

**UNIVERSIDADE ESTADUAL PAULISTA  
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS  
CÂMPUS DE JABOTICABAL**

**ANÁLISE GENÔMICA DA ESTRUTURA POPULACIONAL EM  
CAVALOS DA RAÇA BRASILEIRA MANGALARGA MARCHADOR**

**Bruna Aparecida dos Santos**  
Zootecnista

**2018**

**UNIVERSIDADE ESTADUAL PAULISTA  
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS  
CÂMPUS DE JABOTICABAL**

**ANÁLISE GENÔMICA DA ESTRUTURA POPULACIONAL EM  
CAVALOS DA RAÇA BRASILEIRA MANGALARGA MARCHADOR**

**Bruna Aparecida dos Santos**

**Orientador: Prof. Dr. Rogério Abdallah Curi**

**Coorientadores: Prof. Dr. Guilherme Luis Pereira**

**Prof. Dr. Júlio Cesar de Carvalho Balieiro**

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

**2018**

S237a

Santos, Bruna Aparecida dos

Análise genômica da estrutura populacional em cavalos da raça brasileira Mangalarga Marchador / Bruna Aparecida dos Santos. -- Jaboticabal, 2019

54 p. : tabs.

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

Orientador: Rogério Abdallah Curi

Coorientador: Guilherme Luis Pereira

1. Desequilíbrio de ligação. 2. Estruturas populacionais. 3. SNP. 4. Tamanho efetivo. I. Título.

CDU 636.082:636.1

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

Essa ficha não pode ser modificada.

**CERTIFICADO DE APROVAÇÃO**

**TÍTULO DA DISSERTAÇÃO: ANÁLISE GENÔMICA DA ESTRUTURA POPULACIONAL EM CAVALOS DA RAÇA BRASILEIRA MANGALARGA MARCHADOR**

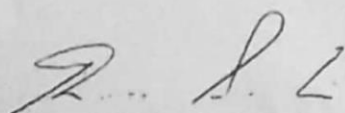
**AUTORA: BRUNA APARECIDA DOS SANTOS**

**ORIENTADOR: ROGERIO ABDALLAH CURI**

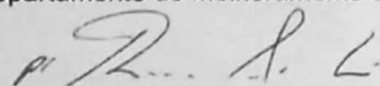
**COORIENTADOR: GUILHERME LUIS PEREIRA**

**COORIENTADOR: JULIO CESAR DE CARVALHO BALIEIRO**

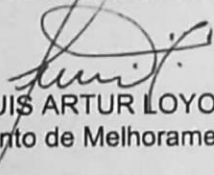
Aprovada como parte das exigências para obtenção do Título de Mestra em GENÉTICA E MELHORAMENTO ANIMAL, pela Comissão Examinadora:



Prof. Dr. ROGERIO ABDALLAH CURI  
Departamento de Melhoramento e Nutrição Animal / FMVZ / UNESP - Botucatu



Profa. Dra. LAURA LEANDRO DA ROCHA (Videoconferência)  
Departamento de Zootecnia/UFRPE / Recife/PE



Prof. Dr. LUIS ARTUR LOYOLA CHARDULO  
Departamento de Melhoramento e Nutrição Animal / FMVZ/UNESP - Botucatu

Jaboticabal, 28 de fevereiro de 2019

## **DADOS CURRICULARES DO AUTOR**

BRUNA APARECIDA DOS SANTOS – filha de Genivaldo Crispiliano dos Santos e Maria José da Silva Santos, nasceu em Vitória de Santo Antão – PE, em 12 de março de 1990. Formou-se em técnica em Agropecuária em 2007, pelo Instituto Federal de Pernambuco - Câmpus Vitória de Santo Antão. Ingressou no curso de Zootecnia em 2011 na Universidade Federal Rural de Pernambuco, Unidade Acadêmica de Serra Talhada (UFRPE/UAST), onde foi bolsista PIC e PIBIC, monitora de Ciências do solo e de Genética e melhoramento genético animal, fundou uma empresa Júnior (CIAGRO Jr.) junto com os colegas do curso de Zootecnia e Agronomia, ocupando a função de Diretora Vice-presidente e posteriormente Diretora presidente. Foi da Comissão de Orientação e Acompanhamento Acadêmico (COAA) da UFRPE/UAST, representando os alunos da graduação. Recebeu o título de Bacharela em Zootecnia em 23 de setembro de 2016. Em março de 2017 iniciou o mestrado em Genética e Melhoramento Animal (PPGMA), na Faculdade de Ciências Agrárias e Veterinárias da Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP), Câmpus de Jaboticabal, sob orientação do professor Rogério Abdallah Curi e coorientação dos professores Guilherme Luis Pereira e Júlio Balieiro.

## AGRADECIMENTOS

À Deus, o autor da minha vida, dono de toda ciência, sabedoria e poder. Toda honra e toda glória sejam dadas a ti Senhor, que me protege, abençoa e ampara em todos os momentos, principalmente nos mais difíceis.

À minha amada família, meus pais Genivaldo e Maria, e minhas irmãs Juliana e Vanessa. Vocês são o meu pilar, minha fortaleza, meu maior amor e a minha maior saúde. Obrigada por se fazerem presente mesmo estando tão distante, pelo amor incondicional, carinho, suporte, compreensão, torcida, por confiarem tanto em mim e sonharem junto comigo, e também por entenderem a minha ausência.

Agradeço a minha amiga Érika Nayara, por toda amizade, carinho, compreensão, por me receber inúmeras vezes em sua casa quando precisei, por cuidar de mim, pelas orações, por caminhar comigo e segurar minha mão desde a graduação. Você é a família que Deus me permitiu escolher, te amo.

À minha amiga Fernanda Rodrigues (Fer), por todos os momentos bons e ruins que passamos juntas e superamos de cabeça erguida, por todas conversas, risadas, cantorias e danças, obrigada por ouvir meus desabafos e por cuidar de mim, me mandar colocar o casaco (1 milhão de vezes), me dar remédio, me aconselhar...em fim, sou grata por ter convivido com você durante esses anos, minha grande parceira de quarto. Vou te levar pra sempre no meu coração.

Minhas meninas da República CP de Jaboticabal, e do Bond 858 de Botucatu, Bru e Nadja, e minha “cãopanheira” Weed, a dog mais linda e carinhosa, muito obrigada por me acolherem em suas casas, pela convivência maravilhosa que tivemos, pela amizade, confidências, pelas milhares gargalhadas, pelos momentos de descontração, companheirismo, por amenizarem a minha dor de estar longe de casa, e por tudo que passamos juntas, minha eterna gratidão. Vocês são muito especiais. Lar é onde o amor está, e vocês me fizeram sentir que eu vivia em um.

Ao meu amigo de pós-graduação Fernando, por todas as etapas do projeto que desenvolvemos juntos, pela parceria, e amizade que foi construída, por me receber em sua casa, pelas conversas e cumplicidade trocadas.

Agradeço ao meu orientador Rogério Curi e coorientador Guilherme Pereira, pela orientação, ensinamentos, ajuda e por estarem sempre presentes durante todo o

desenvolvimento deste trabalho. Ao meu coorientador Júlio Balieiro, pela ótima recepção na USP de Pirassununga, por se preocupar com a minha estadia, e pela disponibilidade em ajudar no que fosse preciso.

Obrigada a todos os professores, amigos e parentes que fizeram parte de toda minha trajetória, que contribuíram de alguma maneira, direta ou indiretamente para que eu pudesse chegar até aqui, muito obrigada a todos.

Ao Programa de Pós-Graduação em Genética e Melhoramento Animal (PPGMA), pela oportunidade de cursar o mestrado, foi uma honra poder fazer parte dessa renomada instituição. Não poderia deixar de agradecer ao prof. Dr. Danisio Prado Munari, por sua eficiência e colaboração durante meu processo seletivo, pois precisei fazer a seleção em Pernambuco.

Agradeço aos criadores e seus animais pela concessão do material genético para realização desse estudo. À organização da Exposição Nacional do Mangalarga Marchador (2017), que permitiu nossa entrada no Parque da Gameleira, para realização das últimas coletas.

À médica que me atendeu no posto de saúde no final do ano de 2017, obrigada pelo profissionalismo, por diagnosticar meu problema de forma precisa, e por me fazer entender que minha saúde deveria ser minha prioridade.

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

## SUMÁRIO

CAPÍTULO 1 – CONSIDERAÇÕES GERAIS.....	1
1. INTRODUÇÃO.....	1
2. OBJETIVOS.....	3
3. REVISÃO BIBLIOGRÁFICA.....	3
3.1. Panorama da equinocultura mundial.....	3
3.2. O Mangalarga Marchador.....	5
3.3. Chips de genotipagem de SNP para equinos.....	7
3.4. Desequilíbrio de ligação e tamanho efetivo populacional.....	9
3.5. Estrutura e subestruturas populacionais.....	11
4. BIBLIOGRAFIA CITADA.....	12
CAPÍTULO 2 – Análise genômica da estrutura populacional em cavalos da raça brasileira Mangalarga Marchador.....	20
1. INTRODUÇÃO.....	21
2. MATERIAL E MÉTODOS.....	22
2.1. Animais.....	22
2.2. Coleta de sangue e extração do DNA.....	23
2.3. Genotipagem e controle de qualidade.....	23
2.4. Imputação entre os arranjos de 65k e 670k.....	24
2.5. Estimação do desequilíbrio de ligação e do tamanho efetivo da população.....	25
2.6. Estrutura e subestruturas populacionais.....	26
3. RESULTADOS.....	27
3.1. Avaliação do arranjo de 670k e imputação dos genótipos.....	27
3.2. Desequilíbrio de ligação.....	28
3.3. Estimação do tamanho efetivo da população.....	32
3.4. Estrutura e subestruturas populacionais.....	32
4. DISCUSSÃO.....	35
5. CONCLUSÃO.....	39
6. BIBLIOGRAFIA CITADA.....	39



## **ANÁLISE GENÔMICA DA ESTRUTURA POPULACIONAL EM CAVALOS DA RAÇA BRASILEIRA MANGALARGA MARCHADOR**

**RESUMO** – O Mangalarga Marchador é o cavalo de sela brasileiro, possui dois tipos de andamentos característicos, a marcha batida e a marcha picada, que proporcionam maior comodidade ao cavaleiro durante a cavalgada e o trabalho. É principalmente utilizado para trabalho em fazendas de gado de corte e vem se destacando em diferentes modalidades de esportes hípicas. Este estudo teve como objetivo caracterizar, por meio da genotipagem de SNP em larga escala, o desequilíbrio de ligação (LD), calculado por  $r^2$ , de equinos da raça brasileira Mangalarga Marchador criados no Brasil. Também foi investigado o tamanho efetivo ( $N_e$ ) da população, bem como as suas estruturas e relações. Foram utilizados 240 equinos Mangalarga Marchador, de ambos os sexos, e registrados na associação brasileira de criadores da raça (ABCCMM). O número de SNP informativos foi de 377.308. Análises de componentes principais mostraram que cavalos Mangalarga Marchador de marcha batida e de marcha picada pertencem a uma mesma população, ou seja, estes grupos não segregaram de forma significativa dentro da raça, o que deve ser levado em consideração nos estudos genético-populacionais. O  $r^2$  genômico calculado foi de  $0,096 \pm 0,166$ . O LD decaiu consideravelmente a partir de distâncias superiores a 15 e 20 Kb, apresentando valores inferiores a 0,3 e 0,2, respectivamente. O  $N_e$  atual foi de 99 animais. Houve acentuada redução neste parâmetro ao se tomar as estimativas de 16 gerações passadas, em que o  $N_e$  estimado foi de 650 animais. Estes resultados podem estar ligados à uma base genética ampla e parcialmente aberta e à pressões de seleção crescentes na raça. Por fim, observou-se subestruturas populacionais definidas relacionadas mais à importantes reprodutores da raça do que propriamente ao tipo de marcha.

**Palavras-chave:** desequilíbrio de ligação, estruturas populacionais, SNP, tamanho efetivo

## **GENOMIC ANALYSIS OF THE POPULATION STRUCTURE IN HORSES OF THE BRAZILIAN MANGALARGA MARCHADOR BREED**

**ABSTRACT** – The Mangalarga Marchador is the Brazilian saddle horse, has two types of characteristic movements, the batida and the picada gait, that provide greater comfort to the rider during the cavalcade and the work. It is mainly used for work on beef cattle farms and has been emphasizing different modalities of equestrian sports. The objective of this study was to characterize, by means of large scale SNP genotyping, the linkage disequilibrium (LD), calculated by  $r^2$ , of Brazilian Mangalarga Marchador breed horses raised in Brazil. We also investigated the effective size ( $N_e$ ) of the population, as well as their structures and relationships. A total of 240 Mangalarga Marchador horses, of both sexes, and registered in the Brazilian association of breeders (ABCCMM) were used. The number of informative SNPs was 377,308. Principal component analyzes showed that Mangalarga marchador of the two diferente gaits belong to the same population, that is, these groups did not segregate significantly within the breed, which should be taken into account in the genetic-population studies. The calculated  $r^2$  genomic was  $0.096 \pm 0.166$ . The LD declined considerably from distances greater than 15 and 20 Kb, presenting values lower than 0.3 and 0.2, respectively. The current  $N_e$  was 99 animals. There was a marked reduction in this parameter when taking the estimates of 16 generations passed, in which the estimated  $N_e$  was 650 animals. These results may be linked to a broad and partially open genetic basis and to increasing selection pressures in the breed. Finally, we observed defined population substructures related to the breed's important breeders rather than to the type of gait.

**Keywords:** linkage disequilibrium, population structures, SNP, effective size

## CAPÍTULO 1 – CONSIDERAÇÕES GERAIS

### 1. INTRODUÇÃO

A raça Mangalarga Marchador, tipicamente brasileira, que surgiu há cerca de 200 anos na Comarca do Rio das Mortes, no sul de Minas Gerais, por meio de acasalamentos de cavalos da raça Alter, trazidos da Coudelaria de Alter do Chão, em Portugal, com éguas nativas selecionadas pelos criadores daquela região mineira. A base de formação dos cavalos Alter foi a raça espanhola Andaluza, cuja origem étnica vem de cavalos nativos da Península Ibérica, germânicos e berberes. Os cruzamentos dessas raças deram origem a animais de porte elegante, beleza plástica, temperamento dócil e próprios para a montaria (ABCCMM, 2018).

Desde a definição do padrão racial, esses animais têm sido avaliados por técnicos credenciados no que diz respeito à sua conformação e qualidade como animais de sela, principalmente em relação ao seu andamento, marcha batida ou marcha picada (ANDRADE, 1984; MEIRA, 2010). Assim, o cavalo marchador deve ser portador, em seu genótipo, do gene da marcha e há de apresentar, em seu fenótipo, as angulações ósseas e as proporções lineares específicas do marchador (SANTIAGO, 2013).

O melhoramento animal implica em manipulação genética com o envolvimento de vários aspectos do ser vivo. A seleção do Mangalarga Marchador, por exemplo, com o objetivo de melhorar seu desempenho na comodidade do passeio, de apresentar uma marcha de qualidade e ter um excelente rendimento nas provas hípicas, não é tão simples de ser conseguido, pois é difícil reunir todas essas características. A maior causa dessa dificuldade é a baixa herdabilidade do desempenho atlético e os valores de correlações genéticas existentes entre as características mais almejadas nos cavalos (REGATIERI; MOTA, 2012).

