, and H_E= 0.392 ± 0.133. Concerning genotype imputation, 83% of imputed SNPs (12K) exhibited an accuracy greater than 0.8, underscoring the potential of this approach in genomic selection studies. The mean accuracy of imputation for all imputed SNPs in each chromosome is available on Supplementary 1.

Table 1. Summary of each genotyping technology and platform used in genomic prediction for a tambaqui population challenged against *Neoechinorhynchus buttnerae*.

População[1]	Plataforma[2]	N_SNPs[3]	N_animais[4]	MAF[5]	HO[6]	HE[7]
Desafio	Agriseq tGBS	1018	717	0.333 ± 0.114	0.435 ± 0.130	0.418 ± 0.097
Desafio	Agriseq tGBS +Imputation	12107	717	0.311± 0.135	0.397 ± 0.164	0.392 ± 0.133

[1]Challenge populations used as reference and target animals for genotype imputation and construction of genomic prediction scenarios.
[2] Platform used for genotyping reference and target individuals.
[3] Number of SNPs used in each scenario after quality control.
[4] Number of animals.
[5] Minor allele frequency.
[6] Heterozygosity observed.
[7] Expected heterozygosity.

3.2.Genetic parameters and genomic prediction accuracy

Estimates of variance components for all models and traits (RS_f, RS_t, and RT_{ct}) are presented in Table 2. The heritability estimates were lower for the PBLUP method concerning the resilience traits RS_f (0.07 ± 0.050) and RS_t (0.121 ± 0.064). In contrast, higher values were observed for the GBLUP_1K methods, specifically with the RS_F trait (0.147 ± 0.056), and for the GBLUP_12K_imp method with the RS_T trait (0.187 ± 0.061). Lower values were noted for the resistance trait (RT_{ct}) with the GBLUP_1K and GBLUP_12K_imp methods (0.208 ± 0.058 and 0.232 ± 0.064 respectively), while a higher heritability was evident for the PBLUP method (0.284 ± 0.106).

Table 2. Estimates of heritability (h^2), genetic variance and residual variance with their respective standard error (SE) values, for resistance and resilience traits against *Neoechinorhynchus buttnerae* in tambaqui (*Colossoma macropomum)* were obtained using different statistical methods and densities of SNPs.

	RS_f			RS_t			RT_{ct}		
Método	h^2 **(SE)**	σ_α^2**(SE)**	σ_e^2**(SE)**	h^2 **(SE)**	σ_α^2**(SE)**	σ_e^2**(SE)**	h^2 **(SE)**	σ_α^2**(SE)**	σ_e^2**(SE)**
PBLUP	0.073	141.54	1743.4	0.121	283.99	2010.5	0.284	7262.8	17657.
	(0.050)	(96.236)	(112.21)	(0.064)	(154.07)	(143.37)	(0.106)	(2991.2)	(1933.9)
GBLUP1k	0.147	284.22	1619.9	0.182	425.48	1880.0	0.208	5046.9	18872.
	(0.056)	(112.77)	(110.23)	(0.057)	(143.55)	(130.24)	(0.0582)	(1529.9)	(1316.7)
GBLUP_12k_imp	0.142	270.84	1611.4	0.187	430.84	1844.7	0.232	5547.8	18105.
	(0.058)	(112.87)	(117.86)	(0.061)	(147.48)	(140.89)	(0.064)	(1643.9)	(1446.2)

RS_F: Final Resilience; RS_T: Total Resilience; RT_{CT}: Resistance by parasite counting

In cross-validation analyses aimed at estimating the accuracy of genomic prediction for resilience (RS_f-RS_T) and resistance (RT_{ct}) traits, it was observed that, in general, the accuracy in predicting genetic values using genomic methods surpassed that obtained with the PBLUP method. The accuracy values for the GBLUP_12K_imp method were higher for the resilience traits RS_f-RS_t (0.57 ± 0.20) and (0.64 ± 0.08) respectively, and were slightly lower than the accuracy presented for resistance in the GBLUP_1K method (0.42 ± 0.04). The average prediction accuracy values for resilience and resistance traits are described in Table 3 and Figure 1

Table 3. Mean accuracy values and standard deviation (SD) of EBV and GEBV for the traits of resistance and resilience to the endoparasite *Neoechinorhynchus buttnerae*.

