

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

**ASSINATURAS DE SELAÇÃO NA LINHAGEM DE TRABALHO
DE EQUINOS DA RAÇA QUARTO DE MILHA**

**Natalia Andrea Rincón Beltrán
Zootecnista**

JABOTICABAL – SÃO PAULO – BRASIL

2014

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

**ASSINATURAS DE SELAÇÃO NA LINHAGEM DE TRABALHO
DE EQUINOS DA RAÇA QUARTO DE MILHA**

Natalia Andrea Rincón Beltrán

Orientador: Luis Artur Loyola Chardulo

Co-orientador: Prof. Dr. Rogério Abdallah Curi

Tese apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Câmpus de Jaboticabal, como parte das exigências para a obtenção do título de Doutor em Genética e Melhoramento Animal.

JABOTICABAL – SÃO PAULO – BRASIL

2014

DADOS CURRICULARES DO AUTOR

NATALIA ANDREA RINCÓN BELTRÁN – nascida em 8 de julho de 1983, na cidade de Bogotá - Cundinamarca, Colômbia, é formada em Zootecnia pela UDEC - Universidad De Cundinamarca, em janeiro de 2006 e, Mestre em Produção Animal pela Universidade Federal dos Vales de Jequitinhonha e Mucuri, sob orientação da Profa. Maria Margarida Figueiredo de Oliverira. Em agosto de 2011, iniciou o doutorado em Genética e Melhoramento Animal, na Faculdade de Ciências Agrárias e Veterinárias, UNESP, Campus de Jaboticabal, sob a orientação do Prof. Dr. Marcilio Dias Silveira da MOTA e a Co-orientação: Prof. Dr. Rogério Abdallah Curi. Devido ao falecimento do Prof. Dr. Marcilio Dias Silveira da Mota foi concluído o doutorado sob a orientação do Prof. Dr. Luis Artur Loyola Chardulo

Dedicatória

*Ào criador de todas as coisas,
ao que tem dado uma nova oportunidade de viver
com fortaleza para continuar e não decair,
guiando-me e ajudando-me a superar
obstáculos e dificuldades ao longo da minha vida
por isso, dedico meu trabalho a Deus*

*Al creador de todas las cosas, el que me ha dado
una nueva oportunidad de vivir
con fortaleza, para continuar y no decaer guiándome y
ayudándome a superar obstáculos y dificultades
a lo largo de toda mi vida por ello,
dedico mi trabajo a Dios.*

AGRADECIMENTOS

À minha mãe Ines Beltrán por que graças a ela sei que a responsabilidade se deve viver como um compromisso de dedicação e esforço. A meu pai Alberto Rincón, que seu viver demonstrou-me que no caminho para a meta se precisa da doce fortaleza para aceitar as derrotas e da sutil coragem para derrotar os medos. A eles por que, apesar da distância física, sinto que estão comigo sempre, sei que neste momento estarmos juntos seria tão especial para vocês como para mim.

A esta pessoa não tenho como agradecer de ter-me visto quando era invisível, meu ponto de vida, meu esposo e companheiro politicamente evolucionado, Luis Ernesto Barnabé, muitíssimo obrigada por acompanhar-me neste processo, sobretudo, teu amor, tua compreensão, paciência e fortaleza que permitiram que pudesse, não só trabalhar, mas também chegar a bom porto. Como em tudo o que escrevo estás sempre na minha mente e na minha alma, nestas líneas. Contigo aprendo constantemente. Amo viver e ser contigo. Amo saber que tua companhia se estenderá muito além deste período, chegando inclusive ao que temos imaginado, um grande Maracatu.

Quero agradecer a meus sogros Rita e Lineu, por ser totalmente contrário que dessa expressão representa no sentido comum. Obrigada pelo seu carinho, compreensão, apoio e paciência. Por querer-me e me cuidar como uma filha. A minha cunhada Ana seus filhos e esposo por suas conversações e companhia. Por cada dia perguntar como ia minha tese.

Aos meus familiares de coração e vida e àqueles que recém se somaram a ela para fazer-me companhia com seus sorrisos de ânimo, em especial a Julia, Camila, Carlão, Edna, Graciene e Michel por que ao longo deste trabalho aprendemos que nossas diferenças se convertem em riquezas quando existe respeito e verdadeira amizade.

Para meus amigos na distância Yohan, Fausto Alfonso, Nancy, Edgar, Joo, Adriano, Tseng, Pablo, Luciana y Gabriela pelo seu apoio, conselhos, compreensão, amor, ajuda

nos momentos difíceis. Têm-me aceitado como sou mesmo com a distância, e apesar dos anos segue firme e forte nossa amizade.

Quero dar um agradecimento especial a meu primeiro orientador Marcilio Dias, que me ensinou a ver-me primeiro antes que os demais. Por não aceitar minhas cartas de desistência do doutorado, oferecendo uma amizade. Nunca esquecerei o senhor falando: “Calma Natalia as coisas não são assim”, depois dava grandes explicações sorrindo e sempre renovava minhas esperanças. Obrigada por me permitir a contar com o senhor quando mais precisei. Ainda dói a sua rápida partida.

De igual maneira agradeço a meu co-orientador o doutor Rogério Abdallah Curi pela sua visão crítica de muitos aspectos cotidianos da vida, pela sua retidão enquanto docente, pelos seus conselhos, que ajudaram a formar-me como pessoa e pesquisadora.

São muitas as pessoas que têm formado parte da minha vida profissional, as quais gostaria agradecer pela sua amizade, conselhos, apoio, ânimo e companhia nos momentos mais difíceis da minha vida. Algumas estão aqui comigo e outras em minhas lembranças e em meu coração. Não importa onde estejam, quero agradecer por formar parte de mim, por tudo o que me brindaram e por todas as suas bênçãos.

SUMÁRIO

1. INTRODUÇÃO.....	1
2. HIPÓTESE.....	3
3. OBJETIVOS.....	3
4. REVISÃO DE LITERATURA.....	4
4.1. Os equinos no Brasil e no mundo.....	4
4.2. Origem e situação atual da raça Quarto de Milha no Brasil	5
4.3. Melhoramento genético de equinos	8
4.4. Marcadores moleculares.....	9
4.5. Marcadores SNPs.....	12
4.6. Métodos de genotipagem de SNPs.....	14
4.7. Chips de genotipagem de SNPs de alta densidade.....	15
4.8. Análise de SNPs em larga escala.....	17
5. MATERIAIS E METODOS.....	24
5.1. Animais.....	24
5.2. Genotipagem dos SNPs.....	25
5.3. Controle de qualidade dos dados de genotipagem.....	26
5.4. Análise dentro da linhagem de trabalho.....	27
5.5. Análise entre as linhagens de corrida e trabalho.....	29
5.6. Anotação dos genes.....	31
6. RESULTADOS E DISCUSSÃO.....	31
6.1. Controle de qualidade dos dados de genotipagem.....	31
6.2. Análise dentro da linhagem de trabalho.....	32
6.3. Análise entre as linhagens de corrida e trabalho.....	37
6.4. Anotação de genes em regiões genômicas selecionadas na linhagem de trabalho de forma divergente em relação à de corrida.....	39
7. CONCLUSÕES.....	48
REFERÊNCIAS.....	49

LISTA DE FIGURAS

Figura 1 Mapa das prováveis assinaturas de seleção no genoma de equinos da linhagem de trabalho da raça Quarto de Milha. Somente os haplótipos com frequência maior que 25% foram considerados. Linhas tracejadas indicam $P < 0,01$ 37

LISTA DE TABELAS

Tabela 1. Estatística descritiva dos polimorfismos de nucleotídeo único (SNPs); informativos e das regiões núcleo (CR) em cavalos da linhagem de trabalho da raça Quarto de Milha.....	34
Tabela 2. Estatística da aplicação do teste do haplótipo de longo alcance (LRH) no genoma de equinos da linhagem de trabalho da raça Quarto de Milha.....	36
Tabela 3. Número de SNPs identificados pela análise F_{ST} com probabilidades de 50 a 95% de pertencer ao <i>cluster 4</i>	39
Tabela 4. Regiões do genoma de animais de trabalho da raça Quarto de Milha identificadas como de coincidência pelas análises REHH e F_{ST} , total de genes identificados nessas regiões, genes com função conhecida e genes candidatos possivelmente relacionados a desempenho na linhagem.....	43
Tabela 5. Genes relacionados com processos biológicos dos sistemas muscular, esquelético, cardiovascular, respiratório, nervoso, atividade motora, visão, audição, função cognitiva, e metabolismo energético muscular, localizados nos cromossomos 1, 4 e 7 próximos as regiões de assinaturas de seleção do genoma de equinos da linhagem de trabalho da raça Quarto de Milha, e o número de SNPs identificado nesses genes.....	46
Tabela 6. Genes relacionados com processos biológicos dos sistemas muscular, esquelético, cardiovascular, respiratório, nervoso, atividade motora, visão, audição, função cognitiva, e metabolismo energético muscular, localizados nos cromossomos 10, 11, 13, 14, 16, 17, 22, 24, 25, e 31 próximos às regiões de assinaturas de seleção do genoma de equinos da linhagem de trabalho da raça Quarto de Milha, e o número de SNPs identificado nesses genes.....	47

LISTA DE ABREVIATURAS

ABQM	Associação Brasileira de Criadores de Cavalos Quarto de Milha
AQHA	<i>American Quarter Horse Association</i>
cM	centrimorgan
Compr	Cumprimento
CR	Regiões Núcleo
c_t	número de amostras de um haplótipo núcleo particular t
DHPLC	Cromatografia Líquida de Alta Pressão Desnaturante
ECA 5	Cromossomo autossômico equino 5
EHH	Homozigosidade do Haplótipo Estendido
EM	Maximização de Expectativa
<i>EquCab 2.0</i>	Genoma equino
e_{ij}	número de amostras de um haplótipo estendido particular i
F_{ST}	índice de fixação
Kb	Mil pares de bases
LD	<i>linkage disequilibrium</i> - Desequilíbrio de Ligação
LRH	<i>Long Range Haplotype</i> - Haplótipo de Longo Alcance
MAF	<i>Minor Allele Frequency</i>
Mbp	Mega bases
N (Dado)	Conjunto de alelos observado na amostra
n°	Número
n_{ilk}	número de alelos observados no <i>locus</i> bialélico
p	vetor das frequências alélicas
QTL	<i>Quantitative Trait Loci</i>
REHH	Homozigose Relativa do Haplótipo Estendido
s	número de haplótipos estendidos únicos
SNPs	<i>Single Nucleotide Polymorphisms</i> - Polimorfismos de nucleotídeo único
t	dado
θ	teta

ASSINATURAS DE SELAÇÃO NA LINHAGEM DE TRABALHO DE EQUINOS DA RAÇA QUARTO DE MILHA

RESUMO:

O objetivo desta pesquisa foi identificar, por meio da utilização e análise de painéis de genotipagem de SNPs de alta densidade e da aplicação das estatísticas homozigose relativa do haplótipo estendido (REHH), uma extensão da análise EHH, e índice de fixação (F_{ST}), regiões do genoma selecionadas na linhagem de trabalho da raça Quarto de Milha de forma divergente em relação à de corrida. Foram utilizados 188 cavalos de ambos os sexos, nascidos entre 1985 e 2009, registrados na Associação Brasileira de Criadores de Cavalo Quarto de Milha (ABQM), sendo 68 da linhagem de trabalho e 120 da de corrida. A partir de 36 regiões do genoma, consideradas assinaturas de seleção por ambas as estatísticas utilizadas, foi feita a anotação funcional de genes a fim de identificar aqueles que possam ter sido importantes ao longo do processo de formação da linhagem de trabalho. Quarenta e cinco genes foram identificados em processos biológicos relacionados aos sistemas muscular, esquelético, cardiovascular, respiratório e nervoso, neurotransmissão, metabolismo energético muscular, atividade motora, visão, audição, e função cognitiva. Os genes relacionados aos últimos quatro processos, merecem destaque pelo fato de que, em confluência, podem estar relacionados ao *cow sense* ou habilidade de apartação, característica de grande importância para a utilização do equino no manejo de bovinos de corte.

