

Universidade Federal do Amazonas – UFAM
Programa de Pós-graduação em Ciência Animal e Recursos Pesqueiros –
PPGCARP

ERIKA WALESKA CORDEIRO ROMALHO

**Identificação e análise de polimorfismos relacionados a genes de
resistência a doença em tambaqui (*Colossoma macropomum*)**

MANAUS – AMAZONAS

2025

ERIKA WALESKA CORDEIRO ROMALHO

Identificação e análise de polimorfismos relacionados a genes de resistência a doença em tambaqui (*Colossoma macropomum*)

Orientador: José de Ribamar da Silva Nunes

Dissertação apresentada ao Programa de Pós-Graduação em Ciência e Recursos Pesqueiros – PPGCARP da Universidade Federal do Amazonas – UFAM, como requisito final para obtenção do grau de Mestre em Ciência Animal e Recursos Pesqueiros.

MANAUS – AMAZONAS

2025

Ficha Catalográfica

Elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

-
- R757i Romalho, Erika Walesca Cordeiro
 Identificação e análise de polimorfismos relacionados a genes de
 resistência a doença em tambaqui (*Colossoma macropomum*) / Erika
 Walesca Cordeiro Romalho. - 2025.
 88 f. ; 31 cm.
- Orientador(a): José de Ribamar da Silva.
 Dissertação (mestrado) - Universidade Federal do Amazonas, Programa
 de Pós-Graduação em Ciência Animal e Recursos Pesqueiros, Manaus,
 2025.
1. Polimorfismos. 2. Resistência Imune. 3. Aquicultura Sustentável. 4.
 Adaptação imunológica. I. Silva, José de Ribamar da. II. Universidade
 Federal do Amazonas. Programa de Pós-Graduação em Ciência Animal e
 Recursos Pesqueiros. III. Título
-

ERIKA WALESKA CORDEIRO ROMALHO

**IDENTIFICAÇÃO E ANÁLISE DE POLIMORFISMOS RELACIONADOS A
GENES DE RESISTÊNCIA A DOENÇA EM TAMBAQUI (*COLOSSOMA
MACROPOMUM*)**

Dissertação apresentada ao Programa de Pós-graduação em Ciência Animal e Recursos Pesqueiros da Universidade Federal do Amazonas, como requisito para obtenção do título de Mestre em Ciência Animal e Recursos Pesqueiros, área de concentração em Produção Animal.

Aprovado em 11 de abril de 2025.

BANCA EXAMINADORA

Documento assinado digitalmente
 **JOSE DE RIBAMAR DA SILVA NUNES**
Data: 11/04/2025 11:24:01-0300
Verifique em <https://validar.itl.gov.br>

Dr. José de Ribamar da Silva Nunes
Presidente da banca
Universidade Federal do Amazonas

Documento assinado digitalmente
 **FABIO PERTILLE**
Data: 11/04/2025 11:43:48-0300
Verifique em <https://validar.itl.gov.br>

Dr. Fábio Pertille
Membro titular
Uppsala University

Documento assinado digitalmente
 **KALIANE NASCIMENTO DE OLIVEIRA**
Data: 23/04/2025 15:22:28-0300
Verifique em <https://validar.itl.gov.br>

Dra. Kaliane Nascimento de Oliveira
Membro titular
Universidade Federal do Amazonas

Dedico esta dissertação ao meu maior amor, a minha melhor amiga.

Mamãe, essa foi por você!

Agradecimentos

Dedico este trabalho ao meu namorado, Carlos Eduardo, que desde o início da minha dissertação tem sido meu apoio incondicional, oferecendo amor, carinho e ajuda em todos os momentos. Te amo!

Agradeço profundamente à minha mãe, Rose, por todo o suporte necessário, tanto financeiro quanto emocional, que me permitiu superar os desafios ao longo desta jornada.

Aos meus amigos Wallacy e Janaína, vocês foram meu porto seguro, riram e choraram comigo ao longo dessa jornada. Vocês foram essenciais para que essa jornada chegasse ao fim.

À minha amiga Miyuki, minha imensa gratidão por ter confiado o seu notebook, essencial para que meu projeto pudesse começar e se desenvolver.

Agradeço também à professora Flavia Souza, que, como coordenadora do programa, me auxiliou com os equipamentos necessários e o suporte indispensável para a realização deste trabalho.

Expresso meu sincero agradecimento à CAPES/CNPQ pela bolsa que possibilitou toda a pesquisa.

E por fim, mas não menos importante agradeço ao meu orientador José de Ribamar, por acreditar em mim e oferecer todo o suporte necessário para a conclusão da minha dissertação.

*Ao verme que primeiro roeu as frias carnes do meu cadáver dedico como
saudosa lembrança estas memórias póstumas.*

Machado de Assis

RESUMO

A diversidade genética e a adaptação imunológica são componentes centrais para a sustentabilidade da aquicultura de espécies nativas. Neste estudo, foram analisados 239 indivíduos de *Colossoma macropomum* (tambaqui) oriundos de diferentes subpopulações cultivadas no Brasil, com o objetivo de identificar e caracterizar polimorfismos de nucleotídeo único (SNPs) associados a genes relacionados à resistência a doenças. A partir da genotipagem por sequenciamento (GBS), foram detectados 182.416 SNPs polimórficos. As análises filogenética e de componentes principais (PCA) revelaram clara estruturação genética entre as subpopulações, refletindo ausência de fluxo gênico recente e possíveis efeitos de deriva genética e manejo reprodutivo isolado. Foram identificados SNPs com valores elevados de F_{ST} localizados em regiões codificantes de genes funcionalmente relevantes, como *unc5da*, *il16*, *epha2b*, *cd209* e *rnf175*, os quais participam de vias metabólicas críticas como sinalização por netrina, ativação de linfócitos T, sinalização *Ephrin*, reconhecimento de patógenos e apresentação antigênica via MHC I. A análise funcional das mutações revelou a presença de variantes sinônimas, não sinônimas e nonsense, sugerindo impactos potenciais sobre a estrutura e função das proteínas envolvidas na resposta imune. Esses resultados destacam loci candidatos à seleção natural e reforçam o papel dessas variantes como marcadores promissores em programas de melhoramento genético orientados à resistência a doenças. O presente estudo contribui para o avanço do conhecimento genômico do tambaqui e oferece bases para a aplicação de estratégias de seleção assistida por marcadores na aquicultura de espécies nativas.

Palavras-chave: Polimorfismos; Resistência Imune; Aquicultura Sustentável.

ABSTRACT

Genetic diversity and immune adaptation are central components for the sustainability of aquaculture involving native species. In this study, 239 individuals of *Colossoma macropomum* (tambaqui) from different farmed subpopulations in Brazil were analyzed to identify and characterize single nucleotide polymorphisms (SNPs) associated with disease resistance-related genes. Using genotyping-by-sequencing (GBS), a total of 182,416 polymorphic SNPs were detected. Phylogenetic and principal component analyses (PCA) revealed a clear genetic structure among subpopulations, indicating a lack of recent gene flow and suggesting the influence of genetic drift and isolated reproductive management. High F_{ST} SNPs were identified in coding regions of functionally relevant genes such as *unc5da*, *il16*, *epha2b*, *cd209*, and *rnf175*, which are involved in key metabolic pathways including netrin signaling, T cell activation, Ephrin signaling, pathogen recognition, and antigen presentation via MHC I. Functional analysis of the mutations revealed the presence of synonymous, non-synonymous, and nonsense variants, indicating potential impacts on the structure and function of proteins involved in immune responses. These findings highlight candidate loci under natural selection and reinforce the potential of these variants as valuable markers in genetic improvement programs aimed at enhancing disease resistance. This study contributes to advancing the genomic knowledge of tambaqui and provides a foundation for the application of marker-assisted selection strategies in native species aquaculture.

Keywords: Polymorphisms; Immune Resistance; Sustainable Aquaculture

SUMÁRIO

1. INTRODUÇÃO	12
2. REVISÃO DE LITERATURA	15
2.1. Tambaqui (<i>Colossoma macropomum</i>): Criação, Importância econômica, ecológica e principais doenças que afetam a espécie	15
2.2. Bases genéticas da resistência a doenças em peixes e sua importância na aquicultura	16
2.3. Polimorfismos em genes relacionados à resposta imune em peixes e sua associação com a resistência a doenças	18
2.4. Técnicas de sequenciamento de nova geração (NGS) e sua aplicação na identificação de polimorfismos em genes de resistência a doenças em peixes	20
2.5. Limitações e desafios na identificação e análise de polimorfismos em genes de resistência a doenças em peixes	23
3. OBJETIVOS	29
3.1. Objetivo Geral	29
3.2. Objetivos específicos	29
4. MATERIAL E MÉTODOS	30
4.1. Material Genético	30
4.2. Preparação das bibliotecas e sequenciamento das amostras	31
4.3. Tratamento das sequências geradas	31
4.4. Alinhamento das sequências	32
4.5. Filtragens	32
4.6. Análise de parentesco	33
4.7. Identificação de SNPs sob seleção natural	33
4.7.1 Cálculo das frequências alélicas	33
4.7.2 Cálculo de FST	34
4.7.3 Identificação de SNPs sob seleção natural	34
4.8. Localização dos SNPs nos genes	34
4.9. Caracterização funcional dos SNPs	35
4.10. Identificação de vias metabólicas	35
5. RESULTADOS	36
5.1. Descoberta de SNPs	36
5.2. Alinhamento das sequências	36
5.3. Análise de parentesco	37
5.3.1 Análise de parentesco com base na árvore filogenética	37
5.3.2 Análise de parentesco com base no PCA	38

5.4. Identificação de SNPs sob seleção natural.....	39
5.5 Localização dos SNPs nos genes	40
5.6 Caracterização funcional dos SNPs candidatos.....	41
5.7. Identificação de vias metabólicas.....	42
6. DISCUSSÃO.....	43
7. CONSIDERAÇÕES FINAIS	50
8. REFERÊNCIAS BIBLIOGRÁFICAS.....	51
9. ANEXOS.....	60
9.1 Anexo 1: Script de filtragem de sequencias em Python	60
9.1 Anexo 2: Script para análise de FST em Python	68
9.1 Anexo 3: Script para conversão dos códons em aminoácidos	72
9.1 Anexo 4: Sequencias FASTA alinhadas em genes relacionados a importantes funções imunológicas e de sinalização celular.	74

1. INTRODUÇÃO

A aquicultura tem experimentado um crescimento expressivo nas últimas décadas, desempenhando um papel importante na segurança alimentar e na economia global (FAO, 2020). Entre as espécies de peixes cultivadas, o tambaqui (*Colossoma macropomum*) destaca-se como um dos principais peixes de água doce nativos da América do Sul, tanto por seu valor comercial quanto por sua relevância ecológica (Hancz, 1993).

A intensificação da piscicultura, embora essencial para atender à crescente demanda por alimentos, frequentemente expõe os peixes a condições de estresse, como superlotação, má qualidade da água, manuseio frequente e alterações bruscas de temperatura. Essas condições não apenas comprometem o bem-estar dos peixes, mas também aumentam sua vulnerabilidade a infecções e doenças, criando desafios significativos para a produção sustentável. Reduzir esses fatores de estresse é fundamental para manter a saúde dos peixes e garantir a viabilidade econômica e ambiental da piscicultura (Lafferty et al., 2015).

Ao longo dos últimos onze anos, pesquisas relacionadas ao tambaqui têm revelado um cenário bastante diversificado na parasitologia, com o registro de inúmeros grupos de parasitos, entre eles protozoários, crustáceos e helmintos, conforme documentado por Valladão et al. (2014) e Jerônimo et al. (2017). Essa diversidade ilustra não apenas a abrangência dos parasitos que acometem essa espécie, mas também enfatiza que a natureza da relação entre parasitos e hospedeiros é complexa, dependendo intrinsecamente da espécie do parasito, do local da infecção e dos mecanismos pelos quais esses agentes atuam no organismo do tambaqui.

Essa complexidade das relações patogênicas encontra um paralelo interessante nas pesquisas desenvolvidas em outros organismos. O trabalho de Fukuoka (2019), por exemplo, mostrou que múltiplos polimorfismos funcionais presentes em um único gene podem, de maneira sinérgica, aumentar a resistência a determinadas doenças. Esse achado abre espaço para a hipótese de que mecanismos semelhantes de resistência podem operar em peixes e

outros animais aquáticos, sugerindo que a variabilidade genética desempenha, de fato, um papel crucial na capacidade de responder a patógenos.

Nesse mesmo sentido, estudos mais recentes, como os realizados por Houston et al. (2020), reforçam a importância dos polimorfismos genéticos na resposta imunológica de diversas espécies de peixes. A identificação de marcadores genéticos associados à resistência a doenças desponta, assim, como uma ferramenta essencial para o desenvolvimento de estratégias de melhoramento genético em tambaqui, visando não só a potencialização da resistência imunológica, mas também a promoção de uma aquicultura mais sustentável e produtiva.

A integração dos achados parasitológicos com os avanços na compreensão dos mecanismos imunológicos e genéticos fornece uma perspectiva abrangente e integrada. Essa abordagem multidisciplinar não apenas aprofunda nosso entendimento sobre as interações patogênicas que afetam o tambaqui, mas também impulsiona inovações tecnológicas que podem transformar as práticas de manejo e melhoramento genético em piscicultura.

A pesquisa em genética molecular tem avançado a passos largos, permitindo a identificação de genes e polimorfismos ligados à resistência a doenças em diversas espécies animais, inclusive entre os peixes (Houston et al., 2008). Essas descobertas têm sido fundamentais para o aprimoramento das técnicas de melhoramento genético, possibilitando a seleção de linhagens mais resistentes a patógenos específicos, como ressaltado por Gjerdem et al. (2012).

Em virtude desse cenário promissor, o presente estudo propõe-se a identificar e analisar polimorfismos relacionados a genes de resistência a doenças em tambaqui. Ao integrar essas análises, busca-se fornecer informações cruciais que possam orientar e aperfeiçoar as práticas de manejo e as estratégias de melhoramento genético para a espécie. Compreender os mecanismos moleculares por meio da análise dos polimorfismos genéticos é um passo importante para desenvolver abordagens mais eficazes e sustentáveis na produção de tambaqui e na implementação de programas de melhoramento

genético, visando à seleção de linhagens mais resistentes e adaptadas às condições de cultivo (Aguiar et al., 2018; Agudelo et al., 2022).

2. REVISÃO DE LITERATURA

2.1. Tambaqui (*Colossoma macropomum*): Criação, Importância econômica, ecológica e principais doenças que afetam a espécie

O tambaqui (*Colossoma macropomum*) é uma das espécies mais importantes para a piscicultura amazônica, justamente por sua rusticidade e versatilidade. Sua habilidade de migrar sazonalmente entre as águas brancas e pretas, aliada à excelente adaptação a diferentes condições físico-químicas, faz com que ele seja amplamente cultivado em viveiros escavados e em tanques-rede (Aride et al., 2007). Essa rusticidade natural, no entanto, precisa ser potencializada para atender à demanda de produções em larga escala, especialmente em sistemas que podem expor os animais a condições de estresse variáveis.

Os sistemas de cultivo variam em suas condições, afetando diretamente o bem-estar dos peixes. Por exemplo, viveiros escavados, comuns no Amazonas, são economicamente vantajosos, mas podem gerar alta densidade, com problemas como baixa oxigenação e acúmulo de matéria orgânica. Em contrapartida, tanques-rede oferecem alta produtividade e sustentabilidade, porém exigem manejo intensivo e dietas específicas, fatores que podem aumentar o estresse dos peixes. (Chagas et al., 2003; Gomes et al., 2006; Merola & Souza, 1988).

Com o aumento da intensificação dos sistemas de cultivo, observa-se um crescimento nos problemas relacionados a doenças, que em muitos casos são exacerbados pelo estresse associado a altas densidades e condições ambientais adversas. Patógenos como vírus e bactérias (por exemplo, infecções por *Aeromonas* spp., identificadas como as mais comuns em peixes de água doce) e parasitas – como os acantocéfalos que causam inflamações locais e generalizadas no trato digestivo – têm provocado perdas significativas durante todas as fases de criação (Merola & Souza, 1988; Noga, 1996; Lima et al., 2024; Figueiredo & Leal, 2008; Chagas et al., 2015). Além disso, doenças como a columnariose, desencadeadas por fatores de estresse decorrentes de manejo inadequado, transporte prolongado e baixa qualidade da água, aumentam o risco de surtos em sistemas intensivos (Onaka & Moraes, 2004).

Nesse contexto, os avanços na pesquisa em genética molecular vêm ganhando destaque ao possibilitar a identificação de polimorfismos genéticos relacionados à resistência a doenças em peixes, inclusive no tambaqui (Brito et al., 2012). O conhecimento desses marcadores genéticos é fundamental para o desenvolvimento de programas de melhoramento que selecionem linhagens mais resistentes a patógenos. Ao identificar esses polimorfismos, é possível entender melhor os mecanismos moleculares que conferem imunidade e, assim, reduzir a vulnerabilidade dos peixes ao estresse gerado por diferentes sistemas de cultivo. Tal abordagem não só aprimora a saúde dos animais, mas também garante maior produtividade e sustentabilidade em produções de larga escala, contribuindo decisivamente para a economia regional (Agudelo et al., 2022; Aguiar et al., 2018).

2.2. Bases genéticas da resistência a doenças em peixes e sua importância na aquicultura

A resistência a doenças é uma característica crucial na aquicultura, uma vez que surtos infecciosos podem gerar prejuízos econômicos significativos e comprometer a produtividade dos sistemas aquícolas (Potts et al., 2021). Em resposta a esses desafios, o melhoramento genético tem sido amplamente aplicado para ampliar a resistência em espécies de importância agrícola e aquícola, demonstrando grande potencial também para os peixes (Gomes et al., 2006; Lima et al., 2024; Liu et al., 2024).

O estudo das bases genéticas da resistência em peixes permite uma compreensão mais profunda dos mecanismos imunológicos que os protegem de patógenos específicos. Essa abordagem contribui não apenas para a seleção de indivíduos mais resistentes, aumentando a produtividade e garantindo o bem-estar dos animais, mas também para o desenvolvimento de estratégias de melhoramento adaptadas às demandas de uma produção intensiva (Bishop & Mackenzie, 2003; Stearn et al., 2001).

É importante ressaltar que a resposta imune dos peixes está sujeita a uma interação dinâmica entre fatores genéticos e ambientais. Alterações de temperatura, variações na qualidade da água, regimes alimentares e estresses

diversos interagem intimamente com o sistema imunológico, influenciando a capacidade dos animais de combater infecções (Urbinati & Carneiro, 2004; Mastrochirico-Filho et al., 2019; Van Muiswinkel & Nakao, 2014). Dessa forma, a avaliação da resistência a doenças exige a integração desses múltiplos fatores, bem como a elaboração de estratégias de manejo que promovam condições ambientais favoráveis.

