

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

**VARIANTES DA REGIÃO CROMOSSÔMICA DO GENE DA
MIOSTATINA E SUAS RELAÇÕES COM LINHAGENS,
DESEMPENHO E MEDIDAS CORPORAIS NA RAÇA
QUARTO DE MILHA**

**Rafael de Matteis
Zootecnista**

**Jaboticabal
2017**

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

**VARIANTES DA REGIÃO CROMOSSÔMICA DO GENE DA
MIOSTATINA E SUAS RELAÇÕES COM LINHAGENS,
DESEMPENHO E MEDIDAS CORPORAIS NA RAÇA
QUARTO DE MILHA**

**Rafael de Matteis
Orientador: Prof. Dr. Rogério Abdallah Curi**

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

Jaboticabal

2017

Matteis, Rafael de
M435v Variantes da região cromossômica do gene da miostatina e suas
relações com linhagens, desempenho e medidas corporais da raça
quarto de milha / Rafael de Matteis. – – Jaboticabal, 2017
iv, 50 p. : il. ; 29 cm

Dissertação (mestrado) - Universidade Estadual Paulista,
Faculdade de Ciências Agrárias e Veterinárias, 2017
Orientadora: Rogério Abdallah Curi
Banca examinadora: Guilherme Costa Venturini, Simone Cristina
Méo Niciura
Bibliografia

1. Equino. 2. ERE1. 3. Marcadores moleculares. I. Título. II.
Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 636.082:636.1

Ficha catalográfica elaborada pela Seção Técnica de Aquisição e Tratamento da Informação –
Diretoria Técnica de Biblioteca e Documentação - UNESP, Câmpus de Jaboticabal.



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal



CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: VARIANTES DA REGIÃO CROMOSSÔMICA DO GENE DA
MIOSTATINA E SUAS RELAÇÕES COM LINHAGENS,
DESEMPENHO E MEDIDAS CORPORAIS NA RAÇA QUARTO DE
MILHA

AUTOR: RAFAEL DE MATTEIS

ORIENTADOR: ROGERIO ABDALLAH CURI

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

Prof. Dr. ROGERIO ABDALLAH CURI

Departamento de Melhoramento e Nutrição Animal / FMVZ / UNESP - Botucatu

Prof. Dr. GUILHERME COSTA VENTURINI

Departamento de Melhoramento e Nutrição Animal / FMVZ / UNESP - Botucatu

Pesquisadora Dra. SIMONE CRISTINA MÉO NICIURA

Empresa Brasileira de Pesquisa Agropecuária / EMBRAPA / São Carlos/SP

Jaboticabal, 11 de outubro de 2017

DADOS CURRICULARES DO AUTOR

RAFAEL DE MATTEIS – filho de Enzo de Matteis e Tereza Cristina Arbulu de Matteis. Nasceu em 08 de maio de 1992 na cidade de São Paulo, estado de São Paulo (Brasil). Em março de 2010 iniciou o curso de Zootecnia na Universidade Estadual Paulista “Julio de Mesquita Filho” (UNESP), Campus Botucatu – SP, obtendo o título de Zootecnista em dezembro de 2014. Em agosto de 2015 ingressou no Programa de Pós-Graduação em Genética e Melhoramento Animal na Faculdade de Ciências Agrárias e Veterinárias, UNESP, Campus Jaboticabal – SP, sob orientação do Prof. Dr. Rogério Abdallah Curi.

AGRADECIMENTOS

Aos meus pais, Enzo e Tereza, e meus irmãos Luiz Filipe e Luiz Fernando, que sempre acreditaram na minha capacidade e me fizeram chegar até aqui.

Aos meus tios, Ângelo e Cristina, que me ajudaram de todas as formas possíveis.

Aos meus avós, Rubens, Giovanni, Inez e Giovanna, que onde estiverem, serei eternamente grato por tudo que aprendi.

Aos meus amigos da graduação, Camilla, Cíntia, Lucas, Daniel e Felipe.

Aos meus amigos, André e Bianca que me acolheram e fizeram minha passagem por Jaboticabal muito melhor.

Ao meu amigo Wellison (piá), por me acompanhar e me auxiliar desde os tempos de estágio em São Carlos.

Ao meu amigo Guilherme, por sempre estar presente e ter me ajudado em todas as dificuldades que o mestrado me proporcionou.

Ao meu orientador, Rogério Abdallah Curi, pelos ensinamentos, paciência e dedicação na elaboração do projeto.

A CAPES, pela concessão da bolsa de estudo.

A todas as pessoas que, de alguma maneira, contribuíram para o meu crescimento pessoal e profissional e me ajudaram neste período.

Finalmente, agradeço a Faculdade de Ciências Agrárias e Veterinárias (FCAV), pela oportunidade de desenvolver minha dissertação de mestrado.

Sumário

CAPÍTULO 1 – CONSIDERAÇÕES GERAIS.....	3
1.Introdução	3
2. Objetivos	4
2.1. Objetivo geral	4
2.2. Objetivos específicos	5
3. Revisão bibliográfica	5
3.1. Os equinos no Brasil e no mundo	5
3.2. A raça Quarto de Milha no mundo e no Brasil	6
3.3. As linhagens de corrida e de trabalho da raça Quarto de Milha	7
3.4. Diferenças entre as linhagens de corrida e de trabalho da raça Quarto de Milha	9
3.5. Melhoramento genético de equinos	10
3.6. Marcadores de DNA.....	11
3.7. Polimorfismos de nucleotídeo único (SNP).....	11
3.8. Arranjos de genotipagem de SNP.....	12
3.9. Desequilíbrio de Ligação (DL).....	13
3.10. Gene da miostatina (<i>MSTN</i>)	14
3.11. Elemento Repetitivo Equino 1 (ERE1) do gene <i>MSTN</i> equino	15
4. Bibliografia citada	16
CAPÍTULO 2 – VARIANTES DA REGIÃO CROMOSSÔMICA DO GENE DA MIOSTATINA E SUAS RELAÇÕES COM LINHAGENS, DESEMPENHO E MEDIDAS CORPORAIS NA RAÇA QUARDO DE MILHA.....	24
1. Introdução	26
2. Material e métodos.....	28
2.1. Animais amostrados e coleta dos fenótipos.....	28

2.2. Extração do DNA e genotipagem de polimorfismos da região do <i>MSTN</i> no ECA18.....	29
2.3. Imputação de genótipos na linhagem de corrida	30
2.4. Variantes genômicas da região do <i>MSTN</i> no ECA18	31
2.5. Associação entre variantes genômicas e características de interesse	32
3. Resultados e discussão	33
3.1. Variantes genômicas da região do <i>MSTN</i>	33
3.2. Variantes genômicas da região do <i>MSTN</i> associadas à linhagem	35
3.3. Associação entre as variantes genômicas e as características de interesse	40
4. Conclusão	45
5. Bibliografia citada.....	46

Variantes da região cromossômica do gene da miostatina e suas relações com linhagens, desempenho e medidas corporais na raça Quarto de Milha

RESUMO - Ao longo de várias décadas a raça Quarto de Milha foi selecionada para diferentes objetivos, formando grupos com aptidões ou habilidades distintas, como a linhagem de corrida, que apresenta melhor desempenho que qualquer outra linhagem ou raça em corridas de curta distância e a linhagem de trabalho, utilizada no manejo de bovinos a campo e em provas de caráter funcional. Considerando as diferenças nas medidas corporais e na musculatura que ocorrem entre as linhagens de corrida e de trabalho da raça Quarto de Milha, o objetivo deste trabalho foi analisar a ocorrência de relações entre alelos de polimorfismos da região cromossômica do gene da miostatina *MSTN* (ECA18), um regulador negativo do desenvolvimento muscular, e as duas linhagens da raça. Outro objetivo foi realizar análises de associação de polimorfismos dessa região cromossômica com valor genético estimado (EBV) do desempenho em corridas, dado pelo índice de velocidade máximo (IV max), e EBVs da altura à cernelha (AC), perímetro torácico (PT) e comprimento corporal (CC) na linhagem de corrida, e com EBVs da AC, PT e CC na linhagem de trabalho. Foram utilizadas informações genômicas, fenotípicas e de pedigree de 420 equinos, de ambos os sexos, registrados na associação Brasileira de criadores (ABQM), sendo 352 da linhagem de corrida e 68 da de trabalho. Na região genômica estudada (gene *MSTN* $\pm$ 2Mb), foram identificados 46 SNPs e 1 SINE – ERE1 comuns às linhagens de trabalho e de corrida, dos quais 32 SNPs e 1 SINE apresentaram alelos com frequências significativamente diferentes entre as mesmas. Também foi constatado que a porção dessa região mais próxima ao gene *MSTN* ($\pm$ 1Mb) é menos polimórfica na linhagem de corrida em relação à de trabalho, o que pode ser consequência de maior pressão de seleção sobre essa região na linhagem de corrida e/ou em função do uso de animais da raça Puro Sangue Inglês na mesma. Em relação aos polimorfismos associados à fenótipos (p não ajustado $< 0,05$), destacaram-se os relacionados à altura de cernelha na linhagem de corrida devido ao seu grande número (p não ajustado $< 0,05$) e alto desequilíbrio de ligação. Ademais estes marcadores apresentaram frequências do alelo favorável significativamente maiores em relação à linhagem de trabalho. Tendo se em vista a possibilidade de associação genética entre características morfológicas e desempenho em equinos, esse resultado é de grande importância, fornecendo informações para futuros estudos de melhoramento genético animal.

Palavras-chave: equino, ERE1, marcadores moleculares, QTL, SNP, velocidade.

Variants of the chromosomal region of the myostatin gene and its relationships with lineages, performance and body measurements in the quarter-mile breed

ABSTRACT - For several decades the Quarter Horse breed was selected for different purposes, forming groups with different abilities, such as the racing line, which performs better than any other lineage or breed in short distance races and the lineage of work, used in the management of field cattle and in functional tests. Considering the differences in body measurements and musculature that occur between racing and working strains of the Quarter-Mile breed, the objective of this study was to analyse the occurrence of relations between polymorphic alleles of the chromosomal region of the myostatin (*MSTN*) gene (ECA18), a negative regulator of muscle development, and the two lineages of the breed. Other objective was to perform association analyses of these polymorphisms with estimated genetic value (EBV) of the performance in lineages, given by the maximum velocity index (IV max), and EBVs of height at withers (HW), heart girth (HG) and body length (BL) in the racing line, and with EBVs of AC, PT and CC in the working line. We used genomic, phenotypic and pedigree information from 420 horses of both sexes, recorded in the Brazilian breeders' association (ABQM), 352 and 68 of the racing and working line, respectively. In the studied genomic region (*MSTN* gene $\pm$ 2Mb), 46 SNPs and 1 SINE - ERE1 were identified, common to the working and running strains, of which 32 SNPs and one SINE presented alleles with significantly different frequencies between them. It has also been found that the portion of this region closest to the *MSTN* gene ($\pm$ 1Mb) is less polymorphic in the racing versus the working strain, which may be a consequence of higher selection pressure on that region in the racing line and / or due to the use of purebred English animals in it. In relation to the polymorphisms associated with the phenotypes (p not adjusted <0.05), those related to the height of the withers in the race line due to their large number (p not adjusted <0.05) and high linkage disequilibrium were the most prominent. In addition, these markers presented frequencies of the favorable allele significantly higher in relation to the working lineage. Considering the possibility of genetic association between morphological characteristics and performance in equines, this is a result of great importance, providing information for future animal breeding studies.

Key words: equine, ERE1, molecular markers, QTL, SNP, speed.

CAPÍTULO 1 – Considerações Gerais

1.Introdução

Equinos da raça Quarto de Milha são provenientes da América do Norte e no decorrer dos anos houve a seleção de animais para provas de velocidade em pistas de curta distância, cerca de um quarto de milha (402m), o que deu o nome à raça. Entretanto, ao longo de várias décadas seguintes, o Quarto de Milha foi selecionado para diferentes objetivos, formando grupos com aptidões ou habilidades distintas (ABQM, 2017), entre as quais as linhagens de corrida e de trabalho.

O principal atributo para a seleção de reprodutores, utilizado pelos criadores do cavalo Quarto de Milha de corrida, é a pontuação denominada índice de velocidade, a qual é obtida durante a campanha do animal com o objetivo de classificar o seu desempenho em diferentes condições (distâncias, hipódromo, clima, país) (EVANS, 1996). Os animais da linhagem de trabalho podem competir em diversas provas esportivas, como balizas, rédeas, tambores e apartação. Estes equinos são destinados às provas de caráter funcional, que exploram habilidades como agilidade e obediência, características de grande importância no manejo de bovinos a campo (MEIRA et al., 2014a).

Cavalos Quarto de Milha podem realizar variados tipos de exercícios físicos, indo do predominantemente aeróbio ao predominantemente anaeróbio. Nas corridas de curtas distâncias, o Quarto de Milha da linhagem de corrida apresenta metabolismo muscular predominantemente anaeróbio. Já em algumas classes de exercícios de arena, como apartação e rédeas, modalidades desempenhadas pela linhagem de trabalho, os animais intercalam curtos turnos de exercício anaeróbio com maiores períodos de atividade aeróbia (FREEMAN, 2013). Além de variações fisiológicas, incluindo o metabolismo energético muscular, as linhagens do cavalo Quarto de Milha apresentam diferenças em relação à morfologia. Segundo Beltrán et al. (2015) em geral, o cavalo de corrida da raça Quarto de Milha é mais alto que o de trabalho devido, em parte, à influência de animais Puro-Sangue Inglês (PSI), utilizados para melhorar a sua velocidade. Animais da linhagem de trabalho têm o corpo mais curto e compacto, com os posteriores musculosos, enquanto que os de corrida são mais

altos e com pernas mais longas, e musculatura menos proeminente (BELTRÁN et al., 2015).

O potencial atlético em mamíferos é influenciado por complexa inter-relação entre conjunto de genes e fatores ambientais (HILL et al., 2010c). Entretanto, até o momento poucas variantes genéticas foram relacionadas ao potencial atlético em equinos, e exclusivamente em animais Puro-Sangue Inglês (PSI). Entre estas variantes tem-se um polimorfismo de nucleotídeo único (SNP) no gene da miostatina (*myostatin* – *MSTN*; 66490208..66495180) (HILL et al., 2010a;b e BINNS et al., 2010, TOZAKI et al., 2010), o qual se encontra localizado no cromossomo equino 18 (ECA18), sendo expresso no tecido muscular esquelético, e atua como regulador negativo do crescimento da massa muscular (TOZAKI et al., 2010).

Entretanto, este polimorfismo não é o único do gene *MSTN* a se destacar em equinos. Santagostino et al. (2015) estudaram o efeito de um elemento curto intercalado (SINE) chamado de Elemento Repetitivo Equino 1 (ERE1) na região promotora do *MSTN*, e mostraram que a sua inserção/ocorrência provoca redução da expressão gênica. Os autores sugeriram que a inserção ERE1 (alelo ERE1+) pode diminuir os níveis de miostatina, e assim incrementar o desenvolvimento muscular e o desempenho atlético em cavalos.

2. Objetivos

2.1. Objetivo geral

Este trabalho teve como objetivo geral desenvolver estudo genômico nas linhagens de corrida e de trabalho da raça Quarto de Milha, com foco em região do cromossomo equino 18 (ECA18), onde se encontra mapeado o gene da miostatina (*MSTN*), dada a importância do produto proteico deste gene no desenvolvimento corporal e, conseqüentemente, na aptidão esportiva.

2.2. Objetivos específicos

Em vista do papel do gene da miostatina (*MSTN*) na fisiologia do músculo esquelético e considerando as diferenças nas medidas corporais e na musculatura que ocorrem entre as linhagens de corrida e de trabalho do Quarto de Milha, os objetivos específicos deste trabalho foram analisar a ocorrência de relações entre polimorfismos da região cromossômica do *MSTN* e as duas linhagens da raça.

