

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA GOIANO
- CAMPUS RIO VERDE

DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOTECNIA

**TÉCNICAS MULTIVARIADAS ASSISTIDA POR FERRAMENTAS
MOLECULARES NA CLASSIFICAÇÃO DE SERRASALMIDEOS**

Autor: Reyger Rayson Bittar Silva

Orientador: Prof. Dr. Adriano Carvalho Costa

Rio Verde - GO

Maior – 2019

TÉCNICAS MULTIVARIADAS ASSISTIDA POR FERRAMENTAS MOLECULARES NA CLASSIFICAÇÃO DE SERRASALMIDEOS

Autor: Reyger Rayson Bittar Silva

Orientador: Prof. Dr. Adriano Carvalho Costa

Dissertação apresentada como parte das exigências para obtenção do título de MESTRE EM ZOOTECNIA, ao Programade Pós-Graduação em Zootecnia do Instituto Federal de Educação, Ciência e Tecnologia Goiano – Campus Rio Verde – Área de concentração Zootecnia/Recursos Pesqueiros.

Rio Verde - GO

Maio – 2019

Sistema desenvolvido pelo ICMC/USP
Dados Internacionais de Catalogação na Publicação (CIP)
Sistema Integrado de Bibliotecas - Instituto Federal Goiano

Silva, Reyger Rayson Bittar
S586t TÉCNICAS MULTIVARIADAS ASSISTIDA POR FERRAMENTAS
MOLECULARES NA CLASSIFICAÇÃO DE SERRASALMIDEOS /
Reyger Rayson Bittar Silva; orientador Adriano
Carvalho Costa; co-orientador Marco Antônio Pereira
da Silva. -- Rio Verde, 2019.
60 p.

Dissertação (em PROGRAMA DE PÓS-GRADUAÇÃO EM
ZOOTECNIA) -- Instituto Federal Goiano, Campus Rio
Verde, 2019.

1. Colossoma Macropomum. 2. Modelo de mistura. 3.
Peixes nativos. 4. Serrasalmideos. 5. Piaractus
Brachypomum. I. Costa, Adriano Carvalho, orient. II.
Silva, Marco Antônio Pereira da, co-orient. III.
Título.

Responsável: Johnathan Pereira Alves Diniz - Bibliotecário-Documentalista CRB-1 nº2376



Repositório Institucional do IF Goiano - RIIF Goiano
Sistema Integrado de Bibliotecas

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR PRODUÇÕES TÉCNICO-CIENTÍFICAS NO REPOSITÓRIO INSTITUCIONAL DO IF GOIANO

Com base no disposto na Lei Federal nº 9.610/98, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia Goiano, a disponibilizar gratuitamente o documento no Repositório Institucional do IF Goiano (RIIF Goiano), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IF Goiano.

Identificação da Produção Técnico-Científica

- | | |
|--|---|
| <input type="checkbox"/> Tese | <input type="checkbox"/> Artigo Científico |
| <input checked="" type="checkbox"/> Dissertação | <input type="checkbox"/> Capítulo de Livro |
| <input type="checkbox"/> Monografia – Especialização | <input type="checkbox"/> Livro |
| <input type="checkbox"/> TCC - Graduação | <input type="checkbox"/> Trabalho Apresentado em Evento |
| <input type="checkbox"/> Produto Técnico e Educacional - Tipo: _____ | |

Nome Completo do Autor: Reygen Raysson Bittar Silva
 Matrícula: 2017102310210070
 Título do Trabalho: Técnicas multivariadas assistida por Invariantes moleculares na classificação de cereais amidos

Restrições de Acesso ao Documento

Documento confidencial: Não Sim, justifique: _____

Informe a data que poderá ser disponibilizado no RIIF Goiano: 01/01/2020

O documento está sujeito a registro de patente? Sim Não
 O documento pode vir a ser publicado como livro? Sim Não

DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

O/A referido/a autor/a declara que:

- o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;
- obteve autorização de quaisquer materiais incluídos no documento do qual não detém os direitos de autor/a, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia Goiano os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;
- cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia Goiano.

Rio Verde, 20/12/2019
Local Data

Reygen Raysson Bittar Silva
Assinatura do Autor e/ou Detentor dos Direitos Autorais

Ciente e de acordo:

[Assinatura]
Assinatura do(a) orientador(a)

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
GOIANO – CÂMPUS RIO VERDE
DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOTECNIA

**TÉCNICAS MULTIVARIADAS ASSISTIDA POR
FERRAMENTAS MOLECULARES NA CLASSIFICAÇÃO DE
SERRASALMÍDEOS**

Autor: Reyger Rayson Bittar Silva
Orientador: Adriano Carvalho Costa

TITULAÇÃO: Mestre em Zootecnia – Área de concentração Zootecnia
– Zootecnia e Recursos Pesqueiros.

APROVADO em 27 de junho de 2019.



Prof. Dr. Fernando Higino de Lima e Silva
Avaliador externo
IF Goiano/ RV



Prof.^a Dra. Kátia Aparecida de Pinho Costa
Avaliadora interna
IF Goiano/ RV



Prof. Dr. Adriano Carvalho Costa
Presidente da banca
IF Goiano/RV



Prof. Dr. Roniel Geraldo Ávila
Avaliador externo
IF Goiano/ RV

AGRADECIMENTOS

Ao Senhor Deus, por sempre interceder na minha vida me guiando no caminho do bem e dando força para conquistar meus objetivos.

Aos meus pais, Cloves Batista e Maria Elza, por todo sacrifício para que eu e meus irmãos pudéssemos ter chances de crescer profissionalmente nos ensinando o caminho correto a seguir, sempre com muita humildade e respeito. Vocês são meu maior exemplo!!

Aos meus irmãos de sangue Shesniel e Cloves Jr, e irmãos de coração Willyan, Pablo, Fernando e Rodrigo, pelo apoio, carinho e incentivo durante minha carreira acadêmica, e em todos os momentos da minha vida.

Ao meu orientador, professor Dr. Adriano Carvalho Costa, por disponibilizar a base de dados pelas orientações e dedicação no decorrer de todo o mestrado, que sempre fez muito além de suas obrigações como orientador, sendo um grande amigo e irmão de coração. Você é fera!!

Aos meus colegas de curso e professores do Instituto Federal de Educação, Ciência e Tecnologia Goiano - campus Rio Verde do Programa de Pós-Graduação em Zootecnia, pelas experiências e aprendizados adquiridos.

À FAPEG - Fundação de Amparo à Pesquisa do Estado de Goiás, pela concessão da bolsa de mestrado através do Edital 03/2018.

Ao Instituto Federal Goiano – Campus Rio Verde, que possibilitou a conclusão de mais esta etapa na minha formação acadêmica.

À Universidade Federal de Lavras (UFLA) e ao Programa de Pós-Graduação de Zootecnia (DZO), que forneceram a estrutura para realização do estudo.

Ao Professor Dr. Moacyr Serafini, por disponibilizar a sua base de dados para o desenvolvimento desse trabalho.

ÍNDICE GERAL

	Página
PRIMEIRA PARTE – CONSIDERAÇÕES INICIAIS	11
1. INTRODUÇÃO.....	11
2. REFERENCIAL TEÓRICO.....	12
2.1 CARACTERÍSTICAS DAS ESPÉCIES.....	12
2.2 PACU (<i>Piaractus mesopotamicus</i>).....	13
2.3 PIRAPITINGA (<i>Piaractus brachypomus</i>)	14
2.4 TAMBACUI (<i>Colossoma Macropomum</i>).....	15
2.5 IDENTIFICAÇÃO DOS ANIMAIS	16
2.6 MODELO DE MISTURA.....	18
2.7 ALGORITMO NO AJUSTE DE MODELOS DE MISTURA	19
2.8 ANÁLISE DISCRIMINANTE	21
3 REFERENCIAL TEÓRICO.....	25
SEGUNDA PARTE – ARTIGO I.....	31
RESUMO	31
1 INTRODUÇÃO.....	33
2 MATERIAL E MÉTODOS.....	34
2.1 MATERIAL BIOLÓGICO E PROCEDIMENTOS EXPERIMENTAIS.....	34
2.2 AVALIAÇÃO MORFOMÉTRICA	35
2.3 ANÁLISE MOLECULAR	35
2.4 ANÁLISE ESTATÍSTICA.....	35
3 RESULTADOS	38
4 DISCUSSÃO.....	45
5 REFERENCIAL TEÓRICO.....	48
TERCEIRA PARTE – ARTIGO II.....	51
RESUMO	51
1 INTRODUÇÃO.....	53
2 MATERIAL E MÉTODOS.....	54
2.1 OBTENÇÃO DO MATERIAL BIOLÓGICO.....	54
2.2 AVALIAÇÃO MORFOMÉTRICA	54
2.3 ANÁLISE ESTATÍSTICA.....	55
3 RESULTADOS	57
4 DISCUSSÃO.....	58
5 REFERENCIAL TEÓRICO.....	60

ÍNDICE DE FIGURAS

Página

Figura 1 - Pacu (<i>Piaractus mesopotamicus</i>).	13
Figura 2 - Pirapitinga (<i>Piaractus brachypomus</i>).	14
Figura 3 - Tambaqui (<i>Colossoma macropomum</i>).	15

ÍNDICE DE TABELAS

Página

SEGUNDA PARTE – ARTIGO I

Tabela 1 - Correlação de Pearson usando o teste de Mantel entre genótipo verdadeiro e genótipo imputado para cada característica e combinadas.....	35
Tabela 2 – Correlação de Pearson usando o teste de Mantel entre a matrizes desconhecidas e imputadas no modelo unicaráter mediante uso de dados morfométricos.....	36
Tabela 3 – Correlação de Pearson usando o teste de Mantel entre a matrizes desconhecidas e imputadas no modelo bicaráter mediante uso de dados morfométricos.....	36
Tabela 4 – Correlação de Pearson usando o teste de Mantel entre a matrizes desconhecidas e imputadas no modelo tricaráter mediante uso de dados morfométricos.	37
Tabela 5 – Correlação de Pearson usando o teste de Mantel entre a matrizes desconhecidas e imputadas no modelo tetracaráter e pentacaráter mediante uso de dados morfométricos.	38
Tabela 6 – Valores médios de Correlação de Pearson usando o teste de Mantel entre a matrizes desconhecidas e imputadas nos diferentes modelos mediante uso de dados morfométricos.....	39

TERCEIRA PARTE – ARTIGO II

Tabela 1 - Coeficientes de discriminantes lineares para os grupos genéticos, obtidos a partir das medidas lineares.....	50
Tabela 2 – Porcentagem da variância total explicada pelos autovalores.....	51
Tabela 3 - Classificação dos grupos genéticos, em função das medidas lineares gerados com dados morfométricos e peso dos animais.....	51

LISTA DE ABREVIATURAS

- AC: Altura do corpo;
- CC: Comprimento de cabeça;
- CP: Comprimento padrão;
- LD: Largura do corpo;
- CC/CP: Comprimento da cabeça / comprimento padrão;
- AD/CP: Altura do corpo na linha do 1º raio da nadadeira dorsal / comprimento padrão;
- LD/CP: Largura do corpo na linha do 1º raio da nadadeira dorsal / comprimento padrão;
- CC/AD: Comprimento da cabeça / altura do corpo na linha do 1º raio da nadadeira dorsal;
- LD/CC: Largura do corpo na linha do 1º raio da nadadeira dorsal / comprimento da cabeça;
- LD/AD: Largura do corpo na linha do 1º raio da nadadeira dorsal / altura do corpo na linha do 1º raio da nadadeira dorsal.

CAPÍTULO I – CONSIDERAÇÕES INICIAIS

1. INTRODUÇÃO GERAL

A aquicultura é uma atividade promissora que tem ganhado cada vez mais espaço no agronegócio. Esse desenvolvimento é influenciado por vários fatores como o aumento da demanda pelos consumidores de alimentos nutricionalmente saudáveis como os peixes, e que tem expectativa de aumento de produção na piscicultura de 581,000 toneladas para 1.097,000 toneladas de 2016 até 2030 (FAO, 2018).

Entre as espécies nativas cultivadas que possuem importância econômica para aquicultura continental do Brasil, pode-se destacar os peixes redondos, composto principalmente pelo pacu, (*Piaractus mesopotamicus*), a pirapitinga (*Piaractus brachypomus*) e o tambaqui (*Colossoma macropomum*) e seus híbridos (PeixeBR, 2018).

O mercado aquícola tem influenciado a produção de indivíduos com melhor custo benefício possível, em virtude da concorrência entre as diversas cadeias produtivas de carne existentes no mercado mundial. Uma alternativa economicamente viável e que geralmente é empregada por produtores, é através da formação de híbridos interespecíficos, em que se busca melhorar características de importância econômica, como o ganho de peso, resistência as oscilações ambientais, rusticidade e adaptação à alimentação artificial (Prado, 2012).

Essa hibridação em peixes redondos tem sido realizada de forma indiscriminada, e produtores de alevinos têm comercializado animais não compatíveis com sua caracterização genética verdadeira, levando os piscicultores ao perigo de usar de forma indevida, híbridos no lugar dos parentais como reprodutores, indicando alerta sobre o monitoramento da produção e comercialização destas espécies e seus híbridos (Hashimoto, 2016).

A identificação desses indivíduos com base na visualização da morfologia é dificultada pela semelhança na coloração e formato entre as espécies puras e seus híbridos, podendo ser influenciada pela ação genica ou por efeitos ambientais (Scribner et al., 2001; Gomes et al., 2012).

Marcadores moleculares têm sido utilizados com grande precisão para identificar animais puros e híbridos simples em peixes. No entanto, o baixo número de marcadores disponíveis, não é suficiente para identificar cruzamentos mais complexos, como pós-F1 em que a informação sobre a origem dos cruzamentos é desconhecida (Costa et al., 2017).

Além disso, marcadores moleculares que amplificam para estas espécies, apresentam custo relativamente alto, demandam tempo para realização das análises laboratoriais e não são facilmente acessíveis para produtores.

A aplicação de métodos estatísticos, podem ser utilizados para classificação de grupos genéticos de animais com base nos caracteres fenotípicos, possibilitando fazer inferências sobre a composição genética de animais pós-F1 (Costa et al., 2017). Entre os procedimentos têm-se o modelo de mistura, que busca identificar a combinação linear de variáveis independentes que irão descrever as categorias da variável dependente de maneira ideal.

Esse procedimento que é frequentemente usado na estatística, como modelo probabilístico para representar subpopulações dentro de uma população geral, sem exigir que o conjunto de dados observados identifiquem as subpopulações que pertençam a observação individual (Assis, 2017).

A análise discriminante é outra técnica multivariada que pode ser usada para prever ou discriminar a composição genética dos animais, e tem o objetivo de estabelecer as funções das variáveis observadas que sejam responsáveis ou possam explicar as diferenças entre os grupos, permitindo assim ao pesquisador examinar se existem diferenças significativas entre os grupos, em termos de variáveis preditoras e também avaliar a precisão da classificação (Michelotti, 2018; Hair et al., 2009).

