UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"

FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS

CÂMPUS DE JABOTICABAL

CORRIDAS DE HOMOZIGOSE EM TILÁPIAS DO NILO (*Oreochromis niloticus*): IMPACTO DOS CRITÉRIOS DE DETECÇÃO

Isabella Almeida Ferreira

Zootecnista

2022

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"

FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS

CÂMPUS DE JABOTICABAL

CORRIDAS DE HOMOZIGOSE EM TILÁPIAS DO NILO (*Oreochromis niloticus*): IMPACTO DOS CRITÉRIOS DE DETECÇÃO

Isabella Almeida Ferreira

Orientador: Dr. Roberto Carvalheiro

Coorientador: Prof. Dr. José Manuel Yáñez

Dissertação apresentada à Faculdade de Ciências Agrárias e Veterinárias – UNESP, Campus de Jaboticabal, como parte das exigências para a obtenção do título de Mestre em Genética e Melhoramento Animal.

2022

F383c Ferreira, Isabella Almeida

Corridas de homozigose em tilápias do Nilo (Oreochromis niloticus): impacto dos critérios de detecção / Isabella Almeida Ferreira. -Jaboticabal, 2022

61 p.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal

Orientador: Roberto Carvalheiro

Coorientador: José Manuel Yáñez

1. Genética Animal. 2. Melhoramento genético. 3. Endogamia. 4. Tilápia-do-Nilo. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada

unesp

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: CORRIDAS DE HOMOZIGOSE EM TILÁPIAS DO NILO (*Oreochromis niloticus*): IMPACTO DOS CRITÉRIOS DE DETECÇÃO

AUTORA: ISABELLA ALMEIDA FERREIRA
ORIENTADOR: ROBERTO CARVALHEIRO
COORIENTADOR: JOSÉ MANUEL YÁÑEZ LÓPEZ

Aprovada como parte das exigências para obtenção do Título de Mestra em GENÉTICA E MELHORAMENTO ANIMAL, área: Zootecnia pela Comissão Examinadora:

Pesquisador Dr. ROBERTO CARVALHEIRO (Participaçao Virtual)
Departamento de Zootecnia / FCAV - UNESP Jaboticabal

Prof.Dr. FERNANDO SEBASTIAN BALDI REY (Participaçao Virtual)
Departamento de Zootecnia / FCAV / UNESP - Jaboticabal

Pós-Doutoranda ELISA PERIPOLLI (Participaçao Virtual)
Faculdade de Zootecnia e Engenharia de Alimentos - USP - FZEA / Pirassununga/SP

Jaboticabal, 24 de agosto de 2022

DADOS CURRICULARES DO AUTOR

Isabella Almeida Ferreira, filha de Francisco Lima Ferreira e Ana Paula Carvalho de Almeida Ferreira, nasceu em São Paulo, São Paulo, aos 29 de setembro de 1997. Em março de 2015, ingressou no curso de Zootecnia na Universidade Federal de Pelotas, UFPel, Rio Grande do Sul, foi bolsista de ensino em estatística pelo Projeto Galton. Também, foi bolsista de iniciação científica em três projetos do Grupo de Pesquisa em Melhoramento Animal Aplicado a Bovinos da UFPEL. Em dezembro de 2019, obteve título de Zootecnista. Em março de 2020, ingressou no curso de pós-graduação da FCAV (Faculdade de Ciências Agrárias e Veterinária) - Unesp/Jaboticabal, para se habilitar ao Título de Mestre em Genética e Melhoramento Animal, onde concluiu em agosto de 2022.

“Por que dar armas é tão fácil, mas dar livros é tão difícil?
Por que produzir tanques é tão fácil e
construir escolas é tão difícil?”
Malala Yousafsi

Dedico este trabalho a todos pesquisadores (especial as mulheres) que resistem no Brasil e que através de suas pesquisas contribuem com o desenvolvimento do nosso País.

AGRADECIMENTOS

Agradeço à minha família por ter me incentivado e apoiado a viver meus sonhos, mesmo que distante dos olhos de cada um deles. Em especial aos meus pais, Ana Paula Almeida Ferreira e Francisco Lima Ferreira, pelo suporte emocional e financeiro. À minha irmã Vitória de Almeida Ferreira por toda cumplicidade, videochamadas e esculta ativa (como ela diz). Dedico também à minha avó Maria Eliane Lima que acha que estou fazendo outra “faculdade” (risos) e que sempre lembra de mim nos almoços de domingo.

Ao meu orientador Dr. Roberto Carvalheiro pela oportunidade de trabalhar junto ao Programa de pós-graduação em Genética e Melhoramento Animal. Agradeço também por ter ajudado essa jornada ser tão agradável sempre com muitos ensinamentos, paciência e confiança. Sou muito grata e sortuda por essa parceria. Obrigada!

A todos os meus professores de graduação, em especial, a Dr. Arione Augusti Boligon e Dr. Daniel Duarte da Silveira, por toda a atenção que dedicaram nos meus projetos de iniciação científica.

Aos meus colegas do curso de pós-graduação, em especial Bruna, Baltasar, Caio, Gabi, Ivan e Kétuly por sempre estarem à disposição para me ensinar e ajudar. Obrigado pelas conversas no “whats” durante a pandemia.

Agradeço aos meus amigos de graduação da Universidade Federal de Pelotas (UFPel), Érika Madruga, Ezequiel Formetim e Francisca Baldino por terem feito parte da minha caminhada em Pelotas bem como pelos choros e risadas compartilhadas.

À minha melhor amiga Sara Suellen (minha *roomate* preferida) que acompanha minha vida acadêmica desde 2016 e sempre me incentivou.

Ao meu namorado Paulo Henrique Floriano, por ter sido o melhor presente que já ganhei e por tornar os meus dias caóticos em dias mais leves.

Aos docentes do programa da pós-graduação em Genética e Melhoramento animal, pela atenção, ensinamentos e pela preocupação com a formação acadêmica e científica de seus alunos, sempre os incentivando ao constante aprendizado.

A todos que, direta ou indiretamente, contribuíram para a realização deste trabalho.

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

SUMÁRIO

RESUMO - x
ABSTRACT xi
CAPÍTULO 1 - CONSIDERAÇÕES GERAIS 1
1.1 INTRODUÇÃO 1
1.2 OBJETIVOS 3
1.2.1 Objetivos gerais 3
1.2.2 Objetivos específicos 3
1.3 REVISÃO DE LITERATURA 4
1.3.1 Produção nacional e melhoramento genético da tilápia 4
1.3.2 Coeficientes de endogamia em tilápias 7
1.3.3 Corridas de homozigose e coeficiente de endogamia baseado em ROH (F_{ROH}) 9
1.3.4 Critérios para detecção de ROH 11
1.3.5 Efeito da densidade de painéis de SNP nas análises de ROH 12
1.4 REFERÊNCIAS BIBLIOGRÁFICAS 14
CAPÍTULO 2 - Critérios utilizados na detecção de corridas de homozigose (ROH) em tilápias do Nilo (*Oreochromis niloticus*) e suas implicações nos coeficientes de endogamia genômico (F_{ROH}) 26
RESUMO - 26
ABSTRACT - 27
2.1 INTRODUÇÃO 28
2.2 MATERIAL E MÉTODOS 32
2.2.1 Populações 32
2.2.2 Genótipos e controle de qualidade 33
2.2.3 Corridas de homozigose (ROH) e coeficientes de endogamia (F_{ROH} e F_{PED}) 33
2.3 RESULTADOS 36
2.3.1 Tamanho de Janela e número mínimo de SNPs 36
2.3.3 Genótipos ausentes e heterozigotos 38
2.3.4 GAP máximo 39
2.3.5 Comprimento mínimo e desequilíbrio de ligação (LD) 41
2.3.6 Densidade mínima 41
2.3.7 F_{ROH} e F_{ped} 41

2.4 DISCUSSÃO 45
2.5 CONCLUSÃO 50
2.6 REFERÊNCIAS BIBLIOGRÁFICAS 51
2.7 FIGURAS SUPLEMENTARES 56

CORRIDAS DE HOMOZIGOSE EM TILÁPIAS DO NILO *(Oreochromis niloticus)*: IMPACTO DOS CRITÉRIOS DE DETECÇÃO

RESUMO - Os objetivos com o presente trabalho foram: (i) avaliar o impacto de diferentes critérios na detecção de corridas de homozigose (ROH) e estimativas de endogamia genômica (FROH) em painéis de média densidade (50K) de tilápias do Nilo (Oreochromis niloticus); (ii) apresentar o conjunto de valores dos critérios que fornecem estimativas de FROH mais plausíveis; (iii) discutir os níveis de FROH no cenário que apresenta os valores mais adequados para os critérios que determinam um segmento como ROH; e (iv) estimar a endogamia a partir de informações do pedigree (FPED) e comparar com FROH. Foram utilizados genótipos de 2.848 indivíduos de duas populações distintas (POP_A= 1.376 e POP_B= 1.472) usando um painel comercial de 50K SNPs (Single Nucleotide Polymorphisms). Como controle de qualidade, foram removidos marcadores com base no equilíbrio de Hardy-Weinberg (p-value<10^{-15}) e taxa de genotipagem (call rate) menor que 0,95, totalizando 36.080 SNPs para as análises. A análise de ROH foi realizada com uso do pacote detectRUNS implementado no software R e o FROH foi calculado de acordo com a proporção autozigótica do genoma. Os parâmetros que tiveram seus critérios avaliados foram: tamanho de janela e número mínimo de SNPs (10, 15, 20 e 50 SNPs); número máximo de genótipos ausentes permitidos (1, 2, 3 e 5); número máximo de genótipos heterozigotos permitidos (0, 1, 2, 3 e 5); espaçamento máximo entre SNPs (GAP) (50, 125, 370, 500 e 1000 Kb); comprimento mínimo de ROH (350, 500 e 1000 Kb); e densidade mínima de SNP (1/25, 1/50, 1/75 e 1/100 SNP/Kb). A estimação da endogamia do pedigree foi realizada considerando 130.049 animais e o uso do programa BLUPF90. Os resultados dos diferentes cenários foram comparados utilizando estatísticas descritivas do número de segmentos de ROH e os valores estimados de FROH. Os resultados indicaram que, com exceção do critério número máximo de genótipos ausentes, todos os critérios tenderam a afetar tanto o número de ROH quanto as estimativas de FROH. O cenário controle foi o que apresentou estimativas de FROH mais plausíveis, na qual os resultados foram moderados (0,08233 para POP_A e 0,09107 para POP_B) e estão próximos do limite máximo aceitável para tilápias. As estimativas de FPED para a POP_A e POP_B foram iguais a 2,01% e 2,18% (respectivamente) e sugeriram que este coeficiente tende a subestimar a endogamia quando comparado ao FROH do cenário controle. Os resultados aqui apresentados indicam que os critérios utilizados para detectar as corridas de homozigose tem grande impacto nas análises, e consequentemente, nas estimativas de FROH devido a super ou subestimação dos segmentos em homozigose.

Palavras–chave: autozigosidade, corridas de homozigose, genômica, tilapicultura

RUNS OF HOMOZYGOSITY IN NILE TILAPIA (*Oreochromis niloticus*): IMPACT OF DETECTION CRITERIA

ABSTRACT - This study aimed to: (i) evaluate the impact of different criteria in the detection of runs of homozygosity (ROH) and estimates of genomic inbreeding (F_{ROH}) using medium density panels (50K) in Nile tilapia (*Oreochromis niloticus*); (ii) discuss the set of criteria that provides the most plausible F_{ROH} estimates; (iii) discuss the F_{ROH} levels in the scenario that presents the most plausible criteria to determine a ROH; and (iv) estimate inbreeding coefficients from pedigree data (F_{PED}) and compare to F_{ROH}. Genotypes from 2,848 individuals from two distinct populations (POP_A= 1,376 e POP_B= 1,472) were used using a commercial panel of 50K SNPs (Single Nucleotide Polymorphisms). For quality control, markers were removed based on Hardy-Weinberg equilibrium (p-value<10^{-15}) and call rate lower than 0.95, totaling 36,080 SNPs for subsequent analysis. ROH analysis was performed using the detectRUNS package implemented in the R software and the F_{ROH} was calculated according to the autozygosity proportion of the genome. The evaluated parameters were: window size and minimum number of SNPs (10, 15, 20, and 50 SNPs); maximum number of missing genotypes (1, 2, 3, and 5); maximum number of allowed heterozygous genotypes (0, 1, 2, 3, and 5); maximum gap between SNPs (50, 125, 370, 500, and 1000 Kb); minimum ROH length (350, 500, and 1000 Kb; and minimum SNP density (1/25, 1/50, 1/75, and 1/100 SNP/Kb). The F_{PED} was performed in the BLUPF90 program by considering 130,049 animals. The results of the different scenarios were compared using descriptive statistics encompassing the number of ROH, the estimated values of F_{ROH}. The results indicated that all criteria tended to affect both the ROH number and F_{ROH} estimates, except the criterion of maximum number of missing genotypes. The control scenario presented the most plausible F_{ROH} estimates, in which the results were moderate (0.08233 for POP_A and 0.09107 for POP_B) and close to the maximum acceptable inbreeding limit for tilapia. F_{PED} estimates for POP_A and POP_B were equal to 2.01% and 2.18% (respectively) and suggested that this coefficient tends to underestimate inbreeding when compared to the F_{ROH} of the control scenario. Therefore, the results presented herein indicate that the criteria used to detect ROH have a great impact on the analyses, and consequently, on the F_{ROH} estimates due to the over or underestimation of the homozygous segments.

Keywords: autozygosity, genomics, runs of homozygosity, tilaculture

CAPÍTULO 1 - CONSIDERAÇÕES GERAIS

1.1 INTRODUÇÃO

A demanda de alimentos para a população faz com que os programas de melhoramento genético sejam cada vez mais importantes para garantir maior produtividade na agropecuária. O cultivo de espécies aquícolas é uma das atividades da produção animal que mais se desenvolve mundialmente, ainda que poucos indivíduos sejam melhorados geneticamente (Gjedrem, Robinson e Rye, 2012). A tilápia se destaca dentre as espécies de peixes mais produzidas a nível mundial (FAO, 2020). Em vista disso, é crescente os esforços para promover o melhoramento genético dessa espécie.

Em 1971, a tilápia do Nilo foi introduzida na região nordeste do Brasil e a partir de então, distribuída nacionalmente. Entre 2002 e 2005, ocorreu a importação das linhagens *GenoMar Supreme Tilápia* (GST), produzida na Noruega pela empresa Genomar e a Genetic Improved Farmed Tilapia (GIFT) com origem na Malásia. A linhagem GIFT deu início ao programa de melhoramento genético de tilápias do Nilo implantado pela Universidade Estadual de Maringá (UEM), sendo o primeiro programa do Brasil (Ribeiro *et al.*, 2012). Entre os resultados obtidos com a GIFT melhorada pela UEM estão o ganho genético de 15% para crescimento em duas gerações de seleção, comprovando uma eficiente resposta à seleção (Oliveira, de *et al.*, 2014). Na atualidade, a tilapicultura demonstra-se como uma das culturas mais relevantes para a produção aquícola nacional, representando 60,6% da produção nacional da piscicultura e colocando o país como o quarto maior produtor mundial de tilápia (PEIXE, 2021).

Entretanto, altas taxas de endogamia podem ser encontradas em programas de melhoramento genético de tilápias, decorrentes principalmente da restrição quanto ao número de famílias disponíveis para reprodução (Petersen et al., 2012). O incremento da endogamia em uma população provoca perdas na

variância genética aditiva (Falconer e Mackay, 1996). Além disso, características que são fortemente influenciadas por efeitos genéticos não aditivos podem ser mais susceptíveis a combinações gênicas desfavoráveis, gerando depressão endogâmica e, consequentemente, a queda no desempenho reprodutivo, produtivo e na capacidade adaptativa (Mi, Chapman e Tyler, 1965; Nowicki, 1963; Pirlea e Ilea, 1970; Strewe, 1974). Portanto, investigar níveis de endogamia em programas de melhoramento genético é essencial para garantir o ganho genético nas populações melhoradas.

Existem diferentes formas de se calcular coeficientes de endogamia, uma delas pode ser a avaliação de segmentos autozigotos, dado que eles são formados quando indivíduos que apresentam ancestrais comuns são acasalados e transferem segmentos cromossômicos idênticos por descendência (IBD) para o genoma de sua progênie (Wright, 1922). Para estimar a autozigosidade de uma população, as informações genotípicas podem ser utilizadas, apresentando vantagens em relação a informações de pedigree como rapidez, precisão e baixo custo (Silva et al., 2015). As informações de pedigree também podem conter erros aleatórios decorrentes da coleta, anotação, registro e armazenamento das informações de parentesco podendo provocar acréscimos indesejados nos níveis de autozigosidade (Curik, Ferenčaković e Sölkner, 2014; Hudson et al., 2015; Zavarez et al., 2015). Dentre as possíveis análises que utilizam informações genômicas estão as corridas de homozigose ROH (do inglês "*Runs of Homozygosity*"), sendo caracterizadas por longos trechos genômicos em homozigose, que são comuns entre indivíduos e populações (Peripolli *et al.*, 2017).

Diversos fatores podem afetar o comprimento, a frequência e a localização de ROH, logo, as estimativas de coeficientes de endogamia baseados em ROH (F_{ROH}) podem ser subestimadas ou superestimadas. A variação da densidade dos painéis de genotipagem é um dos principais fatores que pode afetar a estimação de F_{ROH}. Painéis de menor densidade (50K), por exemplo, não são recomendados para a identificação de segmentos menores

que 4 Mb uma vez que tendem a superestimá-los (Ferenčaković, Sölkner e Curik, 2013). Os critérios utilizados para determinar um segmento como ROH também podem ter efeito sob os resultados das análises de ROH. Diversos estudos identificaram que esses critérios quando não determinados com cautela tendem a viesar coeficientes de endogamia calculados com base em ROH, principalmente quando utilizadas informações de painéis de SNPs de baixa a média densidade (Ferenčaković, Sölkner e Curik, 2013; Howrigan, Simonson e Keller, 2011; Ku et al., 2011).

Portanto, devido à importância econômica da tilápia do Nilo, o estudo do efeito dos critérios utilizados na detecção de ROH sobre a estimativa de F_{ROH} torna-se importante, tendo em vista que o aumento na precisão da estimativa de F_{ROH} pode representar importante estratégia para o controle da endogamia nos programas de melhoramento de peixes.

1.2 OBJETIVOS

1.2.1 Objetivos gerais

Com este trabalho objetiva-se avaliar os efeitos de diferentes critérios utilizados na detecção de corridas de homozigose (ROH) sobre as estimativas de coeficientes de endogamia genômico (F_{ROH}) de duas populações de tilápias do Nilo (*Oreochromis niloticus*).

1.2.2 Objetivos específicos

- Apresentar o conjunto de valores dos critérios que fornecem estimativas de F_{ROH} mais plausíveis.
- Discutir os níveis de F_{ROH} no cenário que apresenta os valores mais adequados para os critérios que determinam um segmento como ROH.

- Estimar a endogamia a partir de informações do pedigree (F_{PED}) e compará-la com F_{ROH}.

1.3 REVISÃO DE LITERATURA

1.3.1 Produção nacional e melhoramento genético da tilápia

A Organização das Nações Unidas para a Alimentação e Agricultura (FAO) destaca que a criação de animais e plantas aquáticas estão entre os setores de produção de alimentos que mais se desenvolvem a nível mundial. Em 2010, a aquicultura superou a pesca e captura de seres aquáticos, gerando uma produção homogênea, rastreável e lucrativa (Lucas, Southgate e Tucker, 2019). Em 2017, a aquicultura produziu 18,32 milhões de toneladas a mais que a produção extrativista (FAO, 2018). O território brasileiro tem ganhado grande importância na exportação de pescados tendo em vista as melhorias nos sistemas de distribuição bem como aumentos de produção (FAO, 2018).

O Brasil é o quarto maior produtor de espécies aquáticas da América do Sul, com produção de 1,32 milhões de toneladas em 2018, ficando atrás apenas do Peru, Chile e México que produziram 7,27, 3,39 e 1,94 milhões de toneladas, respectivamente nessa mesma época (FAO, 2020). Espera-se que a produção aquícola no Brasil atinja 1,49 milhões de toneladas até 2030, alcançando um crescimento de 12,9% em relação ao ano de 2018 (FAO, 2020). Em 2020, de acordo com a Associação de Brasileira de Piscicultura (PEIXE, 2021), a produção de peixes atingiu 802.930 toneladas, demonstrando o segundo melhor desempenho desde 2014.

A tilápia do Nilo (*Oreochromis niloticus*) está entre as espécies mais produzidas no Brasil e representa 60,6% da produção nacional da piscicultura, colocando o país como o quarto maior produtor mundial de tilápia (PEIXE, 2021). Tamanha importância dada a espécie, torna-se notório a sua contribuição para a expansão da psicultura em território nacional. O estado líder do ranking

nacional na produção desse peixe é o Paraná, com 166 mil toneladas no mercado, seguido por São Paulo, com 70,5 mil toneladas, e Minas Gerais com 42,1 mil toneladas (PEIXE, 2021).