A genômica trouxe soluções para esses problemas, com o desenvolvimento de ferramentas e tecnologias que auxiliam o melhoramento genético. A genética molecular, utiliza diretamente as informações do DNA na seleção, de forma a permitir eficiência seletiva, rapidez na obtenção de ganhos genéticos e baixo custo, em comparação com a tradicional seleção baseada apenas em dados fenotípicos, além de ser eficiente para características de baixa herdabilidade (RESENDE et al., 2008). Os métodos de sequenciamento de nova geração (*next generation sequencing* –

NGS), permitiram a identificação em massa de polimorfismos de DNA distribuídos ao longo dos cromossomos de indivíduos da mesma espécie, resultando no desenvolvimento dos arranjos de SNP, também chamados SNP chips de genotipagem em larga escala (BELTRÁN, 2014). Esses polimorfismos têm possibilitado, entre outros, o estudo da estrutura genética de diferentes populações em várias espécies de animais domésticos (KIJAS et al., 2009; McKAY et al., 2008; GIBBS; TAYLOR; VAN TASSELL, 2009).

Parte significativa destas aplicações requerem o conhecimento ou acesso ao desequilíbrio de ligação (*linkage disequilibrium* – LD) existente entre os marcadores moleculares em dada população da espécie de interesse. O LD é a associação não aleatória entre alelos de diferentes locos do mesmo cromossomo, ou seja, é quando a frequência de um determinado haplótipo é maior do que seria esperado (ARDLIE; KRUGLYAK; SEIELSTAD, 2002; KHATKAR et al., 2006; GODDARD; HAYES et al., 2009).

No nível molecular, a deriva aleatória afeta as frequências alélicas, leva à perda de variação genética neutra, e à fixação de alelos favoráveis ou deletérios. Uma abordagem que auxilia na predição dessas perdas é a determinação do tamanho efetivo da população ( $N_e$ ), definido como o conjunto de indivíduos de uma população que se reproduzem e conseguem deixar descendentes (FALCONER, 1987). Uma abordagem promissora e independente da integralidade do *pedigree* é estimar o  $N_e$  com base em padrões de LD (SVED, 1971; HILL, 1981; HAYES et al., 2003) de milhares de SNP distribuídos ao longo do genoma. Equações deterministas derivadas por Daetwyler et al. (2009) mostraram que, uma vez que o  $N_e$  para uma população é conhecido, a precisão da seleção genômica para uma variedade de cenários pode ser calculada, facilitando assim o entendimento do impacto das estratégias seletivas de reprodução na variação genética presente nas populações, e pode fornecer uma visão do nível de endogamia em populações para as quais os *pedigrees* estão incompletos ou indisponíveis.

O monitoramento do progresso genético deve ser uma preocupação para toda e qualquer raça (EUCLIDES FILHO et al., 2000). Assim, o conhecimento da variabilidade e da evolução genética de uma população tem importância não só para proceder as modificações que se fizerem necessárias, mas também para a avaliação do programa de seleção que vem sendo adotado.

## 2. OBJETIVOS

Com base no exposto e considerando a escassez de informações genômicas referentes à raça brasileira de cavalos Mangalarga Marchador, o objetivo deste estudo foi entender, por meio de informações moleculares, particularidades da estrutura da população de equinos da raça. Isto foi feito pela caracterização, por meio da genotipagem de SNP em larga escala, do desequilíbrio de ligação (LD) de cada cromossomo e do genoma como um todo. Investigou-se também o decaimento dos LDs, ou seja, a diminuição do LD em função do aumento da distância física entre marcadores. Ainda com a informação do LD, estudou-se o tamanho efetivo da população ( $N_e$ ). Por fim, investigou-se subestruturas familiares importantes dentro da população.

## 3. REVISÃO BIBLIOGRÁFICA

### 3.1. Panorama da equinocultura mundial

O número de equinos diminuiu ao longo dos últimos 70 anos, mas o efetivo permaneceu estável nas últimas décadas. A população global de cavalos (*Equus caballus*) foi reduzida em 1950 para 75 milhões, e em 2007 para 58,9 milhões de cabeças (FAO,2014), sendo que a maior parte vive na América, Ásia e alguns países da Europa. Os Estados Unidos é o país com maior número de cabeças (10.260.000), seguido do México (6.355.000) China (6.029.780) e Brasil (5.450.601) (GLIPHA, 2017).

A tropa nacional é superior a cinco milhões de cavalos, e inclui os cavalos de lazer e esporte (1.100.00 cabeças) e de trabalho (3.900.00 cabeças). Estes animais estão distribuídos pelo país inteiro, e os estados com os maiores efetivos são: Minas gerais, ocupando a primeira colocação com 758.880 cabeças, Rio Grande do Sul (535.299 cabeças), Bahia (485.356), Goiás (394.799) e em quinto lugar São Paulo com 347.411 cabeças (MAPA, 2016).

Os agronegócios envolvidos com a criação e a utilização do cavalo ocupam relevante posição tanto nos países desenvolvidos como nos que estão em desenvolvimento, incluindo o Brasil. No país, o complexo do agronegócio do cavalo gera valores superiores a R\$ 16,15 bilhões ao ano e a atividade ainda responde por

aproximadamente 3,2 milhões de empregos diretos e indiretos. O setor de maior destaque é o de utilização dos cavalos na lida com o rebanho, que responde por 8,58 bilhões, o que corresponde cerca de 53% da renda gerada pelo complexo do agronegócio do cavalo e 71,3% dos empregos diretos gerados. Salienta-se a importância do equino na tração animal, que de acordo com o Censo Agropecuário de 2006, cerca de 25% da força total (animal e mecânica) utilizada nas propriedades rurais, ainda é oriunda da tração animal (IBGE, 2016).

Outros setores têm importante participação na equinocultura brasileira, entre os quais os animais dos seguimentos de esporte e lazer, que contribuem com 125.700 ocupações diretas e movimentação econômica de R\$ 5,84 bilhões (MAPA, 2016). No período de 1999 a 2004 o número de eventos envolvendo as diversas modalidades equestres (salto, adestramento, concurso completo de equitação, enduro, vaquejada, competições de marcha, etc.) cresceu 315%, ou seja, aumento médio de 15,3% ao ano (LIMA et al., 2006). O crescimento verificado nesse seguimento desde a realização deste estudo, em que o plantel foi estimado em 800 mil animais, foi de 37% (MAPA, 2016).

A equoterapia é outra atividade que vem se destacando e trazendo benefícios tanto econômicos como sociais. O cavalo é utilizado como recurso terapêutico no tratamento de portadores de dificuldades nas áreas cognitiva, psicomotora e sócio-afetiva. O Brasil possui cerca de 280 centros de equoterapia distribuídos por todo território nacional, a maior parte concentrada nas regiões Sul e Sudeste. São atendidos seis mil praticantes, que pagam valores diferentes de acordo com o centro e a condição econômica do praticante. Os Centros de Equoterapia contam com infraestrutura física, profissionais de diversas formações, e animais, com movimentação financeira anual estimada em R\$ 80 milhões por ano (ANDE-BRASIL, 2018).

Alguns segmentos dentro do agronegócio do cavalo, muitas vezes não são mencionados e divulgados, embora sua importância para economia do país seja marcante. Como por exemplo, o segmento de rações (que gera movimentação econômica de R\$ 53,4 milhões), medicamentos veterinários (R\$ 54,1), selaria (R\$ 174,6), casqueamento e ferrageamento (R\$ 143,6), veterinários (R\$ 20,0), transporte de equinos (R\$ 86,4), vaquejada (R\$ 164,0), turismo equestre (R\$ 21,0), escolas de equitação (R\$ 78,0), exposições e eventos (R\$ 146,1) e leilões (R\$ 54,9) (LIMA et al., 2006). Assim, observa-se a expressividade do cavalo na contribuição para geração

de renda e empregos no Brasil, tendo papel considerável no desenvolvimento do país e na economia.

### **3.2. O Mangalarga Marchador**

Os equinos foram introduzidos no Brasil pelos colonizadores portugueses e em 1808, com a vinda da família Real, chegaram os tradicionais cavalos selecionados pela Real Coudelaria de Alter. Pela tradição, o príncipe regente D. João VI presenteou Gabriel Francisco Junqueira, o Barão de Alfenas, com um garanhão da raça Alter. Este e outros cavalos da coudelaria foram acasalados com éguas nativas formando a base do Mangalarga Marchador (ABCCMM, 1991; CASIUCH, 1997). A raça foi selecionada em Minas Gerais para andamento e morfologia específicos (MOTA; PRADO, 2006). Outros cavalos da raça Alter foram importados para serem usados como melhoradores nos animais nacionais, a fim de refinar a morfologia e gerar animais de aparência mais nobre (ARAÚJO, 2013). A base do Mangalarga Marchador é, portanto, de animais ibérico-nacionais inter cruzados com infusão de genética Alter que, posteriormente, sofreu influência de genética Andaluz, Árabe e Quarto de Milha (CARNEIRO; DUARTE; DIAS, 1991; COSTA et al., 2004; De ASSIS et al., 2009).

Os criadores mineiros, principalmente os Junqueira, descendentes do Barão de Alfenas, continuaram a selecionar o Mangalarga Marchador, e formaram importantes fazendas que constituíram a base inicial da formação da raça, como a fazenda Campo Alegre, Campo Lindo, Traituba, Favacho, Engenho de Serra e Fazenda Angahy (ANDRADE, 1984). De acordo com Costa et al. (2005), a atual população da raça Mangalarga Marchador é constituída, em ordem de importância, por cinco grupos genéticos descendentes dos animais Herdade Alteza e Seta Caxias, Providência Itu e Tabatinga Predileto, Abaíba Marengo, Tabatinga Cossaco e Angaí Miron. A linhagem Catuni também trouxe importante contribuição para a raça, que começou a seleção com o garanhão Casemiro Colares, que se destacou pela qualidade da tropa de cavalos de sela que produziu. Em 1950, a fazenda Favacho adquiriu o garanhão Pedra Estanho, que permaneceu na mesma por 18 anos, com grande parte dos animais do plantel atual sendo seus descendentes (CASIUCH, 1997).

Segundo Costa et al. (2004), a raça Mangalarga Marchador distribuiu-se por todos os estados da federação, exceto Roraima e Amapá. Por ser o berço da raça, Minas Gerais concentra 39,1% do contingente nacional, representando 12,2% do

rebanho equino do estado. Ainda de acordo com os referidos autores, os estados de Minas Gerais, Rio de Janeiro, São Paulo e Bahia detêm 80,0% de todos os animais registrados e, apesar do grande número de criadores e proprietários, a maioria deles é detentora de pequenos plantéis. De acordo com a associação de criadores (ABCCMM, 2018), o efetivo da raça no Brasil está em torno de 480.000 cabeças.

O padrão racial definido pela ABCCMM, que desde 1949 mantém os livros de registros da raça, baseia-se na aparência externa e na mensuração da altura da cernelha dos animais, nos quais a presença do andamento marchado é obrigatória e sua ausência é motivo de desclassificação. Assim, são animais que se caracterizam por apresentar porte médio, agilidade, estrutura forte e bem proporcionada, expressão vigorosa e sadia, visualmente leve na aparência, pele fina e lisa, pelos finos e lisos, temperamento ativo e dócil (MEIRA, 2010).

De acordo com Andrade (2011), a principal característica dos cavalos Mangalarga Marchador é a marcha. As marchas batida e picada são os únicos andamentos naturais de velocidades intermediárias permitidos na raça e os cavalos são classificados de acordo com a marcha pela avaliação visual dos técnicos oficiais da ABCCMM no momento do registro do animal (ABCCMM, 2018). A marcha picada é uma marcha lateral a quatro tempos, caracterizada, quando de boa qualidade, por um tríplice apoio contínuo intercalado de apoios bipedais diagonais e laterais (MOREIRA, 1994). A marcha batida é um andamento de média velocidade, a quatro tempos, com deslocamentos alternados de bípedes em lateral e em diagonal, registrando em cada passada quatro apoios tripedais, onde dois são apoios bipedais diagonais e, dois apoios bipedais laterais (MOREIRA, 1994). Há relatos de que o comprimento de sua passada está próximo de 2,0 metros e a velocidade, entre 12 e 14 km/h (NASCIMENTO, 1999).

A raça Mangalarga Marchador vem ganhando espaço e grande importância no país e no mundo, e a comprovação disso são os números expressivos que ostenta. Como por exemplo, é considerada a maior entidade de criadores de equinos de uma mesma raça da América Latina, com mais de 15 mil associados, tendo 70 núcleos e associações de criadores nos principais estados. É considerada oficialmente “a raça nacional”, depois que a Lei nº 12.975 foi sancionada em 19 de maio de 2014. No exterior, tem representações oficiais na Alemanha, Itália, Estados Unidos e Argentina. A ABCCMM chega a realizar cerca de 240 eventos nos diversos Estados do País, que geram 40 mil empregos diretos e mobilizam, indiretamente, 200 mil pessoas no



campo. A 37ª exposição nacional, por exemplo, realizada em 2018, movimentou R\$ 25 milhões com leilão e vendas diretas entre os criadores. (ABCCMM, 2018).

Devido ao seu crescimento e participação no cenário nacional e mundial, a raça Mangalarga Marchador vem despertando o interesse de pesquisadores em diferentes áreas. Existem poucos trabalhos relacionados com genética e estudos populacionais da mesma, e os que já estão disponíveis na literatura, são em sua grande maioria referentes a herdabilidade e correlações fenotípicas (ZAMBORLINI et al.,1996), endogamia e tamanho efetivo, mas com a utilização de dados de pedigree (COSTA et al, 2005; GONÇALVES et al.,2012), estimação de parâmetros genéticos (MEIRA, 2010) e estrutura populacional, também com dados de pedigree (MOTA; PRADO; SOBREIRO, 2006; COSTA et al, 2005). São escassas as informações genômicas e a utilização da genotipagem de variantes do DNA para os cavalos da raça brasileira Mangalarga Marchador.

A importância das informações obtidas com a genotipagem simultânea de milhares de SNP é facilitar muitas aplicações da genética em equinos, incluindo a identificação de genes envolvidos com características de desempenho, sanidade e estudos sobre as origens do cavalo doméstico, a diversidade dentro das raças e as relações evolutivas entre espécies relacionadas. Esse recurso permitiu que o cavalo se tornasse viável como modelo animal para pesquisas genéticas. Os Pesquisadores que trabalham diretamente com equinos agora possuem informações importantes e essenciais para avaliar a estrutura do genoma dentro e entre as raças de cavalo, bem como a de espécies relacionadas (MCCUE et al, 2012).

### **3.3. Chips de genotipagem de SNP para equinos**

À medida que as sequências de nucleotídeos dos genomas foram sendo desvendadas, observou-se grande número de variações de uma base encontradas ao se comparar segmentos correspondentes do mesmo genoma, ocorrendo, aproximadamente, a cada 500 bases (PEREIRA, 2017).

Essas alterações, quase sempre bialélicas, podem ser provocadas por erros de incorporação de bases durante a replicação do DNA ou em outros casos, são causadas por agentes ambientais. Caso essas alterações ocorram em células germinativas e sejam transmitidas às gerações seguintes passam a ser denominadas mutações pontuais herdáveis. Caso fixem-se na população a uma frequência mínima

de 0,01, passam a ser denominadas de polimorfismos de nucleotídeo único (*single nucleotide polymorphisms* – SNP) (KWOK; GU, 1999).

Os SNP podem ocorrer em regiões gênicas codificadoras ou com função regulatória, bem como em espaços intergênicos. Em regiões codificadoras e regulatórias, podem provocar a substituição de aminoácido na sequência proteica ou alterar a quantidade de proteína expressa. Nesses casos, pode haver modificações fenotípicas. Embora a função da maior parte das regiões intergênicas ainda seja desconhecida, a sua importância no controle da regulação da expressão gênica é cada vez mais conhecida. Neste sentido, polimorfismos localizados nestas regiões também podem estar associados à variações no fenótipo (GUIMARÃES; COSTA, 2002).