Palavras chave: arranjo de SNPs, assinaturas de seleção, equinos, homozigose relativa do haplótipo estendido, índice de fixação

SIGNATURE OF SELECTION IN CUTTING LINE OF QUARTER HORSES

ABSTRACT:

The objective of this study was to identify genomic regions selected in cutting line of Quarter Horses divergently in relation to racing line using high-density SNP genotyping arrays and the relative extended haplotype homozygosity (REHH) test, an extension of EHH analysis, and the fixation index (F_{ST}) as statistical methods. A total of 188 horses of both sexes, born between 1985 and 2009 and registered with the Brazilian Association of Quarter Horse Breeders (ABQM), were used. Of these, 68 horses were from the cutting line and 120 from the racing line. On the basis of 36 genomic regions classified as selection signatures by the two statistical methods, functional annotation of genes was performed in order to identify those that might have been important during formation of the cutting line. Forty-five genes were found to be involved in biological processes related to the muscle, skeletal, cardiovascular, respiratory and nervous systems, neurotransmission, muscle energy metabolism, motor activity, vision, hearing, and cognitive function. The genes related to the last four processes are particularly interesting because these genes together may be involved in *cow sense* or cutting ability, an important characteristic for the management of beef cattle.

Keywords: SNP array, selection signature, horse, relative extended haplotype homozygosity, fixation index.

1. Introdução

O cavalo Quarto de Milha é considerado o mais versátil do mundo no que diz respeito às provas esportivas. Para que se tenha conhecimento de sua popularidade, no século XVII, nos EUA em um intervalo de vinte anos, a população Quarto de Milha aumentou de 200 para 10.000 indivíduos. Atualmente há aproximadamente cinco milhões de cavalos inscritos na *American Quarter Horse Association* (AQHA) em 75 países (ABQM, 2012). No Brasil existem aproximadamente 358.000 equinos da raça registrados junto à Associação Brasileira de Criadores de Quarto de Milha (ABQM), tendo impacto relevante no agronegócio nacional (ABQM, 2012).

Dentro da raça Quarto de Milha se encontram subdivisões em diferentes segmentos de aptidão, provenientes de distintos objetivos de seleção (EVANS, 1996), consideradas linhagens, entre as quais a de trabalho e a de corrida. A linhagem de trabalho destina-se às provas de caráter funcional, explorando habilidades como agilidade, obediência e o *cow sense*.

Animais com esta última qualidade apresentam habilidade inata de trabalhar com rebanho, característica considerada de grande importância no manejo do gado a campo. A linhagem de corrida explora a aptidão dos animais quanto à velocidade em pistas retas e de curta distância.

Avanços tecnológicos recentes na área de Genética Molecular culminaram com o desenvolvimento dos chamados *chips* de genotipagem de *Single Nucleotide Polymorphisms* (SNPs) de alta densidade, os quais permitem a análise de dezenas de milhares desses polimorfismos num único ensaio a baixos custos, trazendo novas perspectivas para aplicações já consolidadas, além do desenvolvimento de novas aplicações, produtos e processos, com uso direto nas cadeias produtivas de animais

de interesse zootécnico (CAETANO, 2009). Neste sentido, o estudo simultâneo de milhares de SNPs espalhados ao longo do genoma tem possibilitado: analisar a estrutura genética de diferentes populações em várias espécies de animais domésticos (MCKAY *et al.*, 2008; KIJAS *et al.*, 2009; GIBBS *et al.*, 2009), estimaram o grau de diversidade dentro e divergência genética entre populações (ZENGER *et al.*, 2007) e, recentemente, identificar e localizar regiões do genoma submetidas à seleção positiva (HAYES *et al.*, 2008; PRASAD *et al.*, 2008; BARENDSE *et al.*, 2009; MACEACHERN *et al.*, 2009).

A identificação de alterações provocadas pela seleção positiva ao longo do genoma pode ser realizada por meio da aplicação de algumas metodologias principais, entre as quais a estatística homozigosidade do haplótipo estendido - EHH (SABETI *et al.*, 2002), utilizada para a identificação dessas regiões dentro de populações, e o índice de fixação, F_{ST} , (WRIGHT, 1951; COCKERHAM, 1969; WEIR e HILL, 2002), utilizado para a identificação de regiões submetidas à seleção, considerando-se múltiplas populações (SIMIANER *et al.*, 2010).

O conceito básico da estatística EHH, introduzida por SABETI *et al.* (2002), se baseia no fato de que uma nova mutação submetida à pressão seletiva positiva irá aumentar rapidamente sua frequência, carregando consigo os alelos vizinhos em desequilíbrio de ligação (*linkage disequilibrium* - LD), o que é conhecido como varredura seletiva (MAYNARD SMITH & HAIGH, 1974). Como se a seleção pode carregar um alelo em um haplótipo específico a altas frequências mais rápido do que ele é quebrado por recombinação, haplótipos com alta frequência serão mais longos do que o esperado sob neutralidade (SIMIANER *et al.*, 2010). Nesse sentido

haplótipos longos com frequência e homozigose elevadas podem indicar regiões do genoma selecionadas recentemente.

A estatística F_{ST} mede a relação entre os pares de alelos de polimorfismos dentro de uma sub-população em relação à população como um todo (WRIGHT, 1951; COCKERHAM, 1969; WEIR & HILL, 2002), e pode ser interpretada como medida da dispersão de frequências alélicas entre grupos em relação à variação esperada na população a partir do qual derivaram esses grupos.

2. HIPÓTESE

A seleção artificial teria provocado modificações nas frequências alélicas de regiões genômicas em equinos de trabalho e de corrida da raça Quarto de Milha.

3. OBJETIVOS

O objetivo desta pesquisa foi identificar, por meio da utilização de painéis de genotipagem de SNPs de alta densidade e da aplicação das estatísticas homozigose relativa do haplótipo estendido (REHH) e índice de fixação, F_{ST} , regiões do genoma selecionadas na linhagem de trabalho da raça Quarto de Milha de forma divergente em relação à de corrida.

4. REVISÃO DA LITERATURA

4.1. Os equinos no Brasil e no mundo

Atualmente, há cerca de 60 milhões de cavalos no mundo, a maioria deles vivendo nas Américas, Ásia e alguns países da Europa. O Estados Unidos é o país com maior contingente (aproximadamente 10,5 milhões), seguindo-se da China (6,7 milhões), México (pouco mais de 6 milhões) e Brasil (pouco menos de 6 milhões). Estes quatro países, possuem cerca de 50% da população global de cavalos (GLIPHA, 2011). O impacto do agronegócio do cavalo no Brasil é bastante expressivo, tendo movimentado em 2006 valores econômicos superiores a R\$ 7,5 bilhões ao ano e abrangendo cerca de 3,2 milhões de empregos diretos e indiretos (LIMA *et al.*, 2006). Considerando-se a estimativa de que o país possui 6.000.000 cabeças (FAO, 2014), constata-se que aproximadamente há geração de um emprego para cada dois cavalos.

A utilização de animais na lida com rebanhos bovinos vem sendo o seguimento de maior destaque dentro do agronegócio do cavalo no Brasil, representando 52% da movimentação monetária e 78% dos empregos diretos gerados (LIMA *et al.*, 2006). Esta intensa relação tem feito os cavalos acompanharem os bovinos em seu deslocamento para as regiões Centro-Oeste e Norte do país. Embora a utilização do cavalo na lida com bovinos seja marcante no contexto da equinocultura nacional, o seguimento que envolve os esportes equestres, acompanhando tendência mundial, tem crescido acentuadamente nos últimos anos. Nesse sentido, no período de 1999 a 2004 o número de eventos envolvendo as diversas modalidades equestres (salto, adestramento, enduro, concurso completo de equitação, etc.) cresceu 315%, ou seja, aumento médio de

15,3% ao ano (FAO 2006). Além da utilização do cavalo na lida da fazenda e nas provas equestres, a tendência de sua utilização para lazer vem aumentando significativamente. Muitas pessoas desejam ter cavalo para dispor dos momentos de descanso longe dos grandes centros, estar em contato com a natureza e, desta forma, se desfazer do estresse. O cavalo também é utilizado na equoterapia, modalidade disponível há milhares de anos e, agora, reconhecida como de grande eficácia para o tratamento de inúmeros males físicos, psíquicos e comportamentais.

4.2. Origem e situação atual da raça Quarto de Milha no Brasil

O cavalo da raça Quarto de Milha teve sua formação na América do Norte, a partir do século XVII, com a introdução de equinos de origem árabe e turca, trazidos por colonizadores europeus. Contudo, seu maior desenvolvimento ocorreu com a ocupação do oeste Norte Americano, devido à necessidade de cavalos robustos e versáteis, com aptidão à sela e tração, visto a dificuldade de se manter plantel variado de animais para atender as diversas necessidades. A população desses equinos cresceu e em 1940 fundou-se a *American Quarter Horse Association* (AQHA), primeira envolvendo a raça. Atualmente conta com aproximadamente cinco milhões de cavalos registrados em todo mundo (ABQM, 2012).

No Brasil existem aproximadamente 358.000 animais registrados junto a Associação Brasileira de Criadores de Cavalo Quarto de Milha (ABQM), desde sua fundação em 1969, tendo impacto relevante no agronegócio nacional do cavalo, visto que estão avaliados em aproximadamente US\$ 785,5 milhões, ocupando área de 491,6 mil hectares de propriedades rurais, estimadas em US\$ 685,7 milhões (ABQM, 2012).