Até o presente momento, trabalhos avaliando a composição genética de pacu, pirapitinga, tambaqui e seus híbridos utilizando modelo misto de mistura multivariada e a análise discriminante, são escassos na literatura científica. Dessa forma, o presente trabalho tem como objetivo avaliar a aplicação de modelo misto de mistura multivariada e análise discriminante na identificação da composição genética de Serrasalmídeos.

2. REFERENCIAL TEÓRICO

2.1 Características das espécies

O Brasil possui aproximadamente três mil espécies de peixes, dos quais grande número com potencial para utilização dentro da piscicultura, a utilização dessas espécies na piscicultura depende de muitos fatores, tais como adaptação dos indivíduos ao novo ambiente, saindo do ambiente natural para o fechado, da relação custo benefício e a aceitação de mercado (Pinheiro et, 2014).

Dependendo da região, o cultivo é predominado por algumas espécies. Na região Norte predominam peixes como o tambaqui e o pirarucu. No Nordeste e Sudeste, a

preferência é pela tilápia, espécie mais cultivada no Brasil, no Sul tem predominância das carpas e tilápias e no Centro-Oeste, são o tambaqui e pacu (Brasil, 2014).

A produção em cativeiro de peixes redondos, tem crescido expressivamente nos últimos anos no Brasil, sendo os peixes nativos com a maior produção na aquicultura brasileira (Filho et al., 2016; Brasil, 2011).

2.2 Pacu (*Piaractus mesopotamicus*)

O pacu (Figura 1) é um teleósteo representante da superordem Ostariophysi, da ordem Characiformes, família Characidae e da subfamília Myleinae. Espécie nativa largamente encontrada nas bacias do rio Paraná, Paraguai e Uruguai, com maior distribuição nas planícies alagadas da região Centro-Oeste, e no Pantanal do Mato Grosso (Petreire, 1989).

É um peixe de escamas, com corpo romboide, achatado e robusto, possuindo coloração uniforme, variando do castanho ao cinza escuro, com o ventre amarelado. A espécie pode chegar até a 50 cm de comprimento (Borges, 2014).

Essa espécie tem hábito alimentar onívora com forte tendência a herbívoro, alimentando-se de frutos, detritos orgânicos, crustáceos, moluscos e pequenos peixes, e as condições ideais para seu desenvolvimento é água, de pH ligeiramente ácido e temperatura em torno de 26°C (Dairiki et al., 2009).



Figura 1 - Pacu (*Piaractus mesopotamicus*).

Fonte: (Aquishow, 2018).

O pacu se apresenta com grande potencial no uso da piscicultura intensiva, possuindo características importantes como, sabor agradável de sua carne, baixo custo de manutenção, resistência a patógenos, baixa exigência quanto à qualidade da água, alta

adaptabilidade ao cultivo em tanques e viveiros, além de estar entre as espécies mais apreciadas na pesca esportiva (Abimorad & Carneiro, 2008; Jomori et al., 2005).

Segundo os dados mais recentes sobre a produção de peixes nativos no Brasil, em 2016, foram produzidos em conjunto 13.084,744 toneladas de pacu e o híbrido patinga. Isso representa aproximadamente 7% da produção de peixes nativos pela aquicultura continental (IBGE, 2016).

2.3 Pirapitinga (*Piaractus brachypomus*)

A pirapitinga (Figura 3), é uma espécie natural dos rios Solimões, Amazonas, Orinoco e seus afluentes, e pode atingir 80 cm de comprimento e 20 kg de peso. Essa espécie possui hábito alimentar onívoro e se alimenta basicamente de frutas, folhas, sementes e microcrustáceos. É considerada espécie rústica, valorizada pela sua carne e de crescimento rápido, e com isso apresenta excelente condição para o cultivo em sistemas de produção fechados (Fresneda et al., 2004; Lima, 2014; Jorge, 2016).



Figura 2 - Pirapitinga (*Piaractus brachypomus*).

Fonte: (Fernando, 2011).

Essa espécie possui coloração avermelhada na região peitoral e prateada nas demais partes do corpo. Apresentam atributos importantes comercialmente como alto rendimento de carcaça (86%), carne de qualidade e saborosa, boa aceitação pelo mercado consumidor, além de possuir também coloração avermelhada (Mora, 2005; Abad et al., 2014).

No Brasil, a produção da pirapitinga (*Piaractus brachypomus*) foi de 2.099,685 toneladas, segundo o levantamento do IBGE em 2016. Em relação ao ano anterior houve redução de 41,2% da produção dessa espécie, que pode estar associada a intensificação da produção de híbridos na aquicultura continental (IBGE, 2016; Brasil, 2011). A

produção dessa espécie não é restrita as regiões em que são endêmicas. A espécie foi introduzida para uso na aquicultura em vários países da América do Sul (Colômbia, Peru e Venezuela), e também em países asiáticos (China, Mianmar, Tailândia e Vietnã) (Flores nava, 2007; Honglang 2007; Jorge, 2016).

2.4 Tambaqui (*Colossoma Macropomum*)

A espécie (Figura 2), conhecida popularmente como tambaqui, pertence à ordem Characiformes, da família Serrasalmidæ. É um peixe de escamas com corpo romboidal. Em sua fase adulta apresenta manchas escuras irregulares ventrais e caudais, com dorso em tonalidade esverdeada, mas pode variar para mais clara ou para escura dependendo da cor da água (Lima & Gomes, 2005; Borges, 2013).



Figura 3 - Tambaqui (*Colossoma macropomum*).

Fonte: (Colares, 2015).

O tambaqui é nativo da bacia amazônica, está presente no rio Solimões-Amazonas e também nos principais rios da bacia do rio Orenoco na Venezuela, cuja reprodução e criação em cativeiro se desenvolveram na região amazônica (Inoue & Boijink, 2011).

De acordo com o último levantamento de aquicultura no Brasil em 2016, o tambaqui é de longe a espécie nativa mais produzida, com cerca de 132.616,2 toneladas cultivadas, com aumento de 16% de produção em relação ao ano de 2011. Além disso, seus híbridos Tambacu e Tambatinga produziram cerca de 44.953,272 toneladas, demonstrando assim, grande importância econômica para a piscicultura nacional (IBGE, 2016; Brasil, 2011). Esta espécie tem como principais qualidades a consistência e o sabor da carne, adaptabilidade ao cativeiro, boa aceitação do mercado consumidor, alta prolificidade, hábito alimentar onívoro, além de adapta-se muito bem ao policultivo tornando essa espécie relevante para a piscicultura no Brasil por apresentar características zootécnicas favoráveis à criação e comercialização (Santos, 2009).

2.5 Identificação dos animais

Outrossim, em geral a identificação de diversos animais pode ser feita através de análises morfológicas, fisiológicas e comportamentais, embora seja recomendado utilizar técnicas que agregam maior confiabilidade ao estudo, como por exemplo técnicas moleculares. Diferentes marcadores moleculares e para diferentes espécies estão disponíveis na literatura.

Cada método possui vantagens, desvantagens e diferentes características genéticas, podendo ser marcadores dominantes ou codominantes, de regiões do DNA nuclear ou mitocondrial, de herança maternal ou biparental entre outros, deste modo, cada marcador possui diferentes procedimentos e técnicas de análise de acordo com o objetivo e técnicas que serão aplicados (Prado, 2012).

A variabilidade ou diversidade genética é a variedade de alelos e genótipos, e constitui a força que dá às populações, capacidade de suportar pressões ambientais, diminuindo o risco de extinção. A determinação da variabilidade genética é importante para populações naturais e cultivadas, visto que essas informações podem mostrar a relação genética entre indivíduos, populações e até mesmo em espécies diferentes (Santos, 2009).

Existem evidências de que muitos estoques estabelecidos, foram modificados em sua estrutura, composição gênica e tem se tornado endogâmicos, seja pelo desbalanço entre o número de machos e fêmeas para a formação da população base ou seleção intencional. Além disso, a reprodução em cativeiro eleva as chances de acasalamento entre indivíduos aparentados resultando na redução da variabilidade genética (Norris et al., 1999; Hallerman, 2003; Melo et al., 2006).

As semelhanças marcadas entre os híbridos e sua espécie-mãe, dificulta a identificação destes animais com base na morfologia, ou mesmo seu comportamento ou ecologia, devido a isso a caracterização genética de existências que passaram por alguma forma de manipulação artificial tornou-se cada vez mais importante nas explorações piscícolas em que é praticado (Gomes et al., 2012).

A identificação de peixes redondos fenotipicamente se torna um desafio, visto que dependendo do ambiente de cultivo e do nível de estresse dos animais a sua coloração pode variar, dificultando ainda mais a identificação do animal pelo fenótipo, principalmente quando há híbridos, assim, técnicas baseadas em marcadores moleculares

têm possibilitado a identificação de erros na classificação de peixes na aquicultura brasileira (Gomes et al., 2012; Hashimoto et al., 2014).

Em suma, a perda de dados é um problema que faz parte do cotidiano da maioria dos pesquisadores, seja na coleta, tabulação e análise. Visto isso, a manipulação e interpretação dessas informações devem ser feitas de maneira crítica e metódica, pois a utilização de métodos inadequados pode levar a deduções imprecisas sobre o conjunto de dados (Nunes et al., 2009). Dessa forma, pode-se utilizar a imputação, em que o processo é usado para substituir dados ausentes por estimativas dos mesmos.

Valores podem estar ausentes por várias razões, dependendo da fonte dos dados, erros de medições, mau funcionamento dos equipamentos utilizados para obtenção e falha do operador. Esse descarte das amostras ou variáveis que possuem componentes faltantes, muitas vezes significa jogar fora grande parte dos dados que poderiam ser úteis para o modelo (Costa et al., 2017).

Os primeiros estudos sobre imputação de dados faltantes foram realizados por Freeman (1975), em que gerou valores de modo iterativo minimizando a soma de quadrados do erro. Posteriormente Dempster et al., (1977), propôs os primeiros estudos sobre imputação multivariada de dados faltantes, o estudo foi elaborado seguindo alguns princípios importantes sendo eles.

1. Imputar os valores ausentes usando modelo apropriado que considera variação aleatória;
2. Repetir o primeiro passo de 3 a 5 vezes;
3. Executar a análise desejada em cada conjunto de dados usando métodos de dados completos e padronizados;
4. Calcular a média dos valores das estimativas de parâmetros nas amostras de valores faltantes, a fim de obter uma estimativa pontual única;
5. Calcular o erro padrão calculando a média dos erros padrão ao quadrado das estimativas dos dados faltantes. Depois disso, deve-se calcular a variância do parâmetro de valor ausente entre as amostras. E, posteriormente é necessário combinar os dois valores para calcular os erros padrões.

Esse método possui vários recursos que podem ser usados, como:

- Obter de estimativas quase que imparciais de todos os parâmetros do erro aleatório;
- Obter boas estimativas dos erros padrão;
- O procedimento pode ser feito com qualquer tipo de dado e nas variadas análises.

Porém, é preciso seguir algumas regras antes de aplicar o método de imputação multivariada para dados faltantes, sendo eles:

- Os dados devem estar ausentes aleatoriamente, ou seja, a probabilidade dos dados estarem, ausentes em determinada variável pode depender de outras variáveis observadas, mas não pode depender de si mesma;
- A segunda condição para a imputação multivariada de dados perdidos é que o modelo usado pelo pesquisador para imputar os valores deve ser apropriado;
- A terceira condição é que o modelo usado pelo pesquisador deve corresponder ao outro modelo usado para a imputação multivariada de dados ausentes.

Dessa forma, usando esses modelos para imputar dados composição gênica, são substituídos os valores ausentes por valores estimados. A maioria dos métodos de imputação fornecem as probabilidades de predizer os possíveis genótipos de cada marcador molecular imputado, e estas podem ser utilizadas para prever os valores genômicos dos animais nos estudos de associação genômica (Marchini & Howie, 2010; Hayes et al., 2011).

2.6 Modelo de mistura

A importância dos modelos de mistura se deve a sua capacidade de representar subpopulações com características específicas (nesse caso, características genéticas) dentro de uma população maior, através do modelo probabilístico (Assis, 2017). A alternativa que recentemente vem sendo muito utilizada quando há dados faltantes é a imputação por meio de modelos de misturas (Costa, 2015). Há diversos métodos que podem ser utilizados para estimar os parâmetros via verossimilhança. Sendo o mais conhecido o algoritmo EM, publicado por Dempster, Laird e Rubin (1977).

Em uma mistura finita com k distribuições Gaussianas, considerando para cada grupo ($g = 1, 2, \dots, k$), as proporções, as médias e as matrizes de covariâncias das misturas são simbolizados por π_g , μ_g e Σ_g , respectivamente. O valor da função de distribuição de probabilidade de uma observação y_j é dado por (Mclachlan; Peel, 2000):

$$f(y_j | \theta) = \sum_{g=1}^k \pi_g (y_j | \mu_g, \Sigma_g)$$

Em que $\theta = (\pi, \mu, \Sigma)$ é o vetor de parâmetros do modelo, contendo as proporções das misturas $\pi = \pi_1, \dots, \pi_k$, os parâmetros $\mu = \mu_1, \dots, \mu_k$ e $\Sigma = \Sigma_1, \dots, \Sigma_k$ que formam a probabilidade da função de distribuição Gaussiana:

$$f(y|\mu, \Sigma) = \frac{1}{2\pi^{d/2} |\Sigma|^{1/2}} e^{-\frac{(y-\mu)^T \Sigma^{-1} (y-\mu)}{2}}$$

Como os k grupos de dados são mutuamente exclusivos e exaustivos, a densidade conjunta dos valores observados de y_j é a soma das densidades observadas em cada grupo ponderadas pelas proporções dos grupos.

Assumindo que os indivíduos são independentes um do outro, o valor da função de probabilidade para o conjunto é simplesmente o produto das probabilidades para as observações:

$$f(Y|\theta) = \prod_{j=1}^N f(y_j|\theta) = \prod_{j=1}^N \left(\sum_{i=0}^n \pi_g(y_j | \mu_g, \Sigma_g) \right)$$

2.7 Algoritmo em no ajuste de modelos de mistura

A abordagem geral para estimar os parâmetros do modelo que vem sendo utilizada é o método da máxima verossimilhança, utilizando o algoritmo EM (E: esperança. M: maximização) proposto por Dempster, Laird e Rubin em 1977 (Dempster et al., 1977). Este foi estendido para estimar modelos de misturas Gaussianas para conjuntos de dados com os valores em falta (Ghahramani; Jordan, 1994). Recentemente, modelos de mistura de Gauss aplicados aos problemas de dados faltantes têm sido estudados extensivamente (Costa, 2017).