Além disso, considerando que os efeitos de polimorfismos de DNA sobre fenótipos são parâmetros intrínsecos de cada linhagem ou raça em determinado ambiente, foram realizadas análises de associação de polimorfismos da região cromossômica do gene *MSTN* com EBVs do índice de velocidade máximo (IV max), altura de cernelha (AC), perímetro torácico (PT) e comprimento corporal (CC) na linhagem de corrida, e com EBVs de AC, PT e CC na linhagem de trabalho.

3. Revisão bibliográfica

3.1. Os equinos no Brasil e no mundo

Há, aproximadamente 58,9 milhões de cavalos no mundo, com a maior parte vivendo na América, Ásia e alguns países da Europa (FAO, 2014). Estados Unidos é o país com maior número de indivíduos (10.260.000), seguido do México (6.355.000), da China (6.024.000) e do Brasil (5.450.601) (GLIPHA, 2014).

O complexo do agronegócio do cavalo no Brasil é expressivo, movimentando valor econômico superior a R\$ 16,15 bilhões ao ano e gerando aproximadamente 3,2 milhões de empregos diretos e indiretos (MAPA, 2015). A utilização de animais no manejo de rebanhos bovinos (aproximadamente quatro milhões) vem sendo o seguimento de maior destaque no país. Somados os valores anuais do custo de manutenção, da tropa e a renda associada à mão de obra, observa-se que o segmento de manejo ligado ao rebanho responde por R\$ 8,58 bilhões, ou seja, aproximadamente 53% da renda gerada pelo complexo do cavalo (MAPA, 2015). Esta intensa relação tem feito estes animais acompanharem os bovinos em seu deslocamento para as regiões centro-oeste e norte do país.

Apesar da utilização do cavalo no manejo com bovinos ser marcante no contexto da equinocultura nacional, o seguimento que envolve os esportes equestres tem crescido acentuadamente nos últimos anos, acompanhando tendência mundial. Calcula-se a existência de 1,1 milhões de animais no segmento de esportes e lazer, com 125.700 ocupações diretas e movimentação econômica de R\$ 5,84 bilhões (MAPA, 2015). Entre os anos de 1999 e 2004, o número de eventos envolvendo as diversas modalidades equestres (corrida, salto, adestramento, concurso completo de equitação, enduro, competições de marcha, etc.) cresceu 315%, com aumento médio de 15,3% ao ano (LIMA et al., 2006). O crescimento verificado nessa categoria desde o estudo realizado por Lima et al. (2006), em que o plantel foi estimado em 800 mil animais, foi de 37% (MAPA, 2015).

Além da utilização do cavalo no manejo da fazenda e nas provas equestres, a tendência de sua utilização para lazer vem aumentando significativamente nos últimos anos. O cavalo também é utilizado na equoterapia, modalidade disponível há milhares de anos e, agora, reconhecida como de grande eficácia para o tratamento em humanos de inúmeros males físicos, psíquicos e comportamentais (BELTRÁN, 2014).

3.2. A raça Quarto de Milha no mundo e no Brasil

A raça Quarto de Milha foi a primeira a ser introduzida na América do Norte, a partir do século XVII, com a chegada de equinos de origem árabe e turca, trazidos por colonizadores europeus. O seu desenvolvimento deveu-se à ocupação do oeste norte-americano, em razão da necessidade de cavalos robustos e versáteis, com aptidão à sela e tração, visto a dificuldade de manutenção de plantel variado de animais para atender a diferentes necessidades (ABQM, 2016).

Em 1940, foi fundada, nos Estados Unidos, a *American Quarter Horse Association* (AQHA), primeira no mundo envolvendo a raça. Atualmente, a AQHA é tida como a maior associação de criadores do mundo, com aproximadamente 400 mil sócios e mais de cinco milhões de animais registrados em 43 países, compreendendo 52% dos equinos de todo o mundo (ABQM, 2016).

No ano de 1969 foi anunciada a fundação da Associação Brasileira de Criadores de Cavalo Quarto de Milha (ABQM), que conta nos dias de hoje com plantel de 480.000 animais registrados, 97.900 criadores, proprietários e associados cadastrados distribuídos nos 26 estados brasileiros mais Distrito Federal. No Brasil, a

raça tem impacto relevante no agronegócio do cavalo, visto que seus haras ocupam área de aproximadamente um milhão de hectares, estimados em mais de R\$ 19,8 bilhões (ABQM, 2016).

3.3. As linhagens de corrida e de trabalho da raça Quarto de Milha

A seleção para diferentes propostas na raça Quarto de Milha (EVANS, 1996) levou à formação das linhagens de corrida, de trabalho e de conformação.

Animais de corrida da raça Quarto de Milha dispõem de aptidão para desempenhar grande velocidade em pistas retas e de curta distância. Estes equinos apresentam melhor desempenho em corridas de curtas distâncias do que qualquer outra raça (ABQM, 2016), e são considerados os mais velozes cavalos do mundo e um dos mais velozes dentre todos os animais, podendo alcançar velocidades de até 88 km/h e percorrer $\frac{1}{4}$ de milha (402 metros) em menos de 21 segundos (AMERICA'S HORSE DAILY, 2008).

Apesar do efetivo de animais ser relativamente menor na linhagem de corrida do que nas demais, sua importância econômica é significativa, não somente por gerar renda por meio de premiações e apostas (US\$ 2,5 milhões – LIMA et al., 2006), mas também pelo elevado custo de manutenção destes animais em modalidades esportivas.

O principal quesito de seleção utilizado pelos criadores do cavalo Quarto de Milha de corrida é a pontuação conhecida como índice de velocidade (IV). Este índice é obtido ao longo da campanha do animal com o intuito de classificá-lo em relação ao seu desempenho em diferentes condições (distâncias, hipódromo, clima, país) (EVANS, 1996). Cada hipódromo tem sua tabela de IV, a qual é elaborada a partir da média simples dos tempos das três vitórias mais rápidas (três melhores tempos) para cada um dos três últimos anos, em cada distância, sendo que o valor da média dos nove tempos corresponderá ao IV de valor 100 (JCS, 2002). Os pontos de IV são inteiros, variam de acordo com o tempo, ao nível de centésimos de segundo, sendo ajustados de acordo com a distância percorrida.

Na linhagem de trabalho o Quarto de Milha é reconhecido pela docilidade, facilidade de ser domado, inteligência, e pela grande capacidade de mudar de direção, com partidas rápidas e bruscas. Destina-se às modalidades esportivas de caráter funcional, como apartação, laço e rédeas, as quais exploram habilidades como

agilidade e obediência, características consideradas de grande importância no manejo do gado a campo. Por outro lado, prova como a dos três tambores exige, além da agilidade e obediência, a velocidade.

A linhagem de trabalho é utilizada nas principais modalidades esportivas destinadas a equinos, como, por exemplo, as provas de apartação, onde cavalo e cavaleiro devem se mover calmamente para dentro do rebanho, apartar o animal escolhido e dirigi-lo ao centro da arena a fim de mantê-lo afastado do rebanho. O cavaleiro deve impedir que o animal, já apartado, volte a compor o rebanho. Isso acontece num tempo limite de 2 minutos e 30 segundos (ABQM, 2017). Dentro desta modalidade sempre houve grande interesse em cavalos com *cow sense* superior (ELLERSIECK et al., 1985). *Cow sense*, ou habilidade de trabalhar com bovinos, é medida pela capacidade do cavalo de cercar o rebanho e apartar um animal escolhido (HINTZ, 1980), com pouca ou nenhuma assistência do cavaleiro (ELLERSIECK et al., 1985).

Na modalidade laço, a prova pode ser de laço em dupla, laço cabeça e pé e laço individual cronometrado e técnico. As provas têm o objetivo de laçar um bezerro no menor tempo possível, o que exige velocidade e habilidade do cavalo. A modalidade dos três tambores é uma das competições de velocidade de maior interesse dentre todas as provas cronometradas do Quarto de Milha. Os competidores correm contra o cronômetro, seguindo percurso que consiste em três tambores dispostos triangularmente. Nos últimos anos, animais que participam desta modalidade têm recebido influência genética de animais da linhagem de corrida com o intuito de melhorar o equilíbrio e a velocidade. Cavalos de corrida que apresentam estas características, além de temperamento dócil, são utilizados em acasalamentos com animais de trabalho, para melhor desempenho na prova. Criadores têm optado por utilizar estes cavalos por apresentarem maior arranque, agilidade em curvas e obediência em relação ao cavaleiro.

Outra modalidade esportiva tradicional no Quarto de Milha de trabalho é a prova de rédeas, na qual o cavaleiro precisa dominar todos os movimentos de seu cavalo. O cavaleiro executa um dos 13 percursos existentes pré-estabelecidos, os quais incluem: manobras prescritas de esbarros; *spins* (giros sobre si de 360 graus); *rollbacks* (esbarro com mudança de direção em 180 graus saindo ao galope); mudança de mão; e círculos ao galope. Nesta prova, o cavalo deve ser guiado voluntariamente com pouca ou nenhuma resistência (ABQM, 2017). Completam as

modalidades que os cavalos de trabalho participam o *team penning*, *ranch sorting* e *working cow horse*.

Por fim, criadores da linhagem de conformação da raça Quarto de Milha têm como propósito preservar as características da raça, selecionando os animais mais apropriados, de acordo com o padrão considerado ideal. Essa seleção busca a combinação e equilíbrio entre estrutura corporal, características raciais e sexuais e musculatura bem distribuída (ABQM, 2017).

3.4. Diferenças entre as linhagens de corrida e de trabalho da raça Quarto de Milha

Processos fisiológicos ligados ao desempenho atlético têm mostrado aspectos semelhantes em humanos e equinos (DIAS et al., 2007; SCHÖRODER et al., 2011). Segundo Breazile (1996), as concentrações de oxigênio no tecido muscular podem variar e atuar sobre a regulação de seu metabolismo, aeróbico ou anaeróbico, em função do tipo de treinamento físico.

Cavalos da raça Quarto de Milha realizam variados tipos de exercícios físicos, indo do predominantemente aeróbico até o predominantemente anaeróbico. As corridas curtas do Quarto de milha da linhagem de corrida apresentam metabolismo muscular predominantemente anaeróbico. Já algumas classes de exercícios de arena, como apartação e rédeas, modalidades desempenhadas pela linhagem de trabalho, intercalam turnos curtos de exercício anaeróbico com maiores períodos de atividade aeróbia (FREEMAN, 2013), assim como as corridas longas da raça PSI.

Embora existam diferenças fisiológicas, incluindo o metabolismo energético muscular, o que é mais evidente e visual é a diversidade morfológica que ocorre entre os animais das linhagens de corrida e de trabalho da raça Quarto de Milha. Meira et al. (2013) colocaram que, em geral, o cavalo Quarto de Miha de corrida é mais alto que o de trabalho devido à influência de animais PSI, utilizados para a melhora de sua velocidade. Entretanto, explicam também que, a seleção direcionada a diferentes objetivos (corrida ou trabalho) colaborou para promover mudanças significativas nas características corporais dos equinos Quarto de Milha. Meira et al. (2013), demonstraram a existência de diferenças significativas entre as linhagens com relação às características morfológicas, entre elas: altura à cernelha; comprimento corporal e perímetro torácico. Outras características morfológicas também foram comparadas

entre as linhagens pelos mesmos autores, como: peso, comprimentos de canela, da quartela, da garupa, da cabeça, e do pescoço e do casco, e todas estas medidas foram superiores em cavalos da linhagem de corrida. Expuseram ainda que a linhagem de trabalho tem o comprimento corporal mais curto e compacto, com os posteriores musculosos, enquanto os de corrida foram mais altos e suas pernas mais longas, com musculatura menos proeminente.

3.5. Melhoramento genético de equinos

Pesquisas na área de melhoramento genético são, no mundo todo, escassas em equinos em relação à outras espécies de interesse zootécnico. No Brasil, em particular, dada a grandeza do plantel, esta escassez é ainda maior (BELTRÁN, 2014).

Embora algumas pesquisas publicadas com equinos envolvam melhoramento genético, ainda não existem programas consistentes de seleção para as diferentes raças criadas no país. O que têm sido vistos, são pequenos grupos de pesquisadores trabalhando especialmente aspectos conservacionistas em raças nacionais ou quantitativos de caracteres de interesse econômico em raças nacionais e importadas. Apenas recentemente pesquisas em Genéticas e Biologia Molecular têm sido realizadas em algumas raças da espécie no Brasil (ARNEIRO et al., 2012; MEIRA et al., 2014a,b,c; PEREIRA et al., 2016a,b; FONSECA et al., 2016).

Programas efetivos de seleção e melhoramento genético em equinos vêm sendo realizados na Europa, com a utilização de modelos estatísticos para a obtenção de valores genéticos individuais (ARNASON, 2013), em cavalos destinados a esportes olímpicos como o *cross country*, o adestramento e o salto (PEREIRA, 2017).

A maior causa desta dificuldade de aplicação da seleção baseada em estimação de valores genéticos é a baixa herdabilidade do desempenho atlético e os valores de correlações genéticas existentes entre as características de maior interesse nos cavalos (KOENEN et al., 1995; WALLIN et al., 2003; BOKOR et al., 2005). Entretanto, a baixa herdabilidade do desempenho atlético em equinos justifica a utilização de novas ferramentas disponíveis para a seleção e o melhoramento genético, como os marcadores de DNA (PEREIRA, 2017).

3.6. Marcadores de DNA

Marcador molecular é toda e qualquer variação oriunda de um gene expresso ou de um segmento específico de DNA, correspondente a regiões expressas ou não do genoma. Quando esses marcadores segregam de acordo com as leis mendelianas para características monogênicas, ou apresentam distribuições compatíveis com as esperadas para características poligênicas, um marcador molecular é definido também como marcador genético (FERREIRA & GRATTAPAGLIA, 1998).

Em relação ao DNA, ocorrem três tipos principais de variações, as regiões repetitivas (minissatélites e microssatélites), as inserções e deleções (InDels) e as alterações de uma base (polimorfismos de nucleotídeo único – SNP). À medida que as sequências de nucleotídeos dos genomas foram sendo desvendadas, uma característica rapidamente observada foi o grande número de variações de uma base encontradas ao se comparar segmentos correspondentes do genoma de indivíduos na mesma espécie, ocorrendo, a aproximadamente, cada 500 bases (PEREIRA, 2017).

3.7. Polimorfismos de nucleotídeo único (SNP)

As substituições mais frequentemente observadas no DNA envolvem os nucleotídeos de bases nitrogenadas de mesma característica estrutural, ou seja, são trocas entre duas purinas (A/G ou G/A) ou duas pirimidinas (C/T ou T/C) e são denominadas transições. As transversões são substituições de uma purina por uma pirimidina ou o contrário. Essas alterações podem ser provocadas por erros de incorporação de bases durante a replicação do DNA ou em outros casos, são causadas por agentes ambientais (químicos, físicos e biológicos). Quando essas alterações ocorrem em células sexuais e são transmitidas às gerações seguintes são denominadas mutações germinativas. Caso fixem-se na população com frequência mínima de 0,01 ou mais (*minor allele frequency* – MAF $\geq 1\%$), passam a ser denominadas de polimorfismos (KWOK & GU, 1999). Neste sentido, as mutações de uma base passam a ser denominadas SNP.

Os SNP podem ocorrer em regiões gênicas codificadoras ou com função regulatória, bem como em espaços intergênicos, sem função especificada. Em regiões codificadoras, quando resultam em substituição de aminoácido na sequência proteica, são denominados não sinônimos, e a substituição pode ser conservativa ou não

conservativa em função das características dos aminoácidos envolvidos na troca. Nesses casos, pode haver modificações estruturais e funcionais na proteína e alterações fenotípicas.

