



MARIELLE MOURA BAENA

**POPULATION STRUCTURE AND GENETIC
CHARACTERIZATION OF THE MANGALARGA
MARCHADOR HORSE BREED**

**LAVRAS - MG
2019**

MARIELLE MOURA BAENA

**POPULATION STRUCTURE AND GENETIC CHARACTERIZATION OF THE
MANGALARGA MARCHADOR HORSE BREED**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Zootecnia, para a obtenção do título de Doutor.

Dra. Sarah Laguna Conceição Meirelles
Orientadora

Dra. Raquel Silva de Moura
Coorientadora

**LAVRAS – MG
2019**

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Baena, Marielle Moura.

Population Structure and Genetic Characterization of the
Mangalarga Marchador Horse Breed/ Marielle Moura Baena. –
2019.

162 p. : il.

Orientador(a): Sarah Laguna Conceição Meirelles.

Coorientador(a): Raquel Silva de Moura.

Tese (doutorado) -Universidade Federal de Lavras, 2019.

Bibliografia.

1. equideocultura. 2. marcador molecular. 3. pedigree. I.
Meirelles, Sarah Laguna Conceição. II. Moura, Raquel Silva
de. III. Título.

MARIELLE MOURA BAENA

**POPULATION STRUCTURE AND GENETIC CHARACTERIZATION OF THE
MANGALARGA MARCHADOR HORSE BREED**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Zootecnia para a obtenção do título de Doutor.

APROVADA em 15 de fevereiro de 2019.

Dr ^a . Sarah Laguna Conceição Meirelles	UFLA
Dr. Alessandro Moreira Procópio	FEAD
Dr ^a . Elaine Maria Seles Dorneles	UFLA
Dr ^a . Luciana Correia de Almeida Regitano	EMBRAPA/Pecuária Sudeste
Dr ^a . Silvina Diaz	UNLP/CONICET

Orientadora
Dra. Sarah Laguna Conceição Meirelles

Dra. Raquel Silva de Moura
Coorientadora

**LAVRAS – MG
2019**

AGRADECIMENTOS

A Deus por todas as bênçãos, oportunidades, força e serenidade para realizar esse trabalho e aos meus mentores e amigos espirituais que me guiam e intuem a respeito da escolha do melhor caminho a seguir e a me proteger.

À Universidade Federal de Lavras, ao Programa de Pós-Graduação em Zootecnia e ao CNPq, pela oportunidade concedida para a realização do doutorado e concessão de bolsa de estudos.

A todos do Grupo de Estudos em Melhoramento Animal e Biotecnologia (GMAB) pela amizade e pela parceria nos estudos.

Aos meus amigos e companheiros de trabalho, Felipe Amorim, Yasmin Ribeiro e Cássia Assis pela boa convivência e ajuda nos experimentos realizados.

Às minhas amigas, irmãs de república, Amanda e Sarita pelos bons e maus momentos e, por cuidarem de mim como minha família.

Aos meus pais, Haroldo e Glória, e aos meus irmãos, Delanno e Gabrielle e ao Thales por acreditarem em mim e por sempre me incentivarem.

A todos os colegas do Instituto de Genética Veterinária da Faculdade de Ciências Veterinárias da Universidade Nacional de La Plata, pela paciência e colaboração para a realização do experimento, principalmente a Dra. Silvina Diaz e Dr. Herman Oliveira.

À minha queridíssima orientadora Professora Sarah pela sua dedicação, amizade, disposição e paciência em me orientar e à minha coorientadora Professora Raquel por todos seus esforços para o desenvolvimento do projeto.

Aos haras de Lineu Machado, El Far, Henrique, Recanto da Mata, Santa Hedwiges, Três Irmãos, Santa Helena, Setecentos, Floresta, Santa Clara, Santa Lucia, Nova Esperança e à organização dos eventos de competições de marcha da raça realizadas em Caxambú, na Venda do Chico (Três Corações) e na Expolavras (Lavras), pela colaboração nas coletas das amostras dos animais.

A todos que contribuíram direta ou indiretamente para a realização deste trabalho, o meu muito OBRIGADA!

RESUMO GERAL

O Mangalarga Marchador (MM) é a maior raça equina do Brasil, e com a análise de *pedigree* e de marcadores moleculares tornou-se possível a melhor investigação genética na raça. Os objetivos neste trabalho foram avaliar a estrutura populacional e pré-caracterizar geneticamente o MM com dados de *pedigree* e genotipagens. *Pedigree* de 514.283 animais foram utilizados para estimativas da endogamia (F), probabilidade de origem gênica e de estruturação da raça. Genotiparam-se 3.193 equinos para 13 microssatélites (SSR) e em chip de polimorfismo de nucleotídeo único, SNP70K, 244 animais das raças MM, Andaluz (AND), Puro Sangue Lusitano (PSL), Puro Sangue Inglês (PSI), Árabe (ARA), Campolina (CAM), Mangalarga (MAN), Puro Sangue Espanhol (PSE) e Sorraia (SOR). Assim, calcularam-se o número de alelos raros (AR), heterozigosidades esperadas (He) e observadas (Ho), homozigosidades esperadas (Home) e observadas (Homo), frequência de alelos menores (MAF), estatísticas F (Fis, Fit e Fst), distância genética (DG), análise de componentes principais (ACP), equilíbrio de Hardy-Weinberg (EHW), desequilíbrio de ligação (DL), diferenciação populacional, estruturação populacional por agrupamento e presença de haplótipos comuns entre as raças. Por análise de *pedigree* da raça MM, o intervalo de geração foi de 9,48 a 10,43, F de 1,02% a 1,10%, relação genética aditiva de 0,0023 a 0,0045, números efetivos de ancestrais, fundadores e genomas fundadores foram de 198 a 1.038, 225 a 1.062 e 133 a 150, números de fundadores que apresentam 50% e 10% do *pool* genético representaram 7,94% a 19,09% e 0,69% a 6,64% dos fundadores. Por análise de SSR, o AR médio no MM foi de dois alelos/*locus*, médias de He e Ho oscilaram de 0,5767 a 0,8450 e 0,5331 a 0,7949, médias de Fis (-0,0195), Fit (0,0566) e Fst (0,0748) e observou-se desvios de EHW, além de diferenciação populacional nos *loci* e 60,26% dos alelos tinham forte DL. As DG foram maiores no ARA e PSI com as demais raças e menores entre o MM, MAN e CAM e estas com o AND e PSL, resultado corroborado pelo ACP. O MM foi uma entidade genética subestruturada, porém com alguns haras mais estruturados. Com os SNP, no MM obteve-se o maior valor de Homo (34.916,70), 51% dos animais possuíam perda de heterozigosidade, MAF médio de 0,2010, houve desvios de EHW nos cromossomos, He variou de 0,2940 a 0,3231 e extensões menores de DL. Encontrou-se bloco de haplótipo semelhante entre o MAN e MM. A menor diferenciação ocorreu entre o CAM e MM (Fst = 0,0411), e a maior entre PSI e SOR (Fst = 0,2284). A raça MM foi pouco estruturada ao contrário do PSL, PSE, PSI e SOR, e observou-se o compartilhamento de alelos entre MM e MAN. ACP discriminou em maior magnitude a SOR e PSI das demais raças, moderadamente o PSL e PSE, e menor entre MM, MAN e CAM. O MM não está em severa endogamia, o que não prejudica sua evolução como espécie. Contudo, é necessário o monitoramento contínuo da diversidade genética com o uso de marcadores e registros precisos de *pedigree*.

Palavras-chave: Equideocultura. Marcador molecular. *Pedigree*. Variabilidade genética.

GENERAL ABSTRACT

Mangalarga Marchador (MM) is the largest equine breed in Brazil, and the analysis of pedigree and molecular markers has made possible the best genetic research in the breed. The objectives of this work were to evaluate the population structure and to pre-characterize the MM with pedigree and genotyping data. Pedigree of 514,283 animals were used for estimates of inbreeding (F), probability of gene origin and breed structuring. Genotyped 3193 equines for 13 microsatellites (SSR) and in chip of the single nucleotide polymorphism, SNP70K, 244 animals of the breeds MM, Andalusian (AND), Lusitano (LUS), Thoroughbred (THO), Arabian (ARA), Campolina (CAM), Mangalarga (MAN), Puro Sangue Espanhol (PSE) and Sorraia (SOR). Thus, calculated the number of rare alleles (RA), expected heterozygosities (He) and observed (Ho), expected homozygosity (Home) and observed (Homo), minor allele frequency (MAF), F statistics (Fis, Fit e Fst), genetic distance (GD), principal component analysis (PCA), Hardy-Weinberg equilibrium (HWE), linkage disequilibrium (LD), population differentiation, population structuring by clustering and presence of common haplotypes between breeds. By pedigree analysis in the MM, the generation interval was from 9.48 to 10.43, F from 1.02% to 1.10, additive genetic ratio from 0.0023 to 0.0045, effective numbers of ancestors, founders, and genomes founders were from 198 to 1038, 225 to 1062 and 133 to 150, numbers of founders who presented 50% and 10% of the genetic pool accounted for 7.94% to 19.09% and 0.69% to 6.64% of the founders. For SSR analysis, the mean RA in the MM was two alleles/locus, He and Ho mean values ranged from 0.5767 to 0.8450 and 0.5331 to 0.7949, Fis averages (-0.0195), Fit (0.0566) and Fst (0.0748) and deviations of HWE were observed such as population differentiation in the loci and 60.26% of the alleles had strong LD. The GD were higher in the ARA and THO with the other breeds and minors between the MM, MAN and CAM and these with the AND and LUS, a result corroborated by the PCA. The MM was a substructured genetic entity, however with some more structured farms. With the SNP, in the MM the highest value of Homo (34,916.70) was found, 51% of the animals had loss of heterozygosity, mean MAF of 0.2010, there were deviations of EHW in the chromosomes, He ranged from 0.2940 to 0.3231 and smaller extensions of DL. A similar haplotype block was found between MAN and MM. The lowest differentiation occurred between CAM and MM (Fst = 0.0411), and the highest between THO and SOR (Fst = 0.2284). The MM was poorly defined unlike the LUS, PSE, THO and SOR, and the alleles sharing between MM and MAN was observed. PCA discriminated to a greater extent the SOR and THO of the other breeds, moderately the LUS and PSE, and smaller between MM, MAN and CAM. The MM is not in severe endogamy, which does not affect its evolution as a species. However, continuous monitoring of genetic diversity is necessary with the use of molecular and accurate pedigree records.

Keywords: Equideoculture. Genetic variability. Molecular marker. Pedigree.

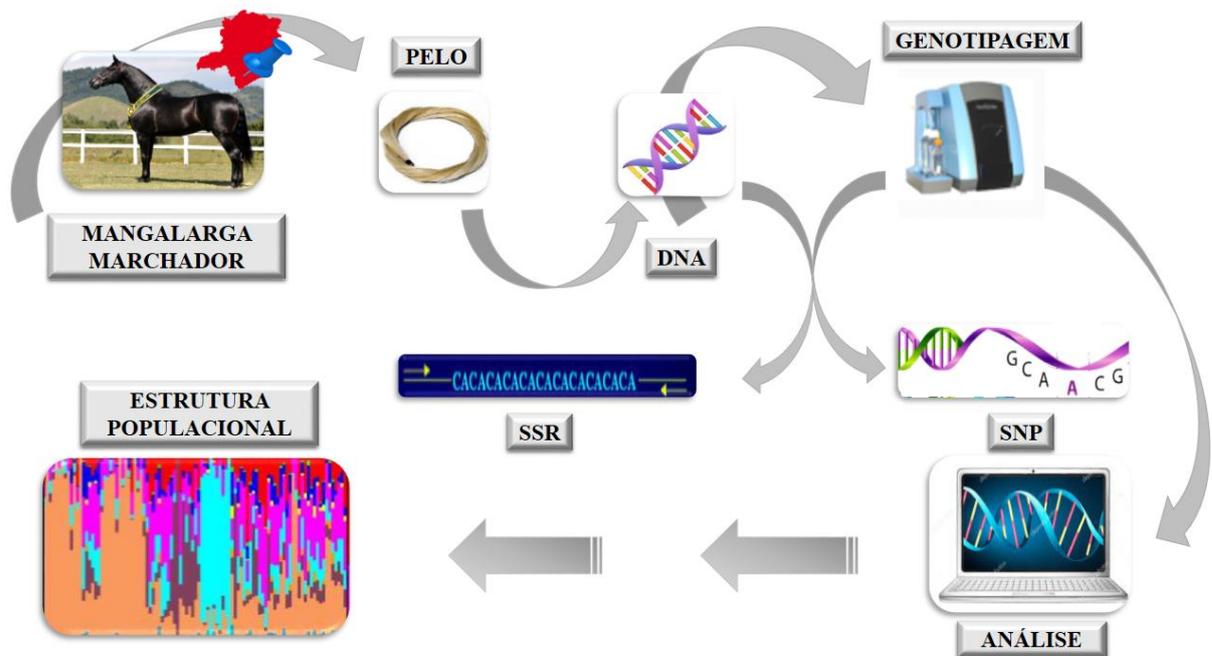
SUMÁRIO

	RESUMO INTERPRETATIVO E RESUMO GRÁFICO.....	10
	PRIMEIRA PARTE – REFERENCIAL TEÓRICO.....	11
1	INTRODUÇÃO	11
2	REFERENCIAL TEÓRICO.....	13
2.1	A raça Mangalarga Marchador	13
2.2	Melhoramento genético em equinos	14
2.3	Conceito de população	15
2.4	Estrutura populacional e <i>pedigree</i>	16
2.5	Variabilidade genética e sua estimação	18
2.6	Seleção artificial e suas consequências	21
2.7	Endogamia e seus efeitos.....	22
2.8	Parâmetros demográficos	24
2.8.1	Intervalo de geração.....	24
2.8.2	Coeficiente e taxa de endogamia	24
2.8.3	Relação genética aditiva.....	25
2.8.4	Tamanho efetivo da população	25
2.8.5	Probabilidade de origem gênica.....	26
2.9	Caracterização genética.....	26
2.10	Marcadores moleculares.....	27
2.10.1	Microssatélites (SSR)	28
2.10.2	Polimorfismo de nucleotídeo único (SNP).....	30
2.11	Análises estatística com uso de marcadores moleculares	31
2.11.1	Equilíbrio de Hardy-Weinberg	31
2.11.2	Desequilíbrio de ligação	32
2.11.3	Heterozigosidades esperada e observada	33
2.11.4	Conteúdo de informação polimórfica	34
2.11.5	Estatísticas F de Wright.....	34
2.11.6	Distância genética	35
	REFERÊNCIAS	37

SEGUNDA PARTE – ARTIGOS	44
ARTICLE 1 – <i>Population structure and analysis of genetic diversity of Mangalarga Marchador horses based on pedigree information.....</i>	44
ARTIGO 2 – Caracterização genética de equinos da raça Mangalarga Marchador baseado em marcadores moleculares microssatélites.....	101
ARTIGO 3 – Pré-caracterização genética de equinos da raça Mangalarga Marchador baseado em marcadores moleculares SNP	131

RESUMO INTERPRETATIVO E RESUMO GRÁFICO

A compreensão sobre a estrutura e caracterização genética fornece novas informações importantes sobre mecanismos e evolução genética que ocorrem na raça Mangalarga Marchador. Aqui, parâmetros comumente estimados foram obtidos a partir de informações derivadas de marcadores moleculares dos tipos microssatélite e polimorfismo de nucleotídeo único, além de análise de *pedigree*. Novas informações sobre o histórico e estrutura genética da raça foram observadas, como a semelhança do Mangalarga Marchador com outras raças brasileiras e falta de uma estrutura genética mais definida. Estes resultados podem dar subsídios para o estabelecimento de um futuro programa de melhoramento genético na raça, além de auxiliar a monitorar e manter a variabilidade genética do plantel, garantindo assim a conservação de padrões específicos, sobrevivência e bom desempenho de suas características de interesse zootécnico.



Resumo gráfico. Coleta de pelo, genotipagem de marcadores moleculares microssatélite (SSR) e polimorfismo de nucleotídeo único (SNP) e análises computacionais para obtenção da estruturação genética populacional de equinos da raça Mangalarga Marchador.

PRIMEIRA PARTE – REFERENCIAL TEÓRICO

1 INTRODUÇÃO

A raça equina Mangalarga Marchador (MM) é de origem brasileira, e sua formação ocorreu na região sul do estado de Minas Gerais há mais de 200 anos. Estes animais sempre foram utilizados nas lidas do campo e como meio de transporte de famílias da região, desempenhando, portanto, importante função econômica e social no país. Atualmente, o MM é considerado uma das principais raças brasileiras de equinos e está entre as maiores do mundo, o que conferiu o título de Raça Nacional por meio do sancionamento da Lei nº 12.975 em 2014, fato que demonstra o seu grande reconhecimento econômico e social no país (ASSOCIAÇÃO BRASILEIRA DOS CRIADORES DO CAVALO MANGALARGA MARCHADOR - ABCCMM, 2018).

As pesquisas na área de melhoramento genético animal são proporcionalmente menores em equinos se comparada a outras espécies de interesse zootécnico. No Brasil, não diferente do restante dos outros países e dada à grandeza de sua tropa, esta distância é ainda mais evidente. Algumas pesquisas vêm sendo publicadas e envolvem a área de melhoramento genético, porém ainda não existem, efetivamente, programas de seleção de raças equinas criadas no Brasil. Deste modo, pequenos grupos de pesquisadores ligados às universidades e institutos de pesquisas têm trabalhado na área, especialmente em relação a aspectos conservacionistas ou quantitativos de caracteres de interesse econômico em raças nacionais e também importadas (PEREIRA et al., 2015).

Algumas pesquisas de melhoramento genético em equinos relatam sobre a estrutura populacional de raças utilizando dados de *pedigree*, sendo esse estudo o primeiro passo para o desenvolvimento de estratégias de seleção (TEEGEN et al., 2009). O conhecimento da estrutura populacional genética e demográfica de uma raça permite entender como ocorreu o fluxo de genes em uma população, bem como obter informações sobre o grau de variabilidade genética, o que contribui para a tomada de decisões em relação à sua gestão genética (MILHO, 2015).

Atualmente, com a disponibilidade de novas ferramentas moleculares, as pesquisas na área de melhoramento genético em equinos, que eram escassas, foram se modificando e tornando-se mais relevantes (PEREIRA, 2017). A partir da identificação de marcadores moleculares como os microssatélites (SSR) e polimorfismo de nucleotídeo único (SNP), tornou-se possível a discriminação genotípica, o que permitiu o estudo da variação genética a

nível de DNA. Assim, estudos de caracterização genética de raças, que verificam os níveis de variabilidade genética dentro e entre populações, além de fornecerem a estrutura genética da mesma e utilizando-se desses marcadores, passaram a ser realizados em distintas populações de animais domésticos (TAMBASCO et al., 2000).

Deste modo, os objetivos deste estudo foram analisar a estrutura genética da população de equinos Mangalarga Marchador do Brasil baseado na análise de *pedigree*, identificando fatores que podem afetar sua variabilidade genética, além de investigar a diversidade genética da raça MM e compará-lo com raças relacionadas ou não à sua história de formação a partir do estudo de pré-caracterização genética com dados de genotipagens de SSR e SNP.

2 REFERÊNCIAL TEÓRICO

2.1 A raça Mangalarga Marchador

A raça de cavalos Mangalarga Marchador (MM) é de origem brasileira e sua formação ocorreu na região sul do estado de Minas Gerais há mais de 200 anos. A raça sempre foi utilizada nas lidas do campo e como meio de transporte das famílias da região, desempenhando, portanto, importante função econômica e social no país. Atualmente, é considerada uma das principais raças brasileiras de equinos criada no país e está entre as maiores do mundo em contingente populacional, o que conferiu o título de “Raça Nacional” por meio do sancionamento da Lei nº 12.975 em 2014, demonstrando o seu grande reconhecimento econômico e social no Brasil (ASSOCIAÇÃO BRASILEIRA DOS CRIADORES DO CAVALO MANGALARGA MARCHADOR - ABCCMM, 2018).

Segundo Costa et al. (2004), as principais características destes animais são o andamento marchado, a docilidade e a rusticidade nas quais permitem sua utilização para o lazer, esporte e trabalho nas propriedades rurais, principalmente como auxílio no manejo com os bovinos.

A formação da raça está diretamente ligada à história da colonização do Brasil, pois em 1808, devido à invasão de Napoleão Bonaparte na coroa portuguesa, a família real comandada pelo príncipe regente Dom João VI fugiu para a colônia brasileira. Com a vinda da família real, vieram também animais da corte portuguesa como os cavalos da raça Puro Sangue Lusitano, cuja origem étnica é de cavalos nativos da península Ibérica, germânicos e berberes como o Sorraia, Marismeño, Garrano, Berbere ou Barbo. O Puro Sangue Lusitano é originário da Coudelaria de Alter do Chão, famoso centro de produção de cavalos da coroa portuguesa, que na época fornecia animais para toda a nobreza europeia. Consta na história que um dos garanhões lusitanos vindos com a família real foi doado pelo príncipe a Gabriel Francisco Junqueira, conhecido como Barão de Alfenas, proprietário da fazenda Campo Alegre em Cruzília no estado de Minas Gerais, considerada hoje o berço da raça MM. Deste modo, Barão de Alfenas realizou cruzamentos deste cavalo lusitano com éguas nativas denominadas “crioulas”, existentes em sua propriedade no sul do estado mineiro. Estas éguas crioulas, provavelmente, eram descendentes dos primeiros equinos trazidos, também da Península Ibérica, no início da colonização brasileira, em 1500. Contudo, a partir destes cruzamentos formou-se a raça Mangalarga Marchador, com animais de porte elegante, com

beleza plástica, temperamento dócil e próprio para montaria (ASSOCIAÇÃO BRASILEIRA DOS CRIADORES DO CAVALO MANGALARGA MARCHADOR - ABCCMM, 2018).

A Associação Brasileira dos Criadores do Cavalo Mangalarga Marchador (ABCCMM) foi fundada em 16 de julho de 1949, em reunião realizada em Caxambu – MG, atualmente com sede em Belo Horizonte. A partir de sua fundação iniciaram-se os trabalhos de registros genealógicos dos animais, com isso, foram estabelecidos padrões morfológicos dentro da raça, algumas avaliadas subjetivamente, o que conseqüentemente direcionou a seleção desses animais. Esta entidade é considerada a maior associação de uma única raça da América Latina, com aproximadamente 11 mil associados em 2016, cinco mil destes no estado de Minas Gerais. A raça, além de difundida por todo o Brasil, tem também criadores em vários países como os Estados Unidos, Alemanha, Holanda, Bélgica, Portugal, Itália, Canadá, Israel, Peru, Uruguai, Argentina, Congo, França e Dinamarca (SOUZA, 2016).

De acordo com a Associação Brasileira dos Criadores do Cavalo Mangalarga Marchador (ABCCMM, 2018) há mais de 600.000 animais registrados; 300 veterinários cadastrados para realizarem transferência de embriões; 56 técnicos de registro; 24 árbitros e 16 instrutores; 70 núcleos de criadores no Brasil e quatro associações de criadores da raça no exterior. São realizados mais de 200 eventos durante o ano nos diversos estados brasileiros e com participação de mais 9000 expositores. São comercializados em torno 4.500 animais em 50 leilões da raça, além da realização de 25 eventos no exterior, nos quais comprovam a relevância da raça que gera cerca de 40 mil empregos diretos e mobiliza 200 mil pessoas indiretamente.

2.2 Melhoramento genético em equinos

As pesquisas na área de melhoramento genético são proporcionalmente menores em equinos, se comparada a outras espécies de interesse zootécnico. No Brasil, não diferente do restante dos outros países e dada à grandeza de sua tropa, esta distância é ainda mais evidente. Algumas pesquisas vêm sendo publicadas em cavalos e envolvem a área de melhoramento genético, porém ainda não existem, efetivamente, programas de seleção de raças equinas criadas no Brasil. Deste modo, pequenos grupos de pesquisadores ligados às universidades e institutos de pesquisa têm trabalhado na área, especialmente em relação a aspectos conservacionistas ou quantitativos de caracteres de interesse econômico em raças nacionais e também de raças importadas. O que se faz ultimamente, quando se pensa em melhoramento genético em equinos, é a estimação de parâmetros genéticos populacionais tais como

herdabilidade e correlação genética entre características de interesse, em razão à disponibilidade de informações fenotípicas e de genealogia aprofundada que se tem em algumas raças (PEREIRA et al., 2015).

Apesar dos criadores de cavalos, de modo geral, almejarem animais geneticamente superiores, existe um distanciamento entre pesquisa e aplicação prática que impede a realização de programas de melhoramento. Mesmo ocorrendo um aumento significativo das publicações que tratam a respeito desse tema em equinos, o melhoramento genético dos cavalos continua sendo muito complexo, fazendo com que a obtenção de um animal geneticamente superior para as provas equestres ou mesmo para o trabalho seja tarefa complicada. A maior causa desta dificuldade é a baixa herdabilidade do desempenho atlético e os valores baixos a moderados de correlações genéticas existentes entre as características de maiores interesses nos cavalos. A baixa herdabilidade do desempenho atlético em equinos, por exemplo, pode ser um argumento a mais para a utilização de novas ferramentas disponíveis para a seleção e o melhoramento genético, como os marcadores moleculares. Existem também dificuldades que são inerentes à própria espécie equina como o baixo índice reprodutivo de seus animais, o grande intervalo de geração, o baixo número de progênes por parição e o longo período de gestação, dificuldades em função de aspectos operacionais como informações escassas e imprecisas de caracteres reprodutivos, comportamentais e de desempenho em grande parte das raças, baixa receptividade das associações de criadores às tecnologias reprodutivas, e a relação superficial entre órgãos técnicos e criadores (MEIRA, 2010).

A seleção aplicada aos equinos, em geral, tem como objetivo o melhoramento de características ligadas à conformação e ao andamento dos animais. Porém, os critérios geralmente utilizados baseiam-se em características avaliadas subjetivamente ou, até mesmo, de forma empírica pelos próprios criadores. Segundo Costa et al. (1998), as associações de raças realizam avaliações morfológicas e de andamento dos animais para a geração dos registros oficiais, quando os mesmos devem atingir pontuação mínima associada à caracterização da raça. Os registros zootécnicos resultantes deste controle não têm sido muito utilizados pela comunidade científica para caracterização e avaliação das raças equinas nacionais.

2.3 Conceito de população

Define-se como população, um grupo de indivíduos da mesma espécie que ocupa o mesmo local com continuidade no tempo e que se acasalam, trocando alelos entre si (GIANNONI, 1987). Nas populações de animais domésticos a condição de acasalamento aleatório normalmente é limitada, já que a permanência de reprodutores em um rebanho está sujeita a interferências, principalmente do homem, devido a critérios de seleção artificial adotados (FALCONER; MACKAY, 1996).

Uma população, do ponto de vista genético, é um conjunto de genes que se forma ao longo do tempo sob a influência de quatro processos evolutivos: a seleção natural, a deriva genética, a migração e a mutação (RODRIGUES, 2009). Toda população possui um reservatório gênico que também é chamado de *pool* gênico específico. Este representa todos os genes presentes na população em uma dada geração ou período. Para cada gene da população pode-se estimar as frequências de seus alelos. A frequência alélica é a proporção de um dado alelo em relação ao total de alelos situados no mesmo *locus* cromossômico. Já a frequência genotípica representa a proporção de um determinado genótipo em relação ao número total de genótipos para dado *locus* e relaciona-se com a frequência alélica da geração anterior, pois na reprodução o que são transmitidos são os alelos e não os genótipos (PINTO, 2010).

2.4 Estrutura populacional e *pedigree*

A estrutura populacional de uma espécie, ou seja, seu conjunto de características genéticas e demográficas, é o resultado da ação e interação de uma série de mecanismos evolutivos e ecológicos. Segundo Cruz (2005), a estrutura genética de uma população pode ser definida por sua frequência de alelos que compõem os diferentes genótipos dos vários indivíduos que a integram. Um estudo de estrutura populacional de uma raça deveria ser o primeiro passo para o desenvolvimento de estratégias de seleção. Porém, é necessário uma melhor monitorização de parâmetros obtidos por esse estudo, que leve a um aumento do tamanho efetivo da população, sem reduzir o progresso genético, enquanto se procede a um trabalho de melhoramento sustentável (TEEGEN et al., 2009). Este estudo permite entender como ocorreu o fluxo de genes em uma população, o que proporciona informação acerca do número de genes que a originou e estima em que medida estes participam na população mais recente. Assim, é possível obter informações acerca do grau de variabilidade genética, o que permite tomar decisões em relação à sua gestão genética. Por outro lado, também nos permite obter informação sobre a presença de linhas genéticas que estão a influenciar de forma mais

significativa a população atual ou um determinado grupo de animais que são conhecidos por alguma qualidade especial, que apresentem uma base genética e um nível adequado para a herança dentro desta espécie, como determinadas características morfológicas ou uma habilidade atlética/desportiva específica (MILHO, 2015).

As frequências alélicas de uma população podem ser influenciadas de quatro processos evolutivos: a seleção natural, a deriva genética, a migração e a mutação. Ao contrário da seleção natural, a deriva genética, causada pela amostragem aleatória dos alelos transmitidos para a próxima geração, é um mecanismo pouco valorizado como força evolutiva. No entanto, e particularmente quando as populações são pequenas, é um mecanismo poderoso, que pode alterar profundamente as frequências genéticas de uma população. O papel da deriva genética é também importante para a genética da conservação uma vez que as populações ameaçadas de extinção têm habitualmente um efetivo populacional baixo ou em declínio, fazendo com que estejam mais sujeitas à ação da deriva genética. Por outro lado, os parâmetros demográficos das populações determinados pela sua análise genética são estimados assumindo a neutralidade, ou seja, que as alterações na composição genética das populações ocorrem por deriva genética (FRANKHAN, 2005).

Um aspecto importante no melhoramento genético animal está no estudo de informações de *pedigree* que têm sido bastante utilizadas para avaliar a diversidade genética e também a estrutura populacional em diversas espécies de animais domésticos (FARIA et al., 2001; VERCESI FILHO et al., 2002).

Segundo Carneiro et al. (2009), apesar de várias técnicas da genética molecular estarem disponíveis, os estudos da estrutura populacional de uma espécie com base em informações de *pedigree* continuam sendo relevantes. O *pedigree* tem como vantagem o baixo custo e a simplicidade de obtenção, quando comparada ao uso de marcadores moleculares. Nos últimos anos, diversos estudos para avaliar a estrutura populacional de animais de produção têm sido realizados no Brasil e na América do Sul. Estes estudos, mediante as informações de *pedigree*, podem mostrar circunstâncias que afetam o histórico genético dessas populações (VALERA et al., 2005). Por meio dessas informações, aliada às informações fenotípicas, é possível realizar avaliações genéticas de reprodutores no intuito de gerar um progresso na elaboração de programas de melhoramento genético de qualquer espécie. Por isso, a coleta acurada da genealogia dos animais é de fundamental importância, pois as avaliações dependem de informações contidas nos *pedigree*.

O *pedigree* deve ser íntegro para avaliar a quantidade de descendentes presentes em cada geração de um determinado indivíduo, o que proporciona porcentagens de pais, avós,

bisavós, etc., em uma representação gráfica das gerações mais recentes, distinguindo as vias paternas e maternas (MACCLUER et al., 1983; NAVARRO, 2008). A integridade do *pedigree* é importante em estudos de estrutura de populações, uma vez que a qualidade e precisão das estimativas de parâmetros dependem da qualidade dos *pedigree*. As informações errôneas, *pedigree* incompleto e introdução recente de animais na população podem acarretar uma subestimação ou superestimação da endogamia dificultando a interpretação dos resultados (BOICHARD et al., 1997; GOYACHE et al., 2003).

2.5 Variabilidade genética e sua estimação

Dentro das diversas espécies domésticas de animais exploradas pelo homem atualmente, existe uma grande quantidade de raças e variedades. De acordo com Falconer e Mackay (1996), raça é o grupamento de indivíduos de uma mesma espécie que apresentam características fenotípicas semelhantes, mantidas nas diversas gerações pela fixação de alelos na população. A manutenção da variabilidade genética dentro de uma raça é primordial para manter o potencial produtivo, e evitar a erosão genética causada pela fixação excessiva de certos alelos (PINTO, 2010).

A variabilidade genética representa o conjunto de diferenças gênicas que pode existir em uma população, normalmente expresso em pequenas variações fenotípicas de morfologia ou de produção (RAMALHO et al., 2004). Segundo Euclides Filho (1999) é na existência dessa variabilidade que reside toda a capacidade de se realizar seleção e melhoramento animal, a sua perda acarreta o que se conhece como erosão genética. Contudo, não há melhoramento sem alguma perda de variabilidade genética, sendo necessário sua constante monitoração.

O risco de perda da variabilidade genética é grande em populações pequenas, mas pode se tornar muito maior em raças que vêm sendo melhoradas geneticamente por muito tempo, uma vez que nos programas de seleção, normalmente, busca-se aumento de produtividade, o que remete a uma utilização desenfreada de poucos ou até mesmo um único patriarca, com conseqüente fixação de certos alelos. Assim muitas vezes encontramos uma raça com grande quantidade de indivíduos e populações bem caracterizadas e desenvolvidas, mas com uma variabilidade genética pequena, já que praticamente todos os indivíduos são aparentados (MARIZ, 2010).

De acordo com Finland (2007), para que não ocorra perda de variabilidade ou erosão genética dentro de uma população são necessárias algumas medidas para manter essa

diversidade. Dentre elas pode-se citar: seleção com um uso maior de progenitores, diminuição da pressão de seleção, e monitoramento do tamanho efetivo e do coeficiente de endogamia na população.

As técnicas de seleção e acasalamento empregadas na formação de uma raça são as responsáveis diretas pela manutenção ou perda da sua variabilidade genética, segundo Euclides Filho (1999). A endogamia foi usada como técnica de acasalamento em vários programas de melhoramento animal em um passado recente, sendo ela o fator que mais contribuiu para a redução da variabilidade genética dentro de uma população ou mesmo de uma raça (GAMA, 2006). Porém, para que ocorresse a formação da maioria das raças equinas a endogamia foi intensamente utilizada, a fim de fixar nas progênes certas características desejáveis e/ou influências de determinados ancestrais (WAGONER, 1978). A endogamia é ainda preferida por vários criadores como estratégia de acasalamento devido à velocidade de melhoria do padrão racial e produtivo que ela imprime em um rebanho (FALCONER e MACKEY, 1996). Contudo, ela também é responsável pelo aparecimento de genes deletérios que podem ser expostos na população e posteriormente serem extintos, com a não seleção de animais afetados por tais genes.

Egito et al. (2002) comentam que a busca por uma padronização e alta produtividade foi trabalhada ao longo de muitos anos em algumas raças, através do uso massivo de possíveis reprodutores geneticamente superiores em várias populações ao mesmo tempo. Isso ocasionou e ainda continua ocasionando uma perda de variabilidade genética em diversos grupamentos genéticos, fato que pode chegar a comprometer sua exploração em pouco tempo, caso não sejam tomadas providências.

São diversos os métodos para estimar a variabilidade genética, porém a maioria baseia-se na avaliação demográfica da população e no estudo de suas frequências alélicas e genotípicas (FALCONER; MACKEY, 1996; RAMALHO et al., 2004). Segundo Navarro (2008), para medir a variabilidade e conhecer as consequências dos acasalamentos adotados em uma população, uma das informações mais significativas procede das análises de contribuição genética, atribuídas pelo coeficiente médio de parentesco e pelo número de animais fundadores e ancestrais.

Outra análise que pode ser realizada para determinar a variabilidade genética de uma raça ou população é a avaliação das condições de equilíbrio de Hardy-Weinberg que em dadas condições como ausência de migração, seleção, mutação e com acasalamento ao acaso, uma população grande manteria inalterada a frequência de determinados alelos, de forma que o termo equilíbrio sinaliza uma constância nas frequências alélicas e genotípicas com o passar

das gerações (HARDY; WEINGBERG, 1908). De acordo com Araújo (2001), uma população em equilíbrio apresenta duas propriedades básicas: a proporção de indivíduos heterozigotos nunca excede 0,5 (50%), e a proporção de heterozigotos é duas vezes a raiz quadrada do produto das duas proporções dos homozigotos.

Frankhan (2005) deixa claro que seleção, migração, mutação e deriva genética são os quatro fatores que podem interferir no equilíbrio de Hardy-Weinberg de uma população, ou seja, são as quatro forças que podem interferir na variabilidade genética da mesma.

A seleção pode agir negativamente, fixa alelos na população devido à utilização de um ou poucos indivíduos nos programas de acasalamento. Esse aspecto deletério surge com maior importância e intensidade na seleção artificial, guiada pelo homem, do que na seleção natural, na qual a fixação de alguns alelos na população refletiria em ganho adaptativo mais interessante (FRANKHAN, 2005). Já a migração genética está relacionada com a inserção de genes distintos na população e também pode contribuir para erosão genética da mesma, principalmente quando empregada através de ações de difusão de um genótipo exótico, tido como melhorador produtivo sobre raças locais, o que gera o desaparecimento desse material genético por uso indiscriminado de cruzamentos absorventes (SERENO 2002; EGITO et al., 2005; NIMBKAR et al., 2008).

A mutação é uma variação brusca e hereditária, fonte geradora da variabilidade genética, que pode ocorrer especificamente em um gene, o que pode determinar novas reações fisiológicas por ele comandadas, ou então em um cromossomo como um todo, devido a aberrações cromossômicas, sendo, contudo um fenômeno raro e casual. A maioria delas passa despercebida, ou por ocorrer em genes sem expressão direta ou por ser incompatíveis com a vida, mas algumas podem ser tidas como vantajosas, de acordo com Torres (1981) e Ramalho et al. (2004). A deriva genética é causada pela amostragem aleatória dos alelos transmitidos para a próxima geração. Há dois processos de amostragem alélica que causam a deriva genética. A primeira é a amostragem entre indivíduos que é devido à seleção dos pais. Nela apenas os alelos dos pais selecionados são transmitidos para a próxima geração. O segundo processo é a amostragem Mendeliana na qual é transmitido um dos dois alelos parentais para a progênie. Ambas as amostragens são completamente aleatórias, e assim a mudança na frequência alélica é também aleatória (GONÇALVES; OLIVEIRA, 2009).

Além de todas as ferramentas já citadas para avaliar a variabilidade genética também são bastante utilizadas a estimação do coeficiente de endogamia, do tamanho efetivo da população, estatísticas F de Wright e o intervalo de gerações (MARIZ, 2010).

Seja qual for a ferramenta estatística ou matemática escolhida para estudar a variabilidade genética de uma população, todos os dados utilizados terão de trazer informações a respeito de frequências alélicas ou genotípicas, pois essa é a base de trabalho de todos os métodos. Contudo, os dados colhidos em uma população podem ter duas origens principais, sendo a primeira os registros genealógicos ou *pedigree*, e a segunda, os dados demográficos ou moleculares desta população (MARIZ, 2010).

Em populações nas quais existem informações de *pedigree*, é possível utilizá-las como base de dados para verificar vários aspectos relacionados com a estrutura genética da população, através de banco de dados e submissão a programas computacionais com dados simples como nome de pai, nome da mãe e data de nascimento do animal. Com isso, estimam-se vários parâmetros como intervalos de gerações, número efetivo de fundadores, número efetivo de ancestrais, tamanho efetivo da população e coeficiente de endogamia, entre outros (GOYACHE et al., 2003). Bodó (1990) explica que em populações que não têm informações por *pedigree*, deve-se proceder à avaliação demográfica através da análise de idade, distribuição por sexo e número de machos e fêmeas que estão em reprodução que possibilitam estimar o coeficiente de endogamia, tamanho efetivo e intervalo de gerações que são os principais parâmetros a serem analisados em populações não genealógicas. Uma outra alternativa ao desconhecimento do *pedigree* é a análise de material genético por técnicas de biologia molecular, como por exemplo a utilização de marcadores moleculares que possibilitam a elaboração de dendrogramas de distância genética, além de vários estudos relacionados à detecção de genes de pureza racial ou que tem relação com algum elemento produtivo (RAMALHO et al., 2004).

2.6 Seleção artificial e suas consequências

A seleção é o processo que determina quais indivíduos se tornarão os progenitores da próxima geração, quantas progênes eles poderão produzir e por quanto tempo eles poderão permanecer na população. Ela é usada para fazer mudanças genéticas na população a longo prazo. A ideia geral da seleção é simples: indivíduos com os melhores conjuntos de alelos reproduzem, e desta forma, a próxima geração terá em média mais genes desejáveis. O resultado bem sucedido da seleção é a melhora genética das gerações futuras através do incremento da proporção de alelos desejáveis ao longo do tempo (GONÇALVES; OLIVEIRA, 2009).

Sendo a seleção uma das forças que afetam a mudança da frequência alélica, então podem ser fixados todos os alelos favoráveis, ao passo que todos os alelos não favoráveis podem se perderem com o passar do tempo, ou seja, a frequência dos alelos favoráveis tenderão a um, e a dos não favoráveis tenderão a zero. Portanto, a característica selecionada irá atingir um nível máximo quando todos os *locis* que afetam essa característica forem homozigotos para o alelo favorável (GONÇALVES; OLIVEIRA, 2009).

A seleção artificial contínua para características de produção pode resultar em mudanças fisiológicas através das quais outras características podem também ser afetadas positivamente ou negativamente. Portanto, a seleção artificial interrompida pode gerar uma resposta correlacionada indesejável, além de alterar as herdabilidades das características ao longo do tempo (GONÇALVES; OLIVEIRA, 2009).

As frequências alélicas variam de zero a um, sendo que estes valores podem ser atingidos em função dos métodos de seleção. Quando estas frequências extremas são alcançadas tornam-se irreversíveis. Assim, quando um alelo atinge a frequência um, ele é chamado de alelo fixado, e quando ele atinge a frequência zero, ele é chamado de alelo perdido. Quando um alelo atinge a fixação, nenhum outro alelo pode estar presente naquela população, nesta situação todos os indivíduos de uma linhagem são geneticamente idênticos. Esta é à base da uniformidade genética dentro das subpopulações (PINTO, 2010).

A probabilidade de fixação de um alelo vantajoso aumenta com o tamanho da população. Contudo, a probabilidade de perda deste alelo vantajoso é alta em populações pequenas (PINTO, 2010). Assim, a fixação de alelos é causada por uma seleção que utiliza um ou poucos progenitores nos programas de acasalamento o que pode levar à erosão genética de uma população.

2.7 Endogamia e seus efeitos

A definição de endogamia é o acasalamento entre indivíduos em que o grau de parentesco é maior que a média do parentesco da população, segundo Gonçalves e Oliveira (2009). A presença de ancestrais comuns na genealogia de um determinado animal causa um aumento na probabilidade desse animal receber cópias idênticas de um determinado gene a partir dos gametas de seus pais (PINTO, 2010).

O efeito primário da endogamia é provocar o surgimento de mais pares de genes em homozigoses, diminuindo assim, a porcentagem de heterozigotos, acarretando manifestações de genes recessivos que normalmente geram degenerações na média do mérito individual do

animal (ALBUQUERQUE, 2010). Navarro (2008) comenta que a endogamia promove a redução na capacidade adaptativa das raças e espécies, fato que diminui então sua capacidade de responder às adversidades do ambiente em gerações futuras.

Na formação da maioria das raças equinas, a endogamia foi intensamente utilizada, a fim de fixar nas progênes certas características desejáveis e/ou influências de determinados ancestrais (WAGONER, 1978). A endogamia nesta espécie foi primeiramente considerada por Steele (1944); Kieffer e Stewart (1976) e Kownacki e Jezierski (1979).

Não existe evolução genética sem variação genética. A endogamia excessiva promove degeneração produtiva racial devido à fixação de genes deletérios na população, no âmbito de desenvolvimento e reprodução que só podem ser revertidos com um aumento efetivo da heterozigotidade na mesma (FALCONER; MACKAY, 1996; QUEIROZ et al., 2000).

A endogamia é ainda preferida por vários criadores, como estratégia de acasalamento devido à velocidade de melhoria do padrão racial e de características produtivas que ela imprime em um rebanho. Contudo, ela também é responsável pela redução da variabilidade genética, redução no número de linhagens e de genótipos fundadores, aparecimento de genes deletérios, perda de genes e muitas vezes diminuição do desempenho médio em relação a determinadas características produtivas, principalmente ganho de peso e fertilidade, fatos relacionados com a depressão endogâmica (FALCONER; MACKAY, 1996; QUEIROZ et al., 2000).

De acordo com Weigel e Lin (2000), o nível de endogamia em rebanhos deve constituir uma preocupação constante, uma vez que os modernos programas de seleção identificam acuradamente famílias com potencial genético superior e tecnologias reprodutivas são usadas eficientemente para que ocorra difusão desses genótipos na população.

A depressão endogâmica é a ação conjunta de genes deletérios em uma população que diminui significativamente o desempenho dos animais (GONÇALVES; OLIVEIRA, 2009). Johanson e Rendel (1971) comentam que existem três causas que podem provocar a depressão endogâmica. A primeira é devido à segregação de genes recessivos que normalmente são desfavoráveis quando em homozigose, a segunda causa pode ser devido às frequências reduzidas de genótipos heterozigotos, e a terceira causa à interação entre genes situados em diferentes *locis*, ou seja, efeito epistático.

De acordo com Cassel et al. (2003), a detecção da depressão endogâmica é importante e depende da acurácia das estimativas de coeficiente de endogamia da população. Contudo, fica clara a importância que a gestão genética populacional tem para o cenário produtivo

mundial, não só para raças em risco de extinção, mas também para raças com grande contingente populacional e de grande magnitude exploratória.

2.8 Parâmetros demográficos

A seguir descreve-se alguns parâmetros que caracterizam a estrutura genética de uma população.

2.8.1 Intervalo de geração

O intervalo de gerações refere-se à idade média dos progenitores ao nascimento dos descendentes que serão utilizados como reprodutores na geração subsequente. Esse intervalo pode ser avaliado separadamente para machos e fêmeas, permitindo melhor diagnóstico da estrutura de acasalamentos dos animais (FALCONER; MACKAY, 1996; JAMES, 1977). O parâmetro permite verificar o tempo de transferência dos genes entre pais e filhos. Por esse motivo, é necessário ser valorizado, uma vez que, é importante estimar o avanço genético nas características de interesse. (FALCONER; MACKAY, 1996; PEREIRA, 2012).

Na análise de estrutura populacional, o intervalo de gerações é importante uma vez que as perdas de variabilidade genética ou o melhoramento genético de uma população podem ser avaliados por geração. Contudo, se o intervalo de geração é longo, as perdas de variabilidade genética ocorrem em menor velocidade e o melhoramento genético vai ser também mais demorado (FALCONER; MACKAY, 1996).

2.8.2 Coeficiente e taxa de endogamia

O coeficiente de endogamia é outra ferramenta bastante utilizada para estimar a variabilidade e o grau de homozigose em uma população. Ele é normalmente representado pela letra F, e reflete a probabilidade de um indivíduo, em um determinado *locus*, apresentar dois alelos idênticos devido a sua ascendência genealógica. Assim, pode-se dizer que o coeficiente de endogamia indicará a porcentagem de genes que eram heterozigotos e que se tornaram homozigotos nos filhos devido ao parentesco entre seus progenitores (GIANNONI, 1987; PEREIRA, 2001).