A compreensão das bases genéticas da resistência a doenças em peixes se configura como uma ferramenta estratégica para a modernização da aquicultura. Ao desvendar os mecanismos moleculares e imunológicos que permitem a um peixe combater infecções, pesquisadores têm sido capazes de identificar variantes genéticas, por meio de estudos como o sequenciamento de nova geração e análises de associação genômica (GWAS), que correlacionam determinadas variações com uma resposta imune mais robusta. Esses polimorfismos, quando validados em estudos de seleção assistida por marcadores moleculares, permitem a implementação de programas de melhoramento genético que direcionam a seleção de linhagens resistentes, reduzindo significativamente os prejuízos causados por surtos infecciosos (Houston, 2017). Apesar dos desafios, como a necessidade de infecções experimentais em laboratório e a limitação de informações dos parentais (Shen et al., 2015; Yáñez & Martínez, 2010), o emprego de dados de gerações anteriores aliado aos métodos avançados de genômica vem contribuindo para superar essas barreiras.

Programas de melhoramento genético bem-sucedidos, como exemplificado com a tilápia – por meio do programa GIFT, originado na Malásia – demonstram como a integração de estratégias genômicas pode resultar em melhorias significativas na taxa de crescimento e na resistência dos animais (Gjerdrem, 2012; Khaw et al., 2015). De forma semelhante, a identificação de genes relacionados à imunidade inata em peixes, como demonstrado em estudos com tambaqui, é fundamental para elucidar a evolução do sistema imune e orientar intervenções terapêuticas e estratégias de manejo (Gjerdrem, 2009; Newman, 2014). Investigações envolvendo espécies como salmão, truta e bacalhau também ampliam o entendimento das bases genéticas e

imunológicas da resistência, facilitando a aplicação dessas estratégias em uma diversidade de espécies (Houston, 2017; Gjøen et al., 2016).

Além disso, a integração de abordagens genômicas com práticas de manejo integradas, que consideram fatores ambientais como a qualidade da água, temperatura e estratégias nutricionais, é essencial para reduzir o estresse e, conseqüentemente, a suscetibilidade dos peixes a doenças. Por exemplo, ao alinhar a seleção genética com práticas rigorosas de biossegurança, é possível diminuir a incidência de patógenos e mitigar os impactos dos fatores ambientais adversos, promovendo não apenas um ambiente mais saudável para os animais, mas também aumentando a lucratividade dos sistemas de produção (Urbinati & Carneiro, 2004; Mastrochirico-Filho et al., 2019). Contudo, o melhoramento genético não pode ser encarado como solução isolada para os desafios sanitários da aquicultura. É imprescindível adotar uma abordagem integrada que combine os avanços genômicos com a melhoria das condições ambientais, práticas de manejo rigorosas e medidas eficazes de biossegurança, garantindo assim a saúde dos peixes e a sustentabilidade dos sistemas produtivos (Urbinati & Carneiro, 2004; Van Muiswinkel & Nakao, 2014).

Inovações recentes, como o emprego da tecnologia CRISPR-Cas9, abrem novas perspectivas para o manejo da saúde dos peixes ao possibilitar modificações precisas em genes-chave da imunidade. Essa abordagem tem o potencial de acelerar o desenvolvimento de linhagens ainda mais resistentes, contribuindo para a adaptação dos sistemas aquícolas frente aos desafios impostos pelos patógenos e pelas condições ambientais. Dessa forma, a aplicação dessas técnicas moleculares pode reduzir perdas econômicas e contribuir para uma aquicultura mais sustentável e resiliente aos desafios atuais e futuros (Liu et al., 2024).

2.3. Polimorfismos em genes relacionados à resposta imune em peixes e sua associação com a resistência a doenças

Os polimorfismos em genes relacionados à resposta imune em peixes têm se destacado como um campo essencial para compreender e aprimorar a resistência a doenças nessas espécies, o que é vital para a sustentabilidade na

aquicultura. De acordo com Houston (2017), variações genéticas, como os SNPs, influenciam diretamente os mecanismos imunológicos, modulando a eficácia do reconhecimento e combate a patógenos.

Estudos de Bishop & Mackenzie (2003) e Gomes et al. (2006) reforçam que a integração de dados genômicos com estratégias de melhoramento genético possibilita a seleção de linhagens mais robustas, reduzindo os prejuízos provocados por surtos infecciosos, conforme evidenciado por Potts et al. (2021), Lima et al. (2024) e Liu et al. (2022). Além disso, fatores ambientais que interagem com a base genética, como indicado por Urbinati & Carneiro (2004), Mastrochirico-Filho et al. (2019) e Van Muiswinkel & Nakao (2014), demonstram a necessidade de uma abordagem integrada para a saúde dos peixes.

Complementando essa perspectiva, os desafios práticos na implementação desses programas, discutidos por Shen et al. (2015) e Yáñez & Martínez (2010), e os avanços observados em espécies como a tilápia (Gjerdrem, 2009; Khaw et al., 2015), salmão, truta e bacalhau (GjØen et al., 2016), fornecem uma base sólida para futuras intervenções voltadas à resistência a doenças na aquicultura.

A resposta imune em peixes é constituída por distintas vias de defesa, tais como a resposta inflamatória, antiviral, antibacteriana e antiparasitária, cada qual mediada por genes específicos (Secombes et al., 2019; Rauta et al., 2012; Gómez et al., 2013). Polimorfismos nesses genes podem modular a eficácia dessas respostas imunológicas, tendo, assim, um efeito direto na resistência dos peixes a diferentes agentes patogênicos (Li et al., 2020; Wang et al., 2015).

Por exemplo, a resposta inflamatória envolve genes como *TNF-α*, *IL-1β* e *IL-8*, e variações genéticas nesses loci podem tanto potencializar quanto comprometer a capacidade dos peixes de enfrentar desafios infecciosos (Li et al., 2020). No contexto da resposta antiviral, os genes *Mx* e *IFN-γ* desempenham papéis cruciais, e polimorfismos nesses genes podem alterar a capacidade dos peixes de eliminar vírus, influenciando, assim, sua suscetibilidade a infecções virais (Secombes et al., 2019).

Relativamente à resposta antibacteriana, os genes responsáveis pela produção de imunoglobulinas, tais como *IgM*, *IgT* e *IgD*, são essenciais para a neutralização de patógenos bacterianos. Polimorfismos nessas regiões genéticas podem modificar a eficácia imune dos peixes contra infecções bacterianas (Rauta et al., 2012). Por sua vez, a resposta antiparasitária, mediada por genes codificadores de Toll-like receptors (TLRs) e lectinas, também é suscetível às variações genéticas, que podem afetar a habilidade dos peixes em controlar infecções parasitárias (Gómez et al., 2013; Wang et al., 2015).

A compreensão da interação entre os polimorfismos em genes da resposta imune e a resistência a doenças possui implicações significativas para a aquicultura e a conservação. O conhecimento gerado a partir desses estudos contribui para o aprimoramento dos programas seletivos, direcionados à seleção de linhagens resistentes, além de orientar estratégias de manejo que promovam a manutenção da diversidade genética nas populações de peixes (Houston et al., 2020). Ademais, a genotipagem desses polimorfismos oferece informações valiosas para elucidar as associações entre variações genéticas e a capacidade dos peixes de resistir a infecções (Larhammar et al., 2018).

2.4 Técnicas de sequenciamento de nova geração (NGS) e sua aplicação na identificação de polimorfismos em genes de resistência a doenças em peixes

As técnicas de sequenciamento de nova geração (NGS) têm revolucionado a pesquisa em genômica, permitindo a obtenção de grandes volumes de dados genéticos de forma rápida e econômica (Neira, 2009). Essa revolução tecnológica tem sido fundamental para avançar a identificação de polimorfismos em genes que conferem resistência a patógenos, contribuindo significativamente para o desenvolvimento de estratégias de manejo, melhoramento genético, aquicultura sustentável e conservação. Em particular, a aplicação de NGS tem permitido elucidar os mecanismos moleculares subjacentes à defesa imunológica em peixes, oferecendo novas perspectivas para aumentar a resiliência dos organismos frente a infecções (Houston et al., 2020).

O sequenciamento de DNA de nova geração engloba diversas plataformas que têm sido amplamente empregadas na pesquisa genômica. Entre as principais, destacam-se os métodos Illumina, Ion Torrent, Pacific Biosciences (PacBio) e Oxford Nanopore. Esses métodos possibilitam a leitura simultânea de múltiplas sequências com alta cobertura e resolução, acelerando a obtenção de resultados e permitindo a detecção de variações genéticas em larga escala (Mardis, 2008; Goodwin et al., 2016).

Além do sequenciamento de DNA, a técnica de RNA-Seq tem proporcionado uma análise abrangente do transcriptoma. Por meio do sequenciamento do RNA mensageiro (mRNA), torna-se possível investigar os perfis de expressão gênica sob diferentes condições experimentais. Essa abordagem é crucial para a identificação dos mecanismos moleculares que regulam a resposta imune dos peixes frente a agentes patogênicos, permitindo a identificação de genes cuja expressão varia em resposta a estressores ou infecções (Wang et al., 2009; Krasnov et al., 2016).

Outro avanço significativo é o sequenciamento do exoma e do genoma completo, que possibilita a identificação de polimorfismos em genes relacionados à defesa imunológica. Essa técnica tem permitido descobrir variantes funcionais que influenciam a resposta imune e a suscetibilidade a doenças, oferecendo um panorama detalhado das alterações genéticas que podem ser exploradas em programas de seleção (Clark et al., 2011; Larhammar et al., 2018). Complementarmente, métodos como o Sequenciamento de Regiões Adjacentes a Restrição (RAD-Seq) têm sido empregados para mapear marcadores genéticos em loci específicos. Essa abordagem facilita a realização de estudos de associação, permitindo a identificação de loci de genes que estão relacionados à proteção contra patógenos em peixes (Baird et al., 2008; Nowak et al., 2021).

A detecção de polimorfismos em genes que atuam na proteção contra patógenos é uma das aplicações mais promissoras das técnicas de NGS. A identificação dessas variações genéticas é essencial para a implementação de programas de melhoramento genético, que podem empregar a seleção assistida

por marcadores (MAS) para produzir populações de peixes mais saudáveis e resilientes (Gjedrem et al., 2012; Tsai et al., 2015).

Após a identificação dos polimorfismos, seguidores dos estudos avançam para a análise de genes candidatos, com ênfase na validação funcional dessas variantes. Técnicas como *knockdown* e métodos de edição genética, exemplificados pela tecnologia CRISPR/Cas9, são aplicadas para confirmar a relevância funcional dos genes identificados, o que permite que intervenções sejam direcionadas e otimizadas de maneira mais precisa (Yáñez et al., 2014).

Os avanços proporcionados pelas técnicas de NGS não se restringem ao melhoramento genético para a aquicultura; eles se estendem também à conservação e manejo de populações de peixes selvagens. A identificação de polimorfismos em genes de proteção contra patógenos permite avaliar a vulnerabilidade das populações naturais e orientar estratégias de manejo que visam a manutenção da diversidade genética. Essa abordagem é fundamental para o desenvolvimento de medidas de biossegurança e para a formulação de estratégias de conservação que garantam a sustentabilidade a longo prazo dos recursos pesqueiros (Bernatchez et al., 2017; Allendorf et al., 2010).

Outra vertente inovadora derivada do uso das técnicas de NGS é a integração dos dados genômicos com abordagens de biologia de sistemas. A análise de redes gênicas e de sinalização permite uma visão integrada das interações entre os genes que conferem resistência a patógenos e os processos biológicos envolvidos. Essa abordagem sistêmica tem o potencial de identificar alvos específicos para terapias ou intervenções genéticas, melhorando nossa compreensão dos mecanismos de defesa dos peixes (Barabási et al., 2011; Ideker et al., 2001).

As técnicas de NGS também têm sido aplicadas em estudos de metagenômica, que examinam a diversidade e a função das comunidades microbianas associadas aos peixes. Esse tipo de análise auxilia na identificação de microrganismos que podem desempenhar papéis benéficos na saúde dos peixes, modulando a resposta imune e contribuindo para a resistência a doenças. Esses insights são particularmente valiosos para o desenvolvimento

de estratégias de manejo baseadas em probióticos ou prebióticos, que visam melhorar o equilíbrio da microbiota em sistemas aquícolas (Caporaso et al., 2011; Merrifield et al., 2010).

A integração dos dados genômicos com informações fenotípicas e ambientais é outro ponto de destaque nas pesquisas atuais. Essa abordagem permite compreender as interações entre o genótipo e o ambiente – um fator crucial para a modulação da expressão gênica e para determinar a resistência a patógenos em peixes. Estudos de genômica ambiental, como aqueles que utilizam análises de eQTL, possibilitam a identificação de loci que afetam a expressão de genes em resposta a estressores ambientais, como infecções e alterações climáticas. Essa compreensão pode, por sua vez, orientar estratégias de manejo que minimizem os impactos negativos desses estressores, contribuindo para a formação de populações mais resistentes (Lande et al., 2015; Rockman et al., 2008).

O contínuo avanço das técnicas de NGS, aliado à redução dos custos de sequenciamento, está expandindo as fronteiras da pesquisa genômica para uma variedade maior de espécies de peixes. Essa acessibilidade ampliada permite que estudos sejam realizados em espécies menos estudadas ou de difícil acesso, contribuindo para a identificação de novos genes envolvidos na resistência a patógenos e para um melhor entendimento da evolução e conservação dos mecanismos imunológicos em diferentes contextos ecológicos (Koboldt et al., 2013; Lien et al., 2016).

2.5. Limitações e desafios na identificação e análise de polimorfismos em genes de resistência a doenças em peixes

A identificação e análise de polimorfismos em genes de resistência a doenças em peixes é uma área de estudo que enfrenta desafios complexos, contribuindo para a limitação dos avanços no melhoramento genético e na conservação dessas espécies. Embora as técnicas de sequenciamento de nova geração (NGS) tenham impulsionado significativamente a genômica dos peixes, diversos fatores intrínsecos e extrínsecos complicam a detecção e a interpretação das variantes genéticas associadas à imunidade.

Uma das principais limitações na identificação de polimorfismos está na própria diversidade genômica dos peixes. Espécies diferentes apresentam genomas com tamanhos variados, níveis distintos de repetitividade e graus de complexidade que podem afetar diretamente a qualidade dos dados gerados pelo sequenciamento (Lien et al., 2016). Enquanto em alguns organismos, como mamíferos e plantas, existem protocolos padronizados bem estabelecidos, a grande heterogeneidade dos genomas dos peixes dificulta a aplicação uniforme de métodos genômicos. Essa variabilidade não apenas torna o mapeamento e a montagem dos genomas mais desafiadores, mas também impõe barreiras à identificação precisa dos polimorfismos, sobretudo em regiões altamente repetitivas, onde as taxas de erro podem ser elevadas (Koboldt et al., 2013).

Além disso, a alta taxa de evolução observada em muitos genomas de peixes pode resultar em uma abundância de mutações neutras ou de baixo impacto funcional. Essas mutações, que não alteram significativamente a função proteica, podem mascarar a presença de variantes de real interesse funcional para a resposta imunológica. Assim, distinguir quais polimorfismos representam verdadeiros alvos para o melhoramento da resistência a doenças torna-se uma tarefa complexa, exigindo abordagens analíticas sofisticadas capazes de filtrar o “ruído” genético predominante (Neira, 2009; Santos, M. C. F.; Ruffino; Farias, 2007).

Outro desafio significativo é a escassez de genomas de referência de alta qualidade para inúmeras espécies de peixes. Genomas bem anotados são fundamentais para métodos de mapeamento que permitem a identificação de polimorfismos, especialmente quando se estudam genes de interesse nas respostas imunes. Em comparação com mamíferos e plantas, onde os genomas de referência contam com extensiva validação e refinamento, o número limitado de genomas disponíveis para peixes impede a correta delimitação de regiões genômicas e dificulta a identificação de variantes relevantes (Lien et al., 2016). Essa limitação é agravada pela existência de linhagens selvagens com alta diversidade genética, onde as variações inerentes não são bem representadas nos genomas de referência disponíveis, comprometendo a eficácia dos estudos comparativos e de associação.

A detecção de polimorfismos em genes de resistência depende fortemente da resolução e sensibilidade das técnicas de NGS. Entretanto, a alta taxa de evolução dos genomas dos peixes, aliada à diversidade de mutações, pode dificultar a distinção entre variantes funcionais e aquelas sem efeitos fenotípicos aparentes. A presença de um grande número de variantes neutras cria um cenário onde as análises computacionais precisam trabalhar com altos índices de falsos-positivos e falsos-negativos, comprometendo a robustez das conclusões sobre a associação entre um polimorfismo específico e a resistência a patógenos (Neira, 2009; Santos, M. C. F.; Ruffino; Farias, 2007).

A ausência de um conhecimento aprofundado sobre os genes e as vias de sinalização que mediam a resposta imune em peixes também dificulta a interpretação dos dados. Sem uma compreensão clara das funções gênicas e da interconexão entre diferentes vias de sinal, fica desafiador identificar variantes que realmente impactam a resistência a doenças. Estudos que buscam estabelecer essa relação frequentemente esbarram em lacunas no conhecimento, o que reforça a necessidade de pesquisas mais detalhadas e integrativas nesse campo (Nowak et al., 2021; Clark et al., 2011).

A análise de polimorfismos é frequentemente comprometida por fatores relacionados à estrutura populacional. Em populações de peixes, a variabilidade genética interna pode gerar efeitos de confusão, onde diferenças entre subpopulações podem ser erroneamente interpretadas como associações causais com a resistência a doenças. Essas discrepâncias levam à ocorrência de falso-positivos e falso-negativos, interferindo na síntese de conclusões robustas. Estudos de associação genética, por exemplo, necessitam de amostras cuidadosamente selecionadas para reduzir esses efeitos, o que pode se tornar um obstáculo em espécies com ampla distribuição geográfica e alta diversidade genética (Houston et al., 2020; Bernatchez et al., 2017).

Além disso, a interação genótipo-ambiente adiciona uma camada extra de complexidade. Fatores ambientais – como variações de temperatura, salinidade e exposição a patógenos – influenciam diretamente a expressão dos genes e podem mascarar ou modificar a associação entre um polimorfismo e a resistência a doenças (Lande et al., 2015; Rockman et al., 2008). Essa

heterogeneidade ambiental necessita ser quantificada e integrada aos estudos, por meio de abordagens que combinem dados fenotípicos e ambientais com os resultados do sequenciamento, para que se obtenha uma visão completa do cenário de resistência imunológica.