Dentro da raça Quarto de Milha há subdivisões em diferentes segmentos de aptidão, provenientes de distintos objetivos de seleção, consideradas linhagens, entre as quais: a de trabalho, a de conformação e a de corrida. A linhagem de trabalho destina-se às provas de caráter funcional, explorando habilidades como velocidade, agilidade, obediência, e o *cow sense*, características consideradas de grande importância no manejo do gado a campo. A linhagem de conformação enfatiza a morfologia do padrão racial. A linhagem de corrida explora a aptidão dos animais quanto à velocidade em pistas retas e de curta distância. Apesar do efetivo de animais ser relativamente menor nesta que nas demais linhagens, sua importância econômica é substancial, não somente por gerar renda por meio de premiações e apostas (US\$ 2,5 milhões – LIMA *et al.*, 2006), mas também pelo elevado custo gerado na manutenção destes animais dentro desta modalidade esportiva (entre R\$ 800,00 e R\$ 1.400,00, em média mensal, excluindo-se medicamentos e procedimentos veterinários).

O cavalo Quarto de Milha da linhagem de trabalho apresenta de seus ancestrais a agilidade dos cavalos de raças sanguíneas e a estabilidade emocional dos cavalos de raças linfáticas, possui condição atlética, temperamento dócil e rápida aprendizagem. É, na América do Norte, a linhagem que melhor se adapta a diferentes condições climáticas de trabalho e esporte, demonstrando boa velocidade e excelente disposição para treinamento (NCHA, 2013).

Os cavalos da raça Quarto de Milha são muito utilizados para trabalho a campo com gado nos Estados Unidos, principalmente em razão do *cow sense*. O *cow sense* pode ser visto como a demonstração da capacidade dos equinos de agir com antecedência aos bovinos, somada ao seu perfil de boa convivência e

amabilidade (uma vez que estabelece boa relação com os bovinos). Diante de uma situação de apartação, o cavalo com *cow sense* bem desenvolvido apresenta iniciativa própria para dominar o rebanho e dele separar o animal desejado pelo cavaleiro.

Animais da linhagem de trabalho caracterizam-se por uma flexibilidade que permite movimentos com velocidade em qualquer direção: são capazes de girar, colocando seu peso sobre os membros pélvicos e mover-se primeiramente com o focinho. Isso ajuda o cavalo a girar rapidamente enquanto permanece focado na vaca. Nas competições se leva em conta o desempenho dos cavalos a partir de uma série de fatores tais como: coragem, apelo visual, trabalho com o rebanho, controle do bovino, grau de dificuldade da execução, tempo de trabalho, e rédeas soltas. Com base nisso, pontos são adicionados ou subtraídos (NCHA, 2013).

Nos Estados Unidos, a primeira competição registrada de cavalos de apartação foi realizada em Haskell, Texas, em 1898. O interesse por este tipo de competição aumentou tanto que, em 1946, foi criada a *National Cutting Horse Association* (NCHA) a fim de padronizar as regras da competição e criar uma modalidade esportiva (ELLERSIECK *et al.*, 1985).

Atualmente, a competição de cavalos de apartação é esporte equestre de rápido crescimento e altos investimentos. Um exemplo claro disso é que, nos EUA, em 1984, a premiação girava em torno de US\$ 237.000, enquanto que recentemente se compete em torneios da *NCHA Futurity* por mais de 3,7 milhões de dólares (NCHA, 2013).

4.3. Melhoramento genético de equinos

Embora algumas pesquisas publicadas em cavalos envolvam a área de melhoramento genético, ainda não existem, efetivamente, programas consistentes de seleção nas diferentes raças criadas no País. Em relação a outras espécies de exploração zootécnica, as pesquisas na área de melhoramento genético são proporcionalmente menores em equinos, no mundo todo. Pequenos grupos de pesquisadores ligados às universidades como as Universidades de Missouri em Columbia, Universidade de Arizona nos EUA, Universidade Estadual Paulista (UNESP) e outros institutos de pesquisa têm publicado na área, especialmente aspectos quantitativos de caracteres de interesse econômico ou de conservação.

Existem algumas vantagens quando se pesquisa na área de melhoramento genético de cavalos, utilizando-se princípios de genética quantitativa. A profundidade da genealogia na maioria das raças é alta, características de desempenho geralmente podem ser medidas em ambos os sexos, repetidamente, e em períodos de tempo relativamente curtos. Por outro lado, aspectos inerentes à espécie equina (baixos índices reprodutivos, altos intervalos de geração e parto, baixo número de progênes por parição e longo período de gestação), e em função de aspectos operacionais (informações escassas e imprecisas de caracteres reprodutivos, comportamentais e de desempenho em grande parte das raças, baixa receptividade das associações de criadores às tecnologias reprodutivas, relação superficial entre órgãos técnicos e criadores) têm sido algumas das dificuldades ao se pesquisar melhoramento genético de cavalos no Brasil.

Com relação ao Quarto de Milha, não há estudos de melhoramento genético envolvendo a linhagem de trabalho, basicamente devido à escassez/inconsistência

de informação de desempenho armazenada atualmente no *Stud Book* da ABQM. No entanto, desde 2009, esta associação vem de modo regular, registrando dados de desempenho de todos os animais participantes de provas homologadas, o que permitirá, em curto prazo, a realização de pesquisas. Pelos mesmos motivos, que na linhagem de trabalho, não há pesquisas na área de melhoramento envolvendo a linhagem de conformação. Uma vez que apenas na linhagem de corrida encontram-se disponíveis dados de desempenho em quantidade e consistência suficientes, somente nela foram realizadas pesquisas envolvendo melhoramento genético (VILELA *et al.* 2002, MOTA e ABRAHÃO, 2004, CORRÊA e MOTA, 2007).

4.4. Marcadores moleculares

Marcador molecular é toda e qualquer variação oriunda que pode detectar um gene expresso ou de um segmento específico de DNA (correspondente a regiões expressas ou não do genoma). Ao se verificar que 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 como marcador genético (FERREIRA e GRATTAGAGLIA, 1998). Do ponto de vista molecular ocorrem dois tipos principais de variações na molécula de DNA. As variações de comprimento e as variações de sequências.

Marcadores moleculares permitem que os genótipos dos indivíduos sejam determinados em muitos *loci* para várias espécies diferentes (O' BRIEN e GRAVES, 1991), possibilitando que parâmetros genético-populacionais como frequências alélicas e genotípicas sejam estimados. Estas informações permitem a comparação

de frequências entre populações e revelam diferenças em suas composições genéticas que podem contribuir para variações fenotípicas (MOODY *et al.*, 1996).

Para a identificação das regiões do DNA responsáveis por características de interesse, entretanto, geralmente, recorre-se a algumas estratégias diferentes, como a localização de QTL (*Quantitative Trait Loci*) pela análise de regiões microssatélites do maior número possível de indivíduos de grandes famílias, originárias de acasalamentos direcionados (HALEY, 1995).

A aplicação dessa ferramenta porém, esbarra na dificuldade de se constituir famílias para o estudo, no elevado custo para a manutenção dos animais, principalmente em espécies, cujo intervalo entre gerações é grande (caso dos bovinos e equinos), e na quantidade de trabalho e tempo necessários para a coleta dos dados moleculares. Além disso, após a identificação de QTL ainda há um longo caminho a percorrer até a descoberta dos genes diretamente implicados com o fenótipo.

Uma metodologia alternativa para minimizar essas dificuldades é a busca de genes candidatos principais, em que o objetivo é estudar os mecanismos fisiológicos envolvidos com a manifestação das características de produção em questão, na tentativa de pesquisar variações de genes específicos entre indivíduos que apresentam fenótipos diferentes (WOMACK, 1993). A estratégia do gene candidato tem como maior vantagem o fato de não ser obrigatória a genotipagem de um grande número de indivíduos pertencentes as grandes famílias originados de acasalamentos direcionados, uma vez que, a princípio, espera-se que o polimorfismo estudado seja determinante direto o variação fenotípica.

Rocha *et al.* (1994) utilizaram o termo “Ilusão Aditiva” ao se referir a um exagerado otimismo ao fácil emprego de marcadores com o propósito de seleção, considerando que alelos isolados não podem ser classificados como bons ou ruins e que somente o genoma como um todo pode receber este tipo de classificação, ainda assim, quando relacionado à determinada condição ambiental.

Neste sentido, a maior dificuldade da seleção assistida por marcadores reside no fato de que somente uma pequena proporção da variância genética total da característica de interesse é capturada pelos marcadores moleculares, limitando desta forma o progresso ou ganho genético (SOLBERG *et al.*, 2008). Segundo MEUWISSEN *et al.* (2001), a solução para este problema é identificar a maior parte dos genes e marcadores (polimorfismos) ou os genes e marcadores principais que afetam tais características. Entretanto, este anseio esteve por muito tempo limitado devido à restrições tecnológicas.

O recente desenvolvimento de novas metodologias de sequenciamento de DNA vem mudando radicalmente este quadro. As chamadas “tecnologias de sequenciamento de segunda geração” (Roche 454 – MARGULIES *et al.*, 2005, Solexa-Illumina – Bennett, 2004 e ABI Solid – VALOUEV *et al.*, 2008) são capazes de produzir vastos conjuntos de dados (da ordem de milhões de bases seqüenciadas) em um único experimento altamente automatizado, em curto espaço de tempo.

Embora muito mais eficiente que o sequenciamento tradicional, a montagem utilizada nos sequenciamentos de segunda geração depende da disponibilidade de uma seqüência referência do genoma da espécie de interesse. Toda essa situação mudou drasticamente nos últimos anos a velocidade de descoberta de polimorfismos

do tipo SNP no genoma humano e de espécies de exploração zootécnica ou de interesse para o homem.

4.5. Marcadores SNPs

À medida que as sequências de nucleotídeos dos genomas foram sendo desvendadas, uma característica observada foi o grande número de variações de ponto encontradas ao se comparar segmentos correspondentes do mesmo genoma. As mais comuns ocorrem, aproximadamente, a cada 600 bases e são denominadas polimorfismos de nucleotídeos únicos ou SNPs. 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.

As substituições mais frequentes observadas no DNA envolvem 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. Caso mutações ocorram em células germinativas, sejam transmitidas às gerações seguintes e permaneçam na população a uma frequência mínima de 1%, passam a ser denominadas de polimorfismos (KWOK & GU, 1999).

Os SNPs podem ocorrer em regiões codificadoras ou com função regulatória, bem como em espaços intergênicos, sem função determinada. Em regiões codificadoras, quando resultam em uma substituição de aminoácido na sequência protéica, são denominados não-sinônimos, podendo a substituição ser conservativa ou não conservativa em função das características dos aminoácidos envolvidos na troca. Nesses casos, pode haver modificações estruturais e funcionais na proteína.