O valor de F pode variar de 0 a 100% e, quanto maior for o valor de F, maior terá sido a dissipação ou perda da heterose na população (MARIZ, 2010).

De acordo com Wright (1923), o parâmetro F mede a correlação entre os gametas que se unem para formar um zigoto, e equivale à metade do coeficiente de parentesco dos pais, quando um ancestral comum é não endogâmico.

Somente o valor de F não é suficiente para compreensão do comportamento da endogamia na população. O “incremento da endogamia” ou “aumento da endogamia” (ΔF), que mede quanto o coeficiente de endogamia varia proporcionalmente de uma geração para outra subsequente, torna possível comparar os efeitos da endogamia sob diferentes sistemas de acasalamento (FALCONER, 1989).

De acordo com a FAO (1998), para se ter uma variabilidade genética em nível aceitável sem prejudicar os ganhos de seleção o ΔF entre duas gerações não deve ser superior a um por cento, e um tamanho efetivo mínimo de 50 animais.

2.8.3 Relação genética aditiva

A relação genética aditiva pode ser interpretada como uma representação do indivíduo na linhagem ou no *pedigree*, independentemente da geração no qual está presente. Portanto, é como uma medida da proporção de cada indivíduo com outros indivíduos na árvore genealógica. O coeficiente de relação genética aditiva é o dobro do coeficiente de parentesco, em que o coeficiente de parentesco entre dois indivíduos é igual ao coeficiente de endogamia de um descendente de ambos (MILHO, 2015).

2.8.4 Tamanho efetivo da população

O tamanho efetivo é definido como o tamanho de uma população idealizada que leva em consideração o número de machos e de fêmeas que estão efetivamente se reproduzindo no rebanho. Este é o número de indivíduos que dariam origem à variância da amostragem ou a taxa de endogamia apropriada às condições em consideração, se eles fossem criados na forma da população idealizada (FALCONER, 1981). Wright (1938) define o tamanho efetivo da população como um parâmetro que quantifica a perda de alelos entre gerações.

O conceito de tamanho efetivo tem papel importante nas teorias evolutivas (WRIGHT, 1938), na genética de populações, na genética quantitativa (CABALLERO; HILL, 1992) e conservação genética (SUGG; CHESSER, 1994). O tamanho efetivo teórico difere do tamanho censo de uma população, sendo muito menor (WRIGHT, 1938).

A seleção de reprodutores interfere no tamanho efetivo, e causa um decréscimo perigoso no número de indivíduos difusores de genes na população (FARIA et al., 2002).

Muitos estudos têm sido realizados para determinar o tamanho efetivo ótimo para diminuir o efeito da seleção sobre a variabilidade genética.

2.8.5 Probabilidade de origem gênica

Os genes totais presentes em uma população são provenientes de seus fundadores. Contudo, a representação de cada um dos fundadores nos indivíduos tomados como referência é muito variável, o que depende da utilização preferencial ou não dos mesmos. Assim, a probabilidade de origem dos genes de uma população proporciona uma informação importante para a compreensão do fluxo gênico, e conseqüentemente, da estrutura genética da população, segundo Falconer e MacKay (1996). Deste modo, para estimar a variabilidade genética e conhecer as conseqüências da política de cruzamentos realizada na população, uma das informações mais interessantes vem da análise da contribuição genética proveniente de animais fundadores e de ancestrais.

O conceito de fundador e ancestral são normalmente confundidos. O fundador é um indivíduo cujos pais são desconhecidos. Quando um indivíduo tem um único parente desconhecido este considera-se um “Fundador Fantasma”, são aqueles que formam a população base do rebanho (RODRIGUES, 2009). Já o ancestral é o indivíduo que tem influenciado mais do que os seus ascendentes na variabilidade genética presente na população, e um ancestral também pode ser um fundador (FALCONER; MACKAY, 1996).

A variabilidade genética é influenciada pelo número de fundadores de uma população. Teoricamente quanto maior o número de fundadores, maior será a variabilidade genética. De acordo com Boichard et al. (1997) pode-se calcular as contribuições genéticas dos fundadores para saber como o *pool* gênico original que vem sendo mantido através das gerações. A contribuição genética é maior à medida que se produzem mais descendentes (LAAT, 2001).

2.9 Caracterização genética

A caracterização da diversidade de uma raça pode ser realizada por suas propriedades fenotípicas, morfológicas e fisiológicas, origem, habitat, distribuição geográfica, produção ou

parâmetros genéticos (MOHAMMED, 1997; VALE FILHO et al., 1997; Del LAMA et al., 1996; HERNANDEZ et al., 1997).

Ao contrário da caracterização fenotípica, a caracterização genética não tem influências do meio ambiente, o que propicia uma maior acurácia dos dados produzidos, sendo importante nas decisões a serem tomadas em programas de conservação e de melhoramento (FITZHUGH; STRAUSS, 1992). Além disso, possibilita a indicação de acasalamentos ou cruzamentos que irão manter a máxima variabilidade genética dentro de uma raça, sendo fundamental para o acompanhamento racional de futuros programas de melhoramento animal, facilitando assim, a rápida adaptação às mudanças necessárias e imprevistas para o desenvolvimento dos sistemas de produção, uma vez que não é possível prever com objetividade quais características poderão ser necessárias no futuro. Esta diversidade é refletida na variedade de tipos e raças que existem e na variação presente dentro de cada uma (EGITO, 2007).

Com a descoberta e a utilização de marcadores moleculares, com a finalidade de discriminar humanos ou animais, foi possível um relevante avanço na área de identificação e caracterização genética de uma raça nas duas últimas décadas. De fato, na maioria das vezes, nesta área da ciência, o conhecimento desenvolvido para os estudos de animais foi adquirido através das técnicas já utilizadas em seres humanos, fazendo-se assim uma adaptação para a espécie pesquisada (GASTALDO, 2017).

2.10 Marcadores moleculares

Os marcadores genéticos são utilizados para marcar/etiquetar/identificar alelos cuja expressão seja de difícil identificação. Com isso, pode-se selecionar o alelo de interesse de forma indireta, por meio do marcador, auxiliando no processo de seleção. Os marcadores genéticos podem ser morfológicos ou moleculares.

Os marcadores moleculares são aqueles à base de proteínas (chamados de bioquímicos) ou os que usam o próprio ácido desoxirribonucleico - DNA (marcadores moleculares, propriamente ditos). Os marcadores moleculares de proteínas são os produtos diretos dos alelos, os mais usados são as isoenzimas - esterase, fosfatase, peroxidase, grupos sanguíneos, antígenos linfocitários e imunoglobulinas, entre outros. Como eles são os produtos diretos dos alelos de interesse, basta identificá-los para selecionarmos o indivíduo com o fenótipo desejado. Os grupos sanguíneos foram usados, por exemplo, no diagnóstico de freemartinismo em bovinos. Assim como os marcadores morfológicos, os bioquímicos

possuem reduzida variabilidade, ou seja, o número de marcadores em uma espécie é relativamente pequeno. Portanto, os alelos responsáveis pelas proteínas facilmente identificáveis, que são os marcadores, não ocorrem em número suficiente para marcar um grande número de alelos de interesse de vários genes, o que torna sua utilidade reduzida. Contudo, são marcadores simples, de fácil identificação, de menores custos além de, serem codominantes (identificam-se os homozigotos e heterozigotos) (LEMOS,1995).

Os marcadores moleculares de DNA são os próprios genes ou seus vizinhos, isto é sequências de DNA situadas próximas ao gene que se deseja marcar. A maior vantagem do DNA é a grande variabilidade que se observa entre os indivíduos de certa espécie, o que equivale a dizer que utilizando o DNA consegue-se um número de marcadores ou “etiquetas” suficientes para marcar todos os alelos de todos os genes da espécie, além de não ser objeto de influências ambientais. A desvantagem dos marcadores de DNA é o fato de as técnicas laboratoriais e os equipamentos serem mais caras quando comparadas aos outros tipos de marcadores. Os marcadores moleculares mais comuns são citados a seguir: **RFLP**, que corresponde à expressão em inglês *Restriction Fragment Length Polymorphism* e significa polimorfismo de comprimentos de fragmentos de restrição; **VNTR**, Minissatélites que corresponde a *Variable Number of Tandem Repeat*; **RAPD** cuja expressão significa polimorfismo de fragmentos aleatórios e amplificados de DNA, da expressão em inglês *Random Amplified Polymorphic DNA*; **SCAR** do inglês *Sequence Characterized Amplified Regions* que que significa sequências amplificadas e caracterizadas; **SSR** (*Simple Sequence Repeat*), que significa repetições de sequências simples ou microssatélites; **CAPS** (*Cleaved Amplified Polymorphic Sequence*) que significa sequência polimórfica amplificada e clivada; **AFLP**, da expressão *Amplified Fragment Length Polymorphism*, que significa polimorfismo de comprimento de fragmentos amplificados e **SNP** de *Single Nucleotide Polymorphism* que que é polimorfismo de nucleotídeo único (SANTOS, 2014).

2.10.1 Microssatélites (SSR)

Os SSR são repetições de sequências em *tandem* geralmente de dois a seis nucleotídeos ao longo da molécula de DNA, ocupando uma extensão de até 100 pares de bases. A sequência mais comumente repetida em animais é a (AC)_n. Essas sequências são distribuídas em todo o genoma de forma aleatória e são flanqueadas por sequências altamente conservadas. As sequências de SSR ocorrem tanto nas regiões codificadoras quanto naquelas que não codificadoras, porém, suas frequências são maiores nas sequências transcritas e

existem algumas regiões com maiores abundâncias, como as regiões próximas aos centrômero e telômeros (GUIMARÃES, 2009).

A origem do SSR ocorre durante a replicação do DNA devido ao pareamento desalinhado de sequências repetidas, chamado também de “escorregamento”. A consequência desse escorregamento é a produção de números variados das unidades repetidas, de forma que cada fragmento de microssatélite assume comprimentos diferentes, que podem ser identificados em eletroforese. Cada um desses fragmentos de um dado *locus* é equivalente a um alelo de um gene, tornando assim um dos marcadores com maior variabilidade alélica, ou seja, muito polimórfico.

Os fragmentos utilizados como marcadores devem ser aqueles flanqueados por sequências conservadas do DNA. Com isso, são sequenciados esses flancos e construídos *primers* complementares a eles e que irão amplificar os microssatélites.

Para identificar os microssatélites existem duas estratégias básicas: uma por busca de sequências que possuem microssatélites nas bases de dados na internet e outra por construção e seleção de livraria/biblioteca genômica ou de ESTs (*Expressed Sequence Tag*), que são sequências de cDNA provenientes de RNA. A primeira estratégia básica é menos onerosa e rápida, já que as sequências já estão disponíveis nos bancos de dados. Para se construir a livraria, um dos métodos, chamando de “não enriquecido”, é isolar o DNA da espécie de interesse, realizar a digestão do DNA com a enzima de restrição apropriada e separação dos fragmentos por eletroforese, selecionar os fragmentos entre 300 a 1000 pb, fazer ligação em vetores para clonagem, hibridizar os clones com sondas constituídas de várias repetições, sequenciar os fragmentos positivos e, finalmente desenhar os *primers* complementares às regiões flanqueadoras. Esse procedimento é considerado pouco eficiente devido ao fato de apenas uma pequena porcentagem dos clones resultarem em *loci* marcadores.

Outro procedimento que tem sido preferido é a construção de biblioteca “enriquecida”, na qual é realizada a hibridização seletiva com fragmentos de DNA, usando esferas magnéticas cobertas com estreptoavidina ou membranas de nylon. Nesse procedimento, ocorre a digestão do DNA e a ligação dos fragmentos a adaptadores, hibridização com sondas de microssatélites com biotina, seguida pela ligação às esferas com estreptoavida, eluição dos fragmentos de DNA ligados às esferas e amplificação com *primers* complementares às sequências adaptadoras, clonagem dos produtos amplificados em vetores, transformação de *E.coli* e sequenciamento dos clones positivos. Estes métodos citados são geralmente demorados, assim uma proposta alternativa consiste na identificação de fragmentos de RAPD ou de AFLP que contém os microssatélites, por meio da hibridização com sondas de

microssatélites e clonagem dos fragmentos positivos. O processo de obtenção de *primers* é semelhante aos procedimentos anteriores (SANTOS, 2014).

Os marcadores microssatélites foram descobertos em 1992 e desde então passaram a ser o marcador mais estudado. Todavia, com o desenvolvimento de métodos de genotipagem de marcadores do tipo Polimorfismo de nucleotídeo único (SNP), os SSR passaram a ser menos utilizados.

As principais vantagens do uso deste marcador são sua herança de codominância, alto nível de polimorfismo, frequência e distribuição no genoma, boa repetibilidade e possibilidade de detecção em sistemas *multiplex*. Como desvantagem pode-se citar o custo para desenvolver *primers* específicos de espécies que não estão disponíveis no banco de dados ou que não foram estudadas (FALEIRO et al., 2004).

2.10.2 Polimorfismo de nucleotídeo único (SNP)

O SNP é uma variação genética em um único nucleotídeo. Um exemplo de um SNP é a alteração do segmento de DNA “AAGGTTA” para “ATGGTTA”, no qual o segundo A (adenina) da primeira sequência foi substituído por um T (timina). Não somente a substituição pode ocorrer, como também pode ter uma mutação por adição ou deleção de uma única base, chamado de *indels*. Estas mutações ocorrem raramente e a maioria desses polimorfismos ocorrem em regiões que não codificam uma proteína (CHO et al., 1999).

Segundo Arbex (2009), os SNP são polimorfismos de evolução lenta, responsáveis pela formação de novos alelos, os quais representam diferentes formas com que um gene pode se apresentar. Essas formas podem ser bi, tri ou tetra-alélicas, ou seja, podem possuir duas, três ou quatro formas diferentes, sendo que os SNP bi-alélicos são os mais comuns (CRISTIANINI; HAHN, 2007).

A substituição de um único nucleotídeo pode levar à substituição de um aminoácido na proteína, e conseqüentemente a perda da função da mesma e alteração fenotípica (SOORI et al., 1999).

Para identificar os polimorfismos é necessário sequenciar várias sequências homólogas na população, e a partir daí, ter-se a ideia das mutações de ponto que ocorreram. As mutações mais encontradas são as do tipo transição, em que há troca de purina por outra purina ($A \leftrightarrow G$) ou de uma pirimidina por outra pirimidina ($C \leftrightarrow T$). As mutações de transversões são menos frequentes e acontecem quando há troca de uma purina por uma pirimidina, ou vice-versa ($C \leftrightarrow G$, $C \leftrightarrow A$, $T \leftrightarrow G$ ou $T \leftrightarrow A$). Caso essas sequências homólogas

sejam alelos de um gene, pode-se ter ideia das mutações que os geraram. Por serem encontrados em regiões codificadoras, os SNP vêm sendo muito empregados na análise de *EST*, isto é, das sequências expressas identificadas, que são provenientes dos RNA mensageiro. Entretanto, devido ao custo do sequenciamento, técnicas alternativas vêm sendo propostas para reduzir o custo de identificação dos SNP, de acordo com Soleimaniet et al. (2003).

Segundo Kwok et al. (1996), métodos de genotipagem de SNP foram desenvolvidos no final da década de 90 e a partir deste momento passou a ser o marcador molecular mais utilizado devido às suas várias vantagens como a possibilidade de detecção de grande quantidade de polimorfismos de alelos de um mesmo gene, ser codominante e por ser o marcador de maior abundância no genoma, além de ser, entre todos os marcadores, o que permite identificar mais polimorfismos entre os indivíduos. Entretanto, é necessário o conhecimento prévio da sequência do gene de interesse, o que é de custo elevado, devido a toda infraestrutura da técnica envolvida como as etapas de sequenciamento dos fragmentos de interesse (FALEIRO, 2007).

2.11 Análises estatística com uso de marcadores moleculares

Após obtenção dos genótipos dos indivíduos que representam significativamente a população estudada, podem ser utilizados testes estatísticos para a realização do estudo de caracterização genética de uma raça. As análises mais importantes são: o teste de Equilíbrio de Hardy-Weinberg, teste de equilíbrio de ligação, e estimação de parâmetros como as heterozigosidades esperada e observada, conteúdo de informação polimórfica, estatística F e distâncias genéticas que também são índices que devem ser estimados quando são realizados estudos populacionais utilizando marcadores moleculares SSR ou SNP.

2.11.1 Equilíbrio de Hardy-Weinberg

O equilíbrio de Hardy-Weinberg, ou também conhecido como o princípio do equilíbrio gênico, prevê a estabilidade das frequências dos alelos e dos genótipos de uma geração para outra. A razão para realizar esse teste, quando se está estimando frequências alélicas, é determinar se, dentro de um *locus*, estes são independentes, sendo que um *locus* é considerado geneticamente estável se suas frequências não alterarem ao longo das gerações (GASTALDO, 2017).

Para uma população se manter em equilíbrio, segundo a teoria de Hardy e Weinberg, são necessárias as seguintes condições: a população deve ser muito grande, de modo que possam ocorrer todos os tipos de acasalamentos possíveis; a população deve ser panmítica, ou seja, os acasalamentos entre os indivíduos devem ocorrer ao acaso; não ocorrer nenhum fator evolutivo, como mutação, seleção ou migração. Deste modo a população seria ideal e permaneceria em equilíbrio gênico, ou seja, as frequências dos alelos não sofreriam alterações ao longo das gerações. Entretanto, é muito difícil que as populações atendam a todas as condições deste princípio, o que ocasiona a alteração da frequência dos alelos ao longo do tempo. Sendo os alelos herdados de forma mendeliana, é possível prever a ocorrência destes seguindo um padrão previsível de probabilidade. As frequências alélicas são utilizadas para calcular as frequências genotípicas esperadas que são então comparadas com as frequências genotípicas observadas, caso os valores observados e esperados forem semelhantes, então presume-se que os alelos dentro do *locus* são estáveis e estão em equilíbrio. De acordo com Butler (2005), uma outra finalidade do teste de equilíbrio é procurar quaisquer indicações de excesso de homozigose, quando o número de homozigotos observados apresenta diferença significativa em relação número de homozigotos esperados. Qualquer desvio de equilíbrio, por excesso ou deficiência de heterozigose, significa que a população está subdividida e que pode existir uma endogamia significativa ou fluxo de genes com outras populações.

No teste de equilíbrio utiliza-se normalmente o nível de significância (valor de P) igual a 0,05, e se o *locus* apresentar valores maiores que 0,05 ($P > 5\%$) pode-se dizer que a população está em equilíbrio para este marcador genético. Contudo, com valores abaixo de 0,05 ($P < 5\%$) há desequilíbrio, e o resultado é significativo.

2.11.2 Desequilíbrio de ligação

A combinação das informações de vários *loci* ao mesmo tempo pode fortalecer as análises estatísticas e torna o uso dos marcadores moleculares uma poderosa ferramenta de identificação individual ou de uma raça. Porém, é preciso que os *loci* sejam independentes uns dos outros, isto é, que ocorra a recombinação entre eles, durante o processo de divisão celular.

O desequilíbrio de ligação é a associação não aleatória de alelos em dois ou mais *loci*, não necessariamente em um mesmo cromossomo (FALCONER e MACKAY, 1996). Esse parâmetro descreve uma situação em as frequências dos duplos heterozigotos em *trans* e *cis* são iguais. O equilíbrio de ligação pode ser atingido quando todas as condições do equilíbrio de Hardy-Weinberg citadas acima são reais em uma população.

Ao contrário do equilíbrio de Hardy-Weinberg, que precisa de apenas de poucas gerações para ocorrer, o equilíbrio de ligação ocorre de forma mais lenta e gradual, e a velocidade para atingir esse equilíbrio dependerá da taxa de recombinação para os genótipos heterozigotos para ambos os *loci* envolvidos. Já a frequência dessa recombinação entre os *loci* irá depender do fato de estarem localizados em um mesmo cromossomo e da distância física entre os mesmo. Quanto mais próximos os *loci* estiverem um do outro, menor a chance da quebra e de união acontecer na região entre os *loci* envolvidos, e assim a migração destes no processo meiótico ocorrerá em bloco, sem o acontecimento da recombinação cromossômica, ou seja, quanto mais ligados estiverem os genes, mais tempo será necessário para se aproximar do equilíbrio. Quando isto ocorrer, há um desequilíbrio de ligação entre eles, conseqüentemente, quanto maior a distância, mais provável é a ocorrência da recombinação entre os dois *loci*, favorecendo o equilíbrio de ligação. De acordo com Silva (2001), o desequilíbrio de ligação é definido como qualquer desvio das combinações genotípicas de dois *loci* sob acasalamentos ao acaso.

Hartl e Clark (2010) relatam que é importante enfatizar que para ocorrer o desequilíbrio de ligação não é necessário que os *loci* estejam fisicamente ligados, existe também o desequilíbrio de ligação entre *loci* localizados em cromossomos diferentes. Por esse motivo, quando são estimadas as frequências alélicas de um grupo de marcadores moleculares, é aconselhável que seja realizado um teste de desequilíbrio de ligação entre todos os *loci* utilizados para verificar se está ocorrendo ou não. Caso não haja certeza de que todos os marcadores e seus alelos foram herdados independentemente, os resultados estimados podem perder o devido valor e significado.

2.11.3 Heterozigosidades esperada e observada

A heterozigosidade esperada, conhecida também como diversidade genética, é calculada para estudar a variação genética em uma população, sendo ela a probabilidade de um indivíduo ser heterozigoto no *locus* marcador. Esse parâmetro fornece informações sobre a estrutura e a história de uma população na qual valores altos significam que existe maior diversidade de alelos, e valores baixos de heterozigosidade podem indicar que o tamanho populacional é pequeno ou que a variabilidade genética está reduzida (BUTLER, 2005; NEI, 1978). O valor do parâmetro pode variar de zero, ou seja, sem heterozigose, a máxima heterozigose espera de 0,5 (quanto mais alto o valor, mais as frequências alélicas estarão próximas da igualdade). Já a heterozigosidade observada é a proporção de indivíduos

heterozigotos nas amostras da população total, sendo este valor comparado com a heterozigosidade esperada para ajustes ao teste de equilíbrio de Hardy-Weinberg.

2.11.4 Conteúdo de informação polimórfica

Do ponto de vista genético, o polimorfismo é como a variação genotípica existente dentro das populações (HARTL; CLARK, 2010). A nível molecular, o termo é frequentemente mais utilizado para descrever a variação genotípica, o que inclui a variabilidade alélica e as alterações dentro de um gene ou do genoma. O termo “Conteúdo de Informação de Polimorfismo” (PIC) que passou a ser utilizado nas análises genética através do DNA, indica a probabilidade de um parental ser heterozigoto num *locus* e o outro parental ter um genótipo diferente, sendo calculado em função do número de alelos detectados, da sua distribuição e frequência na população estudada (BOTSTEIN et al., 1980). Segundo o mesmo autor, esse parâmetro é um indicador da qualidade do marcador em estudos genéticos, usado, principalmente, para avaliar a diversidade de um gene ou segmento de DNA em uma população.

O valor será de quase zero se não houver variação alélica e pode atingir o máximo de um se um genótipo tiver apenas novos alelos, o que é considerado um fenômeno raro. De acordo com Botstein et al. (1980), o PIC segue a seguinte classificação: valores superiores a 0,5, são considerados muito informativos, valores entre 0,25 e 0,50 mediantemente informativos, e valores inferiores a 0,25, pouco informativos.

2.11.5 Estatísticas F de Wright

As conhecidas “estatísticas F” (Fis, Fst e Fit), também denominadas de índices de fixação, foram elaboradas por Sewall Green Wright na intenção de descrever as propriedades das populações que são subdivididas. A subdivisão de uma população pode gerar um efeito semelhante ao efeito da endogamia, resultando em um excesso de homozigotos dentro da população. Desta forma, é importante medir esse efeito em termos de redução na proporção de genótipos heterozigotos (PINTO, 2010).

Os coeficientes F de Wright (1951) são tipos de “coeficientes de endogamia”, entretanto, eles diferem de acordo com a população de referência. São todos calculados baseados na heterozigosidade esperada e observada. Se considerarmos uma população como

um todo (T), suas subpopulações (S) e seus indivíduos (I), podem ser definidos os três parâmetros de Wright: Fit como a correlação entre os gametas que se unem para formar os indivíduos em relação à população, Fst como a correlação entre gametas tomados ao acaso nas subpopulações e que mede o nível de diferenciação genética entre subpopulações e Fis como a correlação entre gametas que se unem para produzir indivíduos com relação à subpopulação.

O coeficiente de endocruzamento (Fis) pode ser definido também como a probabilidade de que dois alelos de um *locus* em um indivíduo endocruzado sejam idênticos por descendência, ou seja, é o coeficiente de endogamia de um indivíduo em relação à subpopulação da qual ele pertence. Fst é definido também como o coeficiente de endogamia da subpopulação em relação à população total da qual ela é parte, sendo uma estatística informativa utilizada para estimar o nível de divergência genética entre subpopulações ou raças, e embora tenha um mínimo teórico de zero, ou seja, sem divergência genética, e um máximo teórico de um, ou seja, há fixação de alelos alternativos em diferentes populações, na prática em si, os valores observados são muito menores que um (HARTL; CLARK, 2010). A interpretação dos valores de Fst de Wright deve seguir as seguintes orientações (WRIGHT, 1978): a amplitude de 0 a 0,05 pode ser considerada indicativa de pequena diferenciação genética; a amplitude de 0,05 a 0,15 indica moderada diferenciação genética; a amplitude de 0,15 a 0,25 indica grande diferenciação genética; valores de Fst acima de 0,25 indicam diferenciação genética muito grande. Fit também é definido como o coeficiente de endogamia que leva em consideração tanto o efeito do acasalamento não aleatório quanto os efeitos da subdivisão da população. É uma medida de heterozigosidade de um indivíduo em relação ao total da população (PINTO, 2010).

2.11.6 Distância genética

A estimativa da unicidade genética de uma dada raça é, normalmente, obtida a partir da mensuração da distância genética e a construção de uma árvore filogenética simples que expresse as relações entre as diferentes populações de uma maneira gráfica (BARKER, 1994). O estudo da variação genética, envolvendo várias populações e muitos genes, permite avaliar a quantidade de variação comum entre populações. Estas medidas de variação comum entre grupos são chamadas de distâncias genéticas e permitem a visualização da relação entre grupos de indivíduos (PINTO, 2010).

A análise da distância genética demonstra a relação entre populações e é importante para reconstruir a relação entre os indivíduos de uma população. De acordo com Naqvi (2007), similaridades genéticas e dissimilaridades podem ser observadas entre duas populações ou indivíduos, e serem utilizadas para a caracterização de raças distintas e para a avaliação da variação das espécies através do tempo. Em uma análise de distância genética, a diferença genética entre populações é estimada baseada na diferença entre distribuições de frequências alélicas em vários *loci* ou baseada na distribuição de tamanho de alelos (LAVAL; SANCRISTOBAL; CHEVALET, 2002).

Várias medidas de distância genética foram desenvolvidas ao longo do tempo com o uso de marcadores moleculares e podem ser classificadas em duas categorias principais: aquelas sem suposições biológicas, chamadas de distâncias geométricas como a de Cavalli-Sforza (CAVALLI-SFORZA; EDWARDS, 1967), e as com suposições biológicas ou modelos como o de Nei (NEI 1972, 1978), a distância de Reynold (REYNOLDS; WEIR; COCKERHAM, 1983) e a distância de Goldstein (GOLDSTEIN et al., 1995). A estimação das distâncias genéticas pode usar também caracteres morfológicos, fisiológicos ou agrônômicos e ainda podem ser empregadas análises de *pedigree* ou medidas de depressão endogâmica.

REFERÊNCIAS

- ASSOCIAÇÃO BRASILEIRA DOS CRIADORES DO CAVALO MANGALARGA MARCHADOR (ABCCMM). **Mercado do Mangalarga Marchador**. Disponível em: <https://g1.globo.com/sao-paulo/itapetininga-regiao/mfruralagronegocio/noticia/mercado-do-mangalarga-marchador-cresce-10-no-ano.ghtml>. Acesso em: 05 ago. 2018.
- ALBUQUERQUE, A. L. S. **Estrutura populacional de um rebanho leiteiro da raça Pardo-Suíço no Estado do Ceará**. 2010. 48 p. Dissertação (Mestrado em Produção Animal) - Centro de Ciências Agrárias da Universidade Federal do Ceará, Fortaleza. 2010.
- ARAÚJO, A. M. O salto qualitativo em Theodosius Dobzhansky: unindo as tradições naturalista e experimentalista. **História, Ciências, Saúde**, v. 8, n. 3, p. 713-726, 2001.
- ARBEX, W. A. **Modelos Computacionais para Identificação de Informação Genômica Associada à Resistência ao Carrapato Bovino**. 2009. 200 p. Tese (Doutorado em Engenharia de Sistemas de Computação) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.
- BARKER, J. S. F. A global protocol for determining genetic distances among domestic livestock breeds. In: **WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION**, 5., 1994, Guelph. **Proceedings...**Guelph: World Congress on Genetics Applied to Livestock Production, 1994. p. 501-508.
- BODÓ, I. Methods and experiences with *in situ* preservation of farm animals. In: **FAO, ANIMAL PRODUCTION AND HEALTH PAPER**, 1990, Roma. **Proceedings...** FAO, Animal Production and Health Paper, 1990, p. 85-102.
- BOICHARD, D.; MAIGNEL, L.; VERRIER, E. The value of using probabilities of gene origin to measure genetic variability in a population. **Genetics Selection Evolution**, v. 29, p. 5-23, 1997.
- BOTSTEIN, D.; WHITE, R. L.; SKOLMICK, H. et al. Construction of a genetic linkage map in man using restriction fragment length polymorphism. **American Journal of Human Genetics**, v. 32, p. 314-331, 1980.
- BUTLER, J. M. **Forensic DNA Typing: Biology Technology and Genetics of STR Markers**, 2nd ed., Academic Press, 2005.
- CABALLERO, A.; HILL, W. G. A note on the inbreeding effective population size. **Evolution**, v. 46, n. 6, p. 1969-1972, 1992.
- CARNEIRO, H.; LOUVANDINI, H.; PAIVA, S. R.; MACEDO, F.; MERNIES. B.; Mc MANUS, C. Morphological characterization of the sheep breeds in Brazil, Uruguay and Colombia. **Small Ruminant Research**, v. 94, p. 58-65, 2009.
- CASSEL, B. G.; ADAMEC, V.; PEARSON, R. E. Effect of incomplete pedigree on estimate of inbreeding on inbreeding depression for days to first service and summie milk yield in Holsteins and Jerseys. **J. Dairy Sci.**, v. 86, p. 2967-2976, 2003.

CAVALLI-SFORZA, L. L.; EDWARDS, A. W. F. Phylogenetic analysis models and estimation procedures. **American journal of human genetics**, v. 19, n. 3, p. 233–57, 1967.

CHO, R. J.; MINDRINOS, M.; RICHARD, D. R.; SAPOSKY, R. J.; ANDERSEON, M. et al. Genome- wide mapping with biallelic markers in *Arabidopsis thaliana*. **Nature Genetics**, v. 23. p. 203-207, 1999.

COSTA, M. D.; BERGAMANN, J. A. G.; PEREIRA, C. S. et al. Avaliação dos fatores genéticos e de ambiente que interferem nas medidas lineares dos pôneis da raça brasileiras. **Revista Brasileira de Zootecnia**, v. 3, p. 491-497, 1998.

COSTA, M. D. et al. Caracterização demográfica da raça Mangalarga Marchador. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, Belo Horizonte, v. 56, p. 687-690, 2004.

CRISTIANINI, N.; HAHN, M. W. **Introduction to computational genomics: a case studies approach**. Cambridge: Cambridge University Press, 2007.

LEESON, S.; SUMMERS, J. D. **Commercial poultry nutrition**. 3rd ed. Nottingham: Nottingham University Press, 2005. 414 p.

CRUZ, C. D. **Princípios da genética quantitativa**. Viçosa: UFV, 2005, 394 p.

DEL LAMA, S. N.; DEL LAMA, M. A.; MESTRINER, M. A. et al. Esterase D polymorphism and partial characterization of red blood cell esterases in cattle and buffalo populations. **Braz. J. Genet.**, v. 19, p. 243-248, 1996.

EGITO, A. A.; MARIANTE, A. S.; ALBUQUERQUE M. S. M. Programa brasileiro de conservação de recursos genéticos animais. **Archivos de Zootecnia**, v. 51, p. 39-52, 2002.

EGITO, A. A.; MARIANTE, A. S.; CASTRO, S. T. R. et al. Situação atual do banco de DNA de recursos genéticos animais no Brasil. **Archivos de Zootecnia**, v. 54, p. 283-288, 2005.

EGITO, A. A. **Diversidade genética, ancestralidade individual e miscigenação nas raças bovinas no Brasil com base em microssatélites e haplotipos de DNA mitocondrial: Subsídios para a conservação**. 2007. 246 p. Tese (Doutorado em Biologia Molecular) - Universidade de Brasília, Brasília, 2007.

EUCLIDES FILHO, K. **Melhoramento genético animal no Brasil: fundamentos, história e importância**. Embrapa Campo Grande: Embrapa Gado de Corte, 1999. 63 p. (Documentos, 75).

FALCONER, D.S. **Introduction to Quantitative Genetics**. Longman, London. 1981.

FALCONER, D. S. **Introdução a genética quantitativa**. Viçosa: UFV, Impr. Univ., 1989, 279 p.

FALCONER, D. S.; MACKAY, T. F. C. **Introduction to quantitative genetics**. 4rd ed, London: Longman Group Ltd, 1996. 464, p.

FALEIRO, F. G.; FERNANDES, F. D.; KARIA, C. T.; BELLON, G.; RAMOS, A. K. B.; MARTHA, J. G. B.; ANDRADE, R. P.; JANK, L. Diversidade genética de uma coleção de trabalho de *Panicum Maximum* com base em marcadores moleculares. In: **REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA**, 41., 2004. Campo Grande. **Anais eletrônicos...** Campo Grande: SBZ, 2004. 1 CD-ROM.

FOOD AND AGRICULTURE ORGANIZATION (FAO). **Secondary Guidelines for development of national farm animal genetic resources management plans: Management of small populations at site.** Rome, FAO, 1998, 215 p.

FARIA, F. J. C.; VERCESI FILHO, A. E.; MADALENA, F. E. et al. Intervalo de gerações e tamanho efetivo da população da raça Gir. In: **REUNIAO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA**, 38., 2001, Piracicaba. **Anais...** Piracicaba: SBZ, 2001, p.484-485.

FARIA, F. J. C.; VERCESI FILHO, A. E.; MADALENA, F. E. Estrutura populacional da raça Nelore Mocho. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 54, n. 5, p. 501-509, 2002.

FINLAMM, E. Genetic diversity and sustainable management of animal genetic resources, globally. **AGRI**, v. 41, p. 45-52, 2007.

FITZHUGH, H. A.; STRAUSS, M. S. Management of global animal genetic resources organizational and institutional structure. In: **J. HODGES**, 1992. Rome. **Proceedings...** Rome: J. HODGES.1992, p. 309.

BARKER, J. S. F. A global protocol for determining genetic distances among domestic livestock breeds. In: **WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION**, 5., 1994, Guelph. **Proceedings...**Guelph: World Congress on Genetics Applied to Livestock Production, 1994. p. 501-508

FRANKHAM, R. Genetics and extinction. **Biological Conservation**, v. 126, n. 2, p.131- 140, 2005.

GAMA, L. T. Programas de seleção e conservação dos recursos genéticos animais: a experiência da Europa Mediterrânea. In: **REUNIAO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA**, 43., 2006, João Pessoa. **Anais eletrônicos...** João Pessoa: SBZ, 2006. 1.CD ROM.

GASTALDO, A. Z. **Identificação genética e estudos populacionais utilizando microssatélites (STR) em equinos, bovinos e caninos domésticos provenientes do Uruguai, Paraguai e Brasil.** 2017. 137 p. Tese (Doutorado em Biologia Celular e Molecular) - Pontifícia Universidade Católica, Porto Alegre, 2017.

GIANNONI, M. A.; GIANNONI, M. L. **Genética e melhoramento de rebanhos nos trópicos.** Ed. Nobel. São Paulo-SP, 1987. 473 p.

GOLDSTEIN, D. B. et al. An evaluation of genetic distances for use with microsatellite loci. **Genetics**, v. 139, n. 1, p. 463–471, 1995.

GONÇALVES, T. M.; OLIVEIRA, A. I. **Introdução ao melhoramento animal**. Lavras: UFLA, 2009, 6 p. (Textos Acadêmicos).

GOYACHE, F.; GUTIERREZ, J. P.; FERNANDEZ, I. et al. Using pedigree information to monitor genetic variability of endangered population: the Xalda sheep breed of Asturias as an example. **Journal of Animal Breeding and Genetics**, v. 120, p. 95-105, 2003.

GUIMARÃES, C. T.; MAGALHÃES, J. V.; LANZA, M. J.; SCLIUSTER, I. **Marcadores moleculares e suas aplicações no melhoramento genético**. Belo Horizonte. 2009. p. 4 - 6. (Infome agropecuário, 30).

HARDY, G. H. (1908). Mendelian proportions in a mixed population. **Science**, v. 28, p. 49–50, 1908.

HARTL, D. L.; CLARK, A. G. **Princípios de genética de populações**. 4 ed. Porto Alegre: Artmed, 2010.

HERNANDEZ, J.; BENEDITO, J. L.; CASTILLO, C. et al. Use of blood parameters as ethnic characteristics in cattle breeds. **Archiv. Fuer. Tierzucht**, v. 40, p. 521-533, 1997.

JAMES, J.W. (1977). A note on selection differentials and generation length when generations overlap. **Anim. Prod.**, v. 24, p. 109-112.

JOHANSON, I.; RENDEL, J. **Genética y mejora animal**. Zaragoza: Acribia, 1971, 567 p.

KIEFFER, N. M.; STEWART, T. S. Inbreeding and performance in two populations of Thoroughbreds. **J. Anim.Sci.**, v. 43, p. 219, 1976.

KOWNARKI, M.; JEZIERSKI, J. Inbreeding coefficients in English Thoroughbred breeding in Poland and their relationship to performance. **Anim. Breed. Abstr.**, v. 47, p. 52-89, 1979.

KWOK, P.Y.; DENG, Q.; ZAKERI, H.; TAYLOR, S. L.; NICKERSON, D.A. Increasing their formation content of STS-based genome maps: identifying polymorphisms in mapped STSs. **Genomics**, v. 31, p. 123-126, 1996.

LAAT, D. M. **Contribuição genética de fundadores e ancestrais na raça Campolina**. 2001. 41 p. Dissertação (Mestrado em Genética) - Universidade Federal de Minas Gerais, Belo Horizonte, 2001.

LAVAL, G.; SANCRISTOBAL, M.; CHEVALET, C. Measuring genetic distances between breeds: use of some distances in various short term evolution models. **Genetics Selection Evolution**, v. 34, n. 4, p. 481-507, 2002.

LEMOS, A. M. **A utilização dos polimorfismos bioquímicos e sistemas de grupos sanguíneos no melhoramento de bovinos**. Coronel Pacheco: EMBRAPA-CNPGL, 1995. 47p. (Documentos, 56).

MAcCLUER, J.; BOYCE, B.; DYKE, L.; WEITZKAMP, D.; PFENNING, A.; PARSONS, C. Inbreeding and pedigree structure in Standardbred horses. **J. Hered.**, n. 74, p. 394-399, 1983.

MARIZ, T. M. A. **Caracterização zoométrica, Estrutura Populacional e Índices reprodutivos da Raça Sindi no Brasil**. 2010. 121 p. Tese (Doutorado em Zootecnia) - Universidade Federal de Paraíba, Areia, 2010.

MEIRA, C. T. **Avaliação de características morfofuncionais de cavalos da raça Mangalarga Marchador**. 2010. 48 p. Dissertação (Mestrado em Zootecnia) - Universidade Federal dos Vales do Jequitinhonha e Mucuri, Diamantina, 2010.

MILHO, S. C. M. **Análise morfológica e genealógica dos cavalos da raça Puro Sangue Lusitano participantes em provas internacionais de ensino (“Dressage”)**. 2015. 84 p. Dissertação (Mestrado em Medicina Veterinária) - Universidade Lusófona de Humanidades e Tecnologias, Lisboa, 2015.

MOHANMMED, T. A. Phenotypic characterization of the Saudi Arabian Hassawi cattle breed. **Anim. Genet.**, v. 21, p. 35-42, 1997.

NAQVI, A. N. Application of molecular genetic technologies in livestock production: potentials for developing countries. **Advances in Biological Research**, v. 1, p. 3-4, 2007.

NAVARRO, I. C. **Estructura genética del caballo de pura raza Árabe Español y su influencia em razas derivadas: aplicacion de nuevas metodologias em el caçulo del tamano efectivo**. 2008. 181 p. Tese (Doutorado em Produção Animal) - Universidad Complutense de Madrid, Madrid, 2008.

NEI, M. Genetic distance between populations. **The American Naturalist**, v. 106, n. 949, p. 283-292, 1972.

NEI, M. Estimation of average heterozygosity and genetic distance from a small number of individuals. **Genetics**, v. 89, n. 3, p. 583-590, 1978.

NIMBKAR, C.; GIBSON, J.; OKEYO, M. et al. **Sustainable use and genetic improvement**. IN: FAO: ANIMAL GENETIC RESOURCES INFORMATION- Special issue: Scientific forum of Interlaken Conference, 42, 2008. Buletin. **Proceedings...** Scientific forum of Interlaken Conference, 2008, p. 118.

PEREIRA, J. C. C. **Melhoramento genético aplicado a produção animal**. 3 ed. Editora FEPMVZ. 2001. 555 p.

PEREIRA, G.L.; REGATIERI, I.C.; FERRAZ, G. C.; QUEIROZ NETO, A.; CURI, R. A. Perspectivas do uso de marcadores moleculares no melhoramento genético de equinos de corrida da raça Quarto de Milha. **Vet. e Zootec.**, v. 22, n. 3, p. 347-369, 2015.

PEREIRA, G. L.; CHUD, T.; BERNARDES, P.; VENTURINI, G.; CHARDULO, L. A.; CURI, R. Genotype Imputation and Accuracy Evaluation in Racing Quarter Horses Genotyped Using Different Commercial SNP Panels. **Journal of Equine Veterinary Science**, v. 58, p. 89-96, 2017.

- PINTO, C. A. B. **Genética de populações**. Lavras: UFLA, 2010. 1-67 p. (Texto acadêmico).
- QUEIROZ, S. A.; ALBUQUERQUE, L. G.; LANZONI, N. A. Efeito da endogamia sobre características de crescimento de bovinos da raça Gir no Brasil. **Revista Brasileira de Zootecnia**, v. 29, n. 4, p.1014-1019, 2000.
- RAMALHO, M. A. P.; SANTOS, J. B.; PINTO, C. A. B. P. **Genética na agropecuária**. 3, ed Lavras: Ed. UFLA, 2004, 472 p.
- REYNOLDS, J.; WEIR, B. S.; COCKERHAM, C. C. Estimation of the coancestry coefficient: basis for a short-term genetic distance. **Genetics**, v. 105, n. 3, p. 767779, 1983.
- RODRIGUES, D. S. **Estrutura populacional de um rebanho Morada Nova variedade branca no Estado do Ceará**. 2009. 46 p. Dissertação (Mestrado em Produção e Melhoramento Animal) - Centro de Ciências Agrárias da Universidade Federal do Ceará, Fortaleza, 2009.
- SANTOS, J. B. **Marcadores Moleculares**. Lavras: UFLA, 2014. 337 p. (Texto acadêmico).
- SERENO, J. R. B. Status-quo de la conservacion genetica de los caballos y bovinos Pantaneiros en Brasil. **Archivos de Zootecnia**, v. 51, p. 65-82, 2002.
- SILVA, H. D. **Aspectos biométricos na detecção de QTLs (“Quantitative Trait Loci”) em espécies cultivadas**. 2001. 177 p. Tese (Doutorado em Agronomia) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, São Paulo, 2001.
- SOORI, V. A., WATONABE, K. N., VALKONEN, J. P. T. Predicted kinase-3a motif of a resistance gene analogue as a unique marker for vírus resistance. **Theoretical and Applied Genetics**, n. 99, p. 164/170, 1999.
- SOUZA, F. A. C. **Curvas de crescimento em equinos da raça Mangalarga Marchador utilizando método transversal**. 2016. 104 p. Dissertação (Mestrado em Produção e Nutrição de Não Ruminantes) – Universidade Federal de Lavras, Lavras, 2016.
- STEELE, D. A genetic analysis of recent Thoroughbred, Standardbreds and American saddle horse. **Bull Kentucky Agric. Exp. Sta. Bull**, n. 462, p. 27, 1944.
- SUGG, D. W.; CHESSER, R. K. Effective population size with multiple paternity. **Genetics**, v. 137, p. 1147-1155, 1994.
- TAMBASCO, D. D., M. M. A.; COUTINHO, L. L.; TAMBASCO, A. J.; TAMBASCO, M. D.; REGITANO, L. C. A. Caracterização Molecular de Animais da raça Nelore utilizando Microsatélites e Genes Candidatos. **Rev. bras. zootec.**, n. 4, p. 1044-1049, 2000.
- TEEGEN, R.; EDEL, C.; THALLER, G. Population structure of the Trakehner Horse breed. **Animal**, n. 3, p. 6–15, 2009.
- TORRES, A. D. P. **Melhoramento dos rebanhos: noções fundamentais**. 3.ed. São Paulo: Ed. Nobel, 1981. 399 p.

VALE FILHO, V. R.; BERGAMANN, J. A. G.; ANDRADE, V. J. et al. Andrologic characterization of Nelore bulls, selected for the first breed season. **Rev. Bras. Reprod. Anim.**, n. 21, p. 42 - 44, 1997.

VALERA, M.; MOLINA, A.; GUTIERREZ, J. P. et al. Pedigree analysis in the Andalusian horse: population structure, genetic variability and influence of the Carthusian strain. **Livestock Production Science**, v. 95, p. 57-66, 2005.

VERCESI FILHO, A. E.; FARIA, F. J. C.; MADALENA, F. E. et al. Estrutura populacional do rebanho Tabapuã registrado no Brasil. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v. 54, p. 609-617, 2002.

WAGONER, D. M. Equine genetics & selection procedures. **Equine Research Publications**, Dallas, v. 513, p. 2-5, 1978.

WRIGHT, S. Mendelian analysis of the pure breeds of livestock. I. The measurement of inbreeding and relationship, **Journal of Heredity**. v.14, p.339-348. 1923.

WRIGHT, S. Size of population and breeding structure in relation to evolution. **Science**, v. 87, n. 2263, p. 430-2264, 1938.

WRIGHT S. **Evolution and the genetics of populations**: Variability within and among natural populations. University of Chicago Press: Chicago. USA, v. 4. 1978.

WRIGHT, S. The genetical structure of populations. **Annals of Eugenics**, v. 15, n. 4, p. 323-354, 1951.