Mesmo com a queda nos custos de sequenciamento proporcionada pelas tecnologias NGS, o investimento necessário para realizar estudos genômicos abrangentes em peixes permanece elevado, sobretudo em países em desenvolvimento. A infraestrutura de laboratórios equipados com tecnologia de ponta, a capacitação de pessoal especializado e os custos computacionais para processamento dos dados podem representar barreiras significativas para muitos centros de pesquisa (Koboldt et al., 2013). Esse fator limita o acesso a estudos avançados de genômica e restringe o número de espécies e populações que podem ser analisadas de forma aprofundada.

A integração dos dados obtidos por diferentes abordagens – como genômica funcional, comparativa e populacional – também representa um desafio adicional. A utilização de ferramentas computacionais avançadas e o desenvolvimento de pipelines específicos para análise integrada demandam conhecimentos multidisciplinares, combinando biologia, estatística e ciência da computação (Ideker et al., 2001; Yáñez et al., 2014). Essa integração é essencial para aumentar a robustez das análises, porém, requer uma coordenação entre equipes e a disponibilidade de software e hardware adequados para lidar com os grandes volumes de dados gerados.

Após a identificação dos polimorfismos potencialmente associados à resistência a doenças, o próximo passo crítico é a validação funcional dessas variantes. A confirmação de que um polimorfismo realmente altera a resposta imune e a suscetibilidade a patógenos envolve a realização de experimentos funcionais, como knockdown ou edição de genes utilizando a tecnologia CRISPR/Cas9. Esses métodos permitem a modulação direta da expressão dos genes candidatos, possibilitando a observação dos efeitos fenotípicos resultantes (Yáñez et al., 2014).

Contudo, a validação funcional é altamente desafiadora. Ela requer condições experimentais rigorosas e a disponibilidade de modelos experimentais que reproduzam de forma confiável o ambiente natural ou as condições de manejo dos peixes. Além disso, a complexidade dos sistemas imunológicos, que envolvem redes de interações entre múltiplos genes e vias de sinalização, torna difícil isolar o impacto de um único polimorfismo. Essa dificuldade pode levar a ambiguidades na interpretação dos dados e na definição precisa da relevância de cada variante para a resistência imunológica.

Os desafios e limitações na identificação e análise de polimorfismos em genes de resistência a doenças em peixes são multifacetados e abrangem desde a complexidade intrínseca dos genomas até limitações práticas relacionadas à infraestrutura e à integração dos dados. A diversidade genética, a ausência de genomas de referência de alta qualidade, a elevada taxa de evolução e os efeitos de estrutura populacional e interação genótipo-ambiente representam barreiras significativas para o avanço desse campo (Lien et al., 2016; Koboldt et al., 2013; Neira, 2009).

Superar esses desafios requer uma abordagem integradora que combine o avanço das tecnologias de sequenciamento com a implementação de métodos computacionais sofisticados e a colaboração multidisciplinar entre pesquisadores. O uso de técnicas emergentes, como CRISPR/Cas9 para a validação funcional, e a integração de dados fenotípicos, ambientais e genômicos surgem como estratégias promissoras para contornar as limitações atuais (Yáñez et al., 2014).

No contexto da aquicultura, a habilidade de identificar e validar polimorfismos associados à resistência a doenças é crucial para o desenvolvimento de linhagens mais robustas e para a formulação de estratégias de manejo e conservação. Assim, os avanços no campo da genômica não apenas contribuem para o melhoramento genético, mas também têm implicações diretas na sustentabilidade e lucratividade dos sistemas aquícolas.

Em perspectiva, esforços contínuos para aumentar a qualidade dos genomas de referência, a integração de análises multiescala e o fortalecimento

de parcerias internacionais poderão ampliar o acesso e a aplicação das tecnologias NGS para um número maior de espécies. Tal avanço é fundamental para enfrentar os desafios da resistência a doenças em peixes e consolidar a aquicultura como um setor estrategicamente sustentável para a segurança alimentar global.

3. OBJETIVOS

3.1. Objetivo Geral

Identificar e analisar polimorfismos relacionados a genes de resistência a doença em tambaqui (*Colossoma macropomum*).

3.2. Objetivos específicos

1. Fazer uma análise populacional com base nos polimorfismos identificados;
2. Estudar os mecanismos genéticos e as vias bioquímicas associadas a resistência a doenças em tambaqui.
3. Avaliar o potencial dos polimorfismos identificados como biomarcadores para seleção de indivíduos geneticamente mais resistentes a doenças na aquicultura do tambaqui.

4. MATERIAL E MÉTODOS

4.1. Material Genético

Neste estudo, um total de 239 espécimes foram obtidos a partir de sete pisciculturas situadas em três distintas regiões do Brasil, bem como de uma coleta de animais selvagens provenientes do Rio Uatumã-AM. A **Tabela 1** fornece informações detalhadas sobre os locais e a quantidade de amostras coletadas em cada localidade.

Tabela 1. Locais onde as amostras foram coletadas com suas respectivas coordenadas geográficas.

Cód.	Local	Latitude	Longitude	Região	N
AM1	Balbina – AM	1°55'54.4.0"	58°07'32.2"	Norte	38
AM2	Pres. Figueiredo – AM	1°57'14.3"	60°02'45.7"	Norte	30
AM3	Rio Uatumã – AM	2°35'07.0"	58°07'32.2"	Norte	30
PA	Mojuí de Campos – PA	2°53'5.25"	54°36'10"	Norte	30
CE	Pentecoste – CE	3°48'11.2"	39°16'06.2"	Nordeste	31
RO	Porto Velho – RO	8°50'52.8"	63°52'17.8"	Norte	40
SP1	Brumadinho – SP	22°31'16.0"	46°53'5.71"	Sudeste	28
SP2	Pirassununga – SP	21°57'31.9"	47°26'56.0"	Sudeste	12

Amostras da nadadeira caudal de cada indivíduo foram cuidadosamente coletadas e armazenadas em etanol 90% para preservação. Posteriormente, o DNA foi extraído utilizando o protocolo de proteinase K. A integridade do DNA extraído foi verificada por meio de eletroforese em gel de agarose a 1%, garantindo a qualidade do material genético. Todas as amostras foram mantidas em condições ideais a -20°C antes da preparação das bibliotecas de sequenciamento.

É importante ressaltar que todos os protocolos experimentais seguiram rigorosamente a Diretriz Brasileira para o Cuidado e a Utilização de Animais para Fins Científicos e Didáticos – DBCA, estabelecida pelo Conselho Nacional de Controle de Experimentação Animal (CONCEA). Assim, garantiu-se que nenhum dos indivíduos foi submetido a manipulações experimentais invasivas ou eutanásia, utilizando-se exclusivamente sequências genéticas de estudos prévios já realizados.

4.2. Preparação das bibliotecas e sequenciamento das amostras

As bibliotecas de DNA foram preparadas e sequenciadas no Laboratório de Biotecnologia Animal da Universidade de São Paulo (Piracicaba, Brasil), utilizando a metodologia de genotipagem por sequenciamento (GBS) conforme o protocolo apresentado por Elshire et al. (2011), com modificações descritas por Nunes et al. (2017). Essas modificações permitem resultados mais precisos na análise genética.

As bibliotecas foram sequenciadas empregando o kit Illumina TruSeq SBS v3-HS, em conjunto com o sequenciador Illumina HiSeq2500 (Illumina, San Diego). O uso desse sistema de sequenciamento fornece dados de qualidade, assegurando a confiabilidade das análises realizadas neste estudo.

4.3. Tratamento das sequências geradas

O processamento de qualidade foi executado utilizando a ferramenta SeqyClean v. 1.9.10 (<https://bitbucket.org/izhbannikov/seqyclean/>), garantindo a remoção de sequências indesejadas e otimização dos dados. A identificação de SNP ocorreu por meio do pipeline UNEAK (Universal Network Enabled Analysis

Kit) aplicando parâmetros padrão, enquanto a filtragem dos dados de sequenciamento para todas as amostras foi feita com o auxílio do Tassel v.5.2.73.

Para facilitar a manipulação das sequências genômicas geradas, o arquivo FASTA contendo sequências de tabaco foi dividido em arquivos menores, cada um contendo um número definido de sequências. Esse método permitiu uma manipulação mais eficiente do conjunto de dados de sequências, otimizando o processamento no banco de dados do National Center for Biotechnology Information (NCBI).

Utilizando a ferramenta de pesquisa do NCBI (Entrez Query), definimos critérios específicos de busca por sequências genômicas associadas a genes de resistência a doenças no genoma do tabaco. O objetivo deste procedimento é otimizar a criação de consultas personalizadas, permitindo buscas mais precisas de sequências genéticas no banco de dados.

4.4. Alinhamento das sequências

O alinhamento foi realizado usando a Ferramenta Básica de Pesquisa de Alinhamento Local (BLAST). O programa compara sequências de nucleotídeos ou proteínas com bancos de dados de sequências e calcula a significância estatística das correspondências. O limitador de busca foi usado para ajustar o algoritmo de alinhamento da ferramenta BLAST.

Para ampliar a busca, foram realizados alguns ajustes nos parâmetros do BLAST, usando como algoritmo de busca o *discontiguous megablast*, que procura por sequências menos similares e aumenta o número de alinhamentos. Os outros parâmetros ajustados foram: *Max target sequences* (50); *Word size* (11); *Template length* (16).

4.5. Filtragens

Para integrar e processar dados genômicos provenientes do alinhamento realizado usando a ferramenta BLAST, foi desenvolvido um pipeline em Python (versão 3.11) para combinar, filtrar e organizar informações genéticas de forma

automatizada, culminando na extração de sequências relevantes para análises subsequentes (Anexo 1).

4.6. Análise de parentesco

Para avaliar se os polimorfismos estudados estão caracterizando a população, serão realizadas duas análises de parentesco com base nos marcadores anotados em genes relacionados à resistência a doenças.

Inicialmente, foi construída uma árvore filogenética usando o método de agrupamento Neighbor Joining (NJ) com base nos marcadores utilizando o programa TASSEL. O processo de construção de uma árvore filogenética no TASSEL ocorre em duas etapas. Primeiro, o software calcula a distância entre pares de taxas com base nos genótipos, utilizando uma distância euclidiana modificada, onde homozigotos têm 100% de semelhança e heterozigotos 50%. Em seguida, a matriz de distâncias é usada para gerar uma árvore filogenética por meio dos métodos Neighbor-Joining ou UPGMA, conforme a escolha do usuário. A árvore gerada é salva no formato Newick e pode ser visualizada no TASSEL ou exportada para outros programas.

Para complementar o estudo das relações genéticas entre os indivíduos, será realizada uma análise de componentes principais (*principal component analysis* - PCA) e os resultados serão plotados com base em dois componentes principais.

4.7. Identificação de SNPs sob seleção natural

Utilizamos um pipeline em Python (versão 3.11) para identificar SNPs sob seleção natural em populações de Tambaqui com base em análises de FST (Anexo 2).

4.7.1 Cálculo das frequências alélicas

Para calcular as frequências alélicas, foi empregada uma função que mapeia os códigos de nucleotídeos derivados da IUPAC para seus pares de bases correspondentes. Foram consideradas as variantes alélicas A, C, G, T, além de combinações ambíguas (e.g., R para A/G). As frequências alélicas foram

calculadas para cada população com base na contagem total de alelos em cada SNP.

4.7.2 Cálculo de FST

Os valores de FST foram calculados para cada SNP utilizando as frequências alélicas de duas populações. A fórmula aplicada seguiu o cálculo tradicional de FST, com base nos índices de heterozigosidade total (h_t) e heterozigosidade dentro das populações (h_s). O valor de FST foi calculado da seguinte forma:

$$h_t: 1 - \sum((\text{freq_POP1} + \text{freq_POP2})/2)^2$$

$$h_s: \text{Média de } (1 - \sum(\text{freq_POP1}^2)) \text{ e } (1 - \sum(\text{freq_POP2}^2))$$

$$\text{FST: } (h_t - h_s) / h_t, \text{ caso } h_t \text{ seja maior que zero.}$$

O arquivo de entrada contendo as informações genotípicas das populações de tabaqui foram carregados em um dataframe utilizando a biblioteca Pandas.

4.7.3 Identificação de SNPs sob seleção natural

Os SNPs sob potencial seleção natural foram identificados com base em um valor limiar de $\text{FST} > 0.15$, sugerindo forte diferenciação entre populações. Esse limiar foi ajustado conforme a literatura e as necessidades específicas do estudo.

4.8 Localização dos SNPs nos genes

As posições dos SNPs foram determinadas com base nos alinhamentos das sequências polimórficas com os transcritos de referência disponíveis no GenBank (NCBI). A localização dos SNPs dentro das regiões codificadoras (CDS) foi estimada com base na posição média do alinhamento no mRNA e na anotação das CDS.

Para visualização, foram geradas representações gráficas com a região CDS. As extremidades 5' e 3' foram marcadas com base nos limites da CDS, permitindo identificar a distribuição e o contexto genômico dos polimorfismos.

4.9 Caracterização funcional dos SNPs

A caracterização dos SNPs previamente identificados como localizados em regiões gênicas associadas a funções biológicas relevantes e diferencialmente distribuídos entre as populações analisadas ($F_{ST} > 0.15$).

A tradução dos códons de referência e variantes foi realizada utilizando a tabela padrão do código genético universal. A conversão dos códons em aminoácidos foi automatizada por meio de scripts desenvolvidos em Python 3.11, utilizando a biblioteca Biopython (Cock et al., 2009) (Anexo 3). Com base nos aminoácidos resultantes, os SNPs foram classificados como mutações sinônimas (sem alteração no aminoácido), não sinônimas (substituição de aminoácido) ou nonsense (introdução de códon de parada prematuro).

Análises funcionais subsequentes foram realizadas com as ferramentas SIFT, PROVEAN e MutPred2 (Kumar et al., 2009; Choi et al., 2012; Pejaver et al., 2020).

4.10. Identificação de vias metabólicas

Devido à limitação das anotações funcionais específicas para tambaqui, a identificação das vias metabólicas foi realizada com base na inferência por ortologia, utilizando os genes ortólogos de *Danio rerio* (zebrafish), espécie modelo com genoma completamente anotado e amplamente representada nos principais bancos de dados funcionais.

Os genes candidatos contendo SNPs foram convertidos em seus respectivos ortólogos no zebrafish. As vias metabólicas associadas a esses genes foram identificadas com base nas anotações do banco KEGG (Kyoto Encyclopedia of Genes and Genomes), por meio da correspondência com os códigos KEGG Orthology (KO) disponíveis para *Danio rerio*.

5. RESULTADOS

5.1. Descoberta de SNPs

A extração de DNA genômico foi bem-sucedida em todos os 239 espécimes coletados, confirmada pela integridade das bandas de DNA observadas na eletroforese em gel de agarose a 1%. As concentrações de DNA variaram entre as amostras, mas todas apresentaram qualidade e quantidade adequadas para a construção das bibliotecas de sequenciamento.

O sequenciamento das bibliotecas gerou um total de 652 milhões de leituras single-end com 100 pares de bases (pb). O número de leituras obtidas por indivíduo variou entre 1,5 e 7,3 milhões, com uma média de 2,8 milhões de leituras por indivíduo. A identificação de SNPs utilizando o pipeline UNEAK resultou na detecção de 182.416 SNPs polimórficos entre as populações estudadas.

5.2. Alinhamento das sequências

Com a utilização do Entrez gerado foram alinhadas 165 sequencias (Anexo 4) em 114 genes do tambaqui relacionados a importantes funções imunológicas e de sinalização celular.

Quadro 1. Entrez do NCBI para consulta genômica no genoma de *Colossoma macropomum*.

```
(LOC118802146[All Fields] OR LOC118798050[All Fields] OR LOC118800974[All Fields] OR celsr2[All Fields] OR LOC118803720[All Fields] OR LOC118824019[All Fields] OR LOC118800885[All Fields] OR il12rb2[All Fields] OR LOC118809992[All Fields] OR epha2b[All Fields] OR grk5l[All Fields] OR LOC118816945[All Fields] OR LOC118827790[All Fields] OR LOC118818394[All Fields] OR spag9b[All Fields] OR LOC118811225[All Fields] OR traf1[All Fields] OR plgrkt[All Fields] OR sema3h[All Fields] OR il2rgb[All Fields] OR LOC118802193[All Fields] OR LOC118809994[All Fields] OR LOC118803715[All Fields] OR il16[All Fields] OR LOC118825996[All Fields] OR LOC118800838[All Fields] OR LOC118805325[All Fields] OR LOC118797800[All Fields] OR LOC118802834[All Fields] OR LOC118800265[All Fields] OR eps15l1a[All Fields] OR LOC118818210[All Fields] OR LOC118804321[All Fields] OR LOC118800174[All Fields] OR LOC118806494[All Fields] OR ptpn22[All Fields] OR LOC118812529[All Fields] OR LOC118815023[All Fields] OR il6st[All Fields] OR LOC118823508[All Fields] OR LOC118800361[All Fields] OR LOC118814455[All Fields] OR LOC118822342[All Fields] OR stat6[All Fields] OR unc5da[All Fields] OR LOC118801782[All Fields] OR LOC118813864[All Fields] OR tfr2[All Fields] OR eda2r[All Fields] OR esr2b[All Fields] OR LOC118827156[All Fields]) AND "Colossoma macropomum"[porgn]
```

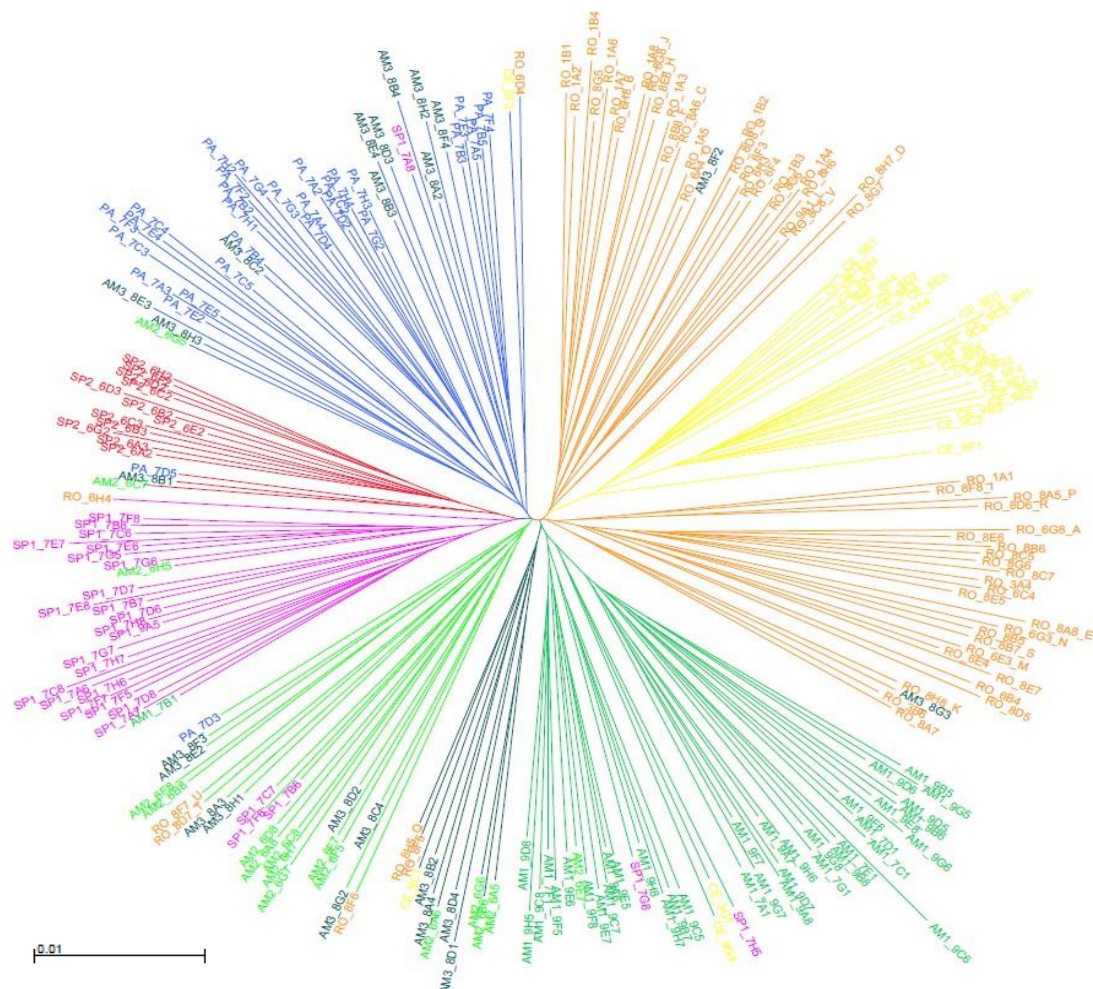
As sequências contendo os SNPs identificado mostraram alinhamentos relevantes no genoma do tambaqui. Os alinhamentos obtidos apresentaram valores de identidade e cobertura satisfatórios, por exemplo, o gene da quinase de receptor acoplado a proteínas G 5 (*grk5l*) e o receptor de interleucina 12, ambos com alinhamentos de alta identidade (97%-100%) e cobertura superior a 50%, confirmando a relação entre as sequências encontradas e os genes previamente anotados no genoma do tambaqui (Hilsdorf *et al.*, 2021).

