



**UNIVERSIDADE FEDERAL DA GRANDE DOURADOS
FACULDADE DE CIÊNCIAS AGRÁRIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOTECNIA**

**ESTRUTURA POPULACIONAL DE BUBALINOS LEITEIROS UTILIZANDO
DADOS GENÔMICOS**

KARINA ROSA DA SILVEIRA

Dourados – MS
Abril de 2019



**UNIVERSIDADE FEDERAL DA GRANDE DOURADOS
FACULDADE DE CIÊNCIAS AGRÁRIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ZOOTECNIA**

**ESTRUTURA POPULACIONAL DE BUBALINOS LEITEIROS UTILIZANDO
DADOS GENÔMICOS**

KARINA ROSA DA SILVEIRA

Zootecnista

Orientador: Prof. Dr. Leonardo de Oliveira Seno

Coorientador: Dr. Rusbel Raul Aspilcueta-Borquis

Dissertação apresentada ao Programa de Pós-Graduação em Zootecnia da Universidade Federal da Grande Dourados, como parte das exigências para obtenção do título de Mestre em Zootecnia.
Área de Concentração: Produção Animal.

Dourados – MS

Abril de 2019

Dados Internacionais de Catalogação na Publicação (CIP).

S587e Silveira, Karina Rosa Da
ESTRUTURA POPULACIONAL DE BUBALINOS LEITEIROS UTILIZANDO DADOS
GENÔMICOS [recurso eletrônico] / Karina Rosa Da Silveira. -- 2019.
Arquivo em formato pdf.

Orientador: Leonardo de Oliveira Seno.
Coorientador: Rusbel Raul Aspilcueta-Borquis.
Dissertação (Mestrado em Zootecnia)-Universidade Federal da Grande Dourados,
2019.
Disponível no Repositório Institucional da UFGD em:
<https://portal.ufgd.edu.br/setor/biblioteca/repositorio>

1. análise de componentes principais. 2. assinatura de seleção. 3. Bubalus bubalis. 4.
genética de populações. 5. SNP. I. Seno, Leonardo De Oliveira. II. Aspilcueta-borquis,
Rusbel Raul. III. Título.

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

©Direitos reservados. Permitido a reprodução parcial desde que citada a fonte.

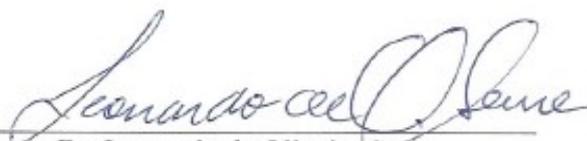
**ESTRUTURA POPULACIONAL DE BUBALINOS LEITEIROS UTILIZANDO
DADOS GENÔMICOS**

por

KARINA ROSA DA SILVEIRA

Dissertação apresentada como parte dos requisitos exigidos para obtenção do título
de MESTRE EM ZOOTECNIA

Aprovado em: 03/05/2019



Dr. Leonardo de Oliveira Seno
Orientador – UFGD/FCA



Dr. André Luiz Julien Ferraz
UEMS/Unidade de Aquidauana



Dr. Rusbel Raúl Aspilcueta Borquis
Autônomo

Agradecimentos

Pai e Mãe, obrigada por cada ensinamento e orientação que vocês me deram sempre. Por todo amor, alegria e união da nossa família. Obrigada por serem meus exemplos de amor e honestidade.

Obrigada minha irmã Patrícia por todo zelo e proteção que tem comigo, pela profissional que você é (a quem eu tenho muito orgulho de me espelhar) e por toda cumplicidade de sempre. Obrigada meu cunhado Saulo, por ser um irmão da vida toda, sempre disposto a caçoar de mim (kkkkk), mas antes de tudo a oferecer sempre sua sinceridade e amizade.

Marcos, meu noivo e melhor amigo, obrigada pela nossa relação de sinceridade e cumplicidade. Por apoiar meus sonhos, mesmo quando isso implica em enfrentar sacrifícios pessoais para me ajudar a alcançá-los.

Minha querida Jessica...minha gema, não me esqueço do primeiro dia em que nos conhecemos, do seu abraço acolhedor e sincero. Desde então não nos desgradamos mais, obrigada por todo seu apoio e amizade.

Professor Leonardo, obrigada pela chance de ser sua orientada sem mesmo nos conhecermos antes. O senhor é para mim um exemplo de pessoa justa e honesta, obrigada por toda orientação profissional, ensinamentos pessoais, pela dedicação, cuidado e amizade. Aprendi muitas coisas com o senhor.

Professor Raul, foi uma honra poder ser sua orientada. O senhor é um grande profissional e me ensinou sempre com muita paciência e dedicação, muito obrigada. Obrigada por toda amizade, cuidado e ensinamentos pessoais, o senhor é uma pessoa incrível e que tem toda a minha admiração.

Dividir uma sala com professor Leonardo, professor Raul, Jessica e Rafael foi uma grande alegria e aprendizado para mim. Obrigada por tudo pessoal!

Minha grande companheira Renata (mozão), obrigada pela parceria e amizade que formamos. Obrigada pela harmonia da nossa casa e todas as gargalhadas que demos juntas (e ainda damos rs).

Cris, você foi uma providência de Deus na minha vida, obrigada por todo cuidado sempre tão amoroso comigo, por me tratar muitas vezes como mais que uma amiga, como uma filha. Obrigada por toda a confiança depositada em mim e todos bons momentos.

Aos amigos de sempre Layla, Bárbara, Raphaela, Daniel e Letícia... obrigada pelos bons conselhos e amizade.

Ao meu querido amigo Mario, obrigada por todas as nossas conversas sobre a vida e por ter se tornado um verdadeiro amigo.

Aos queridos amigos Bianca, Elias e Josi... a cada um de vocês meu obrigada pela amizade e bons momentos juntos.

Obrigada Nina, Kika, Vicky, Pipoca por serem nossas grandes companheirinhas e transbordarem nosso coração de amor e alegria.

Obrigada aos meus antigos, mas sempre mestres Leila, Gabriel e Fausto.

Não poderia deixar de agradecer ao professor Humberto Tonhati por sempre estar disposto a ajudar, ao Jorge Camargo e todos da Nelore Birigui, aprendi muito nas visitas à fazenda. E aos professores Gregório, André e Alexeia pelas contribuições valiosas em meu trabalho.

Obrigada a todos os colegas da pós-graduação, desejo muito sucesso a todos vocês. Obrigada ao Ronaldo por sempre nos ajudar com muita prontidão na secretaria. Obrigada a cada professor da pós-graduação que contribuiu com a minha formação.

Meu agradecimento à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo recebimento da bolsa durante o mestrado.

Obrigada cidade de Dourados e UFGD por me acolher e ser um lar no qual vou sentir muita saudade.

Por último e mais importante, agradeço ao meu Deus querido por colocar cada uma dessas pessoas no meu caminho. Obrigada Senhor por permitir cada alegria, mas também cada tropeço... pois sempre depois da adversidade o Senhor me ampara e me molda para ser uma pessoa melhor. Obrigada Senhor por toda proteção e amor.

"Combati o bom combate, terminei a corrida, guardei a fé"

2 Timóteo 4:7

Sumário

LISTA DE TABELAS	viii
LISTA DE FIGURAS	ix
LISTA DE ANEXOS	ix
LISTA DE ABREVIATURAS	x
CONSIDERAÇÕES INICIAIS	2
CAPÍTULO 1 - REVISÃO BIBLIOGRÁFICA	3
Bubalinocultura no Brasil	3
Avaliação da estrutura populacional.....	3
Medida de variabilidade genética	7
Assinatura de seleção.....	9
OBJETIVO	10
REFERÊNCIAS BIBLIOGRÁFICAS	11
CAPÍTULO 2 - ESTRUTURA POPULACIONAL DE BUBALINOS LEITEIROS UTILIZANDO DADOS GENÔMICOS	19
Resumo	19
Abstract.....	20
Introdução.....	21
Material e Métodos	22
Amostras e controle de qualidade dos marcadores.....	22
Análise de Componentes Principais (PCA)	23
Diversidade genética e análise de diferenciação.....	23
Assinatura de seleção.....	24
Resultados e Discussão.....	24
CONCLUSÃO	32
REFERÊNCIAS BIBLIOGRÁFICAS	33
CONSIDERAÇÕES FINAIS	36
ANEXO 1	37

LISTA DE TABELAS

Tabela 1 – Índices de diversidade genética. H_o = heterozigosidade observada; H_e = heterozigosidade esperada; F_{is} = índice de fixação intra-populacional (coeficiente de endogamia)

LISTA DE FIGURAS

Figura 1 – Representação dos primeiros 4 componentes principais obtidos com os dados genômicos das populações bubalinas. Valores em porcentagem indicam a parcela de variância explicada pelos respectivos componentes principais. PC = *principal component* (componente principal).

Figura 2 – Análise de componentes principais das populações de búfalos em estudo. Indivíduos do mesmo grupo populacional estão representados por pontos de mesma cor; subpopulação 1 (verde); subpopulação 2 (azul); subpopulação 3 (rosa).

Figura 3 – Manhattan *plot* da distribuição genômica dos valores de F_{st} .

LISTA DE ANEXOS

Anexo 1 – *Scree plot* dos 20 primeiros componentes principais e suas respectivas porcentagens de variância explicada.

LISTA DE ABREVIATURAS

CQ – Controle de qualidade

DL – Desequilíbrio de ligação

EHW – Equilíbrio de *Hardy–Weinberg*

Fis – Índice de fixação intra-populacional

Fst – Índice de fixação entre subpopulações

GWAS - *genome wide association studies* (estudo de associação genômica ampla)

He – Heterozigosidade esperada

Ho – Heterozigosidade observada

MAF – *Minor Allele Frequency* (Frequência do Alelo Menor)

PC – *Principal Component* (Componente Principal)

PCA – *Principal Component Analysis* (Análise de Componentes Principais)

SNP – *Single Nucleotide Polymorphism* (Polimorfismo de Nucleotídeo Único)

CONSIDERAÇÕES INICIAIS

A criação de búfalos no Brasil, inicialmente destinava-se para a ocupação de extensões territoriais em que era impraticável o manejo de bovinos, também chamados de “vazios pecuários”. Com o maior conhecimento do perfil produtivo da espécie, foi possível a expansão da cultura para diversas regiões do país (BERNARDES, 2007). Hoje, no Brasil, o principal foco da produção de leite é a fabricação de lácteos tais como a *mozzarella*, diferentemente de alguns outros lugares no mundo que têm a cultura como forma de subsistência. O leite de búfalas possui peculiaridades físico-químicas que enaltecem o seu potencial de produção e qualidade. Entre essas características, inerentes da espécie, estão o elevado teor de sólidos totais como gordura, proteína e minerais que permitem rendimento superior na fabricação de produtos lácteos quando comparado ao leite bovino (TONHATI, 2000)

Apesar de ter grande importância econômica e social no mundo, o búfalo doméstico não recebe grande atenção em termos de melhoramento genético e conservação (HUSSAIN et al., 2017). Desta maneira, considerar a variabilidade genética em animais domésticos é imprescindível, pois, auxilia a mudança genética dentro de uma população, sustentando a capacidade dessa população em responder à seleção para aumento da produtividade e de adaptar-se a mudanças de circunstâncias ambientais sendo estas não só as condições climáticas, como também as associadas às flutuações de mercados, práticas de manejo e desafios patológicos (BOETTCHER et al., 2010). Para tanto, certificar-se se as populações de interesse são homogêneas ou estruturadas, mensurar a diversidade genética dentro e entre os possíveis grupos e entender quais mecanismos biológicos estão agindo para a diferenciação destes grupos, pode auxiliar neste sentido.

O desenvolvimento de biotecnologias como os painéis de genotipagem contendo alta densidade de SNPs (do inglês *Single Nucleotide Polymorphism*), permitiu uma nova perspectiva a respeito do conhecimento das relações genéticas entre indivíduos e em uma maior escala, entre populações, antes possibilitada apenas subjetivamente através da observação fenotípica. Desta maneira, o uso de recursos da bioinformática e de métodos estatísticos, tornou-se essencial para a manipulação e tradução do grande volume de dados em informação interpretável.