Embora SNP sinonímicos não alterem a sequência proteica, podem modificar a estrutura e a estabilidade do RNA mensageiro, e, conseqüentemente, afetar a quantidade de proteína produzida. Esta também pode ser afetada quando ocorrem alterações nas regiões não traduzidas do RNA mensageiro (5' UTR e 3' UTR).

Além disso, polimorfismos gênicos podem promover *splicing* alternativos, formação ou eliminação de códons de terminação, alteração nos códons de início da tradução e alterações no padrão de expressão de genes quando a troca de bases ocorre em sequências promotoras (GUIMARÃES & COSTA, 2002) ou moduladoras.

Polimorfismos em regiões de íntron, embora não relacionados com a sequência de aminoácidos da proteína, também podem ser responsáveis diretos por alterações fenotípicas. Como exemplo, RNAs não codificantes transcritos a partir de regiões de íntrons são processados para formar micro-RNAs, os quais podem estar envolvidos em diferentes processos biológicos, tais como os controles transcricional e pós-transcricional da expressão gênica (NAKAYA et al., 2007). Embora a função da maior parte das regiões intergênicas ainda seja desconhecida, conhece-se cada vez com mais profundidade a sua importância no controle da regulação da expressão gênica. Neste sentido, polimorfismos localizados nestas regiões também podem estar associados à variações fenotípicas (PEREIRA, 2017).

Estudos em humanos e em espécies de interesse zootécnico mostraram a ocorrência de milhões SNP ao longo do genoma de um indivíduo (*Human Genome Project Information, The SNP Consortium LTD, Bovine Genome Sequencing and Analysis Consortium, EquCab2.0 SNP Collection*), ou seja, apenas pela comparação da sequência dos seus dois cromossomos homólogos. Além deste tipo de polimorfismo de DNA ser abundante nos genomas, suas bases moleculares permitem sua distribuição homogênea pelos cromossomos (CAETANO, 2009).

3.8. Arranjos de genotipagem de SNP

O maior destaque proveniente da análise do genoma de equinos foi o seu sequenciamento completo (EquCab2.0) e, a partir deste, a identificação de mais de um milhão de SNP que ocorrem entre diferentes raças (CHOWDHARY & RAUDSEPP,

2008; WADE et al., 2009). A identificação em massa destes polimorfismos de DNA distribuídos uniformemente ao longo dos cromossomos de indivíduos da mesma espécie permitiu, a partir do final dos anos 2000, a construção dos chamados microarranjos de SNP por consórcios estabelecidos entre centros de pesquisa e empresas privadas (BELTRÁN, 2014).

O *Equine SNP50 BeadChip* da empresa Norte Americana Illumina (Illumina Inc., EUA) foi projetado para possibilitar a identificação de regiões genômicas modificadas pela seleção, e a identificação de SNP, regiões cromossômicas e genes que contribuem para a variação de características de interesse nas principais raças de equinos criadas atualmente no mundo. Dessa forma, constituiu poderosa plataforma disponível para a seleção e o melhoramento genético da espécie, habilitando pesquisadores da área a conduzir vasta gama de experimentos em que a aplicação da genotipagem de polimorfismos de DNA é necessária. Em sua segunda geração, o novo arranjo de SNP para cavalos, o *Equine SNP70 BeadChip*, também da empresa Illumina, possui em torno de 65 mil SNP, dos quais 19 mil são novos e 45 mil pertenciam ao *Equine SNP50 BeadChip*.

3.9. Desequilíbrio de Ligação (DL)

A análise simultânea de dezenas a centenas de milhares de polimorfismos espalhados ao longo dos genomas tem possibilitado o estudo da estrutura genética de diferentes populações em diferentes espécies de animais domésticos (KIJAS et al., 2009; McKAY et al., 2008; GIBBS et al., 2009), estimar o grau de diversidade genética dentro e entre populações (ZENGER et al., 2006), determinar a relação entre a perda de alelos e o aumento da endogamia devido à seleção (MUIR et al., 2008), e identificar e mapear regiões do genoma submetidas à seleção recente (HAYES et al., 2006; BARENDSE et al., 2009; MACEachern et al., 2009; PRASAD et al., 2008).

Em geral, estas aplicações necessitam do conhecimento ou do acesso ao desequilíbrio de ligação (DL) existente entre os marcadores moleculares de amostra populacional representativa da linhagem, da raça ou da espécie de interesse. O DL entre dois marcadores de DNA corresponde à correlação entre os seus genótipos, expressando o grau de associação não aleatória entre seus alelos (PORTO-NETO et al., 2014). Esta associação ou ligação é consequência da proximidade física existente entre dois locos (ARDLIE et al., 2002; BOLORMAA et al., 2011) e é influenciado por

eventos como mutação, deriva, seleção, recombinação e tamanho efetivo da população.

As medidas convencionalmente utilizadas para estimar o DL entre marcadores bialélicos são o D' (LEWONTIN, 1964) e o r^2 (HILL & ROBERTSON, 1968), cada qual com diferentes propriedades estatísticas, porém com valores entre 0 e 1 (BOHMANOVA et al., 2010). O r^2 representa a correlação entre dois locos e mostrou-se ser mais adequado para avaliar o DL entre marcadores bialélicos, como os SNPs, ao ser menos sensível em relação à frequência alélica e ao tamanho da amostra (ZHU & ZHAO, 2007; BOHMANOVA et al., 2010). Para um par de locos bialélicos, r^2 é igual a 1 se apenas dois haplótipos estão presentes em uma população (GURGUL et al., 2014), ou seja, se os alelos de dois locos estiverem em completa ligação. Valores de r^2 entre 0,2 e 0,3 são considerados altos e evidenciam o DL entre dois locos.

Os mapas de DL ou de haplótipos são ferramentas fundamentais para explorar a base genética de características economicamente importantes (MEIRA, 2014d). Para melhor estimar o poder de mapear associações entre polimorfismos gênicos e características de importância dentro de raças equinas, a extensão do DL foi caracterizada por meio de r^2 em 14 delas. Em geral, os valores calculados de r^2 caíram abaixo de 0,2 à distâncias entre 100 e 150Kb. Notáveis exceções foram o Puro-Sangue Inglês, o *Standardbred* e o *French Trotter*, nos quais os valores de r^2 permaneceram acima de 0,2 à distâncias superiores à 400Kb. Entre as raças mais difundidas, o Quarto de Milha destacou-se por seus valores relativamente baixos de DL quando comparadas a outras raças equinas, com decaimento de valores de r^2 abaixo de 0,2 entre 50 e 100Kb (McCUE & MICKELSON, 2013). A pequena extensão do DL médio em Quartos de Milha provavelmente é reflexo do seu grande tamanho populacional efetivo e da sua estrutura aberta (MEIRA, 2014d).

3.10. Gene da miostatina (*MSTN*)

Genes candidatos para características de interesse podem ser identificados com base em conhecimentos de fisiologia, análises de expressão gênica diferencial e varredura do genoma (ZHU & ZHAO, 2007). A maior parte das características de importância em animais de interesse zootécnico é governada por grande número de genes de pequeno efeito, ou seja, são quantitativas. Assim, a identificação de locos de características quantitativas (QTL) no genoma destes animais vem sendo focada

nos últimos anos em estudos de genoma inteiro (GWAS), os quais são realizados por meio da genotipagem de dezenas a centenas de milhares de polimorfismos de nucleotídeo único (SNP) utilizando arranjos de pequena, média e de alta densidades (HAYES & GODDARD, 2010). No entanto, existe a possibilidade da existência de *major genes* ou genes principais, que seriam os responsáveis pela maior parte da variação de uma característica (ZHU & ZHAO, 2007).

Mas, o mais provável é que, assim como em humanos (BRAY et al., 2009), o desempenho atlético em equinos seja influenciado por grande número de genes. Entretanto, até o momento poucas variantes genéticas foram relacionadas à característica, exclusivamente em animais PSI, entre estas, SNP do gene da miostatina (*myostatin* – *MSTN*; 66490208..66495180: ECA18) (HILL et al., 2010a;b e BINNS et al., 2010, TOZAKI et al., 2010). O gene *MSTN* é expresso no tecido muscular esquelético e atua como regulador negativo do crescimento da massa muscular. Várias mutações ou polimorfismos foram identificados no *MSTN* de bovinos (GROBET et al. 1997; KAMBADUR et al., 1997; MCPHERRON et al., 1997), ovinos (CLOP et al., 2006), camundongos (NISHI et al., 2002) e humanos (SCHUELKE et al., 2004), resultando em fenótipos de hiperplasia e hipertrofia muscular, ou seja, de aumento de massa corporal.

3.11. Elemento Repetitivo Equino 1 (ERE1) do gene *MSTN* equino

Grande proporção do genoma de mamíferos é preenchida por repetições intercaladas que foram geradas ao longo da evolução pela propagação de elementos transponíveis (DEININGER et al., 2003). Elementos curtos intercalados (SINEs) são retrotransposons não autônomos que fazem uso de um RNA intermediário reversamente transcrito para inserir um cDNA em novo local do genoma (JURKA et al., 2007). A inserção de elementos transposons dentro de genes ou em sua proximidade pode alterar a sua expressão por meio da sua interrupção, e modificações de sequências em promotores ou locais de *splicing* (WHEELAN et al., 2005).

Santagostino et al. (2015) estudaram o efeito de um SINE conhecido como Elemento Repetitivo Equino 1 (ERE1) no promotor do gene *MSTN* (g.66495342:ECA18) em ensaio de gene repórter, e mostraram que a inserção provoca redução da expressão gênica. Os autores sugeriram que a inserção ERE1

(ERE1+) pode diminuir os níveis de miostatina, e assim incrementar o desenvolvimento muscular em cavalos. Ainda de acordo com Santagostino et al. (2015), a inserção ERE1 no promotor do gene da miostatina é polimórfica em cavalos (alelos ERE1+ e ERE1-) e parece estar relacionada à aptidão para corridas específicas. O alelo ERE1+ parece ser comum em raças que correm distâncias curtas, como a Quarto de Milha, e ausente em outras, como a Trotter Italiana, que são utilizadas em corridas de longa distância.

A alta frequência do alelo ERE1+ em cavalos Quarto de Milha sugere que esta variante pode favorecer a capacidade de correr distâncias curtas. Em uma amostra de cavalos Puro-Sangue Inglês de elite, ou seja, classificada nos três primeiros lugares em pelo menos uma corrida de alto nível na Itália, observou-se associação significativa entre a variante ERE1+ e o bom desempenho em corridas de curta distância. Por outro lado, o alelo ERE1- foi mais frequentemente encontrado em Puro-Sangue Ingleses vencedores de corridas de longa distância (SANTAGOSTINO et al., 2015).

4. Bibliografia citada

ABQM. **Associação brasileira dos criadores de cavalos Quarto de Milha**. Disponível em: <http://www.abqm.com.br/item1.htm>. Acesso em: 15 de fevereiro de 2016.

ABQM. **Associação brasileira dos criadores de cavalos Quarto de Milha**. Disponível em: <http://www.abqm.com.br/item1.htm>. Acesso em: 23 de janeiro de 2017.

AMERICA'S HORSE DAILY. **All about the Racing American Quarter Horse**. Disponível em: <http://americashorsedaily.com/all-about-the-racing-american-quarter-horse/>. Acesso em: 26 de fevereiro de 2008.

ARDLIE, K. G.; KRUGLYAK, L.; SEIELSTAD, M. Patterns of linkage disequilibrium in the human genome. **Natural Review Genetics**. V. 3, p. 299:309, 2002.

ARNASON T. Bright future for research in horse breeding! **Journal of Animal Breeding and Genetics**. V. 9, p.130:167, 2013.

ARNEIRO, L.C.M.; CURI, R.A.; SILVA, J.A.II V.; Mota, M.D.S. Caracterização da variabilidade de genes relacionados à fertilidade de machos e ao temperamento em eqüinos da raça brasileira Mangalarga. **Semina: Ciências Agrárias**. Londrina. V.33, n. 5, p. 2001:2010, 2012.

BARENDSE, W.; HARRISON, B. E.; BUNCH, R. J.; THOMAS, M. B.; TURNER, L. B. Genome wide signatures of positive selection: The comparison of independent samples and the identification of regions associated to traits. **BMC Genomics**. V.10, p. 178, 2009.

BELTRÁN, N.A.R. **Assinaturas de seleção na linhagem de trabalho de equinos quarto de milha**. 2014. 59 f. Tese (Doutorado melhoramento animal corrente em Genética e) – Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista "Júlio de Mesquita Filho", Jaboticabal, 2014.

BELTRÁN, N.A.R.; MEIRA, C.T.; OLIVEIRA, H.N.; PEREIRA, G.L.; SILVA, J.A.II V.; MOTA, M.D.S.; CURI, R.A. Prospection of genomic regions divergently selected in cutting line of Quarter Horses in relation to racing line. **Livestock Science**. V. 174, p. 1:9, 2015.

BINNS, M. M.; BOEHLER, D. A.; LAMBERT, D. H. Identification of the myostatin locus (MSTN) as having a major effect on optimum racing distance in the Thoroughbred horse in the USA. **Animal Genetics**. V. 41 (Suppl. 2), p. 154:158, 2010.

BOHMANOVA, J.; SARGOLZAEI, M.; SCHENKEL, F. Characteristics of linkage disequilibrium in North American Holsteins. **BMC Genomics**. V.11, p.421, 2010.

BOKOR, A; BLOUIN, C; LANGLOIS, B; STEFLER, J. Genetic parameters of racing merit of Thoroughbred horses in steeplechase races. **Italian Journal of Animal Science**. V. 5, p. 4:43, 2005.

BOLORMAA, S.; HAYES, B. J.; SAVIN, K.; HAWKEN, R.; BARENDSE, W.; ARTHUR, P. F.; HERD, R. M.; GODDARD, M. E. Genome wide association studies for feedlot and growth traits in cattle. **Journal of Animal Science**. V. 89, p.1684:1697, 2011.

BRAY, M.S.; HAGBERG, J.M.; PÉRUSSE, L.; RANKINEN, T.; ROTH, S.M.; WOLFARTH, B.; BOUCHARD, C. The human gene map for performance and health-related fitness phenotypes: the 2006-2007 update. **Medicine & Science in Sports & Exercise**. V. 73, p. 41:35, 2009.

BREAZILE, J.E. Fisiologia do Músculo Esquelético. In: Swenson MJ, Reece WO. **Fisiologia dos animais domésticos**. Rio de Janeiro: Guanabara; 11a ed., 1996, p. 777:793.

CAETANO, A. R. Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. **Revista Brasileira de Zootecnia**. V. 38 (Suppl. especial), p.64:71, 2009.

CHOWDHARY, B.P., RAUDSEPP, T. The Horse Derby: racing from map to whole genome sequence. **Chromosome Research**. V.16, p.109:127, 2008.

CLOP, A.; MARCQ, F.; TAKEDA, H. et al. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. **Nature Genetics**. V.38, p.813:818, 2006.

DEININGER, P. L.; MORAN, J.V.; BATZER, M.A.; KAZAZIAN JR, H.H. Mobile elements and mammalian genome evolution. **Current Opinion in Genetics & Development**. V. 13, p. 651:658, 2003.

DIAS, R.G.; PEREIRA, A.C.; NEGRÃO, C.E.; KRIEGER, J.E. Polimorfismos genéticos determinantes da performance física em atletas de elite. **Revista Brasileira de Medicina do Esporte**. V. 13, p. 211:216, 2007.

ELLERSIECK, M. R.; LOCK, W. E.; VOGT, D. W.; AIPERSPACH, R. Genetic evaluation of cutting scores in horses. **Equine Veterinary Science**. V.5, p.287:289, 1985.