Embora SNPs sinonímicos não alterem a sequência protéica, podem modificar a estrutura e a estabilidade do RNA mensageiro, e, conseqüentemente afetar a quantidade produzida de proteína. 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 processamentos alternativos, geração ou supressão de códons de terminação, alteração nos códons de iniciação da tradução e alterações no padrão de expressão de genes quando a troca de bases ocorre em sequências promotoras (GUIMARÃES e COSTA, 2002).

Recentemente, polimorfismos intrônicos ganharam importância pelo fato de não mais poderem ser descartados como possíveis responsáveis diretos por alterações fenotípicas. Sabe-se que RNAs não-codificantes transcritos a partir de regiões de íntrons (micro-RNAs) estão envolvidos em diferentes processos biológicos tais como os controles transcricional e pós-transcricional da expressão gênica (NAKAYA *et al.*, 2007).

Estudos em humanos, e em espécies de interesse zootécnico, mostraram a ocorrência de milhões SNPs 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*). Além dos marcadores SNP serem abundantes, suas bases moleculares permitem sua distribuição homogênea pelo genoma (CAETANO, 2009).

4.6. Métodos de genotipagem de SNPs

Em razão da possibilidade de relação entre SNPs e doenças genéticas em humanos e características produtivas e reprodutivas em animais e vegetais, ao longo das últimas décadas muitas tecnologias foram desenvolvidas para sua genotipagem.

Entre as metodologias básicas para genotipagem individual de SNPs, em regiões específicas do genoma, estão a PCR-RFLP (MAEDA *et al.*, 1989), a PCR alelo-específico (SAIKI *et al.*, 1986) e a PCR-SSCP (ORITA *et al.*, 1989). Estes métodos não requerem estruturas laboratoriais sofisticadas e podem ser aplicados a experimentos que envolvam genotipagem em pequena escala, nos quais o número de polimorfismos genotipados seja da ordem de algumas dezenas e o de animais analisados não ultrapasse algumas centenas.

O próximo avanço tecnológico para genotipagem de SNPs surgiu com metodologia que utiliza o sequenciamento direto de produtos de PCR. Neste caso, após a amplificação do fragmento desejado, realiza-se a determinação do genótipo com base na análise da sequência produzida. Embora esta metodologia permita a genotipagem de mais de um SNP em um mesmo fragmento, e seja baseada em equipamentos geralmente encontrados em instituições de ensino/pesquisa (seqüenciadores automáticos de DNA), ainda é necessária grande quantidade de trabalho para a geração de dados. Seguindo a mesma tendência, foram desenvolvidas metodologias de genotipagem de SNPs utilizando equipamentos de PCR em Tempo Real, concebidos inicialmente para quantificação de RNA mensageiros em experimentos de análise da expressão gênica (CAETANO, 2009). Alternativas para a genotipagem de dezenas de SNPs em paralelo ocorrem por meio

das metodologias de extensão (ou sequenciamento) de uma única base, cromatografia líquida de alta pressão desnaturante (DHPLC) e espectrometria de massa (MALDI-TOF).

Avanços tecnológicos recentes culminaram com o desenvolvimento de metodologia para genotipagem de dezenas de milhares de SNPs em um único ensaio a baixos custos (de US\$ 0,10 a US\$ 0,001 por SNP genotipado), trazendo novas perspectivas para aplicações já consolidadas, além do desenvolvimento de novas aplicações, produtos e processos, com uso direto nas cadeias produtivas de animais de interesse zootécnico (CAETANO, 2009). Neste sentido, os chamados chips de genotipagem de SNPs de alta densidade, contendo em média mais de 50 mil SNPs (um polimorfismo a cada 50.000 pares de bases), foram gerados e validados para humanos, bovinos, equinos, ovinos, suínos e caninos, sendo extensivamente utilizados por grupos de pesquisa em todo o mundo para o desenvolvimento de ferramentas de diagnóstico de doenças genéticas e seleção genômica.

4.7. Chips de genotipagem de SNPs de alta densidade

A principal limitação nos estudos de associação entre marcadores moleculares e características de interesse está na necessidade de número suficientemente grande de marcadores para aumentar a possibilidade de identificação de marcadores em desequilíbrio de ligação com QTL. Até recentemente, o escasso número, assim como o alto custo de genotipagem dos marcadores, eram fatores limitantes. Assim sendo, delineamentos experimentais

específicos foram desenvolvidos para reduzir o impacto do pequeno número de marcadores sobre o poder dos testes estatísticos (DAETWYLER *et al.*, 2008).

O desenvolvimento de painéis (ou chips) para genotipagem com grande número de SNPs distribuídos ao longo dos genomas está permitindo avanços importantes nos estudos de localização de genes e polimorfismos que afetam características de importância econômica. Com o desenvolvimento destes painéis com número suficiente de marcadores, maior proporção do genoma é coberta, aumentando as possibilidades de identificação de QTL associados às características de interesse.

Os chips de genotipagem de SNPs de alta densidade permitem a genotipagem de milhares de SNPs distribuídos ao longo de todo o genoma da espécie em estudo, tornando possível a partição de 100% da variabilidade genética aditiva de uma característica, propiciando a estimação do valor de substituição de alelos em cada um dos *loci* envolvidos com o fenótipo. Desta forma, depois de estabelecida a relação entre um conjunto de polimorfismos e uma característica de interesse, é possível estimar o valor genético de um indivíduo com base nos genótipos destes marcadores, ou seja, o valor genômico, sem a necessidade de avaliação fenotípica (MEUWISSEN *et al.*, 2001).

A genotipagem de alta densidade por meio de chips está disponível na forma de serviço terceirizado, eliminando a necessidade de grandes investimentos em equipamentos especializados e em mão-de-obra para a geração dos dados moleculares. Os ensaios são conduzidos com altos níveis de redundância e de forma altamente automatizada, reduzindo erros de genotipagem para índices menores que 0,01%. Estes pontos tornaram possível a coleta de dados moleculares

de um estudo de mapeamento em dias, em vez de meses ou até anos como no caso das pesquisas conduzidas com marcadores do tipo microssatélites (CAETANO, 2009).

4.8. Análise de SNPs em larga escala

O estudo simultâneo de milhares de polimorfismos espalhados ao longo do genoma tem possibilitado o estudo da estrutura genética de diferentes populações em várias espécies de animais domésticos (MCKAY, 2008, KIJAS *et al.*, 2009, GIBBS *et al.*, 2009), determinar a relação entre a perda de alelos e o aumento da endogamia devido à seleção (MUIR *et al.*, 2008), e, recentemente, identificar e localizar regiões do genoma sujeitas à seleção (HAYES *et al.*, 2006; PRASAD *et al.*, 2008; BARENDSE *et al.*, 2009; MACEACHERN *et al.*, 2009).

Ao longo da domesticação e formação das raças, os animais domésticos experimentaram a seleção natural e a artificial. Estas pressões de seleção levaram ao aumento da frequência de algumas mutações em regiões específicas do genoma, as quais tornaram os indivíduos mais adaptados ou deram a eles características favoráveis com base na demanda humana. Ao mesmo tempo, outros polimorfismos apresentaram diminuição de frequência ou eliminação completa. Desta forma, um haplótipo contendo múltiplos genes pode ter se tornado o único ou o mais proeminente na população, situação conhecida como seleção positiva (ANDERSON & GEORGES, 2004).

A identificação de regiões do genoma submetidas à seleção é de grande interesse e pode ser realizada utilizando-se apenas a análise de padrões genômicos, não necessitando de informações fenotípicas. Uma vez identificadas e

mapeadas, é possível anotar o plano funcional dessas regiões do genoma, levando a melhor compreensão de como a seleção atua sobre as características complexas de interesse. Ao mesmo tempo, compreender os mecanismos reais de seleção genômica ocorrida ao longo da história de uma população fornece a base para o desenvolvimento de estratégias mais eficientes de seleção para aplicação futura (SIMIANER *et al.*, 2010).

A comparação das frequências alélicas entre populações selecionadas e não selecionadas ou entre populações selecionadas para diferentes finalidades fornece indícios das regiões do genoma que tenham sido modificadas em função da seleção positiva. Entretanto, a divergência nas frequências alélicas pode ter origem não somente na seleção, como também na deriva genética e na endogamia. Porém, segundo Maceachern *et al.* (2009), a endogamia e a intensidade da deriva genética devem afetar igualmente todos os *loci*, não causando desequilíbrio de ligação entre *loci* adjacentes, como seria esperado no caso da seleção.

Com o intuito de identificar alterações provocadas pela seleção positiva em diferentes regiões do genoma (assinaturas de seleção), alguns métodos têm sido utilizados, entre os quais a estatística *Homozigose Haplótipo Estendido* (EHH) (SABETI *et al.*, 2002), utilizada para a identificação dessas regiões dentro de populações, e o índice de fixação F_{ST} (WRIGHT 1951; COCKERHAM 1969, WEIR e HILL, 2002), utilizado para a identificação de regiões sujeitas à seleção, considerando-se múltiplas populações (SIMIANER *et al.*, 2010).

De acordo com a teoria da seleção artificial, alelos sob seleção positiva apresentam-se em alta frequência e inseridos em extensas regiões de desequilíbrio de ligação. Se a seleção pode carregar um alelo então em um haplótipo específico à

alta frequência mais rápido do que a taxa na qual é desfeito por recombinação, haplótipos em alta frequência serão mais comuns do que o esperado em neutralidade. Nesse sentido, uma “região núcleo” é definida como região de interesse no genoma, presumidamente caracterizada por forte desequilíbrio de ligação entre SNPs e envolve um conjunto de “haplótipos núcleo”.

Segundo Sabeti *et al.* (2002), o teste *Long Range Haplotype* (LRH) utiliza a estatística *Homozigose do Haplótipo Estendido* (EHH), que se baseia na estimação do grau de homozigose entre regiões cromossômicas que carregam um haplótipo comum, sendo, portanto, definido como a probabilidade de que duas regiões cromossômicas homólogas escolhidas aleatoriamente carregando o haplótipo núcleo de interesse, sejam idênticas por descendência para todo o intervalo da região núcleo a uma distância x . O EHH de um haplótipo núcleo testado t é dado por:

$$EHH_t = \frac{\sum_{i=1}^s \binom{e_{it}}{2}}{\binom{c_t}{2}},$$

em que c_t é o número de amostras de um haplótipo núcleo particular t , e_{it} é o número de amostras de um haplótipo estendido particular i , e s é o número de haplótipos estendidos únicos. Taxas de recombinação, entretanto, podem variar consideravelmente ao longo de diferentes regiões do genoma. Isso levanta a possibilidade de que maior estatística EHH pode ser devido às taxas baixas de recombinação em determinada região e não necessariamente a seleção positiva recente.