O Brasil, mesmo apresentando uma aquicultura de destaque, pode obter incrementos de produção através de programas de melhoramento genético. A utilização de peixes como modelos para os programas de melhoramento é extremamente vantajosa, pois as espécies aquícolas apresentam alta fecundidade, fecundação externa, reduzido intervalo de gerações (um a quatro anos), além das características produtivas geralmente apresentam herdabilidade moderada (entre 20% e 40%) e possibilitarem a formação de híbridos (Gjedrem, Robinson e Rye, 2012; Rye, Gjerde e Gjedrem, 2010). De acordo com Gjedrem e Rye (2018), produzir peixes provenientes de programas de melhoramento genético e com taxas de crescimento elevadas permite reduzir o custo da alimentação e o tempo de abate, além de melhorar a qualidade do produto, aumentar a sobrevivência e o bem-estar animal. Dado esse potencial, caso a produção total da aquicultura fosse baseada em animais geneticamente melhorados, a produção poderia até dobrar em 13 anos (Gjedrem, Robinson e Rye, 2012). Contudo, em 2010 estimou-se que 8% da produção mundial de aquicultura vem de animais melhorados (Gjedrem, Robinson e Rye, 2012), indicando que a quantidade de animais provenientes de programas de melhoramento genético no mundo é baixa.

O melhoramento genético da tilápia foi implementado nas Filipinas em abril de 1988 pelo ICLARM (*International Center for Living Aquatic Resources Management*), atual *WorldFish Center*, e dispôs com a colaboração de diversos centros de pesquisa e financiamento (Bentsen et al., 1998; Eknath et al., 1993). O programa considerou na formação da população base quatro linhagens africanas selvagens de tilápias coletadas em 1988-1989 no Egito, Gana, Quênia e Senegal, e quatro linhagens domésticas asiáticas introduzidas nas Filipinas entre os anos de 1979-1984 (Israel, Singapura, Tailândia e Taiwan) e foi nomeado como “Genetic Improvement of Farmed Tilapias” (GIFT). No Brasil, o

uso das ferramentas clássicas de melhoramento em tilápias teve início em 2005 com a importação da linhagem GIFT, por meio de uma cooperação entre Universidade Estadual de Maringá (UEM) e o *WorldFish Center*.

As características associadas ao crescimento, como o peso corporal, são as principais selecionadas no melhoramento de tilápias, algumas exceções selecionam também para rendimento de filé, capacidade de sobrevivência e resistência a doenças (Yáñez, Joshi e Yoshida, 2020). O foco inicial do programa de melhoramento genético da linhagem GIFT foi a seleção combinada entre e dentro de famílias para a característica taxa de crescimento (Elghobashy, 2001). Segundo Bentsen et al. (2017), o ganho acumulado na linhagem GIFT após 5 anos de projeto foi de 86%, aproximadamente 17% por geração. O rendimento de filé é outro objetivo de seleção economicamente importante, entretanto, as características relacionadas ao filé são difíceis e caras de se mensurar, além disso, não podem ser medidas no peixe vivo (Vandeputte et al., 2020). Por isso, pesquisas estão estimando a correlação genética de características de filé com características de crescimento para identificar a possibilidade de se realizar a seleção indireta para essa característica.

Nguyen et al. (2010) observaram que a seleção para alto crescimento pode resultar em um aumento correlacionado significativo no peso do filé em três gerações da linhagem GIFT selecionadas para ganho de peso. Ma et al. (2015) concluíram que após seis gerações selecionadas para diversas características de crescimento houve uma considerável melhora genética do peso de filé. No entanto, existem estudos que encontraram que a seleção indireta para essa característica através da seleção para peso corporal pode ser ineficaz, uma vez que encontraram uma associação genética baixa entre as características (Gjerde et al., 2012; Rutten, Bovenhuis e Komen, 2004; Ma et al., 2011). Atualmente, existem alternativas que estão sendo estudadas para medir com maior eficiência o rendimento de filé, como por exemplo o uso de ultrassom e tomografia para fenotipar os animais (Maas et al., 2020; Prchal et al., 2020; Vandeputte et al., 2017). Além dessas características, também existem estudos nos quais o

objetivo de seleção foi a coloração externa do macho (Rajaee et al., 2010), a tolerância ao frio (Charo-Karisa et al., 2005) e a resistência a doenças (LaFrentz et al., 2016; Shoemaker et al., 2017; Suebsong et al., 2019).

1.3.2 Coeficientes de endogamia em tilápias

Os programas de melhoramento genético de tilápia têm apresentado um expressivo desenvolvimento nos últimos anos, totalizando mais de 20 programas (Cádiz et al., 2020). Entretanto, esses programas podem apresentar altas taxas de endogamia devido principalmente à restrição do número de famílias disponíveis para reprodução por falta de espaço físico e por questões orçamentárias (Petersen et al., 2012; Ponzoni et al., 2010). O incremento da endogamia provoca o aumento de indivíduos homozigotos e eleva as chances de expressão de alelos deletérios indesejáveis (Bentsen e Olesen˚, 2002; Gjedrem e Baranski, 2009; Oliveira, 2013). Concomitante a isso, as características reprodutivas, produtivas e adaptativas também são afetadas pela endogamia, gerando depressão endogâmica e consequentemente perdas na eficiência reprodutiva, produtiva e na capacidade adaptativa (Mi, Chapman e Tyler, 1965; Nowicki, 1963; Pirlea e Ilea, 1970; Strewe, 1974)

A endogamia ocorre quando os indivíduos acasalados são mais aparentados que a média da sua população de origem (Breda et al., 2004; Queiroz, Albuquerque e Lanzoni, 2000). Como consequência desse tipo de acasalamento, cópias de genes presentes nos ancestrais podem ser transmitidas para os seus descendentes (Gutiérrez-Reinoso, Aponte e García-Herreros, 2022). Portanto, indivíduos endogâmicos podem portar dois alelos em um *locus* que sejam réplicas de um alelo da geração anterior. De acordo com Malécot (1948), existem dois tipos de identidade entre alelos, os idênticos em estado (IBS) e os idênticos por descendência (IBD). Os IBS são alelos fisicamente idênticos e que apresentam uma identidade funcional. Por outro lado, os IBD são aqueles que foram originados por um evento de replicação.

Sendo assim, ao calcular a probabilidade de dois alelos serem IBD em um indivíduo podemos quantificar o nível de endogamia desse indivíduo. Comumente, o coeficiente de endogamia é calculado a partir do monitoramento das informações de pedigree (F_{PED}), que são utilizadas para formar a matriz de parentesco “A” (Wright, 1922).

Em tilápias do Nilo, Oliveira et al. (2014) estimaram um coeficiente de endogamia de 0,5% e 0,4% nas segunda e terceiras gerações, respectivamente, de animais da linhagem GIFT. Bolivar E Newkirk (2002) encontraram uma endogamia de 6,3% e uma taxa média de 1,4% por geração após doze gerações de tilápias selecionadas para peso corporal às 16 semanas. Ponzoni et al. (2010) avaliando tilápias melhoradas na Malásia observaram um incremento de 2,14% na décima terceira geração selecionada para crescimento. Bentsen e Olesen (2002), encontraram que são aceitáveis incrementos menores que 1% de endogamia por geração para que não ocorra efeitos deletérios na população. Em contrapartida, Holtsmark et al. (2008) discutem que é aceitável um acréscimo de 0,5% por geração selecionada no nível de autozigosidade.

De acordo com Oliveira et al. (2014), outro ponto relevante é o tamanho efetivo da população (Ne), ou seja, o número de indivíduos que uma população idealizada teria a mesma taxa de aumento na endogamia do que a população em estudo. Segundo o autor, um Ne reduzido pode provocar um aumento na endogamia e deriva genética, podendo a longo prazo dificultar a manutenção de programas de melhoramento. Ponzoni et al. (2010) concluiu que um Ne de 88 indivíduos pode ser o suficiente para conter a endogamia e manter a herdabilidade. Porém, Franklin e Frankham (1998) recomendaram um Ne de 500 indivíduos para manter o potencial evolutivo. Mckinna et al. (2010) estudando GIFT criadas nas ilhas Fiji sugeriram que o declínio da diversidade genética foi devido ao mau manejo do recurso genético.

Uma das soluções que podem ser utilizadas para diminuir os níveis de endogamia das populações selecionadas é a adoção de diferentes estratégias de seleção e de acasalamento, possibilitando manter um programa de

reprodução sustentável a longo prazo em diferentes programas de criação de tilápias (Ponzoni et al., 2010; Yoshida et al., 2017). Dessa forma, estimar as taxas de endogamia da população é de extrema relevância para definir estratégias de controle e maximizar o ganho genético ao longo das gerações.

1.3.3 Corridas de homozigose e coeficiente de endogamia baseado em ROH (F_{ROH})

Atualmente, o controle da endogamia pode ser realizado baseando-se em informações provenientes de painéis de genotipagem. A detecção e caracterização de ROH tem sido amplamente utilizada para explorar os níveis de endogamia populacional e individual em bovinos (Biscarini et al., 2020), frangos de corte (Cendron et al., 2020), ovelhas (Deniskova et al., 2021), búfalos (Ghoreishifar et al., 2020), cabras (Islam et al., 2019), humanos (McQuillan et al., 2008) e salmão (Yoshida et al., 2020). Os coeficientes de endogamia baseados em ROH (F_{ROH}) apresentam vantagens em relação ao calculado com base no pedigree (F_{PED}) como, por exemplo, maior precisão (Gurgul et al., 2016; Marras et al., 2015; Zhang et al., 2015), capacidade de identificar eventos de endogamia remotos (Ferencakovic et al., 2011), fornece informações adicionais sobre endogamia recente (Gomez-Raya et al., 2015) por fim permite ser estimado em qualquer animal genotipado, mesmo quando a informação do pedigree não está disponível (Keller et al. 2011).

Além das estimativas de F_{ROH}, as ROH podem ser utilizadas para a identificação de alelos deletérios recessivos e caracterização de eventos demográficos, estruturais e históricos das populações estudadas (Peripolli et al., 2017; Saura et al., 2015). Esses eventos populacionais podem afetar os padrões de homozigosidade do genoma, podendo assim ser identificados através dos padrões das ROH (Bosse et al., 2012; Purfield et al., 2012; Ramey et al., 2013). A seleção de um número reduzido de indivíduos e o sistema de acasalamento utilizado nos programas de melhoramento genético também podem ser

identificados por meio das ROH, pois esses casos podem gerar padrões de ROH em regiões que controlam características de interesse zootécnico (Bosse et al., 2012; Ferenčaković, Sölkner e Curik, 2013a; Kim et al., 2013; Zhang et al., 2015).

O Plink v1.09 (Purcell et al., 2007) e o SNP & Variation Suite (SVS) da Golden Helix Dois são os dois programas mais utilizados para identificação de ROH (Curik, Ferenčaković e Sölkner, 2014b). Além desses programas, é possível identificar na literatura o Germline (Gusev et al., 2009), Beagle (Browning e Browning, 2010), CGATOH (Zhang et al., 2013), SNP1101 v.1.0 (Sargolzaei, 2014), BCFtools (Narasimhan et al., 2016) e detectRUNS implementado no software R (Biscarini et al., 2019).

Howrigan et al. (2011) compararam os programas Plink, Germline e Beagle e observaram que o software Plink foi capaz de gerar resultados confiáveis na detecção de autozigose, enquanto que o programa Germline apresentou dificuldades na resolução nos pontos iniciais e finais dos segmentos de ROH. Ainda, segundo estes autores, o programa Beagle foi muito conservador na detecção de seguimentos autozigotos. De acordo com Karimi (2013), os programas Plink, SVS e CGATOH possuem semelhanças na posição das ilhas de ROH no genoma, com pequenas diferenças na frequência das assinaturas de seleção de ROH. Comparando o SNP1101, o Plink e o BCFtools, Forutan et al. (2018), reportaram que o SNP1101 fornece as estimativas mais acuradas de endogamia com menos tempo computacional para detectar o segmento ROH em comparação com BCFtools e o Plink.

O detectRUNS é um pacote implementado no software R com o objetivo de detectar ROH e corridas de heterozigozidade (ROHet) em genomas diploides através do método de janelas deslizantes ou janelas consecutivas. O método baseado em janela deslizante é semelhante ao software Plink (Biscarini et al., 2020). Selli et al. (2020) comparou o software Plink e a metodologia de janelas deslizantes do detectRUNS em dados de 50K e 700K de bovinos leiteiros. O autor, observou que os resultados foram semelhantes em relação ao percentual

e o número de ROH por classe de comprimento, além disso ambos os métodos identificaram as mesmas regiões de ROH. Por outro lado, Smaragdov (2021) estudando o genoma de vacas leiteiras identificou que o software Plink obteve um número médio de ROH significativamente menor (P < 0,001) do que os calculados com a metodologia de janelas deslizantes do detectRUNS. O autor identificou a diferença no número de ROHs principalmente em classes de comprimento de 1 a 2 Mb, sugerindo que os resultados obtidos para ROH curtas dependeram do software e dos critérios utilizados para detectar ROH.

1.3.4 Critérios para detecção de ROH

O programa Plink e o detectRUNS utilizam a metodologia de janelas deslizantes para detectar ROH, sendo essa uma das mais utilizadas para esse tipo de análise (Biscarini et al., 2020; Purcell et al., 2007). A janela deslizante utilizadas nesses programas identificam ROH e selecionam apenas aqueles com um número mínimo de SNPs em homozigose, uma distância máxima entre SNPs, um comprimento mínimo, uma densidade mínima e com um número máximo de genótipos heterozigotos e ausentes. Vale ressaltar que cabe ao usuário definir os limites para esses critérios (Curik, Ferenčaković e Sölkner, 2014).

Peripolli et al. (2017) e Meyermans et al. (2020) revisaram diversos trabalhos (N=22, para ambos) e apresentaram os limites para os critérios utilizados nas análises de ROH em bovinos de corte, bovinos de leite, suínos, equinos, ovinos e aves, demonstrando uma limitação em estudos envolvendo ROH. Essa variação demonstra uma falta de consenso sobre quais os valores mais adequados para os critérios que definem ROH, que pode levar a viés nas estimativas de autozigosidade e dificultar a comparação entre estudos (Ferenčaković, Sölkner e Curik, 2013; Howrigan, Simonson e Keller, 2011; Ku et al., 2011). Estabelecer os critérios para definir padrões de ROH tem grande

importância, dado que eles vão identificar se os segmentos de ROH são IBD ou IBS, quando são IBS os segmentos são determinados como não autozigóticos.

Howrigan et al. (2011), considerando genótipos simulados de humanos, avaliaram o efeito das configurações de tamanho de janela, número máximo de genótipos heterozigotos e ausentes na janela, *threshold* e número mínimo de SNPs. O efeito do comprimento mínimo de ROH foi estudado exaustivamente por Purfield et al. (2012) e Ferenčaković et al. (2013). Mastrangelo et al. (2016) estudaram o efeito do número de SNPs heterozigotos permitidos em bovinos de raças italianas (Cinisara, Modicana, Reggiana e Italian Holstein). Meyermans et al. (2020) estudaram os efeitos de tamanho de janela, distância máxima entre SNPs e densidade mínima em diferentes espécies de interesse zootécnico.

Na literatura, foi encontrado um estudo em espécies aquáticas, no qual Chinchilla-Vargas et al. (2021) estudaram os efeitos dos critérios tamanhos de janela e número de genótipos heterozigotos permitidos na janela em dados de sequência de duas populações de Muskellunge (Esox masquinongy), espécie nativa da América do Norte. Sendo assim, são necessários mais estudos voltados aos efeitos desses critérios na detecção de ROH em espécies aquícolas.

1.3.5 Efeito da densidade de painéis de SNP nas análises de ROH

As informações genotípicas utilizadas para identificar as regiões em homozigose, podem ser realizadas por meio de painéis de baixa (BD) a alta densidade (AD), ou ainda com os dados de sequenciamento completo do genoma (em inglês "*whole genome sequence"*, WGS). A variação da densidade dos painéis de genotipagem pode afetar fortemente a qualidade de estimação de F_{ROH} dado que os painéis BD tendem a superestimar o F_{ROH} em comprimentos menores de ROH quando comparado com os painéis AD (Ferenčaković, Sölkner e Curik, 2013; Purfield et al., 2012). Outra consequência da densidade do painel

seria na quantificação das ROH, uma vez que a ampla cobertura do genoma pode ser capaz de detectar um número superior de corridas (Rebelato e Caetano, 2018).

Purfield et al. (2012) compararam a frequência de ROH em dados de média densidade (50K) e alta densidade (777K) de bovinos, concluíram que o painel 50K foi adequado apenas para identificar ROH com mais de 5 Mb. Ferencakovic et al. (2013) também identificaram o efeito da densidade na detecção de ROH e estimativas de $F_{ROH.}$ Segundo os autores, o painel de 50K superestimou o número de segmentos curtos (1–4 Mb) e o painel 777K subestimou segmentos maiores que 8Mb ao limitar o número de SNPs heterozigotos permitidos nos segmentos de ROH. Além disso, observou-se que quando utilizados dados de menor densidade as estimativas de F_{ROH} tendem a ser superestimadas. Em bubalinos, Macciotta et al. (2021) observaram que diferentes números de ROH foram detectados quando comparadas as mesmas populações genotipadas com diferentes densidades de SNPs. No mesmo trabalho, também foram comparados resultados de duas populações distintas no qual foi observada diferença no número de ROH curtos (<4 Mb), sendo que na população com menor densidade de SNPs foi observado uma superestimação desses segmentos. Essas diferenças observadas podem ser devido à ausência distribuição uniforme dos SNPs ao longo do genoma, ao processo de amostragem de SNPs, a uma baixa resolução dos dados e/ou também a uma alta frequência alélica (Kirin et al., 2010; Pemberton et al., 2012; Purfield et al., 2012; Zhang et al., 2015).

Uma possível solução para viabilizar análises de ROH com informações de maior densidade é a técnica de imputação de genótipos, podendo assim diminuir o viés decorrente do tipo de informação utilizada e consequentemente reduzir custos de obtenção das informações genômicas. Essa técnica consiste na genotipagem de animais com painéis de menor densidade e menor custo, e posteriormente imputar o seu genótipo para painéis de alta densidade (ou para

a sequência completa), com base em uma população referência genotipada com o painel mais denso (ou sequência) (Sargolzaei, Chesnais e Schenkel, 2014).

Zhang et al. (2018), compararam dados imputados e dados simulados em suínos, e concluíram que dados com uma taxa de genótipo ausentes moderada de 0,7 e três genótipos heterozigotos permitidos em uma janela deslizante poderiam gerar distribuição de ROH comparável aos dados não imputados. Keller et al. (2012), Christofidou et al. (2015) e Johnson et al. (2016) estudando doenças em humanos utilizaram a imputação para obter números de SNPs semelhantes entre amostras e, portanto, comparar diferentes conjuntos de dados. Os autores observaram que a detecção de ROH variou menos entre os conjuntos de dados em termos de estatísticas descritivas básicas e que F_{ROH} pode ter sofrido menos efeitos de taxas de genótipos ausentes e excesso de heterozigozidade.

Em espécies aquáticas, os estudos de ROH tem utilizado painéis de baixa a média densidade (Ambrosio et al., 2018; Aramburu et al., 2020; Yoshida et al., 2020) e até onde se sabe, não existem estudos que avaliem o efeito da utilização de painéis de diferentes densidades de SNPs e nem de dados imputados.

1.4 REFERÊNCIAS BIBLIOGRÁFICAS

AMBROSIO, D. ’; BESTIN A; FRASLIN C; DECHAMP N; DUPONT-NIVET M. **GENOMIC ESTIMATES OF DIVERSITY BETWEEN AND WITHIN FRENCH RAINBOW TROUT POPULATIONS**Aqua 2018-World Aquaculture Society Meeting. **Anais**...2018

ANUÁRIO 2021 Peixe BR da Piscicultura. . [s.l: s.n.]. Disponível em: <www.peixebr.com.br>.

ARAMBURU, O.; CEBALLOS, F.; CASANOVA, A.; MOAN, A. LE; HEMMER-HANSEN, J.; BEKKEVOLD, D.; BOUZA, C.; MARTÍNEZ, P. Genomic Signatures After Five Generations of Intensive Selective Breeding: Runs of Homozygosity and Genetic Diversity in Representative Domestic and Wild

Populations of Turbot (Scophthalmus maximus). **Frontiers in Genetics**, v. 11, 3 abr. 2020.

BENTSEN, H. B. *et al.* Genetic improvement of farmed tilapias: growth performance in a complete diallel cross experiment with eight strains of Oreochromis niloticus 1. **Aquaculture**, v. 160, p. 145–173, 1998.

BENTSEN, H. B. *et al.* Genetic improvement of farmed tilapias: Response to five generations of selection for increased body weight at harvest in Oreochromis niloticus and the further impact of the project. **Aquaculture**, v. 468, p. 206–217, 1 fev. 2017.