25 **ABSTRACT**

26 Mangalarga Marchador (MM) is the most numerous horse breed in Brazil. The aims of this
27 study were to evaluate the population structure of the MM breed and identify factors that
28 affect its genetic diversity based on pedigree data. Pedigree data from 1906 to 2016 were
29 obtained from the Brazilian Mangalarga Marchador Horse Breeders' Association. Analyses
30 were performed on the programs SAS, Pedig, and PopReport, from which characterization of
31 the population, the quality of the pedigree information, and the relevant parameters of
32 inbreeding and of the probability of gene origin were obtained, considering the total
33 population (TP, n = 514,282) and a reference population (RP, n = 172,508) of animals born
34 from 2005 to 2016. The stallions and mares represented 9.72% and 32.71% of the TP,
35 producing an average of 22.30 and 4.62 foals per reproducer, respectively. The mares
36 produced their first foals at an average age of 11.20 years, whereas the average reproduction
37 age for the stallions was 7.86 years. The generation interval was 10.43 in the TP and 9.48 in
38 the RP. The equivalent number of known generations was 2.38 in the TP and 3.95 generations
39 in the RP. The combined average of pedigree completeness of the five most recent
40 generations was 44.46% for the TP and 84.63% for the RP. The inbreeding coefficient (F) and
41 the rate of F per year were 1.02% and 0.11% in the TP and 1.10% and -0.14% in the RP. The
42 effective population size was 1,471 for the TP and -959 for the RP. The mean of the additive
43 genetic relationship and the rate this relationship per year for animals of the TP were 0.0023
44 and 0.0012, and, for animals of the RP, they were 0.0045 and 0.0017, respectively. Family
45 size for stallions was 27.19, and 4.47 for mares. The effective number of ancestors, founders,
46 and founder genomes were 1,038, 1,062, and 150 in the TP, and 198, 225, and 133 in the RP,
47 respectively. The number of segregating genes was 5,122 for the total population. The
48 numbers of founders that exhibit 50% and 10% of the current gene pool of the breed
49 represented 7.94% and 6.64% of the total founders of the TP, and 19.09% and 0.69% of the
50 RP. The Mangalarga Marchador population does not exhibit severe inbreeding that would
51 harm its evolution as a species. Nevertheless, a possible loss of genetic diversity is notable,
52 due to genetic bottlenecks.

53

54 **Keywords:** equine, inbreeding, demographic parameters, genetic variability.

55 **Funding:** This work was supported by the Conselho Nacional de Desenvolvimento Científico
56 e Tecnológico (CNPq).

57

58 **1. Introduction**

59

60 The Mangalarga Marchador (MM) breed is the largest and most representative native
61 Brazilian equine breeding stock (Lima and Cintra, 2016). In 2017, there were more than 600
62 thousand animals registered, 15 thousand association members, 70 breeding groups in Brazil,
63 and four groups abroad, in Argentina, the United States, Germany, and Italy (Fonseca et al.,
64 2017). MM animals have a dissociated gait and are docile and resistant. They are mainly used
65 in daily tasks on rural properties, including herding of cattle, but with growing participation in
66 sporting events and, more recently, in horse riding as a leisure activity and in ecological
67 tourism (Gonçalves et al., 2011; Souza et al., 2017).

68 Native Brazilian horse breeds arose from the actual needs of the regions of the country or
69 through the preferences of breeders. Thus, at the end of the nineteenth century, the
70 Mangalarga Marchador horse was formed in the southern part of the state of Minas Gerais
71 (Assis et al., 2009). The breed was the result of the cross between Lusitano stallions
72 originating from the Coudelaria de Alter do Chão, introduced in the country by Portuguese
73 colonizers, with native mares, probably descendants of the first horses, also brought from the
74 Iberian Peninsula in 1500 (Meira, 2010). According to Nascimento (1999) and Assis et al.
75 (2009), there was a mixture of diverse breeds, such as Arabian, Andalusian, Criollo, and
76 Quarter Horse, besides the Lusitano, that made genetic contribution to the formation of MM.
77 This fact suggests the hypothesis of wide genetic variability in the current population.

78 After consolidation of MM formation, the first selective breeding efforts began, and
79 unlike in other domestic breeds of horses, selection does not have such well-defined economic
80 objectives; however, strength and gait appear to have always been the main objectives in
81 selection in this breed. Since 1870 in Brazil, the MM has been selected for cavalcades and

82 movement at low speed, but with a dissociated gait, called “*marchado*” gait (Santiago et al.,
83 2016). However, the criteria generally used for selection have been based on pedigree
84 information, and especially on characteristics subjectively evaluated from their morphology or
85 even in an empirical way by the horse breeders themselves, based on choosing reproducers
86 that have won awards in gait competitions (Costa et al., 2005). This selection, practiced over
87 approximately 150 years, as well as artificial reproduction techniques such as insemination
88 and embryo transfer, also practiced for years, and local management that creates genetic
89 subdivisions within the breed, raise the hypothesis that some degree of reproductive isolation
90 may exist and that genetic variability may have been affected (Assis et al., 2009).

91 Since 1906, pedigree records have been maintained by private breeders and initially
92 communicated to the Associação Brasileira de Criadores de Cavalos da Raça Mangalarga
93 (ABCRM), which was first founded in 1934. Only in 1949 was the Brazilian Mangalarga
94 Marchador Horse Breeders’ Association (Associação Brasileira dos Criadores do Cavalos
95 Mangalarga Marchador - ABCCMM) founded, and therefore, the official registry of the breed
96 began. In 1965, the year in which the ABCCMM began administering the Genealogical
97 Registry (Studbook) Service of the breed, parentage testing of the animal was required before
98 the animal could be officially registered. At first, confirmation of kinship was based on blood
99 typing or on protein polymorphism, and beginning in 2005, through genotyping of
100 microsatellite molecular markers (Oliveira and Kuabara, 2001).

101 In the beginning, animals without known genealogy were registered in what was called
102 an “Open Book”, as long as they corresponded to breed standards, as decided by specialized
103 technicians. However, the open studbook for males was closed in 1966, and for females in
104 1984 (Costa et al., 2004). In 2006, after denunciations of fraud in the studbook, the Conduct
105 Adjustment Declaration (Termo de Ajustamento de Conduta - TAC) was established by
106 competent inspection authorities, in which suspect animals or those without confirmation of

107 their genealogy came to be registered in the “Book for Inclusion in the TAC” (“Livro para
108 Inclusão no TAC”). Only in 2015, after establishment of the studbook system, was the open
109 record of males and females reopened.

110 The evolution of genetic variability of a population has always been studied from
111 verification of the level of inbreeding, of genetic relationships among individuals, of
112 probability of gene origin of herds, of the number of founders and ancestors, and changes over
113 time (Falconer and Mackay, 1996; Boichard et al., 1997; Hill, 2000; Vicente et al., 2012). For
114 years and even now, these principles have been applied in studies of genetic structure of
115 different horse breeds from pedigree information, as for example in horse breeds such as
116 Quarter Horse (Tunnel et al., 1983), Thoroughbred (Cunningham et al., 2001), Lipizzan
117 (Zechner et al., 2002), Purebred Spanish (Valera et al., 2005), Mangalarga Marchador (Costa
118 et al., 2005), Mangalarga (Mota et al., 2006), Franches-Montagnes (Poncet et al., 2006),
119 Asturcón (Royo et al., 2007), Hanoverian (Hamann and Distl, 2008), Spanish Arabian
120 (Cervantes et al., 2008), Arabian Trakehner (Cervantes et al., 2009; Teegen et al., 2009);
121 Lusitano (Vicente et al., 2012), Slovakian Warmblood (Pjontek et al., 2012), Spanish Sport
122 Horse (Bartolomé et al., 2011), Maremmano (Giontella et al., 2018), Quarter Horse (Faria et
123 al., 2018), Campolina (Bussiman et al., 2018), Hucul Pony (Somogyvári et al., 2018),
124 Brazilian Criollo (Maciel et al., 2014), and Japanese Native horses (Onogi et al., 2018).

125 The aims of this study were to analyze the initial and current genetic structure of the
126 population of Mangalarga Marchador horses of Brazil based on pedigree analysis provided by
127 the ABCCMM, and identify relevant factors that may affect the genetic variability of the
128 breed.

129

130 2. Materials and Methods

131

132 2.1 Animals

133 The pedigree information used for this study was obtained through the Associação
134 Brasileira de Criadores do Cavalo Mangalarga Marchador (ABCCMM).

135 Of the total animals of the population ($n = 514,283$), 85,958 were registered mares in
136 reproduction and 21,911 were stallions. Artificial insemination in horses is a reproductive
137 technique that has been applied since 1912 and has come to be a routine practice in various
138 countries; however, the ABCCMM first allowed its use in the breed only in 1994 (Costa et al.,
139 2004). Since then, it has been used progressively, as well as embryo transfer, which was
140 allowed in 1995 (Caiado et al., 2007).

141

142 2.2 Database

143 The database was adjusted for greater consistency, in which duplicates and other factors
144 were excluded and individual identification was renumbered for later pedigree analysis. After
145 this editing, the file included data referring to 514,283 male, female, and castrated animals
146 born and registered from 1906 to 2016 coming from 14421 horse breeding farms throughout
147 all Brazilian states. Minas Gerais was most represented (41.72%), followed by Rio de Janeiro
148 (14.26%), Bahia (12.42%), São Paulo (10.76%), Espírito Santo (4.71%), Goiás (3.45%), and
149 Pernambuco (3.28%).

150 Some demographic parameters for the study of population structure were calculated
151 considering not only total population (TP) but also a reference population (RP) that
152 encompasses the last generation and that contains 172,508 registered horses (33.54% of the
153 TP) from 2005 to 2016 – 23833 mares and 9384 stallions in reproduction. The TP was also
154 divided into ten generations that covers the years 1906-1916 (corresponding to generation 1 -

155 Gen 1), 1917-1928 (Gen 2), 1928-1938 (Gen 3), 1939-1949 (Gen 4), 1950-1960 (Gen 5),
156 1961-1971 (Gen 6), 1972-1982 (Gen 7), 1983-1993 (Gen 8), 1994-2004 (Gen 9), and 2005-
157 2016 (Gen 10).

158

159 *2.3 Statistical analyses*

160

161 Statistical analyses were performed using procedures of the SAS program (SAS, 2009)
162 and R program (R Core Team, 2016). The demographic parameters were calculated by the
163 Pedig (Boichard, 2002) and PopReport (Groeneveld et al., 2009) programs, specifically
164 developed for this purpose.

165

166 *2.3.1 Characterization of the population*

167 Number of births and reproductive parameters

168 The number of registered births of males and females over the years 1906-2016 and
169 reproductive parameters such as means, standard deviation, minimum and maximum number
170 of foals produced per stallion and per mare, as well as the mean age of the reproducers when
171 in reproduction, were calculated for the entire population (TP), including the reference
172 population (RP).

173

174 Generation interval

175 The generation intervals (GI) for the TP and RP were calculated for the reproducers
176 (stallions and/or mares) with all the progenies, and also for the four selection methods: the
177 mean ages of stallions at birth of colts, stallions at birth of fillies, mares at birth of colts, and
178 mares at birth of fillies. GI is the mean age of the parents at the birth of their foals, with
179 registration of progeny (Falconer and Mackay, 1996).

180

181 *2.3.2 Quality of pedigree information*

182 Known generations

183 The maximum number of generations traced, which provides the number of generations
 184 that separate the individual from its most distant ancestor, regardless of whether the parents
 185 are known or not; the mean of the number of complete or known generations (G_t), which
 186 corresponds to the number of generations with both parents known; and the equivalent
 187 number of complete or known generations (G) by year of birth and by sex were calculated as
 188 described by Boichard et al. (1997). This last parameter best represents the pedigree
 189 information, according to Maignel et al. (1996).

190

$$191 \quad G = \frac{1}{N} \sum_{j=1}^N \sum_{i=1}^{n_j} \frac{1}{2g_{ij}}$$

192

193 where n_j is the number of ancestors of the animal j in the population and g_{ij} is the number
 194 of generations between j and its ancestor i .

195

196 Pedigree completeness

197 MacCluer et al. (1983) proposed some indices for measuring pedigree completeness.
 198 These indices summarize the proportion of ancestors known in each ascending generation,
 199 and are calculated as follows:

200

$$201 \quad I_d = \frac{4 I_{d_{pat}} I_{d_{mat}}}{I_{d_{pat}} I_{d_{mat}}} \quad \text{and} \quad I_{d_k} = \frac{1}{d} \sum_{i=1}^d a_i$$

202

203 where k represents the maternal (*mat*) line and paternal (*pat*) line of an individual, a_i is the
 204 proportion of ancestors known in the generation, and d is the number of generations
 205 considered in calculation of the pedigree completeness index.

206

207 2.3.3 Inbreeding Parameters

208 Inbreeding coefficient

209 The individual inbreeding coefficient (F) is defined as the probability of an individual
 210 having the two alleles identical by descent, according to Wright (1931). Thus, the individual
 211 F , paternal and maternal F , number of inbred animals, and the mean and maximum of F were
 212 calculated for all the individuals and for the inbred animals according to the method of
 213 Meuwissen and Luo (1992), using the Cholesky factor of the relationship matrix, a
 214 modification of the Quass (1976) method:

215

$$216 \quad F = A_{ii} = \sum_{j=1}^i L_{ij}^2 D_{jj}$$

217

218 where A_{ii} is the i -eth diagonal element of A , which is equal to the inbreeding coefficient
 219 of animal $i + 1$, L is a triangular matrix containing the fraction of the genes of animals that
 220 were derived from their ancestors, and D is the diagonal matrix with the additive genetic
 221 variances within the family.

222

223 Rate of inbreeding

224 The rate of inbreeding (ΔF) indicates the increase of inbreeding in a year or generation, as
 225 proposed by VanRaden (1992), derived from the tabular method from construction of the

226 relationship matrix of each individual and of its ancestors. The ΔF was obtained by the
 227 equation of Falconer and Mackay (1996):

228

$$229 \quad \Delta F = \frac{(F_t - F_{t-1})}{(1 - F_{t-1})}$$

230

231 where F_t is the mean inbreeding coefficient of the generation or year t , estimated in the
 232 second period; and F_{t-1} is the mean inbreeding coefficient, estimated in the first period.

233

234 Effective population size

235 The effective population size (N_e) in the year and through ΔF was determined by the
 236 following equation, defined by Gutiérrez and Goyache (2005):

237

$$238 \quad N_e = \frac{1}{2\Delta F}$$

239

240 According to Falconer and Mackay (1996), N_e represents the size of the genetically active
 241 population, i.e., that could generate the same rate of inbreeding or variance in the allele
 242 frequencies as that observed in the population under consideration if these animals had been
 243 raised under the conditions of an idealized population.

244 The following formula was also used to calculate the N_e in the year, but based on the
 245 number of parents (Falconer and Mackay, 1996):

246

$$247 \quad N_e = \frac{4 N_m N_f}{N_m + N_f}$$

248

249 where N_m and N_f are the number of male and female parents.

250 Rate and mean of the additive genetic relationship

251 The inbreeding coefficient (F) of an individual is the half to the additive genetic
252 relationship (AGR) between its parents or equal to the co-ancestry coefficient ($F_i = f_{sd}$), where
253 i is the individual, and s and d are the mother and the father, respectively (Falconer and
254 Mackay, 1996).

255 Consequently, the mean of the AGR indicates the contribution of the founders to the
256 population and can be used as an index to maintain the initial genetic stock, as well as
257 compare the inbreeding among subpopulations.

258 The additive genetic relationship between pairs of individuals was computed based on the
259 relationship matrix of all the animals of the population. The AGR between animals born in a
260 certain year was calculated as the mean value of inbreeding of the progenies of all the
261 possible matings between individuals. The rate of the additive genetic relationship (Δ AGR)
262 per year was also computed following the same principle as the ΔF cited above.

263

264 Family size

265 Family size was obtained from the number of descendants of individuals that became
266 reproducers in the following generation, according to Falconer and Mackay (1996).

267

268 2.3.4 Probability of gene origin

269 Founders

270 According to Gutiérrez and Goyache (2005), founders are ancestors without their parents
271 being known in the pedigree. The effective number of founders (f_e) is the number of equally
272 contributing founders that are expected to produce the same genetic diversity as the
273 population studied. Thus, the genetic contribution, the number of founders (N_f), and the f_e

274 were calculated, with f_e being obtained from the contributions of all the individuals of the
 275 population (Rochambeau et al., 1989):

276

$$277 \quad f_e = 1 / \sum_{k=1}^f q_k^2$$

278

279 where q_k represents the probability of gene origin and k the ancestor.

280

281 Ancestors

282 Boichard et al. (1997) proposed to identify the main ancestors and to determine their
 283 marginal contributions to the gene pool in a population. The effective number of ancestors
 284 (f_a) is the minimum number of ancestors, not necessarily founders, that explain the entire
 285 genetic diversity of a population, complementing the f_e , and indicating the occurrence of
 286 genetic bottlenecks produced through unbalanced use of some reproducers (Boichard et al.,
 287 1997).

288 The formula below can be applied to these marginal contributions (p_j) to determine the
 289 effective number of ancestors:

290

$$291 \quad f_a = 1 / \sum_{j=1}^f p_j^2$$

292

293 where p_j is the marginal genetic contribution of the ancestor in the population.

294

295 The genetic contributions and the mean numbers of ancestors (N_a) in the year and by sex
 296 were also computed, as well as the f_e/f_a and f_e/N_e ratios, which have implications for the
 297 population genetics bottleneck effect (Somoogyvári et al., 2018).

298

299 Founder genomes

300 The effective number of founder genomes (f_g) estimates the probability of a given gene
 301 present in the founders, i.e., a “founder gene”, still being present in the population under
 302 study, or even defined as the number of founders that would be expected to produce the same
 303 genetic diversity in the population studied if the founders were equally represented and if no
 304 allele were lost (Caballero and Toro, 2000). This parameter was computed according to Lacy
 305 (1989):

306

$$307 \quad f_g = \frac{NA}{2} = 1/2 \sum_{k=1}^{2f} f_k^2, \text{ with } NA = 1/\sum_{k=1}^{2f} f_k^2$$

308

309 where f_k is the gene frequency and NA is the effective number of alleles in the population
 310 under study.

311 The f_g/f_e ratio is also important and indicates whether the population was affected or not
 312 by genetic drift (Boichard et al., 1997). The comparison of f_g and N_f (f_g/N_f) shows the
 313 percentage of founder genomes that were retained in the population.

314 The number of genes of the founders that are in segregation in the population and the
 315 percentage of them that are explained by n ancestors (founders or otherwise) were also
 316 computed.

317

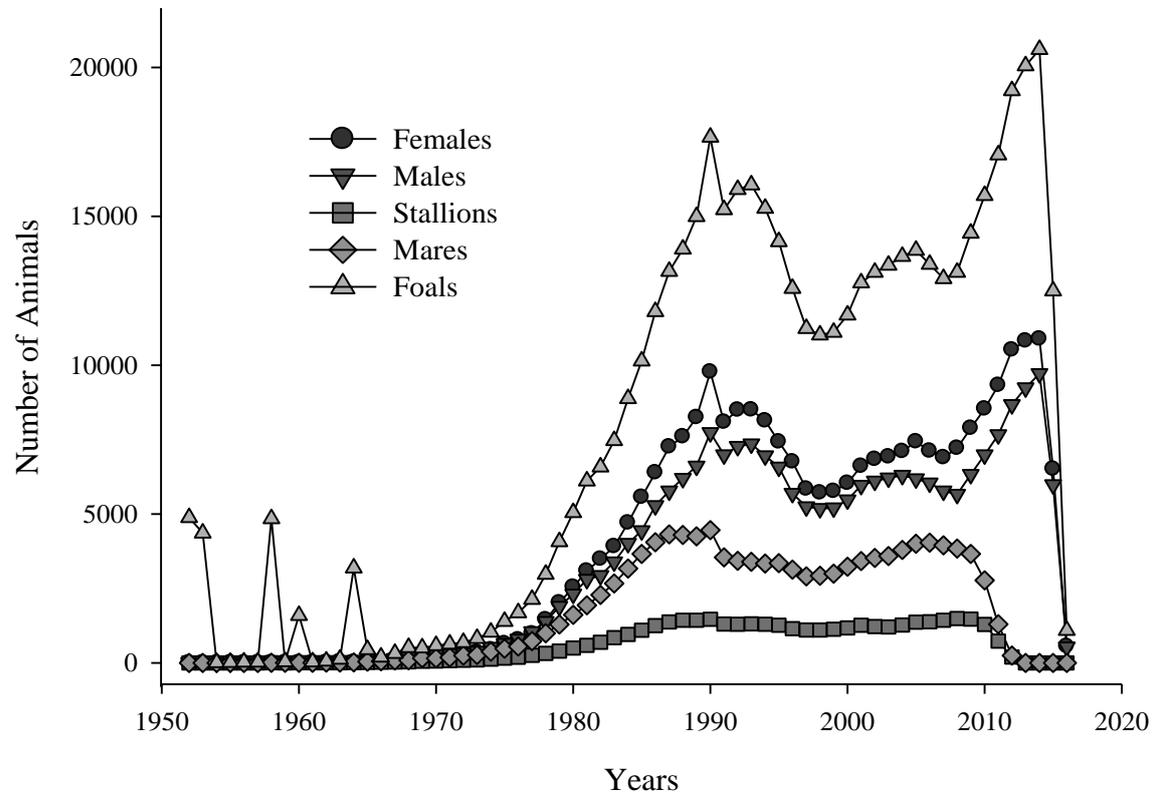
318 **3. Results**

319

320 *3.1 Characterization of the population*

321 Number of births

322 The distribution of the official records of births encompasses the years 1906 to the
323 beginning of 2016. Although the ABCCMM was founded in 1949, it considered some already
324 existing records. Before 1952, fewer than ten animals of each sex were registered per year,
325 with a significant increase in 1977, when the contingent of 1,000 registered animals was
326 achieved (980 females and 1,053 males) (Fig. 1). After that year, the number of registered
327 males rose an average of 18% a year until 1990, and the average increase for females was
328 greater, at 20% up to 1993. This was followed by a decline from 1991 to 1998 for males
329 (mean of -4.72% a year) and for females (-6.23% a year). From 1999 to 2005, registration
330 once more increased (mean of 4% a year), and then once more declined (mean of -2.86% for
331 males and -3.66% for females) from 2006 to 2008. Successive increases followed, up to 2014,
332 the year with the highest number of records of the breed, with 10,884 females and 9,720
333 males. In contrast, in 2015, registrations decreased sharply, with a decline of 40.17% for
334 females (6,511 mares active or not) and 38.40% for males (5,987 stallions and geldings).



335
336
337
338

Fig. 1. Number of females, males, reproducers (mares and stallions), and foals born and registered each year for the Mangalarga Marchador horse population.

339

340 Reproductive parameters

341 The stallions and mares represented 9.72% and 32.71% of the total population of males
342 and females, respectively. The utilization of stallions grew from 1952 on, with a significant
343 increase in 1985, which remained practically constant until 2010, but with an abrupt fall from
344 2011 on. The utilization of mares also increased from 1952 on, with a relevant increase in
345 1979, but with increases and decreases to 2011, the year of a significant decline in their
346 utilization (Fig. 1). Annually, the mean numbers of stallions and mares utilized are
347 approximately 649 and 1,823, respectively, with total numbers of 21,911 stallions and 85,958
348 mares in reproduction in the TP. These reproducers generated an average of 15,582 foals a
349 year, with average production of 4.69 foals per mare and 22.29 foals per stallion. The stallion

350 and the mare were the reproducers with the largest number of foals (1,536 and 73,
 351 respectively). An additional result, the proportion of mares per stallion in the Mangalarga
 352 Marchador population, was 3.9 (Table 1).

353

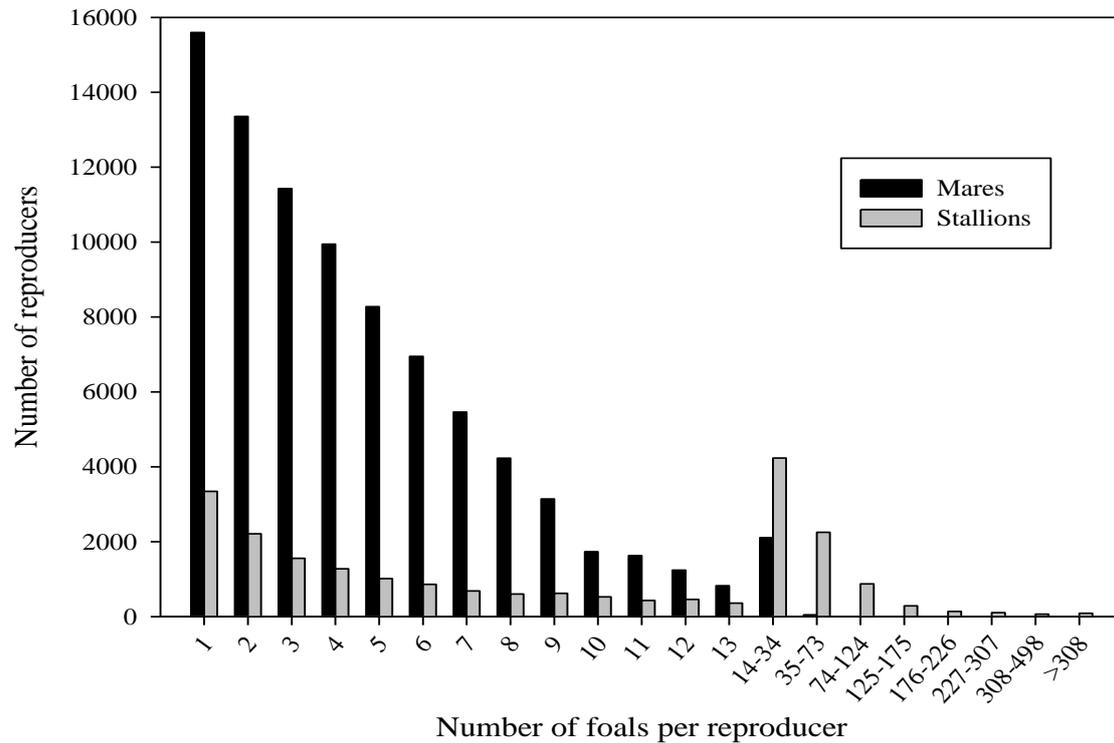
354 **Table 1.** Descriptive statistics of reproductive parameters of the males and females of the
 355 Mangalarga Marchador horse breed population.

Item	Total population	
	Males	Females
Total animals (n)	225,464	262,803
Total reproducers (n)	21,911	85,958
Total foals (n)	488,416	399,113
No. of foals per reproducer		
Mean and standard deviation (n)	22.29 ± 47.13	4.69 ± 3.69
Minimum (n)	1	1
Maximum (n)	1,536	73
Mare/stallion ratio	3.9	

356 (n) number of observations.

357

358 In general, most reproducers, over their reproductive life, produced from 1 to 13 foals,
 359 and above this number of progenies, stallions always had higher production than mares (Fig.
 360 2). However, the number of stallions with production greater than 34 foals decreased,
 361 especially stallions with 176 foals or more.

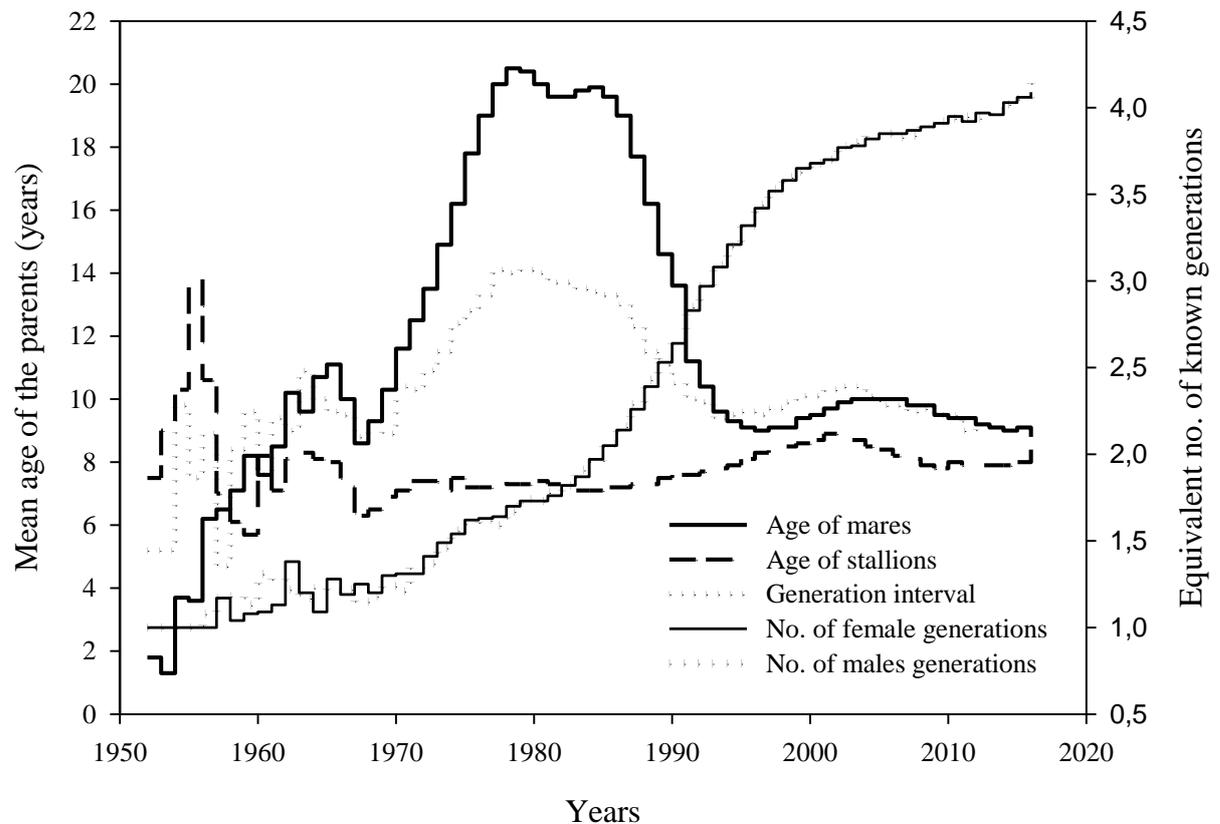


362
363
364
365

Fig. 2. Number of reproducers (mares and stallions) per class of number of foals per reproducer of the Mangalarga Marchador horse breed population.

366
367
368
369
370
371
372
373
374
375
376

The age distribution of the reproducers (Fig. 3) indicates that mares had their first foals on average at 11.2 ± 4.78 years, whereas the stallions at 7.86 ± 1.10 years. For stallions, a decline in their utilization above nine years of age is evident, and over the years of this study, the age of the stallion in reproduction remained more constant. Only from 1954 to 1956 were stallions from 10 to 13 years of age utilized for reproduction. Nevertheless, for mares, age did not remain constant throughout the same period analyzed (data not tabulated), but data show that initially younger mares of even four years of age were utilized, and from 1974 to 1988, the mares were from 14 to 20 years of age. Also, as a result, total foal production from reproducers up to 15 years of age was 92.65% for stallions and 76% for mares, and 7.35% of the foals are descendants of stallions above 16 years of age and 23.34% of the foals were also born from mares above that age.



377
378
379
380
381

Fig 3. Mean values of ages of mares and stallions when in reproduction, of the total generation interval, and of the equivalent number of known generations by sex (male and female) each year of the Mangalarga Marchador horse breed population.

382 Generation Interval

383 At the beginning of the 1970s to the end of the 1980s, greater generation intervals (GI) in
384 the population were observed, and generation interval is more constant from 1992 on (Fig. 3).
385 The GI based on the ages of the reproducers (mares and stallions) with all the progenies in the
386 total Mangalarga Marchador population was 10.43 years; the GI for the stallions was lower
387 (8.84 years). The same is observed for the reference population (Table 2). In general, upon
388 considering the different selection routes for the two populations analyzed (TP and RP), the
389 mean generation interval in the mare-son path is greater than in the stallion-son path, and the
390 mare-daughter path is also greater than the stallion-daughter path. In the TP, the stallion or
391 mare with son path is lower than the stallion or mare with daughter path, once more showing

392 that males enter in reproduction earlier than females. However, the RP shows the contrary; the
 393 stallion-son path is greater than that of mare-daughter, and the mare-son and mare-daughter
 394 paths are similar.

395

396 **Table 2.** Statistics of the generation intervals of the paths between parents and progeny and of
 397 the numbers of known generations by sex of the Mangalarga Marchador horse breed
 398 population.

Parameter	Total Population		Reference Population	
	N	Mean \pm SD	N	Mean \pm SD
Stallion-Son	20,398	8.74 \pm 2.11	5,226	9.24 \pm 0.15
Stallion-Daughter	96,954	9.09 \pm 2.68	20,813	8.99 \pm 0.14
Mare-Son	20,436	11.42 \pm 3.90	5,243	9.96 \pm 0.56
Mare-Daughter	97,011	12.15 \pm 4.25	20,859	9.92 \pm 0.61
Stallion-Progeny	117,352	8.84 \pm 1.58	26,039	9.05 \pm 0.11
Mare-Progeny	117,447	12.02 \pm 4.14	26,102	9.92 \pm 0.60
Reproducers-progeny	234,799	10.43 \pm 2.25	52,141	9.48 \pm 0.62
Mean of the equivalent no. of Known Generations (G)				
Total equivalent	-	2.38 \pm 1.14	-	3.95 \pm 0.09
	Male	Female	Male	Female
Equivalent by sex	(mean \pm SD)	(mean \pm SD)	(mean \pm SD)	(mean \pm SD)
	2.39 \pm 1.15	2.38 \pm 1.15	3.95 \pm 0.11	3.94 \pm 0.08
No. of Known Generations (G_t)				
Total	14		1	

399 (N) number of observations; SD = standard deviation.

400

401 3.2 Quality of the pedigree information

402 Known generations

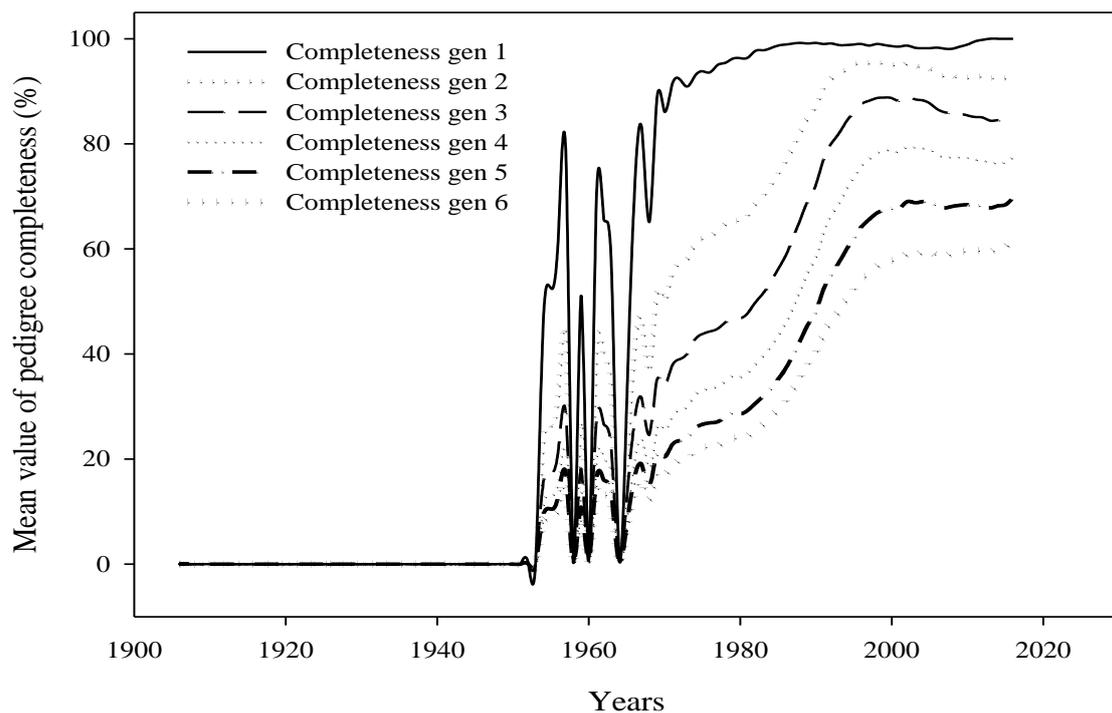
403 The equivalent number of known generations (G) has increased over the years, until
 404 reaching the mean per year of 2.38 generations in the TP and 3.95 in the RP. There was no
 405 difference in the mean of G between males and females in the two populations studied (Table
 406 2). Moreover, an increase in G was observed over the years for both males and females (Fig.
 407 3). The total number of generations traced was 14 for the entire Mangalarga Marchador
 408 population.

409

410 Pedigree completeness

411 The pedigree records date from January 1906 to April 2016, and the mean percentage of
 412 parents (generation 1), grandparents (generation 2), great grandparents (generation 3), great
 413 great grandparents (generation 4), third great grandparents (generation 5), and fourth great
 414 grandparents (generation 6) known in the complete dataset were 63.55%, 50.85%, 42.26%,
 415 35.58%, 30.05%, and 25.6%, respectively (Fig. 4).

416 For the reference population, the mean percentages known were 99.12%, 92.92%,
 417 85.49%, 77.29%, 68.35%, and 59.69%, and only from 2013 on were 100% of the parent
 418 known for all the animals registered. The combined mean value of pedigree completeness of
 419 the five most recent generations was 44.46% for TP and 84.63% for RP.



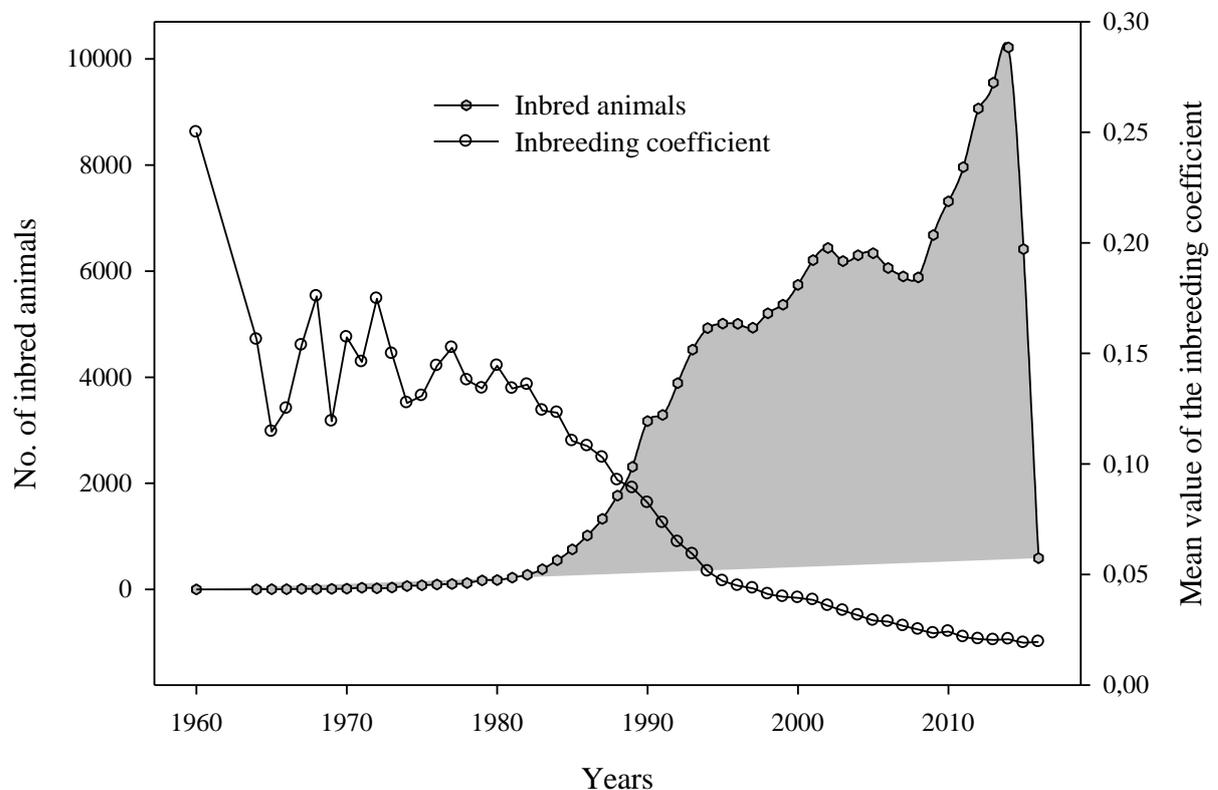
420
 421 **Fig. 4.** Mean value of pedigree completeness from generation 1 to 6 (gen 1 to 6) of the
 422 Mangalarga Marchador horse breed population.
 423

424
 425 *3.3 Inbreeding Parameters*

426 *Inbreeding Coefficient*

427 The mean value of the inbreeding coefficient (F) was $1.02\% \pm 0.03$ for all the animals
 428 registered and considering the entire population, and $1.10\% \pm 0.28$ for the reference
 429 population. However, 32.61% (167,734 animals) of the TP and 47.52% (81,971 animals) of
 430 the RP had an F other than zero, with maximum F of 46.87% for the TP and 37.50% for the
 431 RP. The paternal F (1.18%) was superior to the maternal F (0.80%) for the TP, and paternal F
 432 (1.36%) was superior to the maternal F (1.33%) for the RP.

433 The first inbred animals began to arise in 1960, with a high F of up to 25% (Fig. 5). Over
 434 the years, the number of inbred individuals always grew, with a maximum of 10,215 inbred
 435 animals in 2014, and later reduction of 63% from that year on. Nevertheless, although the
 436 number of inbred animals increased every year, their F level had successive increases and
 437 decreases over the years, and then decreased constantly from 1982 on, with mean values of
 438 $8.70\% \pm 5.79\%$ for the TP and of $2.33\% \pm 0.35\%$ for the RP.



439 **Fig. 5.** Number and mean value each year of the inbreeding coefficient of inbred animals of
 440 the Mangalarga Marchador horse breed population.
 441

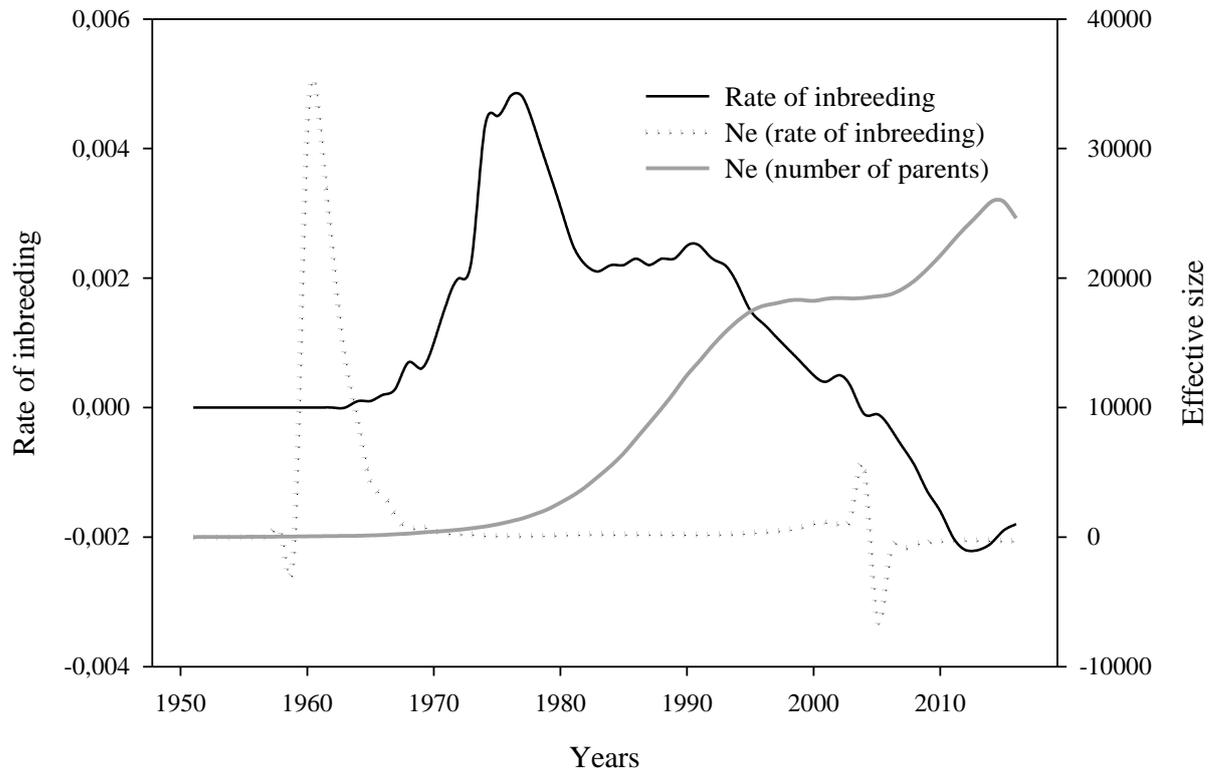
442 Rate of inbreeding

443 The mean value of the $\Delta F/\text{year}$ was $0.11 \pm 0.19\%$ considering the entire population, and -
444 $0.14 \pm 0.08\%$ for the reference population. The first negative values of $\Delta F/\text{year}$ occurred from
445 2004 to 2005 and continued negative up to the last year of the study. The mean value of
446 $\Delta F/\text{generation}$ (assuming a generation interval of 10 years) was $0.12 \pm 0.18\%$, and was
447 negative only in the last complete generation of the population (2005 to 2016). Furthermore,
448 the highest values of $\Delta F/\text{year}$ were found in the years from 1974 to 1980 (Fig. 6), with values
449 from 0.37% to 0.43%, and the highest $\Delta F/\text{generation}$ occurred in the seventh generation,
450 which includes these same years.