O algoritmo EM é um processo iterativo que melhora as estimativas atuais dos parâmetros a cada passo, sendo finalizado quando nenhuma melhoria adicional é obtida. Começando por escolher valores arbitrários para os parâmetros do modelo, em cada interação, o algoritmo substitui os valores de classe por seus valores esperados com base nos parâmetros correntes (Esperança). Em seguida, atualizar os parâmetros (Maximização), ou seja, maximiza a log-verossimilhança dos dados completos, dadas as estimativas dos dados faltantes. Devido às propriedades de ajuste do modelo, essa interação faz com que ocorra melhoria na estimação dos parâmetros em cada passo (Dempster et

al., 1977). O procedimento é repetido várias vezes, até que não se consiga melhoria considerável, ou o número máximo pré-determinado de iterações seja alcançado.

Seja $E[a | b]^{(t)}$ o valor esperado de a dado b na iteração t . Agora, mais formalmente, para a mistura de distribuições Gaussianas, o algoritmo funciona da seguinte forma:

- a) Inicialização: em que $\pi_i^{(0)}, \mu_i^{(0)}$ e $\Sigma_i^{(0)}$ assumem valores arbitrários em $t = 1$.
- b) E-passo: definir os valores esperados:

$$\hat{Z}_j^{(t)} = E \left[z_j \mid \pi_i^{(t-1)}, \mu_i^{(t-1)}, \Sigma_i^{(t-1)} \right] \text{ para cada } j = 1, \dots, N.$$

- c) M-passo: cálculo de $\pi_i^{(t)}, \mu_i^t$ e Σ_i^t para cada $i = 1, \dots, k$ com estimativas de máxima verossimilhança baseada em Y e $\hat{Z}_j^{(t)}$.

O cálculo dos valores necessários para distribuições de Gauss é simples, como se segue (Hunt; Jorgensen, 2003; Mclachlan; Peel, 2000). A esperança de aj observação pertencer ao grupo g é a probabilidade de y_j dado o grupo e os parâmetros do grupo, dividido pela soma das probabilidades de y_j , em cada grupo:

$$Z_{gj} = E [Z_{gj}] = \frac{\pi_g(y_j \mid \mu_g, \Sigma_g)}{\sum_{i=1}^k \pi_i(y_j \mid \mu_i, \Sigma_i)}$$

O vetor médio de cada grupo (g) é a média dos dados para todas as observações, ponderadas pelas probabilidades do grupo:

$$\mu_g = \frac{\sum_{j=1}^N Z_{gj} y_j}{\sum_{j=1}^N Z_{gj}}$$

A covariância de duas variáveis l e m no grupo g é a covariância das duas variáveis de todas as observações, ponderadas pelas probabilidades do grupo:

$$\sigma_{g(lm)} = \frac{\sum_{j=1}^N Z_{gj} (y_{jl} - \mu_g)(y_{jm} - \mu_g)}{\sum_{j=1}^N Z_{gj}}$$

As proporções das misturas são as somas de probabilidades de cada grupo dividido por N:

$$\pi_i = \frac{\sum_{j=1}^N Z_{ij}}{N}$$

Dada a soma das estimativas Z_{ij} para cada observação, os valores finais dos vetores Z_j dos grupos podem ser usados com a probabilidade de que determinada observação pertença a um determinado grupo, portanto informa o agrupamento desejado.

Ademais, esse processo vai sempre convergir para um máximo local da log-verossimilhança (McLachlan; Peel, 2000). Não há garantia de encontrar o máximo global; na verdade, em muitos casos, o máximo global não existe, como certos casos em que se utilizam distribuições cada vez mais estreitas sobre um ponto que pode alcançar probabilidades infinitas. Para combater estes problemas, o algoritmo é reiniciado várias vezes com diferentes valores iniciais, com um número máximo de iterações. No final, os valores mais frequentes (o encontrado mais vezes) ou a melhor (no sentido da probabilidade, dados observados) das soluções convergidas é escolhida como a "correta", sendo que nem sempre o mais frequente e a melhor solução são iguais.

2.8 Análise discriminante

Análise discriminante se fundamenta em técnicas estatísticas multivariadas e geralmente é utilizada para discriminar e classificar grupos, e vem sendo comumente usada com eficácia em diversas áreas de aplicações com o objetivo de discriminar populações pré-definidas. Essa discriminação, forma-se k grupos homogêneos na amostra, sem o conhecimento *a priori* da alocação dos objetos nos grupos (Khattree & Naik 2000).

Na Análise Discriminante, conhece-se *a priori* a que grupos pertencem os elementos da amostra, em que uma amostra de cada um dos grupos é coletada e, além de suas respectivas variáveis observadas considerando cada objeto amostrado. Também é preciso saber de que população cada amostra provém. Sendo os principais objetivos da Análise Discriminante (Lucena, 2012; Hair, 2009):

- Discriminação: encontra-se funções das variáveis observadas que possam explicar as diferenças entre as g populações;
- Classificação ou alocação: determina-se funções das variáveis observadas que permitam classificar novos objetos em alguma das g populações.

Na prática, essas funções são determinadas com base na amostra e são utilizadas para fazer a classificação de objetos extra amostra em uma das populações (Barroso, 2003).

A função discriminante linear de Fisher visa basicamente determinar a combinação linear das variáveis observadas que apresentam maior poder de discriminação entre os grupos e com a propriedade de minimizar as probabilidades de má classificação, quando as populações são normalmente distribuídas com parâmetros de média μ e variância Σ conhecidas (Fisher, 1936).

Se determinada variável está vinculada a um fator é incluída, torna-se menos improvável que variáveis adicionais daquele mesmo fator também sejam, pois são altamente correlacionadas e têm potencialmente menos poder preditivo adicional do que as variáveis que não estão naquele fator, pois seu efeito já está representado pela variável do fator incluída (Hair, 2005). Essa técnica envolvendo dados morfométricos, tem sido utilizada por alguns pesquisadores, compararam o otólito de peixes e formas de escala espécies coletadas em cinco locais diferentes (Ibáñez, et al 2017). Já Silva et al (2009), investigou os padrões de variação morfológica entre populações proveniente de diferentes bacias hidrográficas. De forma semelhante Yakubu, (2011), usou diferenças morfométricas para caracterização de duas espécies de peixes através da análise discriminante.

2.7 FUNÇÃO DISCRIMINANTE LINEAR DE FISHER

A função discriminante linear de Fisher é uma combinação linear de características originais que se caracteriza por produzir separação máxima entre duas populações (Varela, 2004).

Considerando que μ_i e Σ são parâmetros conhecidos e respectivamente, os vetores de médias e a matriz de covariâncias comum das populações π_i . Demonstra-se que a função linear do vetor aleatório X que produz separação máxima entre duas populações é dada por (Abreu, 2016):

$$D(X) = L' \cdot X = [\mu_1 - \mu_2]' \cdot \Sigma^{-1} \cdot X$$

em que,

$$X = [X_1 \ X_2 \ \dots \ X_p] \text{ e } \pi = [\pi_1, \pi_2]$$

- L = vetor discriminante;
 X = vetor aleatório de características das populações;
 μ = vetor de médias p-variado;
 Σ = matriz comum de covariâncias das populações π_1 e π_2 ;

O valor da função discriminante de Fisher para uma dada observação x_0 é:

$$D(x_0) = [\mu_1 - \mu_2]' \cdot \Sigma^{-1} \cdot x_0$$

O ponto médio entre as duas médias populacionais univariadas μ_1 e μ_2 é:

$$m = \frac{1}{2} [\mu_1 - \mu_2]' \cdot \Sigma^{-1} \cdot [\mu_1 + \mu_2], \text{ ou seja;}$$

$$m = \frac{1}{2} [D(\mu_1) + D(\mu_2)]$$

A regra de classificação baseada na função discriminante de Fisher é:

<p>Alocar x_0 em r_1 se $D(x_0) = [\mu_1 - \mu_2]' \cdot \Sigma^{-1} \cdot x_0 \geq m$</p> <p>Alocar x_0 em r_2 se $D(x_0) = [\mu_1 - \mu_2]' \cdot \Sigma^{-1} \cdot x_0 < m$</p>

Assumindo-se que as populações π_1 e π_2 têm a mesma matriz de covariâncias Σ pode-se então estimar uma matriz comum de covariâncias S_c :

$$S_c = \left[\frac{n_1 - 1}{(n_1 - 1) + (n_2 - 1)} \right] \cdot S_1 + \left[\frac{n_2 - 1}{(n_1 - 1) + (n_2 - 1)} \right] \cdot S_2$$

em que,

- S_c = estimativa da matriz comum de covariâncias Σ ;
 n_1 = número de observações da população π_1 ;
 n_2 = número de observações da população π_2 ;
 S_1 = estimativa matriz de covariâncias da população π_1 ;
 S_2 = estimativa matriz de covariâncias da população π_2 ;

A função discriminante linear amostral de Fisher é obtida substituindo-se os parâmetros μ_1 , μ_2 e Σ pelas respectivas quantidades amostrais x_1, x_2 e S_c :

$$D(x) = \hat{L}' \cdot x = [\bar{x}_1 - \bar{x}_2]' \cdot S_c^{-1} \cdot x$$

em que,

$D(x)$ = função discriminante linear amostral de Fisher;

– \hat{L} = estimativa do vetor discriminante;

x_1 = média amostral da população π_1 ;

x_2 = média amostral da população π_2 .

3 REFERÊNCIAL TEÓRICO

- ABAD, D.; RINCÓN, D.; POLEO, G. Índice de rendimento corporal em morocoto *Piaractus brachypomus* cultivados em sistema Biofloc. **Zootecnia Tropical**, Maracay, v. 32, n. 2, p. 119-130, 2014.
- ABIMORAD, E.G.; SQUASSONI, G.H.; CARNEIRO, D. J. Apparent digestibility of protein, energy, and amino acids in some selected feed ingredients for pacu *Piaractus mesopotamicus*. **Aquaculture Nutrition**, v. 14 p. 374-380, 2008.
- ABREU, A. L. E. D. **Bootstrap e modelos de Support Vector Machine-SVM**. 2016.
- ASSIS, R. C. D. **Inferência em modelos de mistura via algoritmo EM estocástico modificado**. Universidade Federal de São Carlos. Dissertação de mestrado. 2017.
- BARROSO, L.; ARTES, R. **Análise Multivariada**. Lavras: UFLA. 156p, 2003.
- BORGES, A. . Parâmetros de qualidade de pacu (*Piaractus mesopotamicus*) e tambaqui (*Colossoma macropomum*) eviscerados e armazenados em gelo por diferentes períodos. *International Food Research Journal*, v. 21, p. 589-596, 2014.
- BRASIL. **Anuário Brasileiro da Pesca e Aquicultura (ABPA). Futuro próspero: Quantidade de água disponível, clima e geografia favoráveis à aquicultura, além de uma grande biodiversidade e fronteiras de pesca extrativa ainda não exploradas podem tornar o Brasil um dos grandes produtores do setor**. Associação Cultural e Educacional Brasil - ACEB. 1ª ed, 2014.
- BRASIL. **Ministério da Pesca e Aquicultura. Boletim estatístico da pesca e aquicultura**. Brasília, p. 60, 2011.
- BRASIL. **Ministério da Pesca e Aquicultura. Boletim estatístico da pesca e aquicultura do Brasil**. Brasília, p. 128, 2012.
- CAMARGO, E. G. **Análise multivariada de características reprodutivas e de crescimento de linhagens comerciais de suínos universidade federal do paraná**. Dissertação (Mestrado em Ciências Veterinárias). **Universidade Federal do Paraná**, Curitiba, v. 53, 2013.

- COSTA, A. C.; BALESTRE, M. ; BOTELHO, H. A. ; FREITAS, Rilke Tadeu Fonseca de ; GOMES, R. C. S. ; CAMPOS, S. A. S. ; FORESTI, F. P. ; HASHIMOTO, D. T. ; MARTINS, D. G. ; PRADO, F. D. ; MENDONCA, M. A. C. . Imputation of genetic composition for missing pedigree data in Serrasalminidae using morphometric data. **SCIENTIA AGRICOLA**, v. 74, p. 443-449, 2017.
- DAIRIKI, J.K.; BALDESSIN JUNIOR, I.; PENA, S.V.; CYRINO, J.E.P. **Manual Técnico de Extensão PACU E TAMBACU**. Universidade de São Paulo, Setor de Piscicultura – Departamento de Zootecnia, f. 221, 2009.
- DEMPSTER, A. P.; Laird, N. M.; Rubin, D. B. Maximum Likelihood from Incomplete Data Via the EM Algorithm. **Journal of the Royal Statistical Society B**, 39, 1– 38, 1977.
- FAO. **The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable development goals**. Rome. Licence: CC BY-NC-SA 3.0 IGO. 2018.
- FILHO, M. X. P.; RODRIGUES, A. P. O.; REZENDE, F. P. Dinâmica da produção de tambaqui e demais peixes redondos no Brasil. **Ativos aquicultura**, ed. 7, 2016.
- FISHER, R.A. **The use of multiple measurements in taxonomic problems**. **Ann. Eugenics**, v. 7, p. 179-188, 1936.
- FLORES, A. N. **Aquaculture seed resources in Latin America: a regional synthesis. In: Assessment of freshwater fish seed resources for sustainable aquaculture (ed Bondad - Reantaso MG)**. FAO Fisheries Technical Paper, n. 501, 2007.
- FRANKHAM, R., BALLOU, J.D., E BRISCOE, D.A. **Introduction to Conservation Genetics**. **Cambridge University Press**, p. 617, 2002.
- FREEMAN, H.G. Analysis of interactions in incomplete two-ways tables. **Applied Statistics, Chichester**. v.24, p.46-55, 1975.
- FRESNEDA, A.; LENIS, G.; AGUDELO, E. et al. Espermiación inducida y crioconservación de semen de cachama blanca (*Piaractus brachypomus*). **Revista Colombiana Ciencias Pecuárias**. v. 17 p. 46-52, 2004.
- GARCÍA, D. L. F.; CANONNE, M.; QUILLET, E.; BONHOMME, F.; CHATAIN B. The application of microsatellite markers to breeding programs in the sea bass, *Dicentrarchus labrax*. **Aquaculture**, v. 159, p. 303-316, 1998.