BENTSEN, H. B.; OLESEN˚, I. O. Designing aquaculture mass selection programs to avoid high inbreeding rates. **Aquaculture**, v. 204, p. 349–359, 2002.

BISCARINI, F.; COZZI, P.; GASPA, G.; MARRAS, G. **Detect Runs of Homozygosity and Runs of Heterozygosity in Diploid Genomes.**, 2019.

BISCARINI, F.; MASTRANGELO, S.; CATILLO, G.; SENCZUK, G.; CIAMPOLINI, R. Insights into genetic diversity, runs of homozygosity and heterozygosity-rich regions in maremmana semi-feral cattle using pedigree and genomic data. **Animals**, v. 10, n. 12, p. 1–17, 2020a.

___. Insights into genetic diversity, runs of homozygosity and heterozygosity-rich regions in maremmana semi-feral cattle using pedigree and genomic data. **Animals**, v. 10, n. 12, p. 1–17, 2020b.

BOLIVAR, R. B.; NEWKIRK, G. F.; ECIJA, N. **Response to within family selection for body weight ž / in Nile tilapia Oreochromis niloticus using a single-trait animal model Aquaculture**. [s.l: s.n.]. Disponível em: <www.elsevier.comrlocateraqua-online>.

BOSSE, M.; MEGENS, H. J.; MADSEN, O.; PAUDEL, Y.; FRANTZ, L. A. F.; SCHOOK, L. B.; CROOIJMANS, R. P. M. A.; GROENEN, M. A. M. Regions of Homozygosity in the Porcine Genome: Consequence of Demography and the Recombination Landscape. **PLoS Genetics**, v. 8, n. 11, nov. 2012.

BREDA, F. C.; EUCLYDES, R. F.; PEREIRA, C. S.; TORRES, R. DE A.; CARNEIRO, P. L. S.; SARMENTO, J. L. R.; FILHO, R. DE A. T.; FRANÇA MOITA, A. K. F. Endogamia e Limite de Seleção em Populações Selecionadas Obtidas por Simulação. **Revista Brasileira de Zootecnia**, v. 33, n. 6, p. 2017–2025, 2004.

BROWNING, S. R.; BROWNING, B. L. High-Resolution Detection of Identity by Descent in Unrelated Individuals. **American Journal of Human Genetics**, v. 86, n. 4, p. 526–539, 9 abr. 2010.

CÁDIZ, M. I.; LÓPEZ, M. E.; DÍAZ-DOMÍNGUEZ, D.; CÁCERES, G.; YOSHIDA, G. M.; GOMEZ-UCHIDA, D.; YÁÑEZ, J. M. Whole genome re-sequencing reveals recent signatures of selection in three strains of farmed Nile tilapia (Oreochromis niloticus). **Scientific Reports**, v. 10, n. 1, 1 dez. 2020.

CENDRON, F. *et al.* Genome-wide snp analysis reveals the population structure and the conservation status of 23 Italian chicken breeds. **Animals**, v. 10, n. 8, p. 1–16, 1 ago. 2020.

CHARO-KARISA, H.; REZK, M. A.; BOVENHUIS, H.; KOMEN, H. Heritability of cold tolerance in Nile tilapia, Oreochromis niloticus, juveniles. **Aquaculture**, v. 249, n. 1–4, p. 115–123, 12 set. 2005.

CHINCHILLA-VARGAS, J.; MEERBEEK, J. R.; ROTHSCHILD, M. F.; BERTOLINI, F. Signatures of selection and genomic diversity of Muskellunge (Esox masquinongy) from two populations in North America. **Research Square**, p. 1–29, 2021.

CHRISTOFIDOU, P. *et al.* Runs of Homozygosity: Association with Coronary Artery Disease and Gene Expression in Monocytes and Macrophages. **American Journal of Human Genetics**, v. 97, n. 2, p. 228–237, 6 ago. 2015.

CURIK, I.; FERENČAKOVIĆ, M.; SÖLKNER, J. Inbreeding and runs of homozygosity: A possible solution to an old problem. **Livestock Science**, v. 166, n. 1, p. 26–34, 2014a.

___. Inbreeding and runs of homozygosity: A possible solution to an old problem. **Livestock Science**, v. 166, n. 1, p. 26–34, 2014b.

DENISKOVA, T.; DOTSEV, A.; SELIONOVA, M.; BREM, G.; ZINOVIEVA, N. Biodiversity of russian local sheep breeds based on pattern of runs of homozygosity†. **Diversity**, v. 13, n. 8, 1 ago. 2021.

EKNATH, A. E. *et al.* Genetic improvement of farmed tilapias: the growth performance of eight strains of Oreochromis niloticus tested in different farm environments. **Genetics in Aquaculture**, p. 171–188, 1993.

ELGHOBASHY, H. **Aquaculture genetics research in Egypt** (B. O. Acosta & M. V. Grupta, Eds.)Fish
genetics research in member countries and institutions of the International

Network on Genetics in Aquaculture. **Anais**...Penang, Malaysia: ICLARM, 2001Disponível em: <http://hdl.handle.net/1834/845>. Acesso em: 8 dez. 2021

FALCONER, D.; MACKAY, T. **Introduction to quantitative genetics**. 4. ed. New York: Logman, 1996.

FAO. **EL ESTADO MUNDIAL DE LA PESCA Y LA ACUICULTURAOrganización de las naciones unidas para la alimentación y la agricultura**FOOD & AGRICULTURE ORG, , 2018.

___. **El estado mundial de la pesca y la acuicultura**Organización de las Naciones Unidas para la Alimentación y la Agricultura . **Anais**...FAO, 19 jun. 2020

FERENCAKOVIC, M.; HAMZIC, E.; GREDLER, B.; CURIK, I.; SÖLKNER, J. Runs of Homozygosity Reveal Genome-wide Autozygosity in the Austrian Fleckvieh Cattle. **Agriculturae Conspectus Scientifi cus**, v. 76, n. 4, p. 325–328, 2011.

FERENČAKOVIĆ, M.; SÖLKNER, J.; CURIK, I. Estimating autozygosity from high-throughput information: Effects of SNP density and genotyping errors. **Genetics Selection Evolution**, v. 45, n. 1, 29 out. 2013a.

___. Estimating autozygosity from high-throughput information: Effects of SNP density and genotyping errors. **Genetics Selection Evolution**, v. 45, n. 1, 29 out. 2013b.

FORUTAN, M.; ANSARI MAHYARI, S.; BAES, C.; MELZER, N.; SCHENKEL, F. S.; SARGOLZAEI, M. Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. **BMC Genomics**, v. 19, n. 1, 27 jan. 2018.

FRANKLIN, I. R.; FRANKHAM, R. How large must populations be to retain evolutionary potential? **Animal Conservation**, v. 1, n. 1, p. 69–70, fev. 1998.

GHOREISHIFAR, S. M.; MORADI-SHAHRBABAK, H.; FALLAHI, M. H.; MORADI-SHAHRBABAK, M.; ABDOLLAHI-ARPANAHI, R.; KHANSEFID, M. Genomic measures of inbreeding coefficients and genome-wide scan for runs of homozygosity islands in Iranian river buffalo, Bubalus bubalis. **BMC Genetics**, v. 21, n. 1, 10 fev. 2020.

GJEDREM, T.; BARANSKI, M. The Success of Selective Breeding in Aquaculture. *Em*: **Reviews: Methods and Technologies in Fish Biology and Fisheries**. Springer, Dordrecht: [s.n.]. v. 10p. 13–23.

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, 20 jun. 2012.

GJEDREM, T.; RYE, M. Selection response in fish and shellfish: a review. **Reviews in Aquaculture**, v. 10, n. 1, p. 168–179, 1 mar. 2018.

GJERDE, B.; MENGISTU, S. B.; ØDEGÅRD, J.; JOHANSEN, H.; ALTAMIRANO, D. S. Quantitative genetics of body weight, fillet weight and fillet yield in Nile tilapia (Oreochromis niloticus). **Aquaculture**, v. 342–343, n. 1, p. 117–124, 15 abr. 2012.

GOMEZ-RAYA, L.; RODRÍGUEZ, C.; BARRAGÁN, C.; SILIÓ, L. Genomic inbreeding coefficients based on the distribution of the length of runs of homozygosity in a closed line of Iberian pigs. **Genetics Selection Evolution**, v. 47, n. 1, 16 out. 2015.

GURGUL, A.; SZMATOŁA, T.; TOPOLSKI, P.; JASIELCZUK, I.; ŻUKOWSKI, K.; BUGNO-PONIEWIERSKA, M. The use of runs of homozygosity for estimation of recent inbreeding in Holstein cattle. **Journal of Applied Genetics**, v. 57, n. 4, p. 527–530, 1 nov. 2016.

GUSEV, A.; LOWE, J. K.; STOFFEL, M.; DALY, M. J.; ALTSHULER, D.; BRESLOW, J. L.; FRIEDMAN, J. M.; PE'ER, I. Whole population, genome-wide mapping of hidden relatedness. **Genome Research**, v. 19, n. 2, p. 318–326, fev. 2009.

GUTIÉRREZ-REINOSO, M. A.; APONTE, P. M.; GARCÍA-HERREROS, M. A review of inbreeding depression in dairy cattle: current status, emerging control strategies, and future prospects. **Journal of Dairy Research**, v. 89, n. 1, p. 3–12, 28 fev. 2022.

HOLTSMARK, M.; KLEMETSDAL, G.; SONESSON, A. K.; WOOLLIAMS, J. A. Establishing a base population for a breeding program in aquaculture, from multiple subpopulations, differentiated by genetic drift: I. Effects of the number of subpopulations, heritability and mating strategies using optimum contribution selection. **Aquaculture**, v. 274, n. 2–4, p. 232–240, 5 fev. 2008.

HOWRIGAN, D. P.; SIMONSON, M. A.; KELLER, M. C. Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. **BMC Genomics**, v. 12, 23 set. 2011.

HUDSON, N. J.; PORTO-NETO, L.; KIJAS, J. W.; REVERTER, A. Compression distance can discriminate animals by genetic profile, build relationship matrices

and estimate breeding values. **Genetics Selection Evolution**, v. 47, n. 1, 13 out. 2015.

ISLAM, R.; LI, Y.; LIU, X.; BERIHULAY, H.; ABIED, A.; GEBRESELASSIE, G.; MA, Q.; MA, Y. Genome-wide runs of homozygosity, effective population size, and detection of positive selection signatures in six chinese goat breeds. **Genes**, v. 10, n. 11, 1 nov. 2019.

JOHNSON, E. C. *et al.* No Reliable Association between Runs of Homozygosity and Schizophrenia in a Well-Powered Replication Study. **PLoS Genetics**, v. 12, n. 10, 1 out. 2016.

KARIMI, Z. **Runs of Homozygosity patterns in Taurine and Indicine cattle breeds**. [s.l: s.n.].

KELLER, M. C. *et al.* Runs of homozygosity implicate autozygosity as a schizophrenia risk factor. **PLoS Genetics**, v. 8, n. 4, 1 abr. 2012.

KIM, E. S.; COLE, J. B.; HUSON, H.; WIGGANS, G. R.; TASSEL, C. P. VAN; CROOKER, B. A.; LIU, G.; DA, Y.; SONSTEGARD, T. S. Effect of artificial selection on runs of homozygosity in U.S. Holstein cattle. **PLoS ONE**, v. 8, n. 11, 14 nov. 2013.

KIRIN, M.; MCQUILLAN, R.; FRANKLIN, C. S.; CAMPBELL, H.; MCKEIGUE, P. M.; WILSON, J. F. Genomic runs of homozygosity record population history and consanguinity. **PLoS ONE**, v. 5, n. 11, 2010.

KU, C. S.; NAIDOO, N.; TEO, S. M.; PAWITAN, Y. Regions of homozygosity and their impact on complex diseases and traits. **Human Genetics**, v. 129, n. 1, p. 1–15, jan. 2011.

LAFRENTZ, B. R.; LOZANO, C. A.; SHOEMAKER, C. A.; GARCÍA, J. C.; XU, D. H.; LØVOLL, M.; RYE, M. Controlled challenge experiment demonstrates substantial additive genetic variation in resistance of Nile tilapia (Oreochromis niloticus) to Streptococcus iniae. **Aquaculture**, v. 458, p. 134–139, 1 maio 2016.

LUCAS, J. S.; SOUTHGATE, P. C.; TUCKER, C. S. **Aquaculture: Farming aquatic animals and plants**. 3rd Edition ed. [s.l: s.n.]. v. 2

MA, Y.; DING, X.; QANBARI, S.; WEIGEND, S.; ZHANG, Q.; SIMIANER, H. Properties of different selection signature statistics and a new strategy for combining them. **Heredity**, v. 115, n. 5, p. 426–436, 1 nov. 2015.

MAAS, P.; GRZEGRZÓŁKA, B.; KRESS, P.; OBERLE, M.; JUDAS, M.; VALERIE KREMER-RÜCKER, P. Prediction of body composition in mirror carp (*Cyprinus carpio*) by using linear measurements in vivo and computed tomography post-mortem. **Archives Animal Breeding**, v. 63, n. 1, p. 69–80, 25 fev. 2020.

MACCIOTTA, N. P. P.; COLLI, L.; CESARANI, A.; AJMONE-MARSAN, P.; LOW, W. Y.; TEARLE, R.; WILLIAMS, J. L. The distribution of runs of homozygosity in the genome of river and swamp buffaloes reveals a history of adaptation, migration and crossbred events. **Genetics Selection Evolution**, v. 53, n. 1, 1 dez. 2021.

MALÉCOT, G. **Les mathématiques de l´hérédité**. Paris: [s.n.]. Disponível em: <https://www.persee.fr/doc/linly_0366-1326_1948_num_17_10_8510_t1_0203_0000_2>.

MARRAS, G.; GASPA, G.; SORBOLINI, S.; DIMAURO, C.; AJMONE-MARSAN, P.; VALENTINI, A.; WILLIAMS, J. L.; MACCIOTTA, N. P. P. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. **Animal Genetics**, v. 46, n. 2, p. 110–121, 1 abr. 2015.

MASTRANGELO, S.; TOLONE, M.; GERLANDO, R. DI; FONTANESI, L.; SARDINA, M. T.; PORTOLANO, B. Genomic inbreeding estimation in small populations: Evaluation of runs of homozygosity in three local dairy cattle breeds. **Animal**, v. 10, n. 5, p. 746–754, 1 maio 2016.

MCQUILLAN, R. *et al.* Runs of Homozygosity in European Populations. **American Journal of Human Genetics**, v. 83, n. 3, p. 359–372, 12 set. 2008.

MEYERMANS, R.; GORSSEN, W.; BUYS, N.; JANSSENS, S. How to study runs of homozygosity using plink? a guide for analyzing medium density snp data in livestock and pet species. **BMC Genomics**, v. 21, n. 1, 29 jan. 2020.

MI, M. P.; CHAPMAN, A. B.; TYLER, W. J. Effects of Mating System on Production Traits in Dairy Cattle. **Journal of Dairy Science**, v. 48, n. 1, p. 77–84, 1965.

___. Effects of Mating System on Production Traits in Dairy Cattle. **Journal of Dairy Science**, v. 48, n. 1, p. 77–84, 1965b.

NARASIMHAN, V.; DANECEK, P.; SCALLY, A.; XUE, Y.; TYLER-SMITH, C.; DURBIN, R. BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. **Bioinformatics**, v. 32, n. 11, p. 1749–1751, 1 jun. 2016.

NGUYEN, N. H.; PONZONI, R. W.; ABU-BAKAR, K. R.; HAMZAH, A.; KHAW, H. L.; YEE, H. Y. Correlated response in fillet weight and yield to selection for increased harvest weight in genetically improved farmed tilapia (GIFT strain), Oreochromis niloticus. **Aquaculture**, v. 305, n. 1–4, p. 1–5, jul. 2010.

NOWICKI, B. The influence of inbreeding on the utility value of black Pied Lowland cows. **Animal Breeding Abctract**, v. 32, p. 966, 1963.

OLIVEIRA, S. N. **Interação genótipo X ambiente para peso vivo e modelagem estatística para seleção genética em tilápias do nilo (Oreochromis niloticus)**. Maringá: [s.n.].

OLIVEIRA, S. N. DE; OLIVEIRA, C. A. L. DE; FILHO, L. A.; RESENDE, E. K. DE; BARRERO, N. M. L.; KUNITA, N. M.; RIBEIRO, R. P.; SANTANDER, V. F. A. Genetic parameters and morphometric characteristics of two generations from the GIFT strain of the Nile Tilapia. **Semina:Ciencias Agrarias**, v. 35, n. 6, p. 3457–3468, 1 nov. 2014.

OLIVEIRA, S. N. DE; OLIVEIRA, C. A. L. DE; FILHO, L. A.; RESENDE, E. K. DE; BARRERO, N. M. L.; KUNITA, N. M.; RIBEIRO, R. P.; SANTANDER, V. F. A. Genetic parameters and morphometric characteristics of two generations from the GIFT strain of the Nile Tilapia. **Semina:Ciencias Agrarias**, v. 35, n. 6, p. 3457–3468, 1 nov. 2014.

PEMBERTON, T. J.; ABSHER, D.; FELDMAN, M. W.; MYERS, R. M.; ROSENBERG, N. A.; LI, J. Z. Genomic patterns of homozygosity in worldwide human populations. **American Journal of Human Genetics**, v. 91, n. 2, p. 275–292, 10 ago. 2012.

PERIPOLLI, E.; MUNARI, D. P.; SILVA, M. V. G. B.; LIMA, A. L. F.; IRGANG, R.; BALDI, F. Runs of homozygosity: current knowledge and applications in livestock. **Animal Genetics**, v. 48, n. 3, p. 255–271, 1 jun. 2017.

PETERSEN, R. L.; GARCIA, J. E.; MELLO, G.; LIEDKE, A. M. R.; SINCERO, T. C. M.; GRISARD, E. C. ANÁLISE DA DIVERSIDADE GENÉTICA DE TILÁPIAS CULTIVADAS NO ESTADO DE SANTA CATARINA (BRASIL) UTILIZANDO MARCADORES MICROSSATÉLITES. **BOLETIM DO INSTITUTO DE PESCA**, v. 38, n. 4, p. 313–321, 2012.

PIRLEA, T.; ILEA, S. Effect of different inbreeding intensities on some production characters in cattle. **Animal Breeding Abstract**, v. 39, p. 3151, 1970.

PONZONI, R. W.; KHAW, H. L.; NGUYEN, N. H.; HAMZAH, A. Inbreeding and effective population size in the Malaysian nucleus of the GIFT strain of Nile tilapia (Oreochromis niloticus). **Aquaculture**, v. 302, n. 1–2, p. 42–48, 1 abr. 2010a.

___. Inbreeding and effective population size in the Malaysian nucleus of the GIFT strain of Nile tilapia (Oreochromis niloticus). **Aquaculture**, v. 302, n. 1–2, p. 42–48, 1 abr. 2010b.

PRCHAL, M. *et al.* Morphological predictors of slaughter yields using 3D digitizer and their use in a common carp breeding program. **Aquaculture**, v. 520, 15 abr. 2020.

PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M. A. R.; BENDER, D.; MALLER, J.; SKLAR, P.; BAKKER, P. I. W. DE; DALY, M. J.; SHAM, P. C. PLINK: A tool set for whole-genome association and population-based linkage analyses. **American Journal of Human Genetics**, v. 81, n. 3, p. 559–575, 2007.

PURFIELD, D. C.; BERRY, D. P.; MCPARLAND, S.; BRADLEY, D. G. Runs of homozygosity and population history in cattle. **BMC Genetics**, v. 13, 14 ago. 2012a.

___. Runs of homozygosity and population history in cattle. **BMC Genetics**, v. 13, 14 ago. 2012b.

QUEIROZ, S. A. DE; ALBUQUERQUE, L. G. DE; LANZONI, N. A. Efeito da Endogamia sobre Características de Crescimento de Bovinos da Raça Gir no Brasil 1. **Revista Brasileira de Zootecnia**, v. 29, n. 4, p. 1014–1019, 2000.

RAJAEE, A. H.; HUNTINGFORD, F. A.; RANSON, K. J.; MCANDREW, B. J.; PENMAN, D. J. The effect of male colouration on reproductive success in Nile tilapia (Oreochromis niloticus). **Aquaculture**, v. 308, n. SUPPL.1, 2010.

RAMEY, H. R.; DECKER, J. E.; MCKAY, S. D.; ROLF, M. M.; SCHNABEL, R. D.; TAYLOR, J. F. Detection of selective sweeps in cattle using genome-wide SNP data. **BMC Genomics**, v. 14, n. 1, 7 jun. 2013.

REBELATO, A. B.; CAETANO, A. R. Runs of homozygosity for autozygosity estimation and genomic analysis in production animals. **Pesquisa Agropecuaria Brasileira**, v. 53, n. 9, p. 975–984, 1 set. 2018.