Estudos em humanos (VENTER et al., 2001; LANDER et al., 2001) e em espécies de interesse zootécnico, como os equinos (WADE et al., 2009; DOAN et al., 2012), mostraram a ocorrência de milhões SNP ao longo do genoma de um indivíduo (*Human Genome Project Information, The SNP Consortium LTD, Bovine Genome Sequencing and Analysis Consortium, EquCab2.0 SNP Collection*), ou seja, apenas pela comparação da sequência dos seus dois cromossomos homólogos. Além dos marcadores SNP serem abundantes, suas bases moleculares permitem sua distribuição homogênea pelo genoma (CAETANO, 2009).

Entre os maiores destaques provenientes da análise do genoma de equinos está o seu sequenciamento completo (*EquCab2.0*) e, a partir deste, a identificação de 1.162.753 de SNP que ocorrem entre diversas raças (CHOWDHARY; RAUDSEPP, 2008, WADE et al., 2009). A partir dessa abundância de marcadores foram construídos os primeiros arranjos ou chips de genotipagem de SNP para equinos.

Projetados para a realização de estudos genéticos em equinos em que a genotipagem de polimorfismos de DNA se faz necessária, encontram-se disponíveis atualmente no mercado diferentes arranjos. O *Equine SNP50 BeadChip* (Illumina Inc., EUA), já fora do mercado, constituiu poderosa plataforma com cerca de 54.000 SNP. Já em sua segunda geração, o novo SNP chip equino da empresa Illumina (*Equine SNP70 BeadChip*; Illumina Inc., EUA) possui 65.157 SNP, dos quais 19 mil são novos marcadores e 45 mil foram validados no *Equine SNP50 BeadChip*.

Com densidade muito superior aos chips da Illumina, precisamente 670.796 SNP, tem-se o *Axiom® Equine Genotyping Array* (Affymetrix, EUA). Os marcadores deste arranjo foram selecionados a partir de amostra de dois milhões de SNP,

objetivando otimizar a cobertura genômica considerando a diversidade genética e da extensão do desequilíbrio de ligação (*linkage disequilibrium* – LD) existente entre as diferentes raças de equinos domésticos. Os 74.000 SNP diferentes presentes nos arranjos da empresa Illumina (Equine SNP50 e Equine SNP70) foram incluídos neste desenho com a intenção de permitir a compatibilidade, comparação e continuidade de projetos existentes. DNA isolado do sangue ou do bulbo capilar de 347 cavalos de 20 raças equinas foram genotipadas com o *Axiom® Equine Genotyping Array* com o intuito de avaliar o seu potencial de aplicação. Entre estas raças, não se encontra a Mangalarga Marchador, por ser uma raça com maior destaque nacional.

### **3.4. Desequilíbrio de ligação e tamanho efetivo populacional**

O desequilíbrio de ligação (*linkage disequilibrium* – LD) é a associação não aleatória entre alelos de diferentes locos do mesmo cromossomo, ou seja, é quando a frequência de um determinado haplótipo é maior do que seria esperado. Ocorre quando dois genes ou locos estão suficientemente próximos no genoma, sendo improvável a recombinação durante a meiose, o que mantém segmentos do cromossomo conservados de uma geração para outra (ARDLIE et al., 2002; KHATKAR et al., 2006; GODDARD; HAYES, 2009). Assim, o LD entre marcadores moleculares descreve a correlação entre os genótipos de dois marcadores, expressando o grau de associação não aleatória entre seus alelos (PORTO-NETO et al., 2014). Quando as frequências gaméticas observadas em determinada população são idênticas às esperadas obtidas pelo produto das frequências alélicas envolvidas, diz-se que os locos estão em equilíbrio de ligação (*linkage equilibrium* – LE). Quando não ocorre, tem-se o LD (LEWONTIN; KOJIMA, 1960; LEWONTIN, 1964).

O estudo do LD nas espécies teve crescimento rápido na década de 1980 (HEDRICK, 1987), devido sua utilidade na genética de populações, por permitir visão detalhada da evolução e divergências genéticas, além de servir como base para o mapeamento de genes que afetam características de importância (SLATKIN, 2008). O LD é influenciado por eventos como mutação, deriva, seleção e recombinação. Alguns eventos são responsáveis pelo aumento na extensão do LD, incluindo autofecundações, pequenos tamanhos populacionais, isolamento genético entre linhagens, subdivisão populacional, baixa taxa de recombinação, seleção artificial e natural, dentre outros. Alguns outros fatores são responsáveis pela queda na extensão

do LD, incluindo acasalamento ao acaso, elevadas taxas de recombinação e mutações, dentre outros (GUPTA et al., 2005). Assim, sua existência acaba interferindo na formação dos genótipos e haplótipos dos indivíduos nas populações. Desta maneira, o estudo do desequilíbrio de ligação pode mostrar quais e o quanto forças evolutivas estão atuando sobre determinada população (OLIVEIRA, 2014).

A informação do LD nos marcadores utilizados para estudos genéticos tem grande importância, já que o mesmo é responsável pela aplicabilidade em pesquisas de estrutura de diferentes populações em várias espécies de animais domésticos (KIJAS et al., 2009; McKAY et al., 2008; GIBBS et al., 2009), estima o grau de diversidade dentro e divergência genética entre populações (ZENGER et al., 2006), determina a relação entre a perda de alelos e o aumento da endogamia devido à seleção (MUIR et al., 2008), e identifica e localiza regiões do genoma sujeitas à seleção (HAYES et al., 2006; BARENDSE et al., 2009; MACEACHERN et al., 2009; PRASAD et al., 2008). Além disso, os mapas de LD são ferramentas fundamentais para explorar a base genética de características economicamente importantes (MEIRA et al., 2014).

Para avaliar o LD entre marcadores as medidas mais utilizadas são o  $D'$  (LEWONTIN, 1964) e o  $r^2$  (HILL; ROBERTSON, 1968), cada uma com diferentes propriedades estatísticas, porém com valores entre 0 e 1 (BOHMANOVA et al., 2010). O  $r^2$  representa a correlação entre alelos de dois locos e demonstrou ser a medida mais adequada para estimar o LD entre marcadores bialélicos, como os SNP, por ser menos sensível à frequência alélica e ao tamanho da amostra (ZHU; ZHAO, 2007; BOHMANOVA et al., 2010). Para um par de locos bialélicos,  $r^2$  será igual a 1 se apenas dois haplótipos estiverem presentes em uma população (GURGUL et al., 2014). Segundo Ardlie et al. (2002), valores de  $r^2$  superiores a 0,33 são suficientes para caracterizar o LD entre alelos de dois locos.

O tamanho efetivo ( $N_e$ ) é definido como o conjunto de indivíduos de uma população que se reproduzem e conseguem deixar descendentes (FALCONER; MACKAY, 1996). O  $N_e$  foi definido por Fisher (1930) e Wright (1931) como a dimensão de uma população idealizada. O valor do  $N_e$  pode sofrer interferência quando acontece alguns eventos como: seleção e migração, e possibilita identificar a quantidade e distribuição da variação genética presente em determinada população. Também prediz a perda e a distribuição da variação genética, as probabilidades de fixação de alelos benéficos ou prejudiciais e a aptidão e sobrevivência de uma

população (WANG, 2005). Logo, a estimativa do  $N_e$  deve ser considerada na tomada de decisões relativas à pressão de seleção. Segundo Daetwyler et al. (2009), quando o  $N_e$  para uma população é conhecido, a precisão da seleção genômica para uma variedade de cenários pode ser calculada, o que facilita compreender o impacto das estratégias seletivas de reprodução na variação genética presente nas populações e pode fornecer uma visão do nível de endogamia em populações para as quais os *pedigrees* estão incompletos ou indisponíveis.

O estudo do LD em diferentes distâncias permite estimar o  $N_e$  em vários momentos/gerações. O  $N_e$  também pode ser estimado pelo uso de informações de *pedigree*, mas isso depende da integralidade das informações de genealogia disponíveis. Com os métodos baseados em dados genômicos, especificamente os com base em padrões de desequilíbrio de ligação, otimizou-se a estimativa do  $N_e$  sem necessitar da informação de parentesco (SVED, 1971; HILL, 1981; HAYES et al., 2003). Esta nova tecnologia é ferramenta útil para estudar algumas características relacionadas à estrutura das populações e permite comparação mais detalhada entre populações do que as abordagens anteriores baseadas em genealogia.

Em geral, o LD aumenta lentamente à medida que diminui a distância física entre dois locos e, eventualmente, atinge alto  $r^2$  (próximo a 1) em distâncias muito pequenas (GODDARD & HAYES, 2012). Quando o valor de  $N_e$  cai drasticamente, isso pode significar a perda de variabilidade e aumento da endogamia com o consequente surgimento de características genéticas recessivas indesejáveis. O  $N_e$  pode diferir entre populações e raças (BIEGELMEYER et al., 2016). Em populações com tamanho efetivo pequeno, um nível substancial de LD deverá abranger grandes distâncias genéticas (FARNIR et al., 2000; McRAE et al., 2002; ODANI et al., 2006). Por outro lado, em populações com tamanho efetivo grande, espera-se uma menor extensão do desequilíbrio de ligação médio.

### **3.5. Estrutura e subestruturas populacionais**

As análises multivariadas são utilizadas para extrair informações de grandes conjuntos de dados, e são ferramentas de grande utilidade no estudo da estrutura genética de populações. A análise discriminante de componentes principais (*discriminant analysis of principal component* – DAPC), é um método multivariado que baseia-se na transformação de dados usando a análise de componentes principais

(*principal component analysis* – PCA) como um passo anterior para a análise discriminante (*discriminat analysis* – DA), o que garante que as variáveis submetidas à DA sejam perfeitamente não correlacionadas e que seu número seja menor que o dos indivíduos analisados. Sem implicar em uma perda necessária de informação genética, esta transformação permite que a DA seja aplicada a qualquer dado genético (JOMBART; DEVILLARD; BALLOUX, 2010).

A DAPC identifica e descreve dentro da população grupos de indivíduos geneticamente relacionados que formam subestruturas. Quando faltam antecedentes do grupo, a DAPC usa *K-means* sequenciais e seleção do modelo para inferir *clusters* genéticos. Essa abordagem permite extrair informações ricas de dados genéticos, fornecer atribuição de indivíduos a grupos, avaliar visualmente a diferenciação entre populações e a contribuição de alelos individuais para a estruturação da população. As ferramentas implementadas na DAPC para a identificação de *clusters* e a representação gráfica de estruturas entre grupos, permitem desvendar estruturas populacionais complexas (JOMBART; DEVILLARD; BALLOUX, 2010).

Após a inferência do número de grupos, é possível analisar subestruturas familiares importantes dentro da população. A classificação de indivíduos com base no grupo genético possibilita, por análise simples do *pedigree*, estabelecer os principais reprodutores dentro de cada grupo, podendo assim aumentar a compreensão da dinâmica e estrutura populacional. Para isso podem ser analisados dentro de cada grupo os indivíduos e ancestrais mais frequentes (pai, avô materno e paterno).

#### 4. BIBLIOGRAFIA CITADA

ABCCMM. Associação Brasileira dos Criadores do Caval Mangalarga Marchador. Disponível em <<http://www.abccmm.org.br/>>. Acesso em: 12 ago. 2018.

ABCCMM. ASSOCIAÇÃO BRASILEIRA DOS CRIADORES DO CAVALO MANGALARGA MARCHADOR. A história do cavalo Mangalarga Marchador. Belo Horizonte: Nova Fronteira, 1991. 89p.

ANDE-BRASIL. Associação nacional de equoterapia. Disponível em: <<http://www.http://equoterapia.org.br/>>. Acesso em: 12 ago. 2018.

- ANDRADE, L.S. As raças: Mangalarga Marchador. In: ANDRADE, L.S. Criação e adestramento de cavalos marchadores. Recife: L.S. Andrade, 1984. P. 29-37.
- ANDRADE, L. S. A herança genética da marcha. *Matéria Técnica*. 2011. Disponível em: <<http://hytalobretas.com.br/site/noticias.php?id=18&tipo=artigos>> Acessado em: 07/03/2016.
- ARAÚJO, N. Cavalos marchadores brasileiros. *Globo Rural*, Matéria Técnica São Paulo, 2013. Disponível em: <[http:// globo.com/rede-globo/globo-rural/t/edicoes/v/computador-analisa-marcha-de-cavalo-com-cameras-de-alta-velocidade/2328452/](http://globo.com/rede-globo/globo-rural/t/edicoes/v/computador-analisa-marcha-de-cavalo-com-cameras-de-alta-velocidade/2328452/)> Acessado em: 25/02/2016.
- ARDLIE, K. G.; KRUGLYAK, L.; SEIELSTAD, M. Patterns of linkage disequilibrium in the human genome. **Natural Review Genetics**, v. 3, p. 299-309, 2002.
- 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**, v. 10, p. 178, 2009.
- BELTRÁN, N. A. R. **Assinaturas de seleção na linhagem de trabalho de equinos quarto de milha**. 2014. 59 f. Tese (Doutorado corrente em Genética e Melhoramento Animal) – Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista "Júlio de Mesquita Filho", Jaboticabal, 2014.
- BIEGELMEYER, P; GULIAS-GOMES, C.C.; CAETANO, A.R. Linkage disequilibrium, persistence of phase and effective population size estimates in Hereford and Braford cattle, **BMC Genetics**, 17, 1-12, 2016.
- BOHMANOVA, J.; SARGOLZAEI, M.; SCHENKEL, F. Characteristics of linkage disequilibrium in North American Holsteins. **BMC Genomics**, v. 11, p. 421, 2010.
- CAETANO, A. R. Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. **Revista Brasileira de Zootecnia**, v. 38, p. 64-71, 2009.
- CARNEIRO, E.; DUARTE, J. S.; DIAS, M. V. A história do cavalo Mangalarga Marchador. Belo Horizonte: Nova Fronteira, 1991.
- CASIUCH, R. O romance da raça: histórias do cavalo Mangalarga Marchador. São Paulo: Empresa das Artes, 1997. 254p.
- CHOWDHARY, B. P.; RAUDSEPP, T. The Horse Derby: racing from map to whole genome sequence. **Chromosome Research**, v. 16, p. 109-127, 2008.