EVANS, J. W. **Horses: a guide to selection, care and enjoyment**. 4. ed. Freeman and Company: New York, 1996, p. 797.

FAO. **Food and Agriculture Organization of United Nations**. Disponível em: <http://www.fao.org/home/en/>. Acesso em 14 de novembro de 2014.

FERREIRA, M. E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. Embrapa-Cenargen: Brasília, 3 ed., 1998, p. 220.

FONSECA, M. G.; FERRAZ, G. C.; PEREIRA, G. L.; CURI, R.A. A genome-wide association study reveals differences in the genetic mechanism of control of the two gait patterns of Brazilian Mangalarga Marchador breed. *Journal of Equine Veterinary Science*. V. 37, p. 1:98, 2016.

FREEMAN, D.W. Physical conditioning of horses. Stillwater, Okla: Oklahoma State University, **Division of Agricultural Sciences and Natural Resources**. 2013.

GIBBS, R. A.; TAYLOR, J. F.; VAN TASSELL, C. P. Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. **Science**. V. 324, p.528:532, 2009.

GLIPHA. Global Live <http://www.agricultura.gov.br/> stock Production and Health Atlas. Disponível em:http://www.agricultura.gov.br/http://www.fao.org/ag/againfo/home/en/news_archive/AGA_in_action/glipha.html>. Acesso em: 15/01/2014.

GROBET, L.; ROYO MARTIN, L.J.; PONCELET, D.; PIROTTIN, D.; BROUWERS, B.; RIQUET, J.; SCHOEBERLEIN, A.; DUNNER, S.; MÉNISSIER, F.; MASSABANDA, J.; FRIES, R.; HANSET, R.; GEORGES, M. A deletion in the myostatin gene causes double muscling in cattle. **Nature Genetics**. V.17, p.71:74, 1997.

GUIMARÃES, P. E. M.; COSTA, M. C. R. SNPs: Sutis diferenças de um código. **Biotecnologia Ciência e Desenvolvimento**. V. 26, p.24:27, 2002.

GURGUL, A.; EMIK, E.; PAWLINA, K.; SZMATOLA, T.; JASIELCZUK, I.; BUGNOPONIEWIERSKA, M. The application of genome-wide SNP genotyping methods in studies on livestock genomes. **Journal of Applied Genetics**. V. 55(2), p.197:208, 2014.

HAYES, B. J.; LIEN, S.; NIELSEN, H.; OLSEN, H. G.; BERG, P.; MACEACHERN, S.; POTTER, S.; MEUWISSEN, T. H. The origin of selection signatures on bovine chromosome 6. **Animal Genetics**. V. 39, p.105:101, 2006.

HAYES, B.; GODDARD, M. Genome-wide association and genomic selection in animal breeding. **Genome**. V. 53, p. 876:883, 2010.

HILL, W. G.; ROBERTSON, A. Linkage Disequilibrium in Finite Populations. **Theoretical and Applied Genetics**. V. 38, p. 226:231, 1968.

HILL, E. W.; MCGIVNEY, B. A.; GU, J.; WHISTON, R.; MACHUGH, D. E. A genome-wide SNP-association study confirms a sequence variant (g.66493737C>T) in the equine myostatin (MSTN) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses. **BMC genomics**. V. 11, n. 552, 2010a.

HILL, E. W.; MCGIVNEY, B. A.; GU J.; WHISTON, R.; MACHUGH D. E. A genome-wide SNP-association study confirms a sequence variant (g.66493737C>T) in the equine myostatin (MSTN) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses. **BMC Genomics**. V.11, n.552, p.1:10, 2010b.

HILL, E. W.; GU J.; MCGIVNEY, B. A.; MACHUGH D. E. Targets of selection in the Thoroughbred genome contain exercise-relevant gene SNPs associated with elite racecourse performance. **Animal Genetics**. V. 41, p. 56:63, 2010c.

HINTZ, R.L. Genetics performance in the horse. **Journal of Animal Science**. V.51, p.582:594, 1980.

JCS. Especial Quarto de Milha de Corrida. **Jockey Club de Sorocaba**, Sorocaba, SP, 2002.

JURKA, J.; KAPITONOV, V.V.; KOHANY, O.; JURKA, M.V. Repetitive sequences in complex genomes: structure and evolution. **Annual Review of Genomics and Human Genetics**. V. 8, p. 241:259, 2007.

KAMBADUR, R.; SHARMA, M.; SMITH, T.P.L.; BASS, J.J. Mutations in myostatin (GDF-8) in double muscled Belgian Blue and Piedmontese cattle. **Genome Research**. V. 7, p.910-916, 1997.

KIJAS, J. W.; TOWNLEY, D.; DALRYMPLE, B. P.; HEATON, M. P.; MADDOX, J. F.; MCGRATH, A.; WILSON, P.; INGERSOLL, R. G.; MCCULLOCH, R.; MCWILLIAM, S.; TANG, D.; MCEWAN, J.; COCKETT, N. V.; ODDY, H.; NICHOLAS, F. W.; RAADSMA, H. A. Genome Wide Survey of SNP Variation Reveals the Genetic Structure of Sheep Breeds. **Plos One**. V. 4(3): n.4668, 2009.

KOENEN, E.P.C.; VAN VELDHUIZEN, A.E.; BRASCAMPA, E.W. Genetic parameters of linear scored conformation traits and their relation to dressage and show-jumping performance in the Dutch Warmblood Riding Horse population. **Livestock Production Science**. V.43, p. 85:94, 1995.

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

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

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

MACEACHERN, S.; HAYES, B.; MC EWAN, J.; GODDARD, M. An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (*Bostaurus*) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in Domestic cattle. **BMC Genomics**. V.10, p.181, 2009.

MAPA. **Ministério da Agricultura, Pecuária e Abastecimento**. Disponível em: <http://www.agricultura.gov.br/>. Acesso em: 18 de agosto de 2015.

McCUE, M.; MICKELSON, J. Genomic Tools and Resources: Development and Applications of an Equine SNP Genotyping Array. **Equine Genomics**. p. 113:124, 2013.

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

MCPHERRON, A. C.; LAWLER, A. M.; LEE, S. J. Regulation of skeletal muscle mass in mice by a new TGF-beta superfamily member. **Nature**. V. 387, p. 83:90, 1997.

MEIRA, C. T.; FORTES, M. R. S.; FARAH, M. M.; PORTO-NETO, L. R.; CURI, R. A.; MOORE, S. S.; et al. A genome-wide association study for height at withers in racing quarter horse.

Association for the Advancement of Animal Breeding and Genetics. V. 20, p.420:423, 2013.

MEIRA, C.T.; FARAH, M.M.; FORTES, M.R.S.; MOORE, S.S.; PEREIRA, G.L.; SILVA, J.A.II V.; MOTA, M.D.S.; CURI, R.A. A Genome-wide association study for morphometric traits in Quarter Horse. **Journal of Equine Veterinary Science.** V. 34 ed. 8, p. 1028:1031, 2014a.

MEIRA, C.T.; FORTES, M.R.S.; FARAH, M.M.; PORTO NETO, L.R.; KELLY, M.; MOORE, S.S.; PEREIRA, G.L.; CHARDULO, L.A.L.; CURI, R.A. Speed Index in the Racing Quarter Horses: A Genome wide association study. **Journal of Equine Veterinary Science.** V. 34 ed. 11-12, p. 1263:1268, 2014b.

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

MEIRA, C.T.; **Identificação de regiões genômicas selecionadas de forma divergente em equinos quarto de milha de corrida e trabalho.** 2014. 95 f. Tese (Doutorado em Genética e melhoramento animal) – Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Jaboticabal, 2014d.

MUIR, W. M.; WONGB, G. K.; ZHANG, Y.; WANGC, J.; GROENEND, M. A. M.; CROOIJMANS, R. P. M. A.; MEGENS, H. J.; ZHANG, H.; OKIMOTO, R.; VEREIJEN, A.; JUNGIERUS, A.; ALBERS, G. A. A.; LAWLEY, C. T.; DELANVI, M. E.; MACEACHERNE, S.; CHENG, H. H. Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. **Proceedings of the National Academy of Sciences of the United States of America.** V.105, p.17312:17317, 2008.

NAKAYA, H. L.; AMARAL, P. P.; LOURO, R.; LOPES, A.; FACHEL, A. A.; MOREIRA, Y. B.; EL-JUNDI, T. A.; SILVA, A. M.; REIS, E. M.; VERJOVSKI-ALMEIDA, S. Comment reviews reports deposited research refereed research interactions information Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. **Genome Biology.** V.8 (R43), p.1-25, 2007.

NISHI, M.; YASUE, A.; NISHIMATU, S.; NOHNO, T.; YAMAOKA, T.; ITAKURA, M.; MORIYAMA, K.; OHUCHI, H.; NOJI, S. A missense mutant myostatin causes hyperplasia without hypertrophy in the mouse muscle. **Biochemical and Biophysical Research Communications.** V. 293, p.247–251, 2002.

PEREIRA, G.L.; MATTEIS, R.; REGITANO, L.C.A.; CHARDULO, L.A.L.; CURI, R.A. MSTN, CKM and DMRT3 gene variants in different lines of Quarter Horses. **Journal of Equine Veterinary Science**. V. 39, p. 33:37, 2016a.

PEREIRA, G.L.; MATTEIS, R.; REGITANO, L.C.A.; MEIRA, C.T.; SILVA, J.A.II.V.; CHARDULO, L.A.L.; CURI, R.A. Comparison of sequence variants in the PDK4 and COX4I2 genes between racing and cutting lines of quarter horses and associations with the speed index. **Journal of Equine Veterinary Science**. V. 39, p. 1:6, 2016b.

PEREIRA, G.L.; **Identificação de regiões cromossômicas, genes e polimorfismos de DNA associados ao desempenho de equinos de corrida da raça quarto de milha**. 2017. 98 f. Tese (Doutorado melhoramento animal corrente em Genética e) – Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista "Júlio de Mesquita Filho", Jaboticabal, 2017.

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

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

SANTAGOSTINO M., KHORIAULI L., GAMBA R., BONUGLIA M., KLIPSTEIN O., PIRAS F.M., VELLA F., RUSSO A., BADIALE C., MAZZAGATTI A., RAIMONDI E., NERGADZE, S.G., GIULOTTO E. Genome-wide evolutionary and functional analysis of the Equine Repetitive Element 1: an insertion in the myostatin promoter affects gene expression. **BMC Genet**. V. 26, p. 16:126, 2015.

SCHUELKE, M.; ANGER, K.R.; STOLZ, L.E.; HUBNER, C.; RIEBEL, T. et al. Myostatin mutation associated with gross muscle hypertrophy in a child. **The New England Journal of Medicine**. V. 350, p. 2682:2688, 2004.

SCHRÖDER, W.; KLOSTERMANN, A.; STOCK, K. F.; DISTL, O. A genome-wide association study for quantitative trait loci of show jumping in Hanoverian Warm blood horses. **Animal Genetics**. V. 43(4), p. 392:400, 2011.

TOZAKI, T.; MIYAKE, T.; KAKOI, H.; GAWAHARA, H.; SUGITA, S.; HASEGAWA, T. et al A genome-wide association study for racing performances in Thoroughbreds clarifies a candidate region near the MSTN gene. **Animal Genetics**. V. 41, p.28-35, 2010.

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

WALLIN, L.; STRANDBERG, E.; PHILIPSSON, J. Genetic correlations between field test results of Swedish Warmblood Riding Horses as 4-year-olds and life time performance results in dressage and show jumping. **Livestock Production Science**. V. 82, p. 61:71, 2003.

WHEELAN, S.J.; AIZAWA, Y.; HAN, J.S.; JEF D. BOEKE, J.D. Gene-breaking: A new paradigm for human retrotransposon-mediated gene evolution. **Genome**. V. 15, p. 1073:1078, 2005.

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

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

Capítulo 2 – Variantes da região cromossômica do gene da miostatina e suas relações com linhagens, desempenho e medidas corporais na raça Quarto de Milha

Resumo - Considerando as diferenças nas medidas corporais e na musculatura que ocorrem entre as linhagens de corrida e de trabalho da raça equina Quarto de Milha, o objetivo deste trabalho foi analisar a ocorrência de relações entre alelos de polimorfismos da região cromossômica do gene da miostatina *MSTN* (ECA18), um regulador negativo do desenvolvimento muscular, e as duas linhagens da raça. Outro objetivo foi realizar análises de associação de polimorfismos dessa região cromossômica com valor genético estimado (EBV) do desempenho em corridas, dado pelo índice de velocidade máximo (IV max), e EBVs da altura à cernelha (AC), perímetro torácico (PT) e comprimento corporal (CC) na linhagem de corrida, e com EBVs da AC, PT e CC na linhagem de trabalho. Foram utilizadas informações genômicas, fenotípicas e de pedigree de 420 equinos, de ambos os sexos, registrados na associação Brasileira de criadores (ABQM), sendo 352 da linhagem de corrida e 68 da de trabalho. Na região genômica estudada (gene *MSTN* $\pm$ 2Mb), foram identificados 46 SNPs e 1 SINE – ERE1 comuns às linhagens de trabalho e de corrida, dos quais 32 SNPs e 1 SINE apresentaram alelos com frequências significativamente diferentes entre às mesmas. Também foi constatado que a porção dessa região mais próxima ao gene *MSTN* ($\pm$ 1Mb) é menos polimórfica na linhagem de corrida em relação à de trabalho, o que pode ser consequência de maior pressão de seleção sobre essa região na linhagem de corrida e/ou em função do uso de animais Puro Sangue Inglês na mesma. Em relação aos polimorfismos associados a fenótipos (p não ajustado $< 0,05$), destacaram-se os relacionados à altura de cernelha na linhagem de corrida devido ao seu grande número (p não ajustado $< 0,05$) e alto desequilíbrio de ligação. Ademais estes marcadores apresentaram frequências do alelo favorável significativamente maiores em relação à linhagem de trabalho. Tendo se em vista a possibilidade de associação genética entre características morfológicas e desempenho em equinos, esse resultado é de grande importância, fornecendo informações para futuros estudos de melhoramento genético animal.

Palavras-chave: cavalo, ERE1, marcadores moleculares, QTL, SNP, velocidade.

Chapter 2 - Variants of the chromosomal region of the myostatin gene and its relationships with lineages, performance and body measurements in the quarter-mile breed

Abstract - Considering the differences in body and muscle measurements that occur between racing and working strains of the Quarter-mile breed, the objective of this work was to analyze the occurrence of relations between polymorphic alleles of the myostatin gene (*MSTN*) (ECA18), a negative regulator of muscle development, and the two lineages of the breed. Other objective was to perform association analyses of these polymorphisms with estimated genetic value (EBV) of the performance in races, given by the maximum velocity index (IV max), and EBVs of height at withers (HW), heart girth (HG) and body length (BL) in the racing lineage, and with EBVs of AC, PT and CC in the working line. Genomic, phenotypic and pedigree information of 420 equines, of both sexes, registered in the Brazilian breeders association (ABQM), 352 of the race line and 68 of the work line were used. In the studied genomic region (*MSTN* gene $\pm$ 2Mb), 46 SNPs and 1 SINE - ERE1 were identified common to the working and racing strains, of which 32 SNPs and 1 SINE presented alleles with significantly different frequencies between them. It has also been found that the portion of this region closest to the *MSTN* gene ($\pm$ 1Mb) is less polymorphic in the racing versus the working strain, which may be consequence of higher selection pressure on that region in the racing line and / or due to the use of purebred English animals in it. In relation to the polymorphisms associated with the phenotypes (p not adjusted <0.05), those related to the height of the withers in the racing line due to their large number (p not adjusted <0.05) and high linkage disequilibrium were the most prominent. In addition, these markers presented favorable allele frequencies significantly higher in relation to the working lineage. Considering the possibility of genetic association between morphological characteristics and performance in equines, this is a result of great importance, providing information for future animal breeding studies.