Em razão disso, a estatística *Homozigose Relativa do Haplótipo Estendido* (REHH), sugerida por Sabeti *et al.* (2002), corrige o valor EHH observado em um haplótipo núcleo para o nível médio dos valores EHH para todos os haplótipos núcleo

relevantes do mesmo cromossomo. REHH é calculado por EHH_t / \overline{EHH} , com \overline{EHH} , definido como a diminuição de EHH em relação a todos os outros haplótipos núcleo combinados dado por:

$$\overline{EHH} = \frac{\sum_{j=1, j \neq t}^n \left[\sum_{l=1}^s \binom{e_l}{2} \right]}{\sum_{i=1, j \neq t}^n \binom{c_i}{2}},$$

em que n é o número de haplótipos núcleo diferentes.

Diferentes métodos têm sido utilizados para a identificação das regiões genômicas submetidas à seleção com base em dados de múltiplas populações. Nesta situação, o pressuposto é que a seleção provoca aumento da diferenciação do genoma, quando populações, linhagens ou raças são selecionadas de forma divergente. A estatística amplamente utilizada para detectar essa diferenciação é a F_{ST} que mede a relação entre os pares de alelos (SNPs) dentro de uma subpopulação em relação à população como um todo (WRIGHT 1951; COCKERHAM 1969, WEIR e HILL, 2002). A estatística F_{ST} pode ser interpretada como uma medida da dispersão de frequências alélicas entre grupos em relação à variação esperada na população a partir do qual estes grupos derivaram.

Gianola *et al.* (2010) sugeriram modelo Bayesiano simples para a extração de amostras das distribuições posteriores dos valores de F_{ST} locus específico. Seguindo a regra de Jeffreys (BERNARDO & SMITH, 1994), uma distribuição *Beta* (0,5; 0,5) é atribuída como prévia para todos os *loci* em todas as populações. Isso leva à densidade posterior conjunta de todas as frequências alélicas como na sequência:

$$g(\mathbf{p}) | \text{Dado} = \prod_{i=1}^N \prod_{l=1}^L \text{Beta}(n_{i1} + 0.5, n_{i2} + 0.5),$$

em que p é o vetor das frequências alélicas, n_{ilk} é o número de alelos observados no *locus* bialélico $1 \leq l \leq L$ no grupo $1 \leq i \leq N$ (ou *Dado*) é o conjunto de alelos observado na amostra. As amostras podem ser extraídas dessa distribuição posterior das estimativas de Monte Carlo como médias posteriores, variâncias, ou intervalos de credibilidade. Com base nisto, também é possível extrair amostras da distribuição posterior de valores F_{ST} . Sendo $p_{il}^{(s)}$, $s = 1, 2, \dots, S$ amostras da distribuição posterior do alelo 1 no locus l no grupo i . Então, uma extração a partir da distribuição posterior de F_{STl} é dada por:

$$F_{STl}^{(s)} = \frac{N \sum_{i=1}^N (p_{il}^{(s)})^2 - \left(\sum_{i=1}^N p_{il}^{(s)} \right)^2}{N \sum_{i=1}^N p_{il}^{(s)} - \left(\sum_{i=1}^N p_{il}^{(s)} \right)^2},$$

que é uma variável aleatória com distribuição (0,1).

A diferenciação de populações, linhagens ou raças por meio do isolamento reprodutivo, consanguinidade ou intensidade da deriva genética, processo que afeta todos os *loci* do genoma de forma similar. Portanto, seria de esperar, na ausência de seleção direcional, que todos os valores de F_{ST} , obtidos ao longo do genoma, provenientes do mesmo processo estocástico compartilhassem a mesma distribuição. Entretanto, se a diferenciação não se baseia apenas em deriva, mas a seleção direcional é ativa e funciona de forma desigual em diferentes regiões do genoma, espera-se maior grau de diferenciação em algumas regiões em comparação com a média da diferenciação causada por deriva ou consanguinidade. Dessa forma, espera-se que os valores de F_{ST} em diferentes regiões do genoma apresentem distribuições diferentes, com valores mais elevados indicando seleção divergente e valores menores refletindo seleção balanceadora.

Os chips de SNPs de alta densidade têm sido utilizados para detectar alterações genéticas provocadas pela seleção, bem como para estudos de associação e identificação de QTL para características de importância produtiva, principalmente em bovinos. Maceachern *et al.* (2009) utilizaram painel de aproximadamente 9.500 SNPs para estudar as diferenças entre bovinos da raça Angus, selecionados para baixa ou alta eficiência de conversão alimentar, e bovinos da raça Holandesa, selecionados para maior produção de leite, proteína e gordura.

Os autores encontraram diferenças importantes nos cromossomos 8, 20 e 24 nos animais da raça Angus, em regiões contendo QTL que afetam a composição corporal (BTA 8 e BTA 24) e rendimento da carcaça (BTA 20). Prasad *et al.* (2008), estimaram o desequilíbrio de ligação entre marcadores (SNP) ao longo dos cromossomos 19 (370 SNPs) e 29 (186 SNPs), identificando regiões sujeitas à seleção que afetam características de importância produtiva em bovinos das raças Angus e Holandesa. Sonstegard *et al.* (2009), utilizando 46.231 SNPs para comparar a linha controle e selecionada para maior produção de leite de vacas da raça Holandesa, identificaram diversas regiões do genoma bovino afetadas pela seleção (BTAs 1, 2, 3, 7, 8, 11 e 26).

Apesar da maior dificuldade de aplicação de ferramentas moleculares em estudos visando a seleção e o melhoramento genético de equinos principalmente, na disponibilidade de animais, as informações referentes ao genoma da espécie vêm experimentando expansão sem precedentes.

Ao longo dos últimos nove ou dez anos, muitas equipes de pesquisa ao redor do mundo analisaram grande quantidade de genes candidatos individuais buscando a identificação de marcadores potencialmente associados à características de

interesse em cavalos tais como: cor da pelagem (BRUNBERG *et al.*, 2006; ROSENGREN PIELBER *et al.*, 2007; REISSMANN *et al.*, 2007); doenças (HANSEN *et al.*, 2007; TRYON *et al.*, 2007; YOUNG *et al.*, 2007); resistência a doenças (SOLBERG *et al.*, 2004; BROWN *et al.*, 2006; RIOS *et al.*, 2007); reprodução e fertilidade (HAMMANN *et al.*, 2007; GIESECKE *et al.*, 2009); comportamento e temperamento (MOMOZAWA *et al.*, 2005, 2006); e desempenho (HANZAWA *et al.*, 2002; REEBEN *et al.*, 2006; MCGIVNEY *et al.*, 2007; HILL *et al.*, 2010a,c; GU *et al.*, 2010).

De acordo com Chowdhary e Raudsepp (2008) entre os maiores destaques provenientes da análise do genoma de equinos estão o seu sequenciamento completo (*EquCab2.0*) e, a partir deste, a identificação de 1,5 milhão de SNPs que ocorrem entre diversas raças. Como resultado destes desenvolvimentos e da aplicação dos chips de SNPs de alta densidade, pode ser antecipado que, nos próximos anos, as bases genéticas de importantes características monogênicas serão analisadas com maior acurácia e rapidez e as características complexas/multigênicas de interesse terão seus componentes genéticos dissecados.

Projetado para permitir a identificação de regiões genômicas modificadas pela seleção e a identificação de SNPs e genes que contribuem para características de interesse nas principais raças de equinos criadas atualmente no mundo, o *Equine SNP50 BeadChip* da empresa Illumina (Illumina Inc., USA) constitui-se em poderosa plataforma 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 um primeiro momento, esses chips serviram com sucesso, principalmente, à identificação de SNPs, regiões genômicas e genes relacionados às importantes doenças e síndromes que acometem determinadas raças tais como lordose (COOK *et al.*, 2009; COOK *et al.*, 2010), osteocondrose (LYKKJEN *et al.*, 2010; TEYSSEDRE *et al.*, 2012), neuropatia laringeal recorrente (DUPUIS *et al.*, 2011), *Dwarfismo* é uma rara doença genética que impede o crescimento normal do indivíduo deixando-o com baixa estatura (EBERTH *et al.*, 2009) e *Lavender Foal* é uma doença genética que afeta recém-nascidos potros de cavalos da raça árabe (BROOKS *et al.*, 2010). Mais recentemente, características complexas relacionadas ao desempenho em provas esportivas e aptidões específicas têm sido alvo de pesquisas por meio de chips de SNPs de alta densidade para equinos (HILL *et al.*, 2010b; BINNS *et al.*, 2010; SCHRODER *et al.*, 2011).

5. MATERIAIS E MÉTODOS

5.1. Animais

Foram utilizados 188 equinos da raça Quarto de Milha criados no Estado de São Paulo, de ambos os sexos, nascidos entre os anos de 1985 e 2009, registrados na Associação Brasileira de Criadores de Cavalos Quarto de Milha (ABQM), sendo 68 e 120 da linhagem de trabalho e de corrida, respectivamente, os animais da linhagem de trabalho, 26 machos e 42 fêmeas, eram filhos de 44 garanhões e 64 éguas, resultando em média de 1,5 progênie/garanhão e 1,06 progênie/égua. Estes animais pertenciam a 42 criadores e encontravam-se alojados em três propriedades no interior do estado de São Paulo, Brasil. Por sua vez, os animais da linhagem de corrida, 18 machos e 102 fêmeas, eram filhos de 48 garanhões e 107 éguas,

resultando em média de 2,5 progênie/garanhão e 1,1 progênie/égua. Estes animais pertenciam a 57 criadores e encontravam-se alojados em cinco propriedades no interior do estado de São Paulo, Brasil. A presença de irmãos completos foi impedida em ambas as linhagens, evitando assim a endogamia. Todos os procedimentos envolvendo os animais utilizados foram realizados de acordo com a legislação brasileira para o bem estar animal.

5.2. Genotipagem dos SNPs

O DNA genômico foi extraído das amostras de sangue dos equinos Quarto de Milha utilizando-se o *Illustra Blood GenomicPrep Mini Spin Kit* (GE Healthcare, EUA), de acordo com as instruções do fabricante. Este DNA foi analisado quanto a sua integridade em gel de agarose a 0,8% e quantificado em espectrofotômetro NanoDrop ND-2000 (Thermo Scientific, EUA). A concentração de DNA das amostras foi ajustada entre 40 e 60 ng/ μ L.