- COSTA, M.D.; BERGMANN, J.A.G.; RESENDE, A.S.C.; FONSECA, C.G.; FARIA, F.J.C. Estudo da subdivisão genética da raça Mangalarga Marchador. **Arq. Bras. Med. Vet. Zootec.**, v.57, n.2, p.272-280, 2005.
- COSTA, M. D. et al. Caracterização demográfica da raça mangalarga marchador. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, Belo Horizonte, v. 56, n. 5, p. 687-690, 2004.
- DAETWYLER H.D., PONG-WONG R., VILLANUEVA B. & WOOLLIAMS J.A. The impact of genetic architecture on genome-wide evaluation methods. In: *Genome-Wide Evaluation of Populations*, Thesis (PhD) of Daetwyler, H. D., Wageningen University, NL.2009.
- De ASSIS, J. B. et al. Genetic diversity and population structure in brazilian Mangalarga Marchador horses. **Genetics and Molecular Research**, Ribeirão Preto, v. 8, p. 1519-1524, 2009.
- DOAN, R.; COHEN, N.D.; SAWYER, J.; GHAFARI, N.; JOHNSON, C.D.; DINDOT, S.V. Whole-Genome Sequencing and Genetic Variant Analysis of a Quarter Horse Mare. **BMC Genomics**, v.13, n.78, p. 1471-2164, 2012.
- EUCLIDES FILHO, K.; SILVA, L. O. C.; ALVES, R. G. O.; FIGUEIREDO, G. R. Tendência genética na raça Gir. **Pesquisa Agropecuária Brasileira**, v. 35, p. 787-791, 2000.
- FALCONER, D. S. Introdução à genética quantitativa. Trad. de Martinho de Almeida e Silva e José Carlos Silva. Viçosa, UFV, Impr. Univ., 1987. 279 p.
- FALCONER, D.S.; MACKAY, T.F.C. Introduction to quantitative genetics. 4.ed. **Edinburgh: Longman Group Limited**. 464p, 1996.
- FAO. Food and Agriculture Organization of United Nations. Disponível em: <<http://www.fao.org/home/en/>>. Acesso em 14 de nov. 2014.
- FARNIR, F.; COPPIETERS, W.; ARRANZ, J. J.; BERZI, P.; CAMBISANO, N.; GRISART, B.; KARIM, L.; MARCQ, F.; MOREAU, L.; MNI, M.; NEZER, C.; SIMON, P.; VANMANSHOVEN, P.; WAGENAAR, D.; GEORGES, M. Extensive genome-wide linkage disequilibrium in cattle. **Genome Research**, v. 10, p. 220-227, 2000.
- FISHER R,A, The genetical theory of natural selection, Oxford University Press, p,308. 1930.
- GIBBS, R. A.; TAYLOR, J. F.; VAN TASSELL, C. P. Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. **Science**, v. 324, p. 528-532, 2009.



- GLIPHA. Global Livestock Production and Health Atlas. Disponível em: <[http://www.fao.org/ag/againfo/home/en/news\\_archive/AGA\\_in\\_action/glipha.html](http://www.fao.org/ag/againfo/home/en/news_archive/AGA_in_action/glipha.html)>. Acesso em: 11 set. 2017.
- GODDARD, M. E.; HAYES, B. J. Mapping genes for complex traits in domestic animals and their use in breeding programmes. **Nat. Rev. Genet.**, v. 10, p. 381-391, 2009.
- GODDARD, M.E.; HAYES, B.J. Genome-Wide Association Studies and Linkage Disequilibrium in Cattle. **Bovine Genomics. Chapter 13**, p.194-195, 2012.
- GONÇALVES, R. W., Costa, M. D., Rezende, A. S. C., Júnior, V. R., & Leite, J. R. A. Efeito da endogamia sobre características morfométricas em cavalos da raça Mangalarga Marchador. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 64, n. 2, p. 419-426, 2012.
- GUIMARÃES, P. E. M.; COSTA, M. C. R. SNPs: Sutis diferenças de um código. **Biotecnologia Ciência e Desenvolvimento**, v. 26, p. 24-27, 2002.
- GUPTA, P.K., RUSTGI, S., KULWAL, P.L. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. **Plant Molecular Biology**, 57: 461-485. 2005.
- GURGUL, A.; EMIK, E.; PAWLINA, K.; SZMATOLA, T.; JASIELCZUK, I.; BUGNOPONIEWIERSKA, M. The application of genome-wide SNP genotyping methods in studies on livestock genomes. **Journal of Applied Genetics**, v. 55, n. 2, p. 197-208, 2014.
- HAYES, B. J.; BOWMAN, P. J.; CHAMBERLAIN, A. J.; GODDARD, M. E. Invited review: Genomic selection in dairy cattle: progress and challenges. **J Dairy Sci**, v. 92, n. 2, p.433-443, 2009.
- HAYES, B. J.; LIEN, S.; NIELSEN, H.; OLSEN, H. G.; BERG, P.; MACEACHERN, S.; POTTER, S.; MEUWISWSEN, T. E. The origin of selection signatures on bovine chromosome 6. **Animal Genetics**, v. 39, p. 105-101, 2006.
- HAYES, B. J.; VISSCHER, P. M.; MCPARTLAN, H. C.; E. GODDARD, M. E. Novel Multilocus Measure of Linkage Disequilibrium to Estimate Past Effective Population Size. **Genome Research.**, v. 13, p. 635-643, 2003.
- HEDRICK, P.W. Genetic disequilibrium measures. **Genetics**. 117, 331-341, 1987.
- HILL, W. Estimation of effective population size from data on linkage disequilibrium. **Genetical Research**, v. 38, n. 3, p. 209-216, 1981.

HILL, W. G.; ROBERTSON, A. Linkage disequilibrium in finite populations. **Theoretical and Applied Genetics**, v. 38, p. 226-231, 1968.

IBGE. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Sistema IBGE de Recuperação Automática – SIDRA. Disponível em: <<http://www.sidra.ibge.gov.br/>>. Acesso em: 7 nov. 2016.

JOMBART, T.; DEVILLARD, S.; BALLOUX, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. **BMC Genet**, v. 11, p. 94-109, 2010.

KHATKAR, M. S.; COLLINS, A.; CAVANAGH, J. A. L.; HAWKEN, R. J.; HOBBS, M.; ZENGER, K. R.; BARRIS, W.; McCLINTOCK, A. E.; THOMSON, P. C.; NICHOLAS, F. W.; RAADSMA, H. W. A first-generation metric linkage disequilibrium map of bovine chromosome 6. **Genetics**, v. 174, p. 79-85, 2006.

KIJAS, J. W.; TOWNLEY, D.; DALRYMPLE, B. P.; HEATON, M. P.; MADDOX, J. F.; MCGRATH, A.; WILSON, P.; INGERSOLL, R. G.; MCCULLOCH, R.; MCWILLIAM, S.; TANG, D.; MCEWAN, J.; COCKETT, N. V.; ODDY, H.; NICHOLAS, F. W.; RAADSMA, H. A. Genome Wide Survey of SNP Variation Reveals the Genetic Structure of Sheep Breeds. **PLoSOne**, 4(3):e4668, 2009.

KWOK, P. Y.; GU, Z. Single nucleotide polymorphism libraries: why and how are we building them? **Mol Med Today**, v. 5, p. 538-543, 1999.

LANDER, E. S., LINTON, L. M., BIRREN, B., NUSBAUM, C., ZODY, M. C., BALDWIN, J., ... & FUNKE, R. Initial sequencing and analysis of the human genome. **Nature** 409(6822):860-921. 2001.

LEWONTIN, R. C. The interaction of selection and linkage. I. General consideration sheterotic models. **Genetics**, v. 49, p. 49-67, 1964.

LEWONTIN, R.C.; KOJIMA, K. The Evolutionary Dynamics of Complex Polymorphisms. **Evolution**. 14, 458-472, 1960.

LIMA, R. A. S.; SHIROTA, R.; BARROS, G. S. C. Estudo do Complexo do Agronegócio Cavalo no Brasil. (Relatório Final), **CEPEA – Centro de Estudos Avançados em Economia Aplicada**, Piracicaba: ESALQ/USP, SP, Brasil, 2006. 250 p.

MACEACHERN, S.; HAYES, B.; MC EWAN, M.C.; J.; GODDARD, M. An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (*Bostaurus*) using a high density SNP genotyping platform and the

contribution of ancient polymorphism to genomic diversity in Domestic cattle. **BMC Genomics**, v. 10, p. 181, 2009.

MAPA. Ministério da Agricultura, Pecuária e Abastecimento. Disponível em: <<http://www.agricultura.gov.br/>>. Acesso em: 18 ago. 2016.

MCCUE, M. E., BANNASCH, D. L., PETERSEN, J. L., GURR, J., BAILEY, E., BINNS, M. M., ... & LEEB, T. A high density SNP array for the domestic horse and extant *Perissodactyla*: utility for association mapping, genetic diversity, and phylogeny studies. **PLoS genetics**, v. 8, n. 1, p. e1002451, 2012.

MCKAY, S. D.; SCHNABEL, R. D.; MURDOCH, B. M.; MATUKUMALLI, L. K.; AERTS, J.; COPPIETERS, W.; CREWS, D.; DIAS NETO, E.; GILL, C. A.; GAO, C.; MANNEN, H.; WANG, Z.; VAN TASSELL, C. P.; WILLIAMS, J. L.; TAYLOR, J. F.; MOORE, S. S. An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. **BMC Genetics**, v. 9, p. 37, 2008.

MCRAE, A. F.; MCEWAN, J. C.; DODDS, K. G.; WILSON, T.; CRAWFORD, A. M.; SLATE, J. Linkage disequilibrium in domestic sheep. **Genetics**, v. 160, p. 1113-1122, 2002.

MEIRA, C. T.; CURTI, R. A.; FARAH, M. M.; OLIVEIRA, H. N.; BELTRAN, N. A. R.; SILVA, J. A. II V.; MOTA, M. D. S. Prospection of genomic regions divergently selected in racing line of Quarter Horses in relation to cutting line. **Animal Journal**, v. 8, ed. 11, p. 1754-1764, 2014.

MEIRA, C. T. **Avaliação de características morfofuncionais de cavalos da raça mangalarga marchador**. – Diamantina: UFVJM, 2010.48 p. Tese (Doutorado em Zootecnia)-Faculdade de Ciências Agrárias, Universidade Federal dos Vales do Jequitinhonha e Mucuri, 2010.

MOREIRA, G. Ensaio sobre andamentos. Belo Horizonte: **ABCCC**, 1994, 39p.

MOTA, M. D. S.; PRADO, R. S. A. Estudo genético da pontuação total em equinos Mangalarga. *Archivos de Zootecnia, Córdoba*, v. 54, p. 25-30, 2006.

MUIR, W. M.; WONGB, G. K.; ZHANG, Y.; WANGC, J.; GROENEND, M. A. M.; CROOIJMANS, R. P. M. A.; MEGENSD, H. J.; ZHANGE, H.; OKIMOTOF, R.; VEREIJENG, A.; JUNGGERIUSG, A.; ALBERSG, G. A. A.; LAWLEYH, C. T.; DELANVI, M. E.; MACEACHERNE, S.; CHENG, H. H. Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. **Proceedings of the National Academy of Sciences of the United States of America**, v. 105, p. 17312-17317, 2008.

- NASCIMENTO, J.F. Mangalarga marchador: tratado morfofuncional. Belo Horizonte: ABCCMM, 1999, 577p.
- ODANI, M.; NARITA, A.; WATANABE, T.; YOKOUCHI, K.; SUGIMOTO, Y.; FUJITA, T.; OGUNI, T.; MATSUMOTO, M.; SASAKI, Y. Genome-wide linkage disequilibrium in two Japanese beef cattle breeds. **Animal Genetics**, v. 37, p. 139-144, 2006.
- OLIVEIRA, P.S. Estrutura populacional de rebanho fechado da raça Nelore da linhagem Lemgruber, **Pesquisa Agropecuária Brasileira**, v,46, p,639-647, 2014.
- PEREIRA, G. L. **Identificação de regiões cromossômicas, genes e polimorfismos de DNA associados ao desempenho de equinos de corrida da raça quarto de milha.** 2017. 98 f. Tese (Doutorado genética e melhoramento animal) – Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista "Júlio de Mesquita Filho", Jaboticabal, 2017.
- PORTO-NETO, L.; KIJAS, J.; REVERTER, A. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. **Genetics Selection Evolution**, v. 46, n. 2, 2014.
- PRASAD, A.; SCHNABEL, R. D.; MCKAY, S. D.; MURDOCH, B.; STOTHARD, P.; KOLBEHDARI, D.; WANG, Z.; TAYLOR, J. F.; MOORE, S. S. Linkage disequilibrium and signatures of selection on chromosomes 19 and 29 in beef and dairy cattle. **Animal Genetics**, v. 39, p. 597-605, 2008.
- REGATIERI, I. C.; MOTA, MDS. HORSE BREEDING PROGRAM: BIOCHEMICAL ASPECTS./Melhoramento genético de equinos: Aspectos bioquímicos. **Ars Veterinaria**, v. 28, n. 4, p. 227-233, 2012.
- RESENDE, M. D. V.; LOPES, P. S.; SILVA, R. L.; PIRES, I. E. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. **Pesquisa Florestal Brasileira, Colombo**, n.56, p.63-77, 2008.
- SANTIAGO, J. M. **Caracterização morfométrica da raça Mangalarga Marchador**– Belo Horizonte: UFMG, 2013. 110 p. Tese (Doutorado em Zootecnia)-Universidade Federal de Minas Gerais, Escola de Veterinária. 2013.
- SLATKIN, M. Linkage disequilibrium understanding the evolutionary past and mapping the medical future. **Nature reviews genetics.**, 9: 477-485, 2008.
- SVED, J. A. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. **Theor. Popul. Biol.**, v. 2, p. 125-141, 1971.

VENTER, J.C., ADAMS, M.D.; MYERS, .E.; LI, P.W.; MURAL, R.J.; SUTTON, G.G.; SMITH, H.O.; et al. The sequence of the human genome. **Science**, v. 291, p. 1304–1351, 2001.

WADE, C. M.; GIULOTTO, E.; SIGURDSSON, S.; ZOLI, M.; GNERRE, S.; IMSLAND, F.; LEAR, T. L.; ADELSON, D. L.; BAILEY, E.; BELLONE, R. R.; BLOCKER, H.; DISTL, O.; EDGAR, R. C.; GARBER, M.; LEEB, T.; MAUCELI, E.; MACLEOD, J. N.; PENEDO, M. C. T.; RAISON, M.; SHARPE, T.; VOGEL, J.; ANDERSSON, L.; ANTCZAK, D. F.; BIAGI, T.; BINNS, M. M.; CHOWDHARY, B. P.; COLEMAN, S. J.; DELLA VALLE, G.; FRYC, S.; GUERIN, G.; HASEGAWA, T.; HILL, E. W.; JURKA, J.; KIIALAINEN, A.; LINDGREN, G.; LIU, J.; MAGNANI, E.; MICKELSON, J. R.; MURRAY, J.; NERGADZE, S. G.; ONOFRIO, R.; PEDRONI, S.; PIRAS, M. F.; RAUDSEPP, T.; ROCCHI, M.; ROED, K. H.; RYDER, O. A.; SEARLE, S.; SKOWN, L.; SWINBURNE, J. E.; SYVANEN, A. C.; TOZAKI, T.; VALBERG, S. J.; VAUDIN, M.; WHITE, J. R.; ZODY, M. C.; Broad Institute Genome Sequencing Platform, Broad Institute Whole Genome Assembly Team; LANDER, E. S.; LINDBLAD-TOH, K. Genome sequence, comparative analysis, and population genetics of the domestic horse. **Science**, v. 326, p. 865-867, 2009.

WANG J.L. Estimation of effective population sizes from data on genetic markers. **Philosophical Transactions of the Royal Society B-Biological Sciences** 360, 1395–409. 2005

WRIGHT, S. Evolution in Mendelian populations, **Genetics**, 16, 97-159, 1931.

ZAMBORLINI, L. C., Bergmann, J. A. G., Pereira, C. S., Fonseca, C. G., & Carneiro, A. S. Estudo genético-quantitativo de medidas lineares de eqüinos da raça Mangalarga Marchador-I. Estimativas dos fatores de ambiente e parâmetros genéticos. **Revista Brasileira de Ciência Veterinária**, v. 3, n. 2, 1996.

ZENGER, K. R.; KHATKAR, M. S.; CAVANAGH J. A. L.; HAWKEN, R. J.; RAADSMA, H. W. Genome-wide genetic diversity of Holstein Friesian cattle reveals new insights into Australian and global population variability, including impact of selection. **Animal Genetics**, v. 38, p. 7-14, 2006.

ZHU, M.; ZHAO, S. Candidate Gene Identification Approach: Progress and Challenges. **International Journal of Biological Sciences**, v. 3, n. 7, p. 420-427, 2007.