- GHAHRAMANI, Z., & JORDAN, M. I. **Learning from incomplete data.** Massachusetts Institute of Technology, 1994.
- GOMES, F. S. H.; BARROS, C.; SAMPAIO, D.; HASHIMOTO, D. T.; PORTO-FORESTI, F.; Sampaio, I. Innovative molecular approach to the identification of *Colossoma macropomum* and its hybrids. *An Acad Bras Ciênc*, v. 84, p. 517-26, 2012.
- HAIR, J.F.; BLACK, W.C.; BABIN, B.J.; ANDERSON, R.E.; TATHAM, R.L. **Análise multivariada de dados.** 6. ed., 2005.
- HALLERMAN, E.M. Population genetics: Principles and practices for fisheries scientists. **American Fisheries Society**, Bethesda, MD, p. 458, 2003.
- HASHIMOTO, D. T. et al. Genetic Identification of F1 and Post-F1 Serrasalmid Juvenile Hybrids in Brazilian Aquaculture, **Plos One**, v. 9, San Francisco. 2014.
- HASHIMOTO, DT; PRADO, FD ; FORESTI, F ; PORTO-FORESTI, F . Molecular identification of intergenus crosses involving catfish hybrids: risks for aquaculture production. **Neotropical Ichthyology (Online)**, v. 14, p. 150-139, 2016.
- HAYES, B.J.; BOWMAN, P.J.; DAETWYLER, H.D.; KIJAS, J.W.; VAN DER WERF, J.H.J. Accuracy of genotype imputation in sheep breeds. **Animal Genetics**, Malden, v. 43, n.1, p. 72-80, 2011.
- HONGLANG, H. Freshwater fish seed resources in China. In *Assessment of Freshwater Fish Seed Resources for Sustainable Aquaculture* (Ed. M. G. BondadReantaso), p. 185–199, 2007.
- HUNT, L.; JORGENSEN, M. Mixture model clustering for mixed data with missing information. **Computational Statistics & Data Analysis**, Amsterdam, t+v. 41, n. 3/4, p. 429-440, 2003.
- IBÁÑEZ, A. L., HERNÁNDEZ-FRAGA, K., & ALVAREZ-HERNÁNDEZ, S. Discrimination analysis of phenotypic stocks comparing fish otolith and scale shapes. **Fisheries research**, v. 185, p. 6-13, 2017.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Quantidade e valor dos produtos de origem animal e variação anual - Brasil - 2015-2016.** Rio de Janeiro: IBGE, 2016.

- INOUE, L. A. K. A.; BOIJINK, C. L. **Manaus a capital do tabaqui. 2011.** Artigo em Hypertexto. Disponível em: <<http://www.infobibos.com/Artigos/2011>>.
- JOMORI, R.K. et al. Economic evaluation of *Piaractus mesopotamicus* juvenile production in different rearing systems. **Aquaculture**, v. 243, p. 175-183, 2005.
- JORGE, P. H. **Sequenciamento do transcriptoma e caracterização de microssatélites na pirapitinga *Piaractus brachypomus* para análises de variabilidade genética.** Dissertação de Mestrado em Ciências Biológicas, **Universidade Estadual Paulista.** Botucatu, 2016.
- KHATTREE, R.; NAIK, D.N. **Multivariate data reduction and discrimination with SAS software.** New York: BBU and John Wiley Sons, p. 574, 2000.
- LIMA, C. A. R. M. A.; GOMES, L. C. **Espécies nativas para piscicultura no Brasil.** Santa Maria: Editora UFSM, p. 349, 2005.
- LIMA, M. C. C. **Desenvolvimento embrionário e larval de colossoma macropomum, piaractus brachypomus e do híbrido tambatinga.** Dissertação de Mestrado em Ciência Animal. **Universidade Federal de Goiás**, f. 60, 2014.
- LUCENA, L. R. R. **Análise de condrocânio de *Rhizoprionodon porosus* (Poey, 1861) da costa nordeste do Brasil. Dissertação (Programa de Pós-Graduação em Biometria e Estatística Aplicada).** Universidade Federal Rural de Pernambuco, Recife. f. 71, 2012.
- MARCHINI, J.; HOWIE, B. Genotype imputation for genome-wide association studies. **Nature Ver. Genet.** v. 11, p. 499-511, 2010.
- MCLACHLAN, G.; PEEL, D. **Finite mixture models.** New York: **J. Wiley.** p. 456, 2000.
- MELO D.C. Caracterização genética de seis plantéis comerciais de tilápia (*Oreochromis*) utilizando marcadores microssatélites. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 58, p. 87-93, 2006.
- MICHELOTTI, V. T. **Avaliação genética de linhagens de poedeiras utilizando componentes principais e função discriminante linear de Fisher.** Dissertação de mestrado. Universidade Federal de Santa Maria, Santa Maria. p. 52, 2018.

- MORA, J. Rendimiento de la canal de la cachama blanca (*Piaractus brachypomus*) y el híbrido *Colossoma macropomun* x *Piaractus brachypomus*. Procesamiento primario y productos con valor agregado. **Bioagro**. v. 17 p. 161- 169. 2005.
- NORRIS, A.T. et al. **Microsatellite genetic variation between cultivate and wild Atlantic salmon (*Salmo salar*) population**. **Aquaculture**, v. 180, p. 247-264, 1999.
- NUNES, L. N. et al. **Multiple imputations for missing data: a simulation with epidemiological data**. **Cad. Saúde Pública, Rio de Janeiro**, v. 25, p. 268-278, 2009.
- PEIXE BR. **Anuário Peixe BR da Piscicultura 2018**. São Paulo: Associação Brasileira de Piscicultura. p. 138, 2018.
- PETRERE, M. J. R. **River Fisheries in Brazil: a review**. **Regulated Rivers: Research and Management**, v. 4 p. 1-16, 1989.
- PINHEIRO, A. Futuro próspero: Espécies. Ministério da Pesca e Aquicultura. 1º Anuário estatístico brasileiro da pesca e aquicultura: 2014. Brasília: ACEB/MPA, 2014. p. 13-14.
- PRADO, F. D. D. et al. **Cytogenetic characterization of F1, F2 and backcross hybrids of the Neotropical catfish species *Pseudoplatystoma corruscans* and *Pseudoplatystoma reticulatum* (Pimelodidae, Siluriformes)**. **Genetics and molecular biology**, v. 35, n. 1, p. 57-64, 2012.
- SANTOS, Maria Da Conceicao F.; HRBEK, Tomas; FARIAS, Izeni P. **Microsatellite markers for the tambaqui (*Colossoma macropomum*, Serrasalminidae, Characiformes), an economically important keystone species of the Amazon River floodplain**. **Molecular ecology resources**, v. 9, n. 3, p. 874-876, 2009.
- SCRIBNER, K.T.; PAGE, K.S.; BARTRON, M.L. 2001. Hybridization in freshwater fishes: a review of case studies and cytonuclear methods of biological inference. **Rev Fish. Biol Fish**, v. 10 p. 293-323, 2001.
- SILVA, E. L., CENTOFANTE, L., & MIYAZAWA, C. S. Análise morfométrica em *Thoracocharax stellatus* (Kner, 1858)(Characiformes, Gasteropelecidae) proveniente de diferentes bacias hidrográficas Sul-americanas. **Biota Neotropica**, v. 9, p. 71-76, 2009.

VARELLA, C. A. A. **Estimativa da produtividade e do estresse nutricional da cultura do milho usando imagens digitais. Universidade Federal de Viçosa. Tese de Doutorado. 2004.**

YAKUBU, A., & OKUNSEBOR, S. A. Morphometric differentiation of two Nigerian fish species (*Oreochromis niloticus* and *Lates niloticus*) using principal components and discriminant analysis. **International Journal of Morphology**. v. 29, p. 1429-1434, 2011.

CAPÍTULO II – ARTIGO I
**MODELO MISTO MULTIVARIADO NA IMPUTAÇÃO DE COMPOSIÇÃO
GENÉTICA EM PEIXES REDONDOS**

RESUMO

O presente trabalho objetivou-se avaliar modelos multivariados na imputação da composição genética de peixes Serrasalminidae com base no peso corporal em medidas morfométricas do corpo. Foram adquiridos 96 juvenis de duas pisciculturas comerciais, sendo 12 de cada grupo genético: pacu, pirapitinga, tambaqui, tambacu, tambatinga, patinga, paqui e piraqui. Os animais foram distribuídos aleatoriamente em 16 caixas d'água em sistema de recirculação, sendo duas caixas por grupo genético com densidade de seis animais por caixa, e foram cultivados até 495 dias de idade, sendo pesados e submetidos à análise morfométrica. Para confirmação da identidade dos animais foram utilizados dois marcadores nucleares e um mitocondrial. Para os animais que foram considerados híbridos avançados (pós-F1) pela análise molecular, foi utilizada a metodologia de modelos mistos de misturas multivariados para imputar a composição genética dos animais com base no peso corporal e nas medidas morfométricas. Utilizou-se a metodologia proposta por Griffing (1956), considerando o modelo misto, em que os efeitos ambientais foram estimados através dos EBLUE, e os efeitos genéticos foram considerados como aleatórios, sendo obtidos os EBLUP dos efeitos gerais e específicos de combinações. Para verificar a concordância da composição genética dos animais obtida via imputação, e a conhecida para as características avaliadas e entre estas, foi realizada a correlação de Pearson utilizando o teste de Mantel. Houve diferença na correlação entre os genótipos verdadeiros e os imputados entre as variáveis utilizadas nos diferentes modelos (uni e multicaráter). Observou-se que com acréscimo de variáveis morfométricas, elevou-se consideravelmente a correlação para a maioria das características, isso devido alta correlação genética e herdabilidade dessas variáveis, melhorando a acurácia de predição da composição genética desses indivíduos. Através da imputação foi possível fazer inferência sobre a composição genética dos animais para as diferentes metodologias e combinação das variáveis.

Palavras-chave: *Colossoma Macropomum*. Modelo de mistura. Peixes nativos. *Piaractus brachypomum*. *Piaractus mesopotamicus*. *Serrasalmideos*.

ABSTRACT

The present research was carried out to impute the genetic composition for missing pedigree data in Serrasalmideos based on body weight and morphometric measurements. A total of 96 juveniles were purchased from two commercial fish farms, 12 of each of the following genetic groups: pacu, pirapitinga, tambaqui, tambacu, tambatinga, patinga, paqui and piraqui. The animals were randomly distributed in 16 water boxes in a recirculation system, two boxes per genetic group and six animals per box density, where they were grown up to 495 days old, being weighed and submitted to morphometric analysis. Two nuclear and one mitochondrial markers were used to confirm the animals identity. For the animals that were considered advanced hybrids (post-F1) by molecular analysis, the mixed models methodology of multivariate mixtures was used to impute the kinship of the animals based on body weight and morphometric measurements. The Griffing (1956) methodology was used, considering the mixed model where the environmental effects were estimated through the EBLUE and the genetic effects were considered as random, obtaining the EBLUP of general and specific effects of combinations. In order to verify the genetic composition concordance of animals obtained via imputation and the known one for the characteristics evaluated and among them, the Pearson correlation was performed using the Mantel test. There was a difference in the correlation between the true and imputed genotypes among the variables used in different models (uni and multivariate). It was observed that, with the morphometric variables addition, the correlation for most of the characteristics was elevated considerably, due to the high genetic correlation and heritability of these variables, and in this way, they had better explain the genetic composition of these individuals. Through imputation it was possible to make inferences about animals genetic composition for different methodologies and variables combination.

Keywords: *Colossoma Macropomum*. Mixing model. Native fish. *Piaractus brachypomum*. *Piaractus mesopotamicus*. *Serrasalmideos*.

1 INTRODUÇÃO

O *Colossoma macropomum*, *Piaractus mesopotamicus* e *Piaractus brachypomus* são espécies de peixes que possuem características nutricionais e sensoriais atraentes para o mercado consumidor, sendo cultivadas em diversos países da América do Sul, especialmente Brasil, Colômbia, Peru e Venezuela, bem como em países asiáticos como China, Tailândia e Vietnã. (Jorge, 2016; Costa et al., 2018). A produção de híbridos interespecíficos entre essas espécies tem sido muito empregada com objetivo de obter heterose e complementariedade para as características de importância econômica (Mourad et al., 2018).

A dificuldade de identificação acurada da composição genética dos indivíduos destas espécies por meio da visualização morfológica vem sendo relatada em vários trabalhos científicos (Hashimoto et al., 2014, Costa et al., 2017). Dependendo da interação alélica e dos efeitos ligados ao sexo, materno, paterno e ambiental, os híbridos podem fenotipicamente assemelhar-se a seus pais, resultando assim em identificações errôneas (Costa et al., 2017).

Marcadores moleculares e mitocondriais têm sido utilizados com grande precisão na identificação de animais puros e híbridos simples (F1) (Gomes et al., 2012; Hashimoto et al., 2011, 2014). Entretanto, para animais oriundos de cruzamentos mais complexos, ou seja, em que pelo menos um dos genitores é um híbrido, não é possível identificar a composição genética do indivíduo, sendo estes classificados como pós-F1 (Costa et al., 2017).

Modelos de misturas finitas têm sido empregados para realizar imputação em diversas áreas. A aplicação de forma direta, é quando tem-se uma população e esta é formada por subpopulações claramente definidas com suas características próprias, apesar de raramente se saber a qual população pertence cada observação. Cada subpopulação é descrita por sua densidade e as ponderações das misturas são as probabilidades de cada observação pertencer a uma certa subpopulação (Luca e Zuccolotto, 2003). Esses modelos já mostraram eficácia na classificação peixes utilizando abordagem univariada (Costa et al., 2017).

A análise multivariada possui vantagens em relação a univariada por considerar a covariância entre as variáveis, resultando em maior precisão e acurácia nos resultados. Modelos de misturas mistos multivariados são promissores para a imputação de

composição genética baseada em fenótipo, especialmente em cruzamentos com pais desconhecidos. Portanto, objetivou-se avaliar modelos multivariados na imputação da composição genética de peixes Serrasalminidae com base no peso corporal em medidas morfométricas do corpo.

2 MATERIAL E MÉTODOS

2.1 Material biológico e procedimentos experimentais

Os dados utilizados no projeto foram coletados no ano 2012 em um experimento conduzido na Estação de Piscicultura da Universidade Federal de Lavras, por 465 dias, dividido entre 180 dias para adaptação e 285 dias experimentais. O experimento foi realizado no laboratório de recirculação de água, equipado com 16 tanques de água com capacidade para 500 litros que consistiu em um biofiltro, sonda pt100, controlador de temperatura (N540) e uma bomba (1/3 hp).

Para este, 96 juvenis de Serrasalminidae foram adquiridos de duas fazendas comerciais de peixe no estado de São Paulo, dos seguintes grupos genéticos (espécies e seus híbridos): pacu, pirapitinga, tambaqui, patinga ($\text{♀pacu} \times \text{♂pirapitinga}$), paqui ($\text{♀pacu} \times \text{♂tambaqui}$), piraqui ($\text{♀pirapitinga} \times \text{♂tambaqui}$), tambacu ($\text{♀tambaqui} \times \text{♂pacu}$) e tambatinga ($\text{♀tambaqui} \times \text{♂pirapitinga}$).

Os peixes foram colocados em 16 tanques internos (500 L) por um período de aclimação de cinco meses. Durante esse período de adaptação, 13 peixes morreram, reduzindo o número de 96 para 83 animais. A densidade de estoque foram seis peixes por tanque, sendo dois por grupo genético. Os tanques foram controlados para que houvesse três renovações totais a cada hora e a temperatura mantida a 28°C ao longo do período de estudo.