5.3. Análise de parentesco

5.3.1 Análise de parentesco com base na árvore filogenética

A árvore filogenética evidenciou agrupamentos genéticos bem definidos entre os indivíduos das diferentes subpopulações analisadas (Figura 1).

Figura 1. Árvore filogenética baseada nas distâncias genéticas entre as populações de *Colossoma macropomum*.

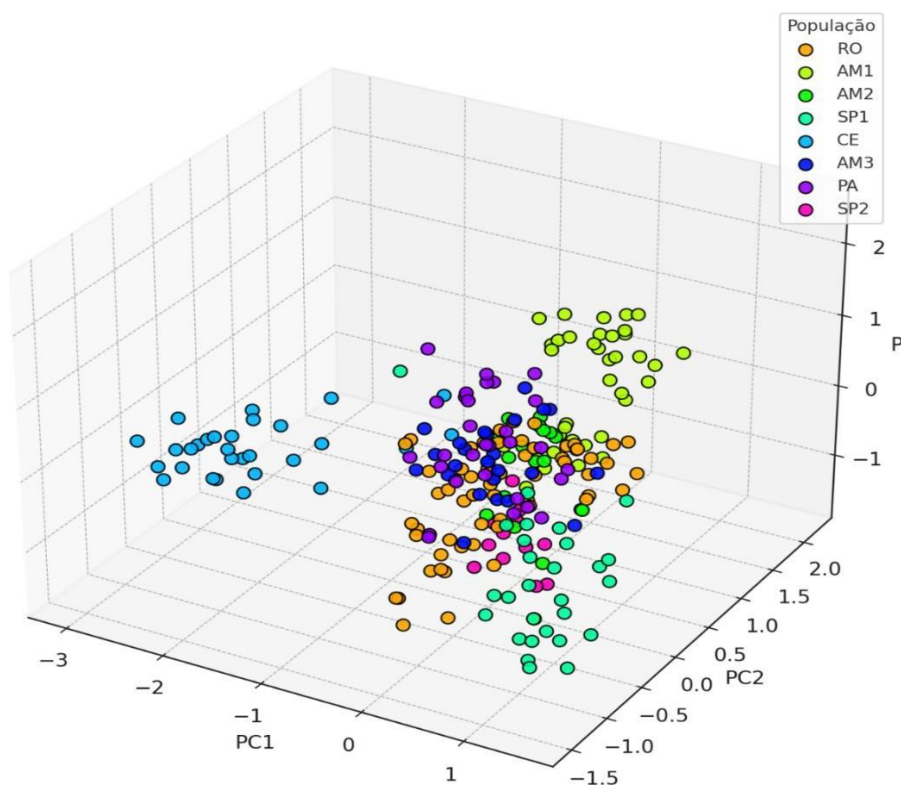


Os grupos RO, CE, PA, SP1, SP2, AM1, AM2 e AM3 formaram clados distintos, reforçando a estrutura genética observada previamente na análise de componentes principais (PCA). Os indivíduos das populações RO e CE apresentaram os ramos mais longos e separados dos demais grupos, indicando maior divergência genética. Já as subpopulações AM1, AM2 e AM3, bem como SP1 e SP2, apresentaram maior proximidade filogenética entre si, organizando-se em ramos agrupados. Esses padrões de agrupamento refletem a heterogeneidade genética entre as subpopulações avaliadas com base nos SNPs associados a processos imunes e de sinalização celular.

5.3.2 Análise de parentesco com base no PCA

A Análise de Componentes Principais (PCA), baseada em 165 SNPs localizados em genes relacionados a funções imunológicas e de sinalização celular, revelou a formação de agrupamentos distintos entre as subpopulações de *Colossoma macropomum* (Figura 2).

Figura 2. Gráfico tridimensional com base nos três componentes principais gerado a partir da Análise de Componentes Principais (PCA) de *Colossoma macropomum*.



Os indivíduos foram distribuídos no espaço tridimensional formado pelos três primeiros componentes principais (PC1, PC2 e PC3), evidenciando separação entre os grupos RO, PA, CE, AM1, AM2, AM3, SP1 e SP2. As subpopulações AM1, AM2 e AM3 apresentaram maior proximidade entre si, assim como SP1 e SP2. Por outro lado, os grupos RO e CE exibiram maior afastamento no espaço multivariado. A distribuição dos indivíduos nos componentes principais indicou variação genética entre as subpopulações analisadas.

5.4. Identificação de SNPs sob seleção natural

A análise dos valores de F_{ST} revelou uma ampla variação entre os SNPs analisados, com valores de F_{ST} variando de 0 a 0.22. Os SNPs com valores mais elevados de F_{ST} indicam uma maior diferenciação genética entre as populações analisadas, sugerindo que esses loci estão sob forte seleção local. Os SNPs com F_{ST} mais elevados incluem:

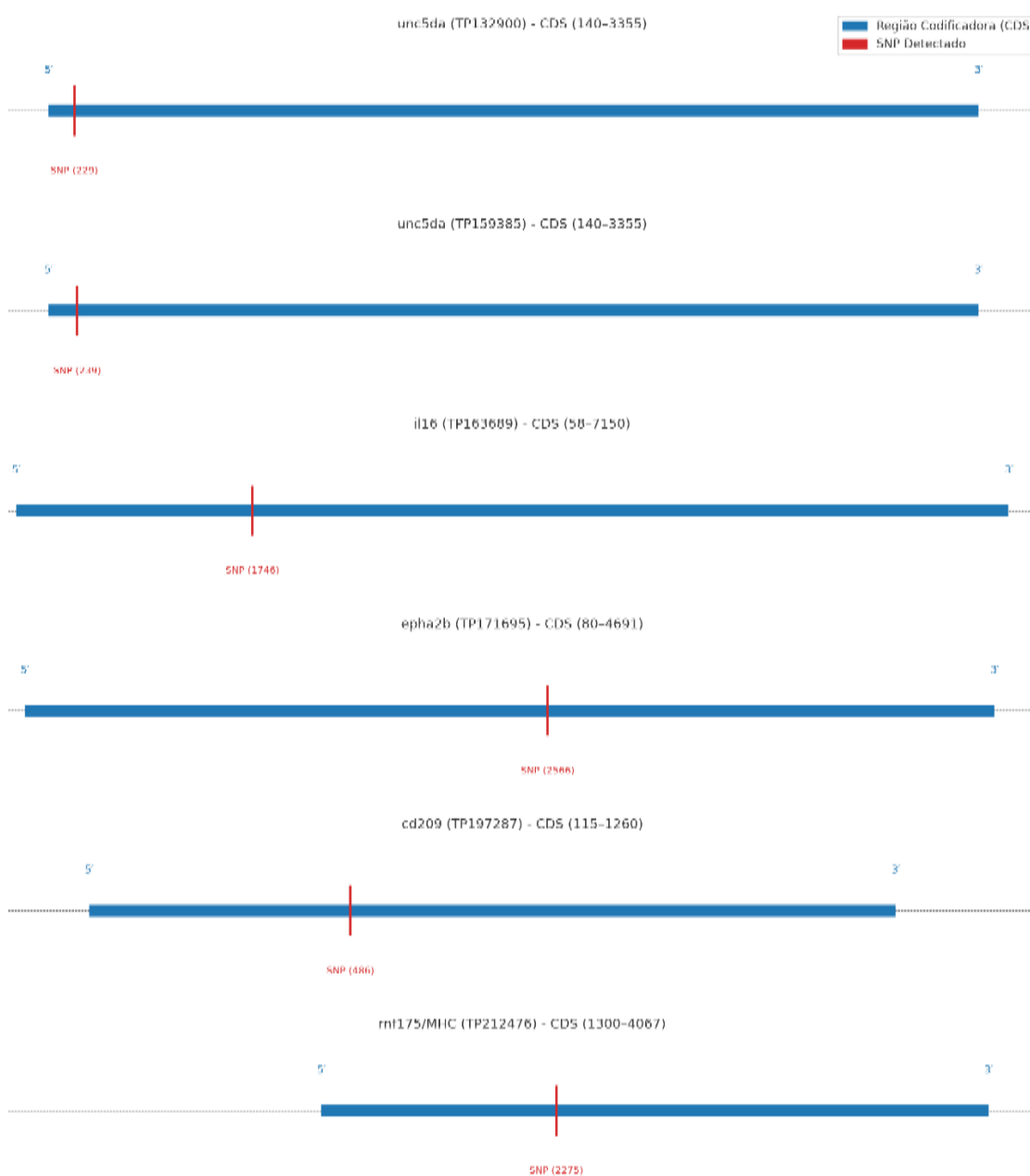
Tabela 2. SNPs candidatos, genes associados, F_{ST} e estatísticas de alinhamento.

Query	F_{ST}	Gene	Accession	Identity (%)	E-value
TP132900	0.167	<i>unc5da</i> (mRNA)	XM_036588857.1	98.44	9,00E-25
TP159385	0.222	<i>unc5da</i> (mRNA)	XM_036588857.1	100.0	2,00E-26
TP163689	0.193	<i>il16</i> (mRNA)	XM_036585977.1	98.36	4,00E-23
TP171695	0.183	<i>epha2b</i> (mRNA)	XM_036591044.1	98.36	4,00E-23
TP197287	0.154	<i>cd209</i> (mRNA)	XM_036582892.1	93.44	4,00E-18
TP212476	0.157	<i>MHC class I-related</i> (mRNA)	XM_036560351.1	94.55	2,00E-16
		<i>rnf175</i> (mRNA)	XM_036557338.1	92.19	4,00E-18

5.5 Localização dos SNPs nos genes

Os seis SNPs candidatos mapeados no genoma de tambaqui foram localizados dentro das regiões codificadoras (CDS) de seus respectivos genes, conforme definido pelos transcritos de referência disponíveis no NCBI (Figura 3), evidenciando sua distribuição ao longo dos transcritos e suas respectivas posições relativas entre as extremidades 5' e 3'.

Figura 3- Representação esquemática da localização dos SNPs nas regiões codificadoras (CDS) dos genes candidatos identificados em *Colossoma macropomum*



No gene *unc5da*, os SNPs TP132900 e TP159385 foram posicionados nas bases 229 e 239 do mRNA, respectivamente, ambos situados no terço inicial da CDS (140–3355). O SNP TP163689 foi mapeado na base 1746 do gene *il16*, cuja CDS compreende as posições 58 a 7150. Para *epha2b*, o SNP TP171695 foi localizado na base 2566, dentro de uma CDS anotada entre as posições 80 e 4691. O SNP TP197287, associado ao gene *cd209*, foi posicionado na base 486 do mRNA, entre as posições codificantes 115 e 1260. Já o SNP TP212476 foi identificado na base 2275 do gene *rnf175/MHC*, com CDS entre as posições 1300 e 4067.

5.6 Caracterização funcional dos SNPs candidatos

A caracterização funcional dos seis SNPs candidatos revelou a presença de diferentes tipos de mutações em genes relacionados à resposta imune e regulação celular em tabaqui (Tabela 3). No total, foram identificadas duas mutações sinônimas, três não sinônimas e duas do tipo nonsense.

Tabela 3. Caracterização dos SNPs candidatos quanto ao códon de referência e variante, aminoácidos correspondentes e tipo de mutação identificada nos genes associados.

SNP ID	Gene	Códon Ref	Códon Var	AA Ref	AA Var	Tipo de Mutação
TP132900	<i>unc5da</i>	CCT	CCC	Pro	Pro	Sinônima
TP159385	<i>unc5da</i>	GAG	AAG	Glu	Lys	Não Sinônima
TP163689	<i>il16</i>	TTA	TAA	Leu	STOP	Nonsense
TP171695	<i>epha2b</i>	GAG	AAG	Glu	Lys	Não Sinônima
TP197287	<i>cd209</i>	GGT	GAT	Gly	Asp	Não Sinônima
TP212476	<i>rnf175 / MHC</i>	CAA	TAA	Gln	STOP	Nonsense

Dois SNPs localizados no gene *unc5da* apresentaram características distintas: TP132900 corresponde a uma mutação sinônima, sem alteração no aminoácido codificado (Pro → Pro), enquanto TP159385 representa uma mutação não sinônima com substituição de Glu por Lys. No gene *il16*, o SNP TP163689 resultou em uma mutação nonsense, com a introdução de um códon

de parada prematuro (TAA), potencialmente afetando a tradução da proteína. O SNP TP171695, localizado em *epha2b*, também foi classificado como não sinônimo, promovendo a substituição do aminoácido Glu por Lys. Os SNPs TP197287 (*cd209*) e TP212476 (*rnf175/MHC*) apresentaram mutações do tipo não sinônima e nonsense, respectivamente.

Essas variações indicam o potencial impacto funcional diferencial dos SNPs nas proteínas codificadas, com possíveis implicações em fenótipos adaptativos e resistência a doenças.

5.7. Identificação de vias metabólicas

A análise funcional baseada nos ortólogos de *Danio rerio* permitiu a identificação das principais vias metabólicas associadas aos genes contendo SNPs candidatos.

Tabela 4. Identificação das principais vias metabólicas associadas aos genes identificados.

Gene	Via Metabólica (KEGG)
<i>unc5da</i>	Axon guidance; Apoptosis; Netrin signaling
<i>il16a</i>	Cytokine-cytokine receptor interaction; T cell activation
<i>epha2b</i>	Cell adhesion molecules (CAMs); Ephrin receptor signaling
<i>cd209e</i>	Pathogen recognition; Lectin pathway of complement activation
<i>rnf175</i>	Ubiquitin mediated proteolysis; MHC class I antigen processing

O gene *unc5da* foi associado a vias relacionadas à orientação axonal, apoptose e sinalização mediada por netrina. O ortólogo *il16a* foi mapeado nas vias de interação citocina–receptor de citocina e ativação de linfócitos T. O gene *epha2b* apresentou associação com moléculas de adesão celular (CAMs) e sinalização por receptores *Ephrin*, vias que desempenham papel essencial na comunicação célula-célula e organização tecidual. O gene *cd209e* está relacionado a vias de reconhecimento de patógenos e ativação da via das lectinas do sistema complemento. Por fim, *rnf175* foi associado a processos de proteólise mediada por ubiquitina e ao processamento de antígenos do MHC classe I.

6. DISCUSSÃO

A quantidade de variantes identificadas (182.416 SNPs) reflete não apenas a capacidade da plataforma GBS (*genotyping-by-sequencing*) em captar polimorfismos em regiões genômicas amplas, mesmo em espécies com poucas informações genéticas disponíveis na literatura, mas também sugere a presença de uma base genética complexa e diversificada nas populações analisadas (Liu et al., 2024). A heterogeneidade genética observada pode estar relacionada a múltiplos fatores, como o histórico de origem dos estoques, práticas de manejo reprodutivo distintas entre as regiões, e possíveis eventos de adaptação local ao ambiente de cultivo (Gjedrem et al., 2012).

Estudos recentes demonstram a relevância do uso de tecnologias de genotipagem em larga escala para desvendar a diversidade genética e a estrutura populacional em peixes. Por exemplo, Mastrochirico-Filho et al. (2020) desenvolveram um mapa de ligação SNP e realizaram um estudo de associação genômica para resistência à *Aeromonas hydrophila* em pacu. De forma complementar, Landínez-García, Narváez e Márquez (2020) investigaram a genética populacional de *Prochilodus magdalenae* utilizando marcadores moleculares específicos, enquanto Sahoo et al. (2023) revelaram, por meio de genotipagem por sequenciamento, a diversidade genética e a estrutura populacional do *Labeo catla*.

Nossos achados confirmam que nível de diversidade genética encontrada em tambaqui é compatível com a de outras espécies de interesse aquícola, o que ressalta a sua importância como modelo regional para estudos genômicos aplicados. Esses resultados também reafirmam o valor dos marcadores SNP como ferramentas robustas para investigação da estrutura genética populacional e para aplicação em programas de seleção assistida por marcadores (MAS) (Yáñez et al., 2015).

A resolução genômica proporcionada pelos SNPs permite não apenas a inferência de relações filogenéticas e padrões de estrutura populacional com maior precisão, mas também viabiliza a identificação de variantes associadas a características adaptativas e produtivas (Banger & Goddard, 2015).