451

452 Effective population size

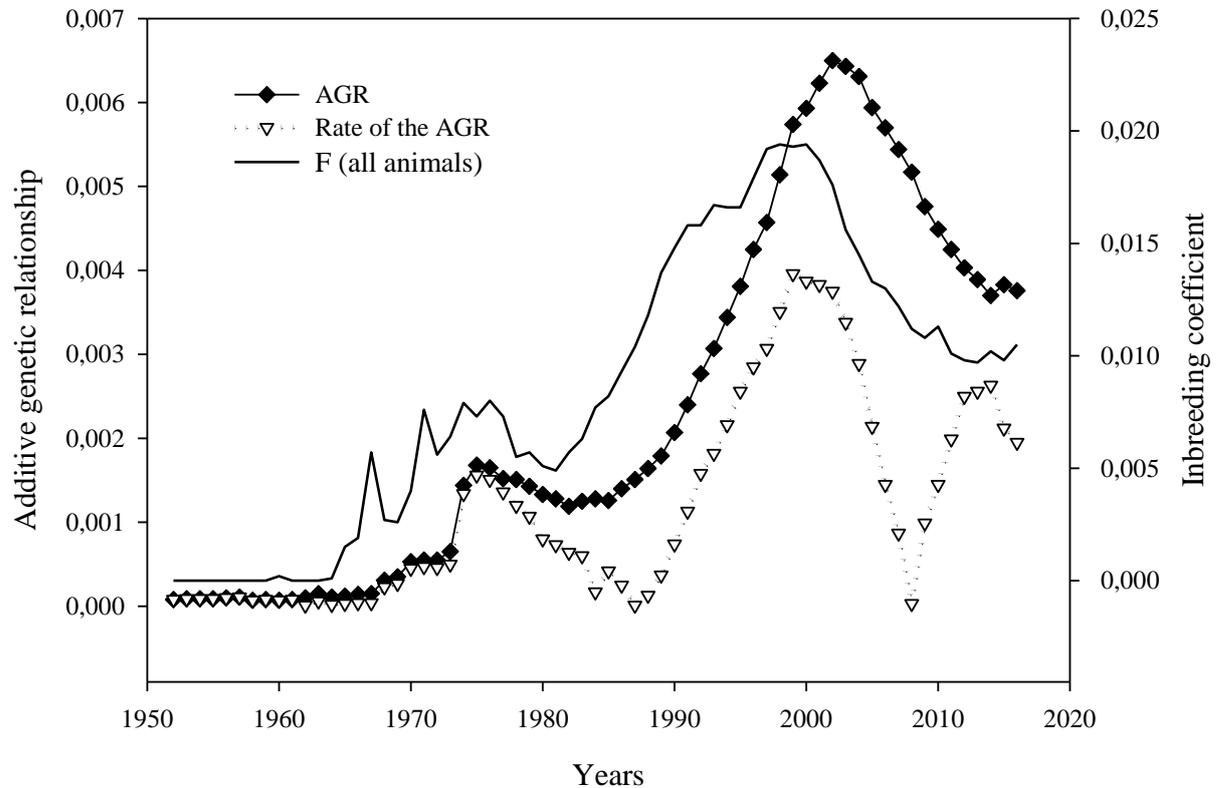
453 The mean of the effective population size (N_e) in the year based on the rate of inbreeding
454 was $1,471 \pm 5,663.30$ for the TP and $-959 \pm 1,772.20$ for the RP. Based on the number of
455 parents, the N_e was $7,039 \pm 8,933.28$ for the TP and $22,232 \pm 2,802.77$ for the RP. The
456 behavior of N_e over the years is represented in Figure 6; $N_e (\Delta F)$ is most used in the literature.
457 The highest value of $N_e (\Delta F)$ was in 1961, with 31,744 animals, with an abrupt decline in the
458 following years and then a constant level at the beginning of the 1970s to the end of the
459 1990s. In 2004, there was a new significant increase from 1,000 to 5,293 animals, and,
460 following that, a new decline.



461
 462 **Fig. 6.** Rate of inbreeding (ΔF), effective sizes in the year based on the ΔF and on the
 463 number of parents of the Mangalarga Marchador horse breed population.
 464

465 Rate and mean of the additive genetic relationship

466 Of the total population ($n = 514,283$) of Mangalarga Marchador, 371,939 animals do not
 467 have a kinship relationship, representing 72.32% of the total population. The mean additive
 468 genetic relationship (AGR) in the year for animals of the TP was 0.0023 ± 0.0021 , whereas
 469 the AGR among the animals of the RP was 0.0045 ± 0.0008 . The rate of the additive genetic
 470 relationship (ΔAGR) in the year had a mean value of 0.0012 ± 0.0012 for the TP, and the
 471 ΔAGR for the RP was 0.0017 ± 0.0008 . Initially, the mean values of the AGR and of the
 472 ΔAGR had low values, but with significant increases in 1974, and later declines. From 1984
 473 on, new increases were observed, which continued until 2002. In the following years, the
 474 AGR and ΔAGR once more came to decline (Fig. 7).



475
 476 **Fig. 7.** Mean value of the additive genetic relationship (AGR), rate of the AGR, and
 477 inbreeding coefficient each year of all the individuals of the Mangalarga Marchador horse
 478 breed population.
 479

480 Family size

481 Variation in family size exhibited a tendency to increase from the year 1948 to the end of
 482 the 1980s, and then changed to a decrease after that date (data not tabulated). The family size
 483 for stallions (27.19 ± 16.23) was much greater than for mares (4.47 ± 1.74).

484

485 3.4 Probability of gene origin

486 Founders, ancestors, and founder genomes

487 The 514,283 Mangalarga Marchador horses registered up to April 2016 result from
 488 genetic contributions from 25,607 founders (1,497 males and 24,110 females), with an annual
 489 mean of 2,930 ancestors registered in the TP. The effective number of founders (f_e)
 490 represented 4.15% and 30.90% of the founders of the TP and RP. The effective number of

491 ancestors (f_a) was less than the f_e for the two populations analyzed. A comparison of N_f and
 492 f_g in the f_g/N_f ratio showed that 18% of the founder genomes were retained in the most recent
 493 population (Table 3). The ratios f_e/f_a , f_e/f_g , f_e/N_e , and f_g/N_f were also presented in Table 3.

494 The number of genes coming from the founders that are in segregation in the population
 495 was 51,622, and 42.49% of these genes are explained by 100 main ancestors and half of those
 496 genes are explained by 19.27% of the effective ancestors alone (Table 3).

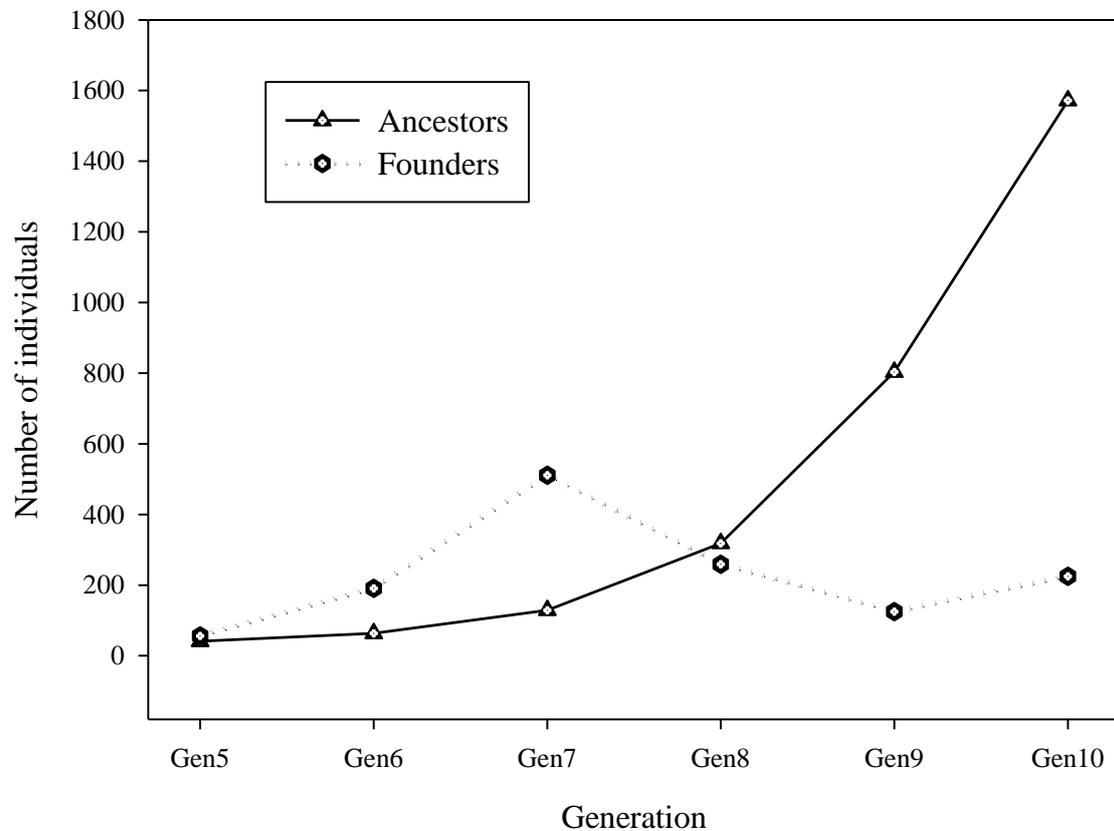
497

498 **Table 3.** Parameters of probability of gene origin, genetic drift, number and percentage of
 499 segregating genes of the Mangalarga Marchador horse breed population.

Parameters	Total Population			Reference Population		
	Male	Female	Total	Male	Female	Total
Probability of gene origin						
Mean no. of ancestors (N_a)	1470	1460	2930	788	784	1572
No. of founders (N_f)	1497	24110	25607	224	504	728
Effective no. of ancestors (f_a)		1038			198	
Effective no. of founders (f_e)		1062			225	
Effective no. of founder genomes (f_g)		150			133	
f_e/f_a		1.02			1.14	
f_g/f_e		0.14			0.60	
f_e/N_e		0.72			0.23	
f_g/N_f		0.006			0.18	
No. of segregating genes			51622			
% of genes explained – 100 ancestors			42.79%			
% of genes explained – 200 ancestors			51.44%			

500

501 Up to the seventh generation (1972-1982), the mean number of ancestors (N_a) and of
 502 effective number of founders (f_e) grew simultaneously; however, only the N_a always
 503 continued growing, with the mean maximum of 1,572 ancestors known in the tenth
 504 generation. The same was not observed for the f_e , which had a decrease of 24.46% after the
 505 seventh generation, but returned to grow from the ninth generation on (Fig. 9).



506
 507 **Fig. 8.** Evolution of the mean numbers of ancestors and effective number of founders in the
 508 last five generations of the Mangalarga Marchador horse breed population.
 509

510 Genetic contribution

511 The numbers of founders that exhibit 50% and 10% of the gene pool of the breed are
 512 small and represent only 7.94% and 6.64% of the total founders of the TP, respectively (Table
 513 4). For RP, these values are 19.09% and 0.69%. The total genetic contributions of the most
 514 influential ancestor and of the 10 most influential ancestors increased in the RP by 2.49% and
 515 12.73% in relation to the TP, and by 1.18% and 8.16% considering the contributions of the
 516 founders. The number of founders with contributions greater than 1.0% were 2 and 10
 517 animals in the TP and in the RP, respectively. The most influential founding mare was
 518 Herdade Alteza, with 1.24% marginal contribution in the TP and with 2.42% in the RP. The
 519 stallion Herdade Cadillac contributed 2.76% in the RP and the mare Azarola de Santa

520 Terezinha 0.27% in the TP. These were the most influential ancestors in the Mangalarga
521 Marchador population.

522

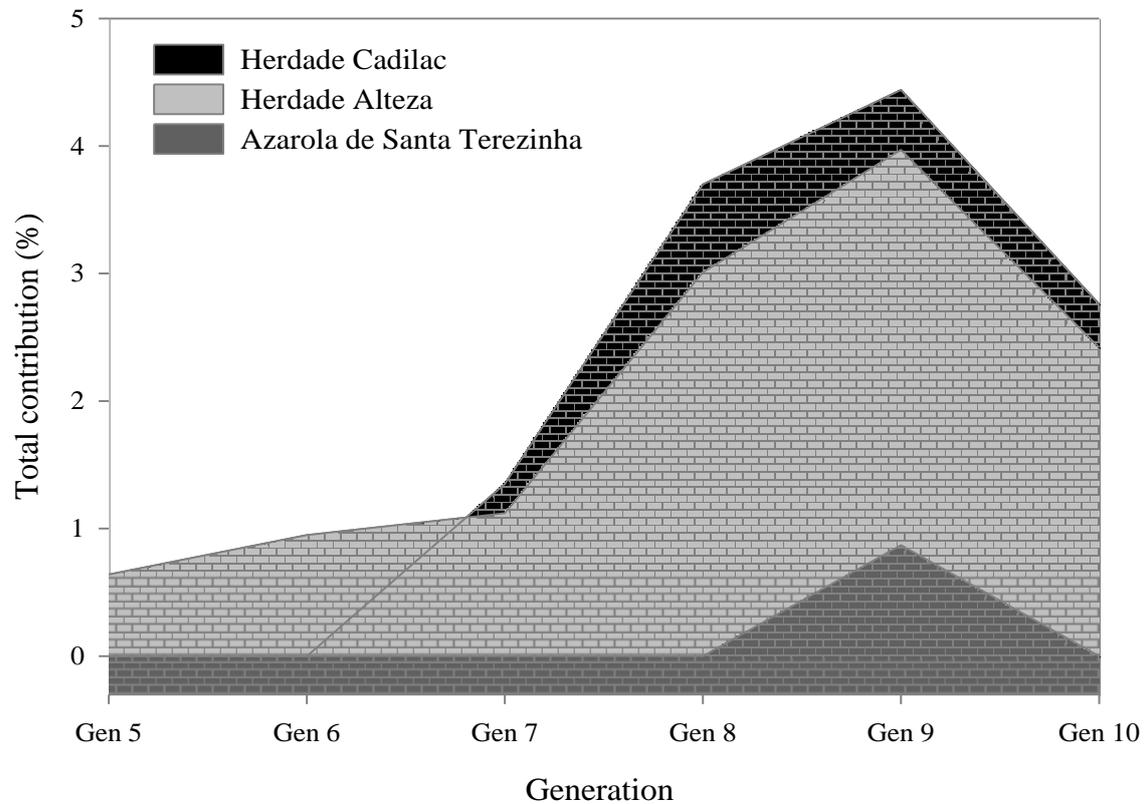
523 **Table 4.** Population parameters and genetic contributions of ancestors and founders of the
524 Mangalarga Marchador horse breed population.

Item	Total Population	Reference Population
No. of animals with both parents known	488,676	133,069
No. of animals with one of the parents known	2,060	10
No. of animals with both parents unknown	25,607	728
Contributions of ancestors and founders		
No of ancestors (50% of the gene pool)	4,774	2,561
No. of ancestors (10% of the gene pool)	774	6
No. of founders (50% of the gene pool)	2,034	139
No. of founders (10% of the gene pool)	17	5
Genetic Contribution (%)		
The most influential ancestor	0.27 ^a	2.76 ^b
The 10 most influential ancestors	1.47	14.20
The most influential founder	1.24 ^c	2.42 ^c
The 10 most influential founders	7.67	15.83

525 a = Herdade Alteza; b = Herdade Cadillac; c = Azarola de Santa Terezinha.

526

527 The evolution of the genetic contributions of the main ancestors and founders from the
528 fifth generation to the tenth generation are shown in Figure 8. The highest genetic
529 contributions for the three most influential individuals occurred in the ninth generation, but
530 with declines from this generation on. The main founding mare (Herdade Alteza) made a
531 greater genetic contribution to the entire population in each new generation, but with a
532 significant decline (56.42%) in the last generation. The same occurrence was observed for the
533 ancestors Herdade Cadillac and Azarola, with declines of 62.16% and 87% in the last
534 generation.



535
 536 **Fig. 9.** Evolution of the genetic contributions of the main ancestors and founders in the last
 537 five generations of the Mangalarga Marchador horse breed population.
 538

539 **4. Discussion**

540 Mangalarga Marchador has had a long history since it was formed more than 200 years
 541 ago, with pedigree information dating back to the nineteenth century. Several generations are
 542 known and, in this study, data from animals registered from 1906 to 2016 were used. These
 543 conditions provide a great deal of information for studying the evolution of genetic diversity
 544 and of factors that affect the breed.

545

546 4.1 Characterization of the population

547 Number of births

548 The number of registered births observed in the Mangalarga Marchador (MM) horse
549 population showed that, although the breed is still relatively recent in history, its size is
550 equivalent to that of older breeds known worldwide.

551 The decline observed in the number of births from 1991 to 1998 (Fig. 1) is similar to that
552 observed in other horse breeds in Brazil, as in Quarter Horse (Faria et al., 2018), Campolina
553 (Procópio et al., 2003; Bussiman et al., 2018), Mangalarga (Mota et al., 2006), and
554 Mangalarga Marchador (Costa et al., 2005). This period is highly correlated with a specific
555 moment in the Brazilian economy that witnessed a change in currency in the economic plan
556 and a crisis that lasted until 1994. This reduced investment in agribusiness drastically affected
557 the horse market for several consecutive years. As time passed after implementation of the
558 new Brazilian economic plan, with inflation under control, and after devaluation of the
559 currency in 1999, agribusiness returned to growth until 2005, according to Barros and Silva
560 (2008). The new sharp decline observed around 2008 coincides with the financial crisis of
561 Europe and the United States, which slowed horse raising in various regions of the world, as
562 in Brazil. The main modifications in the number of births throughout the period of study are
563 mainly due to changes in economic, financial, and social aspects confronted by Brazil.

564 In spite of the constant Brazilian crisis, the horse market grew by 12% in 2015, according
565 to the Instituto Brasileiro de Geografia e Estatística – IBGE (2016), with a turnover of around
566 16 billion reals (4-5 billion dollars) for the economy. However, no increase in the number of
567 births in the MM breed was observed in this period; on the contrary, it was the year with the
568 greatest decline in births (Fig. 1). However, Lima (2015) comments that the number of horses
569 indeed tends to decrease and the quality tends to increase, because horses used exclusively for
570 field tasks are being replaced by those with greater participation in leisure and various sports
571 modalities. Thus, these animals are going through careful quality selection and are directed to

572 events and expos, where horse raisers are able to achieve greater profits from the best animals.
573 This result is also confirmed by Figure 1, which shows that the number of reproducers in this
574 same period underwent a considerable reduction, especially in number of stallions.

575

576 Reproductive parameters

577 A reproductive difference between the sexes was observed, indicating greater
578 participation of females than males in reproduction (Table 1). The need for a smaller number
579 of males in horse breeding may be explained by the minimum score for official registry,
580 which is greater for males (140 points) than females (120 points). Therefore, many producers
581 prefer not to bear the costs of non-reproducing males and continue to raise them without
582 definitive registry, or normally the males considered not suitable for reproduction are
583 immediately castrated, accentuating even more the difference between the sexes. Similar
584 results were also found in the breeds Lusitano (Vicente et al., 2012), Brazilian Criollo (Maciel
585 et al., 2014), Quarter Horse (Faria et al., 2018), and Campolina (Bussiman et al., 2018),
586 among others.

587 There was also a significant difference between the sexes in relation to number of
588 progenies, and this is evident in comparing the mean and maximum number of foals per
589 reproducer in Table 1. Thus, not unlike other horse breeds, the males, though in smaller
590 number, are still more intensely used than females, for example, through the artificial
591 insemination technique adopted by horse breeders and the high standard deviation, the
592 preferential use of some breeders is also observed. This difference in mean number of
593 descendants per reproducer was also reported in the literature in the MM breed by Costa et al.
594 (2004), who observed a higher mean number of foals per stallion of 26.2 and maximum of
595 1,322, and a lower mean number of foals per female of 3.2 and maximum of 22, similar to the
596 present study. According to the same authors, mares came to generate more descendants after
597 official authorization of use of the embryo transfer technique by the ABCCMM. In the

598 Pantaneira breed, McManus et al. (2013) also found higher values in stallions than in mares
599 (12.8 foals per stallion and 2.6 foals per mare), as well as Medeiros et al. (2014) in the
600 Brazilian Sport Horse, with a mean of 2.89 for females and 19.16 for males. In general, the
601 number of foals per mare is low in the MM breed, just as in other horse breeds, in contrast to
602 the number of foals per stallion.

603 The stallion Elo Kafé da Nova and the mare Kibela do Passo Fino (Table 1) were the
604 reproducers with the largest number of progenies generated. This number is explained by the
605 sporting results of these animals, as they have collected Brazilian and regional titles of diverse
606 categories within the breed. The ratio of the number of females per male calculated in this
607 study (3.9) indicates reasonable intensity of use of stallions, which can be explained by the
608 use of artificial insemination in a controlled manner by breeders. This may contribute to
609 reduction in genetic relationship among animals, reduction in the number of inbred matings,
610 and, consequently, low inbreeding coefficients for the animals.

611 The age distribution of the reproducers indicates that the first foals were born to mares
612 with an average age of 11, which is three years older than the stallions in reproduction. This
613 shows long use for reproduction, especially for males, whereas mares are active for a shorter
614 period because they enter in reproduction later. At the end of the 1970s until the beginning of
615 the 1990s, greater use of females in reproduction with the age of 14 to 20 years was notable
616 (Fig. 3); this explains the greater average age of mares in reproduction compared to stallions.
617 These females were probably mares that stood out in competitions in the breed and, for that
618 reason, did not enter in reproduction at that time. Bussiman et al. (2018) also found a higher
619 age of Campolina mares in relation to males, with ages of 8.3 and 8.9 years, respectively.

620

621 Generation Interval

622 The generation interval (GI) in Mangalarga Marchador is generally considered quite long,
623 with a total mean for all four selection routes of 10.43 years in the TP and a little less in the
624 RP, 9.48 years. This result is greater than that found by Costa et al. (2004), who also studied
625 the MM breed, with population data up to the year 2000, and found the mean value of GI of
626 8.4 years. However, the GI of the MM breed is near the mean generation intervals reported for
627 other horse breeds, including Mangalarga (9.49 years, Mota et al., 2006), Hanoverian (10
628 years, Hamann and Distl, 2008), Spanish Arabian (10.1-12 years, Cervantes et al., 2009), and
629 Trakehner (10.2 years, Teegen et al., 2009). The GI in this study was greater than in the draft
630 horse breed of the Austrian Noriker (7.9 anos, Druml et al., 2009). Yamashita et al. (2010)
631 also reported a similar GI, from 10.5 to 11.5 in Thoroughbred horses in Japan. Roos et al.
632 (2015) likewise found a mean generation interval from 10.03 to 10.59 in Holsteiner horses.

633 The lower GI (Table 2) observed in the stallions may be associated with the preference of
634 most horse breeders to begin a foal early, at an age of 2-3 years, in morphological and canter
635 competitions in the breed, but remain in breeding for shortly. The choice of the male for
636 reproduction depends on the performance of the animal in these competitions. Nevertheless,
637 most males are parents at younger ages than females, thus reducing the GI in the stallions.
638 However, from the maternal side, the higher GI may also be motivated by training for
639 competitions. Another reason for the greater GI in females was their being used for a long
640 period (1970 to 1990), or the use of older females, from 14 to 20 years of age (Fig. 3). The
641 GIs of this study contrasted with that of Costa et al. (2004), who observed a greater GI in the
642 stallions (8.9) than in the mares (7.9).

643 Although the total GI is considered long and that of the mare is greater than that of the
644 stallion, a decrease in total GI was observed in the RP, which includes more recent years (Fig.
645 3), and the difference in GI between the sexes was still present, but at lower magnitudes

646 (Table 2). Reducing the generation interval of a population produces greater genetic gain in
647 characteristics of interest. In addition, reproducers remaining for a very long period in the
648 breeding stock increases the probability of mating between individuals with a kinship
649 relationship and, consequently, an increase in the inbreeding coefficient in the population
650 (Gonçalves et al., 2011).

651

652 *4.2 Quality of pedigree information*

653 Known generations

654 The mean values of the equivalent numbers of known generations (G) observed in the TP
655 and RP allowed evaluation of the evolution in genetic diversity that occurred over time (Table
656 2). The evolution of G shows the improvement in pedigree information of each animal in the
657 population. In the literature, values such as 0.7 have been reported in Pantaneiro (McManus et
658 al., 2013), 2.98 in Campolina (Bussiman et al., 2018), 5.7 in Spanish Arabian (Cervantes et
659 al., 2008), 6.24 in Quarter Horse (Faria et al., 2018), and 9.9 in Lusitano (Vicente et al.,
660 2012). The maximum number of generations traced in this study was greater than the 10.50 in
661 Maremmano (Giontella et al., 2018), and less than the 31.41 in Quarter Horse (Faria et al.,
662 2018). These results indicate difference in knowledge of pedigree information for each breed
663 analyzed.

664

665 Pedigree completeness

666 To characterize the genetic diversity of a breed, knowledge of the ancestors in the
667 pedigree is necessary. Incomplete datasets reduce the accuracy of the results, according to
668 Cervantes et al. (2009). The increase in pedigree completeness in the RP suggests
669 improvement in the quality of the information in the records of the ABCCMM in more recent
670 times (Fig. 3). Furthermore, the policy implemented by the association for all animals born to

671 be DNA tested to confirm paternity, and only that way be registered, has contributed to more
672 accurate genealogical construction in the breed. However, these results are still lower than
673 those we found in the literature in Medeiros et al. (2014), Duru (2017), and Faria et al. (2018),
674 who found pedigree completeness for the first (parents), second (grandparents), and third
675 (great grandparents) generation in the RP of 99.99%, 98.0%, and 69.9% (Brazilian Sport
676 Horse); 98.2%, 96.6%, and 95.0% (Arabian); and 100.0%, 100.0%, and 98.5% (Quarter
677 Horse), respectively. The mean values of the last five generations of the TP and RP were also
678 inferior to the 77.9% in the TP and 91.0% in the RP of Quarter Horse animals by Faria et al.
679 (2018) and to the 88% in the RP of Holsteiner horses by Roos et al. (2015). These results
680 suggest that the improvement already observed should continue for the Mangalarga
681 Marchador breed to achieve higher values of pedigree completeness

682

683 *4.3 Inbreeding Parameters*

684 Inbreeding coefficient and rate of inbreeding

685 Costa et al. (2005) studied the MM population with data on 286,047 horses in the years
686 1949-1999 from various states of Brazil and found F for the TP of 1.31%, F of the inbred
687 animals of 5.7%, maximum F of 46.9%, and 22.6% of the population as inbred. A more recent
688 study of Gonçalves et al. (2012) found an F value of 1.45% in the TP, F of 5.28% among the
689 inbred animals, maximum F of 28.13%, and 27.59% of the animals as inbred in a population
690 of 2,186 MM born from 1951 to 2006 in the state of Minas Gerais, Brazil. In comparison with
691 the present study, an increase in inbred animals to 32.61% in the TP is noted. Figure 5 shows
692 that from 2008 on, there was a significant increase, and these years were not analyzed in the
693 studies of the authors mentioned above. However, the F considering all the animals of the
694 population was lower (1.02%); F for the inbred animals (8.7%) was greater compared to these

695 studies; and maximum F was equal (46.87%) to that of Costa et al. (2005) and greater than
696 that of Gonçalves et al. (2012), who considered a smaller MM population.

697 Over the years, there was an increase in inbred animals in the population; however, a
698 small part of these animals had considerably high F, though not greater than had already been
699 observed in the past, and the large majority had low F. This promoted a greater F of the inbred
700 animals and maintained the mean F of the TP still low, comparing these results to the study of
701 Costa et al. (2005). These results imply that there is probably a certain control of horse raisers
702 over mating between animals with a high kinship relationship. However, these inbreeding
703 results of the Mangalarga Marchador might not be highly trustworthy because they are based
704 on pedigree data in which 2.9% of the total population was registered in or had records
705 changed to the Book for Inclusion in the TAC, i.e., animals without genealogy proven by the
706 DNA test. Another fact that supports this assertion is the occurrence of animals that were born
707 before 1949 (only 0.05% of the TP), the date of founding of the ABCCMM. These animals
708 were registered by another association, the ABCRM; thus, the reliability of these records is
709 not known for sure, and whether they are pure animals of the MM breed. Although the
710 percentage of these animals is small considering the total population, the inbreeding
711 coefficients might not reliably represent this parameter. However, the results were discussed
712 in accordance with the findings of this study.

713 Similar values of F were reported in the literature by Bartolomé et al. (2011) in
714 Hanoverian, with 1.33%, and in Quarter Horse, with 1.07% in the TP; and higher values of F
715 were observed in Lusitano, with 9.92% in the TP (Vicente et al., 2012), in Andaluz, with
716 8.48% in the TP (Valera et al., 2005), in Pantaneiro, with 4% (McManus et al., 2013), and in
717 Campolina, with 2.45% in the TP (Bussiman et al., 2018).

718 A difference between the mean values of maternal F and paternal F were observed in the
719 present study. This result may be due to the fact that stallions are generally more important in

720 the structure of the breeding stock. Therefore, a much higher investment is made in
721 acquisition of stallions, which can be bought as adults by other horse breeders, as an embryo,
722 or through acquisition of semen of stallions raised in more distant regions. This facilitates
723 their dissemination, as well as their more extensive use. The same result is observed in the
724 Campolina breed, with paternal F (2.98%) greater than maternal F (2.75) (Bussiman et al.,
725 2018). This is also the case in Costa et al. (2005), who found a higher value of paternal F
726 (2.8%) than maternal F (1.7%) in MM. The authors explained these results by the coincidence
727 that, in the same period, the open book of the males and females was closed, in 1966 and
728 1984, respectively.

729 The number of inbred animals at the end of the 1960s truly showed a slight increase, with
730 greater significance from the beginning of the 1980s. In this period, the F each year were also
731 high, decreasing after that (Fig. 5). This gradual increase in inbred animals was observed up
732 to 2014, years in which the open book was not active in the breed. From the year 2014 on, an
733 abrupt decline in inbred animals in the population, as well as a decrease in their F, is noted,
734 which can be explained by the reopening of the open book in 2015, both for males and for
735 females, which brought new animals and new alternatives for mating to the breed.

736 According to the FAO (1998), to have genetic variability at an acceptable level without
737 impeding gain from selection, the ΔF between two generations should not be greater than one
738 percent (1%), a value that was not exceeded by this study. The mean value of the ΔF in the
739 year was similar to that of Giontella et al. (2018) for the Maremmano breed (0.10%). The
740 highest ΔF in a year and generation occurred from 1974 to 1980, i.e., an interval within the
741 seventh generation, a time in the MM breed in which there were few males and females in
742 reproduction (Fig. 1). The same situation of few active reproducers was observed from 2009-
743 2010 on; however, the ΔF of these years and the ΔF of the last complete generation (2005-

744 2015) were negative. In other words, there was no increase in inbreeding in this period
745 analyzed.

746 In summary, in spite of the slight increase in F in the current population of the MM and
747 of the increase in the number of inbred animals over the years, lower maximum F and F of the
748 inbred animals was recently observed, and the ΔF was within the acceptable limit of
749 inbreeding. This increase in the F of the population was due to the increase in the number of
750 inbred animals; however, these animals mostly obtained a low level of individual inbreeding.
751 All these results once more indicate the care of horse breeders in matings between individuals
752 of the same family, which should always be monitored for the breed.

753 According to Bergmann et al. (1997), an F greater than 10% can lead to inbreeding
754 depression, i.e., a reduction in productive traits and, especially, in adaptive traits. This value
755 was not exceeded in this study. Positive and negative results have been published for the
756 effect of the increase in F on morphological traits in horses (Gandini et al., 1992; Beergmann
757 et al., 1997; Wolc and Balinska, 2010); these traits are also of great importance in the MM
758 breed. According to the selection criteria practiced, high F may cause an increase in the
759 frequency of homozygous genotypes for alleles that are truly favorable, allowing the estimate
760 of the positive effects of inbreeding on some morphological traits. Inbreeding has been used
761 as a tool to achieve more homogeneous groups of progenies and for fixation of some desirable
762 traits. However, the genetic associations between morphological traits and functionality traits,
763 such as health, performance, and survival, can be limiting factors in the abusive use of these
764 inbred matings. Sevinga et al. (2014) reported that the inbreeding found in their study was, at
765 least partially, the cause of the high occurrence of placenta retention in horses of the Friesian
766 breed. This suggests that the increase in inbreeding can cause not only reduction in
767 morphological phenotypes, but also drastic consequences, even though this may be in the long
768 term in a population.

769

770 Effective size

771 The effective size of the population (N_e) recommended to maintain genetic diversity at an
772 acceptable level for its conservation is at least 50 animals (FAO, 1998; Meuwissen and
773 Woolliams, 1994). In this study, the mean values of N_e by ΔF were higher than 50, both in the
774 TP and in the RP. The lowest N_e (ΔF) for the year observed in the MM population occurred
775 from 1968 to 1999, and the minimum value was estimated at 103 in 1997, which is double the
776 minimum N_e recommended for maintaining genetic diversity. The values of the N_e (ΔF) in the
777 MM were much higher than those found in the literature, including the breeds Arabian ($N_e =$
778 110-135, Cervantes et al., 2009), Noriker ($N_e = 157$, Druml et al., 2009), Trakehner ($N_e = 150$,
779 Teegen et al., 2009), Spanish Sport Horse ($N_e = 226$, Bartolomé et al., 2011), Holsteiner ($N_e =$
780 55, Roos et al., 2015), and Campolina ($N_e = 93$, Bussiman et al., 2018).

781 Even considering N_e computed from number of parents, the mean for TP (7,039) and for
782 RP (22,232) was much greater than the minimum recommended for maintaining genetic
783 diversity. However, the lowest values of N_e (number of parents), fewer than 50 animals, only
784 occurred in the years from 1906 to 1959, when the population of the MM breed was still
785 being established, and matings occurred only by natural mounting. However, the values
786 obtained for N_e (ΔF or number of parents) of this study were not a matter of concern for
787 maintaining diversity. Nevertheless, populations should always be monitored since current
788 reproductive techniques can negatively affect N_e and generate inbreeding depression in the
789 morphological and reproductive traits and for survival.

790

791 Rate and mean of the additive genetic relationship

792 The mean of the additive genetic relationship (AGR) among individuals provides
793 complementary information to the inbreeding values originating from matings of individuals

794 with a kinship relationship (Gutierrez et al., 2003). The mean values of the AGR in the TP
795 and RP are considered low; however, in the RP, this value nearly doubled, explained by the
796 increase in the number of inbred animals (Fig. 5) and by the smaller number of active
797 reproducers in the same period (Fig. 1). The increase was also observed in the rate of the
798 AGR (Δ AGR) in the RP. Values considered low were also described in the literature by other
799 authors (0.0013, McManus et al. 2013; 0.0016, Bartolomé et al. 2011; 0.0116, Faria et al.
800 2018), which shows a reduced representation of each individual in the entire MM breed,
801 regardless of pedigree completeness. The low value of the AGR may be associated with the
802 introduction of new reproductive technologies, such as embryo transfer and artificial
803 insemination, which were introduced in the breed in 1994 and 1995, i.e., a little more than
804 twenty years ago, which is relatively recent, since the AGR was obtained for the entire
805 population. For 90 years, these technologies had not yet been applied, explaining a low AGR.
806 Another factor that could affect the results is the increase in the number of exchanges of live
807 animals (by sale / trade) among the breeding stock, due to improvement in transport and
808 management practices after the 1990s. However, though still low, there was an increase in the
809 rate of the AGR recently, which should be an alert for breeders to continue correct monitoring
810 of matings among their animals, since the number of reproducers has decreased. In addition,
811 just as in the estimated inbreeding coefficients, the AGR might not faithfully represent this
812 parameter, due to the animals in the TAC book.

813

814 Family size

815 According to Falconer and Mackay (1996), an ideal condition in the population is when
816 all the parents have the same chance of producing progenies in the following generation.
817 However, in the practice of animal production, this is not observed. The unequal contribution
818 of parents promotes change in family size, resulting in an increase in the rate of inbreeding
819 and reduction in effective population size. A big variation in family size was observed in the
820 Mangalarga Marchador population, with a decrease from the 1980s on and greater family size
821 of stallions in relation to mares. The greater family size in stallions may be due to growing
822 use of artificial insemination.

823 Due to the variation in family size found in this study, it is possible to implement the
824 selection method called "optimal genetic contribution" which increases the response to the
825 selection applied because this method combines families without a kinship relationship,
826 avoiding relationships with descendants and, consequently, promoting low levels of
827 inbreeding in the parents of the next generation (Sonesson and Meuwissen, 2000). Another
828 consequence of the method is maintaining the highest possible values of population size,
829 which balances the number of males and females, standardizing family size and introducing
830 genetic exchange from rotation of males (Teegen et al., 2009). In horse breeding up to now,
831 one observes that the greatest genetic progress is always achieved by applying increasing
832 selection intensities, normally with the use of artificial insemination. However, the selection
833 used is largely based on the animal's own performance in competitions, therefore using mass
834 selection, instead of that based on estimates of genetic merit. Teegen et al. (2009) assert that
835 by increasing selection accuracy, it would be possible to reduce selection intensities without
836 decreasing genetic progress in the population. This selection method could be applied
837 correctly to balance family size in the breed and not increase inbreeding and the effective MM
838 population size.

839

840 *4.4 Probability of gene origin*

841 Founders, ancestors, and founder genomes

842 Evaluation of the genetic contributions of founders and ancestors to the current
843 population can provide not only an analysis of inbreeding, but also better understanding of the
844 changes that occur in the gene pool in a breed, especially when a small number of generations
845 is considered (Boichard et al., 1997). When only the number of founders and ancestors that
846 contribute to the gene pool is considered, and the possibility of unbalanced contributions
847 among the different founders and ancestors is not taken into account, genetic bottlenecks are
848 not well detected in that population under study. However, the use of an effective number of
849 them is normally preferred, which, in the case of founders, corresponds to the equivalent
850 number of founders that would generate the same genetic diversity as that observed in the
851 population under study if all made the same contribution. The number of effective ancestors,
852 for its part, provides more information regarding the genetic changes that have occurred in the
853 population in a short period of time.

854 The effective number of founders (f_e), which is near 1062 in the TP (Table 3), has been
855 increasing (Fig. 8) as of the ninth generation (1994 to 2004); however, there was a significant
856 decline of nearly 25% from the seventh to the ninth generation (1972 to 1994), which may
857 explain the reduction in this parameter in the RP (225). The number of founder genomes (f_g)
858 also decreased among the populations, but with low significance (150 to 133). The f_e and f_g
859 in MM is considered high and much better than most reported in other horse breeds (Vicente
860 et al., 2012; Bussiman et al, 2018; Roos et al., 2015; Delgado et al., 2014) and comparable
861 only to the results obtained in Quarter Horse (Faria et al., 2018) and in Paint Horse in the
862 study of Siderits et al. (2013). However, the ratio of f_e with the total number of founders (N_f)

863 is nearly five times less for Quarter Horse, indicating that its number of founders ($f_e = 1045$
864 for $N_f = 5427$) is better represented in its population than in Mangalarga Marchador.

865 The effective number of ancestors (1038) found for the TP (Table 3) in MM is also
866 considered high and is one of the highest compared to those found in the literature for other
867 breeds, which range from 12 to 156 (Zechner et al., 2002; Valera et al., 2005; Poncet et al.,
868 2006; Cervantes et al., 2008; Hamann and Distl, 2008; Druml et al., 2009; Vicente et al.,
869 2012; Bussiman et al., 2018; Faria et al., 2018;). A higher value was only found for
870 Pantaneiro, with 5086 of f_a (Maciel et al., 2014), a breed that was formed earlier than MM. In
871 spite of the increase in the mean number of ancestors over the years (Fig. 8), the f_a of the RP
872 was less than that of the TP, indicating that most of these new ancestors did not greatly
873 contribute genetically to the population, since in this period there was an important reduction
874 in the use of reproducers (Fig. 1).

875 According to the literature itself, there is great diversity in the horse breeds observed, as
876 well as in the probabilities of gene origin, but it suggests concern in horse breeding, with most
877 of the studies expressing low values for these parameters. Nevertheless, the variation in the
878 three probability parameters (f_a , f_e , f_g) of gene origin obtained in MM was greater than the
879 vast majority described in the literature. However, caution must be taken in comparison of
880 these data because each population has its own characteristics, and comparison is carried out
881 within each breed and always considering the three parameters of each population.

882 Reductions in genetic variability can occur due to unequal contribution of founders,
883 leading to bottlenecks and events such as genetic drift (Duru, 2017). The discrepant
884 reductions in f_e and f_a in comparison to the real values found (N_f and N_a) in this study
885 suggest loss of genetic variability in the breed (Table 3), though the results found in this study
886 are greater than in other studies already cited.

887 In summary, the high values found in the TP for f_e , f_a , and f_g compared to the literature,
888 the increasing values of the mean number of ancestors and of f_e of the most recent
889 generations (Fig. 8), the lower values of f_e , f_a , and f_g found in the RP in relation to the TP
890 (Table 3), and the evident discrepancy of these values with the real values suggest
891 concentration of the population in a reduced number of reproducer animals (Fig. 1) and
892 progressive loss of the genetic contributions of the principal founders and ancestors (Fig. 9).
893 This can cause a reduction in genetic variability in the breed, though this is still not very
894 clearly explained by the values of inbreeding and AGR found in the study, which showed a
895 small increase in the RP.

896 These results confirm the importance of the use of probability analysis and gene origin
897 for monitoring genetic diversity, proposed by Boichard et al. (1997), because they provide
898 information regarding the occurrence of genetic bottlenecks and genetic drift that then
899 generate loss of genetic diversity in a short period of time within a population.

900 The ratios between N_e , N_f , f_e , f_a and f_g are considered metrics of genetic diversity of a
901 population by providing information regarding the occurrence of bottlenecks and genetic drift
902 in the population. The f_e should be near N_e ($f_e/N_e \sim 1$) in a population in which genetic drift
903 has stabilized (Caballero and Toro, 2000; Sorensen et al., 2005) and the f_e/f_a ratio should be
904 near 1. If not, a bottleneck occurs in the population (Boichard et al., 1997), and if f_a is clearly
905 less than f_e (f_e/f_a ratio > 1), there is a bottleneck effect. The f_g/f_e ratio indicates shrinkage of
906 genetic stock, i.e., loss of genetic diversity in the population, whenever f_g is less than f_e (f_g
907 $/f_e < 1$). The comparison of f_g and N_f (f_g/N_f) reveals the percentage of founder genomes that
908 are retained in the population. The greater this is, the more genetic diversity from the founders
909 is maintained.

910 In this study, the f_e/N_e ratio decreased in the RP, indicating that genetic drift was lower
911 and more recently no longer exists, with progressive loss of founder representation (Fig. 9).

912 This is further confirmed by f_g (150), which is nearly one fifth of the f_e in the TP (1,062) and
913 half (133) in the RP (225), which also reflects loss of alleles over time. The f_e/f_a ratio was
914 1.02 in the TP and 1.14 in the RP, confirming the greater recent occurrence of bottlenecks in
915 the population due to preferential use of certain lines of founders; but the genetic bottleneck
916 had also already occurred before. The f_e in the two populations, TP and RP, compared to the
917 number of founders (N_f), reaffirms the preferential use of certain lines of founders. The f_a
918 parameter complements the information from f_e in the f_e/f_a ratio, indicating losses of genetic
919 diversity, produced by unequal use of reproducer animals. The smaller difference between f_e
920 and f_a indicates the greater number of founder animals in the population over generations,
921 allowing the genetic base of the breed and of a large number of founders to be maintained
922 over time, thus reducing loss of genetic diversity, which contrasts with that observed in the
923 MM breed. The results also contrast with those obtained for Holsteiner (1.72), Arabian (2.31),
924 Lusitano (2.33), and Spanish Sport Horse (2.37) breeds in the studies of Roos et al. (2015),
925 Delgado et al. (2014), Vicente et al. (2012), and Bartolomé et al. (2011), respectively, who
926 observed greater bottlenecks in the formation of populations compared to the present study.

927 This progressive reduction in genetic variability was also confirmed by the result of the
928 f_g/f_e ratio in the TP (0.14), and this reduction continues in the RP, though at lower intensity
929 (0.60). Because of this situation, selection planning should once more also consider the need
930 to maintain genetic variability in the population. A smaller bottleneck was estimated in the
931 Austrian Noriker (4.00) (Druml et al., 2009). The f_g/N_f ratio maintains a relatively greater
932 proportion of genetic diversity from the founders in the more recent population (RP) than in
933 the TP; yet it is a considerably low value (18%). A similar value was also found in native
934 Japanese horses (10.6%) (Onogi et al., 2018).

935

936 Genetic contribution

937 Differences were found in the genetic contributions of all the founders and ancestors that
938 gave rise to the Mangalarga Marchador breed, and the founders were responsible for most of
939 the greatest genetic contributions to the breed (Table 4). In the TP, 17 effective founders
940 contributed to 10% of the gene pool, and two of them alone contributed more than 1% each.
941 In the RP, five founders contributed with the same value (10%), and three of them alone with
942 2% each. Other genetic contributions were described by Hamann and Distl (2008), in which
943 15 principal founders contributed up to 35% in the Hanoverian breed. Roos et al. (2015)
944 found that 15 animals are responsible for 50.0% of the total contribution, with a value of
945 11.6% for the most influential animals. Vicente et al. (2012) reported a contribution of the
946 principal founder of 25% in the Lusitano breed. Thus, differences found in the literature
947 indicate that the MM exhibits comparatively greater genetic diversity, because there are many
948 founders and ancestors contributing a little to the gene pool in the breed, and the number of
949 ancestors that explain the number of segregating genes in the population reaffirms this
950 conclusion (Table 3). However, the most significant (high) contributions were observed in the
951 RP (Table 4) and, as observed in Figure 9, the contributions of the principal founders and
952 ancestors has recently decreased, which contrasts with the expectation that the relative
953 contributions of the founder tend to stabilize in the populations after a small number of
954 generations (Caballero and Toro, 2000; McPhee and Wright, 1925), which has already been
955 observed in Thoroughbred (Cunningham et al., 2001).

956 The decline of the genetic contribution of the main ancestors and founders observed from
957 generation 9 (Figure 9) may indicate a bottleneck in the population, which is proven by Figure
958 1. Thus, the genetic contributions should be monitored by the ABCCMM and by breeders in
959 order to direct matings, so as to obtain more balanced values and maintain the genetic
960 diversity of the MM breed in Brazil.

961 **5. Conclusion**

962

963 The Mangalarga Marchador population does not exhibit severe inbreeding that would
964 impair evolution of the breed in Brazil. Nevertheless, analysis of probability of gene origin
965 showed a certain loss of genetic diversity in the population, especially due to genetic
966 bottlenecks.

967 Knowledge of the population structure provides an overall vision of the population and is
968 a first step toward development of optimal mating and strategic selection to minimize or even
969 prevent future loss of genetic diversity. In Mangalarga Marchador, it is still necessary to take
970 care in monitoring genetic variability to prevent change in the level of inbreeding that may
971 result in negative changes in essential traits used in competitions within the breed.

972 Efforts must be made in directing matings considering the relationship between stallions
973 and females. An increase in the contingent of reproducers, reduction in the generation interval
974 by selecting younger animals and animals with a shorter time in the breeding stock, use of the
975 “optimal genetic contribution” method, implementation of modern software for optimization
976 of reproduction programs, and priority in family size are some breeding practices that can be
977 applied in genetic management in the Mangalarga Marchador breed.

978

979 **Acknowledgments**

980 The authors acknowledge Federal University of Lavras, CNPq and ABCCMM for
981 providing the data used in this study.

982

983 **References**

984

985 Associação Brasileira de Criadores de Cavalos Mangalarga Marchador (ABCCMM), 2018.

986 Mercado do Mangalarga Marchador. [https://www.g1.globo.com/sao-](https://www.g1.globo.com/sao-paulo/itapetiningaregiao/especial-publicitario/)

987 paulo/itapetiningaregiao/especial-publicitario/ (accessed 05 August).

988

989 Assis, J.B., Laat, D.M., Peixoto, M.G.C.D., Bergman, J.A.G., Fonseca, C.G., Carcavlo,

990 M.R.S., 2009. Genetic diversity and population structure in Brazilian Mangalarga

991 Marchador horses. *Genetics and Molecular Research*. 8, 1519-1524.

992

993 Barros, G.S.C., Silva, S.F., 2008. A balança comercial do agronegócio brasileiro de 1989 a

994 2005. *Rev. Econ. Sociol. Rural*. 46, 905-936. [http://dx.doi.org/10.1590/S0103-](http://dx.doi.org/10.1590/S0103-20032008000400001)

995 20032008000400001.