Os peixes foram alimentados *ad libitum* com alimentos comerciais contendo 32% de proteína bruta, fornecidos três vezes por dia (08h às 12h e 16h) até atingir 100 g de peso corporal, e depois duas vezes por dia (08h e 14h) depois. O oxigênio dissolvido foi monitorado diariamente e amônia, nitrito, nitrato e pH a cada três dias. Aos 180 dias de idade, os peixes foram contados e os valores foram utilizados como efeitos fixos no modelo estatístico.

Não foi necessário nova aprovação do Comitê de Ética para esse estudo, pois os dados foram obtidos de um banco existente, o mesmo foi aprovado pelo Comitê de Ética

em Experimentação Animal da Universidade de Lavras, Lavras, Brasil (Protocolo 074/13 de 27 de fevereiro de 2014).

2.2 Avaliação morfométrica

No final do período experimental, os peixes foram anestesiados com benzocaína (60 mg L⁻¹), pesados e analisados em termos dos seguintes parâmetros morfométricos:

- Comprimento padrão (CP), entre a extremidade anterior da cabeça até o final do perímetro da nadadeira caudal;
- Comprimento da cabeça (CC), entre a extremidade anterior da cabeça e a borda caudal do opérculo;
- Altura do corpo (AC), medida a frente do primeiro raio de nadadeira dorsal;
- Largura do corpo (LD), medida na região do primeiro raio de nadadeira dorsal.

2.3 Análise molecular

Para determinar o grupo de cada peixe, amostras da nadadeira foram coletadas e fixadas em álcool 100%, posteriormente o material genético foi extraído e purificado com o Wizard Genomic DNA Purification Kit (Promega), seguindo as recomendações do fabricante.

Para a identificação de parentes e híbridos, foram utilizados dois marcadores SNP (polimorfismo de nucleotídeo único) de genes nucleares e fragmentos de genes mitocondriais COI (citocromo c oxidase subunidade I) e CYTB (citocromo B). Os fragmentos amplificados primeiramente foram visualizados sob iluminação UV e depois analisados os diferentes tamanhos de fragmentos de DNA em gel de agarose a 2% corados com brometo de etídio (1 ng mL⁻¹), em que foi feito o diagnóstico para cada espécie parental e seus híbridos interespecíficos (Hashimoto et al., 2011).

2.4 Análise estatística

A análise estatística foi realizada para imputar a composição genética dos animais com base em todos os fenótipos medidos (peso corporal e morfometria). No contexto dialélico, foi aplicado o método proposto por Griffing (1956), adaptado à análise do modelo misto, em que os efeitos ambientais foram estimados através de EBLUE e os efeitos genéticos foram considerados aleatórios, obtendo os EBLUP dos efeitos de capacidade de combinação geral (CGC) e específicos (CEC). No entanto, todas as informações sobre os parentais estão disponíveis, assim, a análise dialélica foi realizada pelo modelo de mistura mista normal, que imputa as matrizes de incidência para os efeitos

aleatórios (CGC e CEC) quando a composição genética do animal é desconhecida. Para fazer inferências sobre a composição genética dos animais amostrados, considera-se a matriz de incidência imputada da CGC porque representa o efeito genético aditivo no modelo. O efeito recíproco não foi considerado no modelo. O modelo misto multivariado segue abaixo:

$$Y = X\beta + E[Z_1]\alpha_1 + E[Z_2]d + \varepsilon$$

Em que: Y é o vetor ou matriz das variáveis fenotípicas (peso corporal e medidas morfométricas), X é a matriz do efeito fixo que incluem o número de peixes em cada tanque; Z_1 e Z_2 são as matrizes CGC e CEC; β é o vetor de efeitos fixos; e α_1 e d são o efeito aleatório relacionado a CGC e CEC, respectivamente. Os efeitos aleatórios (α_1 e d) foram assumidos como $\alpha_1 \sim MVN(0, \sigma^2)$, e $d_1 \sim MVN(0, \sigma_d^2)$, e os resíduos aleatórios ε foram assumidos como sendo normalmente distribuídos, não correlacionados e têm variação homogênea. Se todos os pedigrees são conhecidos, então Z_1 e Z_2 carregam informações completas, mas, como alguns genitores estavam faltando, tem que imputar Z_1 e Z_2 com base na expectativa do genótipo.

A função de verossimilhança total pode ser descrita da seguinte forma:

$$L(\theta | Y) = \prod_{j=1}^N p(y_j | \theta)$$

Em que:

$b + E[Z_1]a_1 + E[Z_2]d, \sigma_e^2$ é o vetor dos parâmetros do modelo.

Os modelos mistos multivariados de misturas normais considerados g como denotando grupo genético ($g = 1, 2, \dots, 6$), n_g como a maximização das proporções (presunção inicial, um dado animal pertence a determinado grupo), μ_g como vetor médio e σ_e^2 como a variância do erro. A probabilidade total para os dados pode ser representada pela seguinte equação:

$$L(\theta | y) = \underbrace{\prod_{j=1}^k p(Y_j | \theta)}_{\text{efeito fixo}} \underbrace{\prod_{i=k+1}^n \left(\sum_{g=1}^l \pi_g (y_j | \mu_g \sigma^2) \right)}_{\text{mistura normal}}$$

Em que: $\pi_g(\pi_1, \dots, \pi_6) = \frac{1}{6}$, ou seja, a probabilidade inicial de um indivíduo com composição genética desconhecida pertencente a qualquer cruzamento é 1/6 e μ_g (μ_1, \dots, μ_6) são os valores genotípicos relacionados ao cruzamento g-ésimo.

Os valores observados foram agrupados em produtos de sequências, em que o contador variou de j para k . Os dados faltantes sobre a composição genética foram agrupados no contador variando de i para n , onde $i = k + 1$. A expectativa da j -ésima observação pertencente ao grupo g é dada pela probabilidade condicional de y_j , dado o grupo e seus parâmetros, normalizados pela soma das probabilidades de y_j . Portanto, é uma variável multinomial:

$$\hat{Z}_{2jg} = E[Z_{gi}] = \frac{\pi_g (y_j | \mu_g, \sigma^2)}{\sum_{i=1}^k r_i (y_j | \mu_i, \sigma^2)}$$

O vetor médio de cada grupo (g), ou seja, a média dos dados para todas as observações, ponderada pelas probabilidades do grupo, foi estimada por:

$$\mu_g = E[Z_1]\alpha_1 + E[Z_2]d$$

Esta parte relacionada ao modelo misto refere-se à expectativa do valor genotípico sob o cruzamento g -ésimo. Por exemplo, se o j -ésimo indivíduo apresentar uma probabilidade de $Z_{2lj} = 0.9$ a serem alocados no primeiro grupo (fundo genotípico puro), então, $E[Z_{1j1}] = 2 * 0.9$, em que 2 é o número de alelos obtidos a partir de um genótipo puro.

Portanto, o BLUP empírico para CGC e CEC são dados, respectivamente, por:

$$\hat{a}_1 = \left(E[Z_1]' E[Z_1] + \frac{\sigma_e^2}{\sigma_a^2} \right)^{-1} E[Z_1]' (y - X\beta - \hat{Z}_{2d})$$

$$d = \left(\hat{Z}_2' \hat{Z}_2 + \frac{\sigma_e^2}{\sigma_a^2} \right)^{-1} \hat{Z}_2' (y - XB - E[Z_1]\hat{a}_1)$$

O programa foi formulado usando PROC IML do SAS v. 8.0. Os componentes de variância foram estimados por RNs (REML) usando um critério de convergência igual a 1×10^{-5} (Patterson & Thompson, 1971). Conforme descrito acima, utilizando a matriz de incidência CGC (parte aditiva do modelo), que foi obtida utilizando a imputação do peso corporal das observações e agrupa os animais com semelhanças de segregação.

O teste da razão de verossimilhança foi realizado para verificar as diferenças entre frequências esperadas e observadas, de acordo com o tipo de cruzamento observado, como animais puros, híbridos de cruzamento simples ou híbridos cruzados de três vias,

entre outros. A validação cruzada foi realizada usando o processo leave-one-out para imputar a composição genética de cada peixe com base no peso corporal e outros traços morfométricos analisados.

Então, foi possível realizar associação entre o genótipo verdadeiro e o genótipo imputado pela correlação de Pearson usando o teste de Mantel, com o pacote vegano 2.0-10 e a versão de software R 3.1.0.

Para a validação cruzada, apenas os genótipos apresentando pedigree conhecido foram utilizados e a probabilidade total foi dada por:

$$L(\theta | Y) = \prod_{j=1}^{k-1} p(y_j | \theta) \left(\sum_{g=1}^l \pi_g (y_k | \mu_g \sigma^2) \right)$$

3 RESULTADOS

O processo leave-one-out foi utilizado para comparar a predição da composição genética de animais dos diferentes modelos e características (peso corporal e nas medidas morfométricas) com composição genética conhecida. Observou-se que houve diferença no coeficiente de correlação entre os genótipos verdadeiros e os imputados entre as variáveis utilizadas nos diferentes modelos (uni e multicaráter), conforme Tabela 1.

No modelo unicaráter observou-se que as correlações foram de baixa a moderada, variando de 0,24 a 0,43, para o LD e o CC, respectivamente. No modelo bicaráter verificou-se que as correlações foram moderadas, variando de 0,3 a 0,68, para as variáveis PESO-AD, CP-AD, respectivamente. Para o modelo tricaráter, constatou-se que as correlações foram de moderada a alta, variando de 0,6 a 0,79, com o menor valor para o modelo com PESO, CP e LD e 0,79 para os modelos PESO-CP-CC e CP-CC-AD. Os modelos tetra e o pentacaráter apresentaram correlações altas entre o genótipo imputado e o verdadeiro.

Por conseguinte, de maneira geral, verificou-se correlações médias moderada para as metodologias utilizando o modelo uni, bi e tricaráter e correlações médias altas para os modelo tetra e pentacaráter.

Nas Tabelas 2, 3 e 4, são apresentadas as correlações lineares Pearson através do teste de mantel entre as matrizes desconhecidas e imputadas utilizando as diferentes metodologias (unicaráter, bicaráter, tricaráter, tetracaráter e pentacaráter) com as combinações das variáveis. Na tabela 5 a média das correlações de mantel entre as combinações das metodologias.

Considerando todas as correlações, 16,36% foram fraca ($r \leq 0,3$), 64,54% moderada ($0,3 < r < 0,7$), sendo que a maioria foram encontradas nos modelos bicaráter e tricaráter. Os 19,1% restantes, apresentaram alta correlação ($r \geq 0,7$), apesar da maior parte ser identificada no modelo unicaráter.

Tabela 1 - Correlação de Pearson usando o teste de Mantel entre genótipo verdadeiro e genótipo imputado para cada característica e combinadas.

GENÓTIPO IMPUTADO	MODELO				
	Unicaráter	Bicaráter	Tricaráter	Tetracaráter	Pentacaráter
peso	0,34*	peso-cp	peso-cp-cc	cp-cc-ad-ld	peso-cp-cc- ad-ld
	0,32*	0,65*	0,79*	0,77*	0,80*
cp		peso-cc	peso-cp-ad	peso-cc-ad-ld	-
	0,43*	0,48*	0,69*	0,72*	-
cc		peso-ad	peso-cp-ld	peso-cp-ad-ld	-
	0,30*	0,30*	0,60*	0,71*	-
ad		peso-ld	peso-cc-ad	peso-cp-cc-ad	-
	0,30*	0,55*	0,49*	0,80*	-
ld		cp-cc	peso-cc-ld	peso-cp-cc-ld	-
	0,24*	0,67*	0,69*	0,80*	-
-		cp-ad	peso-ad-ld	-	-
-		0,68*	0,60*	-	-
-		cp-ld	cp-ad-ld	-	-
-		0,31*	0,63*	-	-
-		cc-ad	cp-cc-ad	-	-
-		0,57*	0,79*	-	-
-		cc-ld	cp-cc-ld	-	-
-		0,65*	0,68*	-	-
-		ad-ld	cc-ad-ld	-	-
-		0,60*	0,66*	-	-
Correlação média	0,33	0,55	0,66	0,76	0,80

CP: comprimento padrão; CC: comprimento de cabeça; AC: altura corporal; LD: largura corporal; *($P < 0,05$)

Na Tabela 2, pode-se verificar a classificação dos animais desconhecidos de acordo com a metodologia utilizada, sendo observado que de acordo com a metodologia, o animal pode ser classificado de forma diferente.

Com base na significância de cada correlação foi possível verificar que apenas três correlações, não foram significativas, sendo que todas a variável LD estava inserida no modelo (Tabela 1).

Tabela 2 – Correlação de Pearson usando o teste de Mantel entre a matrizes desconhecidas e imputadas no modelo unicaráter mediante uso de dados morfométricos.

Modelo	Variáveis	Unicaráter			
		cp	cc	ad	ld
Unicaráter	peso	0,88*	0,70*	0,85*	0,73*
	cp	-	0,55*	0,70*	0,76*
	cc	-	-	0,73*	0,46*
	ad	-	-	-	0,58*

CP: comprimento padrão; CC: comprimento de cabeça; AC: altura corporal;

LD: largura corporal; *(P<0,05)

Tabela 3 – Correlação de Pearson usando o teste de Mantel entre a matrizes desconhecidas e imputadas no modelo bicaráter mediante uso de dados morfométricos.

Variáveis	Bicaráter									
	peso cp	peso cc	peso ad	peso ld	cp cc	cp ad	cp ld	cc ad	cc ld	ad ld
Unicaráter										
peso	0,64*	0,37*	0,67*	0,76*	0,45*	0,56*	0,53*	0,50*	0,32*	0,53*
cp	0,60*	0,30*	0,63*	0,66*	0,43*	0,53*	0,60*	0,39*	0,30*	0,44*
cc	0,47*	0,58*	0,66*	0,61*	0,47*	0,37*	0,34*	0,68*	0,55*	0,41*
ad	0,49*	0,41*	0,77*	0,67*	0,30*	0,54*	0,42*	0,51*	0,33*	0,57*
ld	0,51*	0,24ns	0,46*	0,68*	0,49*	0,42*	0,44*	0,44*	0,29*	0,49*
Bicaráter										
peso cp	-	0,33*	0,48*	0,68*	0,57*	0,44*	0,44*	0,40*	0,33*	0,36*
peso cc	-	-	0,38*	0,32*	0,47*	0,17*	0,13ns	0,57*	0,51*	0,18*
peso ad	-	-	-	0,58*	0,28*	0,64*	0,38*	0,54*	0,40*	0,58*
peso ld	-	-	-	-	0,50*	0,45*	0,50*	0,59*	0,48*	0,55*
cp cc	-	-	-	-	-	0,23*	0,27*	0,66*	0,71*	0,19*
cp ad	-	-	-	-	-	-	0,37*	0,34*	0,19*	0,52*
cp ld	-	-	-	-	-	-	-	0,22*	0,32*	0,50*
cc ad	-	-	-	-	-	-	-	-	0,73*	0,39*
cc ld	-	-	-	-	-	-	-	-	-	0,31*

CP: comprimento padrão; CC: comprimento de cabeça; AC: altura corporal; LD: largura corporal; ns: não significativo; *(P<0,05).