Devido ao crescente interesse pela bubalinocultura no Brasil e também pela importância da conservação dos recursos genéticos animais, o objetivo deste trabalho foi estudar a estrutura da população bubalina leiteira no Brasil e identificar regiões

genômicas que podem ter sido alvo de seleção, com auxílio de informações genômicas.

CAPÍTULO 1 – REVISÃO BIBLIOGRÁFICA

Bubalinocultura no Brasil

O búfalo doméstico (*Bubalus bubalis*) tem uma contribuição significativa para a economia rural de muitos países em desenvolvimento, servindo como fonte de leite, carne, couro e força de trabalho (MISHRA et al., 2015). A domesticação destes animais provavelmente ocorreu no subcontinente indiano há cerca de 5.000 anos (WANG et al., 2017; Barker et al., 1997) e sua grande adaptabilidade a condições adversas de umidade e resistência a doenças (SAFARI et al., 2019), proporcionou sua dispersão para vários lugares no mundo.

No Brasil, a bubalinocultura teve início por volta de 1902, quando se deu o primeiro registro oficial de importação desses animais em Marajó no estado do Pará (SANTIAGO, 2000). Hoje, estima-se que o contingente de búfalos no país seja de aproximadamente 1,3 milhões de cabeças (MAPA, 2017), produzindo cerca de 92,3 milhões de litros de leite de búfala dentre os quais são processados 45 milhões de quilos de leite gerando 18,5 mil toneladas de produtos lácteos como *mozzarella*, queijo frescal, ricota, doce de leite entre outros (BERNARDES, 2007).

Das aproximadamente vinte raças de búfalos existentes no mundo, quatro são criadas no Brasil, são elas: Murrah, Jafarabadi, Mediterrâneo e Carabao (DAMÉ et al., 2013), sendo a Murrah a raça mais requerida para produção de leite no Brasil (ABCB, 2018). A descrição do padrão racial é feita pela Associação Brasileira de Criadores de Búfalos (ABCB), criada em 1960, que tem por objetivo o incentivo da atividade no país, o apoio técnico da produção e o serviço de registro genealógico dos animais.

Avaliação da estrutura populacional

Ao analisar dados genéticos, muitas vezes deseja-se determinar se ao escolher amostras aleatoriamente, estas pertencem a uma população homogênea, ou se os dados implicam que a população é geneticamente estruturada (PATTERSON et al., 2006). A maioria das populações é agrupada em subpopulações menores, nas quais geralmente ocorrem os acasalamentos, esse agrupamento é chamado de estrutura populacional. É muito provável que a subdivisão da população venha acompanhada de alguma diferenciação entre as subpopulações, quando as frequências alélicas destas se tornam diferentes (HARTL; CLARK, 2010).

Segundo Pritchard et al. (2010), a definição de população baseada em padrões físicos, culturais ou geográficos dos indivíduos é um tanto subjetiva, sendo difícil saber se essas atribuições consistem com informações em nível genético. De fato, Waples & Gaggiotti (2006), trazem à luz a complexidade da caracterização do conceito de população mostrando que uma gama de definições existentes para o mesmo termo pode levar a diferentes interpretações por parte do pesquisador, dessa maneira, na ausência de um entendimento comum do que uma população representa, pode ser difícil avaliar ou comparar os resultados de estudos de populações estruturadas. Para tanto, com o intuito de mitigar a subjetividade de interpretações teóricas ou qualitativas é necessário definir métricas que possam ser calculadas para espécies na natureza e que possam ser usadas para determinar se os critérios populacionais foram atendidos.

Então, como poderíamos determinar critérios e definir métricas para que seja praticável a análise das relações genéticas entre indivíduos? Atendendo a este propósito, como será visto adiante desta revisão, a diversidade genética dentro e entre raças na pecuária antes predita por atributos fenotípicos, pode hoje ser validada pelo uso de marcadores moleculares e sua análise com sofisticadas técnicas estatísticas (HUSSAIN et al., 2017).

Por muito tempo a caracterização de diferentes raças de animais domésticos foi baseada em dados fenotípicos (morfologia e produção), também chamados de marcadores morfológicos que são, ainda hoje, a base do melhoramento genético convencional, em que características desejáveis são selecionadas nos genitores para os cruzamentos (MARIANTE et al., 2008; TURCHETTO et al., 2017). No entanto, esse tipo de marcador pode levar a conclusões pouco precisas pela dificuldade de se isolar o efeito ambiental, além disso, a medição e identificação de características morfológicas dos animais geralmente levam muito tempo (YANG et al., 2013).

Com o avançar da tecnologia, marcadores bioquímicos, tais como isoenzimas, obtidos através de eletroforese de proteínas, foram amplamente utilizados na investigação da variação genética dentro e entre espécies e são ainda hoje amplamente utilizados. O uso de marcadores citológicos, baseados na comparação do número e estrutura dos cromossomos entre as espécies também foram e ainda são bastante empregados (YANG et al., 2013, AL-SAMARAI & AL-KAZAS, 2014).

A partir dos anos de 1970 (principalmente a partir de 1990), surgiram os marcadores baseados em DNA, capazes de superar as limitações das tecnologias anteriormente aplicadas (BRUFORD et al., 2003). Desde então, vislumbrou-se a possibilidade de utilizar a informação oriunda do fenótipo com aquela oriunda

diretamente do DNA (genótipo), de forma que os processos de identificação e seleção de plantas e animais com genótipos superiores pudessem ser realizados de forma mais eficiente (ABAD et al., 2014), além da vasta utilização em estudos de filogenética e diversidade de populações (YANG et al., 2013).

Vários são os tipos dos marcadores moleculares e das técnicas utilizadas para sua obtenção, como por exemplo, RFLP (*Restriction Fragment Length Polymorphism*) utilizados na produção do primeiro mapa genético humano (VIGNAL et al., 2002); RAPD (*Random Amplified Polymorphism DNA*) utilizado na classificação de raças de búfalos (PARAGUISON et al., 2012); AFLP (*Amplified Fragment Length Polymorphism*) aplicado na caracterização de bovinos (MALIK et al., 2018); microssatélites (SSR – *Simple Sequence Repeats*) desenvolvidos pela primeira vez para uso em mapeamento genético em humanos e desde então vêm sendo amplamente utilizados em diversas áreas da ciência (TURCHETTO et al., 2017) como na caracterização de raças bubalinas (UFFO et al., 2017; VIJH et al., 2018).

Além das abordagens citadas acima, os marcadores do tipo SNP (*Single Nucleotide Polymorphism*) tem sido uma escolha popular de método para estudar a diversidade de várias espécies de animais (SHARMA et al., 2016). De acordo com Caetano (2009), os marcadores do tipo SNP tem como base as alterações mais elementares da molécula de DNA, ou seja, mutações em bases únicas da cadeia de bases nitrogenadas (Adenina, Citosina, Timina e Guanina). Ao longo dos anos, inovações em protocolos e automação de sequenciamento, biologia molecular e bioinformática aumentaram a capacidade de caracterização e genotipagem de marcadores SNP, permitindo a leitura de centenas de pares de bases de DNA e produzindo rapidamente uma grande quantidade de dados em um curto espaço de tempo (CAETANO, 2009; HEATHER et al., 2016).

Diante disso, fez-se necessário o desenvolvimento de métodos estatísticos e computacionais para a manipulação do grande volume de dados gerados, tornando a interpretação dos resultados mais gerenciável. Os dois métodos mais difundidos para agrupar indivíduos com base em dados genéticos podem ser classificados em duas categorias: abordagens baseadas em modelo, como é implementado no *software Structure* e abordagens baseadas em distâncias, como a Análise de Componentes Principais (PCA – *Principal Component Analysis*) (GREENBAUM et al., 2016).

No método baseado em modelo é suposto que existam K populações distintas, em que os genótipos de cada indivíduo são modelados em sorteios aleatórios, sendo que a inferência para a participação em cluster de cada indivíduo é realizada usando

métodos estatísticos de máxima verossimilhança ou bayesianos (PRITCHARD et al., 2000; LAWSON et al., 2012; BRADBURD et al., 2018). Métodos baseados em distância, como o próprio nome sugere, calculam o grau de similaridade ou dissimilaridade entre os pares de objetos (indivíduos) por meio de uma matriz de distância (dentre os tipos de distâncias que podem ser aplicadas estão Euclidiana, Manhattan, Mahalanobis, entre outras) (PRITCHARD et al., 2000; LINDEN, 2009). Segundo Jombart et al. (2009), o segundo método possui algumas vantagens sobre o agrupamento bayesiano como por exemplo não exigir pressuposições sobre um modelo e poder ser aplicado a grandes conjuntos de dados, o que seria impraticável na análise bayesiana.

Como mencionado anteriormente, a PCA pode ser umas das metodologias aplicadas ao agrupamento de indivíduos baseando-se em dados genéticos. Trata-se de uma técnica estatística multivariada popularmente empregada em diversas áreas do conhecimento (ABDI; WILLIAMS, 2010), na genética de populações foi utilizada pela primeira vez por Cavalli-Sforza (1994) no estudo da estrutura populacional humana e desde então foi aplicada a diversos trabalhos de caracterização genética de animais de produção, inclusive bubalinos (AGUDELO-GÓMEZ et al., 2015; JOSHI et al., 2015; MARQUES et al., 2011; entre outros). O objetivo elementar da PCA está em extrair as informações de maior relevância de um conjunto de dados e representá-las em forma de novas variáveis ortogonais (não correlacionadas) denominadas componentes principais ou variáveis sintéticas, reduzindo sistematicamente o número de dimensões dos dados (ABDI, WILLIAMS, 2010; DAVID, JACOBS, 2015).

Contextualizando para a genética de populações, pode-se entender que uma medida de similaridade genética é aplicada aos dados genômicos, atribuindo uma maior semelhança entre determinado par de indivíduos intimamente relacionados do que os mesmos teriam para com indivíduos geneticamente mais distantes, desta maneira uma matriz de similaridade é formada combinando esta informação para todos os pares de indivíduos (LAWSON; FALUSH, 2012). Assim, ao analisar marcadores genéticos, tais como SNP, o principal interesse é encontrar relações entre os η objetos, que nesse caso podem ser considerados os genótipos de indivíduos de uma população, usando p alelos. Os dados são dispostos em uma nuvem de η pontos embutidos dentro de um espaço p -dimensional, onde cada dimensão é definida por um alelo. A média das distâncias entre os pares de pontos, ou inércia, é calculada, sendo que, quanto maior a dispersão maior a inércia, menor a correlação entre os pontos e, conseqüentemente, maior a variabilidade (e vice-versa). Os eixos principais são,

portanto, extraídos levando-se em consideração as direções dentro do espaço que representem a maior variabilidade (maior inércia), cada um destes eixos são definidos por coordenadas representando as cargas dos alelos (a contribuição de cada alelo para a variabilidade). Por definição, os eixos são ortogonais, ou seja, o segundo componente principal (PC2) deve ser perpendicular ao primeiro (PC1), sendo que PC1 reflete a maior variabilidade dos dados e PC2 a maior variabilidade subsequente e assim por diante até serem obtidos todos os componentes não correlacionados e em ordem decrescente de variação (JOMBART et al., 2009; SAINANI, 2014).

Cada componente principal possui um autovalor associado que representa a fração de toda a variabilidade genética de seu componente correspondente. O gráfico dos autovalores (*screeplot*) descreve como a variabilidade é distribuída em ordem decrescente entre os eixos principais e juntamente com as proporções de inércia associada aos componentes, auxilia a interpretação do que pode vir a ser a estrutura verdadeira dos dados. Contudo, não se pode menosprezar a interpretação da natureza biológica dos dados, visto que em alguns casos, componentes relacionados a baixas variâncias podem indicar informações biologicamente relevantes (JOMBART et al., 2009).

Medidas de variabilidade genética

A domesticação de animais foi um passo essencial no desenvolvimento demográfico e cultural humano. Durante a história subsequente da pecuária, as principais forças evolutivas de mutação, reprodução seletiva, adaptação, isolamento e deriva genética propiciaram a eventual fixação de algumas características específicas levando a formação de raças com variáveis intensidades de fluxo gênico e, conseqüentemente, a subdivisões genéticas em diferentes níveis (MARIANTE et al., 2008; GROENEVELD et al., 2010; LENSTRA et al., 2012).