996

997 Bartolomé, E., Cervantes, I., Valera, M., Gutiérrez, J.P., 2011. Influence of foreign breeds on

998 the genetic structure of the Spanish Sport Horse population. *Livest. Sci*. 142, 70–79.999 <http://dx.doi.org/10.1016/j.livsci.2011.06.021>.

1000

1001 Bergmann, J.A.G., Costa, M.D, Mourão, G.B., Hourí Neto, M., 1997. Formation and breeding

1002 structure of the Brasileira pony breed. *Arquivo Brasileiro de Medicina Veterinária e*1003 *Zootecnia*. 49, 251–259.

1004

1005 Boichard, D., Maignel, L., Verrier, E., 1997. The value of using probabilities of gene origin to

1006 measure the genetic variability in a population. *Genet. Sel Evol*. 29, 5-23. [http://dx.](http://dx.doi.org/10.1186/1297-9686-29-1-5)1007 [doi.org/10.1186/1297-9686-29-1-5](http://dx.doi.org/10.1186/1297-9686-29-1-5).

1008

- 1009 Boichard, D., 2002. Pedig: a Fortran package for pedigree analysis suited for large
1010 populations. In: World Congress on Genetics Applied to Livestock Production, 7., 2002,
1011 Montpellier. Proceedings. Montpellier: Institut National de la Recherche Agronomique.
1012 13-28.
1013
- 1014 Bussiman, F.O., Perez, B.C., Ventura, R.V., Peixoto. M.G.C.D., 2018. Pedigree analysis and
1015 inbreeding effects over morphological traits in Campolina horse population. 22, 1-10.
1016 <https://doi.org/10.1017/S175173111800023X>.
1017
- 1018 Caballero, A., Toro, M., 2000. Interrelations between effective population size and other
1019 pedigree tools for the management of conserved populations. Genet. Res. 75, 331–343.
1020 <http://dx.doi.org/10.1017/S0016672399004449>.
1021
- 1022 Caiado, J. R.C., Fonseca, F.A., Silva, J. F. S., Fontes, R.S. 2007. Tratamento de éguas
1023 receptoras de embriões visando sua utilização no segundo dia pós-ovulação. R. Bras.
1024 Zootec. 36, 360-368, 2007. ISSN impresso: 1516-3598.
1025
- 1026 Cervantes, I., Molina, A., Goyache, F., Gutiérrez, J.P., Valera, M., 2008. Population history
1027 and genetic variability in the Spanish Arab Horse assessed via pedigree analysis. Livest.
1028 Sci. 113, 24–33. <http://dx.doi.org/10.1016/j.livsci.2007.02.011>.
1029
- 1030 Cervantes, I., Gutiérrez, J.P., Molina, A., Goyache, F., Valera, M., 2009. Genealogical
1031 analyses in open populations: the case of three Arab-derived Spanish horse breeds. J.
1032 Anim. Breed. Genet. 126, 335–347. <http://dx.doi.org/10.1111/j.1439-0388.2008.00797.x>.
1033

- 1034 Costa, M.D. et al., 2004. Caracterização demográfica da raça Mangalarga Marchador. Arquivo
1035 Brasileiro de Medicina Veterinária e Zootecnia. 56, 687-690.
1036
- 1037 Costa, M.D., Bergmann, J.A.G., Resende, A.S.C., Fonseca, C.G., 2005. Análise temporal da
1038 endogamia e do tamanho efetivo da população de equinos da raça Mangalarga
1039 Marchador. Arq. Bras. Med. Vet. Zootec. 57, 112-119.
1040
- 1041 Cunningham, E.P., Dooley, J.J., Splan, R.K., Bradley, D.G., 2001. Microsatellite diversity,
1042 pedigree relatedness and the contributions of founder lineages to thoroughbred horses.
1043 Anim. Genet. 32, 360–364. <http://dx.doi.org/10.1046/j.1365-2052.2001.00785.x>.
1044
- 1045 Delgado, J.F., De Andrés, N., Valera, M., Gutiérrez, J.P., Cervantes, I., 2014. Assessment of
1046 population structure depending on breeding objectives in Spanish Arabian horse by
1047 genealogical and molecular information. Livest. Sci. 168, 9–16. <http://dx.doi.org/10.1016/j.livsci.2014.07.012>.
1048
1049
- 1050 Druml, T., Baumung, R., Sölkner, J., 2009. Pedigree analysis in the Austrian Noriker draught
1051 horse: genetic diversity and the impact of breeding for coat colour on population
1052 structure. Journal of Animal Breeding and Genetics. 126, 348–356.
1053
- 1054 Duru, S., 2017. Pedigree analysis of the Turkish Arab horse population: structure, inbreeding
1055 and genetic variability. Animal. 11, 1449–1456.
1056 <http://dx.doi.org/10.1017/S175173111700009x>.
1057

- 1058 Falconer, D., Mackay, T., 1996. Introduction to Quantitative Genetics, 4th ed. Longmans
1059 Green, Harlow, Essex, UK, pp. 464.
1060
- 1061 FAO, 1998. Secondary Guidelines for Development of National Farm Animal Genetic
1062 Resources Management Plans: Management of Small Populations at Risk. FAO, Rome,
1063 Italy.
1064
- 1065 Faria, R.A.S., Maiorano, A.M., Bernardes, P.A., Pereira, G.L.M., Curi, R.A., 2018.
1066 Assessment of pedigree information in the Quarter Horse: Population, breeding and
1067 genetic diversity. *Livestock Science*. 214, 135-141, 2018.
1068 <https://doi.org/10.1016/j.livsci.2018.06.001>.
1069
- 1070 Fonseca, M. G., Ferraz, G. C., Lage, J., Pereira, G. L., Curi, R. A., 2017. A Genome-Wide
1071 Association Study Reveals Differences in the Genetic Mechanism of Control of the Two
1072 Gait Patterns of the Brazilian Mangalarga Marchador Breed. *Journal of Equine
1073 Veterinary Science*. 53, 64-67.
1074
- 1075 Gandini, G.C., Bagnato, A., Miglior, F., Pagnacco, G., 1992. Inbreeding in the Italian
1076 Haflinger horse. *J. Anim. Breed. Genet*. 109, 433–443.
1077
- 1078 Giontella. A., Pieramati, C., Silvestrelli, M., Sarti, F. M., 2018. Analysis of founders and
1079 performance test effects on an autochthonous horse population through pedigree analysis:
1080 structure, genetic variability and inbreeding. *Animal*. 29, 1-10.
1081 <https://doi.org/10.1017/S1751731118001180>.
1082

- 1083 Gonçalves, R.W., Costa, M.D., Rocha Júnior V.R., Costa, M.R., Silva, E.S.P., Ribeiro,
1084 A.M.F., 2011. Inbreeding effect on reproductive traits in a herd of Mangalarga
1085 Marchador Brazilian horses . Rev. Bras. Saúde Prod. An., Salvador. 12, 641-649.
1086
- 1087 Gonçalves, R.W, Costa, M.D, Rezende, A.S.C., Rocha Júnior, V.R., Leite, J.R.A., 2012.
1088 Efeito da endogamia sobre características morfométricas em cavalos da raça Mangalarga
1089 Marchador. Arq. Bras. Med. Vet. Zootec. 64, 419-426.
1090
- 1091 Groeneveld, E., Westhuizen, B.V.D., Maiwashe, A., Voordewind, F., Ferraz, J.B.S., 2009.
1092 Gestão Populacional. 2009 enviado. POPREP: Um Relatório Genérico para Genética e
1093 Pesquisa Molecular. 8, 1158-1178.
1094
- 1095 Gutiérrez, J.P., Altarriba, J., Díaz, C., Quintanilla, R., Cañón, J., Piedrafita, J., 2003. Pedigree
1096 analysis of eight Spanish beef cattle breeds. Genet. Sel. Evol. 35, 43–63.
1097 <http://dx.doi.org/10.1051/gse>.
1098
- 1099 Gutierrez, J.P., Goyache, F., 2005. A note on ENDOG: a computer program for analysing
1100 pedigree information. J. Anim. Breed. Genet. 122, 172–176.
1101
- 1102 Hamann, H., Distl, O., 2008. Genetic variability in Hanoverian warmblood horses using
1103 pedigree analysis. J. Anim. Sci. 86, 1503–1513. <http://dx.doi.org/10.2527/jas.2007-0382>.
1104
- 1105 Hill, W., 2000. Maintenance of quantitative genetic variation in animal breeding programmes.
1106 Livest. Prod. Sci.. 63, 99-109.
1107

- 1108 Instituto Brasileiro de Geografia e Estatística (IBGE), 2018, Produção da pecuária municipal.
1109 <https://www.ibge.gov.br/home/> (accessed 04 August).
1110
- 1111 Lacy, R.C., 1989. Analysis of founder representation in pedigrees: founder equivalents and
1112 founder genome equivalents. *Zoo Biol.* 8, 111-123.
- 1113 Lima, R.A.S., 2015. Qual o tamanho do mercado relacionado à equinocultura no Brasil?
1114 *Revista Brasileira de Medicina Equina.* 10, 40-41.
1115
- 1116 Lima, R.A.S., Cintra, A.G., 2016. Revisão do Estudo do Complexo do Agronegócio do
1117 Cavalo. 1th ed. Brasília, Goiás. pp. 56 p.
1118
- 1119 MacCluer JW, Boyce AJ, Dyke B, Weitkamp LR, Pfennig DW, Parsons CJ. 1983. Inbreeding
1120 and pedigree structure in Standardbred horses. *J Hered.* 74:394–399.
1121
- 1122 Maciel, F.C, Bertoli,C.D, Braccini Neto, J., Cobuci, J.A., Paiva, S.R., McManus. C.M., 2014.
1123 Population structure and genealogical analysis of the Brazilian Crioula Horse. *Animal*
1124 *Genetic Resource.* 54,115-125. <https://doi.org/10.1017/S2078633613000489>.
1125
- 1126 Maignel, L., Boichard, D., Verrier, E., 1996. Genetic variability of French dairy breeds
1127 estimated from pedigree information. *Interbull.* 14, 49-54.
1128
- 1129 McManus, C., Santos, S.A., Lima Dallago, B.S., Rezende Paiva, S., Saraiva Martins, R.F.,
1130 Braccini Neto, J., Rocha Marques, P., Pinto de Abreu, U.G., 2013. Evaluation of
1131 conservation program for the Pantaneiro horse in Brazil. *Revista Brasileira de Zootecnia.*
1132 42, 404–413.

- 1133 McPhee, H., Wright, S., 1925. Mendelian analysis of the pure breeds of livestock III. The
1134 Shorthorns J. Hered. 16, 205–215.
1135
- 1136 Medeiros, B.R., Bertoli, C.D., Garbade, P., McManus, C.M., 2014. Brazilian Sport Horse:
1137 Pedigree Analysis of the Brasileiro de Hipismo Breed. Italian Journal of Animal Science.
1138 13, 31-46. <https://doi.org/10.4081/ijas.2014.3146>.
1139
- 1140 Meuwissen ,T.H.E., Luo, Z., 1992. Computing inbreeding coefficients in large populations.
1141 Genet Sel Evol. 24, 305-313. <https://doi.org/10.1186/1297-9686-24-4-305>.
1142
- 1143 Meuwissen, T.H.E., Woolliams, J.A., 1994. Effective sizes of livestock populations to prevent
1144 a decline in fitness. Theor. Appl. Genet. 89, 1019–1026.
1145
- 1146 Mota, M.D.S., Almeida Prado, R.S., Sobreiro, J., 2006. Characterization of the Mangalarga
1147 horse population in Brazil. Arch. Zootec. 55, 31–37.
1148
- 1149 Nascimento, J.F., 1999. Mangalarga Marchador: Tratado morfofuncional. Belo Horizonte:
1150 Associação Brasileira de Criadores de Cavalo Mangalarga Marchador, 1th ed. Belo
1151 Horizonte, Minas Gerais, pp. 900.
1152
- 1153 Oliveira, D. A. A.; Kuabara, M.Y., 2001. A genealogia, sua importância para o melhoramento
1154 e para produção animal e os métodos de confirmação dos pedigrees. Melhoramento
1155 Genético Aplicado á Produção Animal, 1th ed. Belo Horizonte, Minas Gerais, 435-444.
1156

- 1157 Onogi, A., Shirai, K., Amano, T., 2018. Investigation of genetic diversity and inbreeding in a
1158 Japanese native horse breed for suggestions on its conservation. *Animal Science Journal*.
1159 99, 1902-1910. <https://doi.org/10.1111/asj.12867>.
1160
- 1161 Pjontek, J., Kadlecčik, O., Kasarda, Horný, M., 2012. Pedigree analysis in four Slovak
1162 endangered horse breeds. *Czech Journal of Animal Science*. 57, 54–64.
1163
- 1164 Poncet, P., Pfister, W., Muntwyler, J., Glowatzki-Mullis, M., Gaillard, C., 2006. Analysis of
1165 pedigree and conformation data to explain genetic variability of the horse breed
1166 Franches-Montagnes. *J. Anim. Breed. Genet.* 123, 114–121.
1167
- 1168 Procópio, A.M., Bergmann, J.A.G., Costa, M.D., 2003. Formação e demografia da raça
1169 Campolina. *Arq. Bras. Med. Vet. e Zootec.* 55, 361–365.
1170 <http://dx.doi.org/10.1590/S0102-09352003000300018>.
1171
- 1172 Quaas R.L., 1976. Computing diagonal elements of a large numerator relationship matrix.
1173 *Biometrics*. 92, 949-953.
1174
- 1175 R Core Team. 2016. R: A language and Environment for Statistical Computing. R foundation
1176 for Statistical Computing. <http://www.R-project.or/> (accessed 01 August).
1177
- 1178 Rochambeau, H., La Fuente, L.F., Rouvier, R., Ouhayoun, J., 1989. Sélection sur la vitesse de
1179 croissance post-sevrage chez le lapin. *Genet Sel Evol.* 21, 527-546.
1180

- 1181 Roos, L., Hinrichs, D., Nissen, T., Krieter, J., 2015. Investigations into genetic variability in
1182 Holstein horse breed using pedigree data. *Livest. Sci.* 177, 25–32. [http://dx.doi.](http://dx.doi.org/10.1016/j.livsci.2015.04.013)
1183 [org/10.1016/j.livsci.2015.04.013](http://dx.doi.org/10.1016/j.livsci.2015.04.013).
1184
- 1185 Royo, L.J., Alvarez, I., Gutierrez, J.P., Fernandez, I., Goyache, F., 2007. Genetic variability in
1186 the endangered Asturcon pony assessed using genealogical and molecular information.
1187 *Livest. Sci.* 107, 162–169.
1188
- 1189 Santiago, J M., Rezende, A.S.C., Lana, A.M.Q., Fonseca, M.G., Lage, J., 2016. Evolution of
1190 Morphometric Measures in the Mangalarga Marchador Breed. *Revista Caatinga* 29, 191-
1191 199.
1192
- 1193 SAS Institute, 2004. *SASs 9.3 for Microsoft Windows*. SAS International, Heidelberg,
1194 Germany.
1195
- 1196 Sevinga, M., Vrijenhoek, T., Hesselink, J.W., Barkema, H.W., Groen, A.F., 2014. Effect of
1197 inbreeding on the incidence of retained placenta in Friesian horses. *Journal of Animal*
1198 *Science*. 82, 982–986.
1199
- 1200 Siderits, M., Baumung, R., Fuerst-Waltl, B., 2013. Pedigree analysis in the German Paint
1201 Horse: Genetic variability and the iNFLuence of pedigree quality. *Livest. Sci.* 151, 152–
1202 157. <http://dx.doi.org/10.1016/j.livsci.2012.10.018>.
1203

- 1204 Somogyvári, E., János, P., Mihók, S., 2018. Genetic Analysis of the Hungarian Population of
1205 Endangered Hucul Horses. Czech J. Anim. Sci. 63, 237–246.
1206 <https://doi.org/10.17221/54/2017-CJAS>.
1207
- 1208 Sonesson, A.K., Meuwissen, T.H.E., 2000. Mating schemes for optimum contribution
1209 selection with constrained rates of inbreeding. Genetics Selection Evolution. 32, 231–
1210 248.
1211
- 1212 Sorensen, A., Sorensen, M., Berg, P., 2005. Inbreeding in Danish dairy cattle breeds. J. Dairy
1213 Sci. 88, 1865–1872.
1214
- 1215 Souza, F.A.C., Fernandes, T. J. , Moura, R.S., Meirelles, S.L C., Ribeiro, A.R, Cunha, F.O.,
1216 Augusto, M.J., 2017. Nonlinear modeling growth body weight of Mangalarga Marchador
1217 horses. Ciência Rural, Santa Maria. 47, 2-6. [http://dx.doi.org/10.1590/0103-](http://dx.doi.org/10.1590/0103-8478cr20160636)
1218 [8478cr20160636](http://dx.doi.org/10.1590/0103-8478cr20160636).
1219
- 1220 Teegen, R., Edel, C., Thaller, G., 2009. Population structure of the Trakehner Horse breed.
1221 Animal. 3, 6–15. <http://dx.doi.org/10.1017/S1751731108003273>.
1222
- 1223 Tunnell, J., Sanders, J., Williams, J., Potter, G., 1983. Pedigree analysis of 4 decades of
1224 Quarter Horse breeding. J. Anim. Sci. 57, 585–593.
1225
- 1226 Valera, M., Molina, A., Gutiérrez, J. P., Gómez, J., Goyache, F., 2005. Pedigree analysis in
1227 the Andalusian horse: population structure, genetic variability and influence of the
1228 Carthusian strain. Livestock Production Science. 95, 57-66. ISSN 0301-6226.

1229

1230 VanRaden, P.M., 1992. Accounting for inbreeding and crossbreeding in genetic evaluation for
1231 large populations. *J Dairy Sci.* 75, 3136-3144.

1232

1233 Vicente, A.A., Carolino, N., Gama, L.T., 2012. Genetic diversity in the Lusitano horse breed
1234 assessed by pedigree analysis. *Livest. Sci.* 148, 16–25. [http://dx.doi.org/10.](http://dx.doi.org/10.1016/j.livsci.2012.05.002)
1235 [1016/j.livsci.2012.05.002](http://dx.doi.org/10.1016/j.livsci.2012.05.002).

1236

1237 Wolc, A., Balińska, K., 2010. Inbreeding effects on exterior traits in Polish Konik horses.
1238 *Achiv Tierzucht.* 53, 1–8.

1239

1240 Yamashita, J., Oki, H., Hasegawa, T., Honda, T., Nomura, T., 2010. Demographic analysis of
1241 breeding structure in Japanese Thoroughbred population. *Journal of Equine Science.* 21,
1242 11–16.

1243

1244 Zechner, P., Solkner, J., Bodo, I., Druml, T., Baumung, R., Achmann, R., Marti, E., Habe, F.,
1245 Brem, G., 2002. Analysis of diversity and population structure in the Lipizzan horse
1246 breed based on pedigree information. *Livest.Prod.Sci.* 77, 137–14.

ARTIGO 2 – Caracterização genética de equinos da raça Mangalarga Marchador baseado em marcadores moleculares microssatélites

Formatado de acordo com a norma do periódico *Genomics*.

M.M. Baena^a, S. Diaz^b, R.S.Moura^c, S.L.C., S.L.C. Meirelles^{d*}

^{a,c,d} Universidade Federal de Lavras, Department Animal Science, University Campus, postal code 3037, Downtown, CEP 37200000, Lavras/MG, Brazil; E-mail: marielle_moura@hotmail.com; raquelmourabr@yahoo.com.br; sarah@ufla.br*

^b Universidad Nacional de La Plata, Faculty of Veterinary Sciences, Calle 60 and 118 S/N (1900), Box 296, La Plata, Province of Buenos Aires - Argentina; E-mail: sil_diaz@hotmail.com.

*Corresponding author: Universidade Federal de Lavras, Department Animal Science, University Campus, postal code 3037, Downtown, CEP 37200000, Lavras/MG, Brazil; E-mail: sarah@ufla.br. ¹

¹ MM= Mangalarga Marchador; AND= Andaluz; PSL= Puro Sangue Lusitano; PSI= Puro Sangue Inglês; ARA= Árabe; CAM= Campolina; MAN= Mangalarga; Na= número de alelos absolutos; AR= número de alelos raros; Nm_a= número médio de alelos; PIC= conteúdo de informação polimórfica; He= heterozigosidade esperada; Ho = heterozigosidade observada; Fis, Fit e Fst = estatísticas F; ACP= análise de componentes principais; EHW= Equilíbrio de Hardy-Weinberg, DL= desequilíbrio de ligação

RESUMO

A raça de equinos Mangalarga Marcador (MM) possui um contingente superior a 600 mil animais oficialmente registrados, sendo expressiva sua importância na economia do Brasil. Com o avanço de tecnologias e a identificação de marcadores moleculares, tornou-se possível a quantificação da diversidade genética a partir de estudos de caracterização genética. Com isso, o objetivo neste estudo foi investigar a diversidade na raça MM e compará-la com raças relacionadas ou não à sua formação. Foi utilizado um banco de dados com total de 3.193 equinos genotipados (MM, n= 2.829; Andaluz n= 67; AND; Puro Sangue Lusitano n= 43; PSL; Puro Sangue Inglês n= 54; PSI; Árabe n= 99; ARA; Campolina n= 61; CAM e Mangalarga n= 40; MAN) para 13 *loci* de microssatélites (SSR). Distintos programas foram utilizados para estimar os parâmetros de diversidade como os números de alelos absolutos (N_a), alelos raros (AR), número médio de alelos (N_{m_a}), conteúdo de informação polimórfica (PIC), heterozigosidades esperadas (H_e) e observadas (H_o), estatísticas F (F_{is} , F_{it} e F_{st}), distâncias genéticas, análise de componentes principais (ACP), testes de equilíbrio de Hardy-Weinberg (EHW), de desequilíbrio de ligação (DL) e de diferenciação populacional, além de estruturação populacional pelo método de agrupamento Bayesiano. Foi averiguado um N_a de 646 alelos, considerando todas as raças e $N_a=141$ para o MM, o N_{m_a} variaram de 10,85 alelos (MM) a 5,31(PSI) e o AR prevaleceu na raça MM com média de dois alelos por *locus*. As médias de PIC por raça foram acima de 0,5, e valores máximo de 0,82 (LEX03) e mínimo de 0,53 (HSM6) foi observado no MM. As médias de H_e e H_o por raça foram de H_e (0,7364) e H_o (0,7354) e no MM os valores oscilaram de 0,5767 (HSM6) a 0,8450 (LEX03) de H_e e 0,5331 (HSM6) a 0,7949 (ASB17) de H_o . Todos estes parâmetros citados evidenciam a presença de diversidade genética no MM e o bom poder discriminatório dos SSR. Para a população de MM foram estimadas as médias de F_{is} (-0,0195), F_{it} (0,0566) e F_{st} (0,0748) nos quais indicaram pequeno excesso de heterozigotos nas subpopulações, pouca deficiência de heterozigotos na população total e moderada diferenciação. Observou-se desvios de EHW, diferenciação populacional alélica e genotípica em todos os *loci*, além de 60,26% dos alelos demonstrarem forte DL nos quais demonstram possivelmente a presença de cruzamentos, seleção, fluxo gênico, endogamia e deriva genética. As distâncias genéticas foram maiores comparando as raças ARA e PSI com as demais raças estudadas do que entre as raças brasileiras MM, MAN e CAM e estas com as da Península Ibérica AND e PSL, sendo este resultado corroborado pelo ACP e nos quais comprovam que as raças brasileiras são próximas devido a ancestrais comuns, assim como, estas com as da Península Ibérica. A análise de estruturação populacional inferiu cinco agrupamentos genômicos mais prováveis para as raças e também que o MM foi uma entidade genética subestruturada, porém com presença de alguns haras mais definidos geneticamente. A falta de estruturação populacional valida a ocorrência de cruzamentos e fluxo gênico, porém a gestão genética local de alguns haras demonstram uma melhor estruturação. O estudo contribui para o conhecimento da estrutura genética e caracterização molecular da população MM, além de evidenciar como os SSR podem ser utilizados para estabelecer as relações genéticas entre populações, além de fornecer informações valiosas para monitorar a seleção, tornando-se ferramentas fundamentais para a pesquisa básica em programas de genômica animal.

Palavras-chave: diversidade genética, equideocultura, polimorfismo, SSR

ABSTRACT

The Mangalarga Marcador (MM) equine breed has a quota of more than 600 thousand officially registered animals, and its importance in the Brazilian economy is significant. With the advancement of technologies and the identification of molecular markers, it became possible to quantify genetic diversity from genetic characterization studies. With this, the objective of this study was to investigate the diversity in the MM race and to compare it with races related or not to their formation. A database was used with total of 3,193 horses were genotyped (MM, n = 2829, Andalusian n = 67, AND, Lusitano n = 43, LUS, Thoroughbred n = 54, THO, Arabian n = 99, ARA, Campolina n = 61; CAM and Mangalarga n = 40; MAN) for 13 microsatellite loci (SSR). Different programs were used to estimate diversity parameters such as absolute alleles (Na), rare alleles (RA), mean alleles (Nma), polymorphic information content (PIC), expected heterozygosity (He) and observed (Ho), genetic distances, principal component analysis (PCA), Hardy-Weinberg Equilibrium (HWE), linkage disequilibrium (LD) and population differentiation, in addition to populational structuring by the Bayesian grouping method. A Na of 646 alleles was found, considering all races and Na = 141 for MM, Nma ranged from 10.85 alleles (MM) to 5.31 (THO) and RA prevailed in the MM breed with a mean of two alleles by locus. The mean PIC by race were above 0.5, and maximum values of 0.82 (LEX03) and minimum values of 0.53 (HSM6) were observed in the MM. The mean He and Ho by breed were He (0.7364) and Ho (0.7354) and in MM the values ranged from 0.5767 (HSM6) to 0.8450 (LEX03) of He and 0.5331 (HSM6) at 0.7949 (ASB17) of Ho. All these cited parameters, evidenced the presence of genetic diversity in the MM and the good discriminatory power of the SSR. For the MM population, the means of Fis (-0.0195), Fit (0.0566) and Fst (0.0748) were estimated in which they indicated a small excess of heterozygotes in the subpopulations, a low heterozygote deficiency in the total population, and moderate differentiation. HWE deviations, allelic and genotypic population differentiation were observed in all loci, and 60.26% of the alleles demonstrated strong DL in which they possibly demonstrate the presence of crosses, selection, gene flow, endogamy and genetic drift. Genetic distances were higher comparing the ARA and THO races with the other races studied than between the Brazilian, MM, MAN and CAM races and those with the Iberian Peninsula AND and LUS, and this result is corroborated by the PCA and in which the Brazilian breeds are close because of common ancestors, as well as those of the Iberian Peninsula. The analysis of population structure inferred five genomic groupings more probable for the races and also that the MM was a substructured genetic entity, but with the presence of some more genetically defined herds. The lack of population structure validates the occurrence of crosses and gene flow, but the local genetic management of some herds shows a better structuring. The study contributes to the knowledge of the genetic structure and molecular characterization of the MM population, as well as to show how SSR can be used to establish genetic relationships among populations, as well as providing valuable information to monitor selection, becoming fundamental tools for basic research in animal genomics programs.

Key words: equideoculture, genetic diversity, polymorphism, SSR

1. INTRODUÇÃO

A raça Mangalarga Marchador (MM) em 2014 foi oficialmente declarada como raça nacional brasileira, sendo a mais numerosa do país, além de estar entre as maiores do planeta. Sua formação ocorreu a mais de 200 anos, no sudeste do Brasil, a partir do cruzamento de um garanhão Alter (Puro Sangue Lusitano - PSL) da Coudelaria de Alter do Chão, em Portugal, com éguas nativas brasileiras selecionados por criadores locais, provavelmente descendentes dos primeiros equinos trazidos da Península Ibérica (Meira, 2010). A raça PSL tem descendência do Andaluz Espanhol, cuja origem étnica provém dos equinos nativos das Penínsulas Berbere, Germânica e Ibérica, e de acordo com Nascimento (1999) e Assis et al. (2009), ocorreu uma mescla de outras raças como Árabe, Crioulo e Quarto de Milha que também contribuíram geneticamente para a formação do MM.

Os MM são animais considerados dóceis e resistentes, sendo utilizados principalmente na lida no campo, incluindo o pastoreio de bovinos, mas é crescente sua participação no esporte, em passeios equestres e no turismo ecológico (Gonçalves et al., 2012; Souza et al., 2017). Nos últimos anos contabilizou-se mais de 600 mil animais registrados, com 15 mil associados, 70 núcleos de criadores no país e quatro núcleos no exterior em países como a Argentina, Estados Unidos, Alemanha e Itália. Alguns exemplares dessa raça já foram exportados para a Alemanha, Argentina, Áustria, Bélgica, Canadá, Congo, EUA, França, Holanda, Israel, Itália, Peru, Portugal e Uruguai (Fonseca, 2018).

Na última década, com a disponibilidade de novas ferramentas para a seleção, as pesquisas na área de melhoramento genético envolvendo equinos, que eram escassas, foram se modificando e tornando cada vez mais crescentes (Pereira et al., 2017). Com o avanço de tecnologias que identificavam os marcadores moleculares, tornou-se possível a discriminação genotípica, o que permitiu o estudo da variação genética a nível de DNA. O estudo de caracterização genética, que verifica os níveis de variabilidade genética dentro e entre populações, além de fornecer toda a estrutura genética da mesma, utiliza-se desses marcadores para quantificar a diversidade genética em populações de animais domésticos (Tambasco et al., 2000). Os marcadores permitem a obtenção de estimativas não tendenciosas da variação, pois são independentes da sua função. Segundo Pinto (2008), os marcadores moleculares apresentam vantagens em relação aos outros tipos de marcadores, pois a sua variação não está limitada às regiões codantes do genoma, de modo que todas as categorias de variação, a princípio, podem ser detectadas.

Vários trabalhos de caracterização genética em equinos utilizando marcadores do tipo microssatélites (repetições de sequências simples – SSR) foram realizados utilizando populações de equinos Céticos-espanhóis (Cañon et al., 2000); equinos nativos do Japão (Tozaki et al., 2003); Lipizzaner (Achmam et al., 2004); Campolina (Ribeiro et al., 2005); cavalos selvagens (Veja-Pla et al., 2006); trotadores da Espanha (Azor et al., 2007); Mangalarga (Costa et al., 2009); Crioulo da Venezuela (Cotran et al., 2011); Mangalarga Marchador (Silva et al., 2012a; 2012b); Azorean (Lopes et al., 2015), Bardigiano (Stasiol et al., 2008) e equinos tradicionais do Butão (Dorji et al., 2018).

No entanto, uma análise mais abrangente do MM com raças ibéricas, berbere, inglesa e nativas brasileiras, utilizando marcadores moleculares SSR com análises primordiais para o devido fim, são escassos na literatura.

Deste modo, os objetivos neste estudo foram investigar a diversidade genética da raça MM e compará-la com raças relacionadas ou não à sua história, para obter uma compreensão mais clara de seu status atual e fornecer novas informações sobre as relações entre essas raças e subpopulações, estrutura genética e influências raciais na população de equinos Mangalarga Marchador do Brasil.

2. MATERIAL E MÉTODOS

2.1 Banco de dados de microssatélites

O banco de genotipagens de SSR foi cedido pela Associação Brasileira dos Criadores do Cavalo Mangalarga Marchador (ABCCMM), no qual continha genotipagens de 60.796 equinos da raça MM, obtidas nos anos de 2005 a 2016. Estes animais foram oriundos de diferentes haras localizados em todos os estados brasileiros.

A partir desse banco de genotipagens foi realizado uma filtragem, para que no final permanecessem dados de haras que possuíam maior número de animais genotipados e também dados de animais descendentes dos principais ancestrais da raça obtidos por meio do estudo de estrutura populacional com dados de *pedigree*. Dessa forma, o arquivo final utilizado neste estudo continha 2.829 animais MM de haras localizados em dez distintos estados brasileiros, sendo o estado de Minas Gerais o mais representativo, seguido de São Paulo, Bahia, Rio de Janeiro e Espírito Santo, estados estes que englobavam 83,57% da população total do MM no país.

Um grupo de seis raças de equinos da Espanha (Andaluz $n= 67$; AND), Portugal (Puro Sangue Lusitano $n= 43$; PSL), Inglaterra (Puro Sangue Inglês $n= 54$; PSI), Arábia Saudita (Árabe $n= 99$; ARA) e do Brasil (Campolina $n= 61$; CAM e Mangalarga $n= 40$; MAN) foi também incluído no estudo, totalizando em 3.193 equinos genotipados. Os dados de genotipagens destas seis foram cedidos pelo Instituto de Genética Veterinária da Faculdade de Ciências Veterinárias de Universidade Nacional de La Plata e pela Associação Brasileira dos Criadores do Cavalo da Raça Mangalarga.

2.2 Loci microssatélites

Para a avaliação da diversidade genética dentro e entre as raças de equinos, utilizou-se dados de genotipagens de 13 *loci* microssatélites altamente polimórficos que eram padronizado pela Sociedade Internacional de Genética Animal (ISAG). Os SSR utilizados foram distribuídos em 12 diferentes cromossomos do genoma equino (Tabela 1).

2.3 Análises estatísticas

Vários pacotes de programas foram utilizados para estimar a diversidade genética da população de Mangalarga Marchador com o uso de marcadores microssatélites. Alguns parâmetros refletem essa diversidade genética e foram computados neste trabalho como: o número de alelos absolutos (N_a); número de alelos raros (AR); número médio de alelos (N_{m_a}) por raça e *locus*, sendo o AR o número de alelos com frequência igual ou inferior a 0,05, e o N_{m_a} a média aritmética do número total de alelos observados em cada *locus*. O conteúdo de informação polimórfica (PIC) por *locus* foi obtido de acordo com Botstein et al. (1980), sendo este parâmetro um indicador da qualidade do marcador. As heterozigosidades esperadas (H_e) e observadas (H_o) por raça e *locus* foram calculadas com o programa Excel Microsatellite Toolkit (Park, 2008) seguindo a metodologia de Nei (1987), na qual a heterozigosidade é a probabilidade de um indivíduo ser heterozigoto no *locus* marcador.

Para a obtenção dos parâmetros F_{st} , a população total do MM ($n=2829$) foi dividida em subpopulações nos quais cada subpopulação representava um haras.

Análise de variância com estimativas de estatística F (Wright, 1965) também foi estimada para a população do Mangalarga Marchador. Os coeficientes F de Wright são como “coeficientes de endogamia”, entretanto, eles diferem de acordo com a população de referência. São todos calculados baseados nas heterozigosidades esperada e observada. O

parâmetro F_{is} é o coeficiente de endogamia de um indivíduo em relação à subpopulação da qual ele pertence, e foi computado por *locus* para medir a deficiência (valores positivos) ou excesso (valores negativos) de heterozigotos nas subpopulações; F_{st} é o coeficiente de endogamia da subpopulação em relação à população total da qual ela é parte, e mede o efeito genético da subdivisão total da população como a redução proporcional de toda heterozigose devido à variação nas frequências de alelos SSR entre subpopulações diferentes, e F_{it} é uma medida de heterozigosidade de um indivíduo em relação ao total da população.

As estatísticas F sobre todos os *loci*, assim como, as distâncias genéticas baseadas nos valores de F_{st} entre as raças foram obtidos com o programa GENEPOP 4.6 (Rousset, 2008). O dendrograma de análise de agrupamento e a análise de componentes principais (ACP) foram obtidos com o auxílio do Kit de Ferramentas de *ActionStat* (Equipe Estatcamp, 2014) no Excel baseado no método hierárquico e na matriz de distâncias genéticas, respectivamente.

O teste de equilíbrio de Hardy-Weinberg (EHW) entre os marcadores foi também calculado em todos os *loci* genômicos da raça MM, utilizando novamente o programa GENEPOP 4.6 através do método da cadeia de Markov para estimar os valores de p. O desequilíbrio de ligação (DL) entre todas as combinações de pares de *loci* foi estimado de acordo com Weir (1979), e testado pelo teste exato de Fisher, através do mesmo programa do EHW.

A diferenciação genética populacional foi observada através da diferenciação alélica e genotípica com a utilização de testes de probabilidade exata de Fisher (F) e algoritmo descrito por Raymond e Rousset (1995), e pela estimativa imparcial do valor p de um teste exato baseado em log-verossimilhança (G) para todos os *loci* analisados (Goudet et al., 1996), pelo GENEPOP.

A estrutura populacional da raça Mangalarga Marchador e demais raças foi investigada usando o algoritmo de agrupamento baseado em modelo Bayesiano, implementado no programa STRUCTURE 2.3.4 (Falush et al., 2003), que agrupa os animais de acordo com a combinação e distribuição dos alelos marcadores, exibindo de maneira gráfica os coeficientes individuais de associação de cada população. Neste programa foram realizados 5.000 burn-in, assumindo as frequências alélicas entre as populações sendo correlacionadas e usando um modelo de mistura no qual o indivíduo pode ter ancestralidade mista, além disso, 50.000 repetições foram executadas, com cada faixa de exploração de K abrangendo de 2 a 10. O nível mais alto de estrutura populacional foi identificado usando o método ΔK de Evanno et al. (2005), para encontrar o melhor número de clusters (K).

3. RESULTADOS

3.1 Diversidade genética

As localizações dos microssatélites no genoma equino, o conteúdo de informação polimórfica (PIC), o número de alelos absolutos (N_a), o número de alelos raros (AR), o número médio de alelos (N_{m_a}) por *locus* são apresentados na Tabela 1. Os 13 *loci* estudados para todas as raças tinham o PIC em média com valores acima do limite ($PIC > 0,5$), valor este necessário para sua adequação em estudos genéticos. Valores máximo de 0,82 (LEX03 e HSM3) e mínimo de 0,37 (HSM2) foram também observados, sendo que um dos valores máximos (LEX03) foi encontrado no Mangalarga Marchador (MM), e o mínimo nesta raça de 0,53 no *locus* HSM6. De modo geral, poucos *loci* foram menos informativos ($<0,50$), sendo estes observados nos marcadores ASB2 (0,49), HSM2 (0,37 e 0,43), HSM6 (0,44) e HTG4 (0,45).

O N_a detectado em todos os *locus* foi de 646 nas sete raças estudadas, com valores que variaram de três (HSM2) a 22 alelos (ABS17), sendo este valor máximo observado na raça MM, e o mínimo desta raça foi de sete alelos no *locus* HTG4. No MM o N_a total foi 141 alelos. Por raça, os maiores números médios de alelos (N_{m_a}) foram no MM (10,85), seguida pelo Campolina - CAM (7,38), Andaluz - AND (6,92), Puro Sangue Lusitano - PSL (6,77), Arabe - ARA (6,46), Mangalarga - MAN (6) e a menor no Puro Sangue Inglês - PSI (5,31). De modo geral, os AR prevaleceram no MM, com valor médio por *locus* de $2 \pm 2,58$.

Tabela 1. Descrição dos *loci* microssatélites das populações de equinos Mangalarga Marchador e das demais raças brasileiras, portuguesa, espanhola, árabe e inglesa.

LOCUS	CRO	MM		MAN			CAM		PSL		AND			ARA			PSI					
		PIC	Na	AR	PIC	Na	AR	PIC	Na	AR	PIC	Na	AR	PIC	Na	AR	PIC	Na	AR	PIC	Na	AR
AHT4	24	0,74	8	1	0,70	6	-	0,73	7	-	0,76	7	-	0,68	7	-	0,68	6	-	0,68	4	-
AHT5	8	0,71	9	2	0,70	5	-	0,67	7	-	0,73	6	-	0,65	7	-	0,62	6	1	0,63	4	-
ASB2	15	0,80	15	1	0,63	6	-	0,76	7	-	0,76	8	-	0,78	9	-	0,49	8	1	0,77	7	-
ASB17	2	0,79	22	10	0,79	8	-	0,80	12	-	0,78	11	-	0,79	11	-	0,61	5	-	0,72	6	-
ASB23	3	0,74	9	1	0,75	6	-	0,69	7	-	0,72	5	-	0,73	6	-	0,61	6	-	0,74	5	-
HMS2	10	0,58	11	3	0,37	3	-	0,70	8	-	0,76	7	-	0,60	9	-	0,71	7	-	0,43	5	-
HMS3	9	0,70	9	2	0,71	6	-	0,59	7	-	0,82	7	-	0,62	7	-	0,67	6	-	0,64	6	-
HMS6	4	0,53	10	3	0,59	5	-	0,44	4	-	0,61	4	-	0,69	5	-	0,51	5	-	0,56	5	-
HMS7	1	0,71	8	1	0,68	7	-	0,74	6	-	0,66	6	-	0,61	7	-	0,75	6	-	0,74	5	-
HTG4	9	0,59	7	1	0,71	5	-	0,62	5	-	0,59	5	-	0,58	4	-	0,58	4	-	0,45	4	-
HTG10	21	0,69	10	-	0,73	8	-	0,67	9	-	0,75	8	-	0,74	7	-	0,73	9	1	0,80	7	-
LEX03	X	0,82	11	-	0,68	7	-	0,73	8	-	0,80	7	-	0,78	7	-	0,75	9	-	0,70	7	-
VHL20	30	0,76	12	1	0,61	6	-	0,81	9	-	0,81	7	-	0,66	4	-	0,73	7	-	0,70	4	-
Média/Nm _a	*	0,73	10,85	2	0,67	6	-	0,69	7,38	-	0,73	6,77	-	0,69	6,92	-	0,65	6,46	0,23	0,66	5,31	-
DP	*	0,07	3,93	2,58	0,10	1,35	-	0,10	1,98	-	0,07	1,74	-	0,07	1,98	-	0,09	1,51	0,44	0,12	1,18	-

DP= desvio padrão; CRO= cromossomo; MM= Mangalarga Marchador; MAN= Mangalarga; CAM= Campolina; PSL= Puro Sangue Lusitano; AND= Andaluz; ARA= Árabe; PSI = Puro Sangue Inglês; PIC = Conteúdo de Informação Polimórfica; Na= número de alelos absolutos; AR= número de alelos raros; Nm_a= número médio de alelos.

De acordo com Nei (1987) foram calculados os valores médios de heterozigosidades esperada ($H_e = 0,7364 \pm 0,030$) e observada ($H_o = 0,7354 \pm 0,1202$) para todas as raças. Todas as raças obtiveram valores acima de 0,6926 de H_e , e o valor máximo de 0,7775 foi encontrado na raça PSL. O valor mínimo de H_o foi obtido na raça ARA (0,6466) e o valor máximo de H_o no MAN (0,999). As diferenças entre H_o e H_e em cada raça foram de -0,2306 no MAN; 0,0282 no AND; 0,0386 no PSL; 0,0398 no MM; 0,0413 no CAM; 0,0432 no PSI e 0,0460 no ARA. As H_e na raça MM variaram de 0,5767 (HSM6) a 0,8450 (LEX03) e as H_o nesta mesma raça variaram de 0,5331 (HSM6) a 0,7949 (ASB17), considerando esses parâmetros por *locus* (Tabela 2).

3.2 Estatística F de Wright, equilíbrio de Hardy-Weinberg e diferenciação genética

As estatísticas F de Wright (1965) nas subpopulações (Fis) e considerando a população total (Fit) do MM foram calculados como excesso de heterozigosidade (-) ou deficiência (+) e são apresentadas na Tabela 2. A média de Fis no MM foi de $-0,0195 \pm 0,0254$, oscilando de 0,0232 (ATH5) a 0,1490 (LEX03), e a média de Fit foi de $0,0566 \pm 0,0341$, oscilando de 0,0053 (HTG10) a -0,0538 (HTG4). A média de Fit indicou uma pequena deficiência de heterozigosidade em algumas regiões do genoma do MM, porém o valor negativo de Fis denota pequeno excesso de heterozigose nas subpopulações do MM.

Os parâmetros F_{st} por *locus* do MM são demonstradas na tabela 2. Um grau moderado de diferenciação genética ($F_{st} = 0,0748 \pm 0,0176$) revela que apenas 7,48% da variação genética foi devida a diferenças entre as subpopulações, e 92,52% poderiam ter sido atribuídos à variabilidade genética dentro das subpopulações, ou seja, entre os indivíduos.

Tabela 2. Resumo da estatística de heterozigosidade, F estatístico, equilíbrio de Hardy-Weinberg e diferenciação genética por *locus* SSR de equinos da raça Mangalarga Marchador.

<i>Locus</i>	Heterozigosidade		F estatístico			Equilíbrio de Hardy Weinberg	Diferenciação genética	
	He	Ho	Fis	Fst	Fit	EHW	<i>p</i> -valor	
							F	G
AHT4	0,7671	0,7386	-0,0348	0,0719	0,0396	0,0000	0,0000	0,0000
ATH5	0,7450	0,7292	-0,0378	0,0588	0,0232	0,0000	0,0000	0,0000
ASB17	0,8145	0,7949	-0,0440	0,0674	0,0264	0,0013	0,0000	0,0000
ASB2	0,8193	0,7822	-0,0074	0,0542	0,0472	0,0000	0,0000	0,0000
ASB23	0,7766	0,7445	-0,0310	0,0726	0,0438	0,0104	0,0000	0,0000
HMS2	0,6180	0,5638	-0,0184	0,1077	0,0913	0,0000	0,0000	0,0000
HSM3	0,7445	0,7041	-0,0234	0,0784	0,0569	0,0010	0,0000	0,0000
HSM6	0,5767	0,5331	-0,0012	0,0794	0,0783	0,0000	0,0000	0,0000
HSM7	0,7520	0,7110	-0,0205	0,0759	0,0570	0,0000	0,0000	0,0000
HTG10	0,7318	0,6920	0,0053	0,0531	0,0581	0,0274	0,0000	0,0000
HTG4	0,6488	0,6305	-0,0538	0,0804	0,0309	0,0023	0,0000	0,0000
LEX03	0,8450	0,7226	0,0441	0,1098	0,1490	0,0000	0,0000	0,0000
VHL20	0,7831	0,7582	-0,0308	0,0628	0,0340	0,0000	0,0000	0,0000
Média	0,7402	0,7003	-0,0195	0,0748	0,0566	--	--	--
DP	0,0800	0,0793	0,0254	0,0176	0,0341	--	--	--

He = heterozigose esperada; Ho = heterozigose observada; Fis e Fit = coeficientes de endogamia de Wright; Fst = grau de diferenciação; EHW = equilíbrio de Hardy-Weinberg; (F) = Teste de Fisher; (G) = teste de verossimilhança; DP = desvio padrão.

O teste de equilíbrio de Hardy-Weinberg mostrou desvios significativos para todos os *loci* na raça Mangalarga Marchador, e diferenciação populacional alélica e genotípica foram também observadas em todos os *loci* microssatélites a partir dos testes exatos de Fisher (F) e de log-verossimilhança (G), apresentados na Tabela 2.