Tabela 4 – Correlação de Pearson usando o teste de Mantel entre a matrizes desconhecidas e imputadas no modelo tricaráter mediante uso de dados morfométricos.

Variáveis	Tricaráter									
	peso	peso	peso	peso	peso	peso	cp	cp	cp	cc
	cp	cp	cp	cc	cc	ad	ad	cc	cc	ad
	cc	ad	ld	ad	ld	ld	ld	ad	ld	ld
peso	0,31*	0,60*	0,50*	0,38*	0,49*	0,62*	0,34*	0,43*	0,44*	0,44*
cp	0,29*	0,56*	0,53*	0,35*	0,52*	0,53*	0,36*	0,39*	0,46*	0,30*
cc	0,52*	0,57*	0,43*	0,65*	0,43*	0,52*	0,26*	0,53*	0,47*	0,59*
ad	0,29*	0,57*	0,44*	0,46*	0,43*	0,62*	0,39*	0,30*	0,36*	0,47*
ld	0,24*	0,47*	0,32*	0,33*	0,32*	0,46*	0,15*	0,48*	0,32*	0,26*
peso cp	0,43*	0,89*	0,81*	0,35*	0,81*	0,37*	0,29*	0,54*	0,48*	0,28*
peso cc	0,55*	0,45*	0,26*	0,62*	0,25*	0,24*	0,10ns	0,49*	0,44*	0,47*
peso ad	0,34*	0,57*	0,48*	0,55*	0,47*	0,62*	0,52*	0,31*	0,35*	0,50*
peso ld	0,35*	0,65*	0,66*	0,45*	0,66*	0,57*	0,39*	0,53*	0,44*	0,50*
cp cc	0,68*	0,58*	0,40*	0,39*	0,40*	0,22*	0,10ns	0,94*	0,82*	0,53*
cp ad	0,23*	0,45*	0,40*	0,24*	0,40*	0,56*	0,81*	0,24*	0,28*	0,26*
cp ld	0,24*	0,40*	0,58*	0,18*	0,58*	0,49*	0,50*	0,24*	0,47*	0,23*
cc ad	0,63*	0,43*	0,32*	0,78*	0,32*	0,34*	0,22*	0,76*	0,57*	0,76*
cc ld	0,68*	0,43*	0,44*	0,81*	0,43*	0,29*	0,23*	0,76*	0,72*	0,73*
ad ld	0,16*	0,41*	0,35*	0,32*	0,35*	0,81*	0,44*	0,23*	0,24*	0,46*
peso cp cc	-	0,48*	0,39*	0,76*	0,39*	0,18*	0,19*	0,74*	0,72*	0,52*
peso cp ad	-	-	0,75*	0,51*	0,76*	0,47*	0,34*	0,56*	0,51*	0,39*
peso cp ld	-	-	-	0,29*	1,00*	0,36*	0,52*	0,38*	0,52*	0,35*
peso cc ad	-	-	-	-	0,29*	0,34*	0,17*	0,76*	0,68*	0,71*
peso cc ld	-	-	-	-	-	0,36*	0,52*	0,38*	0,52*	0,35*
peso ad ld	-	-	-	-	-	-	0,45*	0,26*	0,24*	0,44*
cp ad ld	-	-	-	-	-	-	-	0,38*	0,53*	0,35*
cp cc ad	-	-	-	-	-	-	-	-	0,76*	0,62*
cp cc ld	-	-	-	-	-	-	-	-	-	0,52*

CP: comprimento padrão; CC: comprimento de cabeça; AC: altura corporal; LD: largura corporal; ns: não significativo; *(P<0,05).

Tabela 5 – Correlação de Pearson usando o teste de Mantel entre a matrizes desconhecidas e imputadas no modelo tetracaráter e pentacaráter mediante uso de dados morfométricos.

Variáveis	Tetracaráter					Pentacaráter
	peso cp cc ad	peso cc ad ld	peso cp ad ld	peso cp cc ld	cp cc ad ld	peso cp cc ad ld
peso	0,32*	0,39*	0,56*	0,31*	0,38*	0,41*
cp	0,30*	0,32*	0,53*	0,33*	0,32*	0,39*
cc	0,51*	0,57*	0,47*	0,52*	0,50*	0,54*
ad	0,29*	0,47*	0,53*	0,30*	0,37*	0,40*
ld	0,24*	0,24*	0,39*	0,30*	0,21*	0,27*
peso cp	0,43*	0,33*	0,45*	0,45*	0,33*	0,39*
peso cc	0,53*	0,54*	0,24*	0,49*	0,46*	0,48*
peso ad	0,34*	0,54*	0,53*	0,37*	0,37*	0,42*
peso ld	0,35*	0,44*	0,55*	0,43*	0,46*	0,45*
cp cc	0,68*	0,63*	0,21*	0,73*	0,68*	0,70*
cp ad	0,22*	0,29*	0,42*	0,26*	0,22*	0,26*
cp ld	0,25*	0,21*	0,69*	0,30*	0,40*	0,41*
cc ad	0,62*	0,66*	0,28*	0,68*	0,62*	0,62*
cc ld	0,68*	0,66*	0,31*	0,80*	0,72*	0,74*
ad ld	0,17ns	0,40*	0,68*	0,21*	0,33*	0,74*
peso cp cc	0,99*	0,68*	0,26*	0,89*	0,69*	0,74*
peso cp ad	0,48*	0,47*	0,49*	0,50*	0,38*	0,44*
peso cp ld	0,40*	0,32*	0,53*	0,46*	0,45*	0,48*
peso cc ad	0,75*	0,81*	0,26*	0,82*	0,69*	0,74*
peso cc ld	0,40*	0,31*	0,53*	0,46*	0,45*	0,74*
peso ad ld	0,18*	0,47*	0,64*	0,21*	0,31*	0,32*
cp ad ld	0,40*	0,32*	0,54*	0,46*	0,45*	0,34*
cp cc ad	0,73*	0,66*	0,20*	0,79*	0,70*	0,72*
cp cc ld	0,73*	0,69*	0,42*	0,77*	0,85*	0,89*
cc ad ld	0,52*	0,64*	0,35*	0,60*	0,65*	0,62*
peso cp cc ad	-	0,68*	0,26*	0,89*	0,69*	0,74*
peso cc ad ld	-	-	0,32*	0,65*	0,80*	0,80*
peso cp ad ld	-	-	-	0,30*	0,46*	0,50*
peso cp cc ld	-	-	-	-	0,71*	0,78*
cp cc ad ld	-	-	-	-	-	0,94*

CP: comprimento padrão; CC: comprimento de cabeça; AC: altura corporal; LD: largura corporal; ns: não significativo; *(P<0,05).

Destarte que, considerando cada modelo separadamente, os resultados apontaram que o modelo unicaráter obteve maior coerência entre si na imputação, em que 70% das metodologias apresentaram alta correlação ($r > 0,7$). Dentro das metodologias bi, tri e tetracaráter, verificou-se que maioria das correlações foram moderadas ($0,3 > r > 0,7$), 76, 64 e 50%, respectivamente. Indagado também que a correlação média entre o modelo pentacaráter com as demais metodologias foi aumentando à medida que se aumentava e incrementava caracteres (Tabela 6).

Tabela 6 – Valores médios de Correlação de Pearson usando o teste de Mantel entre a matrizes desconhecidas e imputadas nos diferentes modelos mediante uso de dados morfométricos.

Unicaráter	Bicaráter	Tricaráter	Tetracaráter	Pentacaráter
0,69	0,50	0,43	0,39	0,40
-	0,43	0,46	0,45	0,52
-	-	0,48	0,54	0,60
-	-	-	0,52	0,75

Na figura 1, pode-se observar a classificação dos animais com composição genética desconhecidas, mostrando que de acordo com a metodologia o animal pode ser classificado de forma diferente.

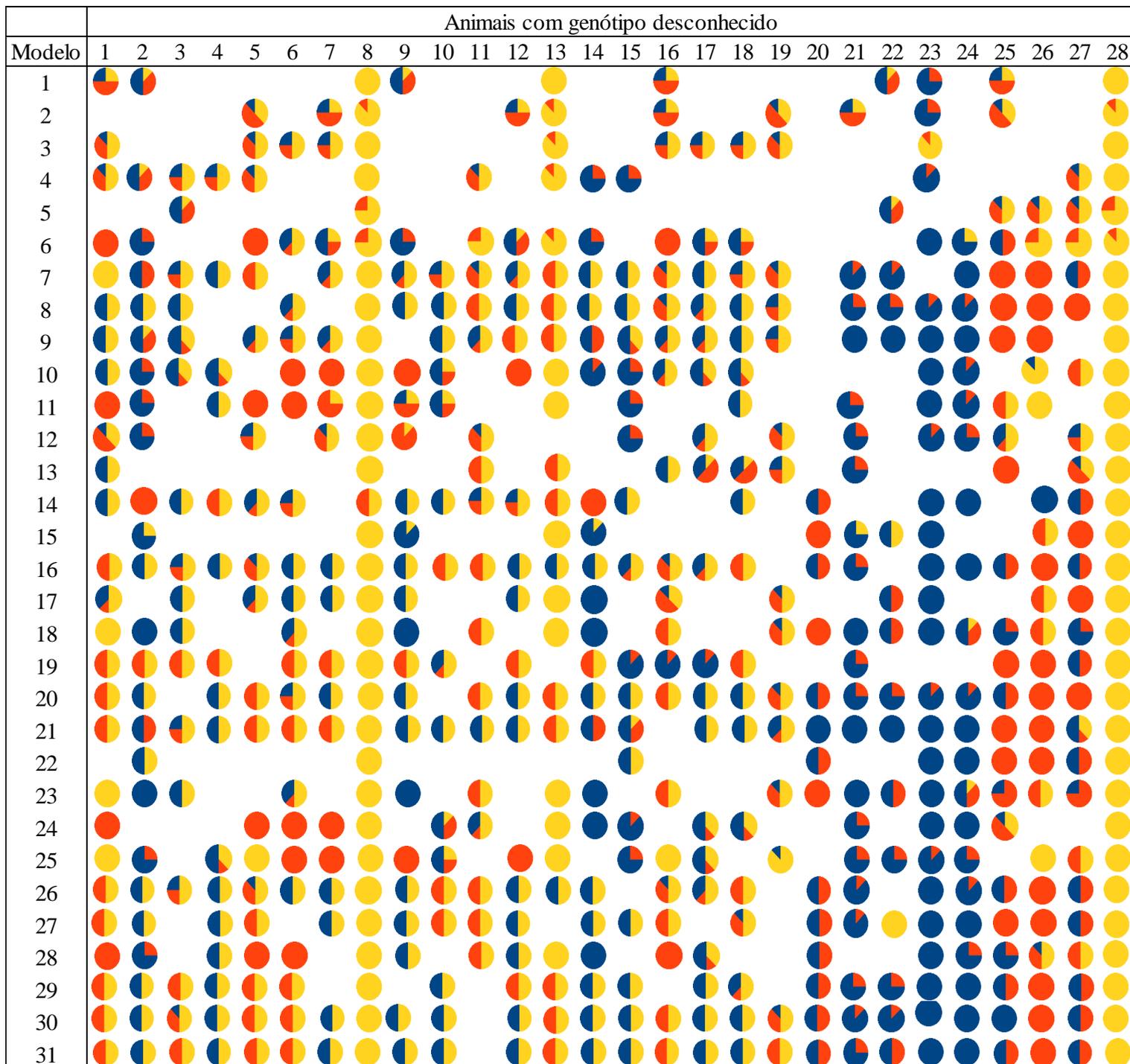


Figura 1 – Classificação dos animais com composição genética desconhecida.

4 DISCUSSÃO

Os resultados obtidos demonstram que, embora todos os coeficientes de correlação obtidos entre o genótipo imputado e o verdadeiro no processo leave-one-out sejam significativos em todas as metodologias (uni e multicaracter) o modelo unicaracter apresentou baixa eficácia. Em síntese, o peso corporal e das medidas morfológicas não devem ser utilizadas isoladamente imputar a composição genética em Serrasalmídeos.

Com acréscimo de variáveis, ou seja, passagem do modelo uni para bicaracter o coeficiente de correlação do genótipo imputado e do verdadeiro aumentaram consideravelmente para maioria das combinações de variáveis. Como por exemplo: CP-AD ($r = 0,68$), CP-CC ($r = 0,67$), CC-LD ($r = 0,65$), PESO-CP ($0,65$), AD-LD ($r = 0,60$), CC e AD ($0,57$), PESO-LD ($r = 0,55$) e PESO-CC ($r = 0,48$). Os modelos utilizando três características também apresentaram aumento significativo, em que oito modelos apresentaram correlação moderada variando de 0,49 a 0,69 e dois com correlação forte PESO-CP-CC ($r = 0,79$) e CP-CC-AD ($r = 0,79$). Todas as correlações dos modelos tetra e pentacaracter obtiveram alta correlação entre o genótipo imputado e o verdadeiro.

Esse aumento do coeficiente de correlação do genótipo verdadeiro e o imputado com a utilização dos modelos multivariados pode ser explicado por haver correlação genética alta e positiva entre características morfométricas e o peso corporal dos animais. Em suma, a utilização de mais características no modelo explicam melhor a composição genética desses indivíduos. A análise multivariada é sempre superior ou igual a univariada, sendo igual quando as variáveis utilizadas não são correlacionadas e melhor quando a quantidade de variáveis é correlacionada (Ferreira, 2008).

Trabalhos avaliando a correlação genética para estas espécies são escassos na literatura científica, sendo observado que estas características apresentam correlações fenotípicas significativas positivas variando de moderada a alta em Serrasalmídeos (Ribeiro et al., 2019). Porém, em outras espécies frequentemente estudadas como a tilápia é possível verificar que as correlações genéticas entre peso corporal e características morfométricas são positivas e fortes independentemente da idade do animal (Fernandes, 2014). Em outro trabalho com tilápias, verificou-se que, estas variáveis também apresentam alta correlação genética, variando de (0,76 a 0,92), e entre (0,89 a 0,99) para o peso de filé e as medidas morfométricas (Velasco et al., 1995). Medidas morfométricas

como altura, comprimento e largura são variáveis longitudinais, simples de serem mensuradas e possuem herdabilidades que vão de moderadas a altas (Rutten et al., 2004).