A caracterização dessa variabilidade genética existente, dentro e entre raças, auxilia na compreensão da existência ou não de famílias ou linhagens com aptidões específicas de produção, como a identificação de indivíduos geneticamente superiores para características de produção de leite, por exemplo. Além disso, com a intensificação da seleção artificial e uso de biotecnologias aplicadas à reprodução, aumentou-se a eficiência reprodutiva dos animais utilizados como reprodutores fazendo com que um menor número de animais sejam selecionados como pais da próxima geração, o que gera aumento da homozigose (ALBARELLA et al., 2009; WHITACRE et al., 2017). Portanto, resultados deste tipo de estudo norteiam programas de seleção mais eficazes, por meio da gestão das estratégias de

acasalamentos que podem controlar os níveis de endogamia (MARCONDES et al., 2010; REZENDE et al., 2017).

Para esse fim, o equilíbrio de Hardy-Weinberg (EHW), as heterozigosidades e as estatísticas F de Wright são métricas úteis na avaliação dos padrões de diversidade genética nas populações.

O equilíbrio de Hardy-Weinberg (EHW), estipula que em uma população infinitamente grande, na ausência de migração, mutação, seleção e com os acasalamentos ocorrendo ao acaso, as frequências genotípicas tendem a permanecer estáveis ao longo do tempo. Entretanto, essas premissas geralmente não são encontradas em populações reais, sendo a interpretação dos desvios do EHW útil para averiguação do controle de qualidade de genotipagem em larga escala, além de testar hipóteses evolutivas sobre demografia e sistemas de acasalamento (WIGGINTON et al., 2005; WITTKÉ-THOMPSON et al., 2005; MEIRMANS, 2019).

A heterozigosidade observada (H_o) pode ser entendida como a razão entre o número total de indivíduos heterozigotos encontrados na população e o número total de indivíduos existentes na população. Enquanto que, a heterozigosidade esperada (H_e) é a probabilidade de dois alelos escolhidos aleatoriamente na população serem diferentes, sendo este o parâmetro mais amplamente utilizado para medir a diversidade genética (TORO et al., 2009). Valores de heterozigosidade esperada altos são indicativos de grande diversidade genética e espera-se que ocorra quando as populações tenham sido isoladas recentemente, pois isso diminui a probabilidade de autozigosidade, mas em qualquer subpopulação de tamanho finito essa probabilidade aumenta ao longo do tempo (HARTL; CLARK, 2010). Quando as heterozigosidades (H_o e H_e) forem equivalentes, considera-se que as populações estão em completo equilíbrio (MENEZES, 2006).

Para quantificar a perda de heterozigosidade, Wright (1951) estabeleceu o denominado índice de fixação. A estatística F, como também é conhecida, decompõe a diversidade genética em três componentes considerando-se populações subdivididas, conforme descrito por Hartl e Clark (2010):

- Dentro da subpopulação ou individual (coeficiente de endogamia F_{is}): redução na heterozigosidade em um indivíduo devido ao endocruzamento, em relação à população total à qual esse indivíduo pertence, ou ainda, reflete a probabilidade de que dois alelos de um locus em um indivíduo sejam idênticos por descendência. Valores positivos podem ser indicativos de acasalamentos endogâmicos enquanto, valores negativos sugerem excesso de heterozigotos.

- Entre subpopulações (índice de fixação F_{st}): abrange a perda de heterozigosidade das subpopulações em relação à população total, ou seja, o nível total de divergência genética entre as subpopulações. Escores de F_{st} são fornecidos para auxílio na comparação dos valores de divergência entre populações, como segue:
 - 0-0,05 – considerada indicativa de pequena diferenciação genética;
 - 0,05-0,15 – indica moderada diferenciação genética;
 - 0,15-0,25 – indica grande diferenciação genética;
 - acima de 0,25 – indica diferenciação genética muito grande.

Porém, vale ressaltar que a divisão desse índice em escalas pode o tornar subjetivo e alguns pesquisadores têm relatado a preferência em comparar os valores de F_{st} com o encontrado em outras pesquisas, como também tentar associá-lo com a história de vida da espécie em estudo. Inclusive Wright sugere que mesmo com valores de F_{st} tão baixos ou menores que 0,05 a diferenciação não pode ser considerada insignificante.

- Na população total (F_{it}) – redução da heterozigosidade de um indivíduo em relação à população total.

Assinatura de seleção

A exposição dos animais domésticos a diferentes pressões de seleção (natural ou artificial) devido à variedade de ambientes e sistemas de criação experimentados ao longo do tempo resultou na formação de diferentes raças pelo mundo (PORTONETO et al., 2013). A seleção positiva, decorrente desta pressão seletiva, faz com que ocorra o aumento da prevalência de características vantajosas na população, este processo acaba por deixar padrões genéticos únicos na sequência do DNA, chamados de assinatura de seleção (RANDHAWA et al., 2016). Em teoria, a assinatura de seleção pode ser caracterizada pela distribuição de nucleotídeos em torno de mutações favoráveis, como a probabilidade de recombinação destes é inversamente proporcional à distância que os separam, os nucleotídeos adjacentes à mutação favorável tendem a aumentar em frequência, em um processo de “carona” denominado desequilíbrio de ligação (STELLA et al., 2010).

Segundo Cheruiyot et al. (2018), as metodologias empregadas para a detecção de assinaturas de seleção podem ser classificadas em três categorias: medidas baseadas na frequência alélicas (métodos D de Tajima; teste Fu e Li); medidas populacionais baseadas em homozigose de haplótipo estendido (método iHS) e

medidas baseadas na diferenciação entre e dentro das espécies (métodos F_{st} ; XPEHH). No presente trabalho, foi utilizado o método F_{st} , em que são fornecidas informações sobre a variação genômica em um locus entre populações em relação àquelas dentro das populações, assim, altos valores de F_{st} (*outliers*) indicam adaptação positiva local, enquanto valores baixos indicam seleção negativa ou neutra (ZHAO et al., 2015). Segundo Barendse et al. (2009), assinaturas úteis de seleção podem ser obtidas em grupos que têm apenas pequenas diferenças genéticas e que não ser claramente diferenciados em uma análise de componentes principais.

É importante salientar que, a identificação de assinaturas de seleção e o estudo de associação genômica ampla (GWAS - *genome wide association studies*), são duas abordagens diferentes para a detecção de possíveis genes candidatos. Enquanto GWAS avalia a relação entre genótipo e fenótipo, a assinatura de seleção avalia parâmetros genéticos da população utilizando apenas informações genômicas (Zhao, et al., 2015).

Uma vez que as regiões genômicas alvos de seleção são detectadas, os genes e processos envolvidos na formação de fenótipos de interesse dentro e entre as subpopulações podem ser mais prontamente identificados (PETERSEN et al., 2013).

OBJETIVO

Estudar a estrutura da população bubalina leiteira no Brasil e identificar regiões genômicas que podem ter sido alvo de seleção, com auxílio de informações genômicas.

OBJETIVO ESPECÍFICO

- Identificar se há a ocorrência de estratificação na população por meio de Análise de Componentes Principais (PCA);
- Estimar a variabilidade genética populacional com a aferição de índices de diversidade tais como H_o , H_e , F_{is} e F_{st} ;
- Identificar regiões genômicas que podem ter sido alvo de seleção, utilizando o método F_{st} .

REFERÊNCIAS BIBLIOGRÁFICAS

ABAD, A.C.A.; LOPES, F.A.; PINHEIRO, J.W.; MOTA, R.A. Marcadores moleculares e suas aplicações nas pesquisas em bovinos. **Acta Veterinaria Brasilica**, v.8, n.1, p.10-18, 2014.

ABCB. Associação Brasileira de Criadores de Búfalos. Disponível em <http://www.bufalo.com.br/home/?page_id=1133> Acesso em: Setembro de 2018.

ABDI, H.; WILLIAM, L. J. Principal component analysis. **WIREs Comp Stat, John Wiley & Sons, Inc.** v.2 p.433–459, 2010. DOI: 10.1002/wics.101.

AGUDELO-GÓMEZ, D.; PINEDA-SIERRA, S.; CERÓN-MUÑOZ, M.F. Genetic evaluation of dual-purpose buffaloes (*bubalus bubalis*) in Colombia using principal component analysis. **Plos One**, 10(7), p.1-9, 2015. Doi:10.1371/journal.pone.0132811.

ALBARELLA, S.; CIOTOLA, F.; DARIO, C.; IANNUZZI, L.; BARBIERI, V.; PERETTI, V. Chromosome instability in Mediterranean Italian buffaloes affected by limb malformation (transversal hemimelia). **Mutagenesis**, v.24, n.6, p.471–474, 2009. Doi:10.1093/mutage/geb030.

AL-SAMARAI, F.R.; AL-KAZAZ, A.A. Applications of molecular markers in animal breeding: a review. **American Journal of Applied Scientific Research**, v.1(1), p.1-5, 2015. doi: 10.11648/j.ajasr.20150101.11.

BARENDSE, W.; HARRISON, B.E.; BUNCH, R.J.; THOMAS, M.B.; TURNER, L.B. Genome wide signatures of positive selection: the comparison of independent samples and the identification of regions associated to traits. **BMC Genomics**, 10:178, p.1-15, 2009. Doi: 10.1186/1471-2164-10-178.

BARKER, J.S.F.; TAN, S.G.; SELVARAJ, O.S.; MUKHERJEE. Genetic variation within and relationships among populations of Asian water buffalo (*Bubalus bubalis*). **Animal Genetics**, v.28, p.1-13, 1997.

BERNARDES, O. Bubalinocultura no Brasil: situação e importância econômica. **Rev. Bras. Reprod. Anim.**, v.31, n.2, 2007.

BOETTCHER, P.J.; TIXIER-BOICHARD, M.; TORO, M.A.; SIMIANER, H.; EDING, H.; GANDINI, G.; JOOS, S.; GARCIA, D.; COLLI, L.; AJMONE-MARSAN, P.; GLOBALDIV CONSORTIUM. Objectives, criteria and methods for using molecular genetic data in priority setting for conservation of animal genetic resources. **Animal Genetics**, v.41(1), p.64–77, 2010. Doi:10.1111/j.1365-2052.2010.02050.x.

BRADBURD, S.G.; COOP, G.M.; RALPH, P.L. Inferring Continuous and Discrete Population Genetic Structure Across Space. **Genetics**, v.210, p.33–52, 2018. Doi: <https://doi.org/10.1534/genetics.118.301333>.

BRUFORD, M.W.; BRADLEY, D.G.; LUIKART, G. Dna markers reveal the complexity of livestock domestication. **Nature Reviews Genetics**, v.4, p.900-910, 2003. Doi:10.1038/nrg1203.

CAETANO, A. Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. **R. Bras. Zootec.**, v.38, p.64-71, 2009. DOI: 10.1590/S1516-35982009001300008.

CAVALLI-SFORZA, L.L.; MENOZZI, P.; PIAZZA, A. **The History and Geography of Human Genes**. Princeton University Press, Princeton, New Jersey, 1994.

CHERUIYOT, E.K.; BETT, R.C.; AMIMO, J.O.; ZHANG, Y.; MRODE, R.; MUJIBI, F.D.N. Signatures of selection in admixed dairy cattle in Tanzania. **Front. Genet.**, v.9, article 607, 2018. Doi: 10.3389/fgene.2018.00607.

DAMÉ, M.C.F.; RIET-CORREA, F.; SCHILD, A.L. Doenças hereditárias e defeitos congênitos diagnosticados em búfalos (*Bubalus bubalis*) no Brasil. **Pesq. Vet. Bras.**, 33(7): 831-839, julho, 2013.

DAVID, C.C.; JACOBS, D.J. Principal component analysis: a method for determining the essential dynamics of proteins. **Methods Mol Biol.** 2014; 1084: 193–226. doi:10.1007/978-1-62703-658-0_11.

GREENBAUM, G.; TEMPLETON, A.R.; BAR-DAVID, S. Inference and analysis of population structure using genetic data and network theory. **Genetics**, v.202, p.1299–1312, 2016. Doi: 10.1534/genetics.115.182626.