A genotipagem dos SNPs foi realizada com o *Equine SNP50 BeadChip* (Illumina Inc., EUA) utilizando-se o equipamento HiScan (Illumina Inc., EUA), junto à Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal/SP. Com intervalo médio de 43.200 pares de bases entre os SNPs distribuídos ao longo dos 31 cromossomos autossômicos, o painel equino é preparado para a análise de 54.602 SNPs altamente informativos distribuídos uniformemente ao longo do genoma de 15 raças de cavalos. Este conteúdo é derivado do *EquCab2.0 SNP Collection* compilado pelo *Broad Institute's Equine Genome Sequencing Project*, no qual foram identificados mais de 940.000 SNPs entre as raças Árabe, Andaluz, Akhal-teke, Islandesa, Standardbred, Puro Sangue Inglês e Quarto de Milha.

5.3. Controle de qualidade dos dados de genotipagem

Os controles de qualidade das informações de genotipagem, para indivíduos e SNPs, foram realizados por meio do programa *Genome Studio* versão 2011.1 (Illumina Inc., USA).

Em relação aos indivíduos foram determinados o *call rate*, a heterozigosidade e a estimação do gênero. Animais com *call rate* abaixo de 0,95; heterozigosidade com ± 3 desvios padrão da média e com inconsistência de gênero entre registro de coleta e dados genômicos, foram excluídos do conjunto amostral. Avaliou-se ainda, as concordâncias de quatro réplicas e as concordâncias das relações de parentesco (compartilhamento de alelos) entre quatro conjuntos garanhão/progênie e três conjuntos garanhão/égua/progênie.

No que diz respeito aos SNPs, para o conjunto total de indivíduos e para cada uma das linhagens foram excluídos (filtrados) os localizados no cromossomo X, os com baixa qualidade de genotipagem (*cluster separation* inferior a 0,3), os com *call frequency* inferior a 0,9; os com *minor allele frequency* (MAF) inferior a 0,05 (incluindo os fixados) e os com *p-value* inferior a 1×10^{-3} para o equilíbrio de Hardy-Weinberg.

Os SNPs mapeados no cromossomo X foram excluídos do conjunto de polimorfismos genotipados uma vez que somente são informativos (podem ser genotipados) em indivíduos do sexo feminino. SNPs com baixos valores de *cluster separation* (inferior a 0,3) foram removidos por apresentarem baixa precisão na diferenciação de indivíduos homozigotos para o primeiro alelo, heterozigotos e homozigotos para o segundo alelo. SNPs com *call frequency* abaixo de 0,9 foram excluídos por apresentarem perda de informação de genótipo em mais de 10% dos

indivíduos estudados. SNPs com MAF inferior a 0,05, incluindo os fixados ou monomórficos, foram removidos por não serem informativos. Por fim, SNPs com P-valor inferior a 1×10^{-3} para o equilíbrio de Hardy-Weinberg foram removidos tendo em vista que desvios extremos podem indicar problemas na precisão da genotipagem bem como polimorfismos de ocorrência duplicada no genoma.

5.4. Análise dentro da linhagem de trabalho

Para avaliar como o LD decai ao longo do genoma foi utilizado, de acordo com SABETI *et al.* (2002), a estatística EHH. Este teste é baseado no contraste de um haplótipo da região núcleo com combinação de alta frequência e homozigose estendida com outros haplótipos do mesmo *locus*. Dessa forma, EHH é a probabilidade que dois haplótipos núcleo escolhidos ao acaso sejam homozigotos para todo um intervalo que abrange a região núcleo em determinado *locus* (SABETI *et al.*, 2002). Porém, como taxas de recombinação podem variar consideravelmente ao longo de diferentes regiões do genoma, existe a possibilidade de que uma maior estatística EHH possa ocorrer devido à baixas taxas de recombinação em determinada região e não necessariamente à seleção positiva recente. Assim, a estatística REHH, corrige o valor EHH observado em um haplótipo núcleo para o nível médio dos valores EHH para todos os haplótipos núcleo relevantes do mesmo cromossomo.

Após o processo de filtragem dos SNPs, considerou-se para aplicação do teste EHH, somente o conjunto de dados referente à linhagem de trabalho. O *software FastPHASE*, configurado com parâmetros padrão, foi utilizado para reconstrução dos haplótipos e para inferir as fases de ligação dos SNPs em cada

cromossomo. Este *software* implementa métodos para estimar genótipos perdidos e reconstruir haplótipos a partir de dados de SNPs de indivíduos não aparentados, baseando-se em um modelo de *cluster* para haplótipos (SCHEET & STEPHENS, 2006).

Para a detecção de prováveis assinaturas de seleção no genoma da linhagem de trabalho utilizou-se o programa *Sweep 1.1* (SABETI *et al.*, 2002), o qual aplica a estatística REHH. Inicialmente foram identificadas regiões núcleo de acordo com o algoritmo sugerido por Gabriel *et al.* (2002), que considera um par de SNPs em forte LD se a estatística D' estiver dentro de um intervalo de confiança de 95%, com limite superior de D' maior que 0,98 e limite inferior acima de 0,7. Segundo QANBARI *et al.* (2010), regiões núcleo são definidas como regiões de interesse no genoma caracterizadas por forte LD entre os SNPs e envolvem um conjunto de haplótipos núcleo. Especificou-se que regiões núcleo tivessem no mínimo 3 e no máximo 20 SNPs.

O teste EHH foi aplicado a uma distância de 1 cM para ambos os lados da região núcleo. Para determinar a significância dos valores de REHH, ou seja, dos valores de EHH corrigidos para cada cromossomo, o programa *Sweep 1.1* atribuiu os haplótipos em 20 classes de frequência e comparou o REHH para cada haplótipo com os outros igualmente frequentes. Os valores de REHH passaram por transformação logarítmica ($-\text{Log}_{10}$) para que apresentassem frequência de distribuição próxima da Normal.

Em seguida foram obtidas as distribuições de probabilidade, as médias, os desvios-padrão e os P -valores, que também foram transformados para $-\text{Log}_{10}(P)$. Haplótipos da região núcleo (*core haplotypes*) com P -valores extremos na

distribuição, ou seja, com $-\text{Log}_{10}(P)$ superiores a 2 ($P < 0,01$) foram considerados significativos. Entretanto, segundo Glick *et al.* (2012), o REHH pode servir como evidência de seleção recente apenas para haplótipos presentes em frequência relativamente alta na população. Neste sentido, considerou-se relevante somente os haplótipos da região núcleo com frequência maior ou igual a 0,25.

5.5. Análise entre as linhagens de corrida e trabalho

Para a detecção de assinaturas de seleção entre populações comparou-se os valores do índice de fixação F_{ST} entre os *loci*, determinando assim, se diferenças nas frequências alélicas dos SNPs entre as linhagens de trabalho e de corrida seriam decorrentes de seleção positiva recente. Segundo Holsinger e Weir (2009), F_{ST} está diretamente relacionado à variância na frequência alélica entre populações e, inversamente ao grau de semelhança entre os indivíduos dentro de populações. Dessa forma, se o valor de F_{ST} é pequeno significa que as frequências alélicas dentro de cada população são similares, porém, se o valor é grande, indica que as frequências alélicas são diferentes. O valor de F_{ST} pode variar de 0 (nenhuma diferenciação) a 1 (diferenciação completa, em que alelos diferentes encontram-se fixados nas populações) (MORADI *et al.*, 2012).

Para o conjunto final de SNPs que passaram pelo controle de qualidade (considerando-se o total de indivíduos) foi estimado o parâmetro $\theta = F_{ST}$ utilizando-se o método Bayesiano proposto por Gianola *et al.* (2010). Este método é um procedimento de dois passos, eventualmente levando a *clusters* de valores θ . No primeiro passo, utilizou-se estrutura Bayesiana simples para a retirada de amostras da distribuição posterior de θ -parâmetros sem construir cadeias de Markov. Neste

passo utilizou-se a regra de *Jeffreys* (1946) para atribuir uma *priori* pouco informativa de distribuição Beta (1/2,1/2) das frequências alélicas e não se fez suposições sobre modelos evolutivos.

O segundo passo do procedimento consistiu no agrupamento de um conjunto de estimativas de médias posteriores de valores $\theta = F_{ST}$ a partir de análise multi-*locus* em grupos de dados dirigidos, considerando se os valores de θ derivam da mesma distribuição ou de diferentes distribuições resultantes da heterogeneidade de processos estocásticos adjacentes. A estrutura de *cluster* foi dada pela montagem de uma sequência de modelos de mistura finita das médias posteriores de valores $\theta = F_{ST}$ para cada *locus*.

Os parâmetros do modelo foram estimados por máxima verossimilhança por meio do algoritmo de maximização de expectativa (EM), implementado no pacote FlexMix (LEISCH, 2004) no R *project* (R development core team, 2008). Se os valores de θ surgem de diferentes processos evolutivos ou artificiais, será esperada a observação de uma mistura de distribuições levando a *clusters* representativos dos tipos de mecanismos operantes, podendo haver vários *clusters* abrigando SNPs que passaram por diferentes processos de seleção. Para cada SNP de cada *cluster* foi calculada a probabilidade condicional de θ pertencer a esse *cluster*. Os SNPs do *cluster* com os maiores valores de θ e com maiores probabilidades de pertencer a este, provavelmente refletem assinaturas de seleção. Neste sentido, foram consideradas assinaturas de seleção os SNPs com elevados valores de θ e com probabilidade condicional de pertencer ao *cluster* maior ou igual a 95%.

5.6. Anotação dos genes

Para a anotação de genes em regiões genômicas selecionadas na linhagem de trabalho de forma divergente em relação à de corrida foram consideradas as regiões núcleo com haplótipos relevantes, ou seja, regiões com valores de REHH significativos ($P < 0,01$) e com frequência de haplótipos maior ou igual a 0,25 estendidas 1 Mb para ambos os lados, que coincidiram com *loci* (SNPs) pertencentes ao *cluster* com os maiores valores de $\theta = F_{ST}$ e probabilidade condicional de pertencer a este *cluster* maior ou igual a 95%. Os genes correspondentes às regiões do genoma contendo assinaturas de seleção segundo ambas as estatísticas, REHH e F_{ST} , e critérios utilizados, foram identificados com base na montagem mais recente da sequência do genoma equino (*EquCab2.0*) por meio da ferramenta *Map Viewer* do NCBI (NCBI, 2013). A partir da lista de genes identificados foi gerada uma segunda lista contendo apenas aqueles com função conhecida. Estes genes foram anotados quanto aos processos biológicos de que participam utilizando-se os bancos de dados NCBI e *Gene Ontology* e considerando-se a espécie *Homo sapiens*. Genes cujos produtos protéicos possam fazer parte de processos biológicos relacionados à características consideradas importantes para a linhagem de trabalho de cavalos Quarto de Milha foram destacados.