Os resultados obtidos neste trabalho, mostram que nos peixes redondos, assim como nas tilápias, muitos dos genes atilado pela variação do peso também são responsáveis pelas variações das medidas morfométricas. Estudos mostraram que geralmente as correlações fenotípicas e genéticas apresentam o mesmo sinal (positivo ou negativo), podendo diferir em magnitude, por causa efeitos ambientais que afetam os caracteres por meio de diferentes mecanismos fisiológicos. Indicando que estas variáveis podem ser utilizadas em estudos futuros de melhoramento genético, uma vez que, possibilitam a seleção indireta do caráter desejado. Apresentando relevância maior quando os mesmos possuem herança complexa e estão correlacionados a um ou mais caracteres identificáveis, podendo resultar em maior eficiência na seleção de características que se pretende melhorar (Filho, 1999; Falconer, 1987).

A correlação de Pearson usando o teste de Mantel entre genótipo verdadeiro e genótipo imputado, demonstrou que dentro de cada metodologia, ou seja, nos modelos uni e multivariados, as características de maior e menor correlação foram o CC e LD respectivamente. E que o uso destas com outras características podem influenciar de forma direta no resultado da análise, uma vez que as médias das correlações na validação são maiores quando contém a medida morfométrica CC e o oposto ocorre quando se utiliza o LD. Isso pode ser explicado pelo fato do CC apresentar maior herdabilidade e LD menor, conforme observado em cruzamento dialélico em espécies de Serralmídeos por meio da capacidade geral de combinação que representa a parte aditiva da variância genética (Costa et al., 2018). Estudos com tilápias relatam que estas características, ou seja, peso corporal e medidas morfométricas apresentam alta herdabilidade e que estas são altamente correlacionadas (0,94 a 0,99) (Nguyen et al., 2007).

Modelos multivariados podem apresentar limitações de uso quando se aumenta a quantidade de variáveis, elevando a quantidade de parâmetros a serem estimados o que pode resultar em uma não convergência, ou mesmo demorar para convergir. Além disso, os esforços computacionais em modelos multivariados são maiores. No presente trabalho, a convergência não foi problema e nem o tempo para convergir os parâmetros.

O presente trabalho demonstra que estas características podem ser utilizadas no modelo com o objetivo de aumentar de forma considerável a acurácia. Validando assim,

o uso do modelo de mistura misto multivariado por meio das características fenotípicas para complementar o uso de marcadores moleculares para predição dos genótipos de Serrasalmídeos, como já observado por Costa et al. (2017) em modelos univariados.

O teste de *Mantel* demonstrou que as correlações encontradas entre as imputações da composição genética dos animais com origem desconhecida apresentaram pouca coerência com o aumento de características utilizadas no modelo. Demonstrando assim que, é importante definir a metodologia empregada no estudo e determinar a quantidade e natureza das características que devem ser utilizadas no modelo.

5 Conclusão

Em geral os resultados demonstraram claramente que com o acréscimo de informações morfométricas e o peso dos indivíduos através da análise multicaráter, pode melhorar de forma gradativa a predição da composição genética em Serrasalmídeos. Validando assim, o uso do modelo de mistura misto multivariado por meio das características fenotípicas para complementar o uso de marcadores moleculares para predição dos genótipos de Serrasalmídeos.

6 REFERÊNCIAL TEÓRICO

- COSTA, A. C.; BOTELHO, H. A.; GOMES, R. C. D. S.; DE SOUSA CAMPOS, S. A., REIS NETO, R. V., BALESTRE, LIMA, M. General and specific combining ability in Serrasalminidae. **Aquaculture Research**, v. 50, n. 3, p. 717-724, 2018.
- COSTA, A. C.; BALESTRE, M.; BOTELHO, H. A.; FREITAS, R. T. F. D.; GOMES, R. C. S.; CAMPOS, S. A. S.; FORESTI, F. P.; HASHIMOTO, D. T.; MARTINS, D. G.; PRADO, F. D.; MENDONÇA, M. A. C. Imputation of genetic composition for missing pedigree data in Serrasalminidae using morphometric data. **Scientia Agricola**, v. 74, n. 6 p. 443-449, 2017.
- FALCONER, D. S. **Introdução à genética quantitativa**. Viçosa: UFV, p. 279, 1987.
- FERNANDES, A. F. A. **Parâmetros genéticos para peso corporal, rendimento de carcaça e características morfométricas em tilápias do Nilo (*Oreochromis niloticus*) em diferentes idades padrão**. Universidade Federal de Minas Gerais. Tese de Mestrado. Belo Horizonte, p. 44, 2014.
- FILHO, D. F.; DE ANDRADE, J. S.; CLEMENT, C. R.; MACHADO, F. M.; NODA, H. Correlações fenotípicas, genéticas e ambientais entre descritores morfológicos e químicos em frutos de cubiu (*Solanum sessiliflorum* Dunal) da amazônia. **Acta Amazônica**, v. 29, n. 4, p. 503-511, 1999.
- FURTADO, F. D. **Estatística multivariada**. Lavras: Editora UFLA, v. 1 p. 29, 2008.
- GOMES, F. S. H.; BARROS, C.; SAMPAIO, D.; HASHIMOTO, D. T.; PORTO-FORESTI, F.; SAMPAIO, I. Innovative molecular approach to the identification of *Colossoma macropomum* and its hybrids. **Anais da Academia Brasileira de Ciências**, v. 84, n. 2, p. 517-26, 2012.
- GRIFFING, B. A. Concept of general and specific combining ability in relation to diallel crossing systems. **Australian Journal of Biological Science**, v. 9, n. 4, p. 463-493, 1956.
- HASHIMOTO, D. T.; SENHORINI, J. A.; FORESTI, F., MARTINEZ, P.; PORTO-FORESTI, F. Genetic Identification of F1 and Post-F1 Serrasalminid Juvenile Hybrids in Brazilian Aquaculture, **Plos One**, v. 9, n. 3, p. e89902, 2014.

- HASHIMOTO, D. T.; MENDONÇA, F. F.; SENHORINI, J. A.; DE OLIVEIRA, C.; FORESTI, F.; PORTO-FORESTI, F. Molecular diagnostic methods for identifying Serrasalmid fish (Pacu, Pirapitinga, and Tambaqui) and their hybrids in the Brazilian aquaculture industry. **Aquaculture**, v.321, n. 1-2, p. 49-53, 2011.
- JORGE, P. H. **Sequenciamento do transcriptoma e caracterização de microssatélites na pirapitinga *Piaractus brachypomus* para análises de variabilidade genética.** Dissertação de Mestrado em Ciências Biológicas, **Universidade Estadual Paulista**, Botucatu, p. 76, 2016.
- KHATKAR, M. S.; MOSER, G.; HAYES, B. J.; RAADSMA, W. Strategies and utility of imputed SNP genotypes for genomic analysis in dairy cattle, **BMC Genomics**, v.13, n. 1, p.538, 2012.
- LUCA, G. D.; ZUCCOLOTTO, P. Finite and infinite mixtures for financial durations. **Metron**. v. 61, n. 3, p. 431-455, 2003.
- MOURAD, N. M. N.; COSTA, A. C.; FREITAS, R. T. F.; SERAFINI, M. A.; REIS NETO, R. V.; FELIZARDO V. O. Weight and morphometric growth of Pacu (*Piaractus mesopotamicus*), Tambaqui (*Colossoma macropomum*) and their hybrids from spring to winter. **Pesquisa Veterinária Brasileira**, v. 38, n. 3 p. 544-550, 2018.
- NGUYEN, N. H.; KHAW, H. L.; PONZONI, R. W.; HAMZAH, A.; KAMARUZZAMAN, K. Can sexual dimorphism and body shape be altered in Nile tilapia (*Oreochromis niloticus*) by genetic means. **Aquaculture**, v.272, n. S1, p. S38-S46, 2007.
- PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block size are unequal. **Biometrika**, v. 58, n. 3, p. 545-554, 1971.
- RIBEIRO, F. M.; LIMA, M.; COSTA, P. A. T. D.; PEREIRA, D. M.; CARVALHO, T. A., DE SOUZA, T. V., COSTA, A. C. Associations between morphometric variables and weight and yields carcass in Pirapitinga *Piaractus brachypomus*. **Aquaculture Research**, v. 50, n. 7, p. 2004-2011, 2019.
- RUTTEN, M. J.; BOVENHUI, 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, 2004.

VELASCO RR, JANAGAP CC, DE VERA MP, AFAN LB, REYES RA, EKNATH AE.
Genetic improvement of farmed tilapias: estimation of heritability of body and carcass traits of Nile tilapia (*Oreochromis niloticus*). **Aquaculture**, v.137, n. 1-4 p.280-281. 1995.

CAPÍTULO III – ARTIGO II

ANÁLISE DISCRIMINANTE EM SERRASALMIDEOS

RESUMO

A produção de espécies nativas como o pacu (*Piaractus mesopotamicus*) a pirapitinga (*Piaractus brachypomus*) e o tambaqui (*Colossoma macropomum*) e seus híbridos tem crescido expressivamente nos últimos anos. A identificação correta do grupo genético que o animal pertence é fundamental para programas de conservação de recursos genéticos em Serrasalmideos, dessa forma o objetivo com o presente estudo foi verificar se a função de discriminação consegue discriminar pacu, tambaqui, tambacu e paqui. Foram utilizados neste estudo 60 exemplares de pacu, 79 tambaqui, 43 tambacu e 43 paqui com 496 dias de idade (contemporâneos). Os peixes foram anestesiados e submetidos à pesagem e avaliação morfométrica. Para análise foi considerada a função discriminante linear de Fisher utilizando o programa computacional R 3.3.3. Os resultados demonstraram que as variáveis mais importantes respectivamente, na primeira função, são RL2CP, RCCAC e RA2CP; na segunda função, RLDCP, RADCP, RLDA2 e RCCAD; na terceira função, RLDCP, LD e RCCA2. A classificação dos grupos genéticos em função da análise discriminante utilizando as medidas lineares, permitiram classificar de forma correta 199 animais dos 225 no total, correspondendo 88,4% de acerto. Os animais puros (Tambaqui e Pacu), apresentaram menor erro de classificação que os híbridos, isso pode ser explicado pois esses indivíduos apresentam alelos dos dois parentais, e desta maneira possuem maior variabilidade genética. Diante do exposto, mesmo indivíduos híbridos possuindo maior erro de classificação, dados morfométricos para discriminação de pacu, tambaqui, tambacu e paqui podem ser utilizados em análises discriminantes para a classificação dos mesmos, uma vez que a taxa de erro foi modesta.

Palavras-chave: Aquicultura. Análise Multivariada. Função discriminante. Peixes redondos.

ABSTRACT

The native species production such as pacu (*Piaractus mesopotamicus*) pirapitinga (*Piaractus brachypomus*) and tambaqui (*Colossoma macropomum*) and their hybrids has grown significantly in recent years. The correct genetic group identification, which the animal belongs, is essential for conservation programs of genetic resources in Serrasalmideos, so the objective of the present study was to verify if the discrimination function can discriminate pacu, tambaqui, tambacu and paqui. There were used 60 specimens of pacu, 79 tambaqui, 43 tambacu and 43 paqui with 496 days of age (contemporaries) in this study. The fish were desensitized through the marrow section, and then they were subjected to weighing and morphometric evaluation. For the analysis, the Fischer linear discriminant function was considered using the software R 3.3.3. The results showed that the most important variables, respectively, in the first function are RL2CP, RCCA2 and RA2CP; in the second function, RLDCP, RADCP, RLDA2 and RCCAD; in the third function, RLDCP, LD and RCCA2. The classification of the genetic groups according to the discriminant analysis using the linear measures allowed to correctly classify 199 animals of the 225 in total, corresponding to 88.4% of correctness. The pure animals (Tambaqui and Pacu) presented lower classification error than the hybrids, this can be explained since these individuals present alleles of the two parental, and in this way have greater genetic variability. In view of the above, even hybrid specimens with the highest classification error, morphometric data for pacu, tambaqui, tambacu and paqui discrimination can be used in discriminant analyzes to classify them, since the error rate was modest.

Keywords: Aquaculture. Multivariate analysis. Discriminant function. Round fish.

1 INTRODUÇÃO

A produção de alimentos se tornou preocupação mundial, isso devido a quantidade de recursos naturais que é limitada, e o crescimento populacional segue aumentando exponencialmente. A aquicultura é a cadeia produtiva de alimentos que mais cresce no planeta e desta forma, está entre as atividades mais importantes no ponto de vista econômico para a sociedade (FAO, 2018).

Os peixes nativos como pacu (*Piaractus mesopotamicus*) a pirapitinga (*Piaractus brachypomus*) e o tambaqui (*Colossoma macropomum*) e seus híbridos apresentam distribuição natural em várias bacias hidrográficas como nas bacias do rio Paraná, Paraguai e Uruguai rio Solimões-Amazonas e também nos principais rios da bacia do rio Orinoco na Venezuela (Inoue & Boijink, 2011; Petreire, 1989). Entretanto, a produção de espécies nativas não é restrita as regiões em que são endêmicas, em que foram introduzidas para uso na aquicultura em vários países da América do Sul (Colômbia, Peru e Venezuela), e também em países asiáticos (China, Mianmar, Tailândia e Vietnã), demonstrando assim grande importância ambiental e econômica no mundo (Flores nava, 2007; Honglang 2007; Jorge, 2016).

A produção de híbridos interespecíficos entre essas espécies tem sido muito empregada com objetivo de obter heterose e complementariedade para as características de importância econômica na aquicultura (Mourad et al., 2018). Em contraste, dependendo de fatores como ambiente, interação alélica, alimentação, esses indivíduos conseguem apresentar muita similaridade com seus parentais em termos morfológicos, e podem ocasionar erros na identificação desses animais.

A identificação correta do grupo genético que o animal pertence é fundamental para programas de conservação de recursos genéticos em Serrasalmideos, que podem ser utilizadas para o monitoramento de estoques naturais, programas de melhoramentos genéticos visando cruzamento ou seleção, e também em piscigranjas que produzem alevinos, fornecendo a visão geral útil da população de cada espécie, para entender o comportamento e interação interespecíficas, que podem ser usadas na elaboração de estratégias de conservação da biodiversidade em ambiente ecológico natural ou em cativeiro (Hasija et.al, 2017).

Marcadores moleculares têm sido utilizados frequentemente na caracterização genética de um indivíduo até a de grupos de organismos (Gasques et al, 2013), porém

essa metodologia tem custo relativamente elevado e não estão prontamente disponíveis no campo.

Características morfológicas podem fornecer informações importantes sobre as espécies, uma vez que podem descrever a origem dessas variações, sejam elas influenciadas por fatores genéticos, ambientais e interação com o mesmo. Visto isso, a caracterização de grupos/espécies pode ser deduzida a partir de dados utilizando essas informações, porém o principal problema na sua aplicação é determinar quais características ou variáveis podem ser utilizadas para determinar de forma plausível a variável resposta desejada.

Recentemente a análise discriminante vem sendo muito utilizada com eficácia em diversas áreas de aplicação com o objetivo de discriminar populações e/ou classificar objetos em populações pré-definidas. A função discriminante linear de Fisher é uma combinação linear das variáveis observadas que apresenta maior poder de discriminação entre os grupos e com a propriedade de minimizar as probabilidades de má classificação, quando as populações são normalmente distribuídas com parâmetros de média e variância conhecidas. Sendo assim, o trabalho foi realizado com o objetivo de classificar em grupos as espécies nativas de maior importância na produção brasileira, sendo elas o pacu, pirapitinga e seus híbridos através da análise discriminante.