Characteristics			
Método	**RSf**	**RSt**	**RTct**
	Accuracy	**Accuracy**	**Accuracy**
PBLUP	0,39 (0,14)	0,55 (0,12)	0,37 (0,05)
GBLUP_1K	0,55 (0,15)	0,68 (0,14)	0,42 (0,04)
GBLUP_12K_imp	0,57 (0,20)	0,70 (0,17)	0,40 (0,03)

RS_F: Final Resilience; RS_T: Total Resilience; RT_{CT}: Resistance by parasite counting

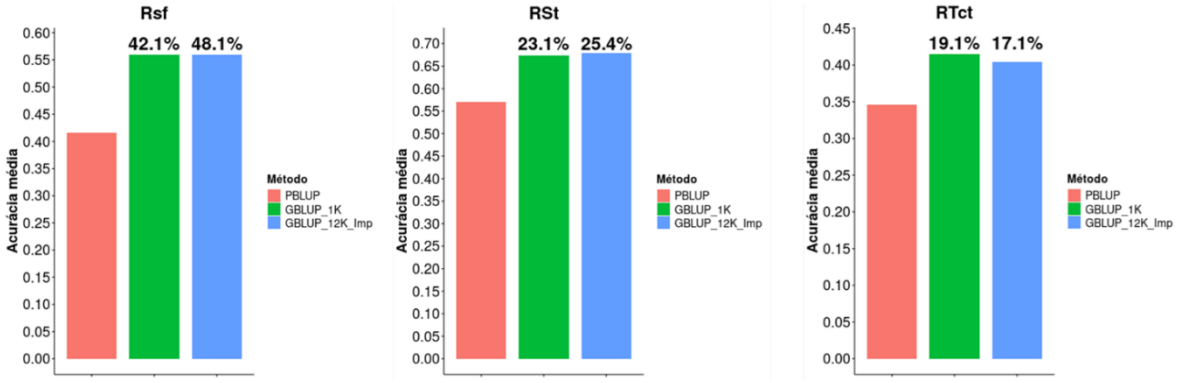


Figure 1. Accuracy of different genomic selection methods for resilience and resistance traits in tambaqui compared to PBLUP, using different densities of SNP arrays 1K and 12K imputed genotypes.

Higher accuracy values for the traits (RS_f, RS_t, and RT_{ct}) were observed for the GBLUP_1K and GBLUP_12K_imp methods compared to the classical pedigree-based selection method (PBLUP). This highlights the greater efficiency and accuracy of genomic information for selecting these traits (Figure 1). The percentage of genetic gain for resilience traits (RS_f and RS_t) was higher with the GBLUP_12k_imp method at 48.1% and 25.4%, respectively, compared to the PBLUP method. For the resistance trait (RT_{ct}), the GBLUP_1k method showed a genetic gain of 19.1% over PBLUP. It was evidenced that for the trait RT_{ct}, GBLUP_1k showed slightly higher accuracy (19.1%) over the PBLUP method compared to the genomic data from GBLUP_12k_imp, which had an accuracy 17% higher over PBLUP. However, no significant differences were observed in the evaluated traits (RS_f, RS_t, and RT_{ct}) regarding accuracy values and genetic gain between the two genomic-based methods.

4. Discussion

Genomic selection is a powerful approach for selective breeding, relying on genomic estimated breeding values (GEBV) to assess the breeding potential of candidate individuals using genome-wide genetic markers. (Meuwissen et al., 2001) The pursuit of cost-effective strategies is driving the increased utilization of GS to enhance economically significant traits in farmed fish. The selection of disease-resistant varieties is imperative for the sustainability of aquaculture. Traditional breeding methods based on pedigree may be limited in some cases, as they exploit only the genetic variation between families. With GS, estimated

breeding values (GEBVs) are predicted based on known genotypes exploiting the genetic variance between and within families leading to enhanced breeding efficiency(Prchal et al., 2023; Qu et al., 2024).