6. Resultados e discussão

6.1. Controle de qualidade dos dados de genotipagem

Quatro indivíduos da linhagem de trabalho foram excluídos do conjunto amostral total por apresentarem *call rate* inferior a 0,95 (95%). As 184 amostras restantes (64 da linhagem de trabalho e 120 da linhagem de corrida) apresentaram

call rate médio de $0,9929 \pm 0,0054$, com valor máximo de 0,9979 e mínimo de 0,9503. A heterozigosidade média estimada para o total de indivíduos foi de $0,3468 \pm 0,0148$, com valor máximo de 0,3788 e mínimo de 0,3049. Neste sentido nenhum animal foi excluído por apresentar heterozigosidade com ± 3 desvios padrão da média ($0,3023 - 0,3914$). Da mesma forma, nenhum indivíduo foi excluído por apresentar estimativa do gênero diferente do esperado. As amostras genotipadas em duplicata apresentaram concordância igual ou superior a ideal (99,8%), com resultados entre 0,9979 e 0,9990. Da mesma forma, as concordâncias das relações de parentesco entre os conjuntos garanhão/progênie e garanhão/égua/progênie foram altas, com resultados entre 0,9985 e 0,9995.

Em relação aos 64 animais da linhagem de trabalho que passaram pelo crivo de qualidade, 13.370 SNPs foram excluídos ao longo da filtragem, permanecendo 41.232 para a aplicação das análises genéticas realizadas com essa linhagem REHH. Considerando-se o conjunto total de indivíduos que passaram pelo controle de qualidade (184 animais), 12.544 SNPs foram excluídos ao longo do processo de filtragem, permanecendo 42.058 para a realização das análises genéticas que envolveram ambas as linhagens.

6.2. Análises dentro de linhagem trabalho

Com os resultados obtidos pelo estudo de mapas de haplótipos (HapMap) realizou-se a estatística REHH para verificar regiões do genoma que apresentam extenso desequilíbrio de ligação e alta frequência e, dessa forma, podem se relacionar à características de importância. A estatística REHH fornece resultados de grande valor ao identificar assinaturas de seleção, e tem sido apresentada em vários

estudos disponíveis na literatura (KAUPE *et al.* 2004; FLORI *et al.* 2009; HAYES *et al.* 2009; MACEACHERN *et al.* 2009; STELLA *et al.* 2010; QANBARI *et al.* 2011).

A descrição dos SNPs informativos e das regiões núcleo definidas pelo programa Sweep 1.1 estão apresentadas na Tabela 1. Foi estabelecido um total de 5.453 regiões núcleo compreendendo 726.943 Kb, ou seja 32% do genoma. A média do comprimento das regiões núcleo foi estimada em $130,73 \pm 124,51$ kb com um máximo de 1.341,00 kb no cromossomo autossômico equino 5 (ECA 5). Encontrou-se 420 regiões núcleo abrangendo 59.823kb no ECA 1, e 58 regiões núcleo no ECA 12, contemplando 7.085 kb, sendo estas as maiores e menores estruturas haplotípicas presentes no genoma.

Tabela 1. Estatística descritiva dos polimorfismos de nucleotídeo único (SNPs); informativos e das regiões núcleo (CR) em cavalos da linhagem de trabalho da raça Quarto de Milha.

ECA	SNP	Compr. ECA (Mbp)	Distância Média (Kb)	N° CR	Compr. Médio CR (Kb)	Compr. de Cobertura CR (Kb)	Compr. Max CR (Kb)	Compr. CR/ Compr. ECA	CR_SNP	Max CR SNPs	CR_SNP/ SNP
1	3.433	186	54,18	420	142,43 ± 130,43	59.823	825,77	0,32	1.874	18	0,54
2	2.197	121	55,07	283	148,49 ± 125,10	42.024	579,52	0,34	1.379	14	0,62
3	2.213	119	53,77	296	133,22 ± 131,22	39.435	793,33	0,33	1.365	15	0,61
4	2.053	109	53,09	263	142,94 ± 144,44	37.593	1055,92	0,34	1.281	20	0,62
5	1.794	100	55,74	227	127,78 ± 140,40	29.007	1341,37	0,29	990	20	0,55
6	1.451	85	58,58	185	125,94 ± 110,71	23.300	514,40	0,27	840	17	0,57
7	1.782	99	55,55	230	137,87 ± 134,52	31.710	678,03	0,32	1.052	17	0,59
8	1.722	94	54,58	224	137,73 ± 152,36	30.851	1127,16	0,32	1.040	17	0,60
9	1.555	84	54,01	203	139,02 ± 125,96	28.221	995,61	0,33	939	20	0,60
10	1.484	84	56,60	196	132,82 ± 130,60	26.033	975,46	0,30	878	17	0,59
11	1.128	61	54,07	144	148,97 ± 143,55	21.451	790,20	0,35	686	18	0,60
12	550	33	60,00	58	122,16 ± 128,61	7.085	643,28	0,21	241	12	0,43
13	754	43	57,02	103	126,09 ± 118,42	12.987	608,46	0,30	431	11	0,57
14	1.777	94	52,89	238	139,55 ± 135,07	33.214	669,59	0,35	1.091	16	0,61
15	1.703	92	54,02	231	132,02 ± 119,41	30.497	868,44	0,33	1.013	15	0,59
16	1.676	87	51,90	230	130,90 ± 138,54	30.108	1186,04	0,34	1.044	20	0,62
17	1.511	81	53,60	201	120,93 ± 115,26	24.308	544,19	0,30	913	15	0,60
18	1.462	83	56,77	187	143,04 ± 134,78	26.749	646,36	0,32	890	15	0,60
19	1.144	60	52,44	159	124,71 ± 116,23	19.829	507,45	0,33	707	14	0,61
20	1.208	64	52,98	161	124,00 ± 112,54	19.964	584,17	0,31	727	12	0,60
21	1.052	58	55,13	148	133,77 ± 113,71	19.799	568,55	0,34	643	11	0,61
22	969	50	51,59	129	116,98 ± 104,09	15.090	550,18	0,30	558	10	0,57
23	986	56	56,79	128	133,59 ± 132,60	17.100	749,20	0,30	553	15	0,56
24	882	47	53,28	124	115,90 ± 125,36	14.372	692,72	0,30	516	11	0,58
25	740	42	54,05	99	145,78 ± 119,60	14.432	609,36	0,36	452	20	0,61
26	765	40	54,90	113	119,00 ± 120,42	13.448	648,28	0,32	488	14	0,63
27	719	40	55,63	101	143,16 ± 142,47	14.459	768,90	0,36	441	13	0,61
28	847	46	54,30	125	139,46 ± 106,41	17.433	452,36	0,37	561	14	0,66
29	616	34	55,19	87	112,06 ± 99,63	9.749	564,78	0,28	359	10	0,58
30	563	30	53,28	86	94,88 ± 94,98	8.160	416,53	0,27	342	9	0,60
31	496	25	50,40	74	117,52 ± 112,38	8.697	563,39	0,34	302	9	0,60
Total	41.232	2.247		5.453		726.943			24.596		
Média			54,56		130,73 ± 124,51			0,32			0,59

Tomando-se os 31 cromossomos, 24.596 SNPs participaram da formação de regiões núcleo, com o mínimo de 9 e o máximo de 20 SNPs. Ainda na Tabela 1 encontram-se apresentados para cada cromossomo a proporção da extensão coberta pelas regiões núcleo versus o total da extensão e o número de SNPs formados por regiões núcleo versus o número total de SNPs informativos.

Considerando-se o total de 5.453 regiões núcleo formadas, foram realizados 38.562 testes EHH, com média de 7,07 testes por região núcleo. Na Tabela 2 estão apresentados resultados da aplicação do teste da homozigosidade do haplótipo estendido, informando, para cada cromossomo, o número de testes e os haplótipos *outliers* da região núcleo. De 14.308 testes nos haplótipos da região núcleo com frequência maior ou igual a 0,25, 100 regiões apresentaram picos *outliers* tomando-se a significância de 0,01, indicando regiões do genoma que poderiam estar sob seleção recente. Essas regiões foram observadas em maior número nos cromossomos 1 e 21 (7), e 4, 7, 10 e 15 (6). Já o cromossomo 12 (ECA 12) não apresentou regiões de importância quando $P < 0,01$.

Para melhor visualização da distribuição das assinaturas de seleção ao longo do genoma, a posição do haplótipo no cromossomo foi plotada contra os P -valores transformados ($-\text{Log}_{10}$) da estatística REHH (Figura 1). Foi observado que a distribuição não é uniforme ao longo do genoma, porém assinaturas de seleção foram identificadas em todos os cromossomos, resultado consistente com a hipótese de que as características sob seleção são complexas e controladas por diferentes regiões do genoma.

Tabela 2. Estatística da aplicação do teste do haplótipo de longo alcance (LRH) no genoma de equinos da linhagem de trabalho da raça Quarto de Milha.

ECA	Testes nos haplótipos da região núcleo*	<i>P</i> -valor < 0,05	<i>P</i> -valor < 0,01
1	1070	53	7
2	736	35	4
3	773	35	4
4	672	33	6
5	604	27	3
6	503	25	4
7	603	28	6
8	566	28	3
9	532	27	4
10	513	27	6
11	353	14	4
12	144	7	0
13	282	15	4
14	617	26	2
15	612	29	6
16	612	34	4
17	516	24	4
18	434	20	1
19	418	19	1
20	421	20	1
21	396	21	7
22	348	15	2
23	319	12	2
24	339	12	4
25	283	14	1
26	310	13	1
27	263	16	1
28	334	19	2
29	255	12	2
30	243	11	2
31	237	13	2
Total	14.308	684	100

* Número de testes nos haplótipos da região núcleo com frequência maior ou igual a 0,25.