3.3 Desequilíbrio de ligação

O resultado do teste de associação entre os pares de *loci* do MM, descrito também como um teste do desequilíbrio de ligação composto, é observado na Tabela 3. Significativos valores de desequilíbrios de ligação na maioria dos pares de marcadores foram encontrados entre os alelos SSR, e em todas as comparações 60,26% mostraram forte DL.

Tabela 3. Informações sobre o desequilíbrio de ligação entre os alelos de marcadores SSR de equinos da raça Mangalarga Marchador: somente valores de $P < 0,01$ são mostrados.

Par de locus		RV	P-valor	Par de locus		RV	P-valor
ATH4	ATH5	Infinito	<0,0001	HSM2	HSM7	Infinito	<0,0001
ATH4	ASB17	Infinito	<0,0001	HSM3	HSM7	Infinito	<0,0001
ATH5	ASB17	Infinito	<0,0001	HSM6	HSM7	Infinito	<0,0001
ATH4	ASB2	Infinito	<0,0001	HSM3	ASB2	Infinito	<0,0001
ATH5	ASB2	Infinito	<0,0001	ATH5	HTG10	Infinito	<0,0001
ASB17	ASB2	Infinito	<0,0001	ASB17	HTG10	13,7370	0,008183
ATH4	ASB23	Infinito	<0,0001	ASB17	VHL20	23,3268	0,000109
ATH5	ASB23	Infinito	<0,0001	ASB23	HTG10	17,9742	0,001248
HSM6	VHL20	19,28873	0,00069	HSM2	HTG10	Infinito	<0,0001
ASB2	ASB23	Infinito	<0,0001	HSM3	HTG10	Infinito	<0,0001
ATH4	HSM2	Infinito	<0,0001	ATH4	VHL20	Infinito	<0,0001
ATH5	HSM2	Infinito	<0,0001	ATH5	VHL20	Infinito	<0,0001
ASB17	HSM2	Infinito	<0,0001	ATH4	HTG4	29,2121	0,000007
ASB2	HSM2	21,8024	0,000219	ATH5	HTG4	Infinito	<0,0001
HSM2	VHL20	Infinito	<0,0001	ASB17	HTG4	19,7077	0,00057
ATH4	HSM3	21,8744	0,000212	ASB2	HTG4	Infinito	<0,0001
ATH5	HSM3	Infinito	<0,0001	ASB23	HTG4	Infinito	<0,0001
ASB17	HSM3	17,7137	0,001404	HSM2	HTG4	Infinito	<0,0001
ASB2	HSM3	Infinito	<0,0001	HSM3	HTG4	Infinito	<0,0001
ASB23	HSM3	Infinito	<0,0001	HSM6	HTG4	Infinito	<0,0001
HSM2	HSM3	Infinito	<0,0001	HSM7	HTG4	Infinito	<0,0001
ATH4	HSM6	20,5314	0,000392	HTG10	HTG4	Infinito	<0,0001
ATH5	HSM6	20,6450	0,000372	HTG10	VHL20	Infinito	<0,0001
ASB17	HSM6	19,9375	0,000514	ATH5	LEX03	Infinito	<0,0001
ASB2	HSM6	Infinito	<0,0001	LEX03	VHL20	Infinito	<0,0001
ASB23	HSM6	29,1006	0,000007	HSM3	VHL20	15,52831	0,003722
HSM2	HSM6	Infinito	<0,0001	ASB23	LEX03	Infinito	<0,0001
HSM3	HSM6	Infinito	<0,0001	HSM2	LEX03	15,4155	0,003913
ATH4	HSM7	Infinito	<0,0001	HSM7	VHL20	Infinito	<0,0001
ATH5	HSM7	Infinito	<0,0001	HSM6	LEX03	15,0163	0,004668
ASB17	HSM7	16,6847	0,002225	HSM7	LEX03	Infinito	<0,0001
ASB2	HSM7	Infinito	<0,0001	ASB23	VHL20	Infinito	<0,0001
ASB23	HSM7	Infinito	<0,0001	HTG4	LEX03	20,4893	0,0004

RV= razão de verossimilhança.

3.4 Distância genética

As distâncias genéticas baseadas em F_{st} entre todas as raças equinas analisadas estão apresentadas na Tabela 4.

Tabela 4. Distâncias genéticas das populações de equinos Mangalarga Marchador e das demais raças brasileiras, portuguesa, espanhola, árabe e inglesa.

RAÇAS	ARA	PSI	AND	PSL	CAM	MAN	MM
ARA	0						
PSI	0,1126	0					
AND	0,1202	0,1121	0				
PSL	0,0955	0,0993	0,0336	0			
CAM	0,1445	0,1564	0,1013	0,0863	0		
MAN	0,1305	0,1057	0,0843	0,0686	0,1051	0	
MM	0,1119	0,1136	0,1058	0,0801	0,0679	0,0458	0

ARA= Árabe; PSI= Puro Sangue Inglês; AND= Andaluz; PSL= Puro Sangue Lusitano; CAM= Campolina; MAN= Mangalarga; MM= Mangalarga Marchador.

A maior distância genética foi observada entre o MAN e PSI (0,1564), e a menor entre MM e MAN (0,0458). De modo geral, as menores distâncias foram observadas entre as raças brasileiras e destas com as raças ibéricas, e as maiores distâncias entre as raças brasileiras e ibéricas com as raças árabe e inglesa. A melhor visualização das distâncias genética entre as raças podem ser observada no dendrograma da Figura 1.

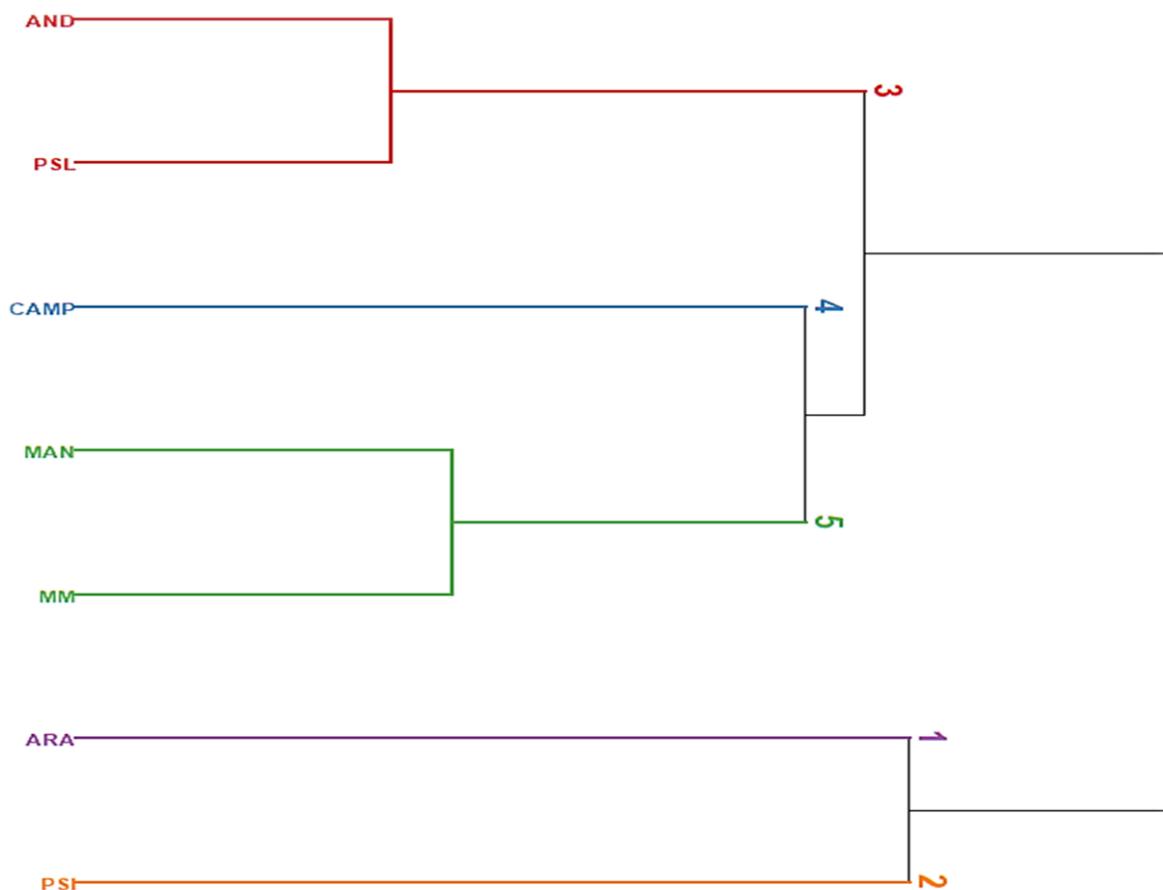


Figura 1. Dendrograma de análise de agrupamento que mostra o nível de relações genéticas entre as sete raças equinas estudadas. AND = Andaluz; ARA = Árabe; CAM= Campolina; MAN = Mangalarga; MM = Mangalarga Marchador; PSI = Puro Sangue Inglês; PSL = Puro Sangue Lusitano.

Pode-se observar a separação entre o grupo formado pelas raças ARA e PSI das demais raças analisadas, no qual não compartilham ancestral comum mais recente. Nota-se também que todas as raças brasileiras e as ibéricas, compartilham de um ancestral comum e portanto, são mais próximas entre si, assim como entre as raças brasileiras, principalmente entre o MAN e MM.

3.5 Componentes principais

Os resultados obtidos nos pares de F_{st} foram comprovados utilizando a análise de componentes principais. Esta análise, baseada nas estatísticas F_{st} , ilustra graficamente as relações entre as raças, condensando as informações de um grande número de alelos e *loci* em algumas variáveis sintéticas, demonstrada na Figura 2. Os dois primeiros componentes

explicaram 40,40% e 26,40%, respectivamente, da variação genética total observada entre raças. O primeiro grupo de raças formado pelo Mangalarga Marchador, Mangalarga e Campolina foram mais próximos entre si, assim como, o segundo grupo, com as raças Andaluz e Puro Sangue Lusitano, e também o terceiro grupo como Puro Sangue Inglês e o Árabe. Esses três grupos, quando comparados, foram mais divergentes entre si, mas nota-se um menor grau de distanciamento do primeiro grupo (MM, MAN e CAMP) com o segundo (AND e PSL), se comparada com o terceiro grupo (ARA e PSI).

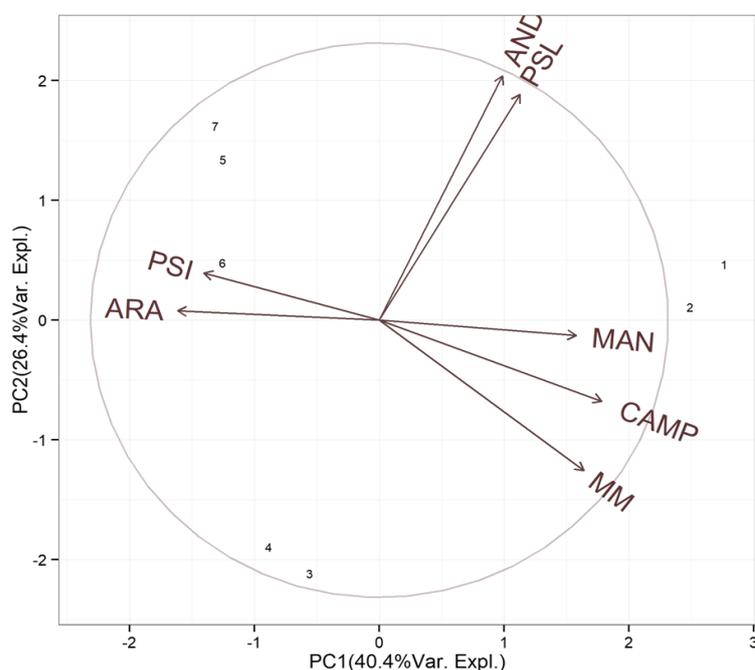


Figura 2. Análise de componentes principais das sete raças equinas estudadas baseadas em valores de F_{st} de 13 microssatélites. AND = Andaluz; ARA = Árabe; CAM= Campolina; MAN = Mangalarga; MM = Mangalarga Marchador; PSI = Puro Sangue Inglês; PSL = Puro Sangue Lusitano; PC1 e PC2 = do inglês *Principal Component*.

3.6 Estrutura populacional

O estudo da estrutura populacional ou análise de agrupamento do Mangalarga Marchador com as demais raças equinas analisadas através da estimação de valores ΔK inferiu $k=5$ como o agrupamento genômico mais provável para as sete raças, com cinco principais alelos marcadores distribuídos pelas populações, como evidenciado na Figura 3. De modo geral, ao observar a raça MM, a sua estrutura populacional baseada no programa STRUCTURE sugere uma falta de estruturação genética bem definida, assim como nas raças MAN e CAM (Figura 3a). Ao contrário das raças ARA e PSI que foram entidade genéticas bem mais definidas.

Entretanto, alguns haras da raça MM foram mais geneticamente definidos (Haras específico - HE1 a HE4) como demonstrado na Figura 4b, apesar da grande maioria das fazendas seguirem o padrão Haras comum (HC). Pelas colorações (verde, amarelo, rosa e azul) evidencia-se o compartilhamento de alelos comuns, principalmente entre CAM, MAN e MM, e em menor grau a coloração vermelha, que foram alelos provenientes de ancestrais destas raças, como o PSL e AND, comum às três raças brasileiras, mas também das raças ARA e PSI.

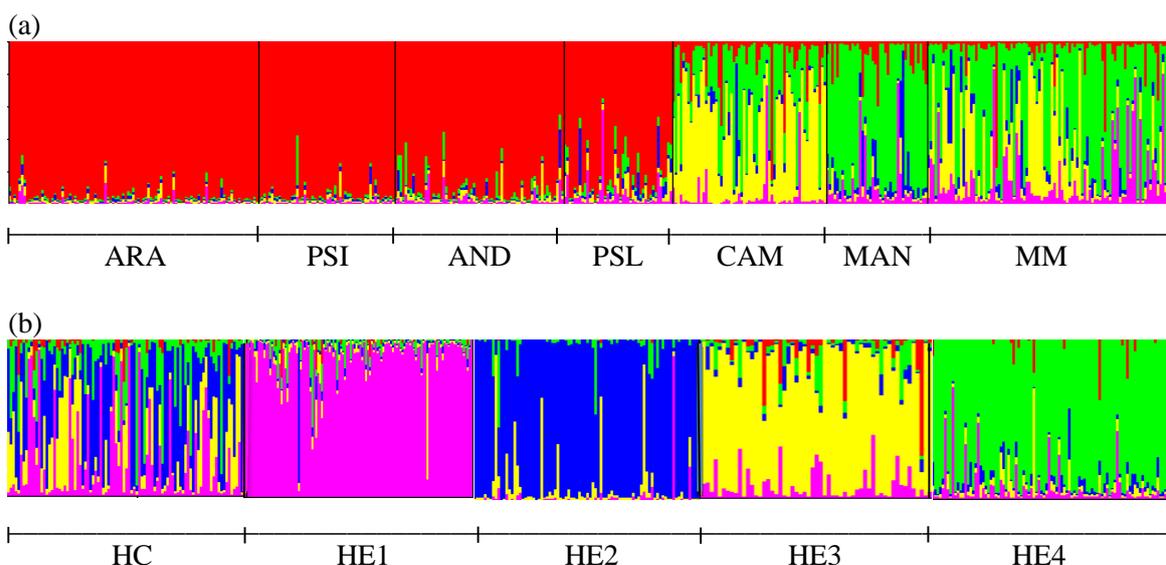


Figura 3. Estrutura populacional das raças estudadas mostrando as relações entre os biótipos. (a) AND = Andaluz; ARA = Árabe; CAM= Campolina; MAN = Mangalarga; MM = Mangalarga Marchador; PSI = Puro Sangue Inglês; PSL = Puro Sangue Lusitano; (b) HC = haras comum do MM; HE1 a HE4 = haras específicos de 1 a 4.

4. DISCUSSÃO

O Mangalarga Marchador (MM) representa uma parte viva do patrimônio cultural brasileiro, principalmente mineiro, uma vez que Minas Gerais é seu estado federativo de origem. Além de motivações científicas, há necessidades histórica, cultural, econômica e étnica para o monitoramento de criação da raça e até mesmo de sua conservação. No contexto de estruturação populacional ou de conservação, a caracterização genética é o primeiro passo para o desenvolvimento de estratégias de gestão genética adequadas e pode ter implicações para futuras estratégias de melhoramento (Solis et al., 2005), e a informação molecular é crucial tanto para detectar e manter a diversidade genética como para a prevenção da perda ou introgressão indesejada de alelos dentro de uma raça.

Vários estudos moleculares têm avaliado o painel de marcadores microsatélites recomendados pela ISAG, para identificação individual ou racial e até mesmo de parentescos entre as raças equinas existentes. Entretanto, o conjunto de marcadores utilizado varia bastante entre as publicações, uma vez que pode incluir outros marcadores, o que dificulta ainda mais a comparação direta entre estes trabalhos (Behl et al., 2006; Luís et al., 2007; Choi et al., 2008; Takasu et al., 2012). Em função disto, serão discutidos preferencialmente artigos com maior número de marcadores semelhantes ao presente estudo.

4.1 Diversidade genética

Segundo Botstein et al. (1980), valores acima de 50% no parâmetro PIC são considerados bastante informativos, o que confere maior confiabilidade na seleção dos mesmos, pois demonstra o quão informativo é o marcador. Os valores de PIC em todos os 13 *loci* de microsatélites, utilizados neste estudo, demonstraram que os marcadores foram adequados para a avaliação genética da raça MM e dos demais grupos genéticos (Tabela 1). Resultados inferiores de PIC foram relatados por Assis et al., 2009 (0,66) e por Silva et al., 2012a (0,68), ambos trabalhos foram com a raça MM e utilizando SSR comuns ao deste estudo. Luís et al. (2002) encontraram resultados semelhantes quando utilizaram dados de animais das raças Garrano (0,73) e PSL (0,74).

A diversidade genética das populações de animais corresponde à variabilidade de alelos presentes nelas, que pode ser estimada a partir do número e média de alelos por *locus*, presença de alelos raros e de parâmetros de heterozigosidades. Neste estudo, os *loci* analisados detectaram de moderado a alto polimorfismos nas sete raças de equinos analisadas (Tabela 1). De acordo com Barker (1994), para estudos de diversidade genética a sugestão é de, no mínimo, quatro diferentes alelos por *locus* para que o marcador seja utilizado para este devido fim. Em particular, o alto nível de polimorfismos ($Nm_a = 10,85$) observado no MM é comparável às outras raças equinas como no equino Miyako (Senju et al, 2017) e ao próprio MM (Silva et al. 2012a) com Nm_a de 9,2; em um estudo com 12 SSR iguais ao deste trabalho e analisando 6.092 equinos; superior a todas as 50 raças equinas estudadas por Cortés et al. (2016) e da raça Nordeste (Pires et al., 2014), com marcadores também similares ao deste trabalho.

A diversidade alélica e a maior ocorrência de AR em praticamente todos os *loci* do MM (Tabela 1) podem estar em função do tamanho populacional da raça e da amostra, e também devido a uma consequência de possíveis cruzamentos com outras raças, de acasalamentos

intrarraciais entre indivíduos poucos ou não aparentados, ou mesmo proveniente da mistura de garanhões PSL e de éguas nativas brasileiras utilizados na formação da raça a mais de 200 anos atrás. No trabalho realizado com o MM por Silva et al. (2012a), foi também encontrado AR alto, principalmente no *locus* ASB17 (11 alelos), assim como o Na neste mesmo *locus* (20 alelos), e ao considerar todos os SSR foi estimado 129 alelos. Valores superiores de AR e de Na no mesmo *locus* ASB17 e nos demais marcadores foram também encontrados neste presente estudo. Em outras raças, o ASB17 também estava entre os marcadores mais informativos (Lee e Cho, 2006; Choi, et al., 2008; Ling, et al., 2010).

O parâmetro mais importante para mensurar a diversidade genética é a heterozigosidade esperada, também conhecida como diversidade de genes ou alélica (Nei,1987), que corresponde à quantidade de heterozigosidade presente em uma população, quando ocorre predominantemente acasalamentos ao acaso (Hedrick, 2005). Esse parâmetro fornece informações sobre a estrutura e a história de uma população, sendo que valores altos significam que existe maior diversidade de alelos naquela população (Butler, 2005; Nei, 1978). De acordo com Ott (1992), um *locus* considerado altamente polimórfico é aquele que possui heterozigosidade observada igual ou superior a 0,70.

De modo geral, os valores de H_e e H_o sugerem uma elevada diversidade genética em todas as raças analisadas neste estudo. Em particular, no MM, estes valores em média, sugeriram pequeno excesso de heterozigotos, comprovados por valores de Fis, discutidos posteriormente (Tabela 2). H_e alta é comumente encontrada em raças equinas mais recentes, pois ainda estão em fase de fixação como um grupo genético definido, e provavelmente ocorrendo ainda, algum fluxo gênico na população, em virtude de possíveis cruzamentos com outras raças. A raça Mangalarga Marchador em comparação com a maioria de outras raças equinas da literatura, mostrou-se também com alto nível de diversidade genética em relação aos parâmetros de H_e , H_o e N_{m_a} como observado em Cañon et al. (2000), Cunningham et al. (2001), Juras et al. (2003), Achmann et al. (2004), Solis et al. (2005), Luís et al. (2007), Berber et al. (2014), Stasiol et al. (2008), embora os valores deste estudo não sejam diretamente comparáveis com os citados acima, devido às diferenças de população e de conjunto de microssatélites utilizados.

Na comparação com trabalhos exclusivos utilizando animais da raça MM, valor superior de H_e (0,839) foi encontrado por de Silva Filho et al. (2007) ao analisarem três marcadores e 79 equinos, valor menor de H_e (0,637) foi relatado por Assis et al. (2009) com um painel de dez microssatélites em 115 indivíduos, e valor semelhante por Silva et al. (2012a) com H_e de 0,70 com 6.092 animais e 14 SSR. Esses resultados evidenciam que não somente o número de

marcadores moleculares influenciam na estimativa da diversidade genética, mas também o tamanho da amostra, e que não há uma correlação clara entre estes parâmetros (N_a , N_{m_a} , AR , H_e e H_o) com esses dois fatores mencionados.

De acordo com Notter (1999), a diversidade genética de animais de interesse zootécnico é necessária para atender às necessidades atuais de produção em vários ambientes, para permitir o aprimoramento genético e facilitar a rápida adaptação a objetivos da reprodução. Em geral, a alta diversidade genética da população de equinos MM, refletida também pela maioria dos *locus* (Tabela 2), é otimista para a resiliência a mudanças ambientais no futuro da raça. Contudo, é sempre importante delinear estratégias de reprodução capazes de preservar a variabilidade genética para evitar efeitos futuros de depressão por endogamia.

4.2 Estatística *F* de Wright

Os parâmetros de Wright são de extrema importância para estudos genéticos populacionais, sendo o F_{is} indicador da proporção da variância contida em subpopulações ou da variabilidade genética intrarracial, e F_{it} indica o coeficiente de endogamia para o total da população. O valor médio de F_{is} , obtida neste trabalho (-0,0195), demonstra que há excesso de heterozigotos, mas em menor grau. F_{is} negativos podem ser explicados pelo efeito de Wahlund na qual a mistura de indivíduos de diferentes populações com distintas frequências alélicas pode promover alterações de heterozigosidade na mesma. Esse ligeiro excesso de heterozigotos indicam que as subpopulações, ao longo dos tempos, nos mais diversos ambientes, conseguiram desenvolver alguns alelos próprios, conferindo características singulares e peculiares, além de que, os criadores da raça em sua maioria se atentaram com relação aos acasalamentos de indivíduos aparentados.

O valor de F_{it} , ao considerar toda a população do MM (0,0566), denotou pequena deficiência de heterozigosidade, o que pode ser devido a diversos fatores como nível de intensidade de seleção mais elevada realizadas em alguns haras, acasalamentos não aleatórios, utilização de técnicas de reprodução assistida que podem ter um forte impacto negativo sobre o tamanho efetivo da população, presença de alelos nulos nos *loci* (alelos que não amplificaram) levando a subestimação de heterozigotos. Resultados diferentes foram estimados por Assis et al. (2009), que ao estudarem o MM em três regiões do estado de Minas Gérias, obtiveram valores de F_{is} (0,1238) e F_{it} (0,1731) denotando endogamia nas três subpopulações e déficit de heterozigotos na população total, porém com somente um SSR semelhante ao deste estudo, e por Silva et al (2012a), com valores de F_{is} (0,0124) e F_{it}

(0,0183) e 12 marcadores iguais ao presente trabalho, porém com subdivisões da população em 15 estados federativos brasileiros, fatores estes que explicam a diferença entre os resultados destes trabalhos citados com o presente estudo. Criscione et al. (2015) obtiveram valor semelhante de Fis (-0,013) e distinto de Fit (0,074) com equinos Salernitano Italiano e com marcadores semelhantes, assim como Dorji et al. (2018), com equinos nativo do Butão, que estimaram valores distintos de Fis (0,019) e Fit (0,025), denotando déficit de heterozigotos nas subpopulações e na população total.

O parâmetro F_{st} indica o nível de diferenciação genética médio que se manifesta nas subpopulações definidas ou entre raças, e é a fração da diversidade total que é causada pela diferença entre populações ou mesmo pela proporção da variância genética total devida à subdivisão. De acordo com Wright (1965), valores de F_{st} entre 0,05 e 0,15 indicam diferenciação moderada, entre 0,15 e 0,25 alta e acima de 0,25 muito alta. Assim, usando o guia de interpretação de Wright, o resultado deste trabalho (0,0748) indica moderada diferenciação entre as subpopulações do MM, o que evidencia um fluxo gênico controlado entre elas, e apesar destas subpopulações serem próximas geneticamente, conseguiram de algum modo, mesmo que em grau moderado, fixar “ecotipos” característicos, de acordo com as condições ambientais de cada local onde se desenvolveram, e devido a estratégias de manejo implementados em cada haras. Além desses fatores, o efeito de fundadores, uso preferencial de alguns reprodutores, a deriva genética e a seleção artificial também estão entre os fatores que podem ter contribuído para a diferenciação genética dentro da raça. Essa diferenciação entre subpopulações é também comprovada pela significância em todos os *loci* obtidos nas estatísticas F e G (Tabela 2).

Assis et al. (2009) também encontraram valor moderado ($F_{st}= 0,0562$) entre três subpopulações do MM, assim como nos trabalhos de Silva et al. (2012a) ($F_{st}= 0,006$) em subpopulações de MM. Silva et al. (2012b) encontraram valores de F_{st} entre 0,074 e 0,16 ao trabalharem com raças criadas no Brasil como a Campeira, Lavradeira, Pantaneira, Baixadeiro, MM, ARA e PSI. Diversos outros trabalhos envolvendo várias raças encontraram valores de F_{st} moderados (Barcaccia et al., 2013 – com a raça Lipizzan; Berber et al., 2014 - equinos criados na Argélia como Barb, Árabe-Barb, Árabe, Puro Sangue Inglês e Trotador Francês; Behl et al., 2007 - raças indianas como o Marwari, Spiti, Bhutia, Manipuri e Zanskari, Dorji et al., 2018 - equinos nativos do Butão). Alguns trabalhos encontraram valores de F_{st} maiores do que este presente estudo como nos trabalhos de Gastaldo et al., 2017 - PSI, Quarto de Milha, Crioulo Uruguaiano, Árabe e Andaluz; Costa et al., 2009 - algumas raças do Brasil como a Marajoara, Puruca, PSI, MM, Brasileiro de Hipismo e Quarto

de Milha, Pires et al. 2014 – com 15 raças equinas entre elas o MM, MAN, CAMP, Nordestino, Pantaneiro, Crioulo Brasileiros, PSL, AND, Garrano, Sorraia e outras raças sul-americanas. Todos estes trabalhos citados acima foram com SSR semelhantes ao deste estudo. A diferença observada, para este parâmetro, está provavelmente relacionada ao fato de que esses estudos investigaram raças de diferentes origens e de funções.

A estrutura populacional e fluxo gênico moderado observado entre as subpopulações do MM sugere, além de todos os fatos mencionados, a presença, ainda que pouco, de cruzamentos com outras raças equinas. No entanto, uma alta diversidade genética observada fornece uma perspectiva otimista para a raça a partir da manutenção e aprimoramento de intervenções de manejo apropriadas.

4.3 Equilíbrio de Hardy-Weinberg

Desvios no equilíbrio de Hardy-Weinberg (EHW) foram detectados em todos os *loci* analisados do MM. Esses desvios podem evidenciar a presença de eventos como cruzamento com outras raças, acasalamento preferencial não aleatório, fluxo gênico, uso preferencial de alguns reprodutores, endogamia ou deriva genética. Uma possível ocorrência de alelos nulos poderia também ter levado a uma falsa observação de homozigotos na população, o que explica mais desvios de EHW, segundo Shahsavarani e Rahimi-Mianji (2010). Silva et al. (2012b) utilizaram SSR semelhantes ao deste estudo, e detectaram que as populações das raças Campeira, Lavradeira, Pantaneira, Baixadeiro, MM, ARA e PSI, globalmente também não se encontravam na condição de equilíbrio de Hardy-Weinberg. No trabalho de Silva et al. (2012a), com apenas MM, também encontraram desvios de equilíbrio, porém somente nos marcadores ATH4, ASB2, HMS2, HSM3, HSM7, e argumentaram que os desvios podem ser devido ao efeito de Wahlund e as diferenças nas frequências alélicas entre essas subpopulações podem ser observadas pela deficiência de heterozigotos. Desvios significativos também foram observados nos *loci* HSM7, HTG4 e VHL20 em 79 animais MM de um único haras no trabalho de Silva Filho et al. (2007).

4.4 Desequilíbrio de ligação

No que diz respeito ao desequilíbrio de ligação (DL), significativos DL foram encontrados entre os pares de marcadores microsatélites do MM. Possivelmente, foram identificadas regiões cromossômicas não homólogas portadoras de *loci* de marcadores

afetados por forte DL, o que podem estar associadas a genes influenciados pela seleção ou por efeito fundador e ter desempenhado um papel importante na evolução da população. Deste modo, além da seleção, fatores de evolução, tais como uniões não aleatórias, mutações, e especialmente, fluxo gênico entre populações e efeitos de deriva genética podem ter afetado as frequências haplotípicas e assim gerar DL entre esses marcadores. DL fortes também foram encontrado por Barcaccia et al. (2013), na raça Lipizzan Italiano, com SSR semelhantes ao deste estudo. Costa et al. (2009) encontram seus marcadores em equilíbrio nas raças equinas como a Pura Raça Espanhola, Menorquina, Potoka, ARA, Puruca, Pantaneiro, Losino, PSI e PSL. Esse parâmetro não foi estimado em populações de MM em outros trabalhos já existentes na literatura.

4.5 Distância genética e componentes principais

A divergência genética entre as raças estudadas foi avaliada por diferentes abordagens (distâncias genéticas, componentes principais e método de agrupamento). Todas essas três estimativas clássicas baseadas em relações genéticas deram resultados semelhantes. Há uma diferenciação genética entre equinos árabes e inglês com as brasileiras e ibéricas. O dendrograma demonstrou uma clara subdivisão dessas raças, e o MM foi mais relacionado com a raça MAN (Figuras 1 e 2), sugerindo que as duas raças repartem o mesmo ou grande parte do mesmo conjunto gênico. Em geral, as raças brasileiras foram mais próximas entre si e também com as raças ibéricas, e todas estas distintas do PSI e ARA. Este resultado pode ser explicado pela influência da raça PSL na formação original das raças brasileiras.

A proximidade genética do MM com CAM, e principalmente com MAN também foi demonstrada pela abordagem de agrupamento bayesiano, que fornece informações mais precisas sobre as relações entre raças. Sabe-se que a formação da maioria das raças equinas brasileiras tem como principais raças formadoras as ibéricas como PSL, Sorraia, Garrano e AND, e também de algumas raças berberes trazidas também pelos colonizadores, fato este descrito por diversos autores com Chieffi (1950), Goulart (1964), Costa, Val e Leite (1974), Beck (1985), Monteiro (1986), Lima, Shirota e Barros (2006), Hendricks (2007), Pires et al. (2014), entre outros. Esta circunstância explica a aproximação entre as raças brasileiras (MM, MAN e CAM) e destas com as raças ibéricas (AND e PSL), observadas nas Figuras 1 e 2.

Pires et al. (2014), em seu trabalho adicionou novas raças equinas a partir de um conjunto de dados de outras raças obtidas por Conant et al. (2011), totalizando em 67 raças distintas, incluindo raças selvagens. A partir disso, construiu-se uma árvore de distâncias genéticas

entre essas raças e foi claro a formação de três grupos bem definidos, o dos pôneis e equinos de tração pesada, o das raças sul-americanas de origem ibérica e o das raças do Velho Mundo e norte-americanas. Como um dos resultados, obteve-se o MM e MAN em um mesmo subgrupo, muitos próximos também do CAM, e estes próximos aos Sorraia, PSL, AND e Garrano. Raças como o Crioulo Brasileiro, Argentino, Chileno e Venezuelano, Passo Peruano, Passo Fino Colombiano e de Porto Rico, e outras raças brasileiras como Nordesteiro e Pantaneiro foram também próximos entre si, participantes de um mesmo grupo a partir de descendentes em comum. Monteiro (1986), relata que outras raças foram trazidas da Argentina e do Paraguai pelos jesuítas que transitavam pelo Brasil, e outros equinos trazidos da América Central passavam pela América do Sul, alcançavam o norte da Argentina e facilmente entraram no território brasileiro, o que justifica a proximidade das raças brasileiras com outras raças da América Central e do Sul.

Lippi e Mortari (2003) compararam a raça MM a outra raça intimamente relacionada, a Mangalarga (MAN), porém com pequenas diferenças de características da marcha e de características morfológicas. No trabalho citado acima foram utilizados quatro grupos sanguíneos e sete polimorfismos bioquímicos de hemácias e soro, e o valor do *Fst* estimado entre essas duas raças foi de 0,117, diferentemente do que foi encontrada neste estudo, com valor de 0,0458 (Tabela 4), o que evidencia a maior proximidade genética entre MM e MAN, e é justificado por fatos históricos, na qual alguns exemplares do MM foram levados para outra região do país, em São Paulo, e a partir disso, esses animais foram submetidos à seleção natural em um diferente ambiente, para os quais desenvolveram características específicas de adaptação e resistência às novas condições (Egito et al. 2007; Meira, 2010).

4. 6 Estrutura populacional

Até o presente momento, este trabalho é o mais abrangente sobre a estrutura e diversidade genética do Mangalarga Marchador. Uma investigação mais detalhada da estrutura populacional usando o programa STRUCTURE mostrou que é mais provável que existam cinco agrupamentos geneticamente distintos ou cinco fontes de diversidade para as sete raças analisadas. Nota-se o maior compartilhamento de alelos entre as raças brasileiras, por serem oriundas de ancestrais comuns e devido a ocorrências de fluxo gênico entre elas, além de aproximação geográfica entre as mesmas, uma vez que o MM e o CAM são raças originárias do estado de Minas Gerais, e o MAN de São Paulo que é um estado adjacente. Outro motivo

do compartilhamento de alelos principalmente entre MAN e MM é o fato de que a raça MAN foi formada a partir de cruzamentos com animias MM.

Se as duas raças, de maiores interesses, como o MM e MAN não parecem mostrar diferenciação expressiva baseada em marcadores do tipo microssatélites, o que pode estar relacionado ao fluxo gênico contínuo entre as duas populações, a diferença na estrutura genética dentro de cada uma destas raças pode eventualmente ser relacionado a métodos de manejo contrastados, o que foi observado no MM (Figura 3b). Apesar destes grupos genéticos serem oficialmente reconhecidos como raças há um bom tempo, o fato de os Serviços de Registros Genealógicos de cada uma destas raças terem sido criados recentemente, e devido a alguns erros cometidos pelos mesmos, estes fatores podem ter impedido que essas raças existissem como populações distintas, o que permitiu uma troca genética mais fácil entre elas. Silva et al. (2012b) encontraram valor de $K=4$, ideal para explicar as fontes de diversidade em raças brasileiras como Campeira, Lavradeira, Pantaneira, Baixadeiro, MM, ARA e PSI. Assis et al. (2009) também encontraram subdivisão na raça MM, na análise agrupamento, com animais provenientes de regiões distintas do estado de Minas Gerais.

De acordo com Berber et al. (2014), as diferentes raças são classificadas, principalmente, baseadas em diferenças morfológicas e/ou estreitamente relacionadas às localizações geográficas, de tal maneira que diferentes nomes podem ser atribuídos a populações próximas geneticamente, mas localizadas em diferentes áreas administrativas. O isolamento reprodutivo por barreiras geográficas ou por considerações sociopolíticas leva a um processo de deriva genética dentro da população que causará a diferenciação genética entre as subpopulações e que serão detectadas pelo uso de marcadores moleculares. As diferenças morfológicas entre as populações que são consequências da seleção artificial, não são levadas em consideração por esses marcadores.

Em adição, diferenciações expressivas foram observadas em alguns haras do MM, denominados de haras específicos (HE), sendo que alguns haras compartilhavam mais alelos com o MAN (HE4 com predominância de coloração verde), e outros haras com o CAM (HE3 com predominância de coloração amarela) (Figura 3b). A participação genética das raças CAMP e MAN no MM já foram acima elucidadas, porém a presença, mesmo que em poucos haras, dessa diferença entre as subpopulações do MM evidencia que é a gestão genética local (direcionamento de acasalamentos, seleção e pressão de seleção, uso preferencial de alguns reprodutores entre outros) que possivelmente criou algum grau de isolamento reprodutivo promovendo observável diferenciação. Os haras específicos, de mesma coloração, também compartilhavam alguns ancestrais comuns entre si, além de que, os HE, em sua grande

maioria, tinham animais descendentes de um único ou no máximo dois ancestrais importantes da raça, ancestrais estes que mais contribuíram geneticamente para o *pool* genético atual do Mangalarga Marchador. Outro fato que explica observável diferenciação dos haras pode ser a alteração, ao longo do tempo, das contribuições genéticas dos principais fundadores da raça, o que é observado no trabalho de Baena (2019).

Além disso, quando realizado a mesma análise em subpopulações divididas de acordo com estados federativos distintos do país não ocorreu a mesma diferenciação, demonstrando que não é a localização geográfica a causadora dessa discriminação.

Em geral, as análises com marcadores moleculares microssatélites foram suficientes para detectarem uma pequena quantidade de diferenciação genética entre as subpopulações do MM, e as principais estatísticas apontam para uma pequena homogeneização genética a nível populacional ao longo do tempo, porém com índice de diversidade genética satisfatório para sua conservação, sobrevivência como espécie e de garantia de seleção.

5. CONCLUSÃO

Os marcadores moleculares revolucionaram nossa capacidade de caracterizar a diversidade genética e racionalizar a seleção genética em animais domésticos, fornecendo informações valiosas para monitorar a seleção artificial, e tornando-se ferramentas fundamentais para a pesquisa básica em programas de genômica animal. Eles prometem encontrar utilidade não só para realizar uma seleção indireta de animais com base em marcadores ligados a traços de interesse (por exemplo, resistência a doenças, tolerância ao estresse, tipo de marcha, entre outros), mas também para a avaliação de prioridades e estratégias de conservação, ou seja, otimizar o plano de acasalamento de uma população e também para evitar a perda de diversidade genética durante a seleção, ajudando na identificação de animais ou conjuntos de animais que devem ser mantidos no processo de reprodução apenas para evitar perda de diversidade genética.

O presente estudo contribui para o conhecimento da estrutura genética e caracterização molecular da população Mangalarga Marchador, além de evidenciar que os microssatélites podem ser utilizados para estabelecer as relações genéticas entre populações e fornecem poder estatístico razoável para a designação de raças, porém quando raças são intimamente relacionadas, esses marcadores não são suficientes, sendo alguns redundantes e não abrangem todo o genoma do equino.

Concomitantemente com os dados de microssatélites do MM, a utilização de informações obtidas a partir de análises de *pedigree* acurados, de medidas morfológicas e de estudos com marcadores moleculares mais informativos e que abrangem todo o genoma equino como o SNP (Polimorfismo de Nucleotídeo Único), podem fornecer a base para o desenvolvimento de estratégias de reprodução clássicas adequadas, bem como, a implementação de um programa de melhoramento genético, visando maximizar sua variabilidade genética, conservar alelos úteis e estruturar geneticamente a raça como uma entidade mais definida.

AGRADECIMENTOS

Os autores agradecem à Universidade Federal de Lavras (UFLA), a CAPES, ao CNPq, à Universidad Nacional de La Plata e à ABCCMM.

REFERÊNCIAS

Achmann, et al., Microsatellite diversity opulation subdivision in gene flow in the Lipizzan horse, *Animal Genetics*, v5, p285-292, 2004.

Assis, J.B., Laat, D.M., Peixoto, M.G.C.D., Bergmann, J.A.G., Fonseca, C.G., Carvalho, M.R.S., Genetic diversity and population structure in Brazilian Mangalarga Marchador horses, *Genetics and Molecular Research*, v8, n4, p1519-1524, 2009.

Azor, P.J., Valera, M., Gómez, D. M., Goyache F., Molina, A., Genetic characterization of the Spanish Trotter horse breed using microsatellite markers, *Genet. Mol. Biol.*, v30, p37-42, 2007.

Barcaccia, G., Felicetti, M., Galla, G., Capomaccio, S., Cappelli, K., Albertini, E., Buttazzoni, L., Pieramati, C., Silvestrelli, M., Supplizi, A.V., Molecular analysis of genetic diversity, population structure and inbreeding level of the Italian Lipizzan horse, *Livestock Science*, v151, p124–133, 2013. <http://dx.doi.org/10.1016/j.livsci.2012.11.022>.

Barker, J. S. F., A global protocol for determining genetic distances among domestic livestock breeds. In: *World Congress on Genetics Applied to Livestock Production*, 5, 1994, Guelph. *Proceedings... Guelph: International Committee for World Congresses on Genetics Applied to Livestock Production*, v5, p501-508, 1994.

Beck, S.L., *Equinos: Raças, Manejo e Equitação*. 2th ed. Editora dos Criadores, São Paulo, 1985.

Behl, R., Behl, J., Gupta, N., Gupta, S. C., Ahlawat, S. P. S., Ragnekar, M., Ahmed, Z., Genetic characterization of Zanskari breed of horse, *Journal of Genetics*, v85, p199-203, 2006.

Behl, R., Behl, J., Gupta, N., Gupta, S.C., Genetic relationships of five Indian horse breeds using microsatellite markers, *Animal*, v1, p483–488, 2007.

Berber, N., Gaouar, S., Leroy, G., Kdidi, S., Aouel, N.T., Mehtar, N.S., Molecular characterization and differentiation of five horse breeds raised in Algeria using polymorphic microsatellite markers, *Anim. Breed. Genet.*, v131, p387–394, 2014.

Botstein, D., White, R.L., Skolnick, H. et al., Construction of a genetic linkage map in man using restriction fragment length polymorphism, *American Journal of Human Genetics*, v32, p314-331, 1980.

Butler, J.M., *Forensic DNA Typing: Biology Technology and Genetics of STR Markers*, 2th ed., Academic Press, p660, 2005.

Cañon, J., Checa, L. M., Carleos, C., Vega-Pla, J.L., Vallejo, M., Dunner, S., The genetic structure of Spanish Celtic horse breeds from microsatellite data, *Animal genetics*, v31, p39-48, 2000.

Chieffi, A., Introdução do equino no Brasil e desenvolvimento da criação, In: *Criemos bons equídeos*, Rio de Janeiro: Ministério da agricultura / Serviço de Informações Agrícolas, p21, 1950.

Choi, S. K., Cho, C. Y., Yeon, S. H., Cho, B. W., Cho, G. J., Genetic characterization and polymorphisms for parentage testing of the Jeju horse using 20 microsatellite loci, *J. Vet. Med., Sci.*, v70, n10, p1111-1115, 2008.

Conant, E.K., Juras, R., Cothran, E.G., A microsatellite analysis of five Colonial Spanish horse populations of the south eastern United States, *Animal Genetics*, v43, p53-62, 2011.

Costa, N., Val, L.L., Leite, G.U., Estudo da preservação do equino Nordeste, Departamento de Produção Animal, Recife, Brasil, v38, p1-15, 1974.

Costa, M. R., Marques, J. R. F., Silva, C. S., Sampaio, M. I. C., Bermejo, J. V. De., Silva, F. K. S., Vega Pla, J. L., Genetic distance of horses (*equus caballus*) using DNA microsatellites, *Revista Biociências*, v15, 2009.

Criscione, A., Moltisanti, V., Chies, L., Marletta, D., Bordonaro, S., A genetic analysis of the Italian Salernitano horse. *Animal*, v9, p1610–1616, 2015.

Cunningham, E.P., Dooley, J.J., Splan, R.K., Bradley, D.G., Microsatellite diversity, pedigree relatedness and the contributions of founder line gesto thoroughbred horses. *Anim. Genet.*, v32, p360–364, 2001. <http://dx.doi.org/10.1046/j.1365-2052.2001.00785.x>.

Dorji, J., Tamang, S., Tshewang, T., Dorji, T., Dorji, T.Y., Genetic diversity and population structure of three traditional horse breeds of Bhutan based on 29 DNA microsatellite markers, *PLoS ONE*, v13, p199-376, 2018. <https://doi.org/10.1371/journal.pone.0199376>.

Equipe Estatcamp (2014). *Software Action. Estatcamp- Consultoria em estatística e qualidade*, São Carlos - SP, Brasil. Disponível em: <http://www.portalaction.com.br/>. 2014.

Egito, A.A., Fuck, B.H., Mcmanusm, C.M., Mariante, A.S., Genetic characterization of Pantaneira horse using RAPD markers, *Rev. Soc. Bras. Zoot.*, v36, p799-806, 2007.

Evanno, G., Regnaut, S., Goudet, J. Detecting the number of clusters of individuals using the software structure: a simulate on study, *Molecular Ecology*, v14, p2611–20, 2005. <https://doi.org/10.1111/j.1365294X.2005.02553.x> PMID: 15969739.

Falush, D., Stephens, M., Pritchard, J.K., Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies, *Genetics*, v164, p1567–1587, 2003.

Gastaldo, A.Z., Rodenbusch, R., Fossati, R, Azambuja, C.J., Alho, C.S., Population genetic study over 32,000 equines from Uruguay using seventeen forensically informative STR loci, *Forensic Sci Int Genet.*, v26, p19-22, 2017.

Gonçalves, R.W, Costa, M.D, Rezende,A.S.C., Rocha Júnior,V.R., Leite, J.R.A., Efeito da endogamia sobre características morfométricas em equinos da raça Mangalarga Marchador, *Arq. Bras. Med. Vet. Zootec.*, v64, p.419-426, 2012.

Goudet, J., Raymond, M., de Meeus, T., Rousset, F., 1996. Testing differentiation in diploid populations, *Genetics*, v144, p1931-1938, 1996.

Goulart, J.A., O equino na formação do Brasil, São Paulo: Letras e Artes, p249, 1964.

Hedrick, P.W. *Genetics of populations*, 3th ed., Massassusetts, Jones and Bartlett Publishers, p736, 2005.