RIBEIRO, C.; URBATZKA, R.; CASTRO, L. F. C.; CARROLA, J.; FONTAINHAS-FERNANDES, A.; MONTEIRO, R. A. F.; ROCHA, E.; ROCHA, M. J. In vitro exposure of Nile tilapia (Oreochromis niloticus) testis to estrogenic

endocrine disrupting chemicals: mRNA expression of genes encoding steroidogenic enzymes. **Toxicology Mechanisms and Methods**, v. 22, n. 1, p. 47–53, jan. 2012.

RUTTEN, M. J. M.; BOVENHUIS, H.; KOMEN, H. Modeling fillet traits based on body measurements in three Nile tilapia strains (Oreochromis niloticus L.). **Aquaculture**, v. 231, n. 1–4, p. 113–122, 5 mar. 2004.

RYE, M.; GJERDE, B.; GJEDREM, T. **Genetic Improvement Programs For Aquaculture Species In Developed Countries**Proceedings of the 9th world congress on genetics applied to livestock production. **Anais**...2010

SARGOLZAEI, M. **SNP1101 User's Guide**CanadaUNIVERSITY OF GUELPH, , 2014.

SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. A new approach for efficient genotype imputation using information from relatives. **BMC Genomics**, v. 15, n. 1, 17 jun. 2014.

SAURA, M.; FERNÁNDEZ, A.; VARONA, L.; FERNÁNDEZ, A. I.; CARA, M. Á. R. DE; BARRAGÁN, C.; VILLANUEVA, B. Detecting inbreeding depression for reproductive traits in Iberian pigs using genome-wide data. **Genetics Selection Evolution**, v. 47, n. 1, 12 dez. 2015.

SELLI, A., BUSSIMAN, F., DE SOUZA, W. L., TASSONI, L., ANDRIETTA, D. L. P., DE CARVALHO BALIEIRO, J. C., & VIEIRA, R. Métodos de detecção de corridas de homozigose (ROH) sob diferentes densidades de marcadores genéticos em bovinos leiteiros.

SHOEMAKER, C. A.; LOZANO, C. A.; LAFRENTZ, B. R.; GARCÍA, J. C.; SOTO, E.; XU, D. H.; BECK, B. H.; RYE, M. Additive genetic variation in resistance of Nile tilapia (Oreochromis niloticus) to Streptococcus iniae and S. agalactiae capsular type Ib: Is genetic resistance correlated? **Aquaculture**, v. 468, p. 193–198, 1 fev. 2017.

SILVA, J. M. DA; GIACHETTO, P. F.; CAMPOS DA SILVA, L. O.; CINTRA, L. C.; PAIVA, S. R.; CAETANO, A. R.; YAMAGISHI, M. E. B. Genomic variants revealed by invariably missing genotypes in Nelore cattle. **PLoS ONE**, v. 10, n. 8, 25 ago. 2015.

SMARAGDOV, M. G. Towards deciphering the structure of long homozygous stretches in cattle genome. **bioRxiv.** 2021.

STREWE, H. Inbreeding as a test of genetic health in cattle. **Dairy Science Abstracts**, v. 36, p. 3572, 1974.

SUEBSONG, W.; POOMPUANG, S.; SRISAPOOME, P.; KOONAWOOTRITTRIRON, S.; LUENGNARUEMITCHAI, A.; JOHANSEN, H.; RYE, M. Selection response for Streptococcus agalactiae resistance in Nile tilapia Oreochromis niloticus. **Journal of Fish Diseases**, v. 42, n. 11, p. 1553–1562, 1 nov. 2019.

THODESEN DA-YONG MA, J.; RYE, M.; WANG, Y. X.; YANG, K. S.; BENTSEN, H. B.; GJEDREM, T. Genetic improvement of tilapias in China: Genetic parameters and selection responses in growth of Nile tilapia (Oreochromis niloticus) after six generations of multi-trait selection for growth and fillet yield. **Aquaculture**, v. 322–323, p. 51–64, 21 dez. 2011.

VANDEPUTTE, M.; FRASLIN, C.; HAFFRAY, P.; BESTIN, A.; ALLAL, F.; KOCOUR, M.; PRCHAL, M.; DUPONT-NIVET, M. How to genetically increase fillet yield in fish: Relevant genetic parameters and methods to predict genetic gain. **Aquaculture**, v. 519, 30 mar. 2020.

VANDEPUTTE, M.; PULEDDA, A.; TYRAN, A. S.; BESTIN, A.; COULOMBET, C.; BAJEK, A.; BALDIT, G.; VERGNET, A.; ALLAL, F.; BUGEON, J.; HAFFRAY, P. Investigation of morphological predictors of fillet and carcass yield in European sea bass (Dicentrarchus labrax) for application in selective breeding. **Aquaculture**, v. 470, p. 40–49, 1 mar. 2017.

WRIGHT, S. COEFFICIENTS OF INBREEDING AND RELATIONSHIP. **The American Naturalist**, v. 26, n. 645, p. 330–338, 1922.

YÁÑEZ, J. M.; JOSHI, R.; YOSHIDA, G. M. Genomics to accelerate genetic improvement in tilapia. **Animal Genetics**, v. 51, n. 5, p. 658–674, 1 out. 2020.

YOSHIDA, G. M.; CÁCERES, P.; MARÍN-NAHUELPI, R.; KOOP, B. F.; YÁÑEZ, J. M. Estimates of autozygosity through runs of homozygosity in farmed coho salmon. **Genes**, v. 11, n. 5, 1 maio 2020a.

___. Estimates of autozygosity through runs of homozygosity in farmed coho salmon. **Genes**, v. 11, n. 5, 1 maio 2020b.

YOSHIDA, G. M.; YÁÑEZ, J. M.; OLIVEIRA, C. A. L. DE; RIBEIRO, R. P.; LHORENTE, J. P.; QUEIROZ, S. A. DE; CARVALHEIRO, R. Mate selection in aquaculture breeding using differential evolution algorithm. **Aquaculture Research**, v. 48, n. 11, p. 5490–5497, 1 nov. 2017.

ZAVAREZ, L. B. *et al.* Assessment of autozygosity in Nellore cows (Bos indicus) through high-density SNP genotypes. **Frontiers in Genetics**, v. 5, n. JAN, 2015.

ZHANG, L.; ORLOFF, M. S.; REBER, S.; LI, S.; ZHAO, Y.; ENG, C. cgaTOH: Extended Approach for Identifying Tracts of Homozygosity. **PLoS ONE**, v. 8, n. 3, 1 mar. 2013.

ZHANG, Q.; CALUS, M. P. L.; GULDBRANDTSEN, B.; LUND, M. S.; SAHANA, G. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. **BMC Genetics**, v. 16, n. 1, 22 jul. 2015.

ZHANG, Z. *et al.* Distribution of runs of homozygosity in Chinese and Western pig breeds evaluated by reduced-representation sequencing data. **Animal Genetics**, v. 49, n. 6, p. 579–591, 1 dez. 2018.

CAPÍTULO 2 - Critérios utilizados na detecção de corridas de homozigose (ROH) em tilápias do Nilo (*Oreochromis niloticus*) e suas implicações nos coeficientes de endogamia genômico (F_{ROH})

RESUMO - As corridas de homozigose (ROH) são trechos do genoma que apresentam genótipos consecutivos em homozigose. A correta identificação desses trechos depende de critérios previamente definidos e que, se não determinados com cautela, podem acarretar na super ou subestimação das análises. Portanto, o objetivo com este trabalho foi avaliar o impacto de diferentes critérios na detecção de ROH bem como na estimação de endogamia baseada em ROH (F_{ROH}) em tilápias do Nilo (*Oreochromis niloticus*). Foram utilizados genótipos de 2,848 indivíduos de duas populações distintas (POP_A= 1.376 e POP_B= 1.472) usando um painel comercial de 50K SNPs (*Single Nucleotide Polymorphisms*). Como controle de qualidade, foram removidos marcadores com base no equilíbrio de Hardy-Weinberg (*p-value*<10^{-15}) e taxa de genotipagem (*call rate*) menor que 0,95, totalizando 36,080 SNPs para as análises. As análises de ROH foram realizadas com uso do pacote detectRUNS implementado no *software* R e o F_{ROH} foi calculado de acordo com a proporção autozigótica do genoma. Os parâmetros que tiveram seus critérios avaliados foram: tamanho de janela e número mínimo de SNPs (10, 15, 20 e 50 SNPs); número máximo de genótipos ausentes permitidos (1, 2, 3 e 5); número máximo de genótipos heterozigotos permitidos (0, 1, 2, 3 e 5); espaçamento máximo entre SNPs (*GAP*) (50, 125, 370, 500 e 1000 Kb); comprimento mínimo de ROH (350, 500 e 1000 Kb; e densidade mínima de SNP (1/25, 1/50, 1/75 e 1/100 SNP/Kb). Os resultados dos diferentes cenários foram comparados utilizando estatísticas descritivas do número de ROH e os valores estimados de F_{ROH}. Observou-se que o F_{ROH} tende a ser superestimado quando se permite um maior número de genótipos heterozigotos e subestimado quando se reduz o GAP máximo permitido entre os SNPs adjacentes. Além disso, para o tamanho de janela e o número mínimo de SNPs, as estimativas são afetadas quando utilizados limites diferentes do proposto pelas fórmulas recomendadas na literatura. Para o comprimento mínimo, observou-se que o ideal é adequá-lo ao desequilíbrio de ligação da população e à densidade do painel de SNPs a fim de evitar resultados viesados. Para a configuração da densidade mínima, observou-se que uma densidade inferior a 1/50 SNP/Kb tende a subestimar as análises de ROH. Portanto, os resultados aqui apresentados indicam que os critérios utilizados para detectar os ROH tem grande impacto nas análises, e consequentemente, nas estimativas de F_{ROH} devido a super ou subestimação dos segmentos em homozigose. Os resultados possibilitaram indicar um conjunto de valores de critérios que fornecem estimativas de F_{ROH} mais plausíveis para as populações estudadas.

Palavras-chave: Corridas de Homozigose, Densidade de SNPs, Endogamia, *Oreochromis niloticus*

CHAPTER 2 - Criteria used to detect homozygosity runs (ROH) in Nile tilapia (*Oreochromis niloticus*) and their genomic inbreeding coefficients (F_{ROH}) implications

ABSTRACT - Runs of homozygosity (ROH) are segments of the genome that present consecutive homozygous genotypes. The correct identification of such segments depends on previously defined criteria since it can lead to over or underestimated analyses if not carefully determined. Therefore, the objective of this study was to evaluate the impact of different criteria on ROH detection as well as ROH-based inbreeding (F_{ROH}) estimation in Nile tilapia (*Oreochromis niloticus*). Genotypes from 2,848 individuals from two distinct populations (POP_A= 1,376 e POP_B= 1,472) were used using a commercial panel of 50K SNPs (Single Nucleotide Polymorphisms). As a quality control, markers were removed based on Hardy-Weinberg equilibrium (p-value<10^{-15}) and call rate lower than 0.95, totaling 36,080 SNPs for subsequent analyses. ROH analyses were performed using the detectRUNS package implemented in the R software, and the F_{ROH} was calculated according to the autozygosity proportion of the genome. The evaluated parameters were: window size and minimum number of SNPs (10, 15, 20, and 50 SNPs); maximum number of missing genotypes (1, 2, 3, and 5); maximum number of allowed heterozygous genotypes (0, 1, 2, 3, and 5); maximum gap between SNPs (50, 125, 370, 500, and 1000 Kb); minimum ROH length (350, 500, and 1000 Kb; and minimum SNP density (1/25, 1/50, 1/75, and 1/100 SNP/Kb). The results of the different scenarios were compared using descriptive statistics of the number ROH segments and the estimated values of F_{ROH}. It was observed that the F_{ROH} tends to be overestimated when allowing a greater number of heterozygous genotypes and underestimated when reducing the maximum GAP between two adjacent SNPs. In addition, for the window size and the minimum number of SNPs, the estimates were affected when setting different thresholds from those proposed by the formulas recommended in the literature. Regarding the minimum ROH length, it was observed that the ideal is to adapt it to the population linkage disequilibrium and to the SNP panel density to avoid biased results. For the configuration of the minimum SNP density, it was observed that a density lower than 1/50 SNP/Kb tends to underestimate the ROH analyses. Therefore, the results presented herein indicate that the criteria used to detect ROH has a great impact on the analyses, and consequently, on the F_{ROH} estimates due to the over or underestimation of the homozygous segments. The results made it possible to indicate a set of criteria values that provide more plausible F_{ROH} estimates for the studied populations.

Keywords: Inbreeding, Nile tilapia, Runs of homozygosity, SNP density

2.1 INTRODUÇÃO

A identificação de marcadores polimórficos de base única, do inglês "*Single nucleotide polymorphisms*" (SNPs), permitiu a elaboração de painéis com milhares de marcadores, e consequentemente, viabilizou a aplicação e o aprimoramento de diversas ferramentas genômicas (Rodrigues Caetano, 2009; Silva et al., 2015). Dentre as possíveis aplicações genômicas, pode-se citar as corridas de homozigose, do inglês *"Runs of homozygosity"* (ROH), que podem ser detectadas com uma maior precisão usando painéis de SNPs de alta a média densidade (Ferenčaković et al., 2013). As ROH se caracterizam por regiões onde o genoma apresenta segmentos em homozigose (Broman & Weber, 1998; Gibson et al., 2006). Esses segmentos ocorrem quando os progenitores, que possuem um ancestral comum, transmitem segmentos em comum para a sua progênie, ou seja, um indivíduo herda haplótipos idênticos por descendência (IBD, do inglês '*Identical by descendent*') (McQuillan et al., 2008), resultando em segmentos homozigotos no genoma da prole. As ROH são resultados da ocorrência combinada de fenômenos populacionais como endogamia, seleção artificial, deriva genética e gargalos populacionais (Falconer & Mackay, 1996).

A partir da identificação das ROH, é possível calcular a autozigosidade dos indivíduos, ou seja, estimar coeficientes de endogamia. Esse parâmetro pode ser considerado um dos mais relevantes na genética de populações, sendo definido pela probabilidade de que dois alelos tomados ao acaso em um *locus* homólogo dentro de um indivíduo sejam idênticos por descendência (IBD). Tradicionalmente, este coeficiente é calculado a partir do monitoramento das informações de pedigree (F_{PED}) (Wright, 1922). Entretanto, essa técnica de estimação é pouco precisa, dado que não considera a endogamia dos fundadores da população, para os quais não se tem informação de pedigree. Além disso, os efeitos amostrais decorrentes de seleção são desconsiderados ao estimar F_{PED} (Curik et al., 2014; Zhang et al., 2015). Os coeficientes de endogamia baseados em ROH (F_{ROH}) são calculados de acordo com a proporção do genoma em homozigose (McQuillan et al., 2008). As estimativas de F_{ROH}

apresentam vantagens em relação ao F_{PED}, dado que medem diretamente a homozigose do genoma (Gurgul et al., 2016; Marras et al., 2015; Zhang et al., 2015). Esses coeficientes também são de fácil interpretação biológica e frequentemente particionados em valores para cada cromossomo e/ou também para tamanhos específicos de segmentos em homozigose (Curik et al., 2014).

O F_{ROH} calculado por classe de comprimento de ROH auxilia a inferir o número de gerações desde o evento de endogamia ao se considerar que 1 cM equivale a 1 Mb (Howrigan et al., 2011). Isso ocorre devido a correlação aproximada entre o comprimento dos ROH e a distância com o antepassado comum devido a eventos de recombinação ao longo do tempo. Segmentos longos (>8Mb) refletem uma endogamia recente (inferior a 6 gerações) uma vez que a recombinação não teve tempo suficiente de quebrá-los, enquanto ROH pequenos e médios (<8Mb) tendem a refletir uma endogamia antiga (superior a 12 gerações) uma vez que esses segmentos foram quebrados por repetidos eventos de meiose (Kirin et al., 2010).

Uma das metodologias mais utilizadas para as análises de ROH é por meio de janelas deslizantes (Purcell et al., 2007) que, no geral, consiste em três etapas. Primeiramente, se determina o tamanho da janela deslizante em número de SNPs, o número máximo de genótipos heterozigotos e ausentes permitidos na janela em decorrência de possíveis erros e falhas de genotipagem. Depois de definidos os critérios da janela, os SNPs são selecionados pela proporção de janelas em homozigose que abrangem esse SNP, se essa proporção for superior a um limite definido, o SNP é declarado como estando em uma ROH. Por último, para selecionar os segmentos finais de ROH são aplicadas outras restrições nos segmentos homozigotos identificados. São verificados o espaçamento máximo entre dois SNPs (*GAP*), a densidade mínima de SNP por Kb, bem como o comprimento mínimo e o número mínimo de SNPs em uma ROH. Os segmentos ROH que não atendem essas condições são removidos.

Nas análises de ROH, para que se obtenha resultados confiáveis, é fundamental distinguir com precisão segmentos autozigóticos dos não

autozigóticos. Apesar da metodologia de janelas deslizantes ser um dos métodos mais recomendados para calcular a autozigosidade, há uma grande variação e uma falta de consenso sobre quais os valores mais adequados dos critérios para detectar uma ROH. Por exemplo, Hulsegge et al. (2022), utilizando dados de sequenciamento completo do genoma (WGS) de bovinos determinaram o número mínimo de SNPs em uma ROH como 50 SNPs, um comprimento mínimo de ROH de 300 Kb e um GAP de 1000 Kb, enquanto Beishova et al. (2022), estudando dados de sequência em bovinos utilizaram um número mínimo de SNPs de 30 SNPs, um comprimento mínimo de ROH de 1 Mb e um GAP de 500 Kb. A discrepância entre esses critérios de ROH torna difícil comparar estudos e definir o conjunto de valores mais adequados para cada população (Howrigan et al., 2011; Ku et al., 2011), podendo ser responsável pelo viés nas estimativas de autozigosidade baseadas em ROH (Ferenčaković et al., 2013).

Howrigan et al. (2011) utilizando genótipos simulados de humanos estudaram os impactos dos critérios número de genótipos heterozigotos e número mínimo de SNPs nas ROH. Os autores concluíram que o mais adequado seria não permitir SNPs heterozigotos e que o número mínimo de SNPs deve ser estabelecido de acordo o objetivo do número de gerações do evento de autozigosidade e o nível do erro de genotipagem dos dados. Resultados de Mastrangelo et al. (2016) em bovinos revelaram diferentes resultados de F_{ROH} dependendo se um, dois ou três genótipos heterozigotos eram permitidos, isto é, variando a tolerância à possíveis erros de genotipagem. Marras et al. (2015), utilizando um painel bovino de 50K SNPs, relataram que quando SNPs heterozigotos foram permitidos, o número de ROH longos aumentou drasticamente, indicando a superestimação do número de segmentos em ROH. Meyermans et al. (2020) identificaram que uma densidade mínima inferior a 1/40 SNP/Kb e um GAP máximo inferior a 500 Kb tende a acarretar na diminuição na cobertura de ROH ao longo do genoma, e consequentemente, na subestimação do F_{ROH} em diferentes espécies.

A adequada identificação de ROH depende também da variação da densidade dos painéis de genotipagem dado que os painéis de média densidade (50K) tendem a superestimar o F_{ROH} e o número de ROH menores que 4Mb quando comparado com os painéis de alta densidade (777K) (Ferenčaković et al., 2013; Purfield et al., 2012). Além disso, Ferenčaković et al (2013) indicaram que os painéis 50K podem falhar na detecção de SNPs heterozigotos presentes em uma ROH observada. Os mesmos autores identificaram que o painel 777K subestimou o número de ROH maior que 8Mb e concluíram que esse impacto pode ser devido à inclusão de genótipos heterozigotos acidentais decorrentes de possíveis erros de genotipagem. Uma possível solução seria tolerar SNPs heterozigotos. Porém, quando se permite um ou mais genótipos heterozigotos por janela, pode ocorrer a fusão de ROH adjacentes acarretando a detecção de segmentos falso-positivos. Portanto, os critérios utilizados na detecção de ROH necessitam ser ajustados de acordo com as particularidades dos dados que estão sendo analisados para obter resultados comparáveis entre os distintos estudos disponíveis na literatura. Além disso, a correta identificação de ROHs é essencial dentro dos programas de melhoramento genético animal, dado que auxilia na determinação de estratégias de acasalamento (Stachowicz et al., 2011) e na preservação da diversidade genética a longo prazo (Zavarez et al., 2015).

Em espécies aquícolas, poucos estudos têm sido descritos na literatura explorando análises de ROH (Aramburu et al., 2020; Chinchilla-Vargas et al., 2021; D'Ambrosio et al., 2019; Paul et al., 2022; Wang et al., 2022; Yoshida et al., 2020). Yoshida et al. (2020), usando um painel de 200K SNPs de três populações de salmão distintas Coho (*Oncorhynchus kisutch*), identificaram um comprimento médio de ROH entre 2.58 e 7.17 Mb e uma variação de F_{ROH} de 0.040 a 0.152. Chinchilla-Vargas et al. (2021) em dados de sequência de genoma completo de duas populações de Muskellunge (*Esox masquinongy*), espécie de peixe nativa da América do Norte, estudaram o efeito de três tamanhos de janelas (5, 10 e 20 SNPs) e genótipos heterozigotos (1, 2 e 3)

permitidos na detecção de ROH e estimação de F_{ROH}. Os resultados de F_{ROH} variaram de 0.00 a 0.44 e os autores concluíram que o resultado mais realista foi reproduzido com tamanho de janela de 20 SNPs e três genótipos heterozigotos permitidos.