## **CAPÍTULO 2 – Análise genômica da estrutura populacional em cavalos da raça brasileira Mangalarga Marchador**

**RESUMO** – O Mangalarga Marchador é o cavalo de sela brasileiro, possui dois tipos de andamentos característicos, a marcha batida e a marcha picada, que proporcionam maior comodidade ao cavaleiro durante a cavalgada e o trabalho. É principalmente utilizado para trabalho em fazendas de gado de corte e vem se destacando em diferentes modalidades de esportes hípicas. Este estudo teve como objetivo caracterizar, por meio da genotipagem de SNP em larga escala, o desequilíbrio de ligação (LD), calculado por  $r^2$ , de equinos da raça brasileira Mangalarga Marchador criados no Brasil. Também foi investigado o tamanho efetivo ( $N_e$ ) da população, bem como as suas estruturas e relações. Foram utilizados 240 equinos Mangalarga Marchador, de ambos os sexos, e registrados na associação brasileira de criadores da raça (ABCCMM). O número de SNP informativos foi de 377.308. Análises de componentes principais mostraram que cavalos Mangalarga Marchador de marcha batida e de marcha picada pertencem a uma mesma população, ou seja, estes grupos não segregaram de forma significativa dentro da raça, o que deve ser levado em consideração nos estudos genético-populacionais. O  $r^2$  genômico calculado foi de  $0,096 \pm 0,166$ . O LD decaiu consideravelmente a partir de distâncias superiores a 15 e 20 Kb, apresentando valores inferiores a 0,3 e 0,2, respectivamente. O  $N_e$  atual foi de 99 animais. Houve acentuada redução neste parâmetro ao se tomar as estimativas de 16 gerações passadas, em que o  $N_e$  estimado foi de 650 animais. Estes resultados podem estar ligados à uma base genética ampla e parcialmente aberta e à pressões de seleção crescentes na raça. Por fim, observou-se subestruturas populacionais definidas relacionadas mais à importantes reprodutores da raça do que propriamente ao tipo de marcha.

**Palavras-chave:** desequilíbrio de ligação, estruturas populacionais, SNP, tamanho efetivo

## 1. INTRODUÇÃO

O Mangalarga Marchador tem distribuição por todo o País e é a mais numerosa raça equina brasileira (COSTA, 2002; COSTA et al., 2004). Esses animais tiveram origem no sul de Minas Gerais, quando houve acasalamentos entre matrizes ibérico-nacionais com garanhões das raças Árabe, Andaluz, Berbere e Alter de Portugal, entre outras. A mistura de todas estas raças constituiu o material genético para o início da seleção dos animais marchadores, após 1870, pelos criadores da região (NASCIMENTO, 1999).

O Mangalarga Marchador foi inicialmente selecionado para ser utilizado como cavalo de sela, principalmente para o manejo de rebanhos em fazendas de bovinos de corte, porém sua participação tem crescido em diferentes atividades (COSTA, 2002). Possui como particularidade um andamento com velocidade intermediária que difere do trote: a marcha, a qual é subdividida em marcha batida e marcha picada. Essa característica é o principal diferencial da raça por gerar comodidade ao cavaleiro e, conseqüentemente, justifica a utilização desses exemplares em cavalgadas de longa duração, concursos de marcha, cavalgadas de regularidade guiadas por planilhas, lida com o gado, equoterapia, dentre outras atividades. Apesar da relevância da raça no Brasil, ainda há demanda por estudos científicos que avaliem as suas particularidades visando melhorar a criação dos exemplares no país, além de fomentar a exportação de animais (FONSECA, 2018).

O sequenciamento completo do genoma de referência equino (*EquCab2.0*), publicado no ano de 2009, permitiu a identificação de mais de um milhão de SNP que ocorrem entre diferentes raças (WADE et al., 2009). A partir dessa abundância de marcadores puderam ser construídos os primeiros arranjos ou chips de genotipagem de SNP para equinos. A genotipagem simultânea de milhares de polimorfismos distribuídos uniformemente ao longo do genoma, por meio desses arranjos, tem possibilitado o estudo da estrutura genética de diferentes populações em várias espécies de animais domésticos (KIJAS et al., 2009; MCKAY et al., 2008; GIBBS et al., 2009), estimar o grau de diversidade dentro e divergência genética entre populações (ZENGER et al., 2006), determinar a relação entre a perda de alelos e o aumento da endogamia devido à seleção (MUIR et al., 2008), e identificar e localizar regiões do genoma sujeitas à seleção (HAYES et al., 2006; BARENDSE et al., 2009;

MACEACHERN et al., 2009; PRASAD et al., 2008), bem como SNP e genes relacionados à características de importância (SAHANA et al. 2010).

A maior parte destas aplicações requerem o conhecimento ou acesso ao desequilíbrio de ligação (LD) existente entre os marcadores moleculares em dada população da espécie de interesse. O LD entre marcadores descreve a correlação entre os genótipos de dois polimorfismos, expressando o grau de associação não aleatória entre seus alelos (PORTO-NETO et al., 2014). Em geral, esta associação ou ligação é consequência da proximidade física existente entre dois locos (ARDLIE et al., 2002; BOLORMAA et al., 2011) e é influenciada por eventos como mutação, deriva, seleção e recombinação.

Ao nível da população, a deriva genética aleatória faz com que ocorra a divisão de uma população em subpopulações com um número menor de pais resultando, portanto, em endogamia (FALCONER, 1987; YOUNG et al., 1996). No nível molecular, a deriva aleatória afeta as frequências alélicas, leva à perda de variação genética neutra, e à fixação de alelos favoráveis ou deletérios. Uma abordagem que auxilia na predição dessas perdas é a determinação do tamanho efetivo da população ( $N_e$ ), definido como o conjunto de indivíduos de uma população que se reproduzem e conseguem deixar descendentes (FALCONER, 1987).

## **2. MATERIAL E MÉTODOS**

### **2.1. Animais**

Foram utilizados 240 equinos da raça brasileira Mangalarga Marchador, de ambos os sexos, e registrados na Associação Brasileira de Criadores de Cavalos Mangalarga Marchador (ABCCMM). Deste total, 48 animais tiveram o sangue colhido no ano de 2015 na 34ª Exposição Nacional da raça, em Belo Horizonte/MG. Alíquotas de sangue dos demais animais ( $n = 192$ ) foram colhidas em haras dos Estados de São Paulo e Minas Gerais e durante a 36ª Exposição Nacional da raça Mangalarga Marchador, realizada no mês de junho de 2017 na cidade de Belo Horizonte/MG. As exposições nacionais ocorrem em período de 15 dias e contam com a participação de aproximadamente 2.000 animais.



Animais de marcha picada (n = 106) e de marcha batida (n =134), com grande número de filhos e descendentes de indivíduos influentes na formação das principais linhagens da raça foram coletados. A presença de animais de linhagens pouco relacionadas foi considerada na composição da amostra. A presença de irmãos completos foi evitada. Assim, a amostra total foi estabelecida de forma a representar, da melhor maneira possível, a diversidade da população da raça no Brasil. Todos os procedimentos envolvendo animais foram realizados segundo as normas brasileiras de bem-estar animal. O presente projeto foi submetido à avaliação da Comissão de Ética no Uso de Animais (CEUA), da FMVZ, Unesp, Botucatu-SP e aprovado sob nº 0029/2017.

## **2.2. Coleta de sangue e extração do DNA**

Amostras de 5 mL de sangue total de cada animal foram colhidas por venipunção da jugular esquerda, na região do pescoço, utilizando-se tubos com vácuo, contendo 7,5 mg de EDTA.

O processo de extração do DNA foi realizado no Laboratório de Genética e Melhoramento Animal (LGMA) da Faculdade de Medicina Veterinária e Zootecnia, da Unesp, Botucatu/SP, utilizando o *Illustra Blood GenomicPrep Mini Spin Kit* (GE Healthcare, EUA), de acordo com as instruções do fabricante. Ao final do procedimento de extração, o DNA foi quantificado por meio do *Qubit® 3.0 Fluorometer* (Invitrogen, EUA) e analisado quanto à pureza (relações 260/280 e 260/230) e integridade no espectrofotômetro de pequenos volumes *NanoDrop Lite* (Thermo Scientific, EUA) e em gel de agarose a 0,8%, respectivamente. Após a diluição para a concentração de trabalho necessária à genotipagem dos SNP (aproximadamente 10 ng/uL), as amostras de DNA foram estocadas em mini tubos de 1,7mL e armazenadas a -20°C até o momento das genotipagens.

## **2.3. Genotipagem e controle de qualidade**

Os 192 animais genotipados neste estudo, 106 de marcha batida e 86 de marcha picada, foram analisados utilizando o *Axiom® Equine Genotyping Array* (Life Technologies, EUA). O chip foi escaneado no sistema *Gene Titan® Instrument* (Life Technologies, EUA). Quarenta e oito animais, 28 de marcha batida e 20 de marcha

picada, foram genotipados no ano de 2015 com o *Equine SNP70 BeadChip*, que possui 65.157 SNP (65K). Dessa forma, os estudos relacionados ao LD foram conduzidos com informações genóticas de 240 animais, sendo 134 de marcha batida e 106 de marcha picada.

Os controles de qualidade das informações de genotipagem, para indivíduos e SNP, foram realizados por meio dos programas *Genome Studio* versão 2011.1 (Illumina, EUA) e PLINK v1.9 (PURCELL et al., 2007). Para animais utilizou-se *call rate* inferior a 0,95. Em relação aos SNP, foram excluídos inicialmente os localizados nos cromossomos sexuais (X e Y) ou sem identificação de posição no genoma, e os com *call rate* inferior a 0,9. A princípio, não se utilizou a MAF (*minor allele frequency*) como critério para exclusão de SNP uma vez que as diferentes faixas de MAF foram utilizadas no processo de avaliação do arranjo de 670k na raça Mangalarga Marchador. O controle de qualidade utilizado para indivíduos não excluiu nenhuma amostra. O número de marcadores do arranjo de 670k que permaneceram após controle de qualidade estão apresentados na Tabela 1.

**Tabela 1.** Número de marcadores do arranjo *Axiom*® *Equine Genotyping Array* e número de SNP excluídos por cada critério de controle de qualidade utilizado.

Critérios utilizados para exclusão de SNP	<i>Axiom Array</i> (n)
SNP genotipados (iniciais)	670.796
Sem localização conhecida	30.863
Localizados no cromossomo X	28.017
Localizados no cromossomo Y	1
<i>Call rate</i> < 0,9	41.514
SNP remanescentes	570.401

#### 2.4. Imputação entre os arranjos de 65k e 670k

Após a aplicação do controle de qualidade de genotipagem e antes do processo de imputação de genótipos foi realizada a avaliação do potencial de aplicação do *Axiom*® *Equine Genotyping Array* na raça Mangalarga Marchador por meio do cálculo

de classes de MAF (< 0,01; 0,01 a 0,05; 0,05 a 0,4; 0,4 a 0,5 e 0,05 a 0,50) e de MAF média. Os SNP informativos (MAF entre 0,05 e 0,50) foram utilizados para o cálculo da cobertura média efetiva do arranjo.

No processo de imputação de genótipos entre os painéis foram utilizados os 48 cavalos genotipados com o *Equine SNP70 BeadChip* (65k) e os 192 genotipados com o *Axiom® Equine Genotyping Array* (670k). A imputação dos genótipos exclusivos dos animais genotipados com o painel de 670k para os animais genotipados com o painel de 65k foi realizada por meio do programa *Beagle* 5.0 (BROWNING; BROWNING, 2016). A fim de verificar a acurácia de imputação, foram tomados 48 animais ao acaso da amostra de animais genotipados em HD e os mesmos tiveram seus genótipos mascarados para 59.164 SNP, estes foram então imputados e calculou-se o índice de concordância por meio da contagem de SNP corretamente imputados dividido pelo número de SNP imputados. Os valores de acurácia foram obtidos para amostras, ou seja, para cada um dos 48 indivíduos genotipados com o *Equine SNP70 BeadChip*. Após o processo de imputação foi aplicada MAF de 0,05 para a realização das análises subsequentes.

## 2.5. Estimação do desequilíbrio de ligação e do tamanho efetivo da população

A análise de componentes principais (PCA) foi conduzida para a matriz *A* do *pedigree* ( $n = 3.172$ ) a fim de verificar se os animais genotipados formavam um grupo isolado ou representavam bem este *pedigree*, ou seja, a população brasileira da raça Mangalarga Marchador. Também foi conduzida PCA para os SNP em comum aos dois painéis (57.506) para verificar o relacionamento genético entre os animais genotipados com os diferentes arranjos. Por fim, uma terceira PCA, agora novamente com os genótipos imputados, foi realizada com o objetivo de verificar a relação entre os animais de marcha batida e marcha picada. As análises de componentes principais foram conduzidas por meio do programa PLINK 1.9 (PURCELL, 2007) e os resultados foram analisados por meio de gráficos gerados pelo pacote ggplot2 do programa R.

O desequilíbrio de ligação na raça Mangalarga Marchador foi avaliado por meio da estatística  $r^2$  (HILL; ROBERTSON, 1968) para todos os pares de SNP em cada cromossomo utilizando o programa estatístico PLINK 1.9 (PURCELL et al., 2007). O LD médio foi calculado para cada cromossomo e para o genoma como um todo. Também foi calculado o LD médio para intervalos de diferentes distâncias físicas entre

os marcadores. Assim, os genótipos para todos os pares de SNP com menos de 1 Mb de distância foram avaliados para estimar o LD em intervalos de 25 a 100 Kb, 100 a 200 Kb, 200 a 500 Kb, e 500 a 1000 Kb. O gráfico do decaimento da média de desequilíbrio de ligação em função da distância física entre os marcadores foi gerado para a raça Mangalarga Marchador com a utilização do programa R (R Core Team, 2016).

A relação entre o LD ( $r^2$ ) e o tamanho efetivo da população ( $N_e$ ) foi expressa pela equação  $r^2 = 1/(4cN_e + 1)$ , onde  $c$  é a distância genética entre dois SNP expressa em centimorgans (cM) (SVED, 1971). O tamanho efetivo da população da raça brasileira Mangalarga Marchador foi estimado com base nos cromossomos autossomos e na distância física entre SNP (<http://www.ensembl.org>). As distâncias genéticas foram convertidas assumindo que 1cM equivale a 1Mb. Os tamanhos efetivos populacionais foram estimados utilizando a fórmula  $N_{eT} = (1/4c) \times (1/r^2 - 1)$ , onde  $N_{eT}$  é o tamanho efetivo populacional há  $T$  gerações, e  $T = 1/2c$  (HAYES et al., 2003; GODDARD et al., 2006). Pares de marcadores com  $c < 10^{-6}$  (~100bp), isto é,  $T > 500.000$ , não foram utilizados uma vez que a aproximação é válida apenas para  $c$  muito maior que a taxa de mutação (~ $10^{-8}$  por loco por geração). As estimativas pontuais finais da  $N_{eT}$  foram obtidas por meio da média da estimativa de valores  $N_e$  dentro de  $T$ . Uma vez que  $r^2$  é parâmetro limitado e as estimativas pontuais de  $N_e$  são infinitas em  $r^2 = 0$  e zero em  $r^2 = 1$ , apenas valores de  $0,01 < r^2 < 0,99$  foram utilizados para estimar  $N_e$ .