Recent studies focusing on genetic selection for disease resistance in the species *C. macropomum* have been conducted. These genetic studies include research on resistance to the bacterium *Aeromonas hydrophila* (Ariede et al., 2020; 2022), resistance to the ectoparasite *Ichthyophthirius multifiliis* (Guerra Lira, 2020; 2023), and resistance to the endoparasite *N. buttnerae* (Agudelo et al., 2024). These studies have shown the feasibility of selecting for disease resistance traits and achieving genetic improvement in this species.

The heritability results in the present study varied from low to moderate for resilience (RSf-RSt) and resistance (RTct) traits in tambaqui. Nonetheless, these heritability values suggest there is ample genetic variation to facilitate selection for resistance and resilience characteristics in this species. It's noteworthy that these estimates of genetic parameters mark the initial exploration of such traits in tambaqui challenged with an endoparasite, setting a foundation for further research in fish genetics.

Similar estimates were derived from pedigree data, where genetic parameters were estimated for tambaqui response to *N. buttnerae* infection. Heritability values ranged from low to moderate, specifically from 0.17 ± 0.040 to 0.20 ± 0.040, for both resistance and resilience traits. It's crucial to clarify that the populations analyzed in this study remain consistent with those examined previously, albeit with variations in the number of individuals assessed. (Agudelo et al., 2024). Heritability estimates for resistance to the ectoparasite

Ichthyophthirius multifiliis, affecting tambaqui, obtained through genomic information (GBLUP), were high and significant for survival status and time of death (0.51 ± 0.16 and 0.50 ± 0.16, respectively). Additionally, moderate heritability was observed for parasite load (0.20 ± 0.12). (Lira et al., 2023) Low heritability estimates (h2=0.07 ± 0.033 to h2=0.10 ± 0.034) were obtained with genomic and pedigree data for sea lice (*Caligus rogercressegi*) resistance, an ectoparasite that affects the Atlantic salmon (Vallejo et al., 2024).

The low heritability in disease resistance studies can be attributed to environmental factors, genetic complexity (e.g., polygenic traits), and genotype-environment interactions. Disease resistance involves intricate gene-environment interactions, making precise genetic isolation challenging. Environmental conditions, such as pathogen exposure and diet, greatly affect resistance phenotypes, complicating heritability estimates (Gjedrem, 2010; Houston, 2017; Ødegård et al., 2014)

To the best of our knowledge, there is solely one report detailing a study conducted on endoparasites, wherein genetic parameters were estimated concerning the response to *N. buttnerae* infection in tambaqui (Agudelo, 2024). While our heritability values may range from low to moderate, it's noteworthy that this study marks the initial attempt to assess genomic prediction for the selection of resistance traits within a tambaqui population challenged with *N. buttnerae*.

Genomic prediction is an indispensable tool in genomic selection and breeding programs. This tool integrates phenotypic information with genomic data and is also dependent on the heritability of the trait, the size and quality of the reference population, and

the genetic relationships between populations (M.H. Kjetså; 2020)). Our results shown that both the GBLUP_1K and GBLUP_12K genomic datasets demonstrated an average increase in accuracy of 11% and 15%, respectively, compared to PBLUP across the three evaluated traits. The difference in accuracy between GBLUP_1k and GBLUP_12k genomic data varied by 2%, with GBLUP_1k showing higher accuracy in the resistance trait (RTct) and GBLUP_12k demonstrating greater accuracy in resilience (RSf-RSt). This suggests a marginal difference in accuracy between the two genomic datasets. Overall, in this study, we observed accuracies of up to 70% with genomic information, underscoring the anticipated advantage in genetic gain facilitated by this source of information. Moreover, our findings suggest that genomic prediction serves as an effective tool for selecting resistance to *N. buttnerae*, potentially yielding genetic gains of up to 48% in resilience and 19% in resistance traits compared to the traditional pedigree-based method. Studies focusing on disease resistance caused by ectoparasites have consistently shown a similar trend, indicating the superiority of genomic-based predictions over those reliant solely on pedigree information. (Tsai et al., 2016; Robledo et al., 2018; Palaiokostas et al., 2016; Bangera et al., 2017; Vallejo et al., 2017; Barria et al., 2018; Yoshida et al., 2019).