Devido à importância econômica da tilápia do Nilo e a ausência de estudos de ROH com dados desta espécie, o objetivo com este trabalho foi avaliar o impacto da definição de diferentes critérios para a detecção de ROH bem como na estimação de F_{ROH} em tilápias do Nilo (*Oreochromis niloticus*), tendo em vista que o aumento na precisão da estimativa de F_{ROH} pode representar importante estratégia para o controle da endogamia nos programas de melhoramento de peixes.

2.2 MATERIAL E MÉTODOS

2.2.1 Populações

Os animais utilizados neste estudo pertencem a duas linhagens comerciais de tilápias do Nilo originárias das Filipinas (estação Carmen aquafarm), introduzidas na Costa Rica em 2005 (Aquacorporacion Internacional) e formadas por oito gerações selecionadas para taxa de crescimento. A primeira população (POP_A) originou-se de um cruzamento da linhagem GIFT (do inglês, '*Genetically Improved Farmed Tilapia*') provenientes de duas populações selvagens do Egito e do Quênia. A segunda população originou-se de tilápias oriundas de um cruzamento de linhagens domésticas de Israel, Cingapura, Taiwan e Tailândia (POP_B). Mais detalhes a respeito das populações estão disponíveis em estudos anteriores (Cádiz et al., 2020; Neira et al., 2016; Yáñez et al., 2020).

2.2.2 Genótipos e controle de qualidade

Foram utilizadas amostras de nadadeiras de 2.848 indivíduos (*n*=1.376 POP_A e *n*=1.472 POP_B) para a obtenção de material genético. O kit Qiagen DNeasy Blood & Tissue (Qiagen, Alemanha) foi utilizado para extrair e purificar as amostras de DNA de acordo com o protocolo do fabricante. Posteriormente, as amostras foram genotipadas usando o painel comercial SNP Illumina BeadChip 50K (Yáñez et al., 2020). Os genótipos de ambas as populações foram submetidos simultaneamente a um controle de qualidade, na qual marcadores foram removidos com base no equilíbrio de Hardy Weinberg (p-value<10^{-15}) e taxa de chamada do genótipo (*call rate*) inferior a 0,95, totalizando 36.080 SNPs para as análises subsequentes.

O controle de qualidade para alto desequilíbrio de ligação (LD) e baixa frequência alélica é amplamente utilizado em análises genômicas, entretanto estudos apontam que estes parâmetros contribuem para uma análise de ROH viesada (Ferenčaković et al., 2013; Hillestad et al., 2017; Kirin et al., 2010; Meyermans et al., 2020). Um alto LD pode aumentar o número de segmentos curtos falso-positivos acarretando resultados viesados. Sendo assim, caracterizamos o LD das duas populações a fim de identificar a necessidade de corrigir ROH falso positivo (causado por alto LD) por meio de um aumento no rigor do critério comprimento mínimo (Purfield et al., 2012). O LD foi calculado conforme o coeficiente de correlação quadrado de Pearson (r^2) obtido separadamente para as duas populações utilizando o software PLINK v.1.9 com os seguintes parâmetros --ld-window-kb 10.000, --ld-window 99.999 e --ld-window-r2 definido como zero.

2.2.3 Corridas de homozigose (ROH) e coeficientes de endogamia (F_{ROH} e F_{PED})

As ROH foram identificadas em cada animal utilizando o método de janelas deslizantes do pacote detectRUNS implementado no software R

(Biscarini et al., 2019). Primeiramente, definiu-se o cenário “Controle” para permitir avaliar individualmente o efeito da variação de cada parâmetro nas análises de ROH. Neste cenário, os seguintes parâmetros foram aplicados à detecção de ROH: (i) número mínimo de SNPs homozigotos consecutivos e tamanho de janela (L) de 15 SNPs com base na fórmula proposta por Lencz et al. (2007) e adaptado por Purfield et al. (2012):

$$L = \frac{log_e \frac{\alpha}{n_s n_i}}{log_e (1 - het)}$$

onde n_s corresponde ao número de SNPs genotipados por indivíduo, n_i é o número de indivíduos genotipados, α é a porcentagem de ROH falso-positivo (0.05), e het corresponde a heterozigosidade média em todos os SNPs; (ii) o limite para considerar um SNP em uma ROH (t) foi de 0,135, com base na fórmula proposta por Meyermans et al. (2020):

$$t = floor\left(\frac{N_{out} + 1}{L}, 3\right)$$

onde N_{out} corresponde ao número desejado de SNPs externos finais em ambos os lados de um segmento de ROH que não deve ser incluso no final de uma ROH, L é o tamanho da janela definido anteriormente; (iii) o comprimento mínimo de ROH de 1000 Kb com o objetivo de corrigir ROH falso positivo causado pelo LD; (iv) distância máxima permitida entre SNPs consecutivos (GAP) de 500 Kb, de acordo com Meyermans et al. (2020); (v) apenas um genótipo ausente permitido na janela, seguindo o *default* sugerido pelo DetectRUNS; (vi) nenhum genótipo heterozigoto permitido na janela, pressupondo ausência de erro de genotipagem; e (vii) densidade mínima de pelo menos 1 SNP por 50 Kb com o objetivo de assumir uma densidade superior a densidade média de 1 SNP por 25 Kb do painel de genotipagem pós controle de qualidade.

Com base no cenário controle, definimos outros cenários no qual cada parâmetro variou individualmente, permitindo assim avaliar o efeito desses parâmetros nas análises de ROH e F_{ROH} (Tabela Suplementar 1). O tamanho de

janela e o número mínimo de SNPs variaram simultaneamente de 10, 20 e 50 SNPs (J10, J20 e J50, respectivamente). O número máximo de genótipos ausentes permitidos foi de 2, 3 e 5 (M2, M3 e M5, respectivamente) e de genótipos heterozigotos foi de 1, 2, 3 e 5 (H1, H2, H3 e H5, respectivamente). Para GAP máximo, os seguintes valores foram adotados: 50, 125, 370 e 1000 Kb (G50, G125, G370 e G1000, respectivamente). Para comprimento mínimo foram usados os valores de 350 e 500 Kb (L350 e L500, respectivamente). Por fim, para densidade mínima os cenários foram de 1/25, 1/75 e 1/100 SNP/Kb (D25, D75 e D100, respectivamente).

Foram determinadas cinco categorias de ROH de acordo com o comprimento (Mb) a seguir: 1–2, 2–4, 4–8, 8–16 e >16 Mb, definidos como $ROH_{1\text{-}2Mb}$, $ROH_{2\text{-}4Mb}$, $ROH_{4\text{-}8Mb}$, $ROH_{8\text{-}16Mb}$ e $ROH_{>16Mb}$, respectivamente. Por fim, o coeficiente de endogamia geral (F_{ROH}^{i}) foi calculado para cada população pela seguinte fórmula (McQuillan et al., 2008):

$$F_{ROH}^{i} = \frac{\Sigma_{ROH}^{i}}{L_{GEN}}$$

em que Σ_{ROH}^{i} é a soma do comprimento de todos os segmentos em homozigose para os i^{th} animais e L_{GEN} é o comprimento total dos autossomos cobertos por marcadores. O L_{GEN} foi considerado de 906.216.665 pares de base, conforme o mapa de referência utilizado no estudo (O_niloticus_UMD_NMBU, GCA_001858045.3) (Conte et al., 2019). As estimativas de F_{ROH} também foram calculadas para cinco classes de comprimento: 1-2, 2-4, 4-8, 8-16 e >16 Mb, definidos como $F_{ROH1\text{-}2}$, $F_{ROH2\text{-}4}$, $F_{ROH4\text{-}8}$, $F_{ROH8\text{-}16}$ e $F_{ROH>16}$ a fim de predizer o número de gerações desde o evento de endogamia ao se considerar que 1 cM equivale a 1 Mb (Howrigan et al., 2011). Por fim, o coeficiente de endogamia com base no pedigree (F_{PED}) e as estatísticas do número de gerações selecionadas foram calculados considerando 130.049 animais, utilizando o programa BLUPF90 (Misztal et al., 2002).

2.3 RESULTADOS

Os resultados do cenário “Controle” foram comparados com os outros cenários utilizando estatísticas descritivas do número de segmentos de ROH, os valores estimados de F_{ROH}, possibilitando assim, identificar o conjunto de valores e parâmetros que fornecem estimativas mais plausíveis para as estimativas dos segmentos em homozigose. Os gráficos violinos para os critérios tamanho de janela, genótipos ausentes, comprimento mínimo e densidade mínima estão inseridos nas Figuras suplementares 1, 2, 3 e 4.

2.3.1 Tamanho de Janela e número mínimo de SNPs

Na tabela 1 (POP_A) e 2 (POP_B) são apresentados os resultados referentes ao número total de corridas de homozigose (*n*ROH) por classe de tamanho e o número médio de ROH por indivíduo (ROH_i). Foi observado um aumento de $ROH_{1–2Mb}$ e ROH_{2-4Mb} no cenário J10 em relação ao controle (J15) em ambas as populações. Por outro lado, observou-se que $ROH_{1–2Mb}$ e ROH_{2-4Mb} foram menos frequentes em relação ao cenário controle quando ocorreu o aumento nos valores determinados para estes dois critérios (J20 e J50).

Tabela 1 - Número total de corridas de homozigose (*n*ROH) por classe de tamanho e o número médio de ROH por indivíduo (ROH_i) para os diferentes cenários de determinação das ROH na população A (POP_A).

	$ROH_{1–2Mb}$	**ROH_{2-4Mb}**	**ROH_{4-8Mb}**	**ROH_{8-16Mb}**	**$ROH_{>16Mb}$**	***n*ROH**	**ROH_i**
Controle[1]	25.940,00	13.655,00	4.840,00	366,00	9,00	44.810,00	32,57
Tamanho de janela[2]							
J10	28.280,00	14.200,00	5.096,00	383,00	10,00	47.969,00	34,86
J20	16.861,00	13.589,00	4.833,00	366,00	9,00	35.658,00	25,91
J50	14.301,00	10.793,00	4.324,00	327,00	8,00	29.753,00	21,62
Genótipos Ausentes[3]							
M2	25.890,00	13.654,00	4.895,00	379,00	9,00	44.827,00	32,58
M3	25.885,00	13.658,00	4.903,00	379,00	9,00	44.834,00	32,58
M5	25.886,00	13.660,00	4.902,00	380,00	9,00	44.837,00	32,59
Genótipos Heterozigotos[4]							
H1	44.985,00	17.600,00	6.718,00	673,00	31,00	70.007,00	50,88
H2	82.517,00	28.930,00	9.205,00	1.060,00	46,00	121.758,00	88,49

H3	117.854,00	52.103,00	15.544,00	2.084,00	71,00	187.656,00	136,38
H5	113.985,00	88.041,00	45.314,00	11.348,00	657,00	259.345,00	188,48
Gap máximo[5]							
G50	6.422,00	131,00	0,00	0,00	0,00	6.553,00	4,76
G125	25.482,00	12.155,00	2.428,00	103,00	0,00	40.168,00	29,19
G370	25.817,00	13.564,00	4.691,00	346,00	9,00	44.427,00	32,29
G1000	25.886,00	13.653,00	4.845,00	384,00	9,00	44.777,00	32,54
Comprimento mínimo[6]							
L350	112.767,00	13.655,00	4.840,00	366,00	9,00	131.637,00	95,67
L500	75.907,00	13.655,00	4.840,00	366,00	9,00	94.777,00	68,88
Densidade mínima[7]							
D25	15.069,00	8.037,00	3.322,00	302,00	9,00	26.739,00	19,43
D75	27.767,00	13.754,00	4.845,00	366,00	9,00	46.741,00	33,97
D100	28.249,00	13.839,00	4.845,00	366,00	9,00	47.308,00	34,38

[1]**Controle:** cenário que permite a comparação com outros cenários nos quais os critérios adotados foram: número mínimo de SNPs e o tamanho de janela de 15 SNPs, threshold de 0.135, comprimento mínimo de 1000 Kb, GAP de 500 Kb, máximo número de genótipos ausentes permitidos na janela de 1, máximo número de genótipos heterozigotos permitidos na janela de 0 e densidade mínima de pelo menos 1 SNP por 50 Kb. [2]**Tamanho de Janela**: cenários com tamanho de janela de 10, 20 e 50 SNPs (J10, J20 e J50, respectivamente); [3]**Genótipos ausentes:** cenários com máximo de 2, 3 e 5 genótipos ausentes (M2, M3 e M5, respectivamente); [4]**Genótipos heterozigotos:** cenários com máximo 1, 2, 3 e 5 genótipos heterozigotos (H1, H2, H3 e H5, respectivamente); [5]**Gap máximo:** cenários com GAP máximo de 50, 125, 370 e 1000 Kb (G50, G125, G370 e G1000, respectivamente); [6]**Comprimento mínimo:** cenários com comprimento mínimo de 350 e 500 Kb (L350 e L500, respectivamente); [7]**Densidade mínima:** cenários com densidade mínima de 1/25, 1/75 e 1/100 SNP/Kb (D25, D75 e D100 respectivamente).

Tabela 2- Número total de corridas de homozigose (*n*ROH) por classe de tamanho e o número médio de ROH por indivíduo (ROH_i) para os diferentes cenários de determinação das ROH na população B (POP_B).

	$ROH_{1–2Mb}$	**ROH_{2-4Mb}**	**ROH_{4-8Mb}**	**ROH_{8-16Mb}**	**$ROH_{>16Mb}$**	***n*ROH**	**ROH_i**
Controle	27.909,00	16.764,00	5.750,00	477,00	15,00	50.915,00	34,59
Tamanho de janela[2]							
J10	29.700,00	17.466,00	6.050,00	514,00	15,00	53.745,00	36,51
J20	19.030,00	16.724,00	5.747,00	476,00	15,00	41.992,00	28,53
J50	15.703,00	13.997,00	5.103,00	422,00	12,00	35.237,00	23,94
Genótipos ausentes[3]							
M2	27.878,00	16.769,00	5.773,00	487,00	15,00	50.922,00	34,59
M3	27.865,00	16.775,00	5.775,00	488,00	15,00	50.918,00	34,59
M5	27.865,00	16.776,00	5.775,00	488,00	15,00	50.919,00	34,59
Genótipos Heterozigotos[4]							
H1	42.744,00	20.261,00	8.211,00	988,00	63,00	72.267,00	49,09
H2	77.779,00	30.288,00	10.558,00	1.525,00	97,00	120.247,00	81,69
H3	113.513,00	51.615,00	16.209,00	2.509,00	161,00	184007,00	125,01
H5	125.057,00	90.606,00	43.926,00	10.750,00	760,00	271099,00	184,17

Gap máximo[5]							
G50	7.962,00	263,00	0,00	0,00	0,00	8.225,00	5,59
G125	28.874,00	15.042,00	2.983,00	2.983,00	0,00	49.882,00	33,89
G370	27.885,00	16.805,00	5.512,00	450,00	12,00	50.664,00	34,42
G1000	27.853,00	16.742,00	5.765,00	490,00	15,00	50.865,00	34,56
Comprimento mínimo[6]							
L350	106.157,00	16.764,00	5.0750,00	477,00	15,00	129.163,00	87,75
L500	75.232,00	16.764,00	5.750,00	477,00	15,00	98.238,00	66,74
Densidade mínima[7]							
D25	16.194,00	10.287,00	3.784,00	360,00	12,00	30.637,00	20,81
D75	29.802,00	16.892,00	5.753,00	477,00	15,00	52.939,00	35,96
D100	30.381,00	17.013,00	5.753,00	477,00	15,00	53.639,00	36,44

Vide nota Tabela 1

2.3.3 Genótipos ausentes e heterozigotos

O acréscimo no número máximo de genótipos ausentes (M2, M3 e M5), quando comparado com o controle (M1) não afetou o número de ROH e a magnitude de F_{ROH} em ambas as populações (Tabelas 1, 2, 3 e 4). Em contrapartida, o aumento no critério número máximo de genótipos heterozigotos permitidos quando comparado ao cenário controle (H0) afetou as estimativas aumentando *nR*OH em todas as classes de comprimento, especialmente nas classes $ROH_{8\text{-}16Mb}$ e $ROH_{>16Mb}$ (Tabelas 1 e 2). Consequentemente, para ambas as populações, houve um aumento nas estimativas de F_{ROH} (Figura 1), cujos valores variaram do cenário controle (H0) para o H5 de 0,08233 a 0,63320 (POP_A) e de 0,09107 a 0,59890 (POP_B), respectivamente.

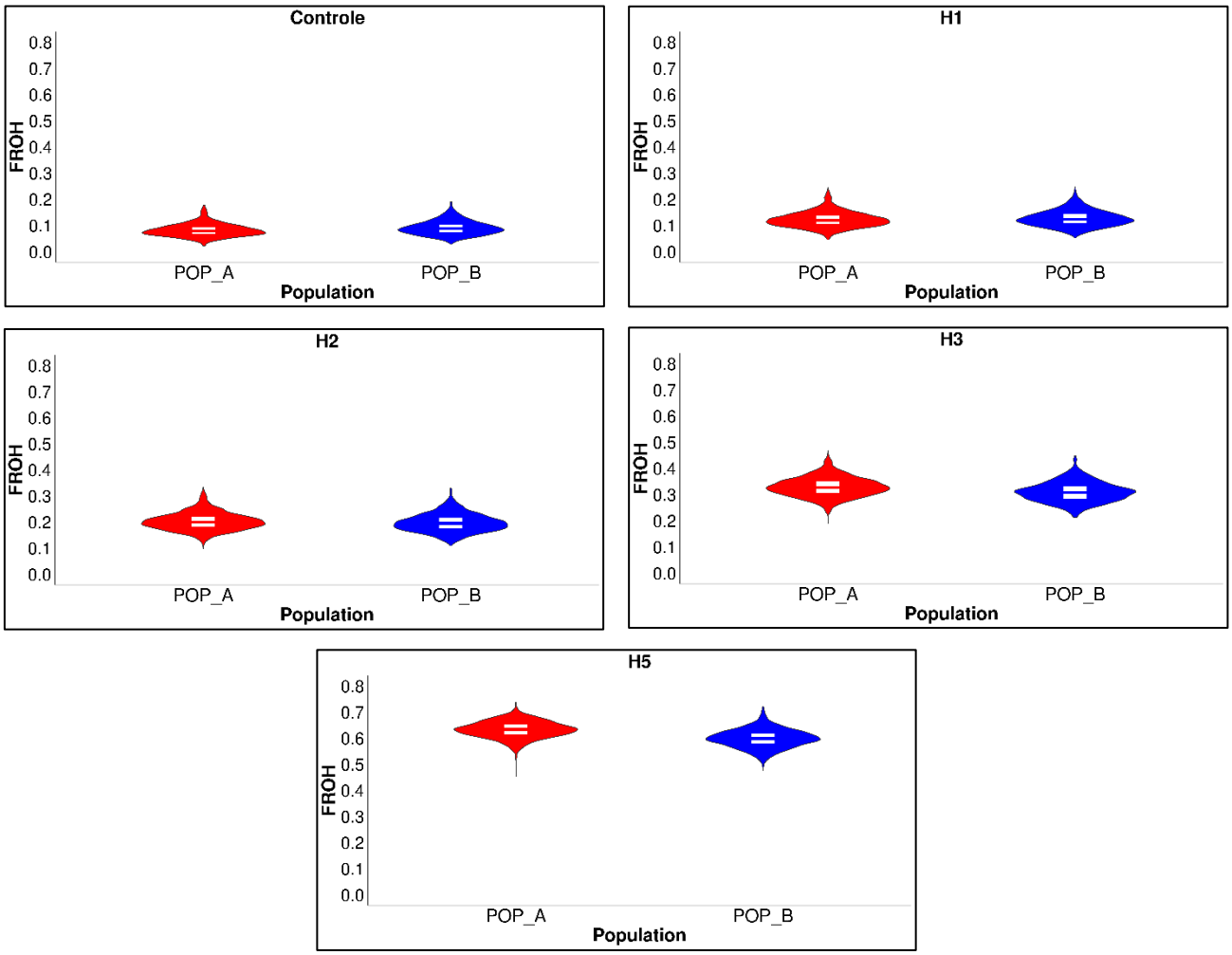


Figura 1- Gráfico violino do coeficiente de endogamia baseado nas corridas de homozigose (FROH) dos cenários controle (H0) e de 1, 2, 3 e 5 genótipos heterozigotos permitidos (H1, H2, H3 e H5) em duas populações de tilápia do Nilo (POP_A e POP_B).