GROENEVELD, L. F., LENSTRA, J. A., EDING, H., TORO, M. A., SCHERF, B., PILLING, D., NEGRINI, R., FINLAY, E. K., JIANLIN, H., GROENEVELD, E.,

WEIGEND, THE GLOBALDIV CONSORTIUM. Genetic diversity in farm animals – a review. **Animal Genetics**, 41 (Suppl. 1), 6–31 2010. Doi: doi:10.1111/j.1365-2052.2010.02038.x.

HARTL, D.; CLARK, A.G. Organização da variação genética, capítulo 2, p.61-99. Endocruzamento, subdivisão populacional e migração, capítulo 6, p.271-322. In: **Princípios de Genética de Populações** – 4.ed. Porto Alegre: Artmed Editora S.A., 2010.

HEATHER, J.M.; CHAIN, B. The sequence of sequencers: The history of sequencing DNA. **Genomics**, v.107, p.1–8, 2016.

HUSSAIN, T.; BABAR, M.E.; ALI, A.; NADEEM, A.; REHMAN, Z.U.; MUSTHAFA, M.M.; MARIKAR, F.M.M.T. Microsatellite based genetic variation among the buffalo breed populations in Pakistan. **J. Vet. Res.**, v.61, p.535-542, 2017. Doi:10.1515/jvetres-2017-0057.

JOMBART, T.; PONTIER, D.; DUFOUR, A.B. Genetic markers in the playground of multivariate analysis. **Heredity**, v.102, p.330–341, 2009.

JOMBART, T.; PONTIER, D.; DUFOUR, A-B. Genetic markers in the playground of multivariate analysis. **Heredity**,v.102, p.330–341, 2009.

JOSHI, J.; SALAR, R.K; BANERJEE, P.; SHARMA, U.; TANTIA, M.S.; VIJH, R.K. Assessment of genetic variability and structuring of riverine buffalo population (*Bubalus bubalis*) of indo-genetic basin. **Animal Biotechnology**, v.26(2), p.148-155, 2015. Doi: 10.1080/10495398.2014.955613.

LAWSON, D.J.; FALUSH, D. Population identification using genetic data. **Annu. Rev. Genomics Hum. Genet.**, 13:14.1–14.25, 2012. Doi: 10.1146/annurev-genom-082410-101510.

LAWSON, D.J.; HELLENTHAL, G.; MYERS, S.; FALUSH, D. Inference of Population Structure using Dense Haplotype Data. **PLoS Genetics**, v.8, p.1-16, 2012.

LENSTRA, J.A.; GROENEVELD, L.F.; EDING, H.; KANTANEN, J.; WILLIAMS, J.L.; TABERLET, P.; NICOLAZZI, E.L.; SOLKNER, J.; SIMIANER, H.; CIANI, E.; GARCIA, J.F.; BRUFORD, M.W.; AJMONE-MARSAN, P.; S. WEIGEND, S. Molecular tools and

analytical approaches for the characterization of farm animal genetic diversity. **Animal Genetics**, v.43, p.483–502, 2012.

LINDEN, R. Técnicas de agrupamento. **Revista de Sistemas de Informação da FSMA**, n. 4, p.18-36, 2009.

MALIK, M.H.; MOAEEN-UD-DIN, M.; BILAL, G.; GHAFAR, A.; MUNER, R.D.; RAJA, G.K.; KHAN, W. A. Development of amplified fragment length polymorphism(AFLP) markers for the identification of Cholistani cattle. **Arch. Anim. Breed.**, 61, 387–394, 2018.

MAPA. Ministério da Agricultura, Pecuária e Abastecimento. Disponível em <http://www.agricultura.gov.br/assuntos/sanidade-animale-vegetal/saude-animale-programas-de-saude-animale-febre-aftosa/documentos-febre-aftosa/DadosderebanhobovinoebubalinodoBrasil_2017.pdf> Acesso em: Setembro de 2018.

MARCONDES, C.R.; VOZZI, P.A.; CUNHA, B.R.N.; LÔBO, R.B.; ARAÚJO, C.V.; MARQUES, J.R.F. Variabilidade genética de búfalos em rebanho-núcleo com base na análise de *pedigree*. **Arq. Bras. Med. Vet. Zootec.**, v.62, n.3, p.706-711, 2010.

MARIANTE, A.S.; EGITO, A.A.; ALBUQUERQUE, M.S.M.; PAIVA, S.R.; RAMOS, A.F. Managing genetic diversity and society needs. **R. Bras. Zootec.**, v.37, suplemento especial p.127-136, 2008.

MARQUES, J.R.F., MARTÍNEZ, A.M., COSTA, M.R., ALBUQUERQUE, M.S.M., QUIROZ, J., VEGA-PLA, J.L., DELGADO, J.V. Genetic Diversity Of Brazilian Buffaloes (*Bubalus Bubalis*) Using Dna Microsatellites. **Arch. Zootec.** 60 (232): 1213-1221. 2011.

MENEZES, M.P.C.; MARTINEZ, A.M.; RIBEIRO, M.N.; PIMENTA FILHO, E.C.; BERMEJO, J.V.D. Caracterização genética de raças caprinas nativas brasileiras utilizando-se 27 marcadores microssatélites. **R. Bras. Zootec.**, v.35, n.4, p.1336-1341, 2006.

MISHRA, B.P.; DUBEY, P.K.; PRAKASH, B.; KATHIRAVAN, P.; GOYAL, S.; SADANA, D.K.; DAS, G.C.; GOSWAMI, R.N.; BHASIN, V.; JOSHI, B.K.; KATARIA, R.S. Genetic analysis of river, swamp and hybrid buffaloes of north-east India throw new light on

phylogeography of water buffalo (*Bubalus bubalis*). **J. Anim. Breed. Genet.**, v.132, p.454–466, 2015. doi:10.1111/jbg.12141.

PARAGUISON, R.C.; FAYLON, M.P.P.; FLORES, E.B.; CRUZ, L.C. Improved RAPD-PCR for Discriminating Breeds of Water Buffalo. **Biochem Genet**, v.50, p.579–584, 2012. DOI 10.1007/s10528-012-9502-8.

PATTERSON, N.; PRICE, A. L.; REICH, D. Population Structure and Eigenanalysis. **PLoS Genet**, v.2, p.22, 2006.

PETERSEN, J.L.; MICKELSON, J.R.; RENDAHL, A.K.; VALBERG, S.J.; ANDERSSON, L.S.; AXELSSON, J.; BAILEY, E.; BANNASCH, D.; BINNS, M.M.; BORGES, A.S.; BRAMA, P.; MACHADO, A.C.; CAPOMACCIO, S.; CAPPELLI, K.; COTHRAN, E.G.; DISTL, O.; FOX-CLIPSHAM, L.; GRAVES, K.T.; GUÉRIN, G.; HAASE, B.; HASEGAWA, T.; HEMMANN, K.; HILL, E.W.; LEEB, T.; LINDGREN, G.; LOHI, H.; LOPES, M.S.; MCGIVNEY, B.A.; MIKKO, S.; ORR, N.; PENEDO, M.C.T.; PIERCY, R.J.; RAEKALLIO, M.; RIEDER, S.; RØED, K.H.; SWINBURNE, J.; TOZAKI, T.; VAUDIN, M.; WADE, C.M.; MCCUE, M.E. Genome-wide analysis reveals selection for important traits in domestic horse breeds. **Plos One**, v.9, e1003211, 2013. Doi:10.1371/journal.pgen.1003211.

PORTO-NETO, L.R.; LEE, S.H.; LEE, H.K.; GONDRO, C. Detection of signatures of selection using FST. In: **Genome-Wide Association Studies and Genomic Prediction, Methods in Molecular Biology**, vol. 1019. Springer Science+Business Media. Capítulo 19. p.423-436, 2013.

PRITCHARD, J.K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, v.155, p.945–959, 2000.

RANDHAWA, I.A.S.; KHATKAR, M.S.; THOMSON, P.C.; RAADSMA, H.W. A meta-assembly of selection signatures in cattle. **Plos One**, 11(4), p.1-30, 2016. Doi: 10.1371/journal.pone.0153013.

REZENDE, M.P.G.; CONDE, E.A.S.L.; BORGES, A.C.; CARNEIRO, P.L.S.; MARTINS FILHO, R.; MALHADO, C.H.M. Estrutura populacional do rebanho Nelore criado no semiárido do nordeste brasileiro. **Cienc. anim. Brás.**, v.18, p.1-14, 2017. Doi: 10.1590/1089-6891v18e-38048.

SAFARI, A.; SHADPARVAR, A.A.; HOSSEIN-ZADEH, N.G.; ABDOLLAHI-ARPANAHI, R. Economic values and selection indices for production and reproduction traits of Iranian buffaloes (*Bubalus bubalis*). **Tropical Animal Health and Production**. Published online, 2019. Doi: <https://doi.org/10.1007/s11250-019-01811-7>.

SAINANI, K.L. Introduction to Principal Components Analysis. **American Academy of Physical Medicine and Rehabilitation**. v.6, p.275-278, 2014.

SANTIAGO, A. A. **Introdução dos búfalos no Brasil**. Associação Brasileira de Criadores de Búfalos, 2000. Disponível em: <http://www.bufalo.com.br/info_criador/historico_bufalos.pdf> Acesso em: Setembro de 2018.

SHARMA, A.; LEE, S-H, LIM, D.; CHAI, H-H.; CHOI, B-H.; CHO, Y. A genome-wide assessment of genetic diversity and population structure of Korean native cattle breeds. **BMC Genetics**, v.17, p.139, 2016. DOI 10.1186/s12863-016-0444-8.

STELLA, A.; AJMONE-MARSAN, P.; LAZZARI, B.; BOETTCHER, P. Identification of selection signatures in cattle breeds selected for dairy production. **Genetics**, v.185, p.1451–1461, 2010. Doi: 10.1534/genetics.110.116111.

TONHATI, H., MUÑOZ, M. F. C., OLIVEIRA, J. A., DUARTE, J. M. C., FURTADO, T. P., TSEIMAZIDES, S. P. Parâmetros Genéticos para a Produção de Leite, Gordura e Proteína em Bubalinos. **Rev. bras. zootec.**, v.29(6), p.2051-2056, 2000.

TORO, M.A.; FERNÁNDEZ, J.; CABALLERO, A. (2009). Caracterização molecular de raças e seu uso em conservação. **Livestock Science**, 120(3), p.174–195. Doi:10.1016/251 j.livsci.2008.07.003.

TURCHETTO, C.; TURCHETTO-ZOLET, A.C.; PASSAIA, G.; ZANELLA, C.M. Marcadores genéticos baseados em DNA. In: **Marcadores Moleculares na Era genômica: Metodologias e Aplicações**.Ribeirão Preto: Sociedade Brasileira de Genética, 2017.Capítulo 1. p.12-20.

UFFO, O.; MARTÍNEZ, N.; ACOSTA, A.; SANZ, A.; MARTÍN-BURRIEL, I.; OSTA, R.; RODELLAR, C.; ZARAGOZA P. Analysis of microsatellite markers in a Cuban water buffalo breed. **Journal of Dairy Research**, v.84, p.289–292, 2017. doi:10.1017/S0022029917000425.

VIGNAL, A.; MILAN, D.; SANCRISTOBAL, M.; EGGEN, A. A review on SNP and other types of molecular markers and their use in animal genetics. **Genet. Sel. Evol.**, v.34, p.275_305, 2002. Doi: 10.1051/gse:2002009.

VIJH, R.K.; SHARMA, U.; BANERJEE, P.; JOSHI, J.; SETTYPALLI, T.B.K.; GOKHALE, S.B.; GAUR, U.; YADAV, D.K.; AHLAWAT, S.; ARORA, R. New insights into the spatial genetic structure of the indian riverine buffalo populations. **Livestock Science**, p.1-20, 2018. Doi: <https://doi.org/10.1016/j.livsci.2018.09.003>.