A estrutura genética observada nas análises filogenética e de componentes principais (PCA) revela padrões consistentes de diferenciação entre as subpopulações de tambaqui. A formação de clados bem definidos para grupos como RO, CE, PA, SP1, SP2, AM1, AM2 e AM3 confirma uma compartimentalização genética clara, indicando que essas populações seguem trajetórias evolutivas relativamente independentes. Essa diferenciação ocorre mesmo entre estoques com origem comum, sugerindo que fatores posteriores à fundação, como isolamento geográfico, manejo reprodutivo e ausência de intercâmbio genético, foram determinantes para o surgimento de perfis genéticos distintos (Waples & Gaggiotti, 2006).

A ausência de fluxo gênico recente entre essas subpopulações é coerente com as práticas produtivas e reprodução frequentemente adotadas na aquicultura, em que cada unidade produtiva mantém seu plantel de forma isolada (Gjedrem et al., 2012). Com o tempo, esse isolamento pode levar à fixação de alelos por deriva genética, especialmente quando há uso restrito de reprodutores ou ausência de renovação genética. Essa hipótese é sustentada por Hrbek et al. (2005), que demonstraram que populações de peixes amazônicos, mesmo derivadas de um ancestral comum, tendem a divergir geneticamente em função do isolamento e da deriva, principalmente quando submetidas a pressões ambientais distintas.

Além disso, os grupos RO e CE destacam-se por apresentarem os ramos mais longos na árvore filogenética, o que indica maior acúmulo de diferenças genéticas. Esse padrão pode refletir não apenas a ausência de fluxo gênico, mas também processos históricos de fundação baseados em um número restrito de indivíduos, o que acarreta gargalos genéticos, ou ainda a exposição a pressões seletivas locais, como diferenças ambientais, sanitárias ou nutricionais, que afetam a frequência alélica ao longo do tempo. Tais mecanismos são amplamente descritos na literatura como motores de estruturação genética em ambientes de cultivo (Urrea-Rojas et al., 2020).

Esse cenário também é análogo ao descrito por Santos et al. (2007), que observaram que a fragmentação de habitats aquáticos e a formação de ambientes isolados, naturais ou artificiais, contribuem para o surgimento de linhagens geneticamente diferenciadas, mesmo entre populações originalmente

conectadas. Aplicado ao contexto da piscicultura, isso ressalta a importância de monitorar e gerenciar a diversidade genética para evitar perdas de variabilidade e promover a sustentabilidade genética das populações em cativeiro (Gjedrem, 2000).

Portanto, os resultados obtidos indicam que a estrutura genética atual das populações de tambaqui de cativeiro é produto de uma combinação entre origem compartilhada e históricos independentes de manejo, e reforçam a necessidade de estratégias de conservação genética que evitem a erosão da diversidade intraespecífica, essencial para a adaptação futura dos estoques a novos desafios ambientais e produtivos.

A variação nos valores de FST, que neste estudo alcançou até 0,22, fornece uma indicação robusta da existência de diferenciação genética entre as subpopulações analisadas. Valores elevados de FST geralmente refletem restrições ao fluxo gênico e indicam a ação de forças evolutivas distintas entre os grupos, como deriva genética ou seleção natural. No contexto deste trabalho, os SNPs com FST mais elevados foram encontrados em genes funcionalmente associados à imunidade e sinalização celular, regiões do genoma que estão frequentemente sob influência de seleção direcional ou balanceada em resposta a pressões ambientais, como patógenos, estresse e condições de produção (Storz, 2005).

Segundo Beaumont e Balding (2004), SNPs com valores de FST elevados representam candidatos a loci sob seleção local, uma vez que sua diferenciação entre populações excede a esperada por deriva genética neutra. No presente estudo, esses marcadores foram mapeados em genes como *unc5da*, *il16*, *epha2b*, *cd209* e *rnf175*, todos envolvidos em processos fundamentais de resposta imune e interação célula-célula. Essa associação funcional fortalece a hipótese de que tais variantes não estão apenas estruturando a diversidade genética, mas também refletindo adaptações diferenciais a desafios imunológicos nos ambientes de criação como exposição a patógenos específicos ou estratégias de manejo distintas entre os viveiros (Yáñez et al., 2015).

Além disso, para espécies cultivadas como o tambaqui, que frequentemente enfrentam condições artificiais de confinamento e alta

densidade populacional, *loci* sob seleção associados à resistência a doenças adquirem ainda mais importância prática. Eles representam potenciais alvos para programas de seleção assistida por marcadores (MAS), permitindo o desenvolvimento de linhagens mais resilientes, com maior adaptabilidade às condições de cultivo intensivo. Assim, a identificação de SNPs com alto FST em genes com função imunológica não apenas esclarece processos microevolutivos em populações cultivadas, mas também fornece subsídios valiosos para a aplicação direta em estratégias de melhoramento genético (Bangera & Goddard, 2015).

A caracterização funcional dos SNPs revelou que as variantes identificadas não estão distribuídas aleatoriamente ao longo do genoma, mas se concentram em regiões codificantes (CDS), o que amplia sua relevância funcional. Alterações em regiões codificadoras tendem a ter maior impacto biológico, pois podem modificar diretamente a estrutura primária das proteínas, afetar a estabilidade do RNA mensageiro, a eficiência da tradução ou a interação com moléculas reguladoras (Kimchi-Sarfaty et al., 2007; Sauna; Kimchi-Sarfaty, 2011; Plotkin; Kudla, 2011). No gene *unc5da*, dois SNPs foram mapeados no terço inicial da CDS, uma região crítica tanto para a formação da estrutura N-terminal da proteína quanto para a eficiência de tradução. Alterações nesta região podem comprometer o correto dobramento da proteína ou modificar sua capacidade de interação com os ligantes, como as netrinas (Mehlen & Mazelin, 2003).

O SNP TP132900, embora sinônimo, pode ainda influenciar o processo de splicing, a estabilidade do RNA mensageiro ou a taxa de tradução, como discutido por Parmley e Hurst (2007), especialmente se ocorrer em códons preferenciais ou em regiões regulatórias internas. Já TP159385, no mesmo gene, resulta em uma mutação não sinônima que substitui o aminoácido ácido glutâmico (Glu) por lisina (Lys), promovendo uma mudança de carga que pode afetar significativamente a interação da proteína com netrinas e, por consequência, alterar sua função em processos inflamatórios e migratórios celulares (Ziegler & Schlegel, 2021).

Em *il16*, o SNP TP163689 introduz um códon de parada prematuro (TAA), caracterizando uma mutação do tipo nonsense. Tal alteração é tipicamente

associada à perda de função da proteína, uma vez que pode resultar na produção de um peptídeo truncado e, portanto, inativo ou instável. Considerando que *IL-16* é uma citocina chave na quimiotaxia de células *T CD4+* e regulação da resposta imune adaptativa, a perda de sua função pode comprometer a resposta inflamatória e a imunocompetência do organismo (Mathy et al., 2000; Zhang et al., 2023).

O SNP TP171695 em *epha2b* promove uma substituição não sinônima de Glu para Lys em um domínio que pode estar envolvido na atividade tirosina-quinase da proteína. Como os receptores *Eph* são fundamentais para a comunicação célula-célula e controle da integridade epitelial, mutações nesse gene têm potencial para afetar processos de adesão celular, apoptose e até mesmo resposta a infecções (Pasquale, 2010).

Os SNPs TP197287 e TP212476, localizados nos genes *cd209* e *rnf175/MHC*, também apresentam mutações potencialmente funcionais. TP197287 promove a troca de um aminoácido glicina por ácido aspártico, o que pode alterar a conformação do sítio de ligação do receptor tipo lectina *CD209*, responsável pelo reconhecimento de patógenos e ativação da via do complemento (Zhou et al., 2010). Já TP212476 resulta em uma mutação nonsense, levando à introdução de um códon de parada prematuro em uma região importante da proteína codificada por *rnf175*, envolvida na via de processamento de antígenos do *MHC* classe I. Esse tipo de mutação pode comprometer a apresentação de peptídeos aos linfócitos T citotóxicos, enfraquecendo a capacidade do sistema imune de eliminar células infectadas ou alteradas (Yamaguchi & Dijkstra, 2019).

Coletivamente, esses achados apontam para um cenário em que as variantes genéticas identificadas, além de estruturar a diversidade populacional, podem influenciar diretamente mecanismos biológicos essenciais à integridade funcional do sistema imune (Magnadóttir, 2006). Tais informações são particularmente valiosas para a piscicultura, pois oferecem alvos genômicos com potencial para aplicação em programas de seleção assistida voltados à resistência a doenças, uma das maiores preocupações na produção intensiva de tambaqui.

As vias metabólicas associadas aos genes que abrigam SNPs funcionais reforçam a importância biológica das variantes identificadas e sustentam sua potencial implicação em processos críticos para a fisiologia e adaptabilidade de tambaqui. Os genes analisados estão envolvidos em vias fundamentais para a sinalização celular, integridade tecidual e resposta imunológica, aspectos essenciais para a homeostase e resistência a agentes patogênicos, especialmente em sistemas de produção intensiva, onde peixes são frequentemente expostos a estressores ambientais e agentes infecciosos (Magnadóttir, 2006).

O gene *unc5da*, por exemplo, participa da via de sinalização mediada por netrina, atuando como receptor de repulsão na orientação axonal e na migração celular. Além de seu papel no desenvolvimento neural, *unc5da* tem sido associado à modulação da resposta inflamatória, especialmente por sua ação em tecidos sob hipóxia ou lesão (Mehlen et al., 1998; Ziegler & Schlegel, 2021). Já o ortólogo *il16* está integrado à via de sinalização de citocinas e à ativação de linfócitos T, desempenhando função-chave no recrutamento e na ativação de células imunes durante processos inflamatórios, como descrito por Chama et al. (2021) e Mathy et al. (2000). O impacto de mutações nesse gene pode comprometer diretamente a eficiência da resposta imune adaptativa.

No caso de *epha2b*, seu envolvimento na via de sinalização Ephrin o coloca como elemento importante na adesão celular e na manutenção da arquitetura tecidual. A sinalização mediada por receptores Eph é essencial para a regeneração e organização de tecidos epiteliais, além de participar da resposta a lesões e estímulos inflamatórios (Pasquale, 2010). *Cd209e*, por sua vez, está vinculado a vias de reconhecimento de padrões moleculares associados a patógenos (PAMPs), e à ativação da via das lectinas do sistema complemento, configurando um componente-chave da imunidade inata (Zhou et al., 2010). Por fim, *rnf175* participa da via de proteólise mediada por ubiquitina e do processamento de antígenos para apresentação via MHC classe I, sendo fundamental para a ativação da resposta imune celular (Kloetzel, 2004; Yamaguchi & Dijkstra, 2019).

A análise conjunta dessas evidências sugere que os SNPs identificados não apenas refletem a diversidade genética entre os estoques de tambaqui, mas

também possuem implicações funcionais relevantes para a adaptação das populações a contextos ambientais e sanitários específicos. A associação de variantes com vias metabólicas críticas confere robustez ao uso desses marcadores em estratégias de seleção genômica, com vistas ao desenvolvimento de linhagens mais resilientes e adaptadas às condições da aquicultura nacional.

7. CONSIDERAÇÕES FINAIS

O presente estudo revelou uma ampla diversidade genética entre populações cultivadas de tambaqui. A estruturação genética observada por meio de análises filogenéticas e de componentes principais demonstrou uma diferenciação consistente entre as subpopulações, em especial nos grupos RO e CE, sugerindo ausência de fluxo gênico recente e possíveis efeitos de deriva genética e manejo reprodutivo isolado.

A análise funcional dos SNPs destacou variantes localizadas em regiões codificadoras de genes associados a processos imunológicos e de sinalização celular, como *unc5da*, *il16*, *epha2b*, *cd209* e *rnf175*. As mutações detectadas incluem alterações sinônimas, não sinônimas e nonsense, com potencial para impactar diretamente a função proteica e, por consequência, a resposta imune e a adaptação fisiológica dos organismos. A integração dessas variantes a vias metabólicas críticas, como sinalização por netrina, ativação de linfócitos T, reconhecimento de patógenos e apresentação antigênica, reforça seu valor como candidatos a marcadores funcionais.

Esses achados contribuem significativamente para o entendimento da base molecular da resistência a doenças em tambaqui, oferecendo subsídios para o desenvolvimento de estratégias de seleção genômica voltadas à aquicultura sustentável. O uso desses SNPs como marcadores em programas de melhoramento pode favorecer a obtenção de linhagens mais resilientes e adaptadas aos desafios sanitários dos sistemas de produção intensiva.

Recomenda-se, para estudos futuros, a validação funcional das variantes mais promissoras por meio de abordagens experimentais e transcriptômicas, bem como a integração de dados genômicos com informações fenotípicas e zootécnicas. Tais avanços permitirão a consolidação de ferramentas genéticas aplicadas à piscicultura nacional, com impacto direto sobre produtividade, sanidade e conservação dos recursos genéticos nativos.

8. REFERÊNCIAS BIBLIOGRÁFICAS

AGUDELO, John Fredy Gómez et al. Genomic selection signatures in farmed *Colossoma macropomum* from tropical and subtropical regions in South America. **Evolutionary Applications**, [s. l.], v. 15, n. 4, p. 679, 2022. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9046916/>. Acesso em: 30 out. 2024.

AGUIAR, Jonas da Paz et al. Loss of genetic variability in the captive stocks of tambaqui, *Colossoma macropomum* (Cuvier, 1818), at breeding centres in Brazil, and their divergence from wild populations. **Aquaculture Research**, [s. l.], v. 49, n. 5, p. 1914–1925, 2018.

ALLENDORF, F. W., HOHENLOHE, P. A., LUIKART, G. Genomics and the future of conservation genetics. **Nature Reviews Genetics**, v.11, n.10, p.697-709, 2010.

ARIDE, P. H. R.; ROUBACH, R.; VAL, A. L. Tolerance response of tambaqui *Colossoma macropomum* (Cuvier) to water pH. **Aquaculture Research**, v.38, p. 588-594, 2007.

BAIRD, N. A., ETTER, P. D., ATWOOD, T. S., CURREY, M. C., SHIVER, A. L., LEWIS, Z. A., JOHNSON, E. A. Rapid SNP discovery and genetic mapping using sequenced RAD markers. **PloS One**, v.3, n.10, e3376, 2008.

BARABÁSI, A.L., GULBAHCE, N., LOSCALZO, J. Network medicine: a network-based approach to human disease. **Nature Reviews Genetics**, v.12, n.1, p.56-68, 2011.

BEAUMONT, M. A.; BALDING, D. J. Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, v. 13, n. 4, p. 969-980, 2004.

BANGERA, R.; GODDARD, M. E. Genomic selection: the future of genetic improvement in aquaculture. *Frontiers in Genetics*, v. 6, p. 104, 2015.

BERNATCHEZ, L., WELLENREUTHER, M., ARANEDA, C., ASHTON, D. T., BARTH, J. M., BEACHAM, T. D., WITHLER, R. E. Harnessing the power of genomics to secure the future of seafood. **Trends in Ecology & Evolution**, v.32, n.9, p.665-680, 2017.

BISHOP, S. C., MACKENZIE, K. Genetic management strategies for controlling infectious diseases in livestock populations. **Genetics Selection Evolution**, v.35, Suppl 1, S3-S17, 2003.

BRITO, N. C.; SANTOS, C. A.; OLIVEIRA, E. R. et al. Expressão de genes do sistema imune em tambaqui (*Colossoma macropomum*) após infecção experimental por *Aeromonas hydrophila*. *Revista Brasileira de Zootecnia*, v. 41, n. 7, p. 1500-1508, 2012.

CHAGAS, E. C.; MACIEL, P. O.; AQUINO-PEREIRA, S. L. Infecções por acantocéfalos: um problema para a produção de peixes. In: *Aquicultura no Brasil: novas perspectivas*. São Carlos: Editora Pedro & João, 2015.

CHAGAS, E. C.; VAL, A. L. Efeito da vitamina C no ganho de peso e em parâmetros hematológicos de tambaqui. **Pesquisa Agropecuária Brasileira**, v.38, p. 397-402, 2003.

CHOI, Y.; SIMS, G. E.; MURPHY, S.; MILLER, J. R.; CHAN, A. P. Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE*, v. 7, n. 10, p. e46688, 2012.

CLARK, M. J., CHEN, R., & LAM, H. Y. An introduction to next-generation sequencing for in vitro fertilization. In **Next Generation Sequencing in Cancer Research** (pp. 191-205). Springer, New York, NY, 2011.

CLARK, T.A., SCHWEITZER, P.A., CHEN, T.X., STAPLES, M.K., LU, G., WANG, H., et al. Discovery of tissue-specific exons using comprehensive human exon microarrays. **Genome Biology**, v.8, n.4, R64, 2011.

COCK, P. J. A.; ANTÃO, T.; CHANG, J. T.; CHAPMAN, B. A.; COX, C. J.; DALKE, A.; FRIEDBERG, I.; HAMELRYCK, T.; KAUFF, F.; WILCZYNSKI, B.; DE HOON, M. J. L. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, v. 25, n. 11, p. 1422-1423, 2009.

CAPORASO, J.G., LAUBER, C.L., WALTERS, W.A., BERG-LYONS, D., HUNTLEY, J., FIERER, N., et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. **Proceedings of the National Academy of Sciences**, v.108, Suppl 1, p.4516-4522, 2011.

ELSHIRE, Robert J. et al. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. **PLoS ONE**, [s. l.], v. 6, n. 5, p. e19379, 2011. Disponível em: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3087801&tool=pmcentrez&rendertype=abstract>. Acesso em: 9 jul. 2014.

FAO (2020). The State of World Fisheries and Aquaculture 2020. **Sustainability in action**. Rome, FAO.

FIGUEIREDO, H. C. P.; LEAL, C. A. G. Tecnologias aplicadas em sanidade de peixes. **Revista Brasileira de Zootecnia**, v.37, suplemento especial, p. 08-14, 2008.

FUKUOKA, H. Recent Progress in Research on the Prevention and Control of Viral Diseases in Aquaculture. **Fish Pathology**, v. 54, n. 1, p. 3-10, 2019.

GJEDREM, T. Genetic improvement of aquaculture species: possibilities and challenges. *Aquaculture Research*, v. 31, n. 3, p. 157-172, 2000.

GJEDREM, T. Genetic improvement for the development of efficient global aquaculture: A personal opinion review. ***Aquaculture***, v. 344–349, n. 44, p. 12–22, 2012.

GJEDREM, T., ROBINSON, N., & RYE, M. The importance of selective breeding in aquaculture to meet future demands for animal protein: a review. ***Aquaculture***, v.350-353, p.117-129, 2012.

GJEDREM, T. The first family-based breeding program in aquaculture. ***Reviews in Aquaculture***, v.1, n.1, p.2-15, 2009.