2.3.4 GAP máximo

Os resultados do *n*ROH para o GAP máximo nas POP_A e POP_B foram subestimados nos cenários G50 e G125 quando comparados ao cenário controle (G500). No cenário G50 poucas corridas foram identificadas, apenas para ROH_{1-}

$_{2Mb}$ e ROH$_{2\text{-}4Mb}$. No cenário G125 não foram detectadas ROH$_{>16Mb}$. Vale ressaltar que a partir do cenário G370, o número de ROH estabilizou-se. O mesmo ocorreu com as estimativas de F_{ROH}, na qual houve pouca variação entre os cenários controle (G500), G370 e G1000. Os valores de F_{ROH} encontrados nesses três cenários variaram de 0,08233 a 0,08244 para POP_A e de 0,08993 a 0,09117 para a POP_B (Figura 2), respectivamente.

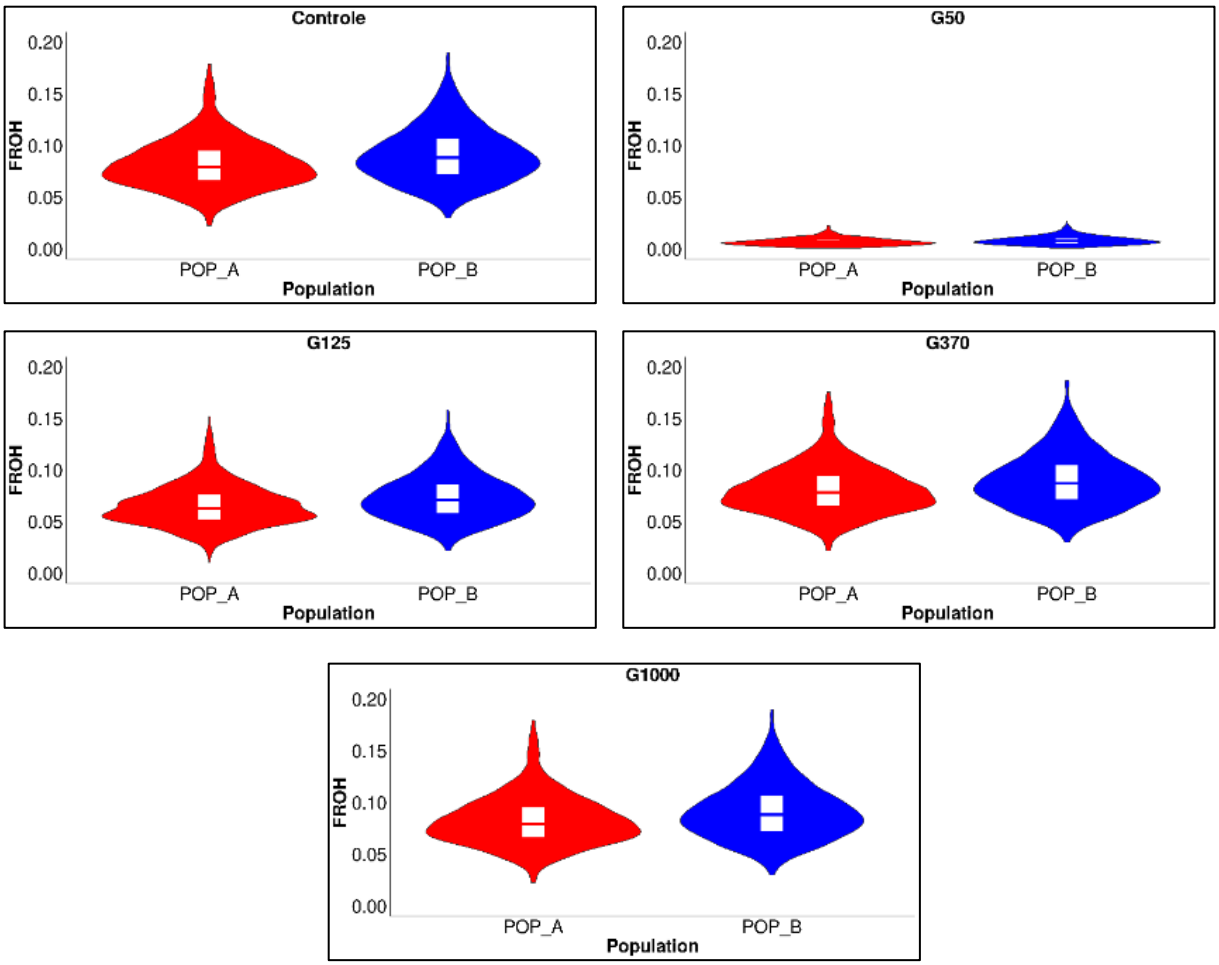


Figura 2- Gráfico violino do coeficiente de endogamia baseado em corridas de homozigose (F_{ROH}) dos cenários controle (G500) e dos cenários com 50 Kb, 125 Kb, 370 Kb e 1000 Kb de distância máxima permitida entre SNPs (G50, G125, G370 e G1000) em duas populações de tilápia do Nilo (POP_A e POP_B).

2.3.5 Comprimento mínimo e desequilíbrio de ligação (LD)

Os cenários de comprimento mínimo afetaram a identificação de corridas de comprimento curto ($ROH_{1\text{-}2Mb}$), em ambas as populações. Notou-se que o aumento do comprimento permitido na detecção dos ROH resultou na diminuição dos segmentos curtos. Foi identificado um r^2 médio de 0,0546 para POP_A e 0,0621 para a POP_B considerando todos os SNPs adjacentes. O decaimento da média de desequilíbrio de ligação (Figura Suplementar 5) revelou um rápido decréscimo de r^2 à medida que a distância entre os marcadores aumentava. O decréscimo mais significativo foi observado nos primeiros 1500 kb de distância e, em seguida, o nível de r^2 diminuiu para valores próximos de 0,05.

2.3.6 Densidade mínima

Nas duas populações, notou-se que o cenário de maior densidade requerida (D25) ocasionou na redução do $ROH_{1\text{-}2M}$ e a partir da densidade mínima permitida de 50 Kb (cenário controle) o número de ROH variou pouco.

2.3.7 F_{ROH} e F_{ped}

Os cenários J20 e J50 quando comparado ao controle (J15) verificou-se um decréscimo no F_{ROH} quando se aumentava o rigor do critério tamanho de janela (Tabela 3 e 4). O acréscimo no número máximo de genótipos ausentes (M2, M3 e M5), quando comparado com o controle (M1) não afetou a magnitude de F_{ROH} em ambas as populações. O F_{ROH} foi superestimado ao passo que se diminuiu o critério comprimento mínimo de ROH, no cenário controle (L1000) o valor estimado foi de 0,08233 e 0,09107, para L350 foi de 0,12246 e 0,12551 e para L500 foi de 0,11006 e 0,11574, para as populações POP_A e POP_B, respectivamente. Nas duas populações, notou-se que o cenário de maior densidade requerida (D25) ocasionou na diminuição da F_{ROH}, em comparação aos demais cenários.

As estimativas médias de F_{ROH} no cenário controle diferiram entre as duas populações sendo a POP_B mais endogâmica que a POP_A. Por classe de ROH, em ambas as populações, o F_{ROH} tendeu a diminuir ao passo que aumentava o tamanho de comprimento médio de ROH com exceção de $F_{ROH2\text{-}4Mb}$ onde foi observado um leve acréscimo de endogamia em comparação ao $F_{ROH1\text{-}2Mb}$. Os valores do coeficiente de endogamia médio F_{PED} para POP_A e POP_B obtidos foram iguais a 2,01% e 2,18%, respectivamente. Por fim, o número médio de gerações foi igual a 5,052 e 5,017, para POP_A e POP_B, respectivamente, e o número máximo para ambos as populações foi de 8 gerações.

Tabela 3- Coeficientes de endogamia médio geral (desvio padrão) baseado nos segmentos em homozigose (F_{ROH}) para os diferentes cenários na população A (POP_A) e diferentes classes de tamanho de ROH ($F_{ROH1-2Mb}$, $F_{ROH2-4Mb}$, $F_{ROH4-8Mb}$, $F_{ROH8-16Mb}$ e $F_{ROH>16Mb}$).

	F_{ROH}	$F_{ROH\ 1–2Mb}$	$F_{ROH\ 2-4Mb}$	$F_{ROH\ 4-8Mb}$	$F_{ROH\ 8-16Mb}$	$F_{ROH\ >16Mb}$
Controle[1]	0,08233 (0,02431)	0,02887 (0,00779)	0,03000 (0,01205)	0,02049 (0,01253)	0,00283 (0,00583)	0,00013 (0,00166)
Tamanho de janela[2]						
J10	0,08727 (0,02514)	0,03145 (0,00828)	0,03119 (0,01239)	0,02153 (0,01282)	0,00297 (0,00597)	0,00015 (0,00173)
J20	0,07334 (0,02338)	0,02003 (0,00635)	0,02988 (0,01203)	0,02046 (0,01252)	0,00283 (0,00582)	0,00013 (0,00166)
J50	0,06182 (0,02109)	0,01713 (0,00615)	0,02394 (0,01020)	0,01812 (0,01161)	0,00252 (0,00540)	0,00012 (0,00157)
Genótipos ausentes[3]						
M2	0,08266 (0,02427)	0,02883 (0,00779)	0,02877 (0,01197)	0,02075 (0,01258)	0,00293 (0,00590)	0,00013 (0,00166)
M3	0,08271 (0,02425)	0,02882 (0,00779)	0,03003 (0,01196)	0,02080 (0,01258)	0,00293 (0,00590)	0,00013 (0,00166)
M5	0,08272 (0,02425)	0,02882 (0,00778)	0,03004 (0,01195)	0,02080 (0,01259)	0,00294 (0,00590)	0,00013 (0,00166)
Genótipos Heterozigotos[4]						
H1	0,12265 (0,02955)	0,04920 (0,01026)	0,03867 (0,01315)	0,02901 (0,01528)	0,00529 (0,00802)	0,00047 (0,00326)
H2	0,20260 (0,03364)	0,09068 (0,01272)	0,06299 (0,01561)	0,03986 (0,01812)	0,00835 (0,01038)	0,0007 (0,00422)
H3	0,32930 (0,03700)	0,13125 (0,01438)	0,11390 (0,01880)	0,06665 (0,02251)	0,01634 (0,01432)	0,00117 (0,00539)
H5	0,63320 (0,03203)	0,13106 (0,01554)	0,19820 (0,02369)	0,20047 (0,03260)	0,09342 (0,03282)	0,01008 (0,01540)
Gap máximo[5]						
G50	0,00701 (0,00370)	0,00413 (0,00430)	0,00023 (0,00073)	0,00000 (0,00000)	0,00000 (0,00000)	0,00000 (0,00000)
G125	0,06544 (0,02017)	0,02845 (0,00806)	0,02631 (0,01063)	0,00992 (0,00839)	0,00076 (0,00289)	0,00000 (0,00000)
G370	0,08119 (0,02411)	0,02874 (0,00781)	0,02975 (0,01190)	0,01989 (0,01220)	0,00268 (0,00569)	0,00013 (0,00166)
G1000	0,08244 (0,02437)	0,02839 (0,00778)	0,03000 (0,01207)	0,02051 (0,01254)	0,00297 (0,00602)	0,00013 (0,00166)
Comprimento mínimo[6]						
L350	0,12246 (0,02685)	0,06901 (0,01104)	0,01104 (0,01205)	0,02049 (0,01253)	0,00283 (0,00583)	0,00013 (0,00166)
L500	0,11006 (0,02639)	0,05661 (0,01043)	0,03000 (0,01205)	0,02049 (0,01253)	0,00283 (0,00583)	0,00013 (0,00166)
Densidade mínima[7]						
D25	0,05136 (0,01777)	0,01679 (0,00548)	0,01772 (0,00860)	0,01437 (0,01023)	0,00235 (0,00530)	0,00013 (0,00166)
D75	0,08447 (0,02479)	0,03079 (0,00812)	0,03020 (0,01216)	0,02052 (0,01254)	0,00283 (0,00583)	0,00013 (0,00166)
D100	0,08520 (0,02498)	0,03134 (0,00829)	0,03038 (0,01221)	0,02052 (0,01254)	0,00283 (0,00583)	0,00013 (0,00166)

Vide nota Tabela 1

Tabela 4- Coeficientes de endogamia médio geral (desvio padrão) baseado nos segmentos em homozigose (F_{ROH}) para os diferentes cenários na população B (POP_B) e diferentes classes de tamanho de ROH ($F_{ROH1-2Mb}$, $F_{ROH2-4Mb}$, $F_{ROH4-8Mb}$, $F_{ROH8-16Mb}$ e $F_{ROH>16Mb}$).

	F_{ROH}	$F_{ROH\ 1–2Mb}$	$F_{ROH\ 2-4Mb}$	$F_{ROH\ 4-8Mb}$	$F_{ROH\ 8-16Mb}$	$F_{ROH\ >16Mb}$
Controle[1]	0,09107 (0,02687)	0,02947 (0,00824)	0,03508 (0,01241)	0,02276 (0,01367)	0,00354 (0,00698)	0,00022 (0,00224)
Tamanho de janela						
J10	0,09559 (0,02757)	0,03132 (0,00835)	0,03643 (0,01276)	0,02384 (0,01397)	0,00379 (0,00717)	0,00022 (0,00225)
J20	0,08287 (0,02598)	0,02138 (0,00681)	0,03500 (0,01237)	0,02274 (0,01367)	0,00353 (0,00698)	0,00022 (0,00224)
J50	0,07037 (0,02352)	0,01780 (0,00604)	0,02925 (0,01107)	0,02000 (0,01292)	0,00313 (0,00656)	0,00018 (0,00208)
Genótipos ausentes						
M2	0,09125 (0,02689)	0,02944 (0,00821)	0,03511 (0,01237)	0,02287 (0,01366)	0,00361 (0,00709)	0,00022 (0,00224)
M3	0,09126 (0,02687)	0,02943 (0,00820)	0,03512 (0,01235)	0,02288 (0,01366)	0,00362 (0,00709)	0,00022 (0,00224)
M5	0,09127 (0,02687)	0,02943 (0,00820)	0,03513 (0,01235)	0,02288 (0,01366)	0,00362 (0,00709)	0,00022 (0,00224)
Genótipos Heterozigotos						
H1	0,12715 (0,03066)	0,04389 (0,00948)	0,04181 (0,01268)	0,03306 (0,01585)	0,00748 (0,01039)	0,00091 (0,00474)
H2	0,1971 (0,034897)	0,07991 (0,01222)	0,06179 (0,01512)	0,04256 (0,01783)	0,01146 (0,01249)	0,00141 (0,00580)
H3	0,31010 (0,03798)	0,11793 (0,01428)	0,10596 (0,01877)	0,06491 (0,02159)	0,01890 (0,01627)	0,00239 (0,00760)
H5	0,59890 (0,03520)	0,13347 (0,01539)	0,19090 (0,02219)	0,18053 (0,03008)	0,08276 (0,03250)	0,01124 (0,01648)
Gap máximo						
G50	0,00833 (0,00370)	0,00783 (0,00402)	0,00043 (0,00100)	0,00000 (0,00000)	0,00000 (0,00000)	0,00000 (0,00000)
G125	0,07348 (0,02017)	0,03026 (0,00889)	0,03076 (0,01158)	0,01136 (0,00909)	0,00109 (0,00363)	0,00000 (0,00000)
G370	0,08993 (0,02410)	0,02939 (0,00823)	0,03513 (0,01248)	0,02189 (0,01329)	0,00335 (0,00687)	0,00016 (0,00182)
G1000	0,09117 (0,02687)	0,00013 (0,00821)	0,03503 (0,01242)	0,02284 (0,01365)	0,00366 (0,00709)	0,00022 (0,00225)
Comprimento mínimo						
L350	0,12551 (0,02684)	0,06392 (0,01087)	0,03508 (0,01241)	0,02276 (0,01367)	0,00354 (0,00698)	0,00022 (0,00224)
L500	0,11574 (0,02856)	0,05414 (0,01038)	0,03508 (0,01241)	0,02276 (0,01367)	0,00354 (0,00698)	0,00022 (0,00224)
Densidade mínima						
D25	0,05658 (0,01872)	0,01717 (0,00604)	0,02149 (0,00883)	0,01505 (0,01044)	0,00269 (0,00586)	0,00018 (0,03401)
D75	0,09313 (0,02479)	0,03128 (0,00853)	0,03532 (0,01248)	0,02277 (0,01367)	0,00354 (0,00698)	0,00022 (0,00224)
D100	0,09400 (0,02762)	0,03193 (0,00878)	0,03554 (0,01258)	0,02277 (0,01367)	0,00354 (0,00698)	0,00022 (0,00224)

Vide nota Tabela 1

0

2.4 DISCUSSÃO

De um modo geral, ao comparar os diferentes cenários utilizados no estudo, foi possível notar que os resultados de ROH e estimativas de F_{ROH} são sensíveis aos critérios utilizados na metodologia de janelas deslizantes. Todos os critérios afetaram os resultados de ROH e F_{ROH} com exceção do critério número máximo de genótipos ausentes permitidos. Observamos também que a densidade de SNPs do painel de genotipagem deve ser levada em consideração antes de determinar o conjunto de critérios adequados para determinar ROH. Foi possível identificar o cenário que apresenta os níveis de autozigosidade mais plausíveis para os dados analisados.

Os resultados encontrados para o critério tamanho de janela (J10, J15, J20 e J50) apontam que ao passo que se aumenta o tamanho de janela ocorre uma diminuição no *n*ROH identificadas bem como a redução nas estimativas de F_{ROH}. O mesmo resultado foi encontrado por Chinchilla-Vargas et al. (2021) ao testar tamanho de janela de 5, 10 e 20 SNPs em uma espécie de peixe nativa da América do Norte. De acordo com Meyermans et al. (2020), esse resultado ocorre devido a um maior rigor aplicado na detecção dos segmentos homozigotos. Em razão da diferença observada nas estimativas, não recomendamos determinar os limites desse critério arbitrariamente. Uma alternativa para determinar o tamanho de janela seria utilizando a fórmula proposta por Lencz et al. (2007) e adaptada por Purfield et al. (2012) que calcula o número mínimo de SNPs homozigotos consecutivos (L). Vale ressaltar que ao determinar o tamanho de janela, deve-se evitar que esse critério não seja maior que o número mínimo de SNPs uma vez que pode impedir a detecção de ROH de tamanho menor ou igual ao determinado para o número mínimo de SNPs (Curik et al., 2014).

Nos cenários H1, H2, H3 e H5 notou-se que o critério número de máximo de genótipos heterozigotos permitidos na janela possui uma grande influência na detecção de ROH e nas estimativas de F_{ROH}, tendendo a superestimar principalmente $ROH_{8\text{-}16Mb}$ e $ROH_{>16Mb}$ quando comparado com o cenário controle (H0). As diferenças de F_{ROH} entre o cenário controle e os demais cenários para esse parâmetro variou de 0,04032 a 0,55087 para a POP_A e de 0,03608 a

0,50783 para POP_B, apresentando uma alta variação. A determinação do número de genótipos heterozigotos permitidos em uma ROH está relacionada com a frequência de erros de genotipagem de SNPs e com a densidade dos painéis de SNPs. Os erros de genotipagem podem determinar que um SNP é heterozigoto quando na realidade é homozigoto, sendo assim, permite-se que genótipos heterozigotos estejam presentes dentro de uma ROH para corrigir esses erros. Mastrangelo et al. (2016) avaliaram o efeito de um, dois ou três genótipos heterozigotos permitidos em análises de F_{ROH} de três raças italianas de bovinos leiteiros com painel de média densidade (54609 SNPs). Os autores encontraram uma variação de 0,003 a 0,007 no F_{ROH} e concluíram que, para os respectivos dados, o mais adequado seria apenas um genótipo heterozigoto permitido. Ferencakovic et al. (2013) em bovinos das raças Pardo Suíço, Pinzgauer e Tyrol Grey sugeriram um ou mais SNPs heterozigotos dentro de uma ROH, principalmente em painéis de alta densidade (777972 SNPs). Os mesmos autores recomendaram, utilizar gráficos para determinar o número máximo de genótipos heterozigotos que devem ser permitidos, e, portanto, evitar segmentos falso-positivos originados de erros de genotipagem.

No geral, as falhas de genotipagem ocorrem com uma frequência de 0,2% a 1% (Rabbee and Speed, 2006). Admitir um número elevado de genótipos heterozigotos nas análises de ROH pode superestimar a identificação de ROH longas ($ROH_{>8Mb}$), uma vez que pode ocorrer a fusão de ROH adjacentes. Deste modo, deve-se ter cautela ao permitir a ocorrência de genótipos heterozigotos, principalmente em painéis de aproximadamente 700 mil marcadores pois nesses painéis os erros de genotipagem ocorrem em maior abundância. Howrigan et al. (2011) e Marras et al. (2015) utilizando dados de média densidade (33040 e 54000 SNPs, respectivamente) recomendaram não permitir a existência de SNPs heterozigotos para classificar um segmento como uma ROH uma vez que o nível de autozigosidade dos indivíduos foi praticamente inalterada pois as ROH longas se dividiram em duas mais curtas. No presente trabalho, considerando que os dados são provenientes de painéis de média densidade (50K) e houve uma grande variação entre cenários, sugere-se que o cenário controle que considera nenhum genótipo heterozigoto seja o mais adequado para evitar superestimação de ROH longos.