2 MATERIAL E MÉTODOS

2.1 Obtenção do material biológico

No presente estudo, foram utilizados 60 exemplares de pacu, 79 tambaqui, 43 tambacu e 43 paqui com 496 dias de idade (contemporâneos) do Setor de Piscicultura da Escola Agrotécnica Federal de Colatina – ES (EAFCOL), situada a 17 km de Colatina. Os peixes foram insensibilizados anestesiados e submetidos a avaliação morfométrica.

2.2 Avaliação morfométrica

A avaliação morfométrica foi utilizada seguindo os seguintes parâmetros morfométricos: Comprimento padrão (CP), compreendido entre a extremidade anterior da cabeça e o menor perímetro do pedúnculo (inserção da nadadeira caudal); comprimento de cabeça (CC), compreendido entre a extremidade anterior da cabeça e a borda caudal do opérculo; altura do corpo medida à frente do 1º raio das nadadeiras dorsal (AD); largura do corpo tomada na região do 1º raio das nadadeiras dorsal (LD);

O comprimento padrão foi medido através do ictiômetro e as demais medidas foram realizadas com auxílio do paquímetro graduado em milímetros (mm). Como complementação, foram calculadas as seguintes razões morfométricas:

CC/CP = comprimento da cabeça / comprimento padrão;

AD/CP = altura do corpo na linha do 1º raio da nadadeira dorsal / comprimento padrão;

LD/CP = largura do corpo na linha do 1º raio da nadadeira dorsal / comprimento padrão;

CC/AD = comprimento da cabeça / altura do corpo na linha do 1º raio da nadadeira dorsal;

LD/CC = largura do corpo na linha do 1º raio da nadadeira dorsal / comprimento da cabeça;

LD/AD = largura do corpo na linha do 1º raio da nadadeira dorsal / altura do corpo na linha do 1º raio da nadadeira dorsal.

2.3 Análise estatística

A análise estatística foi realizada para classificar em grupos as espécies nativas, com base em todos os fenótipos medidos (peso corporal, avaliação morfométrica), utilizando o programa computacional R versão 3.3.3. Para o estudo foi considerada a função discriminante linear de Fisher, que é basicamente a combinação linear de características originais a qual se caracteriza por produzir separação máxima entre duas populações.

Considerando que μ_i e Σ são parâmetros conhecidos e respectivamente, os vetores de médias e a matriz de covariâncias comum das populações π_i . Demonstra-se que a função linear do vetor aleatório X que produz separação máxima entre duas populações é dada por:

$$D(X) = L' \cdot X = [\mu_1 - \mu_2]' \cdot \Sigma^{-1} \cdot X$$

em que,

$$X = [X_1 \ X_2 \ \dots \ X_p] \text{ e } \pi = [\pi_1, \pi_2]$$

L = vetor discriminante;

X = vetor aleatório de características das populações;

μ = vetor de médias p-variado;

Σ = matriz comum de covariâncias das populações π_1 e π_2 ;

O valor da função discriminante de Fisher para uma dada observação x_0 é:

$$D(x_0) = [\mu_1 - \mu_2]' \cdot \Sigma^{-1} \cdot x_0$$

O ponto médio entre as duas médias populacionais univariadas μ_1 e μ_2 é:

$$m = \frac{1}{2} [\mu_1 - \mu_2]' \cdot \Sigma^{-1} \cdot [\mu_1 - \mu_2], \text{ ou seja}$$

$$m = \frac{1}{2} [D(\mu_1) + D(\mu_2)]$$

A regra de classificação baseada na função discriminante de Fisher é:

$\text{Alocar } x_0 \text{ em } r_1 \text{ se } D(x_0) = [\mu_1 - \mu_2]' \cdot \Sigma^{-1} \cdot x_0 \geq m$
$\text{Alocar } x_0 \text{ em } r_2 \text{ se } D(x_0) = [\mu_1 - \mu_2]' \cdot \Sigma^{-1} \cdot x_0 < m$

Assumindo-se que as populações π_1 e π_2 têm a mesma matriz de covariâncias Σ pode-se então estimar uma matriz comum de covariâncias S_c :

$$S_c = \left[\frac{n_1 - 1}{(n_1 - 1) + (n_2 - 1)} \right] \cdot S_1 + \left[\frac{n_2 - 1}{(n_1 - 1) + (n_2 - 1)} \right] \cdot S_2$$

Em que,

S_c = estimativa da matriz comum de covariâncias Σ ;

n_1 = número de observações da população π_1 ;

n_2 = número de observações da população π_2 ;

S_1 = estimativa matriz de covariâncias da população π_1 ;

S_2 = estimativa matriz de covariâncias da população π_2 ;

A função discriminante linear amostral de Fisher é obtida substituindo-se os parâmetros μ_1 , μ_2 e Σ pelas respectivas quantidades amostrais \bar{x}_1, \bar{x}_2 e S_c :

$$D(x) = \hat{L}' \cdot x = [\bar{x}_1 - \bar{x}_2]' \cdot S_c^{-1} \cdot x$$

Em que,

$D(x)$ = função discriminante linear amostral de Fisher;

\hat{L} = estimativa do vetor discriminante;

x_1 = média amostral da população π_1 ;

x_2 = média amostral da população π_2 .

3 Resultados

Destarte, os coeficientes de discriminação linear para a primeira, segunda e terceira função discriminantes estão apresentados na Tabela 1. Cada coeficiente expressa a contribuição relativa da variável associada a cada função. Dessa forma, a análise da Tabela 1 demonstra que as variáveis mais importantes respectivamente, na primeira função, são RLD/CP, RCC/AD e RAD/CP; na segunda função, RLD/CP, RAD/CP, RLDAC e RCC/AD; na terceira função, RLD/CP, LD e RCC/AD.

Tabela 1 - Coeficientes de discriminantes lineares para os grupos genéticos, obtidos a partir das medidas lineares.

Variável	Função Discriminante		
	1	2	3
PESO	-0,0069936	0,00020298	-0,0125987
CP	1,7106187	4,46851447	-1,8276644
CC	-5,5970557	-14,99412	1,2475426
AD	0,6054052	2,53267187	0,14570542
LD	5,0536247	-8,914954	12,1148468
RCC/CP	-20,421294	1,67200975	-101,29219
RAD/CP	46,893915	181,802836	-183,88515
RLD/CP	201,98141	225,930817	567,415469
RCC/AD	72,066005	117,277047	6,40544517
RLD/AD	-61,768706	163,752735	-402,25118
RLD/CC	-74,042321	-94,588671	-33,081269

CP: comprimento padrão; CC: comprimento de cabeça; AC: altura corporal; LD: largura corporal; $R_{variavel^1/variavel^2}$: razões morfométricas.

Quanto ao número de funções discriminantes, pela análise da Tabela 2 verifica-se que o primeiro autovalor representa 83,2% da variância total das variáveis discriminantes, e dessa forma o mesmo será considerado para a classificação dos grupos genéticos, o autovalor, associado com a segunda função discriminante, representa 12,6% da variância residual incorporada ao conjunto de variáveis discriminantes, assim os primeiros

autovalores contribuem com 95,8% da variância total do sistema, por fim verifica-se que o autovalor, associado com a terceira função, representa 2,70% da variância total.

Tabela 2 – Porcentagem da variância total explicada pelos autovalores.

	Função Discriminante		
	1	2	3
Autovalor	0,8324	0,1252	0,0424
Perc. (%)	83,2	12,6	4,2

A classificação dos grupos genéticos em função da análise discriminante utilizando as medidas lineares permitiram classificá-los de forma correta 199 animais dos 225 no total, correspondendo 94,3% de acerto. A análise discriminante conseguiu agrupar 211 indivíduos, em que todos os 55 indivíduos declarados como pacu foram classificados corretamente, no tambaqui apenas um dos 70 indivíduos foi classificado de forma incorreta, representando 98,6% de precisão na discriminação correta do grupo. A partir da classificação do tambacu e paqui (híbridos), os erros de agrupamentos foram ampliados, no grupo genético do tambacu foram classificados oito indivíduos erroneamente como tambaqui e paqui, e no grupo genético do paqui três indivíduos foram classificados equivocadamente como tambaqui e tambacu, representando um percentual de acerto de 81,4% e 93% respectivamente.

Tabela 3 - Classificação dos grupos genéticos, em função das medidas lineares gerados com dados morfométricos e peso dos animais.

Grupos genéticos observados	Grupos genéticos estimados			
	Pacu	Tambaqui	Tambacu	Paqui
Pacu	55 (100%)	0	0	0
Tambaqui	0	69 (98,6%)	2 (4,6%)	1 (2,4%)
Tambacu	0	1 (1,4%)	35 (81,4%)	2 (4,6%)
Paqui	0	0	6 (14%)	40 (93%)

Percentual de acertos = 94,3%

4 DISCUSSÃO

Em suma, verificou-se que os coeficientes de discriminação linear evidenciaram que algumas variáveis conseguem obter uma função discriminante adequada sobre os grupos genéticos já expostos, como por exemplo RLDCP, RCCAD e RADCP, essa discriminação apresenta uma certa concordância entre os coeficientes, podendo indicar

que existe variação da largura e altura dos animais sobre o comprimento padrão, e também comprimento da cabeça nas diferentes espécies. Dessa forma a análise consegue agrupar esses indivíduos por intermédio dessas diferenças e por consequência ser efetiva na discriminação das espécies.

A função discriminante pode ser utilizada para discriminar essas espécies, pois, explicou cerca de 90% da composição genética legítima dos indivíduos usados no estudo. Segundo Lv et.al., (2017), modelos discriminantes combinados com espectroscopia de refletância no infravermelho, podem efetivamente classificar espécies de peixes de água doce. Já Hasija et.al (2017) observaram que a partir de um método computacional automatizado de análise de imagens de peixes através de gráficos é possível classificar espécies de peixes com grande precisão, cerca de 91,66%. De forma semelhante Silva et.al (2009) demonstraram que é possível diferenciar uma espécie proveniente de diferentes bacias hidrográficas a partir de dados morfométricos utilizando análise discriminante. Já Pinto et. al, (2005), verificaram ainda que em outras espécies é possível classificar animais de diferentes faixas etárias e sexo com precisão 92,5% de acertos a partir de medidas morfométricas utilizando a função discriminante.

Esse tipo de estudo é muito valioso do ponto de vista econômico e ecológico, uma vez que é possível discriminar espécies de forma viável economicamente comparado a métodos moleculares por exemplo e desta maneira realizar monitoramento estoques naturais e em cativeiro com o objetivo criar estratégias de preservação e melhoramento nesse grupo pouco influenciado tecnologicamente.

Foi evidenciado que os animais puros (Tambaqui e Pacu), apresentaram menor erro de classificação que os híbridos, isso pode ser explicado pois esses indivíduos apresentam alelos dos dois parentais, e desta maneira possuem maior variabilidade genética.

5 Conclusão

Em geral, mesmo espécimes híbridos possuindo maior erro de classificação, dados morfométricos para discriminação de pacu, tambaqui, tambacu e paqui podem ser utilizados em análises discriminantes para a classificação dos mesmos, uma vez que a taxa de erro foi modesta.

6 REFERÊNCIAL TEÓRICO

- FAO. **The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable development goals**. Rome, 2018.
- FLORES, A. N. Aquaculture seed resources in Latin America: a regional synthesis. **Assessment of freshwater fish seed resources for sustainable aquaculture**. FAO Fisheries Technical Paper, n. 501, p. 91, 2007.
- GASQUES, L. S., BELONI, K. P., & DE OLIVEIRA, J. R. Os marcadores moleculares em peixes e suas aplicações em publicações da base de dados do Scielo. **Arquivos de Ciências Veterinárias e Zoologia da UNIPAR**, v. 16, n. 1, 2014.
- HASIJA, S., BURAGOHAİN, M. J., & INDU, S. Fish species classification using graph embedding discriminant analysis. In **2017 International Conference on Machine Vision and Information Technology (CMVIT)**. p. 81-86, 2017.
- HONGLANG, H. **Freshwater fish seed resources in China**. In *Assessment of Freshwater Fish Seed Resources for Sustainable Aquaculture* (Ed. M. G. BondadReantaso), p. 185–199, 2007.
- INOUE, L. A. K. A.; BOIJINK, C. L. Manaus a capital do tambaqui. **Embrapa Amazonia**, 2011. Artigo em Hypertexto. Disponível em: <<http://www.infobibos.com/Artigos/2011>>.
- JORGE, P. H. **Sequenciamento do transcriptoma e caracterização de microssatélites na pirapitinga *Piaractus brachypomus* para análises de variabilidade genética**. Dissertação de Mestrado em Ciências Biológicas. **Universidade Estadual Paulista**, Botucatu, p. 76, 2016.
- LV, H. XU, W. YOU, J.; XIONG, S. Classification of freshwater fish species by linear discriminant analysis based on near infrared reflectance spectroscopy. **Journal of Near Infrared Spectroscopy**. v. 25, n. 1, p. 54-62, 2017.
- MOURAD, N. M. N. ; COSTA, A. C. ; FREITAS, R. T. F. ; SERAFINI, M. A. ; REIS NETO, R. V. ; Felizardo V. O. . Weight and morphometric growth of Pacu (*Piaractus mesopotamicus*), Tambaqui (*Colossoma macropomum*) and their hybrids from spring to winter. **Pesquisa veterinária brasileira (online)**, v. 38, n. 3, p. 544-550, 2018.
- PETRERE, M. J. R. River Fisheries in Brazil: a review. **Regulated Rivers: Research and Management**, v. 4, n. 1, p. 1-16, 1989.

- PEDRAZZANI, A.S. CARNEIRO, P. C. F., KIRSCHNIK, P. G.; MOLENTO, C. F. M. Impacto negativo de secção de medula e termonarrose no bem-estar e na qualidade da carne da tilápia-do-Nilo. **Revista Brasileira de Saúde e Produção Animal**, v. 10, n. 1, p.188-197, 2009.
- PINTO, L. F. B.; ALMEIDA, F. D., QUIRINO, C. R., CABRAL, G. C., AZEVEDO, P. D., SANTOS, E. M. Análise multivariada das medidas morfométricas de potros da raça Mangalarga Marchador: análise discriminante. **Revista Brasileira de Zootecnia**. v. 34, n. 2, p. 600-612, 2005.
- SILVA, E. L., CENTOFANTE, L., MIYAZAWA, C. S. Análise morfométrica em *Thoracocharax stellatus* (Kner, 1858)(Characiformes, Gasteropelecidae) proveniente de diferentes bacias hidrográficas Sul-americanas. **Biota Neotropica**, v. 9, n. 2 p. 71-76, 2009.