## 2.6. Estrutura e subestruturas populacionais

A análise discriminante dos componentes principais (DAPC) foi implementada por meio do pacote *ade4* (JOMBART, 2008) do programa R. O número de  $K$  grupos e número ótimo de componentes principais (PCs) a serem retidos na análise foram previamente definidos por meio desse mesmo pacote. Para a determinação do número de grupos ( $K$ ) dentro da população utilizou-se o algoritmo de agrupamento *K-means* (JOMBART; DEVILLARD; BALLOUX, 2010). Este algoritmo utiliza os mesmos passos dados na DAPC, em que, de forma prévia, foi feita a transformação dos dados por meio de PCA (principalmente para reduzir os números de variáveis) com posterior análise discriminante, com a ressalva que, neste caso, foi considerado um determinado número de grupos a cada rodada. O algoritmo de agrupamento *K-means*

foi implementado para  $K = 2$  a 20. Para isso foram considerados o número de PCs que explicassem em torno de 70 a 80% da variância acumulada. Para identificar o número ótimo de grupos genéticos foi analisado o critério de informação bayesiana (BIC).

A DAPC baseou-se na transformação de dados por meio de PCA, como um passo anterior e necessário à análise discriminante (DA) propriamente dita. Nesse sentido, foi preciso manter número de PCs suficientes para garantir o poder de discriminação sem, contudo, ter um número excessivo de dimensões, o que poderia levar a um *overfit* das funções discriminantes. Para escolha do número ideal de PCs a serem retidos na análise, foi calculado o  $\alpha$ -score por meio do pacote *adegenet* (JOMBART, 2008) do programa R. O procedimento foi baseado no cálculo do  $\alpha$ -score, em que se mede a diferença entre a proporção da re-atribuição bem-sucedida de indivíduos aos grupos (discriminação observada) e os valores obtidos usando grupos aleatórios (discriminação aleatória). Para diminuir o uso intensivo do processamento, o método utilizou interpolação, reduzindo o número de  $\alpha$ -score estimados na faixa de PCs escolhida, a qual foi estipulada como de 1 a 50.

Após a inferência do número de grupos pelo método *K-means*, foi possível analisar subestruturas familiares importantes dentro da população. A classificação de indivíduos com base no grupo possibilitou, por análise simples do *pedigree*, estabelecer os principais reprodutores dentro destes, possibilitando assim aumentar a compreensão da dinâmica e estrutura populacional de equinos da raça Mangalarga Marchador criados no Brasil. Para isso foram analisados dentro de cada grupo os indivíduos mais frequentes (pai, avô paterno e avô materno).

### 3. RESULTADOS

#### 3.1. Avaliação do arranjo de 670k e imputação dos genótipos

Na Tabela 2 estão apresentadas as quantidades de marcadores do arranjo de 670k distribuídos em diferentes classes de MAFs (<0,01; entre 0,01 e 0,05; entre 0,05 e 0,4; entre 0,4 e 0,5; e entre 0,05 e 0,5) na raça Mangalarga Marchador. A MAF média e desvio padrão encontrados para os marcadores do arranjo de 670k na raça Mangalarga Marchador foi de  $0,1693 \pm 0,1551$ . Considerando os marcadores informativos, ou seja, os com MAF entre 0,05 e 0,5 ( $n = 376.620$ ), a cobertura média

efetiva do *Axiom*® *Equine Genotyping Array* na raça Mangalarga Marchador foi de 1SNP/7.169 pb.

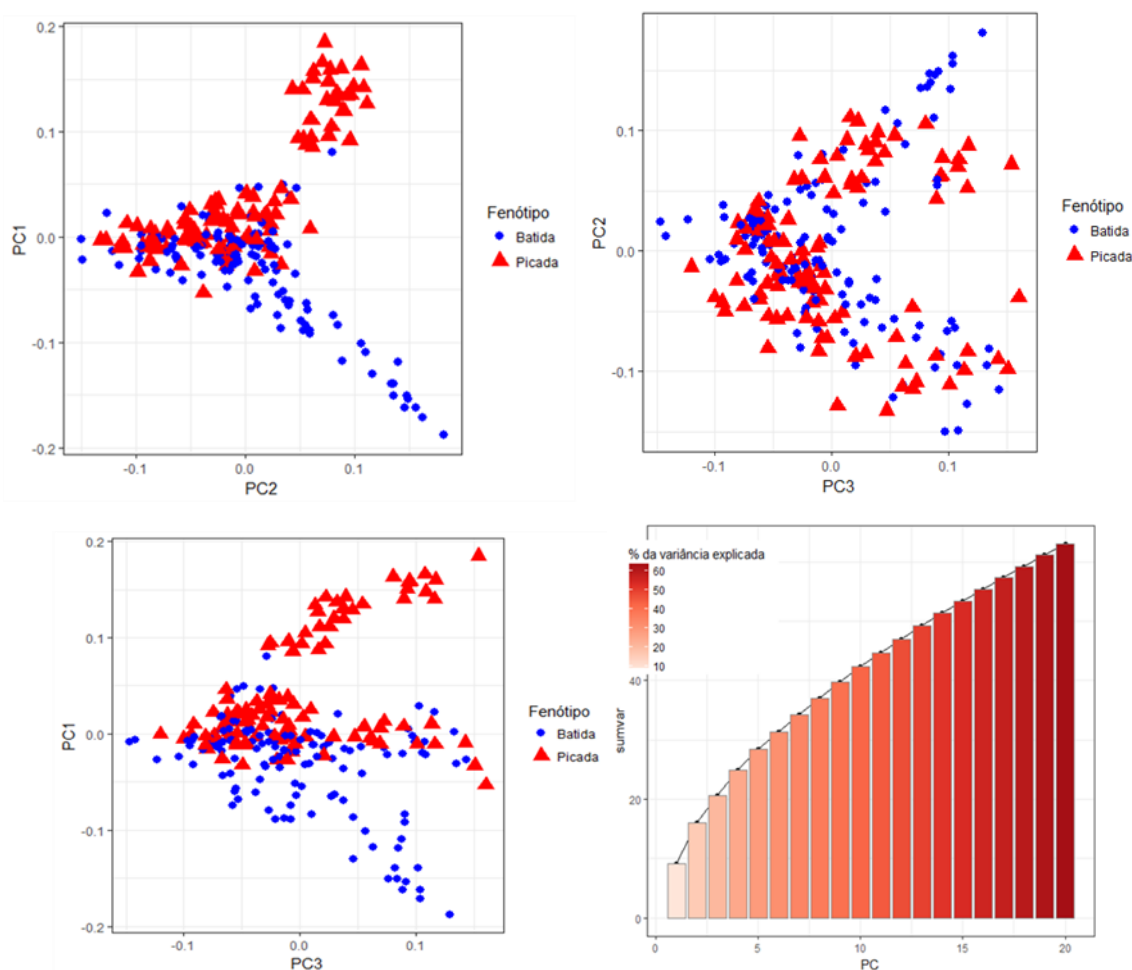
**Tabela 2.** Número e porcentagem de SNP do *Axiom*® *Equine Genotyping Array* distribuídos em quatro classes de *minor allele frequency* (MAF) na raça Mangalarga Marchador.

Classes	<i>Minor allele frequency</i>				
	<0,01	0,01 – 0,05	0,05 – 0,4	0,4 – 0,5	0,05 – 0,5
Número de SNP	107.464	86.317	309.419	67.201	376.620
% de SNP	18,84	15,13	54,25	11,78	66,03

A acurácia de média imputação de genótipos observada para os indivíduos genotipados com o arranjo de 65k foi de 92,63±1,72%, com mediana de 92,91% e valores mínimos e máximos de 84,93 e 95,27%, respectivamente. Após o processo de imputação, o número de SNP eliminados na raça Mangalarga Marchador considerando a MAF de 0,05, foi de 193.093. Desta forma, para as análises genéticas subsequentes foram utilizados 240 indivíduos e 377.308 SNP.

### 3.2. Desequilíbrio de ligação

Os resultados da primeira PCA, a partir da matriz de parentesco A, mostraram que os animais genotipados representam o *pedigree* da raça Mangalarga Marchador no Brasil. Ainda por meio de PCA, observou-se a dispersão dos valores ao longo de quatro quadrantes do gráfico (Figura 1 – PC1xPC2), notando-se segregações (subestruturas) dentro da Mangalarga Marchador, relacionadas, possivelmente, à importantes reprodutores da raça. Entretanto, essas subestruturas não estão relacionadas a animais genotipados com o tipo de andamento (marcha batida e marcha picada). Por fim uma terceira PCA mostrou a inexistência de subestruturas relacionadas aos diferentes arranjos utilizados (54k e 670k).



**Figura 1.** Representação dos três primeiros componentes principais: PC1 e PC2 (superior à esquerda), PC1 e PC3 (superior à direita) e PC2 e PC3 (inferior à esquerda), e variância explicada.

Na Tabela 3 estão descritos os tamanhos dos 31 cromossomos autossomos de equinos, os números de SNP após a realização do processo de imputação de genótipos e os valores médios de  $r^2$  médio por cromossomo na raça Mangalarga Marchador. Observa-se que o maior valor médio de  $r^2$  foi de (0,120) no ECA7 e o menor valor no ECA12 ( $r^2 = 0,077$ ). O  $r^2$  genômico (média geral de todos os  $r^2$  calculados) entre pares de marcadores foi  $0,096 \pm 0,166$  para a raça Mangalarga Marchador.

**Tabela 3.** Tamanho dos cromossomos em megabases (Mb), número de SNP após a imputação e valores médios e desvios padrão e medianas de  $r^2$ , por cromossomo, na amostra estudada de equinos da raça Mangalarga Marchador.

ECA	Tamanho Chr. (Mb)	SNP (n)	Média $r^2$	Desvio padrão	Mediana
1	185,84	35.451	0,102	0,161	0,040
2	120,86	29.488	0,109	0,176	0,042
3	119,48	16.939	0,103	0,154	0,045
4	108,57	15.362	0,100	0,151	0,043
5	99,68	14.612	0,100	0,150	0,044
6	84,72	24.906	0,104	0,166	0,040
7	98,54	13.430	0,120	0,177	0,050
8	94,06	13.782	0,102	0,153	0,044
9	83,56	12.455	0,090	0,143	0,037
10	83,98	12.886	0,096	0,148	0,039
11	61,31	9.172	0,098	0,157	0,039
12	33,09	6.072	0,077	0,122	0,032
13	42,58	7.127	0,088	0,135	0,038
14	93,9	13.791	0,097	0,146	0,043
15	91,57	13.782	0,093	0,144	0,040
16	87,37	13.192	0,098	0,147	0,043
17	80,76	11.901	0,090	0,140	0,038
18	82,53	11.474	0,114	0,163	0,051
19	59,98	9.216	0,093	0,141	0,041
20	64,17	20.265	0,094	0,154	0,036
21	57,72	8.558	0,092	0,139	0,040
22	49,95	7.770	0,089	0,138	0,038
23	55,73	7.628	0,108	0,161	0,045
24	46,75	7.056	0,099	0,147	0,042
25	39,54	6.188	0,096	0,146	0,041
26	41,87	6.604	0,087	0,134	0,038
27	39,96	6.380	0,087	0,132	0,038
28	46,18	6.937	0,094	0,144	0,040
29	33,67	5.450	0,084	0,128	0,038
30	30,06	4.962	0,094	0,140	0,040
31	24,98	4.472	0,082	0,125	0,036

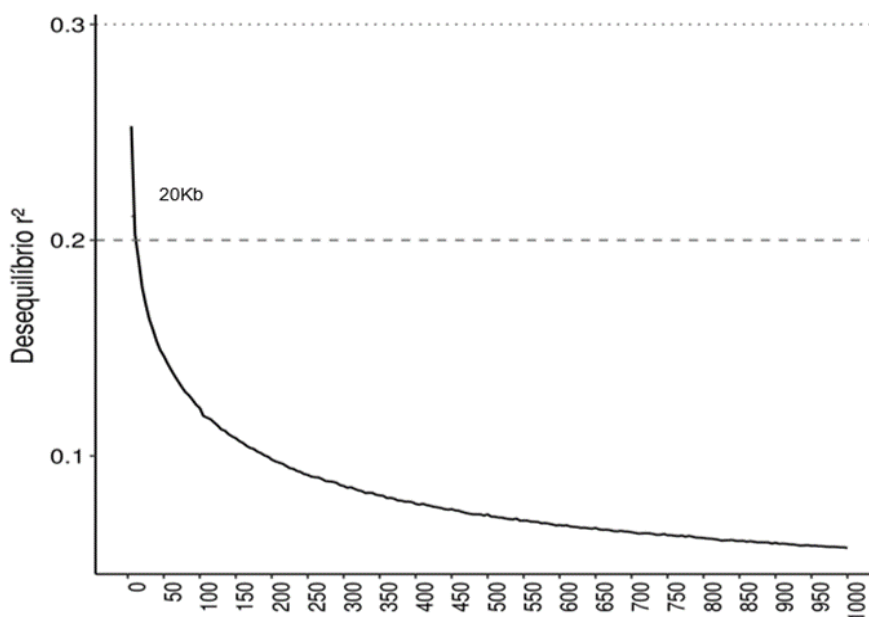
ECA: autossomo *Equus caballus*; Mb= megabases



Os valores médios de  $r^2$  por intervalo considerando todo o genoma estão descritos na Tabela 4. O LD diminuiu à medida que a distância física entre os marcadores aumentou, sendo que o declínio mais acentuado da curva foi observado nos primeiros 50 Kb, decaindo mais lentamente com o aumento da distância (Figura 2). Observa-se ainda na figura 2 que o LD decaiu a curtas distâncias, sendo que a partir de distâncias superiores a 15 e 20 Kb o  $r^2$  médio foi inferior a 0,3 e 0,2, respectivamente.

**Tabela 4.** Valores médios de  $r^2$ , desvios padrão e medianas nos diferentes intervalos de distância física entre SNP na raça de equinos Mangalarga Marchador.

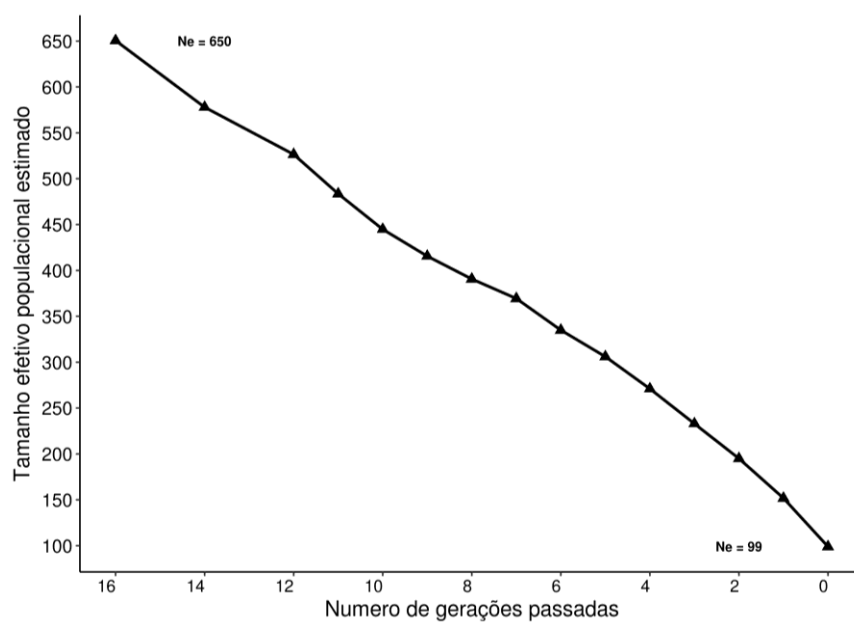
Distância (Kb)	Média $r^2$	Desvio Padrão	Mediana
0-25	0,201	0,258	0,09
25-100	0,140	0,195	0,06
100-200	0,108	0,158	0,05
200-500	0,083	0,125	0,04
500-1000	0,064	0,098	0,03



**Figura 2.** Decaimento da média do desequilíbrio de ligação (LD), analisado por meio de  $r^2$ , na raça Mangalarga Marchador.