GJØEN, H. M., KJØGLUM, S., & BENTSEN, H. B. Genetic variation in resistance to bacterial and viral diseases in fish. ***Aquaculture***, v.459, p. 65-74, 2016.

GOMES, Levy De Carvalho et al. Cage culture of tambaqui (*Colossoma macropomum*) in a central Amazon floodplain lake. ***Aquaculture***, [s. l.], v. 253, n. 1–4, p. 374–384, 2006.

GOODWIN, S., MCPHERSON, J. D., & MCCOMBIE, W. R. Coming of age: ten years of next-generation sequencing technologies. ***Nature Reviews Genetics***, v.17, n.6, p. 333-351, 2016.

GÓMEZ, D., SUNYER, J. O., & SALINAS, I. The mucosal immune system of fish: the evolution of tolerating commensals while fighting pathogens. ***Fish & Shellfish Immunology***, v.35, n.6, p. 1729-1739, 2013.

HANCZ, Csaba. Performance of the amazonian tambaqui, *Colossoma macropomum*, in pond polyculture. ***Aquacultural Engineering***, [s. l.], v. 12, n. 4, p. 245–254, 1993.

HILSDORF, Alexandre Wagner Silva et al. Genome assembly and annotation of the tambaqui (*Colossoma macropomum*): an emblematic fish of the Amazon River Basin. ***GigaByte***, [s. l.], v. 2021, p. gigabyte29, 2021. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9650303/>. Acesso em: 20 nov. 2024.

HOUSTON, R. D.; TAGGART, J. B.; CEZARD, T. et al. Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). *Genetics*, v. 178, n. 3, p. 1109-1115, 2008.

HOUSTON, R. D. Future directions in breeding for disease resistance in aquaculture species. ***Revista Brasileira de Zootecnia***, v.46, n.6, p. 545-551, 2017.

HOUSTON, R. D., MACQUEEN, D. J., & ROBLED, D. Disease resistance in fish: genetics, genomics and new opportunities. **Trends in Genetics**, v.36, n.11, p. 822-832, 2020.

HOUSTON, R. D., MACQUEEN, D. J., & ROBLED, D. Genetic and genomic technologies for sustainable aquaculture. **Nature Reviews Genetics**, v.21, n.11, p. 665-681, 2020.

HRBEK, Tomas et al. Population genetic analysis of *Arapaima gigas*, one of the largest freshwater fishes of the Amazon basin: Implications for its conservation. **Animal Conservation**, [s. l.], v. 8, n. 3, p. 297–308, 2005.

IDEKER, T., GALITSKI, T., & HOOD, L. A new approach to decoding life: systems biology. **Annual Review of Genomics and Human Genetics**, v.2, n.1, p. 343-372, 2001.

JERÔNIMO, G. T.; SPECK, G. M.; MOREIRA, H. L. M. *Aeromonas hydrophila* and *Flavobacterium columnare*: Challenges in the Control of Bacterial Diseases in Fish Farming. **Boletim do Instituto de Pesca**, v. 43, n. 3, p. 1-20, 2017.

KHAW, H. L., PONZONI, R. W., YEE, H. Y., AZIZ, M. A., & MULDER, H. A. Genetic evaluation of the Genetically Improved Farmed Tilapia (GIFT) strain over ten generations of selection in Malaysia. **Aquaculture**, v.448, p. 273-281, 2015.

KIMCHI-SARFATY, C.; OH, J. M.; KIM, I. W.; SAUNA, Z. E.; CALCAGNO, A. M.; AMBUDKAR, S. V.; GOTTESMAN, M. M. A “silent” polymorphism in the MDR1 gene changes substrate specificity. *Science*, v. 315, n. 5811, p. 525-528, 2007.

KLOETZEL, P. M. The proteasome and MHC class I antigen processing. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, v. 1695, n. 1-3, p. 225-233, 2004.

KOBOLDT, D.C., STEINBERG, K.M., LARSON, D.E., WILSON, R.K., & MARDIS, E.R. The next-generation sequencing revolution and its impact on genomics. **Cell**, v.155, n.1, p. 27-38, 2013.

KRASNOV, A., KILENG, Ø., SKUGOR, S., JØRGENSEN, S. M., AFANASYEV, S., & TIMMERHAUS, G. Genomic analysis of the host response to nervous necrosis virus in Atlantic cod (*Gadus morhua*) brain. **Molecular Immunology**, v.73, p. 10-20, 2016.

KUMAR, P.; HENIKOFF, S.; NG, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, v. 4, n. 7, p. 1073-1081, 2009.

LAFFERTY, K. D. et al. Infectious diseases affect marine fisheries and aquaculture economics. **Annual Review of Marine Science**, v. 7, p. 471-496, 2015.

LANDE, R., ARNOLD, S.J., & TURELLI, M. Evolution at the intersection of genetics and ecology. **The American Naturalist**, v.186, S1-S6, 2015.

LANDÍNEZ-GARCÍA, Ricardo M.; NARVÁEZ, Juan Carlos; MÁRQUEZ, Edna J. Population genetics of the freshwater fish *Prochilodus magdalenae* (Characiformes: Prochilodontidae), using species-specific microsatellite loci. *PeerJ*, v. 8, p. e10327, 2020.

LARHAMMAR, M., RISINGER, C., & LÖVGREN, A. The importance of proper bioinformatics analysis and clinical interpretation of tumor genome sequencing: a case study of undifferentiated sarcoma and a constitutional pathogenic BRCA2 mutation and an MLH1 variant of uncertain significance. **Cold Spring Harbor Molecular Case Studies**, v.4, n.5, p. a002410, 2018.

LI, Y., LIU, W., YU, X., & ZHOU, G. Polymorphisms of the immune-related genes and their associations with disease resistance in teleost. **Fish & Shellfish Immunology**, v.98, p. 49-60, 2020.

LIEN, S. et al. The Atlantic salmon genome provides insights into rediploidization. **Nature**, v. 533, n. 7602, p. 200-205, 2016.

LIMA, Adriana Ferreira et al. The effect of nursery production system (in cage and pond) on performance, health status, and plankton ingestion of the low trophic level fish tambaqui, *Colossoma macropomum*. **Aquaculture**, [s. l.], v. 586, p. 740721, 2024.

LIU, Dingrui et al. Identification of SNPs and candidate genes associate with growth performance in all-female mandarin fish (*Siniperca chuatsi*) by a genome-wide association study. **Aquaculture**, [s. l.], v. 586, 2024.

LIU, Zhanjiang; ZHOU, Tao; GAO, Dongya. Genetic and epigenetic regulation of growth, reproduction, disease resistance and stress responses in aquaculture. [s. l.], 2022.

MAGNADÓTTIR, H. Innate immunity of fish (overview). **Fish & Shellfish Immunology**, v. 20, n. 2, p. 137-151, 2006.

MARDIS, E. R. The impact of next-generation sequencing technology on genetics. **Trends in Genetics**, v.24, n.3, p. 133-141, 2008.

MASTROCHIRICO-FILHO, V. A., URBINATI, E. C., & ROMANO, L. A. Stress response to handling in the tambaqui *Colossoma macropomum* (Cuvier, 1816) is time dependent. **Fish Physiology and Biochemistry**, v.45, n.1, p. 297-306, 2019.

MASTROCHIRICO-FILHO, Vito A. et al. Development of a SNP linkage map and genome-wide association study for resistance to *Aeromonas hydrophila* in pacu (*Piaractus mesopotamicus*). *BMC Genomics*, v. 21, p. 1-13, 2020.

MATHY, N. L. et al. Interleukin-16 stimulates the expression and production of pro-inflammatory cytokines by human monocytes. **Immunology**, [s. l.], v. 100, n. 1, p. 63–69, 2000.

MEHLEN, P.; MAZELIN, L. The dependence receptors DCC and UNC5H as a link between neuronal guidance and survival. **Nature Reviews Neuroscience**, v. 4, n. 7, p. 423-436, 2003.

MEHLEN, P. et al. The DCC gene product induces apoptosis by a mechanism requiring receptor proteolysis. **Nature** 1998 395:6704, [s. l.], v. 395, n. 6704, p. 801–804, 1998. Disponível em: <https://www.nature.com/articles/27441>. Acesso em: 14 out. 2024.

MEROLA, Nino; DE SOUZA, J. Henrique. Preliminary studies on the culture of the pacu, *Colossoma mitrei*, in floating cages: Effect of stocking density and feeding rate on growth performance. **Aquaculture**, [s. l.], v. 68, n. 3, p. 243–248, 1988.

MERRIFIELD, D. L. et al. The current status and future focus of probiotic and prebiotic applications for salmonids. **Aquaculture**, v. 302, n. 1-2, p. 1-18, 2010.

NEIRA, Roberto. Breeding in Aquaculture Species : Genetic Improvement Programs in Developing Countries. **9th World Congress on Genetics Applied to Livestock Production**, [s. l.], n. August, p. 8, 2009.

NEWMAN, S. G. Bacterial vaccines for fish. **Annual Review of Fish Diseases**, v. 4, p. 145-185, 2014.

NOGA, E. J. Fish Disease: diagnosis and treatment. St Louis: **Mosby-Year Book**, 367 p., 1996.

NOWAK, B. F., ARCHIBALD, J. M., & IGLESIAS, R. Diagnosis and control of fish diseases: a review of the role of next-generation sequencing. **Reviews in Aquaculture**, v.13, n.3, p. 1601-1620, 2021.

NOWAK, B., THOMPSON, K., & AUSTIN, B. Genetic polymorphism in fish immune response genes: potential association with disease resistance. **Reviews in Aquaculture**, v.13, n.3, p. 1485-1503, 2021.

NUNES, José de Ribamar da Silva et al. Large-scale SNP discovery and construction of a high-density genetic map of *Colossoma macropomum* through genotyping-by-sequencing. **Scientific Reports** 2017 7:1, [s. l.], v. 7, n. 1, p. 1–11, 2017. Disponível em: <https://www.nature.com/articles/srep46112>. Acesso em: 12 jul. 2021.

ONAKA, E. M.; MORAES, F. R. Enfermedades **Parasitárias de Peixes**. , v.1, n.1, 2004.

PASQUALE, Elena B. Eph receptors and ephrins in cancer: bidirectional signalling and beyond. **Nature Reviews Cancer** 2010 10:3, [s. l.], v. 10, n. 3, p. 165–180, 2010. Disponível em: <https://www.nature.com/articles/nrc2806>. Acesso em: 12 jan. 2025.

PARMLEY, C. R.; HURST, L. D. How common are synonymous mutations? **Trends in Genetics**, v. 23, n. 10, p. 362-365, 2007.

PEJAVER, V. et al. MutPred2: inferring the molecular and phenotypic impact of amino acid variants. **Nature Communications**, v. 11, p. 5918, 2020.

PLOTKIN, J. B.; KUDLA, G. Synonymous but not the same: the causes and consequences of codon bias. **Nature Reviews Genetics**, v. 12, p. 32-42, 2011.

POTTS, R. W. A., GUTIERREZ, A. P., PENALOZA, C. S., REGAN, T., BEAN, T. P., & HOUSTON, R. D. Potential of genomic technologies to improve disease resistance in molluscan aquaculture. **Philosophical Transactions of the Royal Society B: Biological Sciences**. 2021.

RAUTA, P. R., NAYAK, B., & DAS, S. Immune system and immune responses in fish and their role in comparative immunity study: a model for higher organisms. **Immunology Letters**, v.148, n.1, p. 23-33, 2012.

ROCKMAN, M. V.; SKROVANEK, S. S.; KRUGLYAK, L. Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. **Science**, v. 319, n. 5862, p. 459-463, 2008.

SAHOO, Bismay et al. Genetic diversity and genome-scale population structure of wild Indian major carp, *Labeo catla* (Hamilton, 1822), revealed by genotyping-by-sequencing. **Frontiers in Genetics**, v. 14, p. 1166385, 2023.

SANTOS, M. C. F.; RUFFINO, M. L.; FARIAS, I. P. High levels of genetic variability and panmixia of the tambaqui *Colossoma macropomum* (Cuvier, 1816) in the main channel of the Amazon River. **Journal of Fish Biology**, [s. l.], v. 71, n. sa, p. 33–44, 2007. Disponível em: <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1095-8649.2007.01514.x>. Acesso em: 26 maio 2014.

SAUNA, Z. E.; KIMCHI-SARFATY, C. Understanding the contribution of synonymous mutations to human disease. **Nature Reviews Genetics**, v. 12, p. 683-691, 2011.

SECOMBES, C. J., WANG, T., & BIRD, S. The interleukins of fish. **Developmental & Comparative Immunology**, v.90, p. 87-102, 2019.

SHEN, Y. et al. Fish & Shellfish Immunology Characterization of the duodenase-1 gene and its associations with resistance to *Streptococcus agalactiae* in hybrid

tilapia (*Oreochromis* spp.). **Fish & Shellfish Immunology**, v. 45, p. 717–724, 2015.

STEARM, M. J., BISHOP, S. C., MALLARD, B. A., & RAADSMA, H. A aplicação de genética para aumentar a resistência a doenças infecciosas em animais. **Revista Brasileira de Zootecnia**, v.30, n.3, p. 70-78, 2001.

STORZ, J. F. Using genome scans of DNA polymorphism to infer adaptive population divergence. **Molecular Ecology**, v. 14, n. 8, p. 671-688, 2005.

TSAI, H. Y., HAMILTON, A., TINCH, A. E., GUY, D. R., GHARBI, K., STEAR, M. J., ... & BISHOP, S. C. Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. **BMC Genomics**, v.16, n.1, p. 969, 2015.

URBINATI, E.C.; CARNEIRO, P.C.F. Práticas de manejo e estresse dos peixes em piscicultura. In: CYRINO, J.E.P.; URBINATI, E.C.; FRACALOSS, D.M.; CASTAGNOLLI, N. (Eds). **Tópicos Especiais em Piscicultura de Água Doce Tropical Intensiva**. São Paulo: Tecart, 2004. p. 171-194.

URREA-ROJAS, Angela Maria et al. Genetic diversity of Tambaqui (Teleostei - Characidae) broodstocks from Northern region of Brazil using microsatellite markers. **Semina: Ciências Agrárias**, [s. l.], v. 41, n. 6, p. 3249–3258, 2020.

VALLADÃO, G. M. R. Potencial de óleos essenciais de plantas para o tratamento de enfermidade em peixes. **Dissertação de mestrado (Universidade Estadual Paulista, Centro de Aquicultura)**, 78 p., 2014.

VAN MUISWINKEL, W. B.; NAKAO, M. A short history of research on immunity to infectious diseases in fish. **Developmental and Comparative Immunology**, v. 43, n. 2, p. 130–150, 2014.

WANG, T., SECOMBES, C. J., & WANG, T. The cytokine networks of adaptive immunity in fish. **Fish & Shellfish Immunology**, v.46, n.1, p. 42-53, 2015.

WANG, Z., GERSTEIN, M., & SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews Genetics**, v.10, n.1, p. 57-63, 2009.

WAPLES, R. S.; GAGGIOTTI, O. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. **Molecular Ecology**, v. 15, p. 1419-1439, 2006.

YAMAGUCHI, Takuya; DIJKSTRA, Johannes M. Major Histocompatibility Complex (MHC) Genes and Disease Resistance in Fish. **Cells** **2019**, Vol. 8, Page 378, [s. l.], v. 8, n. 4, p. 378, 2019. Disponível em: <https://www.mdpi.com/2073-4409/8/4/378/htm>. Acesso em: 12 jan. 2025.

YÁÑEZ, J. M., HOUSTON, R. D., & NEWMAN, S. Genetics and genomics of disease resistance in salmonid species. **Frontiers in Genetics**, v.5, p. 415, 2014.

YÁÑEZ, J. M.; MARTÍNEZ, V. Factores genéticos que inciden en la resistencia a enfermedades infecciosas en salmónidos y su aplicación en programas de mejoramiento. **Archivo de medicina Veterinaria, Universidad Austral de Chile**, v. 13, p. 1–13, 2010.

YÁÑEZ, J. M.; HOUSTON, R. D.; NEWMAN, S. Genomics in aquaculture: applications, opportunities and challenges. **Frontiers in Genetics**, v. 6, p. 129, 2015.

ZHANG, Yilin et al. (*Epinephelus coioides*) in Response to *Vibrio harveyi* Challenge. [s. l.], v. 15, 2023.

ZIEGON, L.; SCHLEGEL, M. Netrin-1: A Modulator of Macrophage Driven Acute and Chronic Inflammation. **International Journal of Molecular Sciences** **2022**, Vol. 23, Page 275, [s. l.], v. 23, n. 1, p. 275, 2021. Disponível em: <https://www.mdpi.com/1422-0067/23/1/275/htm>. Acesso em: 14 out. 2024.

9. ANEXOS

9.1 Anexo 1: Script de filtragem de sequencias em Python

```
import os

import glob

import re

def create_directory(dir_name):

    if not os.path.exists(dir_name):

        os.makedirs(dir_name)

def combine_files(output_dir, pattern="Tambaqui_*.txt"):

    create_directory(output_dir)

    output_filename = os.path.join(output_dir, "Tambaqui_todos.txt")

    # Encontrar todos os arquivos que correspondem ao padrão

    input_filenames = sorted(glob.glob(pattern), key=lambda x:
int(os.path.splitext(os.path.basename(x))[0].split('_')[-1]))

    with open(output_filename, 'w', encoding='utf-8') as outfile:

        for fname in input_filenames:

            with open(fname, 'r', encoding='utf-8') as infile:

                for line in infile:

                    outfile.write(line)

    print(f'Arquivos combinados com sucesso em {output_filename}')

def process_relevant_blocks(input_filepath, output_dir):

    create_directory(output_dir)
```

```

output_filepath = os.path.join(output_dir, "Alinhadas_filtradas.txt")

# Listas para armazenar blocos de dados
current_block = []
relevant_blocks = []

# Abrir o arquivo e processar bloco por bloco
try:
    with open(input_filepath, 'r', encoding='utf-8') as file:
        for line in file:
            if "Query #" in line:
                # Verifica se o bloco atual contém a string desejada antes de iniciar um novo bloco
                if any("Sequences producing significant alignments:" in l for l in current_block):
                    relevant_blocks.append(current_block)
                    current_block = [] # Reinicia o bloco
                current_block.append(line)

            # Verificar o último bloco após terminar a leitura
            if any("Sequences producing significant alignments:" in l for l in current_block):
                relevant_blocks.append(current_block)
except FileNotFoundError:
    print(f"Erro: Arquivo {input_filepath} não encontrado.")
    return

with open(output_filepath, 'w', encoding='utf-8') as outfile:
    for block in relevant_blocks:
        for line in block:
            outfile.write(line)

```

```

print(f"Blocos relevantes processados com sucesso em {output_filepath}")

def filter_relevant_lines(input_filepath, output_dir):
    create_directory(output_dir)

    output_filename = os.path.join(output_dir, "Tambaqui_todos_filtrado.txt")

    # Lista para armazenar linhas relevantes
    relevant_lines = []
    capture_block = False

    # Compilar a expressão regular para verificar as linhas de interesse
    query_regex = re.compile(r"Query #")
    alignment_regex = re.compile(r"Sequences producing significant alignments:")

    # Abrir o arquivo e processar linha por linha
    with open(input_filepath, 'r', encoding='utf-8') as file:
        for line in file:
            if query_regex.search(line):
                capture_block = False # Reinicia a captura de bloco

            if capture_block or alignment_regex.search(line):
                relevant_lines.append(line)
                capture_block = True

    # Escrever as linhas relevantes no arquivo final
    with open(output_filename, 'w', encoding='utf-8') as output_file:
        output_file.writelines(relevant_lines)

    print(f"Processo concluído. {len(relevant_lines)} linhas relevantes salvas.")