The similarities in accuracy observed between the GBLUP_1K and GBLUP_12 genomic datasets could potentially be influenced by genotyping errors in the low-density panel leading to difficulties in the genotype imputation, or by errors introduced during sample identification.

Based on the observed accuracy values, it appears that the 1k panel is adequate for genomic prediction related to the selection of these traits. In this scenario, imputation for the higher density 12k panel seems unnecessary. This finding aligns with numerous studies suggesting that a modest number of markers (around 200 to 500 SNPs) may suffice for accurate genomic prediction. (Gorjanc et al., 2017; Delomas et al., 2023; Kriaridou et al., 2023; Tsai et al., 2017; Yoshida et al., 2018b

In a study conducted by Kriaridou et al. (2020), the author suggests that prediction accuracies in panels of up to 1000 markers can range from 0.44 to 0.75. However, below these densities, accuracies may be lower and more variable, ranging from 0.39 to 0.66 for densities of 100 to 200 markers. Densities of 1.000 to 2.000 markers are deemed sufficient for polygenic trais to achieve maximum accuracies. Conversely, accuracies tend to plateau from densities of around 10.000 to 15.000 markers (Kriaridou et al., 2020) .

Overall, our findings suggest that both the 1k panel and the 12k imputed data exhibited similar accuracy values, highlighting their cost-effectiveness and indicating the feasibility of utilizing low-density panels for prediction and genomic selection for resistance to *N. buttnerae* in Tambaqui.

5. Conclusion

The current study revealed that genomic prediction accuracy surpasses that of the pedigree-based method. Genotype imputation was a cost-effective strategy to expedite genetic gains

while minimizing genotyping expenses for this population. However, the 1K array was enough to produce reliable genomic breeding values estimations, in terms of accuracy. These findings offer insights for devising genotyping strategies aimed at reducing costs for the practical implementation of genomic selection to *N. buttnerae* resistance.

6. References

Agudelo, J. F. G., Neto, B. F. G., Mastrochirico-Filho, V. A., Ariede, R. B., Moreira, L. F. S., da Silva Noronha, L., Yáñez, J. M., Honczaryk, A., Valladão, G. M. R., & Hashimoto, D. T. (2024). Genetic parameters for host-response to acanthocephaliasis caused by the endoparasite *Neoechinorhynchus buttnerae* in the Amazon fish *Colossoma macropomum*. *Aquaculture, 582*. https://doi.org/10.1016/j.aquaculture.2023.740533

Agudelo, J. F. G., Mastrochirico-Filho, V. A., Borges, C. H. de S., Ariede, R. B., Lira, L. V. G., Neto, R. R. de O., de Freitas, M. V., Sucerquia, G. A. L., Vera, M., Berrocal, M. H. M., & Hashimoto, D. T. (2022). Genomic selection signatures in farmed *Colossoma macropomum* from tropical and subtropical regions in South America. Evolutionary Applications, 15(4), 679–693. https://doi.org/10.1111/eva.13351

Ajasa, A. A., Boison, S. A., Gjøen, H. M., & Lillehammer, M. (2024). Genome-assisted prediction of amoebic gill disease resistance in different populations of Atlantic salmon during field outbreak. *Aquaculture, 578*. https://doi.org/10.1016/j.aquaculture.2023.740078

Ariede, R. B., Freitas, M. V., Lira, L. V. G., Agudelo, J. F. G., Borges, C. H. S., Mastrochirico-Filho, V. A., Cáceres, P., Yáñez, J. M., & Hashimoto, D. T. (2022). Linkage map for chromosome-level genome anchoring and genome-wide association study for resistance to *Aeromonas hydrophila* in *Colossoma macropomum*. *Aquaculture, 560*. https://doi.org/10.1016/j.aquaculture.2022.738462

Ariede, R. B., Freitas, M. V., Agudelo, J. F., Borges, C. H., et al. (2020). Genetic (co) variation between resistance to *Aeromonas hydrophila* and growth in tambaqui (*Colossoma macropomum*). Aquaculture, 523, 735225.