Key words: equine, ERE1, molecular markers, QTL, SNP, speed.

1. Introdução

A seleção para diferentes finalidades na raça de cavalos Quarto de Milha (EVANS, 1996) levou à formação de linhagens, com destaque para as de corrida e de trabalho. A linhagem de corrida explora a aptidão dos animais quanto à velocidade em pistas retas e de curta distância. Equinos desta linhagem têm melhor desempenho em corridas de curtas distâncias do que qualquer outra raça ou linhagem (ABQM, 2016). Nestas corridas, o metabolismo energético predominantemente no músculo dos animais é o anaeróbio (FREEMAN, 2013). O principal quesito para a seleção de reprodutores, utilizado pelos criadores do cavalo Quarto de Milha de corrida, é o escore de pontuação conhecido como índice de velocidade (IV). Este índice é obtido durante a campanha do animal com o intuito de classificar o seu desempenho em diferentes condições (distâncias, hipódromo, clima, país) (EVANS, 1996). Por sua vez, a linhagem de trabalho destina-se às provas de caráter funcional, explorando habilidades como agilidade e obediência, características consideradas de grande importância no manejo do gado a campo (MEIRA et al., 2014). Nestas provas o metabolismo energético predominantemente é o aeróbio (FREEMAN, 2013). Assim, a seleção na linhagem é baseada principalmente em características funcionais, premiações e pedigree.

Em geral, o cavalo de corrida da raça Quarto de Milha é mais alto que o de trabalho devido à influência de animais da raça Puro-Sangue Inglês (PSI), utilizados para melhorar a sua velocidade. Entretanto, a seleção direcionada a diferentes objetivos (corrida ou trabalho) também promoveu mudanças significativas nas características corporais dos equinos Quarto de Milha (ABQM, 2017). Em trabalho realizado por Meira et al. (2014), foram observados maiores peso, altura, comprimentos e perímetros corporais para os animais da linhagem de corrida em relação aos de trabalho. Na literatura, há evidências de associação genética entre características morfológicas e desempenho ao trabalho em equinos (JÖNSSON et al., 2014), principalmente com relação às características de conformação de pernas e altura. Nesse sentido, a possibilidade de utilizar características morfológicas como indicadores de performance, cuja mensuração exige maior complexidade, torna o estudo dos seus componentes genéticos relevante em raças de equinos utilizadas para o trabalho.

É provável que, assim como em humanos (BRAY et al., 2009), o desempenho atlético em equinos seja influenciado por grande número de genes. Entretanto, até o momento poucas variantes genéticas foram claramente relacionadas à característica, exclusivamente em animais PSI, entre estas, polimorfismo de nucleotídeo único (SNP) do gene da miostatina (*myostatin* – *MSTN*; 66490208..66495180) (HILL et al., 2010a;b; BINSS et al., 2010, TOZAKI et al., 2010), o qual se encontra mapeado no cromossomo equino 18 (ECA18). O gene *MSTN* é expresso no tecido muscular esquelético e atua como regulador negativo do crescimento da massa muscular. Várias mutações ou polimorfismos foram identificados no *MSTN* em bovinos (GROBET et al. 1997; KAMBADUR et al., 1997; MCPHERRON et al., 1997), ovinos (CLOP et al., 2006), em camundongos (NISHI et al., 2002) e em humanos (SCHUELKE et al., 2004), resultando em fenótipos caracterizados por hiperplasia e hipertrofia muscular, levando ao aumento de massa corporal. Santagostino et al. (2015) estudaram o efeito de um transposon do tipo SINE, conhecido como Elemento Repetitivo Equino 1 (ERE1), nas sequências promotoras do gene *MSTN* (g.66495342:ECA18) em ensaio de gene repórter, e mostraram que a inserção provoca redução da expressão gênica. Os autores sugeriram que a inserção ERE1 (alelo ERE1+) pode diminuir os níveis de miostatina, e, assim, incrementar o desenvolvimento muscular em cavalos. Ainda de acordo com Santagostino et al. (2015), a inserção ERE1 no gene da miostatina é polimórfica em cavalos (alelos ERE1+ e ERE1-) e parece estar relacionada à aptidão para corridas de diferentes distâncias (curtas ou longas).

Tendo em vista o papel do gene da miostatina (*MSTN*) na fisiologia do músculo esquelético e considerando as diferenças de crescimento e de desenvolvimento muscular que ocorrem entre as linhagens de corrida e de trabalho do Quarto de Milha e ainda que outros genes em desequilíbrio de ligação com o *MSTN* também podem afetar estas características, o objetivo deste trabalho foi analisar a ocorrência de diferença de frequências alélicas de polimorfismos da região cromossômica do *MSTN* entre as duas linhagens da raça. Considerando que os efeitos de polimorfismos de DNA sobre fenótipos são parâmetros intrínsecos de cada linhagem ou raça em determinado ambiente, outro objetivo deste trabalho foi realizar análises de associação de polimorfismos da região cromossômica do gene *MSTN* com valores genéticos estimados (EBVs) do índice de velocidade máximo (IV max), altura de

cernelha (AC), perímetro torácico (PT) e comprimento corporal (CC) na linhagem de corrida, e com EBVs de AC, PT e CC na linhagem de trabalho.

2. Material e métodos

2.1. Animais amostrados e coleta dos fenótipos

Nos anos de 2011, 2013 e 2015 foram coletadas amostras de sangue de 420 equinos Quarto de Milha, de ambos os sexos, registrados na associação Brasileira de criadores da raça (ABQM), sendo 352 da linhagem de corrida e 68 da de trabalho. Os animais da linhagem de corrida, 77 machos e 275 fêmeas, eram filhos de 83 garanhões e 246 éguas, resultando em média de 4,2 progênie/garanhão e 1,4 progênie/égua. Estes animais pertenciam a 157 criadores e encontravam-se alojados em 24 propriedades no interior do estado de São Paulo, Brasil e no Jockey Club de Sorocaba, São Paulo, Brasil. Por sua vez, os animais da linhagem de trabalho, 25 machos e 43 fêmeas, eram filhos de 44 garanhões e 65 éguas, resultando em média de 1,5 progênie/garanhão e 1,04 progênie/égua. Estes animais pertenciam a 45 criadores e encontravam-se alojados em três propriedades no interior do estado de São Paulo, Brasil. A coleta e, conseqüentemente, a presença de irmãos completos na amostra foi evitada em ambas as linhagens.

Os dados de desempenho dos animais Quarto de Milha de corrida foram obtidos junto ao Departamento de Estatística e banco de dados online do Jockey Club de Sorocaba (<http://www.jcsorocaba.com.br/indice-de-velocidade/>) além do banco de dados online Equibase (<http://www.equibase.com/homehorseplayer.cfm>). Para a maior parte dos animais utilizados foram encontradas informações de índice de velocidade (IV) de vários anos, obtidas ao longo dos seus históricos de competição. Entretanto, para outros foi encontrada somente a informação de IV máximo (IV max). Como o IV médio apresenta alta correlação com o IV max ($r = 0,8762$; PEREIRA et al., 2016), optou-se, neste estudo, por utilizar o IV max a fim de que não ocorresse perda significativa de informações de desempenho.

As dimensões corporais, altura de cernelha (AC), perímetro torácico (PT) e comprimento corporal (CC), foram medidas de acordo com Torres e Jardim (1992), ou seja, com os animais em estação forçada, isto é, com membros anteriores e

posteriores na perpendicular sobre piso plano, sempre do lado direito do animal, utilizando-se hipômetro e fita métrica.

Todos os procedimentos que envolveram a manipulação de animais foram realizados segundo as normas brasileiras de bem estar animal (protocolo nº 009/2017-CEUA e expedido pela Comissão de Ética no Uso de Animais – CEUA da Faculdade de Medicina Veterinária e Zootecnia da Unesp, Botucatu, São Paulo, Brasil).

2.2. Extração do DNA e genotipagem de polimorfismos da região do *MSTN* no ECA18

O processo de extração do DNA a partir de alíquotas de sangue foi realizado por meio do *Illustra Blood GenomicPrep Mini Spin Kit* (GE Healthcare, EUA), de acordo com as instruções do fabricante.

A determinação dos genótipos ERE1+/ERE1+, ERE1+/ERE1- e ERE1-/ERE1- do Elemento Repetitivo Equino 1 do gene *MSTN* foi realizada por meio da técnica de PCR, na qual fragmentos de 436 e/ou 209pb da região regulatória do gene *MSTN* equino foram amplificados de acordo com Santagostino et al. (2015).

A genotipagem de SNPs mapeados na região de interesse do cromossomo equino 18 (ECA18) foi realizada por meio de arranjos comerciais. Cento e oitenta dos 420 animais utilizados neste estudo, 112 da linhagem de corrida e 68 da de trabalho, foram genotipados utilizando o *EquineSNP50 BeadChip* – 54K (Illumina, Inc., EUA). As demais amostras de DNA da linhagem de corrida (n = 240) foram genotipadas com o *EquineSNP70 BeadChip* – 65K (Illumina, Inc., EUA). Os controles de qualidade (*quality control* – QC) das informações de genotipagem, para indivíduos analisados por ambos os arranjos 54K e 65K (n= 180 e n= 240) e SNPs, foram realizados por meio do programa *R* v.3.3.1 (R Core Team, 2016). Para os SNPs, foram excluídos os localizados no cromossomo X, com *call rate* inferior a 0,9, e com *p-value* inferior a 1×10^{-5} para o equilíbrio de *Hardy-Weinberg* (*Hardy-Weinberg Equilibrium* – HWE).

O sequenciamento do exoma de subamostra de 39 animais da linhagem de corrida da raça Quarto de Milha foi realizado com o intuito de identificar e genotipar novos SNPs e InDels na região do *MSTN* no ECA18. O preparo de bibliotecas, a captura e o enriquecimento de regiões de exon, 5' e 3' UTR para sequenciamento foram realizados em dois ciclos conforme o *User's Guide* do *Seq Cap EZ HyperCap Workflow v1.0* (Roche Inc., EUA). Em cada ciclo foram feitos *pools* com 20 e 19

amostras de DNA multiplexadas. As bibliotecas preparadas foram sequenciadas por meio do equipamento *HiSeq 2500* (Illumina, EUA) em duas corridas *Rapid Run* de leituras 2 x 100 nucleotídeos (*paired end*). O mapeamento das *reads* em relação ao genoma equino de referência (EquCab 2.0) foi realizado para cada amostra (uma por vez), seguindo os padrões (*default*) do item *mapping to reference* do programa *CLC Workbench Genomics v.9.5.2* (Qiagen, EUA). A identificação de polimorfismos (SNPs e InDels), para cada amostra em relação à referência foi realizada utilizando o item *fixed ploidy variants* do programa *CLC Workbench Genomics* com os seguintes parâmetros: probabilidade de variante de 90%, frequência mínima de 20% e qualidade de mapeamento mínima de 20. Adicionalmente, foi utilizado um segundo filtro de qualidade (*marginal variant filter – MVF*) para evitar o *strand bias*. Este filtro excluiu variantes com balanço *forward-reverse* menor que 0,05, o que afeta principalmente polimorfismos do tipo InDel. Considerou-se cobertura mínima de sequenciamento de 20x, frequência mínima de amostras genotipadas para uma variante de 0,5, e desvio de HWE < 0,001.

2.3. Imputação de genótipos na linhagem de corrida

A imputação de genótipos foi realizada nos animais da linhagem de corrida da raça Quarto de Milha por meio do programa FImpute v.2.2 (SARGOLZAEI et al., 2014).

Nos animais genotipados com o arranjo de 54K foram imputados os SNPs, inclusive os da região de interesse do ECA18, exclusivos do arranjo de 65K, e nos indivíduos genotipados com o arranjo de 65K foram imputados os SNPs, inclusive os da região de interesse do ECA18, exclusivos do arranjo de 54K. As informações utilizadas para a realização da imputação de genótipos em duas vias (i.e. do 54K para o de 65K e do 65K para o de 54K), como descrita por Pereira et al. (2016), foram relativas ao número de indivíduos e SNPs obtidos após o QC descrito no item 2.2. Na imputação em duas vias independentes foi utilizado primeiramente o painel de 65K como referência e o 54K como imputado. Para isso foram considerados todos os marcadores presentes no arranjo de 65K e somente os marcadores em comum do SNP50 (45K). Os marcadores não compartilhados (9K) foram novamente acrescentados após a imputação. O mesmo procedimento foi realizado na segunda via, em que houve a inversão dos painéis utilizados como referência e imputado.

Posteriormente, ambos os painéis imputados foram ordenados e concatenados, formando apenas um conjunto com indivíduos genotipados para todos os marcadores disponíveis nos painéis de 54K e 65K.

Por fim, os genótipos dos novos SNPs e InDels da subamostra de 39 animais da linhagem de corrida, identificados da região do *MSTN* do ECA18 por meio da análise do exoma, foram imputados nos cavalos que completam a amostra da linhagem. A acurácia de imputação foi obtida por meio da estimativa de r^2 alélico, o qual foi determinado pelo quadrado da correlação entre a dosagem alélica estimada dos genótipos imputados ($0x(AA)+1x(AB)+2x(BB)$) e da dosagem alélica dos genótipos verdadeiros.

2.4. Variantes genômicas da região do *MSTN* no ECA18

Para a determinação das variantes genômicas, bem como da estrutura haplotípica na região do gene *MSTN* do ECA18, esta última realizada por meio do programa Haploview 4.2 (BARRETT et al, 2005), foram considerados 2Mb *upstream* e 2Mb *downstream* da sequência de nucleotídeos correspondente ao gene da miostatina (66490208..66495180), tendo como base a montagem referência do genoma equino (EquCab 2.0).

Os blocos de haplótipos foram obtidos para esta região para cada linhagem separadamente. A partir do desequilíbrio de ligação estimado por D' , os blocos foram inferidos assumindo que, em cada um, houvesse 95% de marcadores em alto DL (0,7 a 0,98). Foram considerados para essas análises marcadores com $MAF > 0,05$. Foram obtidas as descrições e frequências dos haplótipos existentes dentro de cada bloco. As recombinações entre os haplótipos dos blocos também foram dadas por D' multialélico.

As análises das diferenças de frequências alélicas entre as linhagens de corrida e de trabalho para o ERE1 e os SNPs comuns aos arranjos de 54 e 65K foram realizadas utilizando o teste Caso-controle por meio do pacote *genetics* do *software* R (WARNES, 2013). Para os polimorfismos com diferença de frequências alélicas entre as linhagens estimou-se dois a dois os desequilíbrios de ligação (DL) por meio de r^2 (HILL: ROBERTSON, 1968) utilizando o programa Plink 1.9 (PURCELL et al., 2007).

Análises do equilíbrio de Hardy-Weinberg (*Hardy-Weinberg Equilibrium* – HWE) foram conduzidas nas linhagens de corrida e de trabalho considerando para a primeira o ERE1, os SNPs comuns aos dois chips e os imputados entre chips e os SNPs e InDels imputados dos animais que tiveram o exoma analisado para os genotipados com os chips, e para a segunda o ERE1, e os SNPs do arranjo de 54K. No estudo do HWE foi utilizado o teste χ^2 do pacote *genetics* do *software* R (WARNES, 2013). A correção FDR para comparações múltiplas (BENJAMINI & HOCHBERG, 1995) foi aplicada para ajuste das significâncias obtidas.