```

```

def filter_gene_lines(input_filepath, output_dir):

    create_directory(output_dir)

    output_filename = os.path.join(output_dir, "Lista_genes.txt")

    # Conjunto para armazenar linhas de genes (evita duplicatas)
    gene_lines = set()

    # Abrir o arquivo e processar linha por linha
    with open(input_filepath, 'r', encoding='utf-8') as file:
        for line in file:
            if line.startswith(">"):
                gene_lines.add(line)

    # Escrever as linhas de genes no arquivo final
    with open(output_filename, 'w', encoding='utf-8') as output_file:
        output_file.writelines(gene_lines)

    print(f'Processo concluído. {len(gene_lines)} linhas de genes salvas.')

def filter_sequence_lines(input_filepath, output_dir):

    create_directory(output_dir)

    output_filename = os.path.join(output_dir, "Lista_sequencias.txt")

    # Conjunto para armazenar sequências (evita duplicatas)
    sequence_lines = set()

    # Expressão regular para encontrar sequências do tipo "TP<number>"
    sequence_regex = re.compile(r"(TP\d+)")

```

```

# Abrir o arquivo e processar linha por linha

with open(input_filepath, 'r', encoding='utf-8') as file:

    for line in file:

        match = sequence_regex.search(line)

        if match:

            sequence_lines.add(f">{match.group(1)}\n")


# Escrever as linhas de sequências no arquivo final

with open(output_filename, 'w', encoding='utf-8') as output_file:

    output_file.writelines(sequence_lines)


print(f"Processo concluído. {len(sequence_lines)} sequências salvas.")


def filter_fasta_sequences(output_dir):

    create_directory(output_dir)

    fasta_filepath = "Sequencias.fasta" # Nome do arquivo FASTA

    list_filepath = os.path.join(output_dir, "Lista_sequencias.txt")

    output_filename = os.path.join(output_dir, "Sequencias_filtradas.fasta")


# Ler as sequências da lista

with open(list_filepath, 'r', encoding='utf-8') as file:

    sequence_list = set(line.strip() for line in file)


# Variáveis para armazenar a sequência atual e seu estado

filtered_sequences = []

capture_sequence = False


# Abrir o arquivo FASTA e processar linha por linha

with open(fasta_filepath, 'r', encoding='utf-8') as file:

```

```

for line in file:

    if line.startswith(">"):

        capture_sequence = line.strip() in sequence_list

    if capture_sequence:

        filtered_sequences.append(line)


# Escrever as sequências filtradas no arquivo final
with open(output_filename, 'w', encoding='utf-8') as output_file:

    output_file.writelines(filtered_sequences)


print(f"Processo concluído. {len(filtered_sequences)} linhas de sequências filtradas salvas.")


def filter_hapmap_file(output_dir):

    create_directory(output_dir)

    hapmap_filepath = "HapMap.hmp.txt"

    list_filepath = os.path.join(output_dir, "Lista_sequencias.txt")

    output_filename = os.path.join(output_dir, "HapMap_filtrado.hmp.txt")


    # Ler as sequências da lista

    with open(list_filepath, 'r', encoding='utf-8') as file:

        sequence_list = set(line.strip()[1:] for line in file) # Remover o ">" do início de cada linha


    filtered_lines = []


    # Abrir o arquivo HapMap e processar linha por linha

    with open(hapmap_filepath, 'r', encoding='utf-8') as file:

        header = file.readline()

        filtered_lines.append(header)

```

```

for line in file:

    if line.split('\t')[0] in sequence_list:

        filtered_lines.append(line)


# Escrever as linhas filtradas no arquivo final

with open(output_filename, 'w', encoding='utf-8') as output_file:

    output_file.writelines(filtered_lines)


print(f"Processo concluído. {len(filtered_lines) - 1} linhas de sequências filtradas salvas no
HapMap.")


def main():

    input_combined_path = "Tambaqui_*.txt"

    output_dir = "combined_files"


    # Passo 1: Combinar arquivos

    combine_files(output_dir, input_combined_path)


    # Passo 2: Processar blocos relevantes

    process_relevant_blocks(os.path.join(output_dir, "Tambaqui_todos.txt"), output_dir)


    # Passo 3: Filtrar linhas relevantes

    filter_relevant_lines(os.path.join(output_dir, "Tambaqui_todos.txt"), output_dir)


    # Passo 4: Filtrar linhas de genes

    filter_gene_lines(os.path.join(output_dir, "Tambaqui_todos.txt"), output_dir)


    # Passo 5: Filtrar sequências específicas

    filter_sequence_lines(os.path.join(output_dir, "Alinhadas_filtradas.txt"), output_dir)

```

```
# Passo 6: Filtrar sequências do arquivo FASTA

filter_fasta_sequences(output_dir)


# Passo 7: Filtrar sequências do arquivo HapMap

filter_hapmap_file(output_dir)


# Chamar a função principal

if __name__ == "__main__":
    main()
```

9.1 Anexo 2: Script para análise de FST em Python

```
import pandas as pd

import numpy as np

import os


# Função para encontrar o arquivo no diretório atual
def find_file(filename):

    for root, dirs, files in os.walk('.'):

        if filename in files:

            return os.path.join(root, filename)

    return None


# Nome do arquivo
filename = 'HapMap_filtrado.hmp.txt'


# Encontrar o arquivo
file_path = find_file(filename)


if file_path is None:

    print(f"Arquivo {filename} não encontrado no diretório atual.")

else:

    # Carregar o arquivo de dados
    data = pd.read_csv(file_path, delimiter='\t')


# Definir as populações com base no prefixo dos IDs
populations = {

    'POP1': [col for col in data.columns if col.startswith('AB')],

    'POP2': [col for col in data.columns if col.startswith('AM')],
```

```

'POP3': [col for col in data.columns if col.startswith('BR')],
'POP4': [col for col in data.columns if col.startswith('CE')],
'POP5': [col for col in data.columns if col.startswith('NA')],
'POP6': [col for col in data.columns if col.startswith('PA')],
'POP7': [col for col in data.columns if col.startswith('RO')],
'POP8': [col for col in data.columns if col.startswith('SP')],
}

```

Mapeamento dos códigos de nucleotídeos derivados da IUPAC para pares de bases

```

iupac_alleles = {
    'A': ['A'],
    'C': ['C'],
    'G': ['G'],
    'T': ['T'],
    'R': ['A', 'G'],
    'Y': ['C', 'T'],
    'S': ['C', 'G'],
    'W': ['A', 'T'],
    'K': ['G', 'T'],
    'M': ['A', 'C'],
    'N': ['A', 'C', 'G', 'T'], # Considerar todas as bases para desconhecido
    '+': ['+'],
    '0': ['+', '-'],
    '-': ['-'],
}

```

Função para calcular frequências alélicas considerando códigos IUPAC

```

def allele_frequencies(data, populations):
    freqs = {}

```

```

for pop, samples in populations.items():

    pop_data = data[samples]

    allele_counts = pop_data.apply(lambda x: x.map(lambda allele: iupac_alleles.get(allele,
[])).explode().value_counts(), axis=1).fillna(0)

    total_alleles = allele_counts.sum(axis=1)

    freqs[pop] = allele_counts.div(total_alleles, axis=0)

return freqs

```

Função para calcular FST

```
def calculate_fst(data, freqs, populations):
```

```
    fst_values = []
```

```
    for index, row in data.iterrows():
```

```
        pop1_alleles = freqs['POP1'].loc[index]
```

```
        pop2_alleles = freqs['POP2'].loc[index]
```

```
        h_t = 1 - sum(((pop1_alleles + pop2_alleles) / 2) ** 2)
```

```
        h_s1 = 1 - sum(pop1_alleles ** 2)
```

```
        h_s2 = 1 - sum(pop2_alleles ** 2)
```

```
        h_s = (h_s1 + h_s2) / 2
```

```
        if h_t > 0:
```

```
            fst = (h_t - h_s) / h_t
```

```
        else:
```

```
            fst = 0
```

```
        fst_values.append(fst)
```

```
    return fst_values
```

Calcular frequências alélicas

```
allele_freqs = allele_frequencies(data, populations)
```

```

# Calcular valores de FST

data['FST'] = calculate_fst(data, allele_freqs, populations)


# Identificar SNPs sob seleção natural (FST alto)

threshold = 0.15 # Este valor pode variar dependendo do seu estudo

selected_snps = data[data['FST'] > threshold]


# Salvar os SNPs selecionados em formato HapMap sem a coluna de FST

selected_snps = selected_snps.drop(columns=['FST'])

selected_snps.to_csv('selected_snps_hapmap.txt', sep='\t', index=False)


# Salvar os valores de FST em um arquivo separado em ordem decrescente

fst_values = data[['rs#', 'FST']].sort_values(by='FST', ascending=False)

fst_values.to_csv('fst_values.txt', sep='\t', index=False, header=True)


print(f'Identificamos {len(selected_snps)} SNPs sob seleção natural.')

print("Os resultados foram salvos nos arquivos 'selected_snps_hapmap.txt' e 'fst_values.txt'.")

```

9.1 Anexo 3: Script para conversão dos códons em aminoácidos

```
import pandas as pd

# Dados dos SNPs identificados com suas respectivas informações
snp_data = [
    {
        "SNP ID": "TP132900",
        "Gene": "unc5da",
        "Códon Ref": "CCT",
        "Códon Var": "CCC",
        "AA Ref": "Pro",
        "AA Var": "Pro",
        "Tipo de Mutação": "Sinônima"
    },
    {
        "SNP ID": "TP159385",
        "Gene": "unc5da",
        "Códon Ref": "GAG",
        "Códon Var": "AAG",
        "AA Ref": "Glu",
        "AA Var": "Lys",
        "Tipo de Mutação": "Não Sinônima"
    },
    {
        "SNP ID": "TP163689",
        "Gene": "il16",
        "Códon Ref": "TTA",
        "Códon Var": "TAA",
    }
```

```

        "AA Ref": "Leu",
        "AA Var": "STOP",
        "Tipo de Mutação": "Nonsense"
    },
    {
        "SNP ID": "TP171695",
        "Gene": "epha2b",
        "Códon Ref": "GAG",
        "Códon Var": "AAG",
        "AA Ref": "Glu",
        "AA Var": "Lys",
        "Tipo de Mutação": "Não Sinônima"
    }
]

```

```
df_snps_completo = pd.DataFrame(snp_data)
```

```
import ace_tools as tools; tools.display_dataframe_to_user(name="Tabela Completa de SNPs
Identificados", dataframe=df_snps_completo)
```

9.1 Anexo 4: Sequencias FASTA alinhadas em genes relacionados a importantes funções imunológicas e de sinalização celular.