Aslam, M. L., Carraro, R., Sonesson, A. K., Meuwissen, T., et al. (2020). Genetic variation, GWAS and accuracy of prediction for host resistance to *Sparicotyle chrysophrii* in farmed gilthead sea bream (*Sparus aurata*). Frontiers in Genetics, 11, 594770.

Bangera, R., Correa, K., Lhorente, J. P., Figueroa, R., et al. (2017). Genomic predictions can accelerate selection for resistance against *Piscirickettsia salmonis* in Atlantic salmon (*Salmo salar*). BMC genomics, 18(1), 1-12.

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

Correa, K., Bangera, R., Figueroa, R., Lhorente, J. P., et al. (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.

Delomas, T. A., Hollenbeck, C. M., Matt, J. L., & Thompson, N. F. (2023). Evaluating cost-effective genotyping strategies for genomic selection in oysters. Aquaculture, 562, 738844.

Dufflocq, P., Pérez-Enciso, M., Lhorente, J. P., & Yáñez, J. M. (2019). Accuracy of genomic predictions using different imputation error rates in aquaculture breeding programs: A simulation study. *Aquaculture*, *503*, 225–230. https://doi.org/10.1016/j.aquaculture.2018.12.061

Garcia, A. L., Bosworth, B., Waldbieser, G., Misztal, I., et al. (2018). Development of genomic predictions for harvest and carcass weight in channel catfish. Genetics Selection Evolution, 50, 1-12.

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

Gutierrez, A. P., Matika, O., Bean, T. P., & Houston, R. D. (2018). Genomic selection for growth traits in Pacific oyster (*Crassostrea gigas*): potential of low-density marker panels for breeding value prediction. Frontiers in Genetics, 9, 391.

Gorjanc, G., Dumasy, J. F., Gonen, S., Gaynor, R. C., et al. (2017). Potential of low-coverage genotyping-by-sequencing and imputation for cost-effective genomic selection in biparental segregating populations. Crop Science, 57(3), 1404-1420.

Hilsdorf, A. W. S., Uliano-Silva, M., 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. *Gigabyte*, *2021*, 1–14. https://doi.org/10.46471/gigabyte.29

Houston, R. D. (2017). Future directions in breeding for disease resistance in aquaculture species. In *Revista Brasileira de Zootecnia* (Vol. 46, Issue 6, pp. 545–551). Sociedade Brasileira de Zootecnia. https://doi.org/10.1590/S1806-92902017000600010

Jerônimo, G. T., Pádua, S. B. de, Belo, M. A. de A., Chagas, E. C., Taboga, S. R., Maciel, P. O., & Martins, M. L. (2017). *Neoechinorhynchus buttnerae* (Acanthocephala) infection in farmed *Colossoma macropomum*: A pathological approach. Aquaculture, 469, 124–127. https://doi.org/10.1016/j.aquaculture.2016.11.027

Kjetså, M. H., Ødegård, J., & Meuwissen, T. H. E. (2020). Accuracy of genomic prediction of host resistance to salmon lice in Atlantic salmon (*Salmo salar*) using imputed high-density genotypes. *Aquaculture*, *526*. https://doi.org/10.1016/j.aquaculture.2020.735415

Kriaridou, C., Tsairidou, S., Fraslin, C., Gorjanc, G., Looseley, M. E., Johnston, I. A., Houston, R. D., & Robledo, D. (2023). Evaluation of low-density SNP panels and imputation for cost-effective genomic selection in four aquaculture species. *Frontiers in Genetics*, *14*. https://doi.org/10.3389/fgene.2023.1194266

Kriaridou, C., Tsairidou, S., Houston, R. D., & Robledo, D. (2020). Genomic Prediction Using Low Density Marker Panels in Aquaculture: Performance Across Species, Traits, and Genotyping Platforms. *Frontiers in Genetics*, *11*. https://doi.org/10.3389/fgene.2020.00124