Hendricks BL (2007). *International Encyclopedia of Horse Breeds*. Norman, OK: University of Oklahoma Pres. 486 p.

Juras, R., Cothran, E.G., Klimas, R., Genetic analysis of three Lithuanian horse breeds, *Acta Agric,Scand.*, v53, p180–185, 2003.

Lee, S., Cho, G., Parentage testing of Thoroughbred horse in Korea using microsatellite DNA typing, *J. Vet. Sci.*, v7, p66-67, 2006.

Lima, R.A.S., Shirota, R., Barros, G.S.C., Estudo do complexo do agronegócio equino, Piracicaba: ESALQ/USP, p205, 2006.

Ling, Y. H., Guan, W. J., Cheng, Y. J., Wang, Y. P., Han, J. L., Mang, L., Zhao, Q. J., He, X. H., Pu, Y. B., Fu, B. L., Evaluation of the genetic diversity and population structure of Chinese indigenous horse breeds using 27 microsatellite markers, *Animal Genetics*, v42, p56-65, 2010.

Lippi, A.S., Mortari, N. Studies of blood groups and protein polymorphisms in the Brazilian horse breeds Mangalarga Marchador and Mangalarga (*Equus caballus*), *Genetics and Molecular Biology*, v26, p431-434, 2003.

Lopes, M. S., Mendonça, D., Rojer, H., Cabral, V., Bettencourt S.X., Machado. A.C., Morphological and genetic characterization of na emerging Azorean horse breed: the Terceira Pony, *Frontiers Genetics*, v6, p1-7, 2015.

Luís, C., Cothran, E.G., Oom, M.M., Microsatellites in Portuguese autochthonous horse breeds: usefulness for parentage testing, *Genetics and Molecular Biology*, v25, p131-134, 2002.

Luis, C., Cothran, E.G., Oom, M.M., Inbreeding and genetic structure in the endangered Sorraia horse breed: implications for its conservation and management, *J. Hered.*, v98, p 232-237, 2007.

Monteiro, J., Origem e evolução do equino na Península Ibérica, *Revista dos criadores*, v55, 1986.

Nascimento, J.F, Mangalarga Marchador: Tratado morfofuncional, 1th ed., Belo Horizonte: Associação Brasileira de Criadores de Equino Mangalarga Marchador, p900, 1999.

Nei, M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, v89, p. 583-590, 1978.

Nei, M., *Molecular evolutionary genetics*, 1th ed., Columbia University Press, 512 p, 1987.

Notter, D.R., The importance of genetic diversity in livestock populations of the future, *J. Anim. Sci.*, v77, p61-69, 1999.

Ott, J., Strategies for characterizing highly polymorphic markers in human gene mapping, *Am. S. Hum. Genet.*, v51, p.283-290, 1992.

Park, S. MStools v.3.1.1: Excel Spreadsheet Toolkit for Data Conversion. Animal Genomics Lab, University College, Dublin, Ireland. Acesso em (<http://animalgenomics.ucd.ie/sdepark/ms-toolkit/>) 2008.

Pereira, G. L., Chud, T., Bernardes, P., Venturini, G., Chardulo, L. A., Curi, R., Genotype Imputation and Accuracy Evaluation in Racing Quarter Horses Genotyped Using Different Commercial SNP Panels, *Journal of Equine Veterinary Science*, v58, p89-96, 2017.

Pinto, P.C.A.B., *Genética de Populações*, Universidade Federal de Lavras, Lavras, 2008.

Pires, D.A.F., et al., Genetic diversity and population structure in remnant subpopulations of Nordeste horse breed, *Arch. Zootec*, v63, p.349-358, 2014. <http://dx.doi.org/10.4321/S000405922014000200013>.

Raymond, M., Rousset, F., An exact test for population differentiation. *Evolution*, v49, p1283-1286, 1995.

Ribeiro, M. et al., Allelic diversity and forensic performance of 17 microsatellite markers in Criollo and Campolina horse breeds. In: Congresso Brasileiro de Genética, 51, 2005. Anais... Águas de Lindóia: São Paulo, Sociedade Brasileira de Genética, 2005.

Rousset, F., GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux, *Molecular Ecology Resources*, v8, p103–106, 2008.

Senju, N., Tozaki, T., Kakoi, H., Almunia, J., Maeda, M., Matsuyama, R., Takasu, M., Genetic characterization of the Miyako horse based on polymorphisms of microsatellites and mitochondrial DNA, *J. Vet. Med. Sci.*, v79, p218–223, 2017.

Shahsavarani, H., Rahimi-Mianji, G., Analysis of genetic diversity and estimation of inbreeding coefficient within Caspian horse population using microsatellite markers, *African Journal of Biotechnology*, v9, pp293-299, 2010.

Silva, A.C., Paiva S.R., Albuquerque M.S., Egito A.A., Santos S.A., Lima F.C., Castro S.T., Mariante A.S., Correa P.S., McManus C.M., Genetic variability in local Brazilian horse lines using microsatellite markers, *Genet. Mol. Res.*, v10, p881-90, 2012b.

Silva Filho, E., Schneider, M.P.C., Silva, A.L.C., Variabilidade genética de equinos baseada em DNA microssatélites, *Rev. Trop. Cienc. Agrar. Biol.*, v1, p76-87, 2007.

Solis, A., Jugo, B.M., Mériaux, J.C., Iriondo, M., et al., Genetic diversity within and among four South European Native Horse breeds based on microsatellite DNA analysis: implications for conservation, *J. Hered.*, v96, p670-678, 2005.

Souza, F.A.C., Fernandes, T. J., Moura, R.S., Meirelles, S.L C., Ribeiro, A.R., Cunha, F.O., Augusto, M.J., Non linear modeling growth body weight of Mangalarga Marchador horses. *Ciência Rural*, v47, p.2-6, 2017. <http://dx.doi.org/10.1590/0103-8478cr20160636>.

Stasiol, L.D., Perrotta, G., Blasi, M., Lisa, C., Genetic characterization of the Bardigiano horse using microsatellite markers, *Italian Journal of Animal Science*, v7, p243-250, 2008.

Takasu, M., Hiramatsu, N., Tozaki, T., Kakoi, H., Nakagawa, T., Hasegawa, T., Maeda, M., Murase, T., Mukoyama, H., Genetic characterization of endangered Kiso horse using 31 microsatellite DNAs, *The Journal of Veterinary Medical Science*, v74, pp6-161, 2012.

Tambasco, D. D., M. M. A., Coutinho, L. L., Tambasco, A. J., Tambasco, M. D., Regitano, L. C. A., Caracterização Molecular de Animais da raça Nelore utilizando Microssatélites e Genes Candidatos, *Rev. bras. zootec.*, v4, p1044-1049, 2000.

Tozaki, T., Takezaki, N., Hasegawa, N., Ishida, M., Kurosawa, M., Tomita, N., Saitou, H., Mukoyama. Microsatellite Variation in Japanese and Asian Horses and Their Phylogenetic Relationship Using a European Horse Out group, *Journal of Heredity*, v94, p374–380, 2003. <https://doi.org/10.1093/jhered/esg079>.

Vega-Pla J.L., Calderón J., Rodríguez-Gallardo P.P., Martínez A.M., Rico C., Saving feral horse populations: does it really matter? A case study of wild horses from Doñana National Park in southern Spain, *Anim. Genet.*, v37, p571-8, 2006.

Weir, B.S., Inferences about linkage disequilibrium, *Biometrics.*, v35, p235–254, 1979.

Wright, S., 1965. The interpretation of population structure by F-statistics with special regard to systems of mating, *Evolution*, v19, p395–420, 1965.

ARTIGO 3 – Pré-caracterização genética de equinos da raça Mangalarga Marchador baseado em marcadores moleculares SNP

Formatado de acordo com a norma do periódico *Journal of Animal Breeding and Genetics*.

**Marielle Moura Baena^a, Felipe Amorim Caetano Souza^b, Yasmim Pereira Ribeiro^c,
Silvina Diaz^d, Raquel Silva de Moura^e, Sarah Laguna Conceição Meirelles^{f*}**

^{a,b,c,e,f} Universidade Federal de Lavras, Department Animal Science, University Campus, postal code 3037, Downtown, CEP 37200000, Lavras/MG, Brazil; E-mail: marielle_moura@hotmail.com; felipeuflazootecnia@yahoo.com; yaribeiro.17@gmail.com; raquelmourabr@yahoo.com.br; sarah@ufla.br*;

^d Universidad Nacional de La Plata, Faculty of Veterinary Sciences, Calle 60 and 118 S/N (1900), Box 296, La Plata, Province of Buenos Aires - Argentina; E-mail: sil_diaz@hotmail.com.

*Corresponding author: Universidade Federal de Lavras, Department Animal Science, University Campus, postal code 3037, Downtown, CEP 37200000, Lavras/MG, Brazil; E-mail: sarah@ufla.br

RESUMO

O Mangalarga Marchador (MM) é uma raça brasileira e atualmente há mais de 600 mil animais oficialmente registrados. Assim, o objetivo deste estudo foi realizar a pré-caracterização genética da raça, investigando sua diversidade genética e estrutura populacional por meio da genotipagens de SNP. Para a genotipagens dos animais, em chip de SNP70K, foram coletados amostras de pelo de 244 equinos (MM, n=106; Puro Sangue Espanhol - PSE n= 20; Puro Sangue Lusitano - PSL n= 21; Sorraia - SOR n=16; Puro Sangue Inglês - PSI n= 21; Campolina - CAM n= 22 e Mangalarga - MAN n= 38). Estes dados passaram por filtros de qualidade, e em seguida parâmetros foram calculados pelo programa PLINK, como as homozigosidades esperadas (H_{exp}) e observadas (H_{obs}), coeficientes individuais de endogamia (F), heterozigosidades esperadas (H_e) e observadas (H_o), equilíbrio de Hardy-Weinberg (EHW), número de marcadores em não equilíbrio, frequência de alelos menores (MAF), desequilíbrio de ligação (DL), diferenciação genética (F_{st}) e análise de componentes principais (ACP). Foi também verificado a presença de haplótipos comuns entre raças utilizando o programa HAPLOVIEW, além de avaliar a estruturação populacional pelo STRUCTURE. O MM obteve o maior valor de H_{obs} (34.916,70± 919,0) e a raça SOR (16.764,8±1159,4) a menor. Nos animais MM, 51% possuíam uma perda significativa de heterozigosidade, sendo o maior valor de F de 0,1859. O MAF médio no MM foi de 0,2010. Desvios de EHW foram obtidos em todos os cromossomos. A H_e variou de 0,1835 na SOR a 0,3240 no PSI, e a H_o oscilou de 0,2159 no SOR e 0,3261 no PSI. O número de SNP em não EHW foi maior no MM (756) e menor no SOR (126). A raça SOR apresentou maior extensão de DL e valores mais elevados de r², e as raças CAM e o MM tiveram extensões menores e baixos valores de r². Foi encontrado bloco de haplótipo semelhante nas raças MAN e MM no cromossomo ECA1. O menor nível de diferenciação ocorreu entre o CAM e o MM (F_{st} = 0,0411), e a maior divergência entre PSI e o SOR (F_{st} = 0,2284). A análise de estruturação populacional demonstrou que a raça MM é pouca definida, assim como as demais raças brasileiras MAN e CAM, ao contrário das raças PSL, PSE, PSI e SOR que foram entidade genéticas bem definidas e observou-se o compartilhamento de alelos comuns entre MM e MAN. A ACP discriminou em maior magnitude as raças SOR e PSI das demais raças, em moderada discriminação as raças PSL e PSE, e em menor discriminação as raças MM, MAN e CAM. Na população do MM é necessário o monitoramento contínuo e regular da diversidade genética com o uso de muitos marcadores moleculares, registros precisos de *pedigrees*, maiores controles no número de pais usados para as próximas gerações, nos acasalamentos e nos cruzamentos com outras raças para maximizar sua variabilidade genética, conservar genes úteis e estruturar a raça como uma entidade genética mais definida.

Palavras-chave: diversidade genética, equideocultura, polimorfismo, SNP

ABSTRACT

The Mangalarga Marchador (MM) is a Brazilian breed and currently there are more than 600 thousand animals officially registered. Thus, the objectives of this study were to perform the genetic pre-characterization of the breed, investigating its genetic diversity and population structure through genotyping of SNP. For the genotyping of the animals, on the SNP70K chip, samples of hair of 244 horses were collected (MM, n = 106; Puro Sanguê Espanhol - PSE n = 20; Lusitano - LUS n = 21; Sorraia - SOR n = 16 (N = 21) and Campolina - CAM n = 22 and Mangalarga - MAN n = 38). These data passed through quality filters, and then parameters were calculated by the PLINK program, as homozygous expected (H_{exp}) and observed (H_{obs}), individual inbreeding coefficients (F), expected heterozygosity (H_e) and observed (H_o), Hardy-Weinberg equilibrium (HWE) number of non-equilibrium markers, lower allele frequency (MAF), linkage disequilibrium (LD), genetic differentiation (F_{st}) and principal component analysis (PCA). It was also verified the presence of common haplotypes between breeds using the HAPLOVIEW program, besides evaluating the population structure by STRUCTURE. The MM obtained the highest H_{obs} value (34,916.70 ± 919.0) and the SOR breed (16,764.8 ± 1159.4) the lowest. In the MM animals, 51% had a significant loss of heterozygosity, with the highest F value of 0.1859. The mean MAF in the MM was 0.2010. HWE deviations were obtained on all chromosomes. H_e ranged from 0.1835 in SOR to 0.3240 in THO, and H_o ranged from 0.2159 in SOR and 0.3261 in THO. The number of SNP in non-HWE was higher in the MM (756) and lower in the SOR (126). The SOR breed had a higher LD extent and higher r² values, and the CAM and MM races had smaller extensions and lower r² values. Similar haplotype block was found in the MAN and MM races on the ECA1 chromosome. The lowest level of differentiation occurred between CAM and MM (F_{st} = 0.0411), and the highest divergence between THO and SOR (F_{st} = 0.2284). The population structure analysis showed that the MM breed is poorly defined, as well as the other Brazilian breeds MAN and CAM, unlike the LUS, PSE, THO and SOR breeds that were well defined genetic entity and the common allele sharing was observed between MM and MAN. The PCA discriminated in greater magnitude the SOR and THO races of the other races, in moderate discrimination the LUS and PSE races, and in less discrimination the races MM, MAN and CAM. In the MM population, continuous and regular monitoring of genetic diversity is necessary with the use of many molecular markers, precise pedigree records, greater controls on the number of parents used for the next generations, mating and crossbreeding with other breeds to maximize their genetic variability, conserving useful genes, and structuring the breed as a more defined genetic entity.

keywords: equideoculture, genetic diversity, polymorphism, SNP

1. INTRODUÇÃO

A raça Mangalarga Marchador (MM) é a responsável pelo maior e mais representativo plantel equídeo do Brasil com mais de 600 mil animais registrados, sendo assim, declarada oficialmente como raça nacional no ano de 2014. A raça originou-se no sul do estado de Minas Gerais, a mais de 200 anos, a partir do cruzamento de um garanhão da raça Puro Sangue Lusitano (PSL), proveniente da Península Ibérica, com éguas nativas brasileiras, provavelmente descendentes dos primeiros equinos trazidos pelos colonizadores portugueses em 1500 (Meira, 2010; Lima e Cintra, 2016). Os equinos da raça PSL têm parentesco com a raça Andaluz Espanhol, cuja origem étnica provém de equinos nativos das Penínsulas Berbere, Germânica e Ibérica, e segundo Nascimento (1999) e Assis et al. (2009) ocorreu também uma miscigenação com outras raças como Árabe, Andaluz, Crioulo e Quarto de Milha nas quais também contribuíram geneticamente para a formação do MM.

As pesquisas de melhoramento genético na espécie equina eram escassas até recentemente, mas com novas tecnologias a nível molecular que contribuíram para a melhor seleção destes animais, esse cenário vem se modificando (Pereira, 2017). Com a identificação de polimorfismos presentes na molécula de DNA, chamado de marcadores moleculares, tornou-se possível a discriminação genotípica, o que permitiu a análise da variabilidade genética das populações de animais. O estudo de caracterização genética, que verifica os níveis de variabilidade genética dentro e entre populações, além de fornecer toda a estrutura genética da mesma, utiliza-se desses marcadores para realizar essa verificação (Tambasco et al., 2000). Os marcadores moleculares revolucionaram a capacidade de caracterizar a variação genética e racionalizar a seleção dos animais, e de acordo com Lanteri et al. (2006) e Soattin et al. (2009), os marcadores têm um efeito benéfico e confiável para uma análise de arquiteturas genética e detecção de polimorfismos.

Com uma população mundial acima de 59 milhões de animais e aproximadamente 500 raças distintas, os equinos são animais economicamente importantes e populares para a agricultura, transporte e recreação em todo o planeta, mas apesar da relevância da espécie equina, trabalhos de caracterização genética com o uso de marcadores moleculares do tipo Polimorfismo de Nucleotídeo Único (SNP) são exíguos na espécie, sendo alguns relatados por McCue et al. (2012), Petersen et al. (2013) e Janova et al. (2013), com equinos doméstico e perissodáctilos, raça modernas como Árabe, Mangalarga, Andaluz, PSL, Puro Sangue Inglês, entre outros, e raça ameaçadas de extinção, respectivamente. Na raça MM já foram relatadas pesquisas semelhantes, porém com marcadores do tipo microssatélites como os trabalhos de

Assis et al. (2009), Silva et al (2012a) e Silva et al. (2012b), mas que não permitem a comparação de dados entre os estudos. Assim, grandes investigações de diversidade genética utilizando marcadores SNP são poucas, e uma análise mais abrangente do MM com raças ibéricas, inglesas e nativas brasileiras, utilizando esses marcadores, ainda não foi relatada na literatura.

O *Equine Genetic Diversity Consortium* (EGDC), uma colaboração internacional da comunidade científica de equinos, foi formada para auxiliar na quantificação da diversidade genética e para o conhecimento das relações dentro e entre as populações, mas em uma escala genômica. O desenvolvimento deste consórcio facilitou a coleta de amostras de 36 diferentes raças para a genotipagem do *Illumina SNP Beadchip*. Essas raças utilizadas representavam as mais populares do mundo, bem como classes fenotípicas divergentes, diferentes regiões geográficas e de diferentes histórias de origem racial. Essa plataforma de genotipagem por marcadores SNP padronizada permite a compilação de dados entre as raças a um nível nunca antes alcançado. Os resultados fornecidos por essa colaboração somente agora permitem a descrição fiel da diversidade genética, da avaliação dos efeitos do isolamento genético, endogamia e seleção dentro de raças, e a descrição do relacionamento entre as várias raças (Petersen et al., 2013).

Deste modo, os objetivos deste estudo foram realizar a pré-caracterização genética da raça Mangalarga Marchador, investigando sua diversidade genética e comparando-a com outras raças ligadas ou não à sua história de formação, bem como a investigação da estrutura populacional da raça por meio da genotipagem de SNP em larga escala.

2. MATERIAL E MÉTODOS

2.1 Banco de dados e amostragem

Inicialmente, para dar desenvolvimento ao estudo de pré-caracterização genética da raça Mangalarga Marchador (MM) foi realizada uma análise no banco de dados cedido pela Associação Brasileira dos Criadores do Cavallo Mangalarga Marchador (ABCCMM). Este banco continha distintas informações de 509.428 mil animais oficialmente registrados durante os anos de 1906 a 2016.

Para auxiliar a amostragem, de animais contidos neste banco de dados, foi realizado o estudo da estrutura populacional do MM utilizando o programa PEDIG (Boichard, 2002),

específico para este fim. Com este programa foram obtidos 163 ancestrais mais representativos geneticamente da raça, sendo que esses ancestrais representavam 30% do pool genético da população atual. Dessa forma, foram selecionados os descendentes vivos destes 163 principais ancestrais, mas que atendessem aos seguintes critérios de amostragem: animais que tinham paternidade comprovada por teste de DNA (DNAvp – verificação de parentesco com pai e mãe), animais que tinham até a terceira geração ascendente conhecida, não eram aparentados, possuíam o registro definitivo e não estavam incluídos no “Livro de Inclusão do Termo de Ajustamento de Conduta (TAC)”. Os animais presentes no Livro TAC foram excluídos da amostragem, pois poderiam ter registros definitivo ou provisório, mas que não tiveram sua genealogia comprovada por teste de DNA. Todos estes critérios foram utilizados para que a amostra a ser trabalhada tivesse maior confiabilidade na caracterização genética da raça MM.

2.2 Animais

Após o estabelecimento da amostragem, foram utilizados para o estudo de pré-caracterização genética informações de 106 equinos MM que eram descendentes de 35 ancestrais que mais contribuíram geneticamente para a raça, além de atenderem todos os critérios já mencionados anteriormente. Estes animais eram provenientes de diferentes haras, localizados em seis estados brasileiros, sendo o estado de Minas Gerais o mais representativo, seguido de São Paulo, Rio Grande do Sul, Rio de Janeiro, Bahia e Pernambuco.

Além dos animais MM (n = 106), amostras de pelos (via correspondência) de equinos de seis outras raças foram também utilizadas, nos quais estes equinos não eram aparentados, sendo elas: equinos provenientes da Espanha (Puro Sangue Espanhol n= 20; PSE), de Portugal (Puro Sangue Lusitano n= 21; PSL e Sorraia n=16; SOR), da Inglaterra, porém, nascidos e criados no Brasil (Puro Sangue Inglês n= 21; PSI), e raças brasileiras (Campolina n= 22; CAM e Mangalarga n= 38; MAN), totalizando 244 equinos (n= 66 machos e n= 178 fêmeas).

2.3 Coleta de pelo

Para a realização da genotipagem dos animais, amostras de pelo da cauda e/ou da crina foram coletados de acordo com as orientações fornecidas pelo laboratório *Neogen Corporation, Michigan, USA*. Seguindo essas orientações, os pelos velhos ao redor da área de

amostragem foram eliminados, e de cinco a 10 pelos foram segurados e puxados firmemente em direção do crescimento do pelo para que fossem retirados. Como o DNA foi extraído do bulbo capilar (raiz do pelo), os fios sem raiz foram descartados. Esse procedimento foi repetido até a operação completar aproximadamente 30 fios com as raízes. Logo que retirados, os pelos foram acondicionados em cartuchos de papel específicos para este fim, identificados e enviados ao laboratório.

2.4 Genotipagem de SNP

A genotipagem das amostras foi realizada pelo laboratório *Neogen Corporation* utilizando o chip de alta densidade do painel *GGP Equine SNP70K* da *Illumina Inc., San Diego, CA*, que dispõem, em sua matriz, de 65.157 SNP altamente informativos e uniformemente distribuídos por todos os cromossomos do genoma equino. Esse chip de SNP foi fabricado de acordo com a tecnologia *Infinium*, usando a plataforma *iScan*, e o protocolo de genotipagem foi seguido de acordo com o estabelecido pelo fabricante (disponível em www.illumina.com).

2.5 Controle de qualidade dos dados

O controle de qualidade das informações de genotipagem, para os indivíduos e SNP, foram realizados pelo programa PLINK v.1.9 (Purcell, 2007). Com relação aos indivíduos, foi utilizado o *Call Rate*, pelo comando (--mind 0,05) do programa, em que animais com valores abaixo de 0,95 foram excluídos de todo conjunto amostral. Em relação aos SNP, os considerados não informativos foram definidos como aqueles com MAF (*Minor Allele Frequency* - Frequência de alelos menores) inferior a 0,05 (--maf 0,05) e com *Genotype Rate* inferior a 0,90 (--geno 0,1), além de que SNP localizados nos cromossomos sexuais X e Y foram também retirados.

Após o processo de filtragem de qualidade, o número de SNP informativos, considerando todos os controles de qualidade foi de 50.953 e o número de animais de 236. Desta forma, permaneceram para as análises genéticas subseqüentes genótipos de 20 equinos da raça PSE, 21 PSL, 16 SOR, 21 PSI, 22 CAM, 36 MAN e 100 MM.

2.6 Análises estatísticas

Diversos comandos do programa PLINK v.1.9 (Purcell et al. 2007) foram utilizados para estimar alguns parâmetros genéticos da população de equinos Mangalarga Marchador e nas demais raças com o uso de marcadores moleculares SNP. Os parâmetros calculados foram: as homozigosidades esperadas (Hom_e) e observadas (Hom_o), os coeficientes individuais de endogamia genômica (F) nos quais foram obtidos pelo comando (--het) do programa. Os coeficientes F foram obtidos baseados nos números observados e esperados de genótipos homozigotos, ou seja, $F = \frac{[\text{contagem de genótipos homozigotos observada}] - [\text{contagem esperada}]}{[\text{total de observações}] - [\text{contagem esperada}]}$. Resultados negativos de F são possíveis, indicando que o indivíduo tem menos genótipos homozigóticos do que se esperaria ao acaso a nível genômico.

Outros parâmetros computados neste trabalho contribuíram para caracterizar a diversidade genética da raça MM e dar base de comparação com as demais raças equinas como os parâmetros: heterozigosidades esperadas (He) e observadas (Ho), equilíbrio de Hardy-Weinberg (EHW), número de marcadores em não equilíbrio nos quais foram obtidos pelo comando do PLINK (--hwe), além da frequência de alelos menores (MAF) por (--freq).

O desequilíbrio de ligação (DL) foi avaliado por meio da estatística de r^2 (Hill e Robertson, 1968), entre todos os pares de SNP em cada cromossomo do genoma equino. Também foi calculado o DL médio para intervalos de diferentes distâncias físicas entre os marcadores. Assim, os genótipos para todos os pares de SNP foram avaliados para estimar o DL em intervalos de 0 a 100 quilobase (kb), 100 a 200 kb, 200 a 300 kb, 300 a 400 kb, 400 a 500 kb, 500 a 600 kb, 600 a 700 kb, 700 a 800 kb, 800 a 900 kb e de 900 a 1.000 kb. Os comandos utilizados foram (--r2 e --ld-window) que reportam às correlações quadradas entre os marcadores. Os blocos de haplótipos também foram estimados, através do comando (--blocks), como sugerido por Gabriel et al. (2002), e também foi verificado a presença de haplótipos comuns entre as raças utilizando o programa Haploview 4.2 (Barrett et al, 2005).

Estimativas de diferenciação genética de F_{st} foram calculadas pelo comando (--fst), utilizando todo o conjunto de dados de SNP, cálculo este baseado no método de Weir e Cockerham (1984).

Para descrever melhor a estruturação populacional de cada raça, avaliou-se a mistura genética dentro de cada população, usando o algoritmo de agrupamento bayesiano do programa STRUCTURE (Pritchard et al., 2000). A análise foi realizada utilizando um período de burn-in de 20.000 iterações seguidas de 50.000 repetições. Para a escolha do número ideal

de supostas populações geneticamente definidas, isto é, K , foi utilizada uma faixa de exploração de K abrangendo de 2 a 10.

Com o mesmo conjunto de dados de SNP foi realizado também a análise de componentes principais (ACP), baseada nas variâncias da matriz de relacionamento genético entre os indivíduos a partir do comando (--pca) do PLINK.

3. RESULTADOS

3.1 Homozigosidade e endogamia genômica

A estatística descritiva das homozigosidades observadas (Hom_o) e esperadas (Hom_e), está apresentada na Tabela 1. O maior valor de Hom_o ($34.916,70 \pm 919,0$) foi encontrado no Mangalarga Marchador (MM). Observa-se que as raças brasileiras MM, Campolina (CAM) e Mangalarga (MAN) obtiveram maiores valores de Hom_o com relação às suas raças ascendentes como o Puro Sangue Lusitano (PSL), Puro Sangue Espanhol (PSE) e Sorraia (SOR), e também superior à raça inglesa Puro Sangue Inglês (PSI), sendo a raça Sorraia ($16764,8 \pm 1159,4$) a de menor valor entre todas as raças analisadas no estudo.

Tabela 1. Estatística descritiva das homozigosidades observadas (Hom_o) e esperadas (Hom_e) das populações de equinos Mangalarga Marchador e das demais raças brasileiras, ibéricas e inglesa.

Parâmetros	MM	MAN	CAM	PSL	PSE	SOR	PSI
Hom_o	34916,7	34456,9	32964,41	30713,67	28895,55	16764,8	29827,3
± DP	±919,0	±1205,4	±597,4	±905,2	±1136,3	±1159,4	±771,7
Mínimo	31482,0	32159,0	31846,0	29549,0	26337,0	15589,0	27976,0
Máximo	37912,0	37505,0	34353,0	33453,0	30540,0	19677,0	31062,0
Hom_e	313206,6	316331,4	319010,63	299819,8	267284,7	114274,9	259805,2
± DP	±104834,2	±966269,3	±711775,5	±686272,6	±913115,1	±912737,3	±1086003,7
Mínimo	33800,0	33900,0	33400,0	31500,0	29700,0	18300,0	30000,0
Máximo	348900,0	345300,0	334400,0	315400,0	297600,0	183200,0	305600,0

DP = Desvio Padrão; MM = Mangalarga Marchador; MAN = Mangalarga; CAM= Campolina; PSL = Puro Sangue Lusitano; PSE = Puro Sangue Espanhol; SOR = Sorraia; PSI = Puro Sangue Inglês.

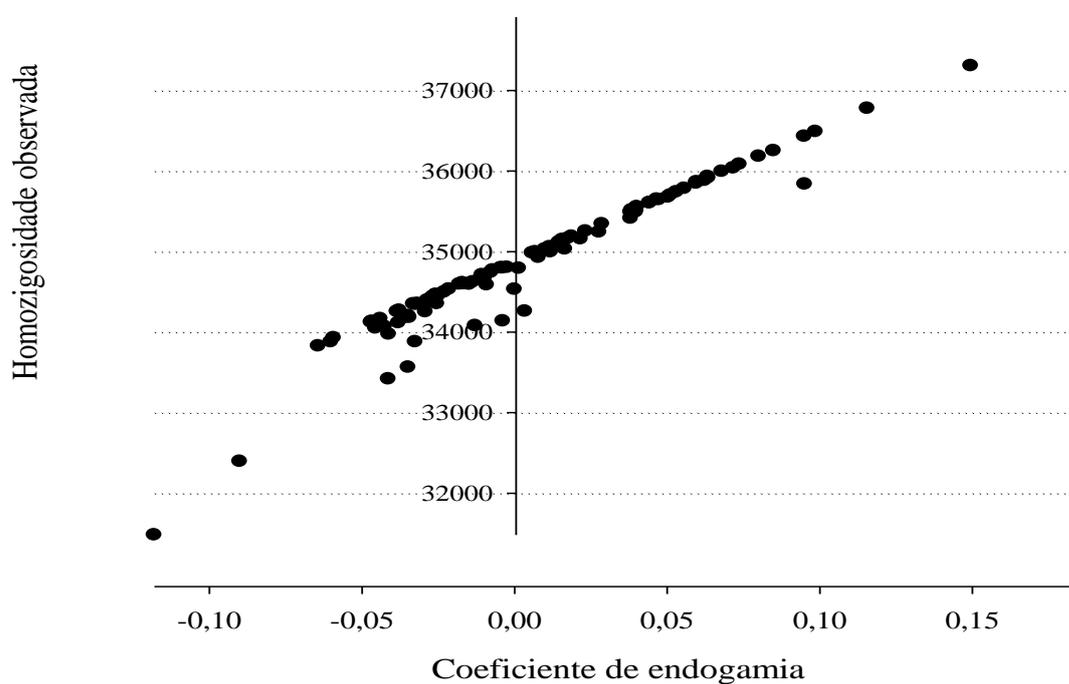


Figura 1. Valores de coeficientes individuais de endogamia baseados em homozigidades observadas na população de equinos da raça Mangalarga Marchador.

Os coeficientes de endogamia (F) calculados para cada indivíduo demonstraram que 51 indivíduos, ou seja, 51% da população total possuíam uma perda significativa de heterozigidade na raça Mangalarga Marchador, sendo o maior valor de F individual de 0,1859 (Figura 1), mínimo de -0,1181 e média de $0,0081 \pm 0,0496$. Como mencionado, valores negativos são possíveis, pois o resultado utiliza-se da estimação de frequências alélicas como descrito por Powell et al. (2010).

3.2 Diversidade genética e equilíbrio de Hardy-Weinberg

As médias de heterozigidades observadas (H_o) e esperadas (H_e), da frequência de alelos menores (MAF), do equilíbrio de Hardy-Weinber (EHW), o número de SNP em não EHW, bem como o número de SNP informativos e a taxa de genotipagem calculados por cromossomo da raça Mangalarga Marchador são apresentados na Tabela 2.

Tabela 2. Valores médios de heterozigosidades observadas (Ho) e esperadas (He), frequência dos alelos menores (MAF), equilíbrio de Hardy-Weinber (EHW), número de SNP informativos e de SNP em não EHW e taxa de genotipagem para cada cromossomo (ECA) da população de equinos da raça Mangalarga Marchador.

ECA	Ho	He	MAF	EHW (p-valor)	N.º de SNP em não EHW	N.º de SNP informativos	Taxa de genotipagem
1	0,3135	0,3164	0,2204	0,6087	56	4048	0,9959
2	0,3209	0,3239	0,2296	0,6223	34	2912	0,9957
3	0,3169	0,3227	0,2021	0,5870	27	2540	0,9969
4	0,3143	0,3149	0,2050	0,6305	32	2509	0,9953
5	0,3083	0,3056	0,2056	0,6259	26	2301	0,9964
6	0,3208	0,3228	0,2241	0,6004	36	2054	0,9957
7	0,3041	0,3071	0,1904	0,6181	27	2192	0,9959
8	0,3231	0,3310	0,2221	0,5761	35	2289	0,9958
9	0,3107	0,3077	0,2210	0,6272	22	1958	0,9966
10	0,3082	0,3138	0,2048	0,6083	29	1955	0,9951
11	0,3138	0,3210	0,2059	0,5998	28	1478	0,9962
12	0,3257	0,3261	0,2320	0,5728	21	754	0,9926
13	0,3114	0,3126	0,1976	0,5831	12	921	0,9940
14	0,3062	0,3074	0,2113	0,6068	28	2257	0,9960
15	0,3230	0,3202	0,1954	0,6107	22	2130	0,9963
16	0,3166	0,3211	0,2196	0,6143	24	2048	0,9960
17	0,3185	0,3221	0,2030	0,5974	29	1925	0,9958
18	0,3167	0,3190	0,2046	0,6234	23	1784	0,9963
19	0,3210	0,3200	0,2040	0,5838	18	1448	0,9958
20	0,3179	0,3136	0,2138	0,6091	25	1490	0,9958
21	0,3065	0,3068	0,2040	0,5890	20	1418	0,9962
22	0,3268	0,3270	0,2111	0,5971	23	1201	0,9954
23	0,3088	0,3195	0,2248	0,5837	24	1231	0,9962
24	0,2997	0,3136	0,1947	0,5788	20	1172	0,9958
25	0,3198	0,3203	0,2245	0,6146	15	906	0,9957
26	0,2940	0,3010	0,1997	0,5805	18	834	0,9960
27	0,3187	0,3245	0,1906	0,6070	18	927	0,9969
28	0,3132	0,3212	0,2044	0,5777	21	1028	0,9963
29	0,2987	0,3011	0,1844	0,6346	15	668	0,9963
30	0,3154	0,3153	0,1991	0,6182	15	726	0,9957
31	0,3180	0,3228	0,2104	0,5642	13	542	0,9952
Média	0,3097	0,3150	0,2010	0,5970	16,87	850,37	0,9960

ECA = cromossomo do *Equus caballus*.

A diversidade genética observada pela H_e foi menor no ECA26 com valor de 0,2940 e maior no ECA8 com valor de 0,3231, e média geral de 0,3150. O MAF médio para toda a população foi de 0,2010, sendo o MAF menor de 0,1844 no ECA29 e o maior de 0,2320 no ECA12. Desvios médios de equilíbrio de Hardy-Weinberg não foram observados por cromossomo, porém alguns SNP em desvio foram observados sendo a média de número de SNP em não EHW de 16,87 por cromossomo, com valor máximo de 56 no ECA1 e mínimo de 12 no ECA13. O número de SNP informativos foi alto no ECA2 com 2.912 marcadores e baixo para ECA31 com 542 marcadores, e média total de 850,37 SNP por cromossomo. As taxas de genotipagens foram altas com média de 0,9960, oscilando de 0,9969 no ECA27 a 0,9926 no ECA12 (Tabela 2).

As H_o e H_e , assim como o número de SNP em não EHW para cada raça equina estão apresentadas na Figura 2. A H_e variou de 0,1835 na raça Sorraia (SOR) a 0,3240 no Puro Sangue Inglês (PSI), e a H_o oscilou de 0,2159 no SOR e 0,3261 no PSI. O número de SNP em não equilíbrio de Hardy-Weinberg foi maior no Mangalarga Marchador (756) e menor no SOR (126).

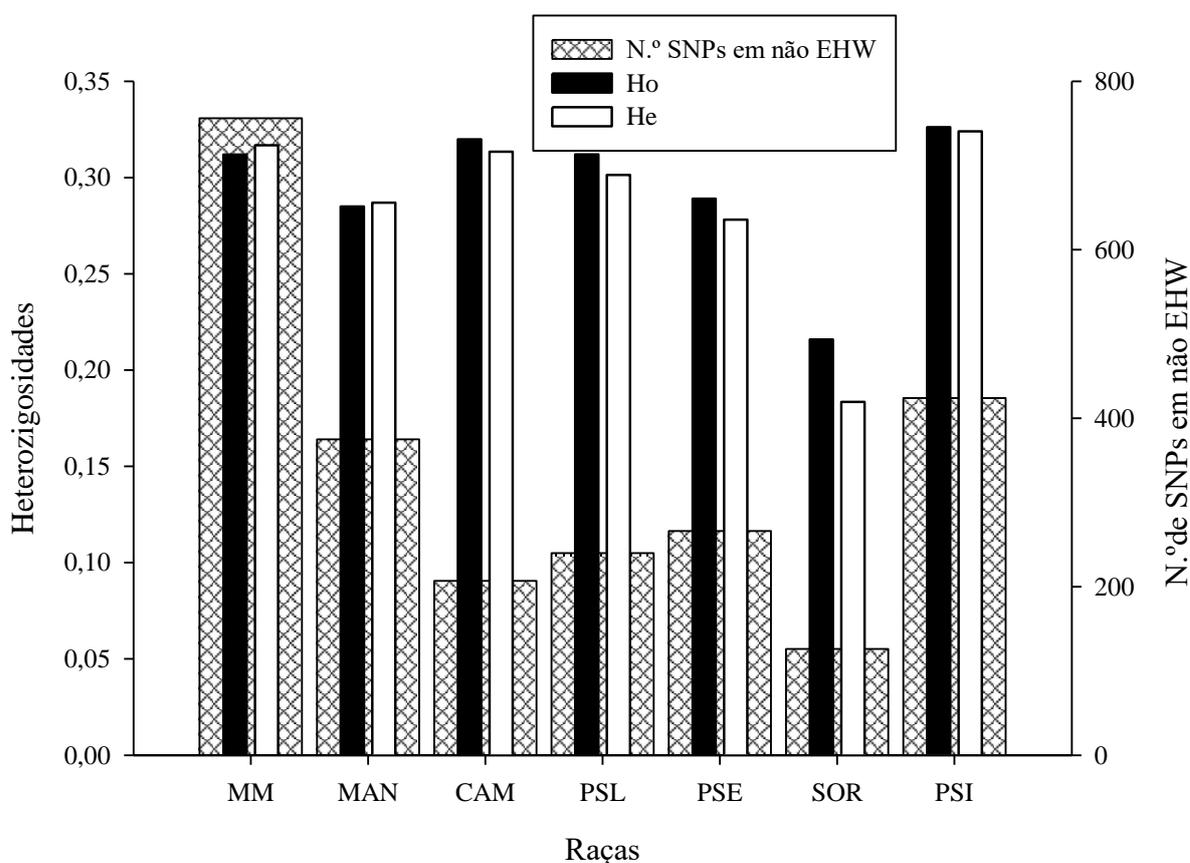


Figura 2. Valores médios de heterozigosidades observadas (H_o) e esperadas (H_e), número de SNP em não equilíbrio de Hardy-Weinberg nas populações de equinos Mangalarga Marchador e das demais raças brasileiras, ibéricas e inglesa. MM = Mangalarga Marchador;

MAN = Mangalarga; CAM= Campolina; PSL = Puro Sangue Lusitano; PSE = Puro Sangue Espanhol; SOR = Sorraia; PSI = Puro Sangue Inglês.

3.3 Desequilíbrio de ligação

Na maioria das raças analisadas, o desequilíbrio de ligação (DL) atingiu um platô em torno de 300 kb (Figura 3). A raça Sorraia apresentou maior extensão de DL e valores mais elevados de r^2 do que as outras raças ao longo de todas as distâncias que separam os *locus*. Entretanto, as raças Campolina e Mangalarga Marchador tiveram extensões menores e baixos valores de r^2 em curtas e em longas distâncias, ao compará-los com as raças restantes.

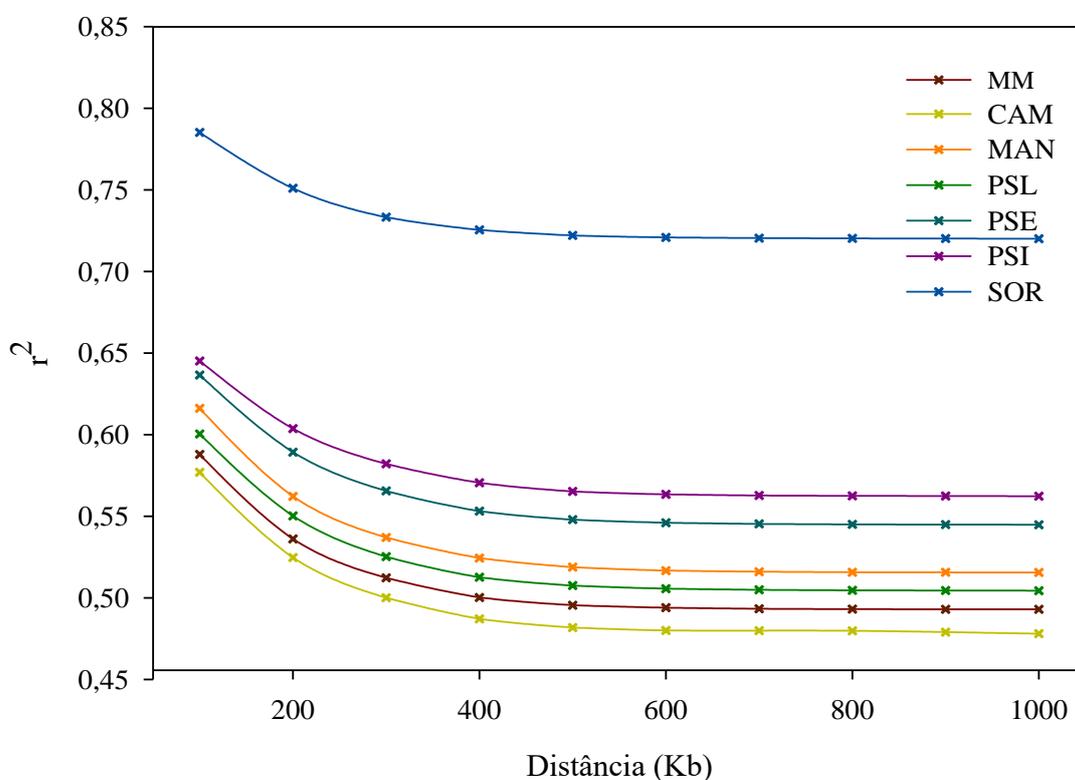


Figura 3. Mudança no desequilíbrio de ligação (r^2) entre os pares de marcadores com o aumento da distância a cada 100kb nas populações de equinos Mangalarga Marchador e das demais raças brasileiras, ibéricas e inglesa. MM = Mangalarga Marchador; MAN = Mangalarga; CAM= Campolina; PSL = Puro Sangue Lusitano; PSE = Puro Sangue Espanhol; SOR = Sorraia; PSI = Puro Sangue Inglês.

Observa-se na Figura 4 que a raça Sorraia apresentou o maior valor médio de r^2 no ECA26 (0,9272), e o menor valor de r^2 foi obtido no ECA31 (0,4303) na raça Campolina. Em particular, no Mangalarga Marchador, o maior valor de r^2 foi em obtido no ECA11 (0,53) e menor valor em ECA12 (0,4478).

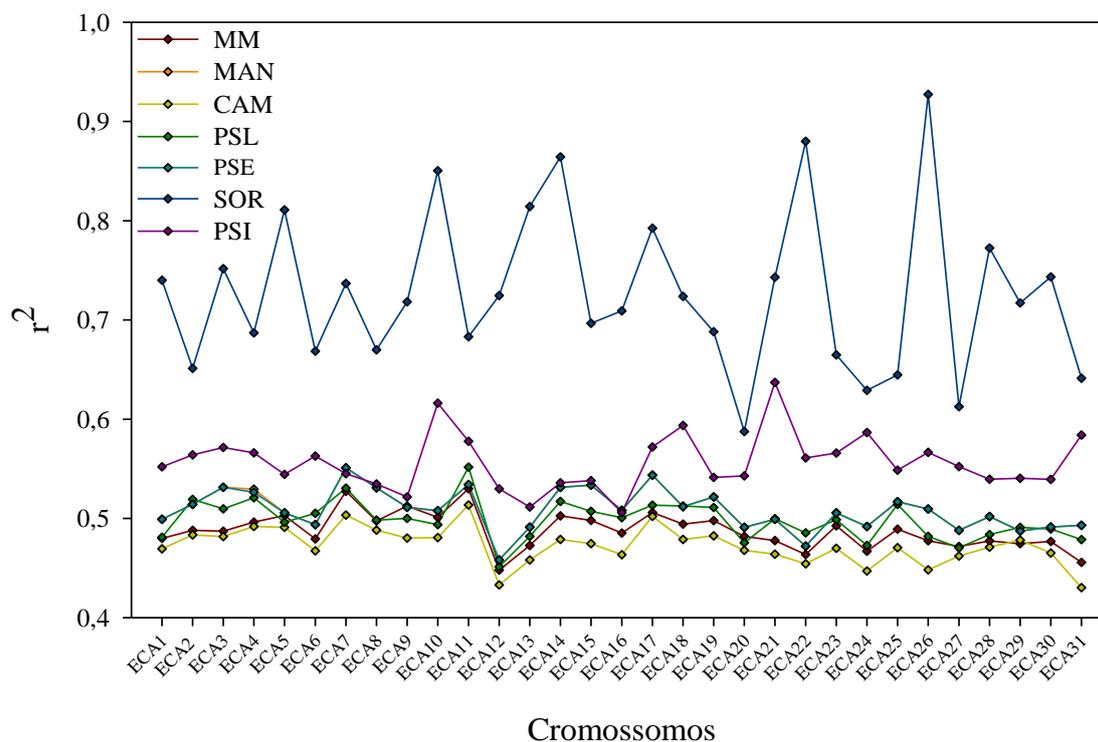


Figura 4. Valores médios de r^2 por cromossomo do genoma equino para as populações de equinos Mangalarga Marchador e das demais raças brasileiras, ibéricas e inglesa. MM = Mangalarga Marchador; MAN = Mangalarga; CAM= Campolina; PSL = Puro Sangue Lusitano; PSE = Puro Sangue Espanhol; SOR = Sorraia; PSI = Puro Sangue Inglês. ECA= cromossomo equino.