### 3.3. Estimação do tamanho efetivo da população

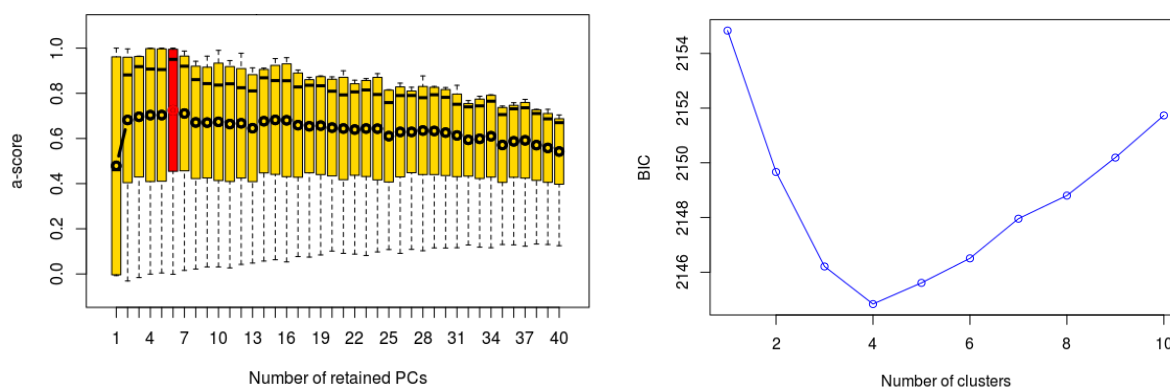
O número efetivo populacional ( $N_e$ ) atual para a raça Mangalarga Marchador, estimado por via genômica, foi de 99 animais (Figura 3). Houve acentuada redução neste número ao se tomar as estimativas de 16 gerações passadas, em que o  $N_e$  estimado foi de 650 animais.



**Figura 3.** Tamanho efetivo populacional estimado ( $N_e$ ) em função do número de gerações passadas em equinos da raça Mangalarga Marchador criados no Brasil.

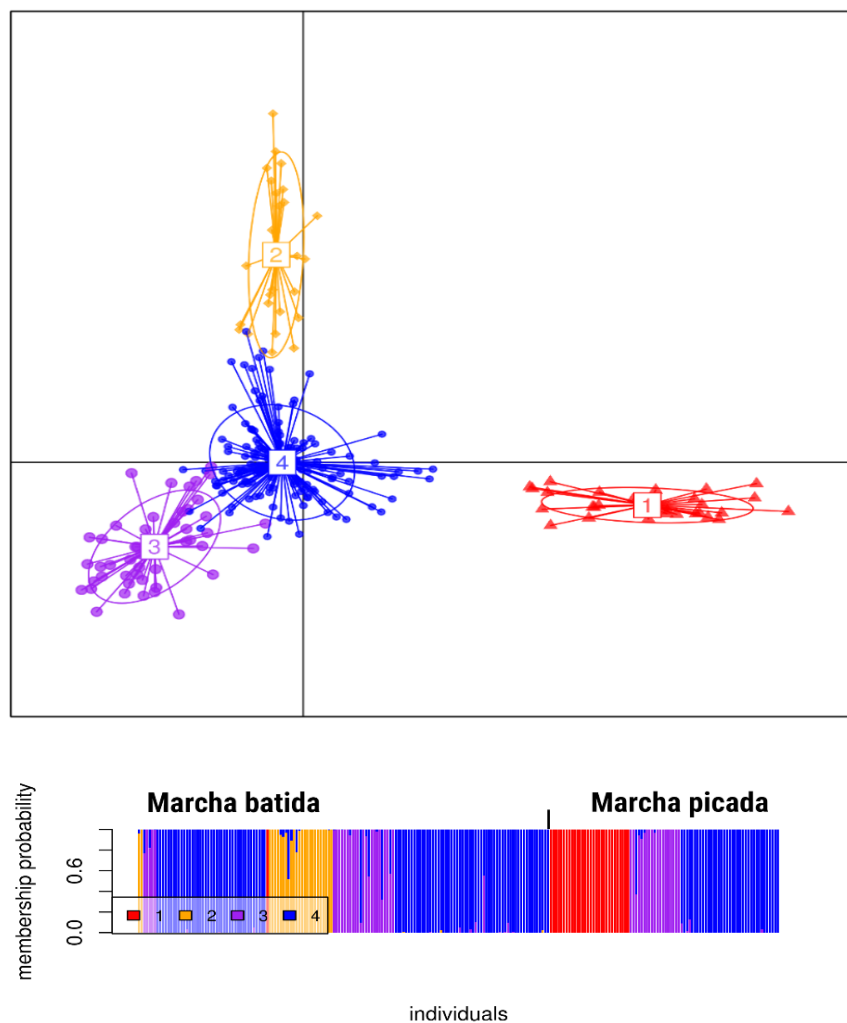
### 3.4. Estrutura e subestruturas populacionais

Com o intuito de esclarecer melhor a estrutura populacional da raça Mangalarga Marchador no Brasil, os dados genômicos foram analisados por meio de DAPC. O número ideal de grupos ( $K = 4$ ) e de componentes principais (PCs = 6) a serem retidos na DAPC foram escolhidos a partir do menor valor de BIC e maior valor de  $\alpha$ -score: 3135 e 0,8 respectivamente (Figura 4). A partir da definição de grupos, foram então analisados os *pedigrees* e os principais reprodutores (os mais frequentes) dentro de cada grupo, os quais notou-se tratar de grupos familiares, com um ou poucos indivíduos aparentados como principais reprodutores.



**Figura 4.** Variações do número ótimo de componentes principais a serem retidos (à esquerda) e número ótimo de grupos (à direita) dentro da amostra populacional de cavalos da raça Mangalarga Marchador.

Ao analisar o gráfico da DAPC notou-se que não houve uma separação entre os animais de marcha batida e marcha picada, mas foi possível observar que se formaram quatro grupos distintos (Figura 5), onde foram demonstradas subestruturas decorrentes da formação de famílias que descendem de diferentes e importantes reprodutores. O *cluster 1* foi o mais afastado geneticamente, os *clusters 2, 3 e 4* que estão mais próximos, representam uma base genética mais homogênea. O *cluster 1* possuiu na sua grande maioria cavalos de marcha picada, tendo apenas um indivíduo de marcha batida, o *cluster 2* teve na sua totalidade equinos de marcha batida e os *clusters 3 e 4* foram representados por animais de ambas as marchas (Tabela 5).



**Figura 5.** Gráfico da análise discriminante dos componentes principais a partir de dados genômicos de cavalos da raça Mangalarga Marchador. Cores e pontos distintos representam cada grupo ( $K$ ) obtido previamente.

**Tabela 5.** Número de cavalos de marcha batida e marcha picada da raça Mangalarga Marchador que compuseram os quatro grupos da análise discriminante de componentes principais.

	Grupos DAPC			
	1	2	3	4
Batida	1	26	25	81
Picada	30	0	22	53

#### 4. DISCUSSÃO

A maior parte dos SNP identificados em equinos, depositados em bancos de dados públicos e utilizados na confecção *Axiom® Equine Genotyping Array* ainda não foi amplamente avaliada. Isto significa que o nível de polimorfismo desse conjunto de SNP ainda é desconhecido para grande parte das raças equinas criadas ao redor do mundo, incluindo a Mangalarga Marchador. Os valores encontrados para o número de SNP nas diferentes classes de MAF mostraram que a utilização do arranjo de 670k é recomendada para a raça Mangalarga Marchador, principalmente para a análise de desequilíbrio de ligação, pois os SNP com MAF superior a 10 e 20% são os mais indicados para estudos de LD, visto que marcadores mais informativos aumentam o poder de detecção de valores mais acurados (SERVICE et al. 2006, KAUPPI et al., 2003) e fornecem estatísticas com menor variância (WEISS; CLARK, 2002; ABECASIS et al. 2001), tornando a determinação da medida mais realista (KE et al., 2004; STENZEL et al., 2004). Os SNP altamente informativos do chip de 670k encontrados para a raça Mangalarga Marchador (com MAF entre 0,4 e 0,5; n = 67.201) apresentam maior potencial para utilização na identificação individual pelo DNA (testes de parentesco e forenses).

A imputação de genótipos dos animais analisados com o arranjo de maior densidade para os animais analisados com o de menor densidade mostrou resultados com acurácia média satisfatória (92,63%) e, portanto, confiáveis. A imputação se baseia na proporção de haplótipos compartilhados entre os animais e esta proporção depende, em última análise, do parentesco entre os indivíduos. Neste sentido, os resultados da imputação mostraram que existe proximidade genética entre os indivíduos genotipados com os diferentes arranjos.

A PCA conduzida para verificar se os animais genotipados utilizados no presente estudo eram representantes do *pedigree* do Mangalarga Marchador criado no Brasil (a partir da matriz A), mostrou que, de fato, esses indivíduos representam muito bem a raça, o que aumentou a confiabilidade dos resultados obtidos para as análises genéticas populacionais. Também deve ser levado em consideração que a ausência de um maior tamanho amostral pode ser compensada pela genotipagem de um grande número de SNP (PRITCHARD; PRZEWORSKI, 2001), o que ocorreu na presente pesquisa. As demais PCAs indicaram que não houve subestruturas relacionadas aos animais com diferentes tipos de marcha ou genotipados com os

diferentes painéis de SNP. Assim, o primeiro resultado mostrou que, embora haja alguma segregação, os animais Mangalarga Marchador de marcha batida e de marcha picada pertencem a uma mesma população e essa informação deve ser levada em consideração nos estudos genético-populacionais da raça. Apesar disso, observou-se subestruturas ligadas a importantes reprodutores da raça, o que será discutido mais adiante. O segundo resultado favoreceu a imputação de genótipos, como já discutido anteriormente.

Os resultados da Tabela 3 mostraram a maior ocorrência de polimorfismos informativos nos cromossomos de maior tamanho e a menor ocorrência nos de menor tamanho na raça Mangalarga Marchador, indicando que o processo de filtragem dos polimorfismos não afetou de forma severa a uniformidade de distribuição dos SNP ao longo do genoma. A mesma tabela mostrou ainda grande diversidade de valores de LD entre os diferentes cromossomos na raça Mangalarga Marchador. Uma possível explicação para o fato é que a grande variabilidade nas taxas de recombinação nos cromossomos autossomos, além de outros fatores, leva à considerável diversidade no padrão de LD em distintas regiões genômicas (ARIAS et al., 2009). As análises de LD por cromossomo mostraram o maior valor médio de  $r^2$  (0,120) no ECA7 e o menor valor (0,077) no ECA12, um dos menores cromossomos equinos. Esses resultados apontam para uma relação entre tamanho do cromossomo e a estimativa média de  $r^2$  em algumas raças equinas, como observado por Marchiori et al. (2019) na linhagem de trabalho da raça Quarto de Milha. Esses autores observaram as maiores e menores estimativas médias de  $r^2$ , de forma respectiva, também nos cromossomos sete e doze.

Observou-se valores médios de  $r^2$  de 0,201, quando considerados intervalos de 0-25 Kb, e 0,064 em intervalos de 500-1000 Kb. Na raça Quarto de Milha, foram, no geral, encontrados valores superiores ao encontrado neste estudo, em que o  $r^2$  médio em extensões menores (0-25kb) foram de 0,427 e 0,475 e extensões maiores (500-1000kb) 0,131 e 0,185, nas linhagens de trabalho e de corrida, respectivamente (MARCHIORI et al., 2019). Esses resultados mostraram que, na raça Mangalarga Marchador, o LD tende a ser menor que outras raças equinas, visto que em estudo conduzido por Mccue e Mickelson (2013), a raça Quarto de Milha já apresentava menores estimativas de LD entre inúmeras raças em diferentes intervalos avaliados.

Neste estudo, conduzido por Mccue e Mickelson (2013), a extensão do LD foi caracterizada em 14 raças equinas. Em geral, os valores calculados de  $r^2$  caíram abaixo de 0,2 à distâncias entre 100 e 150 Kb. Notáveis exceções foram o Puro-

Sangue Inglês, o *Standardbred* e o *French Trotter*, nos quais os valores de  $r^2$  permaneceram acima de 0,2 a distâncias superiores a 400 Kb. Entre as raças mais difundidas analisadas nesse estudo, a Quarto de Milha se destacou por seus valores de LD relativamente baixos quando comparados à outras raças equinas, com decaimento de valores de  $r^2$  abaixo de 0,2 entre 50 e 100 Kb. A pequena extensão do LD e o decaimento de  $r^2$  para valores abaixo de 0,2 a curtas distâncias (20 Kb) encontrados no presente estudo na raça Mangalarga Marchador podem ser explicados inicialmente por um grande número de animais fundadores e pela estrutura parcialmente aberta da raça, pois de acordo com Costa (2002) e Costa et al. (2004), o livro de registro sem genealogia conhecida (livro aberto) de machos da Associação da raça foi fechado em 1966, e o de fêmeas em 1984.

As extensões de LD na raça Mangalarga Marchador descritas neste estudo, fornecem informações estratégicas para futuras pesquisas genômicas na raça, como, por exemplo, os estudos de associação ampla. Segundo Oraguzie et al. (2007), em geral, o LD é variável no genoma e depende da espécie e população sob estudo. Assim, a quantificação do LD é de grande utilidade em estudos evolutivos e possibilita determinar a resolução de mapeamento e a densidade de marcadores necessários na identificação de QTL (NORDBORG et al., 2002; WEI; JACKSON; MCINTYRE, 2006; ZHU et al., 2008). Considerando a cobertura média efetiva do *Axiom*® *Equine Genotyping Array* na raça Mangalarga Marchador (1SNP/7.169 pb) em conjunto com os resultados encontrados para o LD (valores de  $r^2$  inferiores a 0,2 para marcadores distantes 20 Kb), observa-se que o arranjo de 670k é suficiente/adequado para realização de GWAS e seleção genômica, entre outros estudos genéticos e aplicações, na raça Mangalarga Marchador.

As estimativas do presente trabalho mostraram que houve redução acentuada do tamanho efetivo populacional ( $N_e$ ) nas últimas 16 gerações da raça Mangalarga Marchador criada no Brasil. Este fato pode ser atribuído ao aumento gradual da pressão de seleção, devido à diferentes critérios utilizados nas escolhas de reprodutores durante a história da raça, partindo, no início, de uma seleção para trabalho, até a seleção para morfologia e para desempenho em provas de marcha, visto nas gerações mais recentes (BECK, 1992). Marchiori et al. (2019) relataram na raça Quarto de Milha um decaimento menos drástico das estimativas de  $N_e$  ao longo das gerações, com número menor de animais desde sua formação, o que pode ser explicado, possivelmente, pelos eventos distintos de formação da raça. No Quarto de

Milha, a partir de uma base genética mais fechada, selecionou-se animais para modalidades esportivas altamente competitivas (AQHA, 2018), o que levou, naturalmente, desde sua formação, a escolha de poucos reprodutores. Vale destacar que ao longo da criação e da expansão da raça Mangalarga Marchador, foi utilizada uma ampla base genética, em sua grande maioria sem raça definida e genealogia desconhecida (BECK, 1992), o que ajuda entender os altos valores de  $N_e$  nas gerações iniciais e as pequenas extensões cromossômicas em alto LD encontradas atualmente na raça.