2.5. Associação entre variantes genômicas e características de interesse

Para obtenção das estimativas de valor genético (EBV) para IV max foi utilizado o modelo dado por: $y = X\beta + Z\alpha + e$, onde y é o vetor de observações; β é o vetor de efeitos fixos de sexo, idade à corrida (2, 3 e 4 anos) e ano da corrida referentes à observação do IV max (1988 a 2015); α é o vetor de efeitos genéticos aditivos diretos; X é a matriz de incidência; Z é a matriz de incidência dos efeitos aleatórios genéticos aditivos diretos; e é o vetor de efeito residual. Para as características morfométricas de interesse foi utilizado modelo semelhante, no qual os efeitos fixos considerados foram sexo e ano de coleta (2011, 2013 e 2015). A idade do animal à coleta (linhagem corrida - 3 a 26 anos; linhagem de trabalho 2 a 21 anos) foi considerada no modelo como covariável linear. Os EBVs das características morfométricas e de desempenho foram estimados por meio do *software blupf90* (MISZTAL, 2002).

Análises de associação entre genótipos dos polimorfismos individuais e EBVs das características de interesse, realizadas por meio do programa Plink 1.9 (PURCELL et al., 2007), foram conduzidas nas linhagens de corrida e de trabalho da raça Quarto de Milha considerando os mesmos modelos descritos acima. Nas análises de associação na linhagem de corrida foram utilizados os EBVs de IV max, AC, PT e CC e os genótipos dos polimorfismos individuais da região do *MSTN* no ECA18. Nas análises de associação na linhagem de trabalho foram utilizados os EBVs de AC, PT e CC e os genótipos dos polimorfismos individuais da região do *MSTN* no ECA18. A correção FDR para comparações múltiplas (BENJAMINI & HOCHBERG, 1995) foi aplicada para ajuste das significâncias obtidas.

Para os polimorfismos associados à EBVs nas linhagens de corrida e de trabalho, estimou-se o desequilíbrio de ligação (DL) por meio de r^2 (HILL; ROBERTSON, 1968) utilizando o programa Plink 1.9 (PURCELL et al., 2007) a fim de determinar se os efeitos observados dos marcadores sobre as características são independentes ou não.

3. Resultados e discussão

3.1. Variantes genômicas da região do *MSTN*

O controle de qualidade de genotipagem de indivíduos excluiu 13 animais da linhagem de corrida e quatro na linhagem de trabalho. Também em função do QC, foram excluídos, na região de interesse, 53 SNPs dos animais de corrida e 19 dos de trabalho. Os genótipos do ERE1 para estes 17 animais excluídos pelo QC foram desconsiderados. Dessa forma, permaneceram para as análises subsequentes 339 animais da linhagem de corrida e 64 animais da linhagem de trabalho. Na primeira corrida de sequenciamento do exoma (n= 20) foram obtidas 2.498.929.809 sequências pareadas (*paired-end*) e para a segunda (n= 19) 1.629.737.973. Foram observadas leituras de qualidade boa para toda extensão das *reads* sequenciadas (até 100pb) e cobertura média de 43x. Com relação à imputação em duas vias de genótipos exclusivos entre os painéis equinos de 54K e 65K, observou-se aumento considerável de SNPs para todas as amostras (12 a 26%), sem perdas na qualidade das informações. As médias de taxa de concordância do arranjo de 65K para o de 54K e do arranjo de 54K para o de 65K foram 0,9815 e 0,9751 e para r^2 alélico foram 0,9791 e 0,9740, respectivamente. Na imputação dos genótipos do sequenciamento, o r^2 alélico estimado foi de 0,78. Este resultado, consideravelmente menor em relação à imputação dos animais genotipados com arranjos de SNPs, foi devido à MAF com valores baixos e ao pequeno número de animais na população referência.

A região de interesse deste estudo foi concentrada especificamente no loco do gene *MSTN* (66490208..66495180) $\pm$ 2Mb no ECA18, ou seja, em segmento de 4.004.972 pb, com início no nucleotídeo 64.490.208 e término na posição 68.495.180. Dentro desta região, foram identificados, com base no genoma de referência equino (EquCab 2.0), 24 genes conhecidos (incluindo o *MSTN*). Na linhagem de corrida foram identificados 97 polimorfismos, sendo 89 SNPs, sete InDels e um SINE (ERE1).

Destes, 23 (seis InDels e 17 SNPs) são variantes novas identificadas pelo sequenciamento de DNA. Na linhagem de trabalho foram identificados 89 SNPs.

Na linhagem de corrida foram identificados cinco blocos de haplótipos com tamanhos variando entre 12 e 437 Kb (Figura 1). Apesar do alto DL indicado em algumas regiões do heatmap, não foram inferidos blocos nas proximidades do *MSTN* ($66,5 \pm 1\text{Mb}$) devido à frequência extremamente baixa dos alelos menores, os quais estão próximos da fixação. Um grande bloco (B1) foi inferido no início da região estudada ($\sim 64,5$ a 65 Mb). Nota-se que muitos polimorfismos adjacentes a essa região estão em DL com variantes que integram o bloco. Observa-se ainda nesta linhagem, a menor densidade de polimorfismos (após QC) próximo à região do *MSTN* ($\pm 500\text{Kb}$) se comparada a outras regiões. Por outro lado, destaca-se o maior número de polimorfismos na porção final da região genômica de interesse (a partir de 67 Mb), onde foi observado maior concentração de genes e, portanto, maior contribuição de variantes prospectadas por meio do sequenciamento do exoma.

Na linhagem de trabalho foram identificados quatro blocos de haplótipos com tamanhos variando entre 1 e 333 Kb . Os maiores blocos foram encontrados na região do *MSTN* ($\sim 65,8$ a $66,6\text{ Mb}$), a qual ainda se encontra polimórfica nesta linhagem. Os resultados obtidos são importantes para a maior compreensão das fases de ligação entre os polimorfismos e estruturas haplotípicas da região genômica estudada nas linhagens de corrida e de trabalho da raça Quarto de Milha.

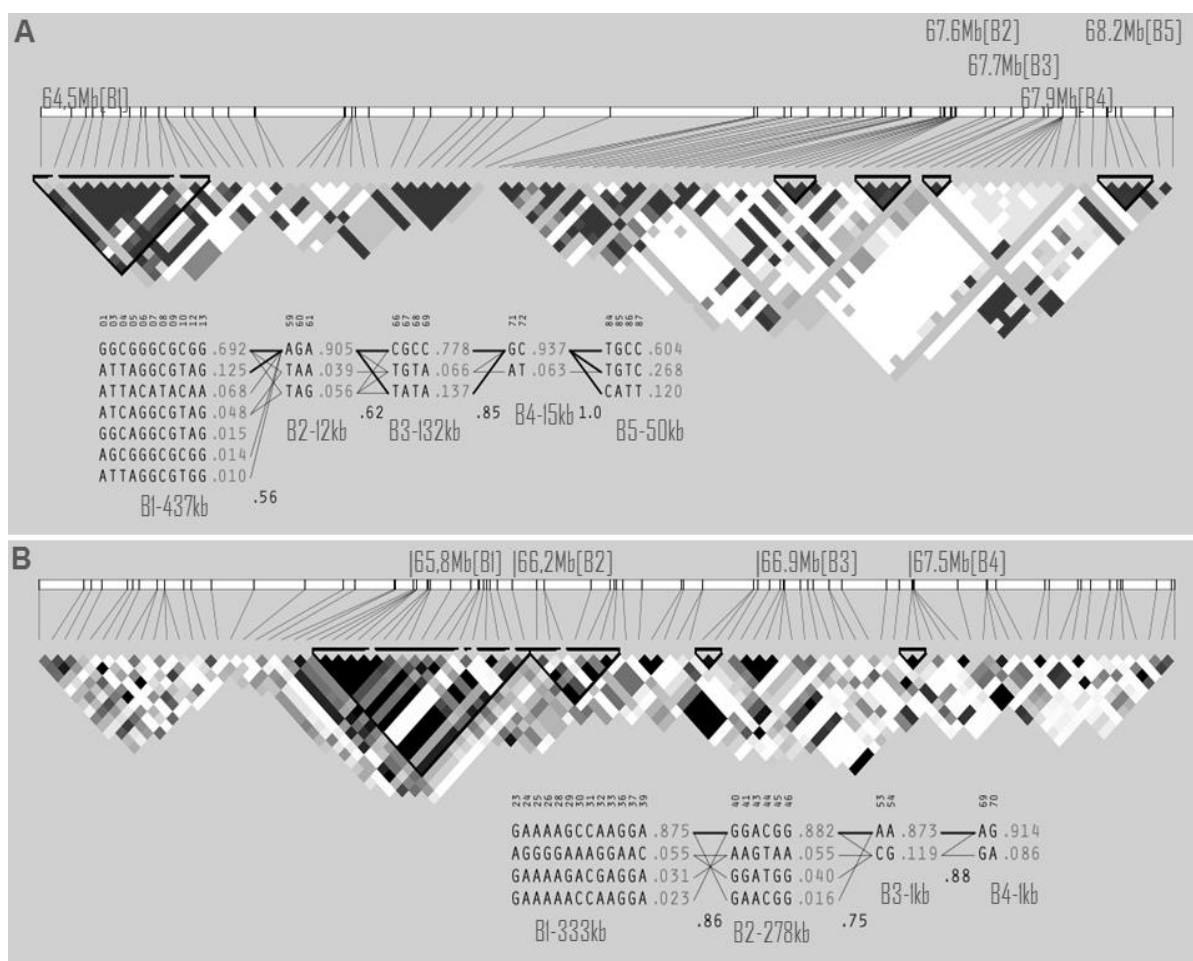


Figura 1 – Blocos de haplótipos estimados a partir do desequilíbrio de ligação (D') entre variantes na região do gene *MSTN* no cromossomo 18 (64,5 - 68,4Mb) em equinos da raça Quarto de Milha das linhagens de corrida (A) e de trabalho (B). A posição (Mb), bem como a identificação de cada bloco ([Bn], $n=1,2,\dots$) são apresentados no topo de cada figura. Diferentes tonalidades de cinza indicam a proporção do D' entre marcadores (cinza claro/branco $< 0,7$ e cinza escuro/preto $> \text{ou} = 0,7$). A identificação, o tamanho e as frequências dos haplótipos de cada bloco são dados abaixo dos *heatmaps*. A recombinação de haplótipos entre os diferentes blocos (valor de D' multialélico) está indicada entre as representações das frequências haplotípicas. Linhas mais grossas indicam as combinações mais frequentes entre os haplótipos dos diferentes blocos.

3.2. Variantes genômicas da região do *MSTN* associadas à linhagem

Na região genômica de interesse, foram identificados 47 polimorfismos comuns às linhagens de trabalho e de corrida (46 SNPs e 1 SINE – ERE1), dos quais 33 (32

SNPs e 1 SINE) foram significativamente associados às mesmas por meio do teste Caso-controle (Tabela 1). Na linhagem de trabalho, os polimorfismos BIEC2-417594 e BIEC2-417604 não se encontram em equilíbrio de Hardy-Weinberg (HWE) ($P < 0,05$). Os demais polimorfismos da região analisada, ou seja, os não significativos, encontram-se em HWE em ambas as linhagens, exceto o BIEC2-417608 ($p = 0,0459$) na linhagem de trabalho.

O dois SNPs significativamente associados às linhagens e que não se encontram em equilíbrio de HW na linhagem de trabalho (BIEC2-417594 e BIEC2-417604) foram mapeados entre os genes Caveolae Associated Protein 2 (*CAVIN2*) e Transmembrane Protein With EGF Like And Two Follistatin Like Domains 2 (*TMEFF2*) e no intron cinco do *TMEFF2*, respectivamente. Em pesquisa conduzida por Meira et al. (2014) com subamostra dos animais utilizados aqui (120 da linhagem de corrida e 64 da linhagem de trabalho), encontrou-se, por meio da estimação do parâmetro $\theta = FST$, via método Bayesiano proposto por Gianola et al. (2010), um cluster constituído de 1333 loci (SNPs), distribuídos ao longo de todo o genoma, com altos valores de θ , refletindo possíveis assinaturas de seleção com probabilidade maior que 95%. Entre estes SNPs estavam o BIEC2-417594 e BIEC2-417604 com probabilidades superiores a 99% de representarem assinaturas de seleção. Destaca-se aqui, em função dos processos biológicos do qual participa, o *CAVIN2*. Chidlow Jr. e Sessa (2010) discorreram que este gene está relacionado a um conjunto de proteínas de revestimento (caveolinas), expressas principalmente em músculo liso vascular, músculo cardíaco e esquelético. Desta forma, a função deste grupo de proteínas está diretamente ligada com a função vascular e a pressão sanguínea.

Tabela 1 – Polimorfismos significativamente associados às linhagens de trabalho e de corrida, posição no genoma, identificação do alelo menor (AM) e sua frequência (MAF), e equilíbrio de Hardy-Weinberg (HWE) para as duas linhagens.

Polimorfismo	Posição	AM	Trabalho		Corrida		<i>p</i>	<i>p</i> ajustado
			MAF	<i>p</i> (HWE)	MAF	<i>p</i> (HWE)		
BIEC2-417034	g.64.505.667 A>G	A	0,148	0,613	0,278	0,142	2,06E-03	4,41E-03
BIEC2-417054	g.64.663.252 T>G	T	0,148	1	0,267	0,133	4,47E-03	8,08E-03
BIEC2-417067	g.64.687.870 T>C	T	0,109	0,549	0,216	0,057	5,48E-03	9,54E-03
BIEC2-417075	g.64.725.066 A>G	A	0,167	1	0,291	0,604	3,89E-03	7,62E-03
BIEC2-417120	g.64.919.213 T>C	T	0,113	0,562	0,213	1	9,80E-03	1,59E-02
BIEC2-417121	g.64.919.265 A>G	A	0,063	0,208	0,023	0,162	1,49E-02	2,33E-02
BIEC2-417124	g.64.943.005 A>G	A	0,094	1	0,259	1	4,72E-05	1,17E-04
BIEC2-417135	g.65.009.739 A>G	A	0,102	1	0,187	1	1,87E-02	2,83E-02
BIEC2-417139	g.65.038.492 A>G	A	0,078	1	0,142	0,826	4,79E-02	6,83E-02
BIEC2-417145	g.65.107.716 A>G	A	0,07	1	0,192	0,607	8,48E-04	1,90E-03
BIEC2-417187	g.65.565.128 T>C	T	0,106	1	0,001	1	5,16E-17	2,43E-15
BIEC2-417192	g.65.606.457 A>G	A	0,085	1	0,001	1	4,55E-11	2,38E-10
BIEC2-417210	g.65.809.482 A>G	A	0,099	1	0,001	1	8,87E-15	1,04E-13
BIEC2-417274	g.65.868.604 T>G	T	0,1	1	0,001	1	5,07E-16	7,95E-15
BIEC2-417291	g.66.010.474 T>C	T	0,09	1	0,001	1	1,45E-14	1,37E-13
BIEC2-417298	g.66.054.465 A>G	A	0,038	1	0,001	1	4,34E-06	1,27E-05
BIEC2-417306	g.66.102.331T>C	T	0,032	1	0,001	1	5,01E-05	1,18E-04
BIEC2-417308	g.66.155.365 A>C	T	0,067	1	0,001	1	1,06E-09	5,00E-09
BIEC2-417328	g.66.265.005 A>G	T	0,091	0,422	0,001	1	3,98E-12	2,34E-11
ERE1	g.66.495.342 ERE	ERE-	0,103	1	0,001	1	4,47E-16	7,95E-15
BIEC2-417453	g.66.996.871 A>C	A	0,115	0,586	0,007	1	7,04E-14	4,72E-13
BIEC2-417454	g.66.996.889 T>C	T	0,123	0,58	0,009	1	3,94E-14	3,09E-13
BIEC2-417458	g.67.010.090 A>C	A	0,067	1	0,007	1	3,86E-05	1,01E-04
BIEC2-417484	g.67.103.337A>G	A	0,078	1	0,028	0,245	6,27E-03	1,05E-02
BIEC2-417495	g.67.186.093 A>G	A	0,131	0,583	0,03	0,268	1,09E-06	4,65E-06
BIEC2-417504	g.67.259.711 T>C	T	0,048	0,117	0,001	1	2,96E-06	1,16E-05
BIEC2-417523	g.67.503.227 T>C	T	0,039	1	0,001	1	4,34E-06	1,27E-05
BIEC2-417524	g.67.545.703 A>G	A	0,086	1	0,032	1	4,08E-03	7,66E-03
BIEC2-417527	g.67.545.806 T>C	T	0,087	1	0,032	1	3,48E-03	7,12E-03
BIEC2-417594	g.68.125.495 A>G	A	0,183	0,003	0,399	0,503	3,42E-06	1,24E-05
BIEC2-417596	g.68.136.361 T>C	T	0,039	1	0,001	1	4,34E-06	1,27E-05
BIEC2-417604	g.68.262.293 A>G	A	0,188	0,005	0,396	0,501	6,45E-06	1,78E-05
BIEC2-417617	g.68.460.098T>C	T	0,023	1	0,071	0,396	4,74E-02	6,83E-02

As frequências dos alelos menores dos SNPs foram mais baixas ($MAF < 0,01$) na região entre 65,5 e 67 Mb na linhagem de corrida em relação à de trabalho, local em que está mapeado o gene *MSTN*. Foi observado nessa região, de aproximadamente 1,5 Mb, marcadores praticamente fixados na linhagem de corrida, mas não na de trabalho. O principal SNP associado (BIEC2-417187) apresentou-se em DL com vários outros marcadores ao longo da região de interesse, a qual inclui o ERE1 do gene *MSTN* (Figura 2). A frequência do alelo menor (ERE-) do polimorfismo ERE1 foi significativamente mais baixa na linhagem de corrida ($MAF < 0,01$) que na de trabalho ($MAF = 0,103$).