WANG, S.; CHEN, N.; CAPODIFERRO, M.R.; ZHANG, T.; LANCIONI, H.; ZHANG, H.; MIAO, Y.; CHANTHAKHOUN, V.; WANAPAT, M.; YINDEE, M.; ZHANG, Y.; LU, H.; CAPORALI, L.; DANG, R.; HUANG, Y.; LAN, X.; PLATH, M.; CHEN, H.; LENSTRA, J.A.; ACHILLI, A.; LEI, C. Whole mitogenomes reveal the history of swamp buffalo: initially shaped by glacial period and eventually modelled by domestication. **Scientific Reports**, 7: 4708.2017. Doi:10.1038/s41598-017-04830-2.

WAPLES, R.S.; R.; GAGGIOTTI, O. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. **Molecular Ecology**, v.15, p.1419–1439, 2006. doi: 10.1111/j.1365-294X.2006.02890.x.

WHITACRE, L.K.; HOFF, J.L.; SCHNABEL, R.D.; ALBARELLA, S.; CIOTOLA, F.; PERETTI, V.; STROZZI, F.; FERRANDI, C.; RAMUNNO, L.; SONSTEGARD, T.S.; WILLIAMS, J.L.; TAYLOR, J.F.; DECKER, J.E. Elucidating the genetic basis of a oligogenic birth defect using whole genome sequence data in a non-model organism, *Bubalus bubalis*. **Scientific Reports**, 7, 39719, 2017. Doi: 10.1038/srep39719.

WIGGINTON, J.E.; CUTLER, D.J.; ABECASIS, G.R. A note on exact tests of Hardy-Weinberg equilibrium. **Am. J. Hum. Genet.**, v.76, p.887–883, 2005.

WITTKÉ-THOMPSON, J.K.; PLUZHNIKOV, A.; COX, N.J. Rational inferences about departures from Hardy-Weinberg equilibrium. **Am. J. Hum. Genet.**, v.76, p.967–986, 2005.

WRIGHT, S. The genetical structure of populations. **Annals of Eugenics**, v.15, p.323-354, 1951.

YANG, W.; KANG, X.; YANG, Q.; LIN, Y.; FANG, M. Review on the development of genotyping methods for assessing farm animal diversity. **Journal of Animal Science and Biotechnology**, 4:2, 2013. doi:10.1186/2049-1891-4-2.

ZHAO, F.; MCPARLAND, S.; KEARNEY, F.; DU, L.; BERRY, D.P. Detection of selection signatures in dairy and beef cattle using high-density genomic information. **Genetics Selection Evolution**, 47:49, p.1-12, 2015. Doi: DOI 10.1186/s12711-015-0127-3.

CAPÍTULO 2: ESTRUTURA POPULACIONAL DE BUBALINOS LEITEIROS UTILIZANDO DADOS GENÔMICOS

Silveira, Karina Rosa. **Estrutura populacional de bubalinos leiteiros utilizando dados genômicos**. 2019. Dissertação (mestrado em zootecnia). Faculdade de Ciências Agrárias, Universidade Federal da Grande Dourados.

RESUMO

Este estudo teve como objetivo descrever a estrutura da população bubalina leiteira no Brasil e identificar regiões genômicas que podem ter sido alvo de seleção, com auxílio de informações genômicas. O banco de dados foi composto por 978 indivíduos distribuídos em 7 fazendas nos estados de São Paulo, Pará e Bahia e Rio Grande do Norte. A genotipagem foi realizada com chip 90K Axion® Buffalo Genotyping). Após o controle de qualidade sobre os dados brutos gerados pela genotipagem, 48.795 marcadores (autossômicos) permaneceram no banco de dados, os quais foram submetidos à Análise de Componentes Principais (PCA - *Principal Component Analysis*) para visualização gráfica da estruturação populacional. Nesta etapa, foram distinguidos 3 grupos genéticos e, posteriormente, foram estimados os parâmetros genéticos de heterozigosidade observada (H_o), heterozigosidade esperada (H_e) e índice de fixação intra-populacional (F_{is}) para cada grupo. O índice de fixação entre subpopulações (F_{st}) foi calculado entre os pares de grupos para medida de diferenciação genética entre as mesmas. Para identificação de regiões genômicas supostamente sob seleção foram estimados valores de F_{st} para cada loco, nos quais foram visualizados em relação à respectiva posição física nos cromossomos. O resultado da PCA foi capaz de demonstrar a presença de subpopulações, mesmo com uma baixa estruturação. Os valores médios de heterozigosidades, observada e esperada, variaram de 0,4143 a 0,4331 e 0,4119 a 0,4373, respectivamente. Estes resultados sugerem variabilidade moderada a alta dentro das subpopulações. O índice de fixação intra-populacional (F_{is}) variou de -0,0294 a 0,0329, o que indica baixo efeito da endogamia nos rebanhos. O valor de fixação entre subpopulações (F_{st}) foi maior entre os grupos 1 e 2 (0,0291), seguido dos grupos 2 e 3 (0,0163) e o menor valor de diferenciação foi entre os grupos 1 e 3 (0,0152), indicando baixa diferenciação genética entre as subpopulações e podendo-se inferir que há mais variabilidade dentro das subpopulações que entre as mesmas. Os resultados apontaram baixa estruturação genética e baixa divergência entre os grupos, indicando que são bastante similares geneticamente. As regiões identificadas como altamente diferenciadas de

acordo com o índice F_{st} forneceram informações a cerca de genes que podem estar sendo alvo de seleção na população.

Palavras-chave: análise de componentes principais, assinatura de seleção, *Bubalus bubalis*, genética de populações, SNP

ABSTRACT

This study aimed to describe the structure of the dairy buffalo population in Brazil and to identify genetic regions that may have been the target of selection, with the aid of genetic information. The database consisted of 978 individuals distributed in 7 farms in the states of São Paulo, Pará, Bahia and Rio Grande do Norte. Genotyping was performed using the 90K Axion® Buffalo Genotyping chip. After quality control over the raw data generated by genotyping, 48,795 (autosomal) markers remained in the database, which were submitted to a Principal Component Analysis (PCA) for graphical visualization of the population structure. At this stage, 3 genetic groups were distinguished and, subsequently, the genetic parameters of observed heterozygosity (H_o), expected heterozygosity (H_e) and intra-population fixation index (F_{is}) were estimated for each group. The fixation index between subpopulations (F_{st}) was calculated between pairs of groups to measure genetic differentiation between them. To identify genomic regions supposedly under selection, F_{st} values were estimated for each locus, in which they were visualized in relation to their physical position on the chromosomes. The result of PCA was able to demonstrate the presence of subpopulations, even with a low structure. Mean observed and expected heterozygosity values ranged from 0.4143 to 0.4331 and 0.4119 to 0.4373, respectively. These results suggest moderate to high variability within subpopulations. The intra-population fixation index (F_{is}) ranged from -0.0294 to 0.0329, which indicates low effect of inbreeding in herds. The fixation value between subpopulations (F_{st}) was higher between groups 1 and 2 (0.0291), followed by groups 2 and 3 (0.0163) and the lowest differentiation value was between groups 1 and 3 (0, 0152), indicating low genetic differentiation between subpopulations and it can be inferred that there is more variability within subpopulations than between them. The results showed low genetic structure and low divergence between the groups, indicating that they are very similar genetically. The regions identified as highly differentiated according to the F_{st} index provided information about genes that may be targeted in the population.

Keywords: *Bubalus bubalis*, population genetics, principal component analysis, selection signature, SNP

INTRODUÇÃO

O búfalo doméstico é uma espécie economicamente importante, correspondendo a segunda maior fonte de leite no mundo, este possui maior teor de gordura, lactose e proteína do que o leite de vaca. Na Ásia, por exemplo, a espécie é responsável por cerca de 50% da produção total de leite, ainda assim, a representatividade da produção no cenário mundial é baixa, com um fornecimento de 13% aproximadamente (DU et al., 2019).

Enquanto que, em muitos países a criação de búfalos é basicamente para subsistência, em outros como, por exemplo, Brasil e Itália, a produção do leite de búfala é destinada em sua maioria para a produção de queijo *mozzarella*. No Brasil, a região sudeste representa um grande centro de produção e consumo deste produto e com o aumento da demanda por produtos lácteos de alta qualidade nas últimas décadas ocorreu também grande valorização da bubalinocultura no país (ASPILCUETA- BORQUIS et al., 2010; SENO et al., 2010).

Com a importância econômica e social da criação de búfalos, torna-se importante elaborar estratégias que promovam o melhoramento da espécie. Para que isso ocorra de maneira eficiente é importante conhecer os padrões da estrutura e da diversidade genética existentes nas populações. Segundo Lenstra et al. (2012), a diversidade genética dentro das espécies refere-se à extensão da variação genética dentro e entre raças e/ou linhagens, sendo que, esta variação é um importante requisito para adaptações às mudanças ambientais como também para o entendimento da variação fenotípica e reconstrução da história da pecuária.

A princípio, na ausência de meios para investigar a diversidade genética, a não ser indiretamente ao nível fenotípico, a extensão e o caráter dessa diversidade permaneceu por algum tempo elusivo (ELLEGREN; GALTIER, 2016). Posteriormente, com a genotipagem de alta densidade baseada em chip, disponibilizando uma grande quantidade de marcadores SNP, os estudos da diversidade genética em todo o genoma tornaram-se possíveis e mais precisos (ALBRECHTSEN et al., 2010). Além disso, os grandes conjuntos de dados SNP também podem ser usados para varredura do genoma em busca de regiões que possam ter sido alvo de seleção (NIELSEN, 2004).

A análise de componentes principais (PCA) é uma técnica multivariada capaz de atender à necessidade de avaliar a estrutura populacional em conjuntos de dados muito grandes (como é o caso de bancos de dados contendo muitos marcadores moleculares). A semelhança genética entre pares de indivíduos é quantificada e as

variáveis são resumidas nos componentes que melhor descrevem a variação dos dados, gerando gráficos de dispersão que fornecem uma avaliação visual direta das estruturas entre os grupos (JOMBART et al., 2010). As análises de heterozigosidades e estatísticas F de Wright também resultam em informações importantes no estudo populacional.

Desta maneira, este trabalho teve por objetivo estudar a estrutura da população bubalina leiteira no Brasil e identificar regiões genômicas que podem ter sido alvo de seleção, com auxílio de informações genômicas.

MATERIAL E MÉTODOS

Amostras e controle de qualidade dos marcadores

Dados genômicos de búfalas com aptidão leiteira, foram cedidos pela Faculdade de Ciências Agrárias e Veterinárias da Universidade Estadual Paulista (UNESP), campus Jaboticabal. Os animais foram genotipados com o chip 90K *Axion*® *Buffalo Genotyping* (101.761 marcadores).

Desta maneira, o banco de dados foi composto de sete fazendas distribuídas nos estados de São Paulo, Pará, Bahia e Rio Grande do Norte. Foram analisados 978 indivíduos, dentre os quais 480 pertencem à fazenda 1, 190 à fazenda 2, 80 à fazenda 3, 20 à fazenda 4, 6 à fazenda 5, 149 à fazenda 6 e 25 à fazenda 7.

O controle de qualidade sobre os dados brutos gerados pela genotipagem foi realizado por meio do *software* PLINK (Shaun Purcell, 2007), para verificação de alelos pouco informativos, correção de possíveis erros do processo de genotipagem e exclusão de marcadores situados nos cromossomos sexuais (para eliminar eventuais problemas entre proporção de machos e fêmeas). Nesta etapa, os dados foram filtrados com base na frequência do alelo menor (MAF) com limiar de 0,05; porcentagem de SNPs válidos na população (*call rate*) de 0,95 e equilíbrio de *Hardy-Weinberg* (EHW) de 10^{-6} . Para averiguar SNPs em desequilíbrio de ligação (DL), considerando $r^2 < 0,2$, utilizou-se o pacote *SNPRelate* (Zheng, 2012) do *software* R.

Após o controle de qualidade 5.119 marcadores foram removidos devido a não atenderem ao requisito de MAF, 34.736 marcadores foram removidos devido ao critério *call rate*, por apresentarem alta taxa de valores perdidos (*missing genotypes*) e 809 marcadores não aderiram ao EHW e foram eliminados. Quanto ao DL, foram excluídos 12.302 marcadores. Desta maneira, permaneceram no banco de dados 48.795 marcadores (autossômicos).