Outros critérios que devem ser determinados de acordo com a densidade do painel de SNPs é o GAP entre SNPs adjacentes e a densidade mínima (SNP/Kb). Em painéis de baixa a média densidade (3K a 50K), o espaçamento entre SNPs é grande em relação aos de alta densidade (700K), deste modo o GAP máximo deve ser determinado com cautela. Entretanto, na literatura poucos estudos discutem a respeito do valor que deve ser adotado para o GAP máximo, existindo uma grande variação de limites impostos (Peripolli et al., 2017). Com relação a densidade mínima definida para as análises de ROH, a maioria dos softwares utilizam uma configuração padrão de densidade mínima de 1/50 SNP/Kb, valor superior ou próximo a densidade dos painéis de genotipagem de baixa a média densidade (7K a 50K), podendo assim afetar a cobertura do genoma nas análises de ROH (Meyermans, et al., 2020). Para as duas populações estudadas, identificamos que ao utilizar GAPs de 50 Kb e 125 Kb, a detecção de ROH e suas respectivas estimativas de F_{ROH} tendem a ser subestimadas quando comparadas ao cenário controle (G500). A configuração da densidade mínima (SNP/Kb) afetou os resultados da análise de ROH no cenário D25 em relação aos demais cenários, subestimando os resultados de F_{ROH}. Meyermans et al. (2020) apontaram que, quando adotado um GAP inferior a 500 Kb, a cobertura do genoma da análise ROH diminui, acarretando assim em uma subestimação dos parâmetros. O mesmo estudo, identificou um aumento nas estimativas de F_{ROH} quando utilizadas densidades entre 1/40 e 1/60 SNP/Kb. Deste modo, o GAP máximo de 500 kb e uma densidade de 1/50 SNP/Kb utilizado no cenário controle parece ser o mais recomendado para os dados de média densidade desse estudo. Vale ressaltar que essa densidade é superior a densidade do painel de SNPs (1/25 SNP/Kb) utilizado para a genotipagem das duas populações.

O critério comprimento mínimo tem sido utilizado para corrigir segmentos curtos de ROH originários de um alto LD (Purfield et al., 2012; Ferenčaković et al., 2013). De acordo com Ai et al. (2013), a magnitude do LD está relacionada com a história populacional e do tamanho efetivo da população. O cruzamento recente de duas raças, por exemplo, pode provocar o aumento de LD e, consequentemente, reduzir o comprimento de segmentos homozigotos no genoma. Além disso, Herrero-Medrano et al. (2013) relataram uma correlação

positiva entre os valores médios de LD e comprimento de ROH por cromossomo em um estudo de populações de suínos domésticos e selvagens da Península Ibérica com um painel de 62.163 SNPs. Sendo assim, o principal motivo para se realizar o controle de qualidade para alto LD seria excluir segmentos curtos determinados como IBD, mas que na realidade foram originados de um alto LD. Entretanto, realizar um controle de qualidade para alto LD pode reduzir a densidade de SNP a ser analisados, provocando uma análise de ROH tendenciosa (Meyermans et al., 2020).

No presente trabalho, o LD foi analisado nas duas populações para verificar a necessidade de um maior rigor no limite determinado para comprimento mínimo de ROH. Os r^2 médios entre SNPs adjacentes estimado no estudo não são elevados de acordo com a literatura (Brito et al., 2015; Conte et al., 2019; Joshi et al., 2018; Khatkar et al., 2008; Yoshida et al., 2019). Além disso, observou-se que quando adotamos um comprimento mínimo com menor rigor (L350 e L500), ocorre uma diminuição no número de segmentos curtos bem como a subestimação da F_{ROH} em comparação ao Controle (L1000). Ferenčaković et al. (2013) sugeriu um comprimento mínimo de 1000 Kb e concluiu que este critério deve ser ajustado à densidade de SNPs do painel. Purfield et al. (2012) recomendaram um comprimento mínimo de ROH de 500 Kb para excluir ROH muito curtas provenientes de um alto LD. Devido a variação dos resultados e a densidade do painel de SNPs utilizados no estudo, acredita-se que seja necessário um maior rigor do que o aplicado nos cenários L350 e L500, adotando um comprimento mínimo de 1000 Kb, como realizado no cenário controle para evitar assim, a superestimação de ROH oriundos de painéis de baixa densidade.

A tilápia do Nilo (*Oreochromis niloticus*) está entre as espécies mais produzidas a nível mundial (FAO, 2020). Em vista disso, desde 1988 esforços têm sido realizados para promover o melhoramento genético dessa espécie (Bentsen et al., 1998; Eknath et al., 1993). Entretanto, os programas de melhoramento genético de tilápia podem apresentar altas taxas de endogamia devido principalmente à restrição do número de famílias disponíveis para

reprodução, à forte pressão de seleção e ao acasalamento de animais aparentados (PETERSEN et al., 2012).

As estimativas de endogamia genômica encontradas no cenário controle são consideradas moderadas (0,08233 para POP_A e 0,09107 para POP_B) e estão próximas do limite máximo aceitável. Resultado semelhante foi encontrado por Ambrosio et al. (2019) em seis linhagens de truta arco-íris (*Oncorhynchus mykiss*) usando um painel de SNPs de 20K, estimando um F_{ROH} de moderado a alto entre as populações (0,086 a 0,195). Chinchilla-Vargas et al. (2021) utilizando a sequência completa do genoma de Muskellunge, espécie nativa da América do Norte (*Esox masquinongy*), estimaram um F_{ROH} de 0,03 para uma população da Lowa State University e 0,32 para uma população do rio Saint Lawrence localizado no Canadá. Segundo os autores, essas diferenças são decorrentes do isolamento de subpopulações do rio Saint Lawrence. Wang et al. (2022), utilizando dados de 39.018 SNPs de uma população comum e outra monosexo de Falso-Alabote-Japonês (*Paralichthys olivaceus*), uma espécie de linguado nativo do noroeste do Oceano Pacífico, encontraram um baixo nível de endogamia genômica na população (0,2171) em comparação com a endogamia genômica da população monosexo (0,9227). As duas populações estudadas passaram por oito gerações de seleção, sendo aceitáveis incrementos de endogamia menores que 1% por geração para evitar que ocorram efeitos deletérios expressivos na população (Ponzoni et al., 2010). Uma possível solução para diminuir a endogamia nessas populações seria o acasalamento de indivíduos de linhagens diferentes, inclusive entre indivíduos da POP_A e da POP_B.

O comprimento de F_{ROH} auxilia a inferir o número de gerações do evento de endogamia, entretanto painéis de média densidade são apropriados para identificar segmentos maiores que 5 Mb (Ferenčaković et al., 2013; Purfield et al., 2012). Sendo assim, as estimativas mais confiáveis de F_{ROH} neste estudo são para segmentos maiores que 5 Mb. Observamos que as estimativas de $F_{ROH8\text{-}16Mb}$ e $F_{ROH>16Mb}$ no cenário controle foram de baixa magnitude nas populações estudadas sugerindo que não houve eventos de endogamia recentes relevantes. Esses resultados podem indicar que a variabilidade

genética vem sendo explorada no melhoramento das populações, evitando o acasalamento entre indivíduos aparentados e utilizando grande número de famílias para controlar o aumento da endogamia. No entanto, a limitação do número de genótipos heterozigotos permitidos nesse cenário pode estar diminuindo a identificação de ROH longos, consequentemente acarretar na subestimação de $F_{ROH8\text{-}16Mb}$ e $F_{ROH>16Mb}$.

A diferença observada entre F_{PED} e F_{ROH} estimado no cenário controle está de acordo com o observado na literatura, sugerindo que o F_{PED} subestima o valor verdadeiro da endogamia. O mesmo resultado foi encontrado por Yoshida et al. (2020) usando um painel de SNPs de 200K em três linhagens de salmão Coho (*Oncorhynchus kisutch*). Os autores também encontraram baixos valores de correlação entre coeficientes de endogamia genômicos e F_{PED}. De acordo com Curik et al. (2014), isso ocorre porque a informação de pedigree possui limitações como não quantificar a proporção exata do genoma que é idêntica por descendência (IBD), além de não levar em consideração as variações decorrentes dos processos de meiose, e de assumir que os animais fundadores são não aparentados e não endogâmicos. Vale ressaltar também que a correlação entre F_{PED} e F_{ROH} tende a aumentar em endogamias calculadas em segmentos longos ($F_{ROH8\text{-}16Mb}$ e $F_{ROH>16Mb}$), uma vez que as informações de pedigree não podem se estender por muitas gerações. No entanto, também sabemos que a variação da densidade dos painéis de genotipagem pode afetar fortemente a qualidade de estimação de F_{ROH} dado que painéis com menores densidades de SNPs tendem a superestimar o F_{ROH} (Ferenčaković et al., 2013). Sendo assim, estudos utilizando uma maior densidade de SNPs poderiam ajudar a estimar a F_{ROH} de maneira mais precisa.

2.5 CONCLUSÃO

Os resultados apresentados neste trabalho apontam que os critérios utilizados para determinar ROH podem afetar fortemente as análises de ROH em genótipos de média densidade de duas populações de tilápias do Nilo. Com exceção do critério número máximo de genótipos ausentes, todos os critérios

tenderam a afetar tanto o número de ROH quanto as estimativas de F_{ROH}. O cenário controle, além de permitir a comparação entre os diversos cenários, também apresentou o conjunto de valores dos critérios que fornecem estimativas de F_{ROH} mais plausíveis. Entretanto, ressaltamos que isso não significa que as ROH observadas no cenário controle não apresentam resultados viesados. Por fim, recomendamos que o efeito dos critérios utilizados nas análises de ROH sejam explorados para que se obtenha estimativas de F_{ROH} confiáveis e com baixo viés.

2.6 REFERÊNCIAS BIBLIOGRÁFICAS

Ai, H., Huang, L., Ren, J., 2013. Genetic Diversity, Linkage Disequilibrium and Selection Signatures in Chinese and Western Pigs Revealed by Genome-Wide SNP Markers. PLoS ONE 8. https://doi.org/10.1371/journal.pone.0056001

Aramburu, O., Ceballos, F., Casanova, A., le Moan, A., Hemmer-Hansen, J., Bekkevold, D., Bouza, C., Martínez, P., 2020. Genomic Signatures After Five Generations of Intensive Selective Breeding: Runs of Homozygosity and Genetic Diversity in Representative Domestic and Wild Populations of Turbot (Scophthalmus maximus). Frontiers in Genetics 11. https://doi.org/10.3389/fgene.2020.00296

Bentsen, H.B., Eknath, A.E., Palada-De Vera, M.S., Danting, J.C., Bolivar, H.L., Reyes, R.A., Dionisio, E.E., Longalong, F.M., Circa, A. v, Tayamen, M.M., Gjerde, B., Philippines˜d, P., 1998. Genetic improvement of farmed tilapias: growth performance in a complete diallel cross experiment with eight strains of Oreochromis niloticus 1. Aquaculture 160, 145–173.

Biscarini, F., Cozzi, P., Gaspa, G., Marras, G., 2019. Detect Runs of Homozygosity and Runs of Heterozygosity in Diploid Genomes.

Brito, L.F., Jafarikia, M., Grossi, D.A., Kijas, J.W., Porto-Neto, L.R., Ventura, R. v., Salgorzaei, M., Schenkel, F.S., 2015. Characterization of linkage disequilibrium, consistency of gametic phase and admixture in Australian and Canadian goats. BMC Genetics 16. https://doi.org/10.1186/s12863-015-0220-1

Broman, K.W., Weber, J.L., 1998. Estimation of Pairwise Relationships in the Presence of Genotyping Errors. The American Journal of Human Genetics 63, 1563. https://doi.org/10.1086/302112

Cádiz, M.I., López, M.E., Díaz-Domínguez, D., Cáceres, G., Yoshida, G.M., Gomez-Uchida, D., Yáñez, J.M., 2020. Whole genome re-sequencing reveals recent signatures of selection in three strains of farmed Nile tilapia (Oreochromis niloticus). Scientific Reports 10. https://doi.org/10.1038/s41598-020-68064-5

Chinchilla-Vargas, J., Meerbeek, J.R., Rothschild, M.F., Bertolini, F., 2021. Signatures of selection and genomic diversity of Muskellunge (Esox masquinongy) from two populations in North America. Research Square 1–29. https://doi.org/10.21203/rs.3.rs-264701/v3

Conte, M.A., Joshi, R., Moore, E.C., Nandamuri, S.P., Gammerdinger, W.J., Roberts, R.B., Carleton, K.L., Lien, S., Kocher, T.D., 2019. Chromosome-scale assemblies reveal the structural evolution of African cichlid genomes. Gigascience 8. https://doi.org/10.1093/gigascience/giz030

Curik, I., Ferenčaković, M., Sölkner, J., 2014. Inbreeding and runs of homozygosity: A possible solution to an old problem. Livestock Science 166, 26–34. https://doi.org/10.1016/j.livsci.2014.05.034

D'Ambrosio, J., Phocas, F., Haffray, P., Bestin, A., Brard-Fudulea, S., Poncet, C., Quillet, E., Dechamp, N., Fraslin, C., Charles, M., Dupont-Nivet, M., 2019. Genome-wide estimates of genetic diversity, inbreeding and effective size of experimental and commercial rainbow trout lines undergoing selective breeding. Genetics Selection Evolution 51. https://doi.org/10.1186/s12711-019-0468-4

Eknath, A.E., Tayamen, M.M., Palada-de Vera, M.S., Danting, J.C., Reyes, R.A., Dionisio, E.E., Capili, J.B., Bolivar, H.L., Abella, T.A., Circa, A. v., Bentsen, H.B., Gjerde, B., Gjedrem, T., Pullin, R.S.V., 1993. Genetic improvement of farmed tilapias: the growth performance of eight strains of Oreochromis niloticus tested in different farm environments. Genetics in Aquaculture 171–188. https://doi.org/10.1016/b978-0-444-81527-9.50021-x

Falconer, D., Mackay, T., 1996. Introduction to quantitative genetics, 4th ed. Logman, New York.

FAO, 2020. El estado mundial de la pesca y la acuicultura 2020, El estado mundial de la pesca y la acuicultura 2020. FAO. https://doi.org/10.4060/ca9229es

Ferenčaković, M., Sölkner, J., Curik, I., 2013. Estimating autozygosity from high-throughput information: Effects of SNP density and genotyping errors. Genetics Selection Evolution 45. https://doi.org/10.1186/1297-9686-45-42

Gibson, J., Morton, N.E., Collins, A., 2006. Extended tracts of homozygosity in outbred human populations. Human Molecular Genetics 15, 789–795. https://doi.org/10.1093/hmg/ddi493

Gurgul, A., Szmatoła, T., Topolski, P., Jasielczuk, I., Żukowski, K., Bugno-Poniewierska, M., 2016. The use of runs of homozygosity for estimation of recent inbreeding in Holstein cattle. Journal of Applied Genetics 57, 527–530. https://doi.org/10.1007/s13353-016-0337-6

Herrero-Medrano, J.M., Megens, H.J., Groenen, M.A.M., Ramis, G., Bosse, M., Pérez-Enciso, M., Crooijmans, R.P.M.A., 2013. Conservation genomic analysis of domestic and wild pig populations from the Iberian Peninsula. BMC Genetics 14. https://doi.org/10.1186/1471-2156-14-106

Hillestad, B., Woolliams, J.A., Boison, S.A., Grove, H., Meuwissen, T., Våge, D.I., Klemetsdal, G., 2017. Detection of runs of homozygosity in Norwegian Red: Density, criteria and genotyping quality control. Acta Agriculturae Scandinavica A: Animal Sciences 67, 107–116. https://doi.org/10.1080/09064702.2018.1501088

Howrigan, D.P., Simonson, M.A., Keller, M.C., 2011. Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. BMC Genomics 12. https://doi.org/10.1186/1471-2164-12-460

Joshi, R., Árnyasi, M., Lien, S., Gjøen, H.M., Alvarez, A.T., Kent, M., 2018. Development and validation of 58K SNP-array and high-density linkage map in Nile tilapia (O. niloticus). Frontiers in Genetics 9. https://doi.org/10.3389/fgene.2018.00472

Khatkar, M.S., Nicholas, F.W., Collins, A.R., Zenger, K.R., Cavanagh, J.A.L., Barris, W., Schnabel, R.D., Taylor, J.F., Raadsma, H.W., 2008. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. BMC Genomics 9. https://doi.org/10.1186/1471-2164-9-187

Kirin, M., McQuillan, R., Franklin, C.S., Campbell, H., Mckeigue, P.M., Wilson, J.F., 2010. Genomic runs of homozygosity record population history and consanguinity. PLoS ONE 5. https://doi.org/10.1371/journal.pone.0013996

Ku, C.S., Naidoo, N., Teo, S.M., Pawitan, Y., 2011. Regions of homozygosity and their impact on complex diseases and traits. Human Genetics. https://doi.org/10.1007/s00439-010-0920-6

Lencz, T., Lambert, C., DeRosse, P., Burdick, K.E., Vance Morgan, T., Kane, J.M., Kucherlapati, R., Malhotra, A.K., 2007. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia.

Marras, G., Gaspa, G., Sorbolini, S., Dimauro, C., Ajmone-Marsan, P., Valentini, A., Williams, J.L., MacCiotta, N.P.P., 2015. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. Animal Genetics 46, 110–121. https://doi.org/10.1111/age.12259

Mastrangelo, S., Tolone, M., di Gerlando, R., Fontanesi, L., Sardina, M.T., Portolano, B., 2016a. Genomic inbreeding estimation in small populations: Evaluation of runs of homozygosity in three local dairy cattle breeds. Animal 10, 746–754. https://doi.org/10.1017/S1751731115002943

McQuillan, R., Leutenegger, A.L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., MacLeod, A.K., Farrington, S.M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S.H., Dunlop, M.G., Wright, A.F., Campbell, H., Wilson, J.F., 2008. Runs of Homozygosity in European Populations. American Journal of Human Genetics 83, 359–372. https://doi.org/10.1016/j.ajhg.2008.08.007

Meyermans, R., Gorssen, W., Buys, N., Janssens, S., 2020. How to study runs of homozygosity using plink? a guide for analyzing medium density snp data in livestock and pet species. BMC Genomics 21. https://doi.org/10.1186/s12864-020-6463-x

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., Lee, D.H., 2002. BLUPF90 AND RELATED PROGRAMS (BGF90), in: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production.

Neira, R., García, X., Lhorente, J.P., Filp, M., Yáñez, J.M., Cascante, A.M., 2016. Evaluation of the growth and carcass quality of diallel crosses of four strains of Nile tilapia (Oerochromis niloticus). Aquaculture 451, 213–222. https://doi.org/10.1016/j.aquaculture.2015.08.033

Paul, K., D'Ambrosio, J., Phocas, F., 2022. Temporal and region-specific variations in genome-wide inbreeding effects on female size and reproduction traits of rainbow trout. Evolutionary Applications 15, 645–662. https://doi.org/10.1111/eva.13308

Peripolli, E., Munari, D.P., Silva, M.V.G.B., Lima, A.L.F., Irgang, R., Baldi, F., 2017. Runs of homozygosity: current knowledge and applications in livestock. Animal Genetics. https://doi.org/10.1111/age.12526

PETERSEN, R.L., GARCIA, J.E., Mello, G., LIEDKE, A.M.R., SINCERO, T.C.M., GRISARD, E.C., 2012. ANÁLISE DA DIVERSIDADE GENÉTICA DE TILÁPIAS CULTIVADAS NO ESTADO DE SANTA CATARINA (BRASIL) UTILIZANDO MARCADORES MICROSSATÉLITES. Boletim do Instituto de Pesca 38, 313–321.

Ponzoni, R.W., Khaw, H.L., Nguyen, N.H., Hamzah, A., 2010. Inbreeding and effective population size in the Malaysian nucleus of the GIFT strain of Nile tilapia (Oreochromis niloticus). Aquaculture 302, 42–48. https://doi.org/10.1016/j.aquaculture.2010.02.009

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

Purfield, D.C., Berry, D.P., McParland, S., Bradley, D.G., 2012. Runs of homozygosity and population history in cattle. BMC Genetics 13. https://doi.org/10.1186/1471-2156-13-70

Rabbee, N., Speed, T.P., 2006. A genotype calling algorithm for affymetrix SNP arrays. Bioinformatics 22, 7–12. https://doi.org/10.1093/bioinformatics/bti741

Rodrigues Caetano, A., 2009. Revista Brasileira de Zootecnia Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro.

Silva, J.M. da, Giachetto, P.F., Campos, L.O.D.S., Cintra, L.C., Paiva, S.R., Caetano, A.R., Yamagishi, M.E.B., 2015. Genomic variants revealed by invariably missing genotypes in Nelore cattle. PLoS ONE 10. https://doi.org/10.1371/journal.pone.0136035

Stachowicz, k.; Sargolzaei, M.; Miglior, F.; Schenkel, F. S, 2011. Rates of inbreeding and genetic diversity in Canadian Holstein and Jersey cattle. Journal of Dairy Science, 94, 10, 5160–5175.