O conhecimento do valor médio aproximado das proporções da população ancestral para cada indivíduo e/ou suas populações de origem é uma parte útil da caracterização das populações em estudo (BIEGELMEYER et al., 2016). Os resultados relacionados à estrutura populacional encontrados neste trabalho diferiram em relação ao estudo conduzido por Costa et al. (2005), que identificaram cinco grupos genéticos na população da raça, que são descendentes dos animais Herdade Alteza e Seta Caxias, Providência Itu e Tabatinga Predileto, Abaíba Marengo, Tabatinga Cossaco e Angaí Miron. É possível que com o passar das gerações dois desses grupos tenham se aproximado por meio de acasalamentos, se tornando geneticamente semelhantes.

Ao analisar o *pedigree*, identificou-se que no *cluster 1*, o qual foi formado quase que exclusivamente por cavalos de marcha picada, os reprodutores de destaque são: Rigoletto da Marcha; Herdade Caxias e Narciso Marca F. Esses animais têm genética Angaí Miron, Herdade Alteza e Seta Caxias. O *cluster 2* foi representado apenas por animais de marcha batida, sendo formado pelos garanhões: Ginete da Embaúba; Incerto H.Z. e Vento J.N.53. Ao observar os *clusters 3* e *4*, nota-se que alguns indivíduos acabaram se aproximando e provocando sobreposição. Estes grupos foram compostos por cavalos de marcha batida e picada, onde observou-se, além dos três grupos genéticos citados anteriormente (Angaí Miron, Herdade Alteza e Seta Caxias), indivíduos com genética Abaíba Marengo e Providência Itu.

Dessa forma foi demonstrada a ocorrência de subestruturas decorrentes da formação de famílias que descendem de diferentes e importantes reprodutores. A distinção genética observada no *cluster 1* se deve ao fato de este ser formado por famílias mais antigas da raça. Os *clusters* mais próximos (2, 3 e 4) representam uma base genética mais homogênea. Isto ocorre, devido, principalmente, à acasalamentos



interfamiliares de seus descendentes e semelhança genética dos principais reprodutores entre grupos.

## 5. CONCLUSÃO

Embora haja alguma segregação, cavalos Mangalarga Marchador de marcha batida e de marcha picada criados no Brasil pertencem a uma mesma população, o que deve ser levado em consideração em estudos genético-populacionais da raça. O decaimento acentuado do desequilíbrio de ligação a curtas distâncias no genoma e o tamanho efetivo, o qual era grande no início da formação da raça e apresentou queda pronunciada no decorrer das últimas gerações, podem estar ligados à uma base genética ampla e parcialmente aberta e à pressões de seleção crescentes na raça. Por fim, foram observadas subestruturas populacionais definidas relacionadas mais à importantes reprodutores da raça do que propriamente ao tipo de marcha.

## 6. BIBLIOGRAFIA CITADA

- ABECASIS, G.R.; NOGUCHI, E.; HEINZMANN, A.; TRACHERNE, J. A.; BHATTACHARYYA, S.; LEAVES, N. I.; CARDON, L. R. Extent and distribution of linkage disequilibrium in three genomic regions. **The American Journal of Human Genetics**, v. 68, n. 1, p. 191-197, 2001.
- ARDLIE, K. G.; KRUGLYAK, L.; SEIELSTAD, M. Patterns of linkage disequilibrium in the human genome. **Natural Review Genetics**, v. 3, p. 299-309, 2002.
- ARIAS, J. A.; KEEHAN, M.; FISHER, P.; COPPIETERS, W., SPELMAN, R. A high density linkage map of the bovine genome. **BMC Genetics**, v. 10, p. 18, 2009.
- AQHA. AMERICAN QUARTER HORSE ASSOCIATION. Horse-Breeding History. Disponível em: <<https://www.aqha.com/daily/breeding/2016/breeding-archive/horsebreeding-history/>>. Acesso em: 29 nov. 2018.
- 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**, v. 10, p. 178, 2009.

- BECK, S. L. **Mangalarga Marchador: Caracterização, história, seleção**. Brasília: Edição dos autores, 1992. p. 1-13.
- BIEGELMEYER, P.; GULIAS-GOMES, C.C.; CAETANO, A.R.; STEIBEL, Y.J.P. Linkage disequilibrium, persistence of phase and effective population size estimates in Hereford and Braford cattle. *BMC Genetics*, 17, 1-12, 2016.
- BOLORMAA, S.; HAYES, B. J.; SAVIN, K.; HAWKEN, R.; BARENDSE, W.; ARTHUR, P. F.; HERD, R. M.; GODDARD, M. E. Genome-wide association studies for feedlot and growth traits in cattle. *Journal of Animal Science*, v. 89, p. 1684-1697, 2011.
- BROWNING, B. L.; BROWNING, S. R. Genotype imputation with millions of reference samples. *The American Journal of Human Genetics*, v. 98, n. 1, p. 116-126, 2016.
- COSTA, M.D. Caracterização demográfica e estrutura genética da raça Mangalarga Marchador. 99f. Tese (Doutorado)-Escola de Veterinária, Universidade Federal de Minas Gerais, Belo Horizonte, 2002.
- COSTA, M.D.; BERGMANN, J.A.G.; REZENDE, A.S.C. et al. Caracterização demográfica da raça Mangalarga Marchador. *Arq. Bras. Med. Vet. Zootec.*, v.56, p.687-690, 2004.
- COSTA, M. D. et al. Inbreeding and effective population size of the Brazilian Mangalarga Marchador horse breed. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia*, v. 57, n. 1, p. 112-119, 2005.
- FALCONER, D. S. Introdução à genética quantitativa. Trad. de Martinho de Almeida e Silva e José Carlos Silva. Viçosa, UFV, Impr. Univ., 1987. 279 p.
- FONSECA, M.G. **Mangalarga Marchador : estudo morfométrico, cinemático e genético da marcha batida e da marcha picada**. 73 p. Tese (Doutorado em Medicina Veterinária) Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Jaboticabal, 2018.
- GIBBS, R. A.; TAYLOR, J. F.; VAN TASSELL, C. P. Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. *Science*, v. 324, p. 528-532, 2009.
- GODDARD, M. E.; HAYES, B. J.; MCPARTLAN, H.; CHAMBERLAIN, A. J. “Can the same genetic markers be used in multiple breeds?”. *In: Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, CD-ROM communication nº 22-16, Brasil: Belo Horizonte, MG, p.13-18, 2006.*
- HAYES, B. J.; LIEN, S.; NIELSEN, H.; OLSEN, H. G.; BERG, P.; MACEACHERN, S.; POTTER, S.; MEUWISWSEN, T. E. The origin of selection signatures on bovine chromosome 6. *Animal Genetics*, v. 39, p. 105-101, 2006.

- HAYES, B. J.; VISSCHER, P. M.; MCPARTLAN, H. C.; E. GODDARD, M. E. Novel Multilocus Measure of Linkage Disequilibrium to Estimate Past Effective Population Size. **Genome Research.**, v. 13, p. 635-643, 2003.
- HILL, W. G.; ROBERTSON, A. Linkage disequilibrium in finite populations. **Theoretical and Applied Genetics**, v. 38, p. 226-231, 1968.
- JOMBART, T.; DEVILLARD, S.; BALLOUX, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. **Bmc Genet**, v. 11, p. 94-109, 2010.
- JOMBART, T. Adegenet: a R package for the multivariate analysis of genetic markers. **Bioinformatics**, v. 24, p. 1403–1405, 2008.
- KAUPPI, L.; SAJANTILA, A.; JEFFREYS, A.J. Recombination hotspots rather than population history dominate linkage disequilibrium in the MHC class II region. **Human Molecular Genetics**, v. 12, n. 1, p. 33-40, 2003.
- KE, X.; HUNT, S.; TAPPER, W.; LAWRENCE, R.; STAVRIDES, G.; GHORI, J.; CARDON, L.R. The impact of SNP density on fine-scale patterns of linkage disequilibrium. **Human Molecular Genetics**, v. 13, n. 6, p. 577-588, 2004.
- KIJAS, J. W.; TOWNLEY, D.; DALRYMPLE, B. P.; HEATON, M. P.; MADDOX, J. F.; MCGRATH, A.; WILSON, P.; INGERSOLL, R. G.; MCCULLOCH, R.; MCWILLIAM, S.; TANG, D.; MCEWAN, J.; COCKETT, N. V.; ODDY, H.; NICHOLAS, F. W.; RAADSMA, H. A. Genome Wide Survey of SNP Variation Reveals the Genetic Structure of Sheep Breeds. **PLoSOne**, 4(3):e4668, 2009.
- MACEACHERN, S.; HAYES, B.; MC EWAN, M.C.; J.; GODDARD, M. An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (*Bos taurus*) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in Domestic cattle. **BMC Genomics**, v. 10, p. 181, 2009.
- MARCHIORI, C. M., PEREIRA, G. L., MAIORANO, A. M., ROGATTO, G. M., ASSONI, A. D., SILVA, J. A. I. V., ... & CURI, R. A. Linkage disequilibrium and population structure characterization in the cutting and racing lines of Quarter Horses bred in Brazil. **Livestock Science**, v. 219, p. 45-51, 2019.
- MCKAY, S. D.; SCHNABEL, R. D.; MURDOCH, B. M.; MATUKUMALLI, L. K.; AERTS, J.; COPPIETERS, W.; CREWS, D.; DIAS NETO, E.; GILL, C. A.; GAO, C.; MANNEN, H.; WANG, Z.; VAN TASSELL, C. P.; WILLIAMS, J. L.; TAYLOR, J. F.; MOORE, S. S. An

assessment of population structure in eight breeds of cattle using a whole genome SNP panel. **BMC Genetics**, v. 9, p. 37, 2008.

McCUE, M.; MICKELSON, J. **Genomic Tools and Resources: Development and Applications of an Equine SNP Genotyping Array**. In: Equine Genomics (Ed. B. P. Chowdhary), Blackwell Publishing Ltd., Oxford, UK, 2013.

MUIR, W. M.; WONGB, G. K.; ZHANG, Y.; WANGC, J.; GROENEND, M. A. M.; CROOIJMANS, R. P. M. A.; MEGENSD, H. J.; ZHANGE, H.; OKIMOTOF, R.; VEREIJENG, A.; JUNGGERIUSG, A.; ALBERSG, G. A. A.; LAWLEYH, C. T.; DELANVI, M. E.; MACEACHERNE, S.; CHENG, H. H. Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. **Proceedings** of the National Academy of Sciences of the United States of America, v. 105, p. 17312-17317, 2008.

NASCIMENTO, J.F. Mangalarga Marchador: Tratado morfofuncional. Belo Horizonte: ABCCMM, 1999. 900p.

NORDBORG, M.; BOREVITZ, J. O.; BERGELSON, J.; BERRY, C. C.; CHORY, J.; HAGENBLAD, J.; KREITMAN, M.; MALOOF, J. N.; NOYES, T.; OEFNER, P. J.; et al. The extent of linkage disequilibrium in *Arabidopsis thaliana*. **Nat Genet.**, v. 30, n. 2, p.190–193, 2002.

ORAGUZIE, N. C., RIKKERINK, E. H. A., GARDINER, S. E., SILVA, H. N. de. (Eds.) Association Mapping in Plants. New York: Springer, p. 11-39, 2007.

PORTO-NETO, L.; KIJAS, J.; REVERTER, A. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. **Genetics Selection Evolution**, v. 46, n. 2, 2014.

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

PRASAD, A.; SCHNABEL, R. D.; MCKAY, S. D.; MURDOCH, B.; STOTHARD, P.; KOLBEHDARI, D.; WANG, Z.; TAYLOR, J. F.; MOORE, S. S. Linkage disequilibrium and signatures of selection on chromosomes 19 and 29 in beef and dairy cattle. **Animal Genetics**, v. 39, p. 597-605, 2008.

PRITCHARD, J.K.; PRZEWORSKI, M. Linkage disequilibrium in humans: models and data. **Am J Hum Genete** 71:1224 – 1247. 2001.

- R CORE TEAM. R: A language and Environment for Statistical Computing. R foundation for Statistical Computing 2016. <<http://www.R-project.org/>>.
- SAHANA, G.; GULDBRANDTSEN, B.; BENDIXEN, C.; LUND, M. S. Genome-wide association mapping for female fertility traits in Danish and Swedish Holstein cattle. **Animal Genetics**, v. 41, p. 579–588, 2010.
- SERVICE, S.; DEYOUNG, J.; KARAYIORGOU, M.; ROOS, J. L.; PRETORIOUS, H.; BEDOYA, G.; HEUTINK, P. Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. **Nature genetics**, 38(5), 556. 2006.
- STENZEL, A.; LU, T.; KOCH, W. A.; HAMPE, J.; GUENTHER, S. M.; FRANCISCO, M.; SCHREIBER, S. Patterns of linkage disequilibrium in the MHC region on human chromosome 6p. **Human genetics**, v. 114, n. 4, p. 377-385, 2004.
- SVED, J. A. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. **Theor. Popul. Biol.**, v. 2, p. 125-141, 1971.
- WADE, C. M.; GIULOTTO, E.; SIGURDSSON, S.; ZOLI, M.; GNERRE, S.; IMSLAND, F.; LEAR, T. L.; ADELSON, D. L.; BAILEY, E.; BELLONE, R. R.; BLOCKER, H.; DISTL, O.; EDGAR, R. C.; GARBER, M.; LEEB, T.; MAUCELI, E.; MACLEOD, J. N.; PENEDO, M. C. T.; RAISON, M.; SHARPE, T.; VOGEL, J.; ANDERSSON, L.; ANTCZAK, D. F.; BIAGI, T.; BINNS, M. M.; CHOWDHARY, B. P.; COLEMAN, S. J.; DELLA VALLE, G.; FRYC, S.; GUERIN, G.; HASEGAWA, T.; HILL, E. W.; JURKA, J.; KIIALAINEN, A.; LINDGREN, G.; LIU, J.; MAGNANI, E.; MICKELSON, J. R.; MURRAY, J.; NERGADZE, S. G.; ONOFRIO, R.; PEDRONI, S.; PIRAS, M. F.; RAUDSEPP, T.; ROCCHI, M.; ROED, K. H.; RYDER, O. A.; SEARLE, S.; SKOWN, L.; SWINBURNE, J. E.; SYVANEN, A. C.; TOZAKI, T.; VALBERG, S. J.; VAUDIN, M.; WHITE, J. R.; ZODY, M. C.; Broad Institute Genome Sequencing Platform, Broad Institute Whole Genome Assembly Team; LANDER, E. S.; LINDBLAD-TOH, K. Genome sequence, comparative analysis, and population genetics of the domestic horse. **Science**, v. 326, p. 865-867, 2009.
- WEI, X.; JACKSON, P. A.; MCINTYRE, C. L. Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure. *Theoretical and Applied Genetics*, New York, v. 114, p. 155-164, 2006.
- WEISS, K.M.; CLARK, A.G. Linkage disequilibrium and the mapping of complex human traits. **TRENDS in Genetics**, v. 18, n. 1, p. 19-24, 2002.

YOUNG, A.; BOYLE, T.; BROWN, T. The population genetic consequences of habitat fragmentation for plants. **Trends Ecol. Evol.**, v. 11, p. 413-418, 1996.

ZENGER, K. R.; KHATKAR, M. S.; CAVANAGH J. A. L.; HAWKEN, R. J.; RAADSMA, H. W. Genome-wide genetic diversity of Holstein Friesian cattle reveals new insights into Australian and global population variability, including impact of selection. **Animal Genetics**, v. 38, p. 7-14, 2006.

ZHU, C.; GORE, M.; BUCKLER IV, E. S.; YU, J. Status and prospects of association mapping in plants. *The Plant Genome*, Madison, v. 1, p. 5-20, 2008.