Análise de Componentes Principais (PCA)

Os dados foram dispostos inicialmente em uma matriz com tantos componentes quanto o número de variáveis existentes nos dados. Com a PCA gerou-se então um conjunto de dados em uma matriz reduzida, na qual foram consideradas as variáveis que proporcionaram maior variação dos dados. Desta maneira, foram obtidos os autovalores e autovetores da matriz de covariância dos dados. O autovetor com o maior autovalor associado correspondeu à componente principal e o autovetor com maior autovalor subsequente correspondeu à componente secundária. Os dados foram avaliados com auxílio de *scree plot* (Anexo 1) do pacote R *pcadapt* (LUU et al., 2016) para que fosse feita a seleção dos componentes que melhor explicassem a variação dos dados, sendo assim, estes foram representados pelas componentes principais 1 e 2. Então, as amostras foram plotadas em função da matriz reduzida em dois componentes (PC1 e PC2), sendo que a distância entre dois pontos quaisquer no espaço amostral representou a similaridade entre as amostras em relação às variáveis. Para a visualização gráfica da relação entre as amostras foi utilizado os pacotes *SNPRelate* e *gdsfmt* (Zheng et al., 2012), do ambiente R.

Posteriormente, foi feita a separação dos grupos identificados pela PCA em respectivos bancos de dados para a realização de estatísticas populacionais para cada subpopulação.

Diversidade genética e análise de diferenciação

Para medida de diversidade genética em cada grupo diferenciado pela PCA, foi utilizado o pacote *hierfstat* (Goudet, 2005) no R, para estimar os seguintes parâmetros genéticos populacionais para cada loco:

- Heterozigosidades:

Heterozigosidade observada (H_o) - é o número de heterozigotos em um determinado loco dividido pelo número total de indivíduos amostrados:

$$H_o = \frac{\text{número de heterozigotos em cada loco}}{\text{número total de indivíduos pesquisados}} \quad (1)$$

Heterozigosidade esperada (H_e) - indica a probabilidade de um indivíduo ser heterozigoto em um loco de acordo com suas frequências alélicas

$$H_e = 1 - \sum P_i^2 \quad (2)$$

em que, P_i é a frequência estimada do i ésimo alelo

- Índices de Fixação de Wright:

Dentro das subpopulações (F_{is}) – redução na heterozigiosidade em um indivíduo devido ao endocruzamento, em relação à população total à qual esse indivíduo pertence:

$$F_{is} = \frac{(H_o - H_e)}{H_e} \quad (3)$$

em que, H_o = heterozigiosidade observada; H_e = heterozigiosidade esperada

Entre as subpopulações (F_{st}) - redução de heterozigiosidade de uma subpopulação relativamente à população total refletindo na diferenciação genética entre subpopulações:

$$F_{st} = \frac{H_t - H_s}{H_t} \quad (4)$$

em que, H_t = heterozigiosidade esperada na população total com acasalamento ao acaso; H_s = heterozigiosidade média esperada dentro de subpopulações com acasalamento ao acaso

Assinatura de seleção

Para identificação de regiões genômicas supostamente sob seleção na população estudada foram estimados os valores de F_{st} para cada loco considerando as três subpopulações juntas, com auxílio do pacote LEA (Frichot & Francois, 2014). Para tanto, os valores de F_{st} foram convertidos em valores absolutos (z -scores). Posteriormente, a partir dos z -scores, foram calculados valores p ajustados em que para considerar um SNP como *outlier* foi considerado um limiar de 0,001, ou seja, os primeiros 0,1% mais significativos de acordo com o valor p ajustado para cada marcador. Optou-se por um valor de limiar mais rigoroso devido à ausência de correção para comparações múltiplas. Desta maneira, foi gerado o *Manhattan plot* com o pacote R qqman (TURNER, 2017), para visualização dos marcadores em relação a sua posição física em cada cromossomo.

RESULTADOS E DISCUSSÃO

A primeira componente (PC1) explicou 3,4% da variação dos dados, enquanto que a componente secundária (PC2) teve uma variância de 2,3% (Figura 1). Como no caso deste trabalho, dados fornecidos por marcadores moleculares, são altamente

dimensionais uma vez que cada marcador se comporta como uma variável, portanto não se deve esperar que PC1 e PC2 expliquem uma alta porcentagem da variância total, pois, quanto maior a quantidade de informações, mais difícil se torna a compactação dos dados em um pequeno número de dimensões (TEICH et al., 2014). Por exemplo, Mokhber et al. (2018), em seu trabalho realizado com 64.866 marcadores SNP de búfalos, obtiveram variação de 3,2% e 0,8% nas componentes principais 1 e 2, respectivamente. Ao passo que Mascioli et al. (2000), obtiveram variação explicada de 74,1% para a primeira componente e 13,9% para a segunda componente principal em estudo no qual avaliou 1.116 observações referentes a característica de peso (pesos ao nascimento, peso à desmama (240 dias), peso aos 12 meses (365 dias), peso aos 18 meses (550 dias) e peso aos 24 meses (730 dias) em bovinos da raça Canchim. O contraste na porcentagem da variação dos dados nos dois exemplos citados pode ser atribuído a diferença de quantidade e tipo de variáveis em cada caso.

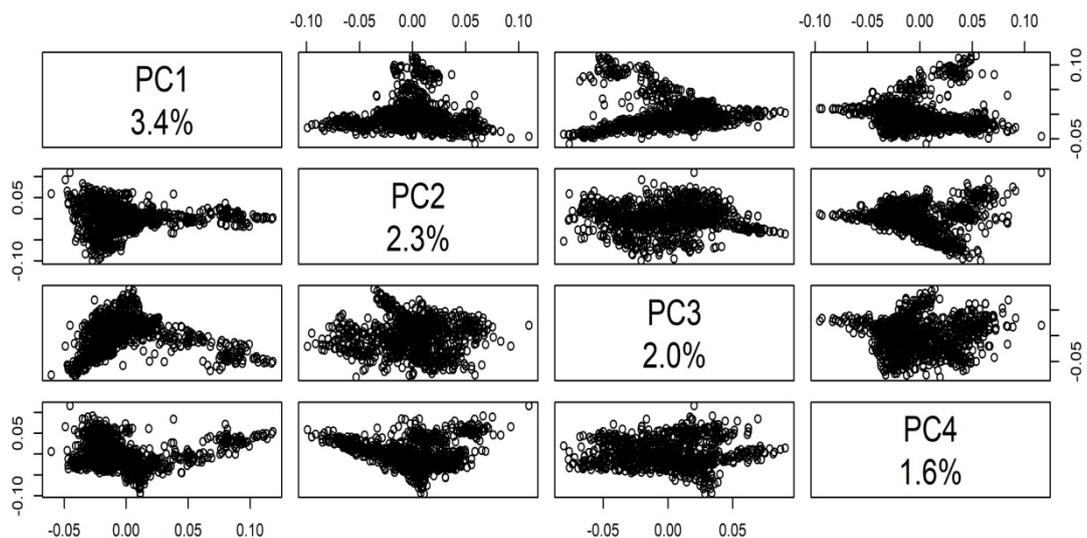


Figura 1 – Representação dos primeiros 4 componentes principais obtidos com os dados genômicos das populações bubalinas. Valores em porcentagem indicam a parcela de variância explicada pelos respectivos componentes principais. PC = *Principal Component* (componente principal).

A ordenação das amostras em função dos componentes que explicaram a maior porção da variação dos dados (PC1 e PC2) pode ser observada na Figura 2. Três grupos (ou subpopulações) foram distinguidos pela PCA sendo, Grupo 1 em verde, Grupo 2 em azul e Grupo 3 em rosa. Como pode ser visto o PC 1 separou a subpopulação 1 das demais, além, de demonstrar a variabilidade entre indivíduos da mesma subpopulação (considerando-se a dispersão dos pontos de mesma cor), sendo

notável a maior variabilidade na subpopulação 1. Enquanto que, o PC 2 fez a separação entre as subpopulações 2 e 3. Foram feitos testes para verificar se a distinção dos grupos poderia ser devido ao fator fazenda ou estado. Os resultados desses testes revelaram que não houve relação dos grupos com estes fatores, uma possível explicação é o pequeno tamanho do rebanho bubalino brasileiro que faz com que os animais procedam, muitas vezes, de um mesmo material genético, causando conectabilidade entre os rebanhos.

A distância entre dois pontos quaisquer no espaço amostral representa a similaridade entre as amostras em relação às variáveis, desta forma, a alta proximidade dos pontos na Figura 2 indica que as populações são bastante similares geneticamente, revelando uma mínima estrutura entre os grupos. Assim, uma provável explicação da baixa estrutura encontrada é que apesar das fazendas que compuseram o banco de dados deste estudo utilizarem seus próprios reprodutores, elas fazem intercâmbio de material genético entre as mesmas, propiciando fluxo gênico entre as subpopulações.

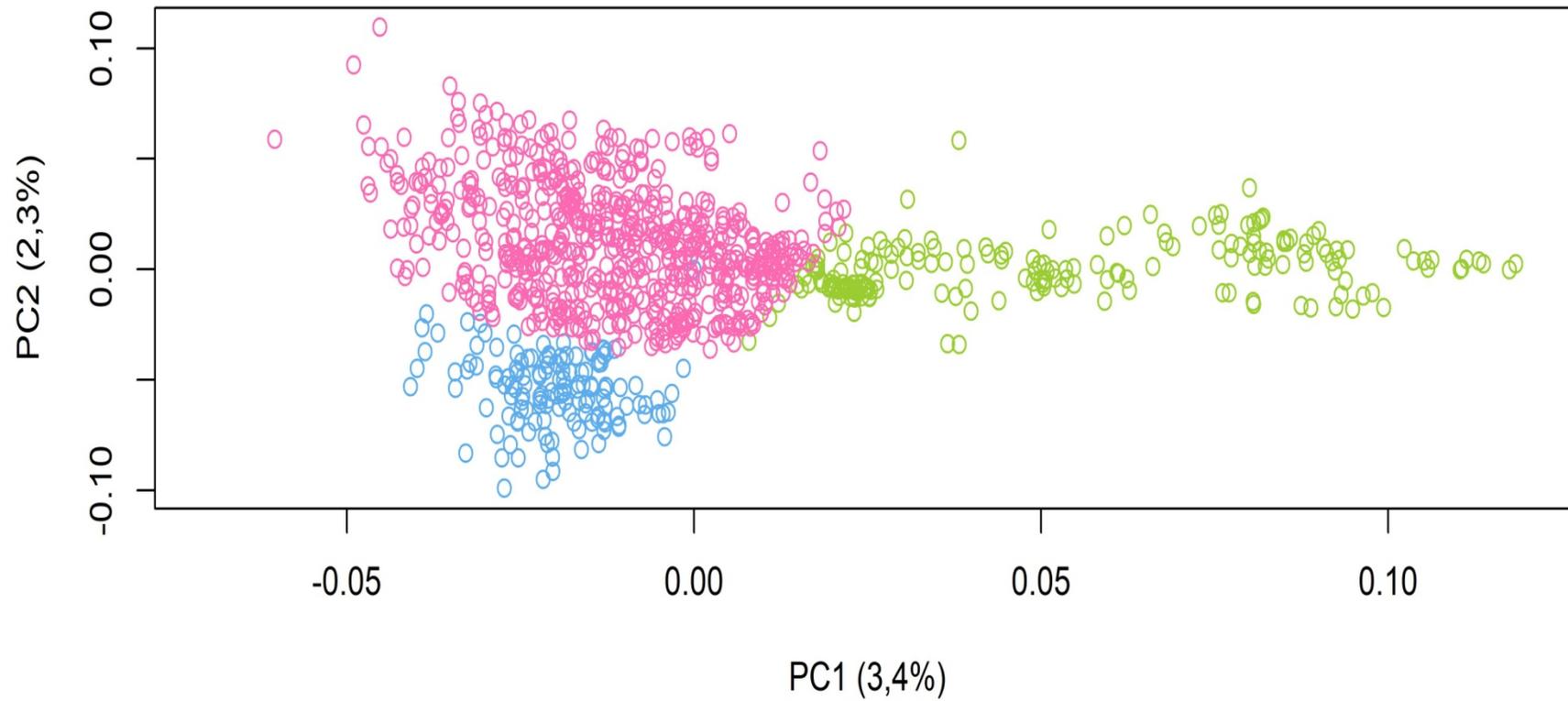


Figura 2 – Análise de componentes principais das populações de búfalos em estudo. Indivíduos do mesmo grupo populacional estão representados por pontos de mesma cor; subpopulação 1 (verde); subpopulação 2 (azul); subpopulação 3 (rosa).