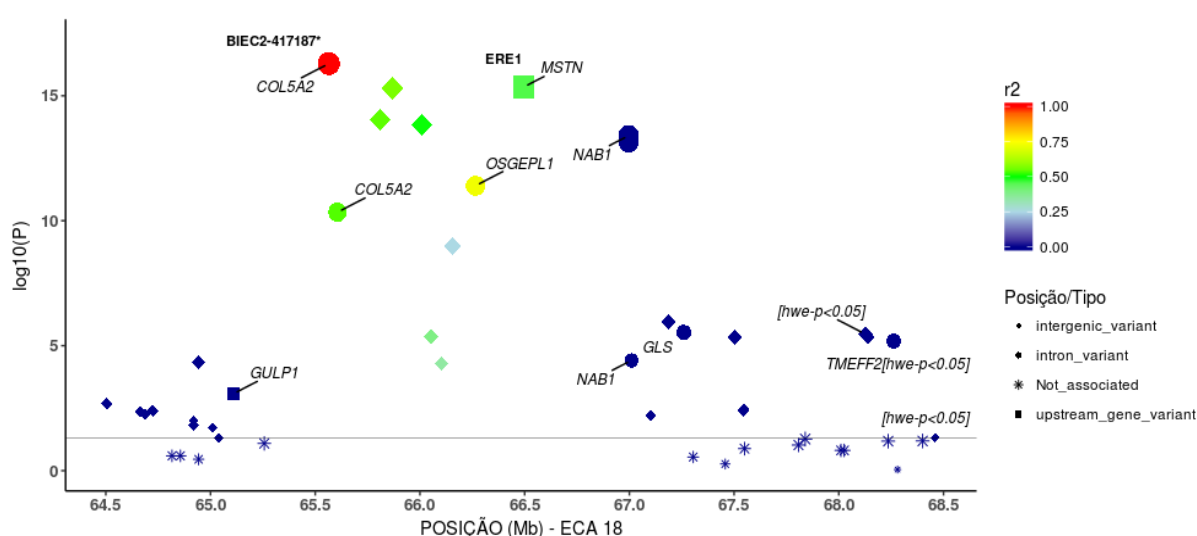


Figura 2 – Polimorfismos da região do gene *MSTN* no cromossomo 18 comuns a cavalos das linhagens de corrida e de trabalho da raça Quarto de Milha. Valores no eixo y indicam a significância da associação Caso-controle, dado em $-\log_{10}(P)$, entre polimorfismos e as linhagens de corrida e trabalho da raça. A linha cinza horizontal indica o limiar de $p = 0,05$. No eixo x estão indicadas posições dentro da região analisada. Diferentes representações de pontos significam diferentes tipos ou posição de variantes. Distintas tonalidades de cor representam o grau de desequilíbrio de ligação (r^2) entre o SNP associado com maior valor de p (BIEC2-417187) e as demais variantes. Os nomes dos genes em que as variantes se encontram estão escritos ao lado de cada polimorfismo associado. Marcadores fora do equilíbrio de HW foram indicados por $hwe-p < 0,05$.

Santagostino et al. (2015) analisaram a frequência do polimorfismo ERE1 em cinco diferentes raças equinas. A frequência do alelo ERE1+ foi particularmente mais

alta na raça Quarto de Milha. Na raça Andaluz, o mesmo alelo foi observado em apenas três indivíduos heterozigotos. Nas raças, nas quais não houve cruzamento com cavalos PSI, o alelo ERE+ não esteve presente. Estes resultados sugerem que a frequência mais alta do alelo maior deste polimorfismo em Quartos de Milha de corrida pode estar ligada à introdução na linhagem de cavalos PSI selecionados para provas de distâncias curtas, com o objetivo de incremento de velocidade. Os polimorfismos desta região apresentaram valores de $-\log_{10}(P)$ superiores aos demais SNPs associados. Isto possivelmente se deve, além da influência do PSI, a seleção indireta de alelos favoráveis ao maior desenvolvimento muscular na linhagem de corrida, principalmente no gene *MSTN*, que, devido ao alto DL na região, levaram à fixação de alelos em toda a adjacência. Dentre os genes descritos nessa região (em alto DL com BIEC2-417187) estão, além do *MSTN*, os genes Collagen Type V Alpha 2 Chain (*COL5A2*) e O-Sialoglycoprotein Endopeptidase Like 1 (*OSGEPL1*).

O *COL5A2* é parte de um grupo de genes que codificam pequenos colágenos fibrilares (VALKKILA et al., 2001). O papel destes colágenos tem importância na montagem, organização e manutenção da estrutura corporal (TSIPOURAS, 1988). Segundo Azuaje et. al. (2013), devido à associação do gene com a cardiopatia isquêmica em humanos, o *COL5A2* pode ser considerado um potencial biomarcador terapêutico.

Na posição aproximada de 67 Mb, dois SNPs no intron 1 do gene *NGFI-A Binding Protein 1 (NAB1)* foram associados com altos valores de $-\log_{10}(P)$. Contudo, não se encontram em desequilíbrio de ligação com o SNP BIEC2-417187. As frequências dos alelos menores de polimorfismos nesta região também foram, em geral, maiores na linhagem de trabalho se comparadas às de corrida. O *NAB1* é altamente expresso no músculo cardíaco e é relatado como regulador do crescimento do órgão (BUIRAGO et al., 2005). Assim, pode ter importância considerável principalmente para o metabolismo aeróbico, mais determinante nas atividades desempenhadas em cavalos da linhagem de trabalho da raça Quarto de Milha. Hill et al. (2010a) investigaram marcadores para distâncias ótimas de corrida em cavalos PSI em região de 1,7 Mb (Chr18: 65809482-67545806) no cromossomo 18. Foram identificados sete SNPs com significância estatística. Nessa região encontram-se mapeados 11 genes codificadores de proteína, incluindo o *MSTN* e o *NGFI-A Binding Protein 1 (NAB1)*, os quais foram considerados genes candidatos funcionais a afetar a característica.

Por fim, na porção inicial da região estuda (64,5 a 65,5 Mb) foram observadas associações significativas, entre as quais a de um SNP mapeado a upstream do gene *GULP*, Engulfment Adaptor PTB Domain Containing 1 (*GULP1*). Nesta região as frequências dos alelos menores de polimorfismos foram maiores na linhagem de corrida em relação à de trabalho. O produto proteico expresso pelo gene *GULP1* é um regulador positivo de TGF- β (Transforming Growth Fator), uma molécula que regula o crescimento celular. Além disso, apresenta outras importantes funções em processos biológicos como a proliferação e a diferenciação celular e a condução na transição epitelio-mesenquima (CHENG et al., 2012).

3.3. Associação entre as variantes genômicas e as características de interesse

Dentre as 97 variantes utilizadas nas análises de associação na linhagem de corrida, apenas o SNP g.67702857A>T, identificado por meio de sequenciamento, foi associado ($p < 0,05$) ao IV max (Tabela 2) e ao PT (Tabela 3). A substituição do alelo A por T apresentou efeito negativo, aproximadamente - 0,32, sobre os valores de EBV de IV max e PT. Apesar das associações encontradas, observa-se que esta variante encontrou-se em HWE nos animais de corrida.

Tabela 2 - Frequências genotípicas do polimorfismo associado à característica índice de velocidade máximo (IV max) na linhagem de corrida da raça Quarto de Milha, médias dos EBVs para os diferentes genótipos, efeito de substituição alélica e variância explicada.

SNP/Posição	Valor	G11	G12	G22	Efeito de substituição alélica (ep ¹)	Var. explicada (%)	p	p ajustado	p (HWE)
g.67702857A>T	Genótipo	AA	AT	TT	-0,317 (0,13)	0,096	0,039	0,971	0,599
67.702.857	Freq.	0,18	0,474	0,345					
	Média (dp*)	95,77 (1,54)	95,5 (1,55)	96,3 (1,74)					

*desvio padrão; ¹erro padrão.

Tabela 3 - Frequências genótípicas do polimorfismo associado à característica perímetro torácico (PT) na linhagem de corrida da raça Quarto de Milha, médias dos EBVs (cm) para os diferentes genótipos, efeito de substituição alélica e variância explicada.

SNP/Posição	Valor	G11	G12	G22	Efeito de substituição alélica (ep ¹)	Var. explicada (%)	<i>p</i>	<i>p</i> ajustado	<i>p</i> (HWE)
g.67702857A>T	Genótipo	AA	AT	TT	-0,320 (0,16)	0,097	0,03	0,965	0,585
67.702.857	Freq.	0,18	0,474	0,345					
	Média (dp*)	192,3 (1,53)	192 (1,55)	192,8 (1,73)					

*desvio padrão; ¹erro padrão.

Ainda na linhagem de corrida, oito polimorfismos foram associados à AC ($p < 0,05$) (Tabela 4). Destes, seis SNPs (BIEC2-438595, BIEC2-417135, BIEC2-417139, BIEC2-417145, BIEC2-417120 e BIEC2-417124) foram genotipados por meio de chip e dois (g.68076310C>G e g.65593080T>A) por meio de sequenciamento. Todos estes polimorfismos encontraram-se em equilíbrio de HW. Para a característica CC (comprimento corporal) não foram identificados polimorfismos associados.

Tabela 4 - Frequências genotípicas dos polimorfismos associados à característica altura de cernelha (AC) na linhagem de corrida da raça Quarto de Milha, médias dos EBVs (cm) para os diferentes genótipos, efeito de substituição alélica e variância explicada.

SNP/Posição	Valor	G11	G12	G22	Efeito de substituição alélica (ep ¹)	Var. explicada (%)	<i>p</i>	<i>p</i> ajustado	<i>p</i> (HWE)
BIEC2-438595	Genótipo	AA	AG	GG	-9,229(1,46)	0,097	1,20E-06	1,13E-04	1
64.611.659	Freq.	0	0,004	0,996					
	Média (dp*)	NA	149,1 (0)	158,3 (1,46)					
BIEC2-417135	Genótipo	AA	AG	GG	0,528(0,19)	0,459	0,005	0,223	1
65.009.739	Freq.	0,024	0,313	0,663					
	Média (dp*)	160,2 (2,38)	158,5 (1,40)	158,1 (1,57)					
BIEC2-417139	Genótipo	AA	AG	GG	0,516(0,2)	0,687	0,011	0,330	0,826
65.038.492	Freq.	0,016	0,265	0,719					
	Média (dp*)	160,3 (3,25)	158,5 (1,46)	158,2 (1,53)					
BIEC2-417145	Genótipo	AA	AG	GG	0,432(0,18)	0,432	0,015	0,361	0,607
65.107.716	Freq.	0,0361	0,309	0,657					
	Média (dp*)	159,4 (2,22)	158,5 (1,45)	158,1 (1,56)					
BIEC2-417120	Genótipo	TT	TC	CC	0,384(0,17)	0,431	0,028	0,454	1
64.919.213	Freq.	0,040	0,333	0,627					
	Média (dp*)	159,1 (2,23)	158,4 (1,41)	158,1 (1,59)					
g.68076310C>G	Genótipo	CC	CG	GG	0,303(0,14)	0,322	0,029	0,454	1
68.076.310	Freq.	0,217	0,482	0,301					
	Média (dp*)	158,6 (1,50)	158,3 (1,65)	158 (1,45)					
g.65593080T>A	Genótipo	TT	TA	AA	0,357(0,17)	0,402	0,036	0,477	0,877
65.593.080	Freq.	0,0482	0,333	0,619					
	Média (dp*)	159,2 (2)	158,4 (1,61)	158,2 (1,49)					
BIEC2-417124	Genótipo	AA	AG	GG	0,325(0,16)	0,375	0,042	0,489	1
64.943.005	Freq.	0,068	0,382	0,550					
	Média (dp*)	158,7 (1,92)	158,5 (1,35)	158,1 (1,65)					

Na linhagem de trabalho, dois polimorfismos (BIEC2-417306 e BIEC2-417479) foram associados à PT ($p < 0,05$) (Tabela 5). Ambos tiveram na substituição do alelo T por C um efeito positivo (1,65 e 1,84, respectivamente). As duas variantes se encontraram em equilíbrio de HW.

Tabela 5 - Frequências genótípicas dos polimorfismos associados à característica perímetro torácico (PT) na linhagem de trabalho da raça Quarto de Milha, médias dos EBVs (cm) para os diferentes genótipos, efeito de substituição alélica e variância explicada.

SNP/Posição	Valor	G11	G12	G22	Efeito de substituição alélica (ep ¹)	Var. explicada (%)	p	p _{ajust.}	p _(HWE)
BIEC2-417306	Genótipo	TT	TC	CC	1,657(0,58)	0,050	0,006	0,294	1
66.102.331	Freq.	0	0,04918	0,9508					
	Média(dp*)	NA	181,5(1,47)	179,8(0,95)					
BIEC2-417479	Genótipo	TT	TC	CC	1,847(0,71)	0,080	0,006	0,294	1
67.099.904	Freq.	0	0,03226	0,9677					
	Média(dp*)	NA	181,7(2)	179,9(0,95)					

*desvio padrão; ¹erro padrão.

Em relação à CC, dois SNPs, BIEC2-417608 e BIEC2-417603, foram associados à característica (Tabela 6). O BIEC2-417608 mapeado no intron 5 do gene TMEFF2, não se encontra em equilíbrio de HW ($p < 0,05$).

Tabela 6 - Frequências genótípicas dos polimorfismos associados à característica comprimento corporal (CC) na linhagem de trabalho da raça Quarto de Milha, médias dos EBVs (cm) para os diferentes genótipos, efeito de substituição alélica e variância explicada.

SNP/Posição	Valor	G11	G12	G22	Efeito de substituição alélica (ep ¹)	Var. explicada (%)	p	p _{ajust.}	p _(HWE)
BIEC2-417608	Genótipo	TT	TC	CC	0,680(0,26)	0,046	0,012	0,314	0,046
68.280.433	Freq.	0,049	0,131	0,820					
	Média(dp*)	164,7(0,67)	163,6(0,86)	163,1(1,13)					
BIEC2-417603	Genótipo	AA	AG	GG	0,500(0,23)	0,084	0,037	0,314	0,206
68.235.850	Freq.	0,066	0,230	0,705					
	Média(dp*)	164,1(1,31)	163,6(1,32)	163,1(1,01)					

*desvio padrão; ¹erro padrão.