>TP101702

TGCAGCGGCTGCATTCAGGCACCCCAATGTAGCTGCTGCACCATTTAAGGTGGAAC
GGGAAAAT

>TP102854

TGCAGCGTAATACTCACAGCCTTCTTTTATCGTGCAGGCTTTAGGTGTCCCTCCACA
CAAAAGC

>TP103053

TGCAGCGTAGCGTAGTGCGAGTCTGTCCAGGTGTGGTCCAGGTGGAGTTTCTGGAA
GCTCTGCA

>TP109285

TGCAGCTCCATCAGTCTCTCCTTGTACTGGTTTCTCTCCATCAGGACCCGGGCCATC
TCTACTC

>TP109663

TGCAGCTCCGGCTCCTGTGAGCAGTCTGCTTTTGCAGAATGATGGAAGCGAGGTCA
CTCTGAAT

>TP114313

TGCAGCTGATTGCTGAGATTTCGTGTTTCTGAGCCACACAGTAGTAGGATCCACT
CTGTAGA

>TP115704

TGCAGCTGCTCGAGGCCTCTGCCCCGTTTGAAGTAAGAGTTTCTCTGGTGACTCCCG
CCATCCA

>TP118386

TGCAGCTGGTTCAGCGACGTCAATGGAGCCGTGAAATTTTTCACCATCGTCGTGGCT
GAGTCTG

>TP118695

TGCAGCTGTAGCGGCGGTCTGGTGTAGGTTTTCTGCTCTGATCTCATGTAGTAATGA
ATAGAGA

>TP120601

TGCAGCTTATACACTGATCAGGCATAACATTATGTATGACCACCTCCTTGTTTCTAC
GCTCATT

>TP121991

TGCAGCTTGACTCCATCTGTTCACTGTTTGTAAGTGGTCTGTAAGTGGTCTCTCTC
TATAAT

>TP122418

TGCAGCTTGGATCGCTGATTCAACTGTAGCCAGAATGGATAAGCACAAAATACTGC
AAGGAGCC

>TP123958

TGCAGCTTTTTCCACACCTCCAAAGTTCAGTCTTAGATGTTGGTTGCAGCCAAGTAA
TTCTAAA

>TP126144

TGCAGGAAGAACTACACTGACCTGGTCAGCATCAGAGATCAGAACCAGAACGAAG
CAGTGAAGA

>TP126404

TGCAGGAAGCAAAAGCCACAAACACATTGCAGCGTTTCTGGGGCTGCTGTGTGTTC
TCCTGCTG

>TP126425

TGCAGGAAGCAGAAACCACAGACTGACCGCAGCGCATCTGGGGCTGCTGTGTGTTC
TCCTGCTG

>TP127524

TGCAGGAATATTTTTCCACACCTCCAAAGTTCAGTCTTAGAAGTTGGTTGCATTTTC
TGCAAGT

>TP127561

TGCAGGAATCAGAGGCTACAGACTGACTGTAGTGTGTCTGGGGCTGCTGTGTGTTC
TCCTGCTG

>TP128734

TGCAGGACAGGGGCAGGGCAGCAGAACAGACTCTCCTGGAGATCTGATGAAAGTT
CTCTCACTG

>TP128909

TGCAGGACATAAGCTGTGAGCTTTTGTGTGGAGGGACACCTAAACCTGCACGATA
AAAGAAGG

>TP130033

TGCAGGACGAACTGCGTCTCCCCCTCGGCAGGGAAGCTCGGGGAGATCAGCGAGA
TGTGCAGAG

>TP131090

TGCAGGACTGAGAATGATCCACCACCCAAATAATACATGCTCTGTGGTGGTCCTGT
GGGGGTTC

>TP132900

TGCAGGAGCCAGACGACGCCTACATCGTCAAGAGCAACCCGATCAAGCTGCGCTGC
CGAGCCCG

>TP13362

TGCAGAAGTTCACACAGTGCAGTGGATGTACGGCTGTGAACTGCATGATGATGGAA
CCAAGAGA

>TP134071

TGCAGGAGGCAGTGTGTAATCTGGAACAGTGCCAGACTCGTCTCATCCAGCGTATT
AAGTCTTG

>TP140784

TGCAGGCATGGCCATCCCAGGTGCAGTAGGGGTCTCGGGCCATACAGCACTCAGCA
CAGACCTG

>TP14152

TGCAGAATCAAACAAAGAAGCTGCTGGTGTTCAGAACAAATGGACTGACCACCAC
AGAGGGAT

>TP143183

TGCAGGCCCTGTCTTTGCACGATCCGAGTCCACACAGCCCCTTCTGGACTGTGAGG
AGCACCA

>TP153204

TGCAGGCTGTGGGCTGAGCCTCAGATGCCTGGTCTTTATATGGATGGAGATTTTGTG
ATTGGGG

>TP154798

TGCAGGGACATTTTTCCACACCTCCAAAGTTCAGTCTTAGAAGTCGGTTGCATTTCC
CCTGGTT

>TP155874

TGCAGGGATATTTTTCCACACCTCCAAAGTTCAGTCTAAGAAGTTGGTTCCAGTTTC
TGCTTCT

>TP155881

TGCAGGGATATTTTTCCACACCTCCAAAGTTCAGTCTTAGAAGTTGGTTGCATTTTT
CGCTTCT

>TP155883

TGCAGGGATATTTTCCACACCTCCAAAGTTCAGTCTTAGAAGTTGGTTGTACTTCC
GCAGTCC

>TP155885

TGCAGGGATATTTTCCACACCTCCAAAGTTCAGTCTTAGACATTTGGTGCACTTTC
TTCATCT

>TP156188

TGCAGGGATGTTTTTCCACACCTCCAAAGTTCAGTCTTAGTTAGTAGTCTAATAGTT
GGTTGCA

>TP159385

TGCAGGGCGGGTCGGGCTCGGCAGCGCAGCTTGATCGGATTGCTCTTGACGATGTA
GGCGTCGT

>TP163689

TGCAGGGTCTGGGCTTCAGTATAGTGGGAGGCAGGGATAGCATGTACGGTCCAATG
GGAATCTA

>TP168172

TGCAGGTCAGAGTCACTGAACTGCCCTCCACTATTTACCAGAGGGACTGATGGAT
GCTGAGAC

>TP16949

TGCAGACAAGCTCAGCGATGTCGCTCACTCACAGGTCATGACAAAAACAAGAAA
TCCAAGAAC

>TP17066

TGCAGACAATGAAGGTGTAAATACATGTATCCATGGGATTTTCTCTCGTGGTTTTAA
CCTCCCT

>TP171695

TGCAGGTGATGAAAGCCATTAATGAAGCTTCCGGCTCCCTGCTCCAATGGACTGC
CCCTCTGC

>TP173357

TGCAGGTGGAGTGGCTGAAGAATGATGAGCCCATCAGCACACTTGATGATGTTAAC
ATTGACAC

>TP175076

TGCAGGTGTTACACAGTGCAGTGGATGTACGGCTGTGAGCTGCATGATGATGGAA
ATAGGAGA

TGCAGTAAAATAAAGCTCTGCTGATCATTACAACACAGTTTAAACAATAAATGATT
TTAAATAC

TGCAGTAAATCCGGAACAGTCAGCAGAGTGGTGGATGACAGGAAATGAAAGAGAT
GCTCAAAAC

TGCAGACAGAAGATGCTACAGGCTGACTGTAGTGTGTCTGGGGCTGCTGTGTGTTC
TCCTGCTG

TGCAGTAACACTGATGTGGTGGTGGTGGTGATGAGGATAACCAACACAAACCATGC
AGCAGCAG

TGCAGTAACAGCCTCTACTCTTCTGGGAAGGCTTTACACTAGATGTTGGAACATTCC
TGCGAGG

TGCAGTAACGTTAACTGCTGCACCATTTAAGGTGGAACGGGAAAATTCGAACAAGA
AGCTTCGT

TGCAGACAGAGAGGAGCAGACCTGGTGATCATAAACAGCAGAGAAGAACAGGTGA
GAGAGAGCA

TGCAGTACATCTGCACTGGAGTAACTCCAGGAATAAACTTCCCAGAGTTCTCTGTG
GTGGGTCT

TGCAGTACGAAAGTTAGATACACCGATCAGGCATAACATTATGACCACCGCCTTGT
TTCTACAC

TGCAGTACTGAGAATGATCCACCACCCAAATAATACCTGCTTTGTGGTGGTACTGT
AGGGGTCC

TGCAGTACTGAGAATGATCGACCACCCAAATAATGCCTGCTCTGTGGTGGTCCTGT
GGGGGTCC

>TP18858

TGCAGACAGCTGGGCTGTGGCTCTGTGATCAGCTTCTCCAACCTCTCTCAATTCAGT
CCTGAAC

>TP18875

TGCAGACAGCTGGGCTGTGGTTCTTTGTTCAGCTTCTCCAGCTCCTCTACATCCAGT
CCTGAAC

>TP190576

TGCAGTATCTGTGTCCACACTGAGTCTGGAGCGCCTCGTTCAGCACGTTGTTGCAGC
TGCTGCA

>TP192044

TGCAGTCAAACATTTGCGCTAACTTGTGATGAGGAGGAATTTTGGTCCATTCCTTAC
AGAATTG

>TP192955

TGCAGTCACCCAAATAATACCTGCACTGAGGTGGTCCTGTGGGGGTCCTGACCATT
GAAGAACA

>TP194087

TGCAGTCAGGCTGGTGTAACCTGGTCTCTCTATAGTCAGGTGGTGTAACCTGGTCTGT
AACTGGT

>TP194131

TGCAGTCAGGGTGTTCAACTTGATCCACAACACTATGATGGCAGTCAGCAGGAGAA
CACACAGC

>TP196157

TGCAGTCCAGGTAAGAGTGTTTCAGATCCAGACTGAGGAAGTCTTCTCCATCATAAC
CAAACCGG

>TP196171

TGCAGTCCAGGTGAGAGTGAATATATCCAGACTGATGAAATCTTCTCCATCATAAC
CGAACTGC

>TP196200

TGCAGTCCAGGTTTCAGTGTTTCAGATCCAGACTGAGGAAGTCTTCTCCATCATAACC
ATATTTG

>TP197287

TGCAGTCCGGGTGTTGAAATTGATCCACTGCACTGTGATGGCAGTCAGCAGGAGAA
CACACAGC

>TP197867

TGCAGTCCTGTCTGCTCTCACTCCAGCTCTTCTTCTCAGTAGAGATGTAGTAAATAT
TAGAGTT

>TP199517

TGCAGTCTAAGGTAGGGATAACATTATGACCACCTCCTTGTTTCTACGCTCATTGTC
CATTTTA

>TP200589

TGCAGTCTCTTACTCTGCTTGCTATGGCCTATGACAGGCTAATTGCTATATGTTTTCC
ACTGAG

>TP202239

TGCAGTCTTCTCCTTCTGTCCCGTTATCTGGAGCCCAGTTCTGGTAGCAGATAGACA
CTCCACT

>TP203330

TGCAGTGAACATGAACCATGGCAGTCAGCAGGAGAACACACAGCAGCCCCAGACA
CACTGCA

>TP203331

TGCAGTGAACATGAACCATGGCAGTCTGATGGCAGTCAGCAGGAGAACACACAGC
AGCTCCAGG

>TP203366

TGCAGTGAACCATGGCAGTGTGATGGCAGTCAGCAGGAGAACACACAGCAGCCCC
AGATGCACT

>TP206485

TGCAGTGAGCTTGAACCACAGCAGGGTGATACCAGCCAGCAGGAGAACACACAGC
AGCCCCAGA

>TP207673

TGCAGTGATGCGACTCAGTCTGAACAGCCAGATCGGAATTCATGCAACTTTTACGT
CAATCCAA

>TP209457

TGCAGTGCAGTTCTGATTTCCACCATAATTATCTGGCTGTGTTGATGCCCAGTATCG
GAACAAG

>TP210099

TGCAGTGCCATTGTCACATACACTACACACCGATCAGGCATAACATTATGACCACC
TCCTTGTT

>TP210313

TGCAGTGCCGAGAATGACCCACCACCCAAATAATACCTGCTCTGTGGGGGTCCCTT
GGGGGTCC

>TP210319

TGCAGTGCCGAGAATGATCCACCACCCAAATAATACCTGCTCTGTGGGGGTCCCTGA
CCATTGAG

>TP211875

TGCAGTGCTGAGAATAATCCACCACCCAAATAATACCTGCTCTGTGGTGGTCCTAT
GGGGGTCC

>TP212161

TGCAGTGCTGAGAATGACCCACCACCCAAATAATACCTGCTCTGTGGTGGTCCTGT
GGGGGTCA

>TP212268

TGCAGTGCTGAGAATGATCCAACACAAAAATAATACCTGCTCTGTGGTGGTCCTGT
GAGCGTCC

>TP212275

TGCAGTGCTGAGAATGATCCACAACCCAAATAATACCTGCTCTGTGGTGGTCCTGT
AAAGGGGG

>TP212325

TGCAGTGCTGAGAATGATCCACCACCCAAAAACATCTGCCCTGTGGTGGTCCTCT
AGGGGTCC

>TP212337

TGCAGTGCTGAGAATGATCCACCACCCAAACAGTACCTGCTCTGTAAGGGTCCTGT
AGGGGTGA

>TP212378

TGCAGTGCTGAGAATGATCCACCACCCAAATAGTACCTGCTCTGAGGTGGTCCTGT
AGGGGTCC

>TP212383

TGCAGTGCTGAGAATGATCCACCACCCAAATAGTACCTGCTCTGTAGTGGTCCTGT
GGGATCCT

>TP212387

TGCAGTGCTGAGAATGATCCACCACCCAAATAGTACCTGCTCTGTGACGGTTCATG
GGGGTCCT

>TP212397

TGCAGTGCTGAGAATGATCCACCACCCAAATAGTACCTGCTCTGTGAGGGTCCACC
TTATCAAC

>TP212410

TGCAGTGCTGAGAATGATCCACCACCCAAATAGTACCTGCTCTGTGAGGGTCCTGT
GGGGGTCC

>TP212413

TGCAGTGCTGAGAATGATCCACCACCCAAATAGTACCTGCTCTGTGAGGTGTGTGA
CCTATATA

>TP212424

TGCAGTGCTGAGAATGATCCACCACCCAAATAGTACCTGCTTTGTGGGGGCCTGA
TCATTGAA

>TP212426

TGCAGTGCTGAGAATGATCCACCACCCAAATAGTACCTGCTTTTGTGAGTGTCCATGG
GAGTCCT

>TP212429

TGCAGTGCTGAGAATGATCCACCACCCAAATAGTACCTGTTCTGTGAGGGTCCATG
AGGGTCCT

>TP212439

TGCAGTGCTGAGAATGATCCACCACCCAAATAGTGTATTGTTGGAGAGTTTGGTCTT
ATAGCGC

>TP212444

TGCAGTGCTGAGAATGATCCACCACCCAAATAACACCTGCTCTGTGGTGGTCCTGG
GGGGTCCT

>TP212445

TGCAGTGCTGAGAATGATCCACCACCCAAATAACATCTGCTCTGTTGTGGTCCTGTG
GGAGTCT

>TP212451

TGCAGTGCTGAGAATGATCCACCACCCAAATAATACCTACTCTGTGGTCCTGTGGG
ATCCTGAT

>TP212461

TGCAGTGCTGAGAATGATCCACCACCCAAATAATACCTGCTCTGTGAGGGTCCTGA
CCATTGAA

>TP212472

TGCAGTGCTGAGAATGATCCACCACCCAAATAATACCTGCTCTGTGTAGGGGCTTTT
GAGGGGT

>TP212476

TGCAGTGCTGAGAATGATCCACCACCCAAATAATACCTTCACTGTGGGGGTCCTGA
CCACTGAT

>TP212481

TGCAGTGCTGAGAATGATCCACCACCCAAATAATATCTGCTCTGTGAGGGTCCTGT
CGGGGTCC

>TP212491

TGCAGTGCTGAGAATGATCCACCACCCAAACAAAGTATACAGAGAAACAGATGGAC
TACAGTCTG

>TP212495

TGCAGTGCTGAGAATGATCCACCACCCAACTGCTCGGTGGTGTGCTTTGGGGGTC
CAGTAAAG

>TP212496

TGCAGTGCTGAGAATGATCCACCACCCAAAGGAGTCTCTTGTCAGTGGTGGTCCAAT
GGTGGTCC

>TP212508

TGCAGTGCTGAGAATGATCCACCACCCAGAGATTCAGGTGTATGAATGGTGGTGT
GTATTATA

>TP212549

TGCAGTGCTGAGAATGATCCACCCTCAAATAATACCTGCTCTGTGGTGGTCCTGTG
AGGGCCT

>TP212727

TGCAGTGCTGAGAATGTCCCACCACCCAAATAATACCTGCTCTGTGGTGGTCCTGTG
GGAGTCC

>TP215029

TGCAGTGGATCTTCTCTCTTCATGTTCAATTACGACTCTGGCATGTGTATTCTCCACTG
TCTCA

>TP217143

TGCAGTGGTCAGGACCCCCTACAGGTCCACCACAGAGCAGGTATTATTTGGGTGGT
GGATCATT

>TP220645

TGCAGTGTGGAGTGGCCCTCAGTGCTCCAGTACCAGCCAGGTGTGGACCTGGAAC
TGACCCAT

>TP220648

TGCAGTGTGGAGTGGCCCTCAGTGCTCCAGTACCAGCCTGGTTTGGTCCTGGAAC
TGACCCAT

>TP220657

TGCAGTGTGGAGTGGCCCTCAGTGTTCTGGTACCAGCCCAGTTTGGACCTGGAAC
TGAGCTAT

>TP220666

TGCAGTGTGGAGTGGCTCTCAGTGCTCCAGTACCAGCCCAGTTTGGACCAGGAAC
TGACCCAT

>TP220756

TGCAGTGTGGCTGGGGCTGCTGTGTGTTCTCCTGCTGACTGTCATCACAATGCTGT
GATAAAG

>TP221133

TGCAGTGTGTCTGGGGCTGCTGTGTGTTCTCCTGCTGACTGCCATCACAGCGCTGG
GGTCAGG

>TP221140

TGCAGTGTGTCTGGGGCTGCTGTGTGTTCTCCTGCTGGCTGCCATCACAGTGCTGT
GATCAAG

>TP221144

TGCAGTGTGTCTGGGGCTGCTGTGTGTTCTCCTGCTGGCTGCCATCATAGTGATGT
GATCAAG

>TP223569

TGCAGTTAAAATACCAGTACTGTACCTTTCTCTCCAGAGTGGATCTGCCATATTCC
ACATACT

>TP224501

TGCAGTTACAGACTAGTCCATCTGTTTGTCTGCATACTTTGTTAGCCCCCTTTACCC
TGCTCT

>TP224596

TGCAGTTACAGCCTCGTCTCTTCTGGGAAGGCTTTACACTAGATGTCGGAACATTGC
TGTGAGG

>TP225536

TGCAGTTAGTCAAAGACGCTGCTGGTGTCTCAGAACAATAGACTGGCCACCCAGA
GCTCAGAC

>TP228482

TGCAGTTCCACAGGGACGTCTCGTTCCCCTCACATTCCTCCTCATTCATCCATATAT
GTCCAGT

>TP228716

TGCAGTTCCCCAGATGAGCTTCACTCCCAGAACAATTGAAACCTGTCATACACACG
CGACTCTG

>TP232967

TGCAGTTGGTTCTATATAGAAGCTTTTTGAAAAGCGTTCTATGTAGCACCAAAAAG
GGTTCTGC

>TP23316

TGCAGACGAGCTACTGCTGCACCATTTAAGGTGGAACGGGAAAATTCTGAACAAGA
AGCTGACTT

>TP27342

TGCAGACTGTGACATACACCGATCAAGCATAACATTATGACCACCTCCTTGTTTCTA
CACTCAT

>TP29108

TGCAGAGAACATCACACAGACCTGGCCAGTGTGAGGAACCAGACTGAGAACCAGC
AGATCTTTA

>TP29717

TGCAGAGACAACCATGTGGATCTGGTCTCTGTTCACTGAGAAAATTCAGCACTG
GGTGGAGT

>TP30638

TGCAGAGAGACGTACACTGACCTGGCTACTGTCCACAGCATGGACGAGATGAAGG
AGCTGAAGG

>TP30706

TGCAGAGAGAGACCAGTTACAGACCAGTTACAACATGCTGACTGTAGAGAATAAC
CAGTTACAG

>TP30708

TGCAGAGAGAGACCAGTTACAGACCAGTTACACCAACCTGACTATAGAGAGACCA
GCTACACCA

>TP36388

TGCAGAGCTTCCAGAACTCCACCTGGACCACACCTGGACAGACTCGCACTACGCT
ACACTGCA

>TP37673

TGCAGAGGCAGAGTCTGGATGTTCAACCAAAACCATTCCAGAAATTTCTCTCTCCTC
ATCTCAG

>TP37674

TGCAGAGGCAGAGTCTGGATGTTGACCAAACACATTCCAGAAATTTCTCTCTCCTC
ATCTCAG

>TP42156

TGCAGAGTTCCGTTGAGGGACACACTGGACTCGGCGTAGGGCAGATGGTACAGAC
GACCATCTA

>TP49379

TGCAGATGAAATTGAGTTTCTGGTCCTCTGGCAGACTGACCCACTGCTTACTTCCAG
ACTGCAC

>TP49637

TGCAGATGACAGAGGCCTCAGACTCACTACAGGAGTCCAGCAGAACTCTCCTCCAG
TCCTTCCT

>TP4967

TGCAGAAATATTTTTCCACACCTCCAAAGTTCAGTCTTAGAAGTTGGTTGCCTTTTC
TGCTTCT

>TP51339

TGCAGATGGACAGATAGTATTATACACAGCATAACCGATCAGGCATAACATTATG
ACCTTGTT

>TP51459

TGCAGATGGAGCTGGTTGTTCTTGATGGTCTTTGTGGTTAAAGGACGTCCATTTTTT
GTACCAG

>TP59570

TGCAGCAACAGCGGACAACTCACTATTAATACCCTTCATTTCAGAAGTATCCTAC
ATTAAAGT

>TP61634

TGCAGCAATGTGACTCAGTCTGAACAGCCAGATCGGAATTCTTACGACTTTTACGTT
AGTCCAA

>TP61636

TGCAGCAATGTGACTCAGTCTGAACAGCCAGATCGGAATTTATGTGACTTTTACGTC
AATCCAA

>TP63502

TGCAGCACATTCTCCATCAATCTGAAGAACATTTACCATGCAAAGAACCACTTAA
GCATGAAA

>TP6361

TGCAGAACAACTCACTATTAATACCCTTCATTTCAGAAGAAACACTGCTGTCTAC
AAATGTAG

>TP6976

TGCAGAACAGATGATTCCAGTGATTGGACATTAACATGGTACCAACAGAAACCTGG
TGAATCTC

>TP70680

TGCAGCAGCAGTTATGTAGATGGTCGTCTGTACCATCTGCCCTACGCCGAGTCCAGT
GTGTCCC

>TP7188

TGCAGAACAGTGGCACTGGAACCATATAAATCTTTGGACAGAACGCGACAGACTTT
TTCTTACC

>TP72469

TGCAGCAGCTGCAACAACGTGCTGAACGAAGCGCTCCAGACTCAGTGTGGACACA
GATACTGCA

>TP76105

TGCAGCAGTGACTCAGTCTGAACAGCCAGATCGGAATTCATGCGACTTTTACGTCA
ATCCAAC

>TP76167

TGCAGCAGTGATGCAAACCCACCTGTGAAAACTATGCCTGGTTTAAAGAAGGTGG
AACCTCAC

>TP76173

TGCAGCAGTGATGCAAACCCACCTGTGGAGAGCTACACCTGGTTTAAGATATCTAG
AACTGAAG

>TP76206

TGCAGCAGTGCAACTCAGTCTGAACAGCCAGATCGGAATTCATGCAACTTTTTACG
TCAATCCA

>TP7661

TGCAGCAGTTACATTCTGGTGCCTCAGGTGTCAAACTGCTGCACCATTTAAGGTG
GAACGGGA

>TP7886

TGCAGAACCCATATACAGACACGTCGCAGCTGCGCGCATCTAACATGGTGTACAAA
GCTACGTA

>TP81662

TGCAGCCAACATGAACCATGGCAGTGTGATGGCAGTCAGCAGGAGAACACACAGC
AGCCCCAGA

>TP85733

TGCAGCCCACATTCTGACCGTTCTCCTGTAGGTACTGTGGACACTCTTGATATGAAT
CCTCATG

>TP85843

TGCAGCCCAGTAGCATATAGCACAGTCACATGCTTGTGAAATTTTGAAAGTAAATT
GTTAAATA

>TP8995

TGCAGAACTATATTTTTAGATAAAAAACAAAATAATCCTCCTTTGCAGTAAAATAAA
ACTGTCCT

>TP93560

TGCAGCCTGTAGTACATTGCACCAAGCTTTGGTTCCGGGTCCAGACAGGTCTCTATC
ACTGGCT

>TP946

TGCAGAAAAGAGACAACTCACTATTAATACCCTTCATTTC AATAGAAAAGCTGGA
GGTGTCCA

>TP9554

TGCAGCGAGCTTTATCCACAGCATTGTGATGACAGTCAGCAGGAGAACACACAGCA
GCCCCAGA

>TP9720

TGCAGAACTTTTCTTTACCGTGTACTTCATTGCTCAGGACCCCCACAGGACCACCAC
AGAGCAG

>TP99948

TGCAGCGCTCTGCACATCTCGCTGATCTCCCCGAGCTTCCCTGCCGAGGGGGGAGAC
GCAGTTCG