Podem ser observados, na Tabela 1, os parâmetros utilizados para medida da diversidade genética dentro das subpopulações estudadas bem como de todos os grupos juntos (considerando uma população sem estrutura). Os valores médios de heterozigosidades, observada e esperada, variaram de 0,4143 a 0,4331 e 0,4119 a 0,4373 respectivamente. A taxa de heterozigotos foi considerada moderada a alta em comparação com a literatura de trabalhos com bubalinos onde foram encontrados valores ligeiramente abaixo e valores acima do encontrado neste estudo (porém, vale ressaltar que diferenças entre os valores da literatura podem ser devido a diferenças de metodologias e análises, número de animais como também diferenças intrínsecas aos próprios rebanhos). Os índices de heterozigosidades descritos por Pérez-Pardal et al. (2017), para búfalos de rio foram de $H_o=0,330$ e $H_e=0,279$. Marques et al. (2011) encontraram $H_o=0,5145$ e $H_e=0,5934$ em rebanho de búfalos Murrah brasileiros. Já Groeneveld et al. (2010), relataram valores de H_e entre 0,630-0,780 em búfalos indianos e 0,577-0,605 em búfalos de países mediterrâneos. Vieira et al. (2011), apresentaram valores de H_e variando de 0 a 0,861 em populações brasileiras de bubalinos da raça Murrah. Colli et al. (2018), atribuíram a diversidade genética na população brasileira da raça Murrah em seu estudo ($H_e=0,403$) à ancestralidade de bubalinos Murrah indianos combinada a cruzamentos com búfalos da raça Mediterrâneo, como também, esta diversidade pode ser devida aos efeitos de migração que decorreram ao longo do tempo na pós-domesticação das espécies, desta forma, sugere-se que a taxa de H_e encontrada neste trabalho, pode estar relacionada com o encontrado por esses mesmos autores.

Os índices de fixação intra-populacional (Fis) são apresentados na tabela 1, pode-se observar nos resultados do grupo 2, uma desproporção entre os valores de H_o e H_e ($H_o > H_e$), o que poderia ser indício da ocorrência de endogamia, porém, o índice Fis nesse grupo apresentou valor negativo (Fis = -0,0294). Valores de Fis negativos ou próximos de zero indicam variabilidade genética na população devido ao maior número de heterozigotos, indicando que nessa população ocorre a incidência de acasalamentos ao acaso, porém, se tratando de um sistema de produção espera-se que os acasalamentos ocorram de maneira direcionada como forma de seleção de características desejáveis por parte do produtor. Com isso, pode ser que este grupo ainda não pronunciou os efeitos de endogamia, mesmo com a diminuição da H_e . O grupo 1 apresentou baixa endogamia (Fis = 0,0096), enquanto o grupo 3 apresentou maior redução na heterozigosidade dos indivíduos devido ao endocruzamento (Fis = 0,0329). Os níveis de endogamia deste trabalho foram mais baixos do que o relatado

na literatura por Marques et al. (2011), que encontraram Fis médio de 0,0741 e Ángel-Marín et al. (2010) com valor médio de 0,0730, ambos em estudo com diversas raças de bubalinos.

Tabela 1 – Índices de diversidade genética*

Grupos	Ho	He	Fis
Grupo 1	0,4331	0,4373	0,0096
Grupo 2	0,4240	0,4119	-0,0294
Grupo 3	0,4143	0,4284	0,0329
Todos os grupos	0,4202	0,4362	0,0365

*Valores médios. Ho = Heterozigosidade observada; He = heterozigosidade esperada; Fis = índice de fixação intra-populacional

Uma estrutura populacional negligenciada pode gerar resultados falsos quanto aos parâmetros genéticos de diversidade, podendo prejudicar o delineamento de programas de melhoramento ou intercorrer em riscos para a conservação da variabilidade de espécies quando, por exemplo, um grupo é considerado altamente diverso, quando na verdade ele não é. Quando os grupos foram considerados pertencentes a uma só população, sem estrutura, os índices de heterozigosidades (observada e esperada) diferiram pouco de quando considerada a subestrutura da população (Tabela1), mas ainda assim pode-se observar que principalmente os grupos 2 e 3, foram considerados com maior variabilidade (He) do que realmente possuem. Nesta mesma situação, verificamos uma grande discrepância entre os valores de Fis considerando-se todos os grupos (Fis = 0,0365) e considerando o grupo 2 (Fis = -0,0294), de acordo com a Tabela 1.

A maior divergência genética foi observada entre os grupos 1 e 2 (Fst = 0,0291), seguido dos grupos 2 e 3 (Fst = 0,0163) e o menor valor de diferenciação foi entre os grupos 1 e 3 (Fst = 0,0152). Valores de Fst próximos de zero indicam uma situação sem divergência genética entre as subpopulações, enquanto que próximos de um indicam completa fixação de alelos diferentes entre os grupos. Os valores apresentados sugerem uma baixa estruturação da população (Fst<0,05). A variação entre os grupos representou, portanto menos que 3% da variação genética total (considerando que os animais são pertencentes a uma população com baixa estruturação, assim espera-se que valores de Fst sejam menores nesse caso que quando comparado entre raças), este resultado está de acordo com muitos estudos que confirmam que, em várias espécies, a variabilidade genética dentro das populações geralmente é maior que entre as populações (CARVALHO et al., 2016).

De acordo com Groeneveld et al. (2010), a maioria da diversidade genética em bubalinos encontra-se dentro de populações, e estimativas de variação entre populações chegam a cerca de 5,7% (comparando-se entre raças), os autores relatam que este valor é baixo comparando-se a outras espécies como bovinos (7,11%); equinos (8%) e suínos (13%), provavelmente pelo fato de que os búfalos não sofreram a mesma intensificação de isolamento e seleção artificial praticada na criação das raças de outras espécies domésticas. Hussain et al. (2017), atribuíram o baixo valor de F_{st} (0,07) entre 5 raças paquistanesas ao uso de um pequeno número de touros reprodutores, que pode ter contribuído para a ocorrência de pressão seletiva sobre os animais. Vijn et al. (2008), encontraram F_{st} de 0,0897 entre as raças Murrah e Jafarabadi em estudo de 12 raças de búfalos na Índia. Marques et al. (2011) e Colli et al. (2018), apresentam valores de 0,065 a 0,334 e 0,004 a 0,199 respectivamente.

O genoma da população foi analisado em busca de assinaturas de seleção pelo método F_{st} . A Figura 3 mostra a distribuição genômica dos valores de F_{st} por cromossomo e destaca os SNP considerados *outliers*.

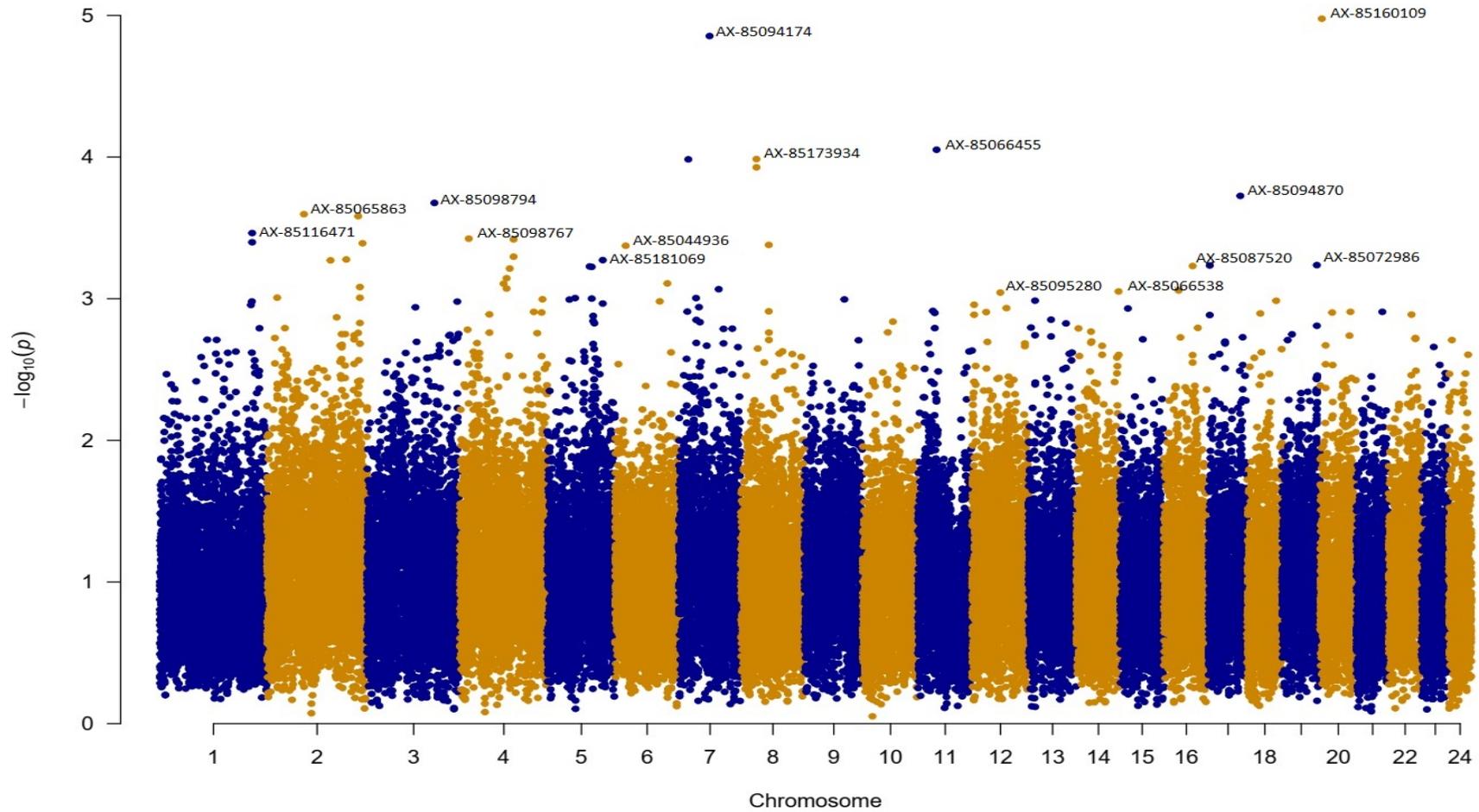


Figura 3 – Manhattan *plot* da distribuição genômica dos valores de Fst para cada marcador.

Foram identificadas 15 regiões genômicas altamente diferenciadas (valor p ajustado $<0,001$), essas regiões foram utilizadas para a pesquisa de genes que podem ter sido alvos de seleção. Os top 3 valores mais significativos de F_{st} foram encontrados nos cromossomos 20, 7 e 11, respectivamente (Tabela 3).