Após a aplicação do critério de MAF ($> 0,05$) e do ajuste FDR para testes múltiplos, as associações observadas entre genótipos de SNPs e os EBVs estimados das características de interesse tornaram-se questionáveis tanto na linhagem de

corrida (IV max, CC e PT) quanto na de trabalho (PT e CC). Contudo, a despeito da significância estatística, maior atenção foi dispendida à altura de cernelha na linhagem de corrida devido ao grande número de marcadores associados (p não ajustado $< 0,05$) com considerável DL entre si e com frequências do alelo favorável significativamente maiores em relação à linhagem de trabalho (Tabela 1). Com a desconsideração do SNP BIEC2-438595 devido à baixa MAF ($< 0,01$), o SNP intergênico BIEC2-417135 foi encontrado associado à AC com o maior valor de $-\log_{10}(P)$ (Figura 3). Todos os SNPs adjacentes associados significativamente ($p < 0,05$) encontram-se em DL com este marcador em região compreendida entre 65 e 65,5 Mb. Dentre os polimorfismos em DL, o BIEC2-417145, o qual apresentou o maior DL com o BIEC2-417135, está localizado a upstream do gene *GULP1*. O SNP g.65593080A>T está localizado no intron 12 do gene *COL5A2*. Ademais, o SNP g.68076310C>G, o qual não está em DL com o principal marcador associado, corresponde a uma variante missense (aGt/aCt) do gene *CAVIN2*.

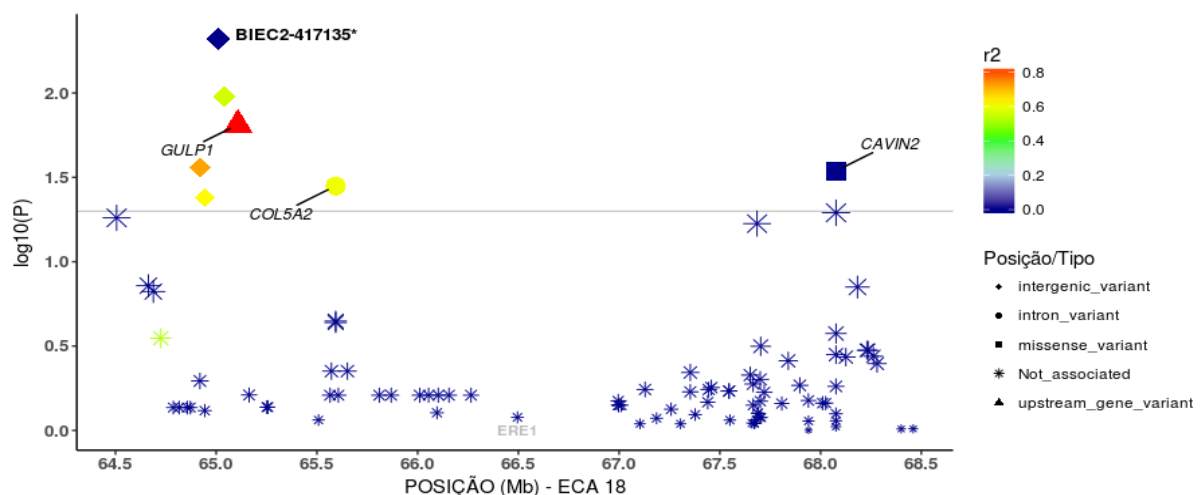


Figura 3 – Polimorfismos da região do gene *MSTN* no cromossomo 18 em cavalos das linhagens de corrida da raça Quarto de Milha. Valores no eixo y indicam a significância da associação, dado em $-\log_{10}(P)$, entre polimorfismos e altura de cernelha (AC). A linha cinza horizontal indica o limiar de $p = 0,05$. No eixo x estão indicadas posições dentro da região analisada. Diferentes representações de pontos significam diferentes tipos ou posição de variantes. Distintas tonalidades de cor representam o grau de desequilíbrio de ligação (r^2) entre o SNP associado com maior valor de p (BIEC2-417135) e as demais variantes. Os nomes dos genes em que as variantes se encontram estão descritos ao lado de cada polimorfismo associado.

4. Conclusão

Na região genômica estudada, gene *MSTN* ± 2 Mb, foram identificados 46 SNPs e 1 SINE – ERE1 comuns às linhagens de trabalho e de corrida da raça Quarto de Milha, dos quais 32 SNPs e 1 SINE apresentam frequências alélicas significativamente diferentes entre as mesmas. Foi constatado que a região mais próxima ao *MSTN* (± 1 Mb) é menos polimórfica na linhagem de corrida em relação à de trabalho. Isso pode ser explicado por uma provável maior pressão de seleção sobre essa região na linhagem de corrida em relação à de trabalho e/ou devido ao uso de animais PSI na linhagem de corrida.

Em relação aos polimorfismos associados à fenótipos (p não ajustado $< 0,05$), sobressaíram os relacionados à altura de cernelha na linhagem de corrida devido ao seu grande número e ao alto desequilíbrio de ligação. Além disso, vários destes marcadores apresentaram frequências do alelo favorável maiores dos que as encontradas na linhagem de trabalho. Tendo se em vista a possibilidade de associação genética entre características morfológicas e desempenho em equinos, esse resultado é de grande importância, fornecendo informações para futuros estudos de melhoramento genético animal.

5. Bibliografia citada

ABQM. **Associação brasileira dos criadores de cavalos Quarto de Milha**. Disponível em: <http://www.abqm.com.br/item1.htm>. Acesso em: 15 de fevereiro de 2016.

ABQM. **Associação brasileira dos criadores de cavalos Quarto de Milha**. Disponível em: <http://www.abqm.com.br/item1.htm>. Acesso em: 23 de janeiro de 2017.

AZUAJE, F.; ZHANG, L.; JEANTY, C; PUHL, S.L.; RODIUS, S.; WAGNER, D.R. Analysis of a gene co-expression network establishes robust association between Col5a2 and ischemic heart disease. **BMC Medical Genomics**. P. 6:13, 2013.

BARRETT, J.C.; FRY, B.; MALLER, J.; DALY, M.J. Haploview: analysis and visualization of LD and haplotype maps. **Bioinformatics**. V. 21, p. 263:265, 2005

BENJAMINI, Y.; HOCHBERG, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple. **Journal of the Royal Statistical Society Testing**. V. 57, p. 289:300, 1995

BINNS, M. M.; BOEHLER, D. A.; LAMBERT, D. H. Identification of the myostatin locus (MSTN) as having a major effect on optimum racing distance in the Thoroughbred horse in the USA. **Animal Genetics**. V. 41 (Suppl. 2), p. 154:158, 2010.

BRAY, M.S.; HAGBERG, J.M.; PÉRUSSE, L.; RANKINEN, T.; ROTH, S.M.; WOLFARTH, B.; BOUCHARD, C. The human gene map for performance and health-related fitness phenotypes: the 2006-2007 update. **Medicine & Science in Sports & Exercise**. V. 73, p. 41:35, 2009.

BITRAGO, M.; LORENZ, K.; MAASS, A.H.; OBERDORF-MAASS, S.; KELLER, U.; SCHMITTECKERT, E.M.; IVASHCHENKO, Y. LOHSE, M.J.; ENGELHARDT, S. The transcriptional repressor Nab1 is a specific regulator of pathological cardiac hypertrophy. **Nature Medicine**. V.11, n.8, p. 837:844, 2005.

CHENG, J.M.; MARTIN, C.; MA, Z.; HAFIANE, A.; DAI, M.; LEBRUN, J.J.; KISS, R.S. Engulfment protein GULP is regulator of transforming growth factor- β response in ovarian cells. **The journal of biological chemistry**. V. 287, p. 2036:20651, 2012.

CHIDLOW JR, J.H.; SESSA, W.C. Caveolae, caveolins, and cavins: complex control of cellular signalling and inflammation. **Cardiovascular Research**. V.86, p.219:225, 2010.

CLOP, A.; MARCQ, F.; TAKEDA, H. et al. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. **Nature Genetics**. V.38, p.813:818, 2006.

EVANS, J. W. **Horses: a guide to selection, care and enjoyment**. 4. ed. Freeman and Company: New York, 1996, p. 797.

FREEMAN, D.W. **Physical conditioning of horses**. Oklahoma State University: Division of Agricultural Sciences and Natural Resources, 2013.

GIANOLA, D.; SIMIANER, H.; QANBARI, S. A two-step method for detecting selection signatures using genetic markers. **Genetics Research Cambridge**. V. 92, p. 141:155. 2010.

GROBET, L.; ROYO MARTIN, L.J.; PONCELET, D.; PIROTTIN, D.; BROUWERS, B.; RIQUET, J.; SCHOEBERLEIN, A.; DUNNER, S.; MÉNISSIER, F.; MASSABANDA, J.; FRIES, R.; HANSET, R.; GEORGES, M. A deletion in the myostatin gene causes double muscling in cattle. **Nature Genetics**. V.17, p.71:74, 1997.

HILL, E. W.; MCGIVNEY, B. A.; GU, J.; WHISTON, R.; MACHUGH, D. E. A genome-wide SNP-association study confirms a sequence variant (g.66493737C>T) in the equine myostatin (MSTN) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses. **BMC genomics**. V. 11, n. 552, 2010a.

HILL, E. W.; MCGIVNEY, B. A.; GU J.; WHISTON, R.; MACHUGH D. E. A genome-wide SNP-association study confirms a sequence variant (g.66493737C>T) in the equine myostatin (MSTN) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses. **BMC Genomics**. V.11, n.552, p.1:10, 2010b.

HILL, W.G.; ROBERTSON, A. Linkage Disequilibrium in Finite Populations. **Theoretical and Applied Genetics**. V. 38, p. 226:231, 1968.

JÖNSSON, L.; NÄSHOLM, A.; ROEPSTORFF, L.; EGENVALL, A.; DALIN, G.; PHILIPSSON, J. Conformation traits and their genetic and phenotypic associations with health in young Sweddish warmblood riding horses, **Livestock Science**, 163, 12-25, 2014.

KAMBADUR, R.; SHARMA, M.; SMITH, T.P.L.; BASS, J.J. Mutations in myostatin (GDF-8) in double muscled Belgian Blue and Piedmontese cattle. **Genome Research**. V. 7, p.910-916, 1997.

KAPRÓN, M.; CZERNIAK, E.; ŁUKASZEWICZ, M.; DANIELEWICZ, A. Genetic parameters of body conformation and performance traits of Wielkopolski horses registered in the successive volumes of the herdbook. *Archives Animal Breeding*. v.56, n.12, p.127-136, 2013.

MCPHERRON A C, LAWLER A M, LEE S J. Regulation of skeletal muscle mass in mice by a new TGF-beta superfamily member. **Nature**. V. 387, p. 83:90, 1997.

MEIRA, C.T.; FARAH, M.M.; FORTES, M.R.S.; MOORE, S.S.; PEREIRA, G.L.; SILVA, J.A.II V.; MOTA, M.D.S.; CURI, R.A. A Genome-wide association study for morphometric traits in Quarter Horse. **Journal of Equine Veterinary Science**. V. 34 ed. 8, p. 1028:1031, 2014.

MISZTAL, I.; TSURUTA, S.; STRABEL, T.; AUVRAY, B.; DRUET, T.; LEE, D. H. 7th World Congress on Genetics Applied to Livestock Production, Agosto 19-23, 2002, Montpellier, França. **BLUPF90 and related programs (BGF90)**, 2003, p. 28.

MISZTAL, I. 2007. Disponível em: <
<http://nce.ads.uga.edu/~ignacy/newprograms.html>>; acesso em: 21/02/2014.

MISERANI, M.G.; MCMANUS, C.; SANTOS, S.A.; SILVA, J.A.; MARIANTE, A.S.; ABREU, U.G.P.; MAZZA, M.C.; SERENO J.R.B. Heritability estimates for biometric measures of the pantaneiro horse. **Arquivos Brasileiro de Zootecia**. V.51, p.107-112, 2002.

NISHI, M.; YASUE, A.; NISHIMATU, S.; NOHNO, T.; YAMAOKA, T.; ITAKURA, M.; MORIYAMA, K.; OHUCHI, H.; NOJI, S. A missense mutant myostatin causes hyperplasia without hypertrophy in the mouse muscle. **Biochemical and Biophysical Research Communications**. V. 293, p.247–251, 2002.

PEREIRA, G. L.; MEIRA, C. T.; SILVA, J.A.II V.; CHARDULO, L.A.L.; CURI, R.A. Estimativas de parâmetros genéticos para características morfométricas em cavalos quarto de milha de corrida. X CURSO DE INVERNO DE GENÉTICA, 2014, FCAV/UNESP, 2014, p, 44-48.

PEREIRA, G.L.; MATTEIS, R.; REGITANO, L.C.A.; CHARDULO, L.A.L.; CURI, R.A. MSTN, CKM and DMRT3 gene variants in different lines of Quarter Horses. **Journal of Equine Veterinary Science**. V. 39, p. 33:37, 2016.

PURCELL, S.; NEALE, B.; BROWN, K.T.; THOMAS, L.; FERREIRA, M. A. R.; BENDER, D.; MALLER, J.; SKALAR, P.; BAKKER, P.I.W.; DALY, M.J.; SHAM, P.C. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. **The American Journal of Human Genetics**. V. 81, p. 559:575, 2007.

SANTAGOSTINO M., KHORIAULI L., GAMBA R., BONUGLIA M., KLIPSTEIN O., PIRAS F.M., VELLA F., RUSSO A., BADIALE C., MAZZAGATTI A., RAIMONDI E., NERGADZE, S.G., GIULOTTO E. Genome-wide evolutionary and functional analysis of the Equine Repetitive Element 1: an insertion in the myostatin promoter affects gene expression. **BMC Genet**. V. 26, p. 16:126, 2015.

SARGOLZAEI, M.; CHESNAIS, J. P.; SCHENKEL, F. S. A new approach for efficient genotype imputation using information from relatives. **BMC Genomics**. V. 15, p. 478, 2014.

SCHUELKE, M.; ANGER, K.R.; STOLZ, L.E.; HUBNER, C.; RIEBEL, T. et al. Myostatin mutation associated with gross muscle hypertrophy in a child. **The New England Journal of Medicine**. V. 350, p. 2682:2688, 2004.

TORRES, A. D. P.; JARDIM, W. R. **Criação do cavalo e de outros equinos**. São Paulo: Editora Nobel, 3ª ed., p. 654, 1992.

TOZAKI, T.; MIYAKE, T.; KAKOI, H.; GAWAHARA, H.; SUGITA, S.; HASEGAWA, T. et al A genome-wide association study for racing performances in Thoroughbreds clarifies a candidate region near the MSTN gene. **Animal Genetics**. V. 41, p.28-35, 2010.

TSIPOURAS, P.; SCHWARTZ R.C.; LIDDELL A.C.; SALKELD, C.S.; WEIL, D.; RAMIREZ, F. Genetic Distance of Two Fibrillar Collagen Loci, COL3A1 and COL5A2, Located on the Long Arm of Human Chromosome 2. **Genomics**. V.3, p275:278, 1988.

VALKKILA, M.; MELKONIEMI, M.; KVIST, L.; KUIVANIEMI, H.; TROMP, G.; ALA-KOKKO, L. Genomic organization of the human COL3A1 and COL5A2 genes: COL5A2 has evolved differently than the other minor fibrillar collagen genes. **Matrix Biology**. V.20, p.357:366, 2001.

WARNES G. Population genetics. Package "genetics" v.1.3.8.1. 2013.