Wang, L., Wu, Z., Zou, C., Lu, Y., Yue, X., Song, Z., Yang, R., You, F., 2022. Genetic diversity and signatures of selection in the mito-gynogenetic olive flounder Paralichthys olivaceus revealed by genome-wide SNP markers. Aquaculture 553. https://doi.org/10.1016/j.aquaculture.2022.738062

Wright, S., 1922. COEFFICIENTS OF INBREEDING AND RELATIONSHIP. The American Naturalist 26, 330–338.

Yáñez, J.M., Yoshida, G., Barria, A., Palma-Véjares, R., Travisany, D., Díaz, D., Cáceres, G., Cádiz, M.I., López, M.E., Lhorente, J.P., Jedlicki, A., Soto, J., Salas, D., Maass, A., 2020. High-Throughput Single Nucleotide Polymorphism (SNP) Discovery and Validation Through Whole-Genome Resequencing in Nile Tilapia (Oreochromis niloticus). Marine Biotechnology 22, 109–117. https://doi.org/10.1007/s10126-019-09935-5

Yoshida, G.M., Barria, A., Correa, K., Cáceres, G., Jedlicki, A., Cadiz, M.I., Lhorente, J.P., Yáñez, J.M., 2019. Genome-Wide Patterns of Population Structure and Linkage Disequilibrium in Farmed Nile Tilapia (Oreochromis niloticus). Frontiers in Genetics 10. https://doi.org/10.3389/fgene.2019.00745

Yoshida, G.M., Cáceres, P., Marín-Nahuelpi, R., Koop, B.F., Yáñez, J.M., 2020. Estimates of autozygosity through runs of homozygosity in farmed coho salmon. Genes (Basel) 11. https://doi.org/10.3390/genes11050490

Zavarez, L. B.; Utsunomiya, Y. T.; Carmo, A. S.; Et Al. Assessment of autozygosity in Nellore cows (Bos indicus) through high-density SNP genotypes.2015. Frontiers in Genetics, 6.

Zhang, Q., Calus, M.P.L., Guldbrandtsen, B., Lund, M.S., Sahana, G., 2015. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. BMC Genetics 16. https://doi.org/10.1186/s12863-015-0227-7

2.7 FIGURAS SUPLEMENTARES

Tabela Suplementar 1- Parâmetros e critérios utilizados nos diferentes cenários para definir as corridas de homozigose (ROH) em duas populações de tilápia do Nilo

Cenários	Tamanho de janela (SNPs)	Threshold	Número mínimo de SNPs	Genótipos Heterozigotos	Genótipos Ausentes	Gap máximo (Kb)	Comprimento mínimo (Kb)	Densidade mínima (SNP/Kb)
Controle	15	0.135	15	0	1	500	1000	1/50
J10	10	0.135	10	0	1	500	1000	1/50
J20	20	0.135	20	0	1	500	1000	1/50
J50	50	0.135	50	0	1	500	1000	1/50
M2	15	0.135	15	0	2	500	1000	1/50
M3	15	0.135	15	0	3	500	1000	1/50
M5	15	0.135	15	0	5	500	1000	1/50
H1	15	0.135	15	1	1	500	1000	1/50
H2	15	0.135	15	2	1	500	1000	1/50
H3	15	0.135	15	3	1	500	1000	1/50
H5	15	0.135	15	5	1	500	1000	1/50
G50	15	0.135	15	0	1	50	1000	1/50
G125	15	0.135	15	0	1	125	1000	1/50
G370	15	0.135	15	0	1	370	1000	1/50
G1000	15	0.135	15	0	1	1000	1000	1/50
L350	15	0.135	15	0	1	500	350	1/50
L500	15	0.135	15	0	1	500	500	1/50
D25	15	0.135	15	0	1	500	1000	1/25
D75	15	0.135	15	0	1	500	1000	1/75
D100	15	0.135	15	0	1	500	1000	1/100

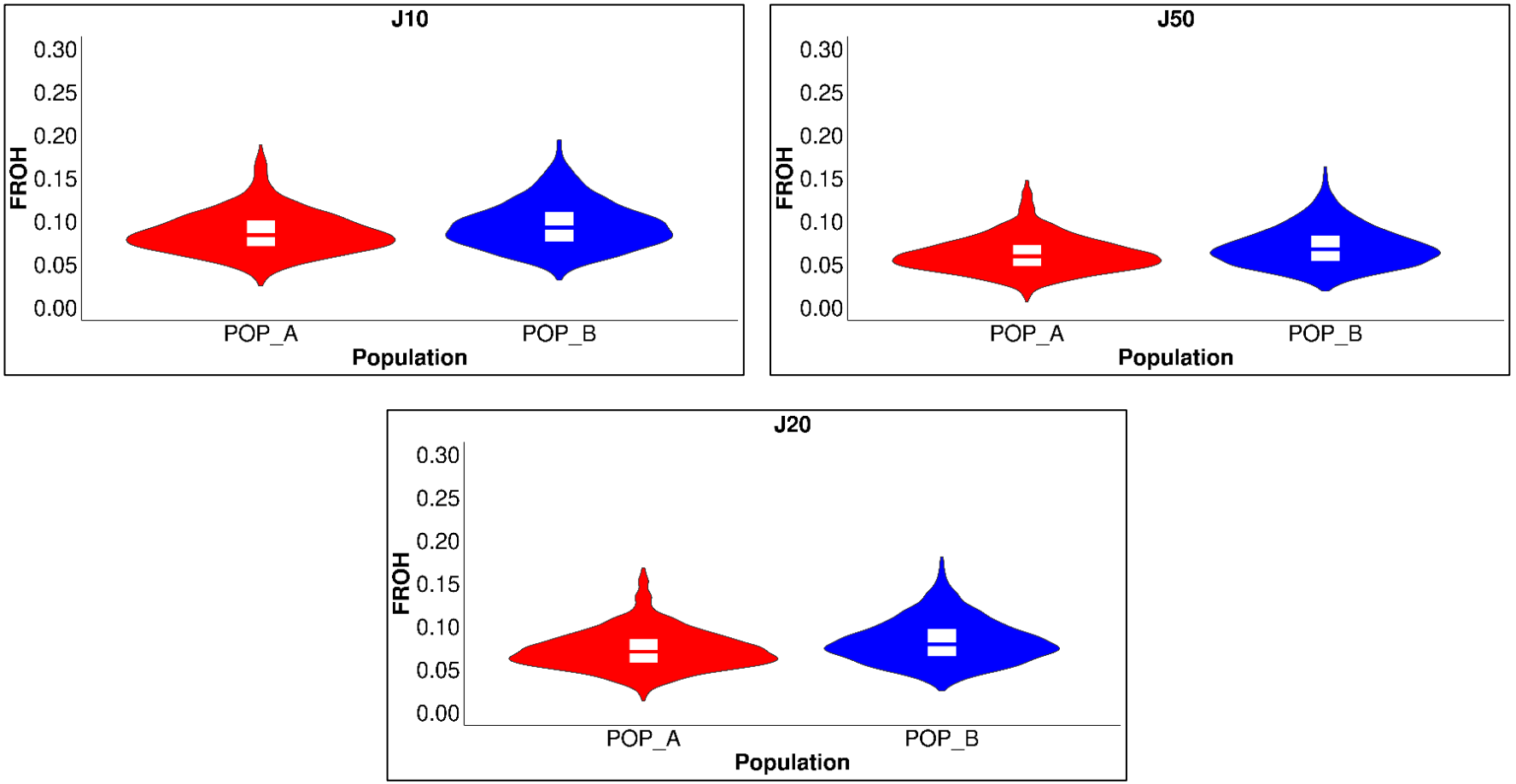


Figura Suplementar 1- Gráfico violino do coeficiente de endogamia baseado em corridas de homozigose (F_{ROH}) dos cenários "Controle" e nos cenários de tamanho de janela de 10, 20 e 50 SNPs (J10, J20 e J50) em duas populações de tilápia do Nilo (POP_A e POP_B).

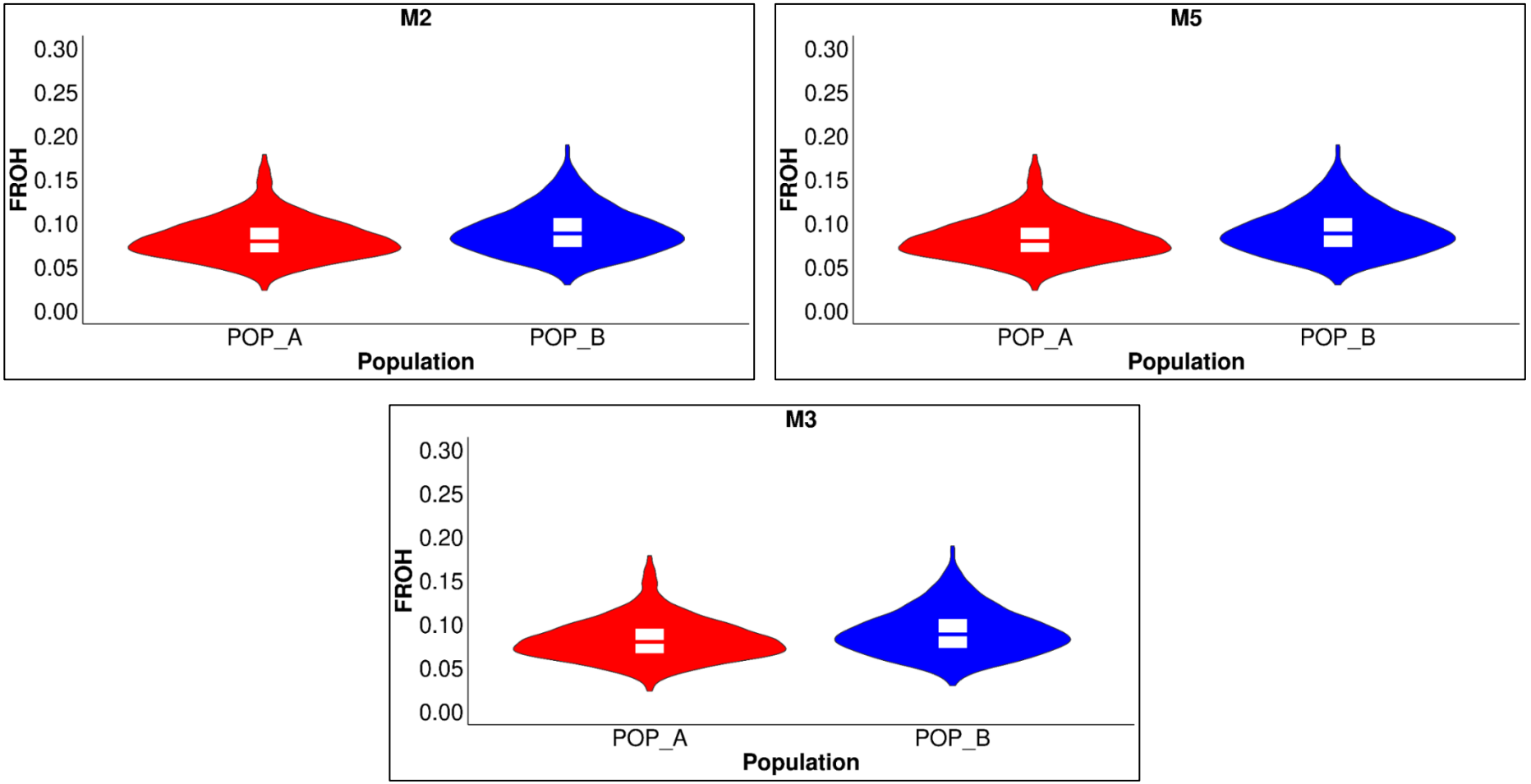


Figura Suplementar 2- Gráfico violino do coeficiente de endogamia baseado em corridas de homozigose (F_{ROH}) dos cenários "Controle" e nos cenários de número de genótipos perdidos na janela 2, 3 e 5 (M2, M3 e M5) em duas populações de tilápia do Nilo (POP_A e POP_B).

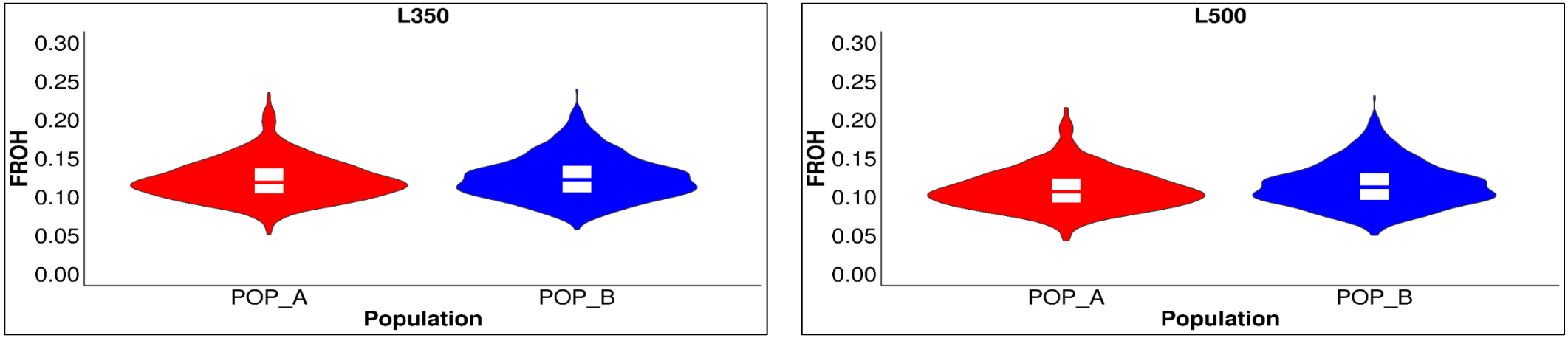


Figura Suplementar 3- Gráfico violino do coeficiente de endogamia baseado em corridas de homozigose (F_{ROH}) dos cenários "Controle" e nos cenários de comprimento mínimo de ROH de 350 Kb e 500 Kb (L350 e L500) em duas populações de tilápia do Nilo (POP_A e POP_B).

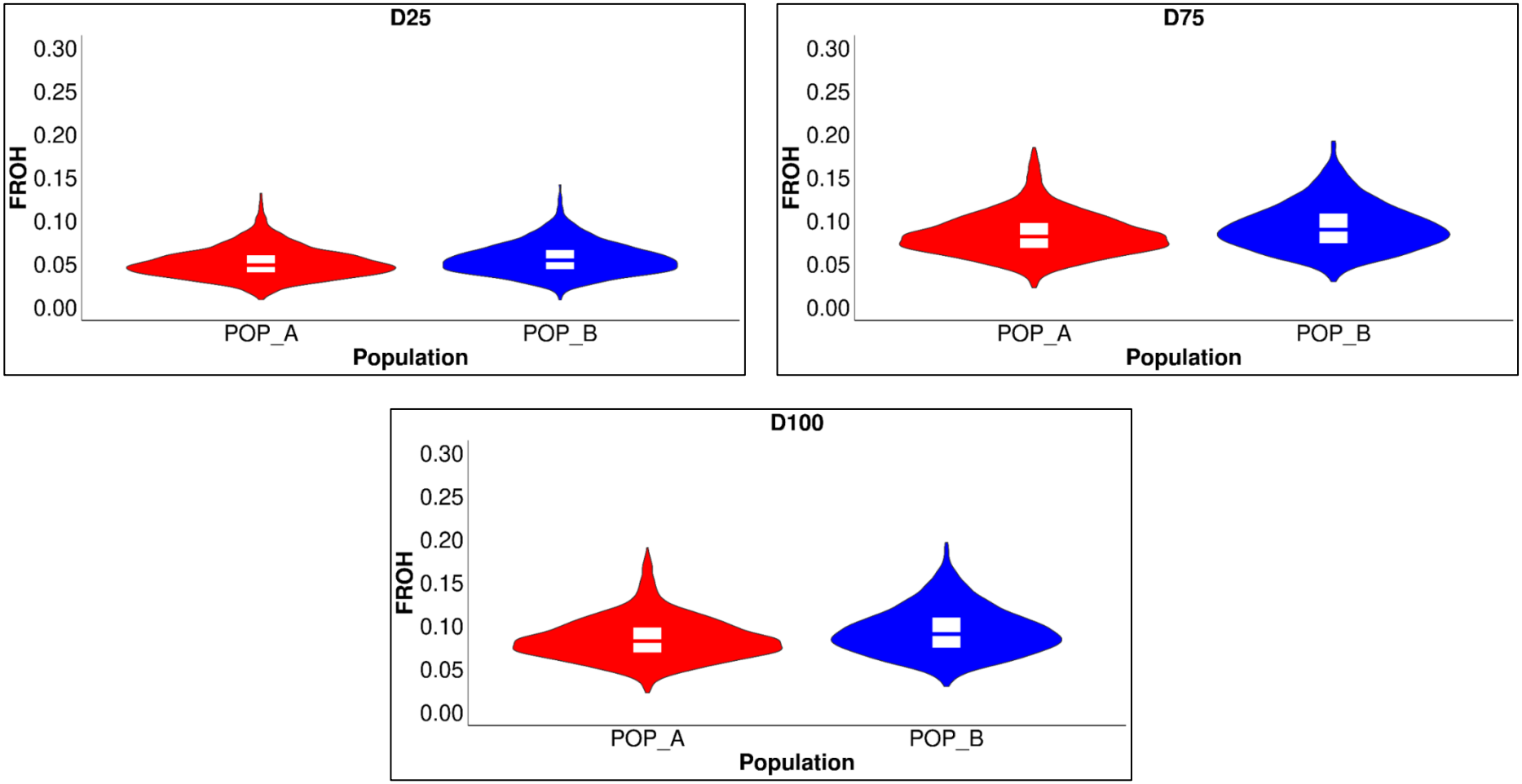


Figura Suplementar 4- Gráfico violino do coeficiente de endogamia baseado em corridas de homozigose (F_{ROH}) dos cenários "Controle" e nos cenários de densidade mínima de ROH de 1/25, 1/75, 1/100 (SNP/Kb) (D25, D75 e D100) em duas populações de tilápia do Nilo (POP_A e POP_B).

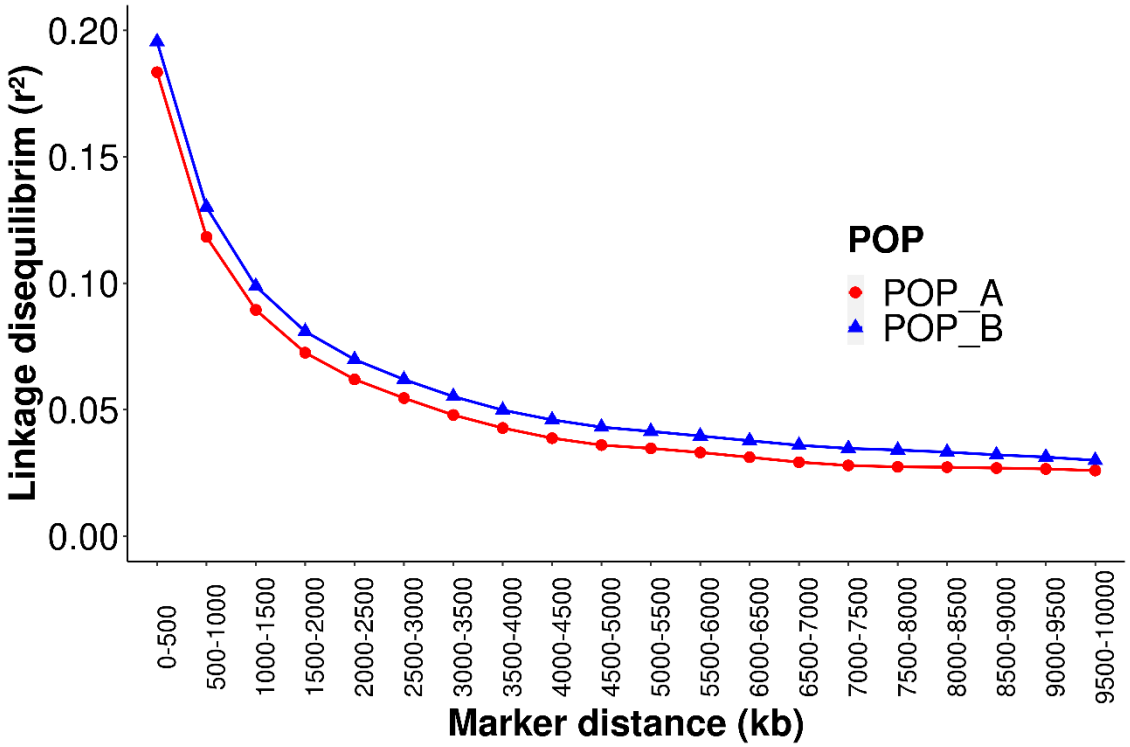


Figura Suplementar 5– Decaimento da média de desequilíbrio de ligação (r^2) em diferentes classes de distância entre marcadores (Kb) para duas populações de tilápia: (POP_A e POP_B).