Tabela 3 – Regiões de alta diferenciação de acordo com a significância do valor de F_{st} (0,1%)

SNP	CHR	Posição	Valor de p	Gene
AX-85160109	20	3115672	1,05E+09	LOC112580945
AX-85094174	7	57667615	1,40E+09	APBB2
AX-85066455	11	34811117	8,88E+09	PELI2
AX-85173934	8	28913055	1,03E+04	ITGB8
AX-85094870	17	59617565	1,87E+04	HHIP
AX-85098794	3	126866790	2,10E+04	COL15A1
AX-85065863	2	69602279	2,53E+04	ZNF385B
AX-85116471	1	173954532	3,43E+04	LOC10240136
AX-85098707	4	16509162	3,76E+04	LAG3
AX-85044936	6	19783114	4,22E+04	GOLPH3L
AX-85181069	5	104048622	5,34E+04	SPA17
AX-85072986	19	65378070	5,79E+04	ADCY2
AX-85087520	16	54189655	5,88E+04	NA
AX-85066538	14	79428776	8,88E+04	LOC112578877
AX-85095280	12	52917729	9,05E+04	NA

SNP= *Single Nucleotide Polymorphism*; CHR= cromossomo; NA= nenhum dado encontrado nesta faixa

CONCLUSÃO

A população de bubalinos leiteiros deste estudo apresentou baixa estruturação genética conforme verificado pela análise de componentes principais, sendo que, os índices de diversidade dentro de cada grupo foram considerados moderados a altos. Quanto à diferenciação entre os grupos, foi constatada baixa divergência entre os mesmos indicando que as subpopulações (grupos) são bastante similares geneticamente. As regiões identificadas como altamente diferenciadas de acordo com o índice F_{st} forneceram informações a cerca de genes que podem estar sendo alvo de seleção na população. Estudos futuros a respeito das funções que esses genes expressam podem contribuir para o melhor entendimento da divergência entre os grupos.

REFERÊNCIAS BIBLIOGRÁFICAS

ALBRECHTSEN, A.; NIELSEN, F.C.; NIELSEN, R. Ascertainment biases in SNP chips affect measures of population divergence. **Mol. Biol. Evol.**, v.7(11), p.2534–2547, 2010. Doi:10.1093/molbev/msq148.

ÁNGEL-MARÍN, P.A; CARDONA, H.; MORENO-OCHOA, M.; CERÓN-MUÑOZ, M.F. Analysis of genetic diversity in Colombian buffalo herds. **Rev. Colomb. Cienc. Pecu.**, v. 23, p.411-421, 2010.

ASPILCUETA-BORQUIS, R.R.; DI PALO, R.; ARAUJO NETO, F.R.; BALDI, F.; DE CAMARGO, G.M.F.; ALBUQUERQUE, L.G.; ZICARELLI, L.; TONHATI, H. Genetic parameter estimates for buffalo milk yield, milk quality and mozzarella production and Bayesian inference analysis of their relationships. **Genet. Mol. Res.**, v.9 (3), p.1636-1644, 2010. Doi 10.4238/vol9-3gmr846.

COLLI L, MILANESI M, VAJANA E, IAMARTINO D, BOMBA L, PUGLISI F, DEL CORVO M, NICOLAZZI EL, AHMED SSE, HERRERA JRV, CRUZ L, ZHANG S, LIANG A, HUA G, YANG L, HAO X, ZUO F, LAI S-J, WANG S, LIU R, GONG Y, MOKHBER M, MAO Y, GUAN F, VLAIC A, VLAIC B, RAMUNNO L, COSENZA G, AHMAD A, SOYSAL I, ÜNAL EÖ, KETUDAT-CAIRNS M, GARCIA JF, UTSUNOMIYA YT, BARUSELLI PS, AMARAL MEJ, PARNPAI R, DRUMMOND MG, GALBUSERA P, BURTON J, HOAL E, YUSNIZAR Y, SUMANTRI C, MOIOLI B, VALENTINI A, STELLA A, WILLIAMS JL, AJMONE-MARSAN P. New Insights on Water Buffalo Genomic Diversity and Post-Domestication Migration Routes From Medium Density SNP Chip Data. **Front. Genet.** 9:53. 2018.

DU, C.; DENG, T.; ZHOU, Y.; YE, T.; ZHOU, Z.; ZHANG, S.; SHAO, B.; WEI, P.; SUN, H.; KHAN, F.A.; YANG, L.; HUA, G. Systematic analyses for candidate genes of milk production traits in water buffalo (*Bubalus Bubalis*). **Animal Genetics**, 2019. doi: 10.1111 / age.12739.

Ellegren, H.; Galtier, N. Determinants of genetic diversity. **Nature**, v.17, p.422-433, 2016. Doi: <http://dx.doi.org/10.1038/nrg.2016.58>.

FRICHOT, E; FRANCOIS, O. (2014). “LEA: an R package for Landscape and Ecological Association studies.”

GROENEVELD, L. F., LENSTRA, J. A., EDING, H., TORO, M. A., SCHERF, B., PILLING, D., NEGRINI, R., FINLAY, E. K., JIANLIN, H., GROENEVELD, E., WEIGEND, THE GLOBALDIV CONSORTIUM. Genetic diversity in farm animals – a review. **Animal Genetics**, 41 (Suppl. 1), 6–31 2010.

HUSSAIN, T.; BABAR, M.E.; ALI, A.; NADEEM, A.; REHMAN, Z.U.; M. MUSTHAFA, M.M.; MARIKAR, F.M.M.T. Microsatellite based genetic variation among the buffalo breed populations in Pakistan. **J. Vet. Res.**, v.61, p.535-542, 2017. Doi:10.1515/jvetres-2017-0057.

JOMBART, T.; Devillard, S.; Balloux, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. **BMC Genetics**, 11:94, 2010 .doi:10.1186/1471-2156-11-94.

MARQUES, J.R.F., MARTÍNEZ, A.M., COSTA, M.R., ALBUQUERQUE, M.S.M., QUIROZ, J., VEGA-PLA, J.L., DELGADO, J.V. Genetic Diversity Of Brazilian Buffaloes (*Bubalus Bubalis*) Using Dna Microsatellites. **Arch. Zootec.** 60 (232): 1213-1221. 2011.

MASCIOLI, A.S.; EL FARO, L.; ALENCAR, M.M.; FRIES, L.A.; BARBOSA, P.F. Estimativas de parâmetros genéticos e fenotípicos e análise de componentes principais para características de crescimento na raça Canchim. **Rev. bras. zootec.**, v. 29(6), p.1654-1660, 2000.

MOKHBER, M.; MORADI-SHAHRBABA, M.; SADEGHI, M.; MORADI-SHAHRBABA, H.; STELLA, A.; NICOLZZI, E.; RAHMANINIA, J.; WILLIAMS, J. L. A genome-wide scan for signatures of selection in Azeri and Khuzestani buffalo breeds. **BMC Genomics**, 19:449, 2018. Doi: <https://doi.org/10.1186/s12864-018-4759-x>.

NIELSEN, R. Population genetic analysis of ascertained SNP data. **Hum. Genomics.**, v.1(3), p.218–224, 2004. Doi: 10.1186/1479-7364-1-3-218.

PÉREZ-PARDAL, L. CHEN, S., COSTA, V., LIU, X., CAVALHEIRA, J., BEJA-PEREIRA, A. Genomic differentiation between swamp and river buffalo using a cattle high-density single nucleotide polymorphisms panel. **Animal**, 12:3, pp 464–471, 2018.

PURCELL, S.; NEALE, B.; TODD-BROWN, K.; THOMAS, L.; FERREIRA, M.A.R.; BENDER, D.; MALLER J.; SKLAR, P.; DE BAKKER, P.I.W.; DALY, M.J.; SHAM, P.C.

(2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. **American Journal of Human Genetics**, 81.

SENO, L.O.; CARDOSO, V.L.; EL FARO, L.; SESANA, R.C.; ASPILCUETA-BORQUIS, R.R.; DE CAMARGO, G.M.F.; TONHATI, H. Genetic parameters for milk yield, age at first calving and interval between first and second calving in milk Murrah buffaloes. **Livestock Research for Rural Development**, v.22, artigo 38, 2010.

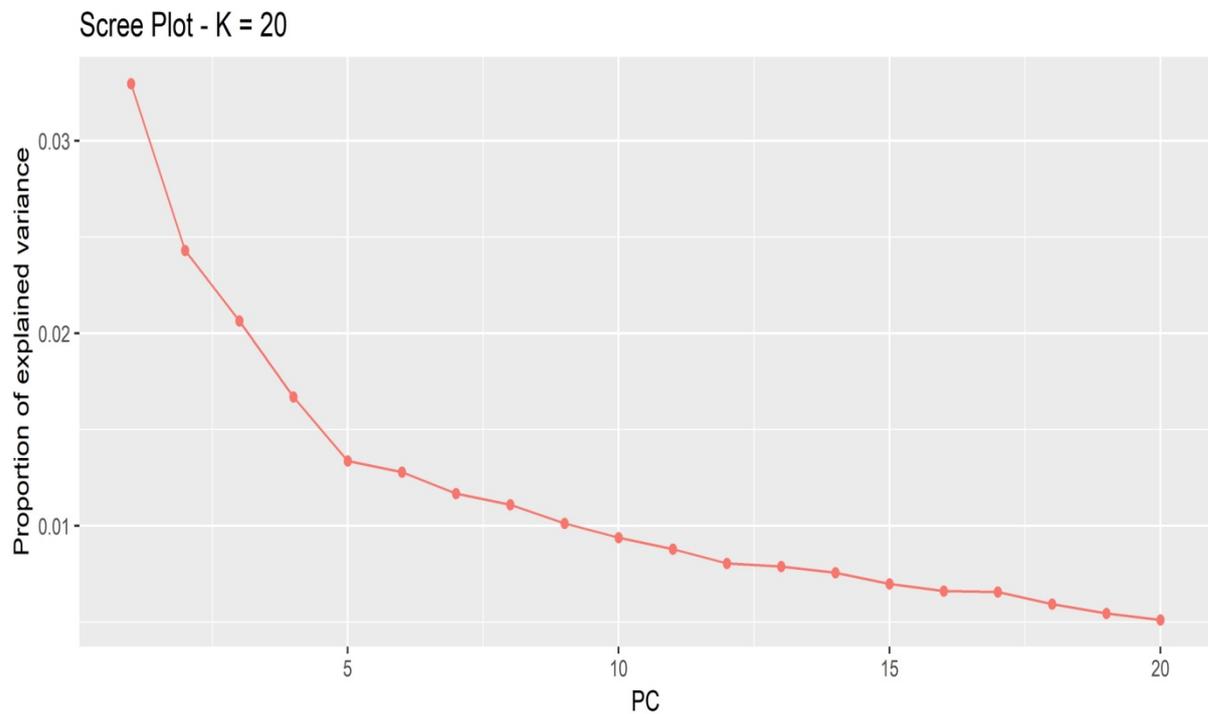
TEICH, I.; VERGA, A.; BALZARINI, M. Assessing spatial genetic structure from molecular marker data via principal component analyses: a case study in a *Prosopis* sp. forest. **Advances in Bioscience and Biotechnology**, v.5, p.89-99, 2014.

VIJH, R. K., TANTIA, M.S., MISHRA, B., BHARANI KUMAR, S. T. Genetic relationship and diversity analysis of Indian water buffalo (*Bubalus bubalis*). **J. ANIM. SCI.**, v.86, p.1495-1502, 2008.

ZHENG, X; LEVINE, D; SHEN, J; GOGARTEN, S; LAURIE, C; WEIR, B. (2012). A High-performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data. **Bioinformatics**, 28(24), 3326-3328. doi: [10.1093/bioinformatics/bts606](https://doi.org/10.1093/bioinformatics/bts606).

CONSIDERAÇÕES FINAIS

A bubalinocultura alcançou bastante espaço desde a sua introdução no Brasil, desempenhando um importante papel no cenário da pecuária leiteira destinada a produtos de alta qualidade. Por isso, o conhecimento do perfil genético do rebanho é importante, tanto para a avaliação de possíveis linhagens geneticamente superiores para produção quanto para a conservação dos recursos genéticos visando produções atuais e futuras. Para este conhecimento, o uso de ferramentas moleculares como, por exemplo, os painéis de SNP fornecem informações em um nível genômico a cerca das relações entre os indivíduos que podem ser consideradas em programas de melhoramento genético. Os resultados deste trabalho indicaram que a população estudada possui maior variabilidade dentro das subpopulações que entre as mesmas. Genes localizados em regiões de alta diferenciação ao longo do genoma foram anotados para que posteriormente (trabalhos futuros), sejam analisadas as suas respectivas funções no intuito de melhor entender o que contribuiu para a diferenciação genética entre as subpopulações estudadas.

ANEXO 1

Anexo 1 – *Scree plot* dos 20 primeiros componentes principais e suas respectivas porcentagens de variância explicada. Onde ocorre o decaimento mais abrupto entre os pontos (cotovelo) é o limite entre os componentes mais explicativos e menos explicativos.