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

Lira, L. V. G., Mastrochirico-Filho, V. A., Mendes, N. J., Ariede, R. B., Yáñez, J. M., & Hashimoto, D. T. (2023). Genome-wide association study of host resistance to the ectoparasite *Ichthyophthirius multifiliis* in the Amazon fish *Colossoma macropomum*. *Molecular Biology Reports*, *50*(1), 599–607. https://doi.org/10.1007/s11033-022-08062-0

Manso, S. C. S., Garcia, B. F., Mastrochirico-Filho, V. A., Porto-Foresti, F., Yáñez, J. M., & Hashimoto, D. T. (2024). Genomic predictions for resistance to *Aeromonas hydrophila* in pacu (*Piaractus mesopotamicus*). *Aquaculture*, *582*. https://doi.org/10.1016/j.aquaculture.2024.740553

Mastrochirico-Filho, V. A., Ariede, R. B., Freitas, M. V., Borges, C. H., et al. (2021). Development of a multi-species SNP array for serrasalmid fish *Colossoma macropomum* and *Piaractus mesopotamicus*. Scientific Reports, 11(1), 19289.

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.

Ødegård, J., Moen, T., Santi, N., Korsvoll, S. A., Kjøglum, S., & Meuwisse, T. H. E. (2014). Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). *Frontiers in Genetics*, *5*(NOV). https://doi.org/10.3389/fgene.2014.00402

Palaiokostas, C., Cariou, S., Bestin, A., Bruant, J. S., et al. (2018). Genome-wide association and genomic prediction of resistance to viral nervous necrosis in European sea bass (*Dicentrarchus labrax*) using RAD sequencing. Genetics Selection Evolution, 50(1), 1-11.

Palaiokostas, C., Ferraresso, S., Franch, R., Houston, R. D., et al. (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.

Prchal, M., Palaiokostas, C., Gela, D., Piačková, V., Reschová, S., & Kocour, M. (2023). Genetic parameters and genomic prediction of resistance to koi herpesvirus disease using a low-density SNP panel on two Amur mirror carp populations. *Aquaculture Reports*, *30*. https://doi.org/10.1016/j.aqrep.2023.101582

Qu, S., Lu, S., Liu, Y., Li, M., & Chen, S. (2024). Accurate genomic selection using low-density SNP panels preselected by maximum likelihood estimation. *Aquaculture*, *579*. https://doi.org/10.1016/j.aquaculture.2023.740154

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.

Sebastião, F. de A., Braga de Oliveira, M. I., Rocha, M. J. S., Souza, D. C. de M., Ribeiro, P., Majolo, C., Crescêncio, R., & Chagas, E. C. (2021). Effect of a food additive in the control of the *acanthocephalan Neoechinorhynchus* buttnerae in *Colossoma macropomum*. Aquaculture Research, 52(2), 635–642. https://doi.org/10.1111/are.14920

Tsai, H. Y., Hamilton, A., Tinch, A. E., Guy, D. R., et al. (2016). Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. Genetics Selection Evolution, 48(1), 1-11.

Vallejo, R. L., Pietrak, M. R., Milligan, M. M., Gao, G., Tsuruta, S., Fragomeni, B. O., Long, R. L., Peterson, B. C., & Palti, Y. (2024). Genetic architecture and accuracy of predicted genomic breeding values for sea lice resistance in the St John River aquaculture strain of North American Atlantic salmon. *Aquaculture*, 740819. https://doi.org/10.1016/j.aquaculture.2024.740819

Vallejo, R. L., Silva, R. M., Evenhuis, J. P., Gao, G., et al. (2018). 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. Journal of Animal Breeding and Genetics, 135(4), 263-274.

Vallejo, R. L., Leeds, T. D., Gao, G., Parsons, J. E., et al. (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.

Yáñez, J. M., Barría, A., López, M. E., Moen, T., Garcia, B. F., Yoshida, G. M., & Xu, P. (2023). Genome-wide association and genomic selection in aquaculture. In *Reviews in Aquaculture* (Vol. 15, Issue 2, pp. 645–675). John Wiley and Sons Inc. https://doi.org/10.1111/raq.12750

Yang, B., Zhi, C., Li, P., Xu, C., Li, Q., & Liu, S. (2024). Genomic selection accelerates genetic improvement of resistance to Vibriosis in the Pacific oyster, *Crassostrea gigas*. *Aquaculture*, *584*. https://doi.org/10.1016/j.aquaculture.2024.740679

Yoshida, G. M., Carvalheiro, R., Rodríguez, F. H., Lhorente, J. P., et al. (2019). Single-step genomic evaluation improves accuracy of breeding value predictions for resistance to infectious pancreatic necrosis virus in rainbow trout. Genomics, 111(2), 127-132.