Alguns blocos de haplótipos semelhantes foram encontrados entre as raças de equinos do presente estudo. Nas raças Campolina e Puro Sangue Espanhol foi observado o mesmo haplótipo no cromossomo ECA1 formado pelos SNP BIEC_2-2597 e BIEC_2-2600, entre as posições (em pares de bases) 7100890 e 7101467, com tamanho de bloco de 0,578 kb. No Mangalarga Marchador (MM) também foi encontrado bloco de haplótipo semelhante ao do Mangalarga (MAN) (Figura 5) no cromossomo ECA1 formado pelos SNP UKUL1, UKUL2 e UKUL3, entre as posições 15656 e 30150 e tamanho de bloco de 14,495 kb.

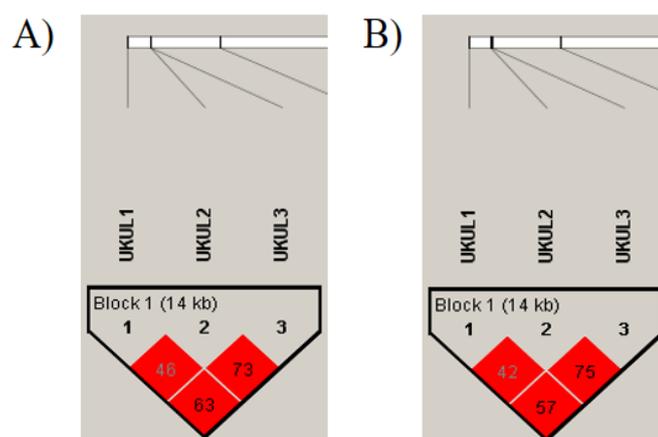


Figura 5. Blocos de haplótipos semelhantes entre as raças Mangalarga Marchador (A) e Mangalarga (B) formados pelos SNP UKUL1, UKUL2 E UKUL3 no ECA1 e seus respectivos graus de ligação (r^2).

As cores vermelhas indicam forte ligação entre os SNP dentro de cada bloco formado, o que é comprovado pelos altos valores de r^2 (Figura 5). De modo geral, foram obtidos maiores r^2 no MM (46, 63 e 73) do que no MAN (42, 57 e 75), e também observou-se diferenças na frequência dos alelos dentro de cada bloco. No MM a combinação de alelos mais frequente foi GCA (frequência de 0,515), seguidos de ATG (0,305), GTG (0,105) e GTA (0,075), no MAN a combinação de maior frequência foi ATG (0,417), seguidos de GCA (0,375), GTG (0,139) e GTA (0,069).

3.4 Diferenciação genética

Os valores de F_{st} foram calculados entre as sete populações de equinos (Figura 6). O menor nível de diferenciação foi encontrado entre as populações do Campolina com o Mangalarga Marchador ($F_{st} = 0,0411$), enquanto a maior divergência foi observada entre Puro Sangue Inglês e o Sorraia ($F_{st} = 0,2284$). De modo geral, o SOR predominou com as maiores divergências em relação às demais raças, e as raças brasileiras (MM, MAN e CAM) possuíram menor diferenciação entre si. O PSL ficou mais próximo de MM e MAN do que do CAM, já o PSE estava mais próximo de MM e CAM do que de MAN. O SOR, outro ancestral comum, ficou com maior diferenciação em relação às raças do Brasil.

	MM	MAN	CAM	PSL	PSE	SOR	PSI
MAN	0,0518						
CAM	0,0411	0,0704					
PSL	0,0801	0,1097	0,1512				
PSE	0,1025	0,1356	0,1003	0,1018			
SOR	0,1935	0,1927	0,1762	0,1938	0,2166		
PSI	0,1427	0,1572	0,1242	0,1359	0,1689	0,2284	

Fst
0,04-0,08
0,08-0,12
0,12-0,16
0,16-0,20
0,20 +

Figura 6. Valores médios de pares de Fst baseados em 50.953 SNP nas populações de equinos Mangalarga Marchador e das demais raças brasileiras, ibéricas e inglesa. MM = Mangalarga Marchador; MAN = Mangalarga; CAM= Campolina; PSL = Puro Sangue Lusitano; PSE = Puro Sangue Espanhol; SOR = Sorraia; PSI = Puro Sangue Inglês.

3.5 Estrutura populacional

A análise de estruturação populacional do Mangalarga Marchador com as demais raças equinas refletiu melhor agrupamento genômico com dez ($K=10$) principais alelos marcadores distribuídos pelas sete populações, como observado na Figura 7. A raça MM demonstra por essa análise, uma estrutura populacional pouca definida, assim como as demais raças brasileiras MAN e CAM, ao contrário das raças PSL, PSE, PSI e SOR que foram entidades genéticas bem mais definidas. A partir das colorações (ex.: laranja claro, amarelo e azul escuro) entre as raças brasileiras, observou-se o compartilhamento de alelos comuns entre elas, principalmente entre MM e MAN. Outras colorações (ex.: roxa e azul claro) também foram comuns entre MM e seus ancestrais como PSL e PSE, mas não se observaram semelhanças com o SOR. Curiosamente, no MM existiu uma coloração expressiva (verde-água) na qual não foi observada, com tal magnitude, nas demais raças brasileiras e de suas raças ancestrais.

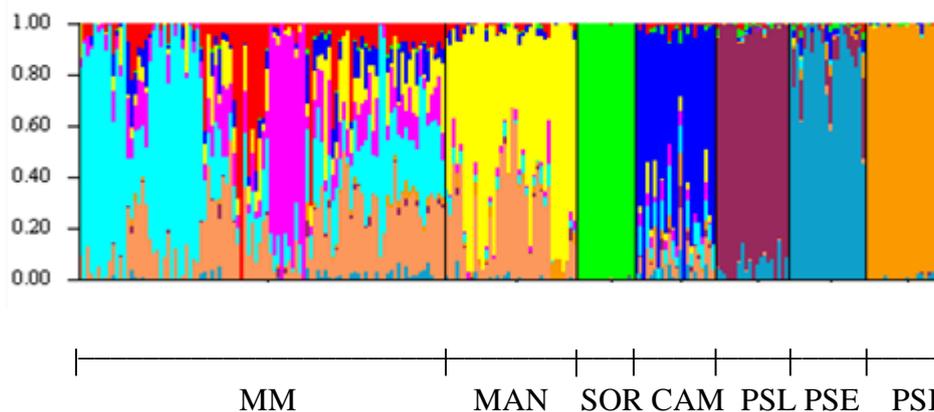


Figura 7. Agrupamento de indivíduos usando o método bayesiano com $K=10$ para as populações de equinos Mangalarga Marchador e demais raças brasileiras, ibéricas e inglesa. Cada indivíduo é representado por uma barra vertical e as raças são separadas por linhas verticais pretas. MM = Mangalarga Marchador; MAN = Mangalarga; CAM= Campolina; PSL = Puro Sangue Lusitano; PSE = Puro Sangue Espanhol; SOR = Sorraia; PSI = Puro Sangue Inglês.

As estruturas populacionais das raças equinas também foram verificadas usando a Análise de Componentes Principais (ACP). A análise discriminou em maior magnitude as raças SOR e PSI das demais raças, em moderada discriminação as raças PSL e PSE, e em menor discriminação as raças MM, MAN e CAM (Figura 8). Em geral, o gráfico ACP seguiu as observações sobre relacionamento entre as raças encontradas na análise de agrupamento Bayesiano (Figura 7).

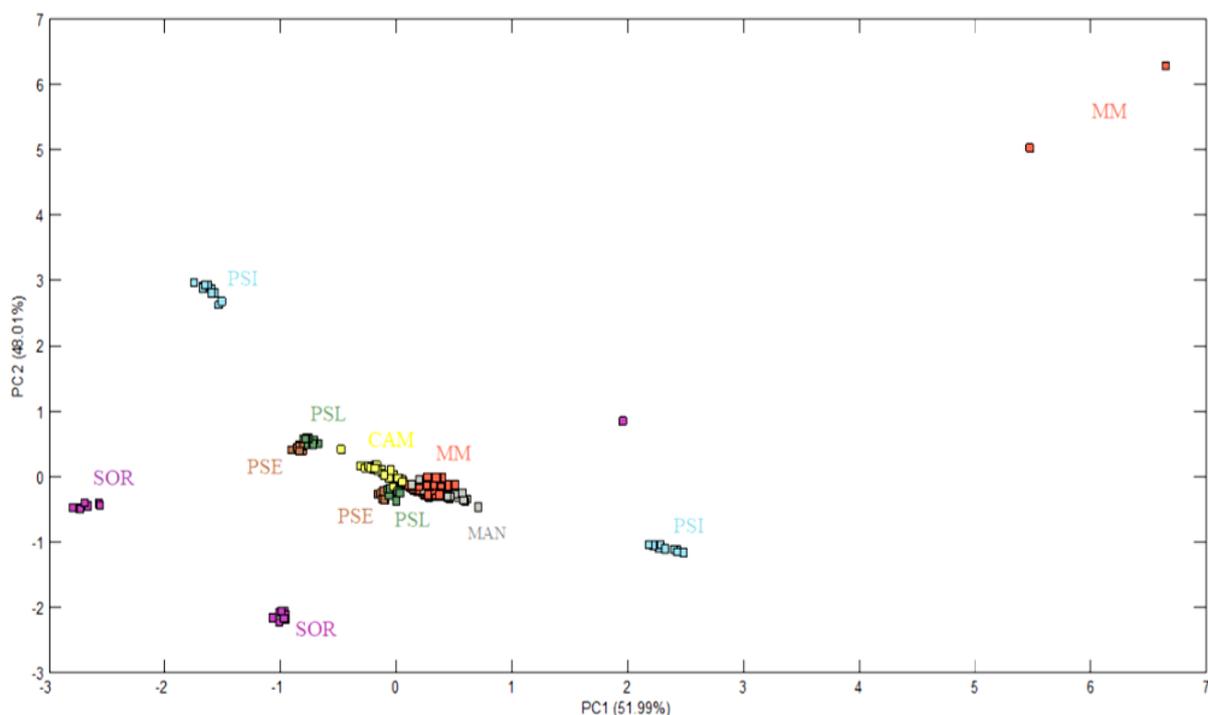


Figura 8. Análise de componentes principais das populações de equino Mangalarga Marchador e das demais raças brasileiras, ibéricas e inglesa. MM = Mangalarga Marchador; MAN = Mangalarga; CAM= Campolina; PSL = Puro Sangue Lusitano; PSE = Puro Sangue Espanhol; SOR = Sorraia; PSI = Puro Sangue Inglês.

4. DISCUSSÃO

A raça Mangalarga Marchador (MM) é considerada um patrimônio cultural vivo do Brasil, além de ter grande destaque na economia no país. Com isso, é prioridade para os órgãos competentes o monitoramento da raça e até mesmo a garantia de sua conservação. De acordo com Solis et al. (2005), em um contexto de estruturação populacional ou de conservação, a caracterização genética de uma raça é essencial para o desenvolvimento de estratégias de gestão genética corretas e pode ter implicações para futuras estratégias de melhoramento, sendo a informação genética molecular crucial para preservar a diversidade genética e prevenir a perda ou introgressão indesejada de alelos oriundos de outras raças.

Apesar da genômica ainda ser uma área relativamente jovem de conhecimento, suas conquistas não devem ser subestimadas. Com a notável redução nos custos das análises desta área, a genômica entrou na área de genética animal, especialmente nos temas de biodiversidade (Twito et al. 2007; Kijas et al. 2009), na produção animal (Hayes et al., 2009; Zhanz et al., 2013), na suscetibilidade a doenças (Zhang et al. 2012; Kizilkaya et al., 2013) e identificação de fatores genéticos associados às características relevantes (Cargill et al. 2008; Ren et al. 2011). Entretanto, a aplicação da genômica em termos de biologia pura, visando à

identificação dos fundamentos de funcionamento, da variabilidade e da estrutura do genoma dos animais é menos popular e muitas vezes ainda subestimada (Gurgul et al., 2014), sendo assim, a importância deste estudo, na qual não há relatos de caracterização genética de equinos da raça MM a nível genômico.

4.1 Homozigidade e endogamia genômica

Valores de coeficientes de endogamia baseados em homozigidades observadas e esperadas indicam que a metade dos equinos da raça Mangalarga Marchador apresentam sinais de endogamia (Figura 1). Valores altos de homozigidades na raça MM podem estar diretamente relacionados à exclusão de animais TAC na amostragem dos animais, o que reduziria a variabilidade possivelmente oriunda de um fluxo gênico.

A endogamia, se não monitorada, leva a uma perda de diversidade, principalmente em populações de tamanho limitado, e essa perda de diversidade quando devido a um gargalo populacional ou ao efeito fundador pode resultar também no aumento da endogamia de uma população (Petersen et al., 2014).

Baena (2019) observou na população do Mangalarga Marchador por análise de *pedigree* que há uso preferencial de alguns reprodutores na população e que 32% dos animais possuíam algum nível de endogamia e 47% da população recente (2005 a 2016) eram endogâmicos. O coeficiente de endogamia baseado em dados de genótipos está correlacionado com as estimativas baseadas em *pedigree*, mas geralmente é mais alta, mesmo considerando que as estimativas baseadas em genótipos podem ser negativas. Discrepâncias nos valores de endogamia por genótipo e *pedigree* podem ser atribuídas às várias características dos dados, sendo um deles a profundidade dos *pedigree* (Maccluer et al., 1983).

4.2 Diversidade genética e equilíbrio de Hardy-Weinberg

A heterozigidade esperada (H_e) corresponde à quantidade de heterozigidade presente em uma população, quando ocorre predominantemente acasalamentos ao acaso (Hedrick, 2005). Essa estimativa fornece informações sobre a estrutura e a história da população, e valores altos significam maior variabilidade genética (Butler, 2005; Nei, 1978).

Os valores de heterozigidades esperadas sugeriram pequena diversidade genética em todas as raças de equinos analisadas (Figura 2), sendo o maior valor de H_e obtido no Puro Sangue Inglês - PSI (0,3290), mas ainda considerado baixo. O PSI apesar de sua grande

população, é amplamente distribuído em uma escala geográfica, seus registros históricos sugerem que apenas um reprodutor é responsável por 95% das linhagens paternas e que apenas 30 fêmeas compõem 94% das linhagens maternas (Cunningham et al., 2001). Em adição, desde 1791 (Weatherby e Sons, 1791), a população tem sido amplamente fechada para o fluxo de genes externos devido ao fechamento do livro genealógico, além dos indivíduos da raça sofrerem forte pressão seletiva para o sucesso nas corridas, assim, a pequena diversidade da raça não é inesperada.

A menor He (0,1835), ou seja, a mais baixa diversidade foi observada principalmente na raça Sorraia, devido ao fato da raça ter pequeno tamanho censitário e ser uma raça relativamente antiga (Luís et al. 2007). Em particular, no MM, o valor encontrado de He (0,3150) foi surpreendente, pois esperava-se o contrário, por ser uma raça mais recente e ainda estar sobre uma fase de fixação como um grupo genético mais definido, além de provavelmente, estar ocorrendo algum fluxo gênico na população em virtude de possíveis cruzamentos com outras raças, e ser uma população relativamente grande com 600 mil animais registrados pela ABCCMM. Contudo, alta heterozigosidade na raça MM foi observada a partir de dados de marcadores microssatélites no trabalho de Baena et al. (2019b).

As razões para a inconsistência com nossa expectativa poderiam ser a alta pressão seletiva nos animais da raça para competição de marcha, e não podemos excluir, a possibilidade de que nossas amostras sub-representam o *pool* genético atual da raça MM, por ser um estudo inicial, embora tenhamos coletados animais não aparentados. Segundo Hendricks (2007) e Weatherley (1978), essa falta de diversidade é provavelmente o resultado de um severo gargalo populacional observado na maioria das raças equinas, que ocorreu com o início da industrialização e após a conclusão da Segunda Guerra Mundial.

Em contrapartida, os altos níveis de diversidade, medidos por He, normalmente ocorrem em raças mais antigas com grande tamanho populacional ou raças mais recentes com ocorrência de cruzamentos, além de populações de alta diversidade fenotípica e menor seleção / manejo artificial (Petersen et al., 2013), o que não foi observado neste estudo. No trabalho destes mesmos autores, valores semelhantes de He foram encontrados nas raças PSI (0,2480-0,3120) e PSL (0,2960-0,3150), e valores inferiores ao deste estudo na raça MAN (0,2350-0,2500), porém com um *Illumina Equine SNP50 Beadchip* e com tamanhos amostrais diferentes.

O número de SNP informativos, em média, em qualquer raça equina doméstica, é normalmente maior do que em outras espécies, como por exemplo em bovinos e cães (Matukumalli et al., 2009; Karlsson et al. 2007). Raças equinas com mistura recente ou ainda

em andamento, como o Quarto de Milha, Hanoverian e Swiss Warmblood, no trabalho de McCue et al (2012), também tiveram maiores números de SNP informativos e maiores médias de MAF, resultados estes semelhantes ao da raça MM. As taxas de genotipagens obtidas no MM foram semelhantes a outros equinos domésticos, sugerindo que o *Illumina Equine SNP70 BeadChip* é uma ferramenta útil para diferentes espécies do *Equus caballus*.

A deficiência heterozigótica, observados nos baixos valores de H_e , contribuem para alguns SNP em desvios de equilíbrio de Hardy-Weinberg, no qual foi observado em todos os cromossomos do MM (Tabela 2) e em todas as raças equinas analisadas (Figura 2). Os desvios de equilíbrio indicam a presença de cruzamento com outras raças, acasalamento preferencial não aleatório, fluxo gênico, endogamia ou mesmo a deriva genética. Desvios de equilíbrio na raça MM também foram observados em estudos com *locus* microssatélites por Silva Filho et al. (2007), Silva et al. (2012a), Silva et al. (2012b) e Baena et al. (2019b).

4.3 Desequilíbrio de ligação

O desequilíbrio de ligação (DL) é definido como a associação não aleatória de alelos em dois ou mais *loci* e é influenciada pela história populacional e de sua evolução (Ardlie et al. 2002; Khatkar et al. 2008). Em adição, o estudo de DL ao longo de um genoma pode ser usado para refletir os sistemas de reprodução e padrões de subdivisão geográfica, e o DL em regiões genômicas mais específicas oferece uma oportunidade de aprender mais sobre a história da seleção natural, conversão gênica, mutações e outros fatores que causam a evolução da frequência gênica (Slatkin, 2008). Regiões com baixa taxa de recombinação, como o cromossomo Y, partes do cromossomo X e regiões próximas ao centrômero são caracterizadas por alta extensão de DL. Por outro lado, a pequena extensão da DL é típica em regiões com alta taxa de recombinação, como a eucromatina e pequenas regiões conhecidas como pontos críticos de recombinação (Jeffreys et al. 2001).

Há grande diversidade de valores de DL entre os diferentes cromossomos, não somente no Mangalarga Marchador, mas também nas demais raças equinas deste estudo (Figura 4). A grande variabilidade nas taxas de recombinação nos cromossomos autossomos pode levar a uma considerável diversidade no padrão de DL em distintas regiões genômicas, de acordo com Ariais et al. (2009). Essa variação foi também encontrada em equinos da raça Quarto de Milha, da linhagem de corrida, no trabalho de Marchiori et al. (2018).

Marcadores moleculares do tipo SNP e blocos de haplótipos associados ao andamento marchado de equinos já foram observados no ECA11 (Staiger et al, 2016), sendo que este

mesmo cromossomo o de maior r^2 no MM (Figura 4). A região gênica do bloco no ECA11 inclui os genes SPNS3, SPNS2, MYBBP1A, GGT6 e SMTNL2, sendo este último gene um novo substrato para proteínas quinases ativadas por mitógenos, que executam processos celulares como a divisão celular, diferenciação e resposta ao estresse (Gordon et al., 2013). Esse gene é expresso em vários tecidos de mamíferos, mas tem alta expressão no músculo esquelético e pode estar envolvido na diferenciação miogênica (Gordon et al., 2013). A rigidez e a flexibilidade da estrutura do corpo do equino podem influenciar a marcha a ser realizada (Ziegler, 2005) e uma diferença genética na composição muscular pode predispor um cavalo a uma estrutura corporal mais fácil de manter e, portanto, um tipo de marcha preferencial.

Observa-se que para todas as raças deste estudo, os maiores valores de r^2 foram encontrados nos primeiros intervalos de 0 até 300 Kb, e os menores valores em distâncias maiores (Figura 3). Esses resultados são similares aos também encontrados por Marchiori et al (2018) à qual evidencia que o DL decai com aumento da distância, independentemente da raça analisada, e esse padrão de declínio inicial mais evidente seguido de menor intensidade ao longo de maiores distâncias foi também similar ao encontrado por Corbin et al. (2010), na raça Puro Sangue Inglês. No estudo de McCue e Mickelson (2013), a extensão do DL foi caracterizada em 14 diferentes raças equinas, e no geral, os valores calculados de r^2 foram inferiores a 0,2 em distâncias entre 100 e 150 Kb, o que difere deste estudo, que apesar do decaimento à medida que se aumentou a distância, os valores de r^2 não foram inferiores a 0,2 em menores distâncias e nem em maiores distâncias do genoma equino. O padrão de decaimento da curva de desequilíbrio de ligação ao longo de maiores distâncias é esperado, uma vez que com o aumento na distância física entre marcadores de um mesmo cromossomo, maior é a probabilidade de ocorrer recombinação genética, portanto, a quebra da ligação entre os marcadores (Du, Clutter e Lohuis, 2007).

A grande maioria dos estudos, com diversas espécies, que avaliaram o DL ao longo de distintas distâncias entre marcadores, observou resultados similares a deste estudo, ou seja, médias de r^2 maiores para distâncias menores (Ai, Huan e Ren, 2013; Mckay et al., 2007; Meadows, Chan e Kijas, 2008). De acordo com Pritchard e Przeworski (2001), a comparação dos resultados com outros estudos é complicada, pois ocorre variação no tamanho da amostra, diferenças no método de mensuração, diferença nos tipos e densidade de marcadores, bem como na história evolutiva de cada população.

Além disso, sabe-se que a extensão do DL em uma população depende da história da mesma, principalmente do tamanho efetivo da população. A raça Sorraia, por exemplo,

apresentou maior extensão de DL do que as demais raças (Figura 3). Isso provavelmente é causado pelo pequeno tamanho da população efetiva da raça e devido a ocorrência de gargalos populacionais severos. Entretanto, a menor extensão do DL observada no MM, sugere maior tamanho efetivo da população e esse fenômeno também pode estar relacionado a efeitos de mistura com outras raças, efeito fundador e alta pressão de seleção nos quais são comprovados pelo estudo de Baena et al. (2019a). De acordo com Touru et al. (2016), valores baixos de r^2 em curtas e longas distâncias indicam a existência de subpopulações dentro de raça, o que é confirmado pela análise do STRUCTURE e ACP (Figuras 7 e 8).

As raças PSI, PSL, PSE, e SOR foram domesticadas a muito mais tempo se comparadas com as raças brasileiras. Assim, mais gerações de recombinação levaram a segmentos cromossômicos mais curtos e idênticos nessas raças e conseqüentemente extensões de DL maiores (Figura 3), mas curiosamente, o MAN obteve padrão de extensão semelhante ao PSE por exemplo, mesmo sendo uma raça mais recente. O segundo maior DL foi claramente observado no PSI, refletindo a baixa diversidade e a alta endogamia da raça. Resultados altos de DL nessa mesma raça também foram relatados por McCue et al. (2012) e Corbin et al. (2010).

O DL é variável no genoma e depende da espécie, da história da população sob estudo (Oraguzie et al. 2007), do tamanho da amostra, do tipo de marcador (microsatélites ou SNP), da densidade e distribuição desses marcadores, do tipo de método utilizado e do rigor da filtragem de SNP. Com isso, a quantificação do DL é de importante utilidade em estudos evolutivos e também possibilita determinar a resolução de mapeamento e a densidade de marcadores necessários na identificação de *Quantitative Trait Loci* - QTL (Norborg et al., 2002; Wei, Jackson; McIntyre, 2006; Zhu et al., 2008), porém é de difícil comparação entres os estudos, como já mencionado.

Segundo Wade et al. (2009), os equinos exibem mais compartilhamentos de haplótipos do que outras espécies, como os cães, por exemplo. Os estudos destes autores demonstram que a história da população de equinos, de modo geral, levou ao compartilhamento desses haplótipos entre as raças, aumentando a possibilidade de mapeamento entre elas, e que os futuros projetos de mapeamento de equinos provavelmente irão se acelerar nos próximos anos e identificarão mutações em genes que são relacionados às características importantes de morfologia, de imunologia e de metabolismo. Haplótipos compartilhados também foram encontrados entre PSI e Quarto de Milha, da linhagem de corrida, por Petersen et al. (2014), resultado este não inesperado, pois o PSI foi fundamental para o desenvolvimento do Quarto de Milha. Assim como, o resultado deste estudo (Figura 5), no qual o MAN foi uma raça

desenvolvida a partir do MM, e com possível fluxo gênico ainda presente entre essas populações.

4.4 Diferenciação genética

O F_{st} reflete o nível de diferenciação genética que ocorre entre subpopulações dentro ou entre raças. De acordo com Janova et al. (2013), valores mais baixos de F_{st} , baseados em marcadores como os SNP, indicam menores distâncias genéticas entre as populações. Os valores relativos (entre raças) de F_{st} são informativos, mas esses valores e dados, que representam raças recentes com estoque limitado de indivíduos, e submetidos a intensas pressões seletivas artificiais, não podem ser usados independentemente para elucidar a evolução de uma raça. Além disso, como é o caso de outras análises de relacionamento, que podem dar uma ideia sobre a história de uma raça, elas não podem distinguir uma mistura recente de uma ancestralidade compartilhada (Petersen et al. 2013).

Em geral, as menores divergências, medidas por F_{st} , ocorreram entre as raças brasileiras e as maiores predominaram entre a raça Sorraia e as demais raças analisadas (Figura 6). Valores baixos de F_{st} entre raças brasileiras iguais a deste estudo também foram observados no estudo de Baena et al. (2019), porém com dados microssatélites.

A raça SOR, em comparação com as outras raças deste estudo, obteve menor diversidade genética (Figura 2), o que era esperado, pois assim é menos provável que a mesma compartilhe a variação genética com outra raça por acaso, especialmente se a raça é também relativamente homogênea. Por outro lado, raças com maiores diversidades, comparativamente, apresentam menores níveis de divergência. Segundo Petersen et al. (2013), esse resultado faz sentido, já que a variação dentro de uma raça pode indicar cruzamentos com outras populações, e também faz com que essas raças tenham maior probabilidade de compartilhar a sua variação com outras raças ao acaso.

4.5 Estrutura populacional

A análise de componentes principais (ACP), tem sido ferramenta eficiente para a detecção da estrutura populacional no estudo de espécies como os equinos (Costa et al., 2005), suínos (Ai; Huan; Ren, 2013), ovinos (Kijas et al., 2012), e humanos (Patterson; Price; Reich, 2006). Pelo ACP, os equinos brasileiros agruparam-se e ficaram próximos de seus

ancestrais como PSL e PSE, mas distantes do SOR (Figura 8). Já as raças PSI e SOR estabeleceram-se como grupos bem definidos e separados dos demais. A ACP no estudo de Baena (2019) também agrupou as mesmas raças brasileiras utilizadas neste trabalho. Observa-se também uma dispersão de valores dentro de raças, com a formação de dois grupos do SOR, PSI, PSL, PSE, e alguns indivíduos do MM foram bem discrepantes de seu grupo, o que evidencia subestruturas dentro dessas raças. Esse fato pode ser explicado pela alteração de contribuições genéticas dos principais fundadores da raça MM ao longo dos anos, como observado no trabalho de Baena (2019).

As raças brasileiras estudadas são descendentes dos mesmos ancestrais e possivelmente sempre houve fluxo gênico entre essas populações, além de aproximação geográfica das mesmas, portanto, era concebível esse resultado. Essa hipótese é apoiada pela análise do programa STRUCTURE (Figura 7), que revelou o sinal de mistura entre as raças brasileiras, e alelos ainda presentes de seus ancestrais (PSL e PSE), mesmo utilizando amostras de animais rigorosamente selecionadas para evitar a utilização de animais mestiços no estudo. A coloração mais específica encontrada no MM pode ser devida a outra raça participante de sua formação, como por exemplo os equinos berberes. Em adição, as raças MM, MAN e CAM provavelmente ainda não tiveram tempo de passar pelos processos evolutivos necessários para serem populações geneticamente distintas e definidas, como é observado em raças com histórias evolutivas mais longas e com livros fechados. A partir da mesma análise, resultados semelhantes no trabalho de Baena et al (2019b) com microssatélites comprovam o compartilhamento de alelos entre as raças brasileiras CAM, MAN e MM e a falta de estruturação populacional dessas raças.

Estruturação populacional dentro da raça pode ser preocupante se sua causa é resultado ou acompanhado por um aumento na endogamia, que tem o potencial de levar a um aumento da incidência de características indesejáveis (Marchiori et al. 2018). Esse fato coloca a raça MM em estado de alerta, com relação aos seus acasalamentos, pois metade da população estudada estava com algum nível de endogamia (Figura 1). Da mesma forma que neste estudo, um trabalho a partir de informações de *pedigree* de animais da raça Mangalarga Marchador evidenciou, a partir de ACP, a existência dessas subestruturas na população, na qual os autores sugeriram a existência de cinco grupos formadores da raça (Costa et al., 2005). Peterson et al. (2013), ao estudarem SNP em análise populacional, também encontraram subestruturas na raça PSL, assim como no Quarto de Milha, nos trabalhos de Tryon et al. (2009), Meira et al. (2013) e Petersen et al. (2014).

Esses métodos demonstram o poder e utilidade dos marcadores SNP como uma ferramenta para reconstruir relações de *pedigree* desconhecidas, pré-históricas ou mesmo errôneas. Segundo Speed e Balding (2015), as medidas de relacionamento derivadas de dados genômicos são mais precisas, pois são derivadas da identidade pelo compartilhamento de alelos por estado, ao invés da probabilidade de compartilhamento por descendência, o que é predito a partir de *pedigree*.

5. CONCLUSÃO

É necessário na população do Mangalarga Marchador, o monitoramento contínuo e regular da diversidade genética com o uso de marcadores moleculares, além de registros mais precisos de *pedigree*, maiores controles no número de pais usados para as próximas gerações, nos acasalamentos e nos cruzamentos com raças geneticamente relacionadas, sendo indispensável a implementação imediata de um programa de melhoramento genético na raça. Fatores estes que visam maximizar sua variabilidade genética, conservar genes úteis e estruturar a raça como uma entidade genética definida para que se possa futuramente diferenciá-la de demais raças equinas.

Este conjunto de dados representa o primeiro estudo específico sobre caracterização genética do equino Mangalarga Marchador por meio da genotipagem de SNP em larga escala e fornece uma visão abrangente da diversidade genética da raça e suas relações com outras populações equíneas.

AGRADECIMENTOS

Os autores agradecem à Universidade Federal de Lavras (UFLA), Universidad Nacional de La Plata, a CAPES, ao CNPq e ABCCMM.

REFERÊNCIAS

- Ai, H.; Huang, L., Ren, J. (2013). Genetic Diversity, Linkage Disequilibrium and Selection Signatures in Chinese and Western Pigs Revealed by Genome-Wide SNP Markers. PLoS ONE, 8, 560-01.
- Ardlie KG, Kruglyak L, Seielstad M (2002). Patterns of linkage disequilibrium in the human genome. Nat Rev Genet., 3, 299–309.

- Arias, J. A., Keehan, M., Fisher, P., Coppieters, W., Spelman, R. (2009). A high density linkage map of the bovine genome. *BMC Genetics*, 10, 18.
- Assis, J.B., Laat, D.M., Peixoto, M.G.C.D., Bergmann, J.A.G., Fonseca, C.G., Carvalho, M.R.S. (2009). Genetic diversity and population structure in Brazilian Mangalarga Marchador horses. *Genetics and Molecular Research*, 8, 1519-1524.
- Barrett, J.C., Fry, B., Maller, J., Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps, *Bioinformatics*, 21, 263–265. <https://doi.org/10.1093/bioinformatics/bth457>.
- Boichard, D. (2002). Pedig: a Fortran package for pedigree analysis suited for large populations. In: *World Congress on Genetics Applied to Livestock Production*, 7., 2002, Montpellier. Proceedings. Montpellier: Institut National de la Recherche Agronomique, 13-28.
- Butler, J.M. (2005). *Forensic DNA Typing: Biology Technology and Genetics of STR Markers*, (2nd ed.). Maryland, Academic Press.
- Cargill, E.J., Nissing, N.J., Grosz, M.D. (2008). Single nucleotide polymorphisms concordant with the horned/polled trait in Holsteins. *BMC Res Notes*, 1, 128.
- Corbin, L. J., Blott, S. C., Swinburne, J. E. (2010). Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Anim Genet*, 41, 8- 15.
- Costa, M. D., Bergmann, J. A. G., Rezende, A. S. C., Fonseca, C. G., Faria, F. J. C. (2005). Estudo da subdivisão genética da raça Mangalarga Marchador. *Arq. Bras. Med. Vet. Zootec.*, 57,272-280.
- Cunningham, E.P., Dooley, J.J., Splan, R.K., Bradley, D.G., (2001). Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to thoroughbred horses. *Anim. Genet.*, 32, 360–364. <http://dx.doi.org/10.1046/j.1365-2052.2001.00785.x>.
- Du, F-X., Clutter, A. C., Lohuis, M. M.(2007). Characterizing Linkage Disequilibrium in Pig Populations. *International Journal of Biological Sciences*, 7, 166-178.
- Gabriel, S.B., Schaffner, S.F, Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S.N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Eric, S. L., Daly, M.J., David, A. (2002). The Structure of Haplotype Blocks in the Human Genome. *Science*, 296, 2225-2229. [http://dx doi: 10.1126/science.1069424](http://dx.doi.org/10.1126/science.1069424).
- Gordon, E. A., T. C. Whisenant, M. Zeller, R. M. Kaake, W. M. Gordon, P. Krotee, V. Patel, L. Huang, P. Baldi, and L. Bardwell. (2013). Combining docking site and phosphosite predictions to find new substrates: Identification of smoothelin-like-2 (SMTNL2) as a c-Jun N-terminal kinase (JNK) substrate. *Cell. Signal.* 25, 2518–2529. [doi:10.1016/j.cellsig.2013.08.004](https://doi.org/10.1016/j.cellsig.2013.08.004).

- Gurgul, A., Semik, E., Pawlina, K., Szmatola, T., Jasielczuk, I., Bugno-Poniewierska, M. (2014). The application of genome-wide SNP genotyping methods in studies on livestock genomes. *J Appl Genetics.*, 55, 197-208. . [http://dx doi.10.1007/s13353-014-0202-4](http://dx.doi.10.1007/s13353-014-0202-4)
- Hayes, B.J., Chamberlain, A.J., Maceachern, S., Savin, K., McPartlan, H., MacLeod, I., Sethuraman, L., Goddard, M.E. (2009) A genome map of divergent artificial selection between *Bos taurus* dairy cattle and *Bos taurus* beef cattle. *Anim Genet.*, 40, 176–184.
- Hedrick, P.W. (2005). *Genetics of populations.* (3rd ed.). Massachusetts. Jones and Bartlett Publishers, 736.
- Hendricks, B.L. (2007). *International Encyclopedia of Horse Breeds.* Norman, OK: University of Oklahoma Pres, 486.
- Hill, W. G., Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, 38, 226-231.
- Janova, E., Futas, J., Klumplerova, M., Putnova, L., Vrtkova, I., Vyskocil. M., Frolkova, P., Horin, P. Horin. (2013). Genetic diversity and conservation in a small endangered horse population. *J. Appl. Genetics*, 54, 285–292. <http://dx.doi.10.1007/s13353-013-0151-3>.
- Jeffreys, A.J., Kauppi, L., Neumann, R. (2001). Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat. Genet.*, 29, 217–222.
- Karlsson, E.K., Baranowska, I., Wade, C.M., Salmon, Hillbertz N.H., Zody, M.C, et al. (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat. Genet.*, 39,1321–1328.
- Khatkar, M.S., Nicholas, F.W., Collins, A.R., Zenger, K.R., Cavanagh, J.A.L., Barri,s W., Schnabel, R.D., Taylor, J.F., Raadsma, H.W. (2008). Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC Genomics*, 9,187.
- 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., Oddy, V.H., Nicholas, F.W., Raadsma, H. (2009). International Sheep Genomics Consortium A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS One*, 4,46-68.
- Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., et al. (2012). Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLoS ONE*, 10, 100-1258.
- Kizilkaya, K., Tait, R.G., Garrick, D.J., Fernando, R.L., Reecy, J.M. (2013). Genome-wide association study of infectious bovine kerato conjunctivitis in Angus cattle. *BMC Genet.* 14, 23.
- Lanteri, S., Barcaccia, G., Ruane, J., Sonnino, A. (2006). Molecular marker based analysis for crop germplasm preservation. In: *The Role of Biotechnology for the Characterisation and*

Conservation of Crop, Forestry, Animal and Fishery Genetic Resources. Turin, Italy, 5–7 March 2005, 105–120.

- Lima, R. A. S., Cintra, A. G. (2016). Revisão do Estudo do Complexo do Agronegócio do Equino. Brasília: Ministério da Agricultura, Pecuária e Abastecimento, 56 .
- Luis, C., Cothran, E.G., Oom, M.M. (2007). Inbreeding and genetic structure in the endangered Sorraia horse breed: implications for its conservation and management. *J. Hered.*, 98, 232-237.
- MacCluer, J.W., Boyce, A.J., Dyke, B., Weitkamp, L.R., Pfennig, D.W., Parsons, C.J. (1983). Inbreeding and pedigree structure in Standardbred horses. *J Hered.*, 74, 394–399.
- Marchiori, C. , Pereira, G. L. , Maiorano, A. M. , Rogatto, G. M. , Assoni, A. D. , Silva, J. A. II V. , Chardulo, L. A. L. , Curi, Rogério Abdallah . (2018). Linkage disequilibrium and population structure characterization in the cutting and racing lines of Quarter Horses bred in Brazil. *Livestock Science*, 219, 45-51.
- Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., et al. (2009). Development and Characterization of a High Density SNP Genotyping Assay for Cattle. *PLoS ONE*, 4, 53-50. <https://doi.org/10.1371/journal.pone.0005350>.
- McCue, M.E., Bannasch, D.L., Petersen, J.L., Gurr, J., Bailey, E., et al. (2012). A High Density SNP Array for the Domestic Horse and Extant Perissodactyla: Utility for Association Mapping, Genetic Diversity, and Phylogeny Studies. *PLoS Genet.*, 8, 100-2451. <https://doi.org/10.1371/journal.pgen.1002451>.
- McCue, M., Mickelson, J. (2013). Genomic Tools and Resources: Development and Applications of an Equine SNP Genotyping Array. In: Chowdary, B.H. (ed.) *Equine Genomics*, Ames: Blackwell Publishing Ltd, 113–14.
- McKay, S. D., Schnabel, R. D., Murdoch, B. M., Matukumalli, L. K., Aerts, J., Wouter Coppieters, W., Crews, D., Dias Neto, E., Gill, C. A., Gao, C., Mannen, H., Stothard, P., Wang, Z., Van Tassell, C.P., Williams, J. L., Taulor, J. F., Moore, S. S. (2007) Whole genome linkage disequilibrium maps in cattle. *BMC Genetics*, 8, 74.
- Meadows, J. R. S., Chan, E. K. F, Kijas, J. W. (2008). Linkage disequilibrium compared between five populations of domestic sheep. *BMC Genetics*, 9, 61.
- Meira, C.T., Curi, R.A., Silva, J. A.V., Correa, M.J.M., Oliveira, H.N., Mota, M.D.S. (2013). Morphological and genomic differences between cutting and racing lines of quarter horses. *J. Equine Vet. Sci.*, 33,244–249.
- Meira, C.T. (2010). Avaliação de características morfofuncionais de equinos da raça Mangalarga Marchador. 2010. 48 f. Dissertação (Mestrado em Zootecnia) - Faculdade de Ciências Agrárias, Universidade Federal dos Vales do Jequitinhonha e Mucuri. Diamantina.
- Nascimento, J.F. (1999). Mangalarga Marchador: Tratado morfofuncional. Belo Horizonte: Associação Brasileira de Criadores de Equino Mangalarga Marchador. 900.

- Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89, 583-590.
- Norborg, 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. (2002). The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.*, 30, 190–193.
- Oraguzie, N. C., Rikkerink, E. H. A., Gardiner, S. E., Silva, H. N. de. (Eds.) (2007). *Association Mapping in Plants*. New York: Springer, 11-39.
- Patterson, N., Price, A. L., Reich, D. (2006). Population structure and eigenanalysis. *Plos Genetics*, 2, 2074-2093.
- Pereira, G. L., Chud, T., Bernardes, P., Venturini, G., Chardulo, L. A., Curi, R. (2017). Genotype Imputation and Accuracy Evaluation in Racing Quarter Horses Genotyped Using Different Commercial SNP Panels. *Journal of Equine Veterinary Science*, 58, 89-96.
- Petersen, J. L., Mickelson, J. R., Leary, K. D., McCue, M. E. (2014). The American Quarter Horse: Population Structure and Relationship to the Thoroughbred. *Journal of Heredity*, 105, 148–162. <https://doi.org/10.1093/jhered/est079>.
- Petersen, J.L., Mickelson, J.R., Cothran, E.G., Andersson, L.S., Axelsson, J., et al. (2013). Genetic Diversity in the Modern Horse Illustrated from Genome-Wide SNP Data. *PLoS ONE*, 8,54-997. <https://doi.org/10.1371/journal.pone.0054997>.
- Powell, J.E., Visscher, P.M., Goddard, M.E. (2010). Reconciling the analysis of IBD and IBS in complex trait studies. *Nat Rev Genet.*, 11,800–805.
- Pritchard, J.K., Stephens, M., Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–59.
- Pritchard, J. K., Przeworski, M. (2001). Linkage disequilibrium in humans: models and data. *American Journal of Human Genetics*, 69, 1–14.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.*, 81, 559–575.
- Ren, J., Mao, H., Zhang, Z., Xiao, S., Ding, N., Huang, L. (2011). A6-bp deletion in the TYRP1 gene causes the brown colouration phenotype in Chinese indigenous pigs. *Heredity (Edinb)* 106,862–868.
- Slatkin, M. (2008). Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genet.*, 9,477–485.
- Silva, E. A. (2012a). Avaliação de um painel de microssatélites para identificação animal/averiguação de paternidade e estrutura genética da população de equinos da raça Mangalarga Marchadro. 2012a. Dissertação (Mestrado em Genética do Instituto de

Ciências Biológicas da Universidade Federal de Minas Gerais). Belo Horizonte. Brasil. 73.

Silva, A.C., Paiva S.R., Albuquerque M.S., Egito A.A., Santos S.A., Lima F.C., Castro S.T., Mariante A.S., Correa P.S., McManus C.M. (2012b). Genetic variability in local Brazilian horse lines using microsatellite markers. *Genet. Mol. Res.*, 10, 881-90. <https://doi.org/10.4238/2012>.

Silva Filho, E., Schneider, M.P.C., Silva, A.L.C. (2007). Variabilidade genética de equinos baseada em DNA microssatélites. *Rev. Trop. Cienc. Agrar. Biol.*, 1, 76-87.

Soattin, M., Barcaccia, G., Dalvit, C., Cassandro, M., Bittante, G., (2009). Genomic DNA fingerprinting of indigenous chicken breeds with molecular markers designed on interspersed repeats. *Hereditas*. 146, 183–197.

Solis, A., Jugo, B.M., Mériaux, J.C., Iriondo, M., et al. (2005). Genetic diversity within and among four South European Native Horse breeds based on microsatellite DNA analysis: implications for conservation. *J. Hered.*, 96, 670-678.

Speed, D., Balding, D.J. (2015). Relatedness in the post-genomic era: is it still useful? *Nat Rev Genet.*, 16, 33–44.

Staiger, A.E., Abri, M.A., Silva, C. A. S., Brooks, S.A. (2016). Loci impacting polymorphic gait in the Tennessee Walking Horse. *J. Anim. Sci.* 94, 1377–1386. doi:10.2527/jas2015-9936.

Tambasco, D. D., M. M. A., Coutinho, L. L., Tambasco, A. J., Tambasco, M. D., Regitano, L. C. A. (2000). Caracterização Molecular de Animais da raça Nelore utilizando Microssatélites e Genes Candidatos. *Rev. bras. zootec.*, 4, 1044-1049.

Touru, T. I. , Tapio, M., Vilkki, J., Kiseleva, T., Ammosov, I., Ivanova, Z., Popov, R., Ozerov, M., Kantane J. (2016). Genetic diversity and genomic signatures of selection among cattle breeds from Siberia, eastern and northern Europe. *Animal Genetics*, 47, 647–657. <https://doi.org/10.1111/age.12473>.

Tryon, R.C., Penedo, M.C.T., McCue, M.E., Valberg, S.J., Mickelson, J.R., Famula, T.R., Wagner, M.L., Jackson, M., Hamilton, M.J., Nooteboom, S. et al. (2009). Evaluation of allele frequencies of inherited disease genes in subgroups of American Quarter Horses. *J Am. Vet. Med. Assoc.*, 234, 120–125.

Twito, T., Weigend, S., Blum, S., Granevitze, Z., Feldman, M.W., Perl-Treves, R., Lavi, U., Hille, L. J. (2007). Biodiversity of 20 chicken breeds assessed by SNP located in gene regions. *Cytogenet Genome Res.* 117, 319–326.

Ziegler, L. (2005). *Easy-gaited horses*. 1st ed. Storey Publishing, North Adams, MA.

Zhang, H., Wang, Z., Wang, S., Li, H. (2012) Progress of genome wide association study in domestic animals. *J. Anim. Sci. Biotechnol*, 3, 26.

- Zhang, L., Liu, J., Zhao, F., Ren, H., Xu, L., Lu, J., Zhang, S., Zhang, X., Weir, C., Lu, G., Zheng, Y., Du, L. (2013). Genome-wide association studies for growth and meat production traits in sheep. *PLoS One*, 8, 66-569.
- Zhu, C., Gore, M., Buckler, E. S., Yu, J. (2008). Status and prospects of association mapping in plants. *The Plant Genome*, Madison, 1, 5-20, 2008.
- Wade, C.M., Giolotto, E., Sigurdsson, S., Zoli, M., Gnerre, S. et al. (2009). Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science*, 326, 865–867. <https://doi.org/10.1126/science.1178158>.
- Weatherby and Sons (1791). *An Introduction to a General Stud Book*. London, UK.
- Weatherley, L. (1978). *Great horses of Britain*. Hindhead: Spur Publications. 269.
- Wei, X., Jackson, P. A., McIntyre, C. L. (2006). Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure. *Theoretical and Applied Genetics*, 114, 155-164.
- Weir, B.S., Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, 38, 1358-1370.