Yoshida, G. M., Carvalheiro, R., Lhorente, J. P., Correa, K., Figueroa, R., Houston, R. D., & Yáñez, J. M. (2018a). Accuracy of genotype imputation and genomic predictions in a two-generation farmed Atlantic salmon population using high-density and low-density SNP panels. *Aquaculture*, *491*, 147–154. https://doi.org/10.1016/j.aquaculture.2018.03.004

Yoshida, G. M., Bangera, R., Carvalheiro, R., Correa, K., et al. (2018b). Genomic prediction accuracy for resistance against *Piscirickettsia salmonis* in farmed rainbow trout. G3: Genes, Genomes, Genetics, 8(2), 719-726.

Zenger, K. R., Khatkar, M. S., Jones, D. B., Khalilisamani, N., et al. (2019). Genomic selection in aquaculture: application, limitations and opportunities with special reference to marine shrimp and pearl oysters. Frontiers in genetics, 9, 411091.

Considerações finais

O presente trabalho descreve um estudo abrangente sobre a resistência e resiliência do tambaqui em relação à doença causada pelo parasita N. buttnerae, com o objetivo de melhorar a produção e promover a sustentabilidade na aquicultura. O estudo envolveu análises genômicas para otimizar a imputação de genótipos e identificar SNPs eficazes para o desenvolvimento de um array comercial. Foram avaliados parâmetros genéticos, demonstrando a possibilidade de selecionar animais com resistência e resiliência. A análise de predição genômica mostrou um aumento na precisão da predição genômica, indicando o potencial de ganho genético por meio de abordagens genômicas. Concluindo que a seleção genômica é uma estratégia viável para melhorar a produção do tambaqui, com benefícios econômicos e ambientais para a indústria da aquicultura

Nosso trabalho aborda a importância do melhoramento genético na aquicultura, destacando a importância de estratégias para melhorar a resistência a doenças e promover a saúde e produtividade dos animais. exploramos como o melhoramento genético tradicional baseado em pedigree pode ter limitações e como a genômica aplicada pode superar essas limitações, permitindo uma seleção mais precisa. Ao desenvolver o primer SNP array de baixa densidade para o tambaqui pretende-se impulsionar o avanço genético da espécie, destacando a importância de abordagens integradas que combinem painéis de média e baixa densidade, conjuntamente com a predição genômica e a imputação de genótipos para melhorar a precisão da seleção genética e reduzir os custos associados à genotipagem. Desta forma mostramos uma alternativa de custo benefício para o melhoramento genético do tambaqui. e melhorar a sustentabilidade e eficiência da produção.

A problemática da acantocefalose tem impactado negativamente a produção de tambaqui, prejudicando seu crescimento e afetando a produtividade da espécie. Diante disso, torna-se crucial buscar soluções que melhorem a eficiência produtiva. Neste trabalho, propomos um programa de seleção genética como uma alternativa viável para enfrentar esse desafio.

Utilizando informações sobre parâmetros genéticos e ferramentas genômicas, buscamos identificar e selecionar animais resistentes ou resilientes ao acantocéfalo. A implementação desta tecnologia tem o potencial de reduzir significativamente as perdas econômicas associadas à acantocefalose. Ao selecionar indivíduos com estas caraterísticas, esperamos minimizar o desperdício de alimentação destinada a peixes que não conseguem alcançar seu peso esperado. Além disso, a redução do uso de produtos para o controle do parasito contribuirá para a produção de um alimento mais saudável e seguro para o consumo humano. Em suma, a seleção genômica é uma estratégia promissora para melhorar a produtividade e a sustentabilidade da produção desta espécie, trazendo benefícios econômicos e ambientais para a indústria da aquicultura do tambaqui.