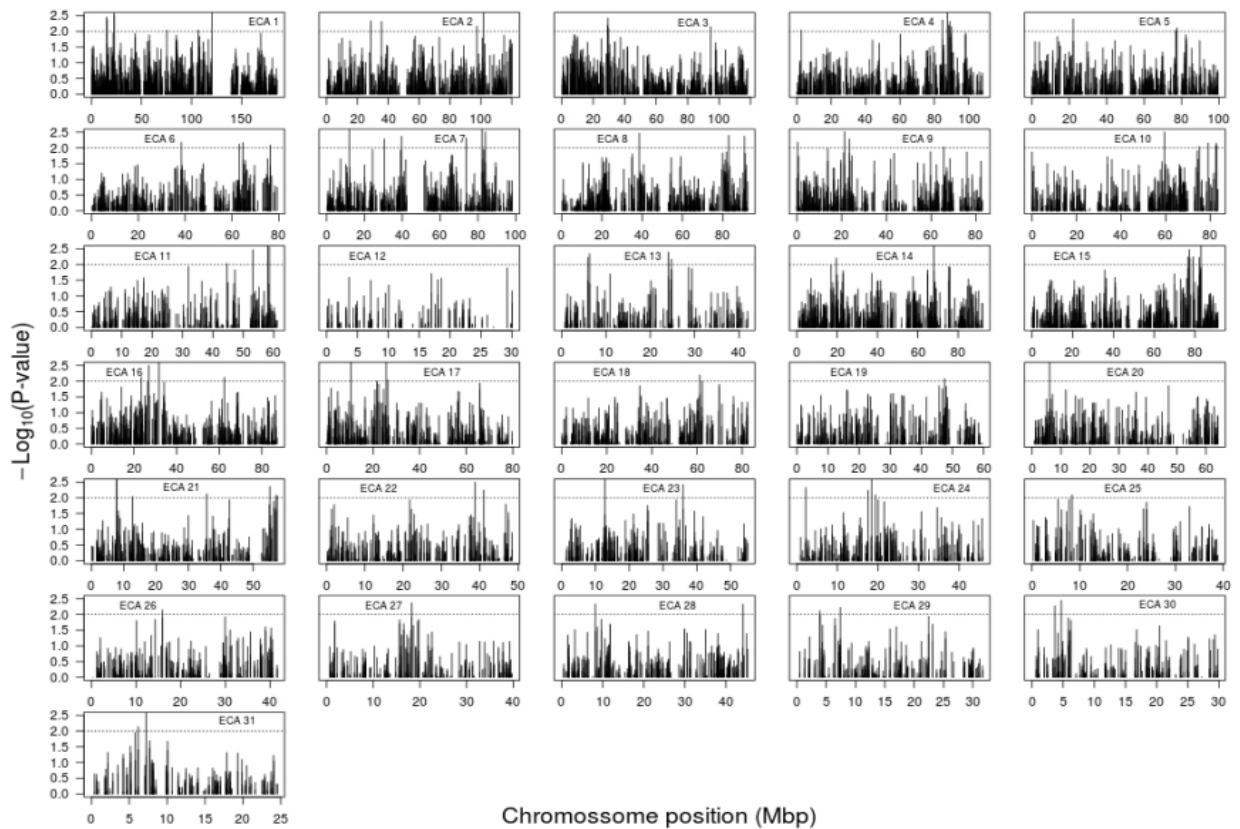


Figura 1. Mapa das prováveis assinaturas de seleção no genoma de equinos da linhagem de trabalho da raça Quarto de Milha. Somente os haplótipos com frequência maior que 25% foram considerados. Linhas tracejadas indicam $P < 0,01$.

6.3. Análises entre as linhagens de trabalho e corrida

A medida de diferenciação genética, F_{ST} , foi realizada com base na análise da distribuição ao nível do genoma. De acordo com Lewontin e Krakauer (1973), Rana *et al.* (1999), Hollox *et al.* (2001), Fullerton *et al.* (2002), Gilad *et al.* (2002) e Hamblin *et al.* (2002), a estatística F_{ST} pode ser considerada estratégia promissora para a detecção de assinaturas de seleção em populações naturais ou selecionadas.

O valor médio obtido do parâmetro $\theta = F_{ST} (\pm \text{desvio-padrão})$ estimado para os 42.058 SNPs informativos, considerando-se ambas as linhagens, foi de $0,0342 \pm$

0,0403, com θ variando de 0,0025 a 0,3556. O *Akaike's information criterion* (AIC) obtido favoreceu o modelo misto finita com sete *clusters* para os 42.058 *loci* (SNPs), sendo os *loci*, normalmente classificados como neutros, sujeitos à seleção balanceadora, ou seleção direcional positiva e/ou negativa. Segundo Gianola *et al.* (2010), se os valores de F_{ST} originam-se de diferentes processos evolutivos ou artificiais ocorre mistura de distribuições, ocasionando a formação dos grupos (*clusters*) que representam os diferentes tipos de mecanismos de funcionamento (operantes). Dos sete *clusters*, o número quatro, constituído de 2558 *loci*, apresentou os maiores valores de F_{ST} , refletindo possíveis assinaturas de seleção. Destes, 1.333 com probabilidade condicional maior ou igual a 95% do *cluster* (Tabela 3). Os valores de F_{ST} , para os 1333 *loci* variaram de 0,1383 a 0,3556, com os 15 maiores, entre 0,3013 e 0,3556, distribuídos nos cromossomos 1, 2, 4, 5, 8, 9, 10, 16, 21 e 30.

Tabela 3. Número de SNPs identificados pela análise F_{ST} com probabilidades de 50 a 95% pertencer ao *cluster* 4.

ECA.	n° SNPs (probabilidade de 50 - 100%)	n° SNPs (probabilidade \geq 95%)
1	343	203
2	85	31
3	63	29
4	144	91
5	141	78
6	82	40
7	120	53
8	93	55
9	79	38
10	222	133
11	102	48
12	26	9
13	35	19
14	113	58
15	45	22
16	52	24
17	71	27
18	110	55
19	41	16
20	46	24
21	91	40
22	114	64
23	33	13
24	46	15
25	74	45
26	38	26
27	20	8
28	63	28
29	13	4
30	45	33
31	8	4
Total	2558	1333

6.4. Anotação de genes em regiões genômicas selecionadas na linhagem de trabalho de forma divergente em relação à de corrida

Considerando-se as regiões coincidentes por ambas a metodologias aplicadas, anotações foram realizadas e, a partir destas, 519 genes identificados, dos quais 205 apresentaram função conhecida. Quarenta e cinco genes apresentaram-se envolvidos em processos biológicos relacionados a metabolismo

energético muscular, neurotransmissão, atividade motora, visão, audição, função cognitiva e sistemas muscular, esquelético, cardiovascular, respiratório e nervoso (Tabelas 5 e 6).

Devido à falta de estudos realizados em cavalos da raça Quarto de Milha, em especial na linhagem de trabalho, e entendendo que as suas habilidades dependem da integração de ações dos diferentes sistemas orgânicos, a presente pesquisa buscou a identificação de assinaturas de seleção e, a partir destas, de genes candidatos posicionais que possam estar relacionados a processos biológicos que determinam características de importância para o desenvolvimento de atividades específicas dessa linhagem. Para identificar regiões do genoma selecionadas na linhagem de trabalho da raça Quarto de Milha de forma divergente em relação à de corrida, foram consideradas as 100 regiões núcleo com haplótipos relevantes provenientes da análise REHH (estendidas 1 Mb para ambos os lados) que coincidiram com os 1333 *loci* (SNPs) da análise F_{ST} .

Dessa sobreposição, resultaram 36 regiões distribuídas nos cromossomos 1, 4, 6, 7, 8, 10, 11, 13, 14, 16, 17, 21, 22, 24, 25, 26, 27, 28, 30, 31 (Tabela 4), as quais, supostamente, abrigam genes relacionados a habilidades importantes para a linhagem de trabalho. A aplicação de duas abordagens diferentes para identificar regiões de assinatura de seleção foi utilizada como procedimento de enriquecimento para corrigir sinais demográficos, de modo que aquelas regiões que apresentam resultados significativos por meio de ambas as estatísticas constituam forte indício de assinaturas de seleção (WALSH, 2010).

Cerca de 200 anos de seleção diferenciaram as linhagens de trabalho e corrida da raça Quarto de Milha em relação as características específicas, que

levaram a fenótipos distintos associados à função a que o animal se destina (EVANS, 1996). Em primeira análise, animais das linhagens de trabalho e de corrida podem ser diferenciados principalmente em função do tipo físico (Meira *et al.*, 2013). Em geral, cavalos de trabalho são mais baixos, compactos e de musculatura pélvica avantajada quando comparados aos de corrida. Por sua vez, cavalos de corrida são mais altos, compridos, de pernas mais longas e musculatura menos protuberante quando comparados aos de trabalho.

O sistema nervoso monitora e coordena a atividade dos músculos, a movimentação dos órgãos, constrói e finaliza estímulos dos sentidos e inicia ações nos animais. De acordo com Strelau (1983), diferentes características do sistema nervoso central formam variadas personalidades e temperamentos. O temperamento de um cavalo pode determinar seu sucesso em uma disciplina esportiva específica ou em determinado tipo de trabalho (MILLS, 1998). Uma vez que é a combinação de temperamento apropriado e habilidades físicas que permitem a formação de um animal vencedor (VISSER *et al.*, 2001).

Esforços na seleção resultaram na formação de raças e linhagens que diferem em temperamento. Cavalos de raças sanguíneas ou enérgicas, tais como animais de corrida, apresentam maior reatividade, enquanto que os linfáticos, como a maioria das raças de tração e trabalho, são mais calmos. Há grande variedade na conformação e temperamentos nos cavalos Quarto de Milha. De acordo com Bertrand Langlois (2012, comunicação pessoal), o cavalo de trabalho deve ser obediente, calmo, fácil de controlar (principalmente em deslocamentos laterais), e muito mais compactos, ou seja, não tão altos à cernelha e com maior largura dos ombros.

Embora diferentes atributos físicos e comportamentais sejam necessários aos vários tipos de equinos de competição, o sucesso em provas depende também da capacidade metabólica do animal para converter energia química em mecânica. Componentes destes processos energéticos incluem a taxa, eficiência e a interação do metabolismo aeróbio e anaeróbio nos músculos e do fornecimento de glicose e oxigênio por meio dos sistemas cardiovascular e respiratório. Segundo Freeman (2013) cavalos desempenham diferentes tipos de exercício físico, variando de predominantemente aeróbio para majoritariamente anaeróbio. Exercícios de arena, como apartação e rédeas, modalidades desempenhadas pela linhagem de trabalho da raça Quarto de Milha, intercalam curtos turnos de exercício anaeróbico com maiores períodos de atividade aeróbia. Corridas de cavalos da raça Quarto de Milha são predominantemente anaeróbicas.

Tabela 4. Regiões do genoma de animais de trabalho da raça Quarto de Milha identificadas como de coincidência pelas análises REHH e F_{ST} , total de genes identificados nessas regiões, genes com função conhecida e genes candidatos possivelmente relacionados a desempenho na linhagem.

ECA	Regiões de coincidência entre as análises REHH e F_{ST}	Total de genes identificados na região	Genes com função conhecida	Genes destacados
1	13982662 .. 16287744	18	8	3
1	105852942 .. 108410726	18	11	4
1	119403352 .. 121554351	28	15	5
4	1646153 .. 3730586	11	6	0
4	86546956 .. 88961701	12	4	1
4	89020999 .. 91022464	17	2	0
4	37421550 .. 39422935	18	8	5
6	75390403 .. 77559160	4	2	0
7	11032566 .. 13323013	25	11	0
7	29488035 .. 31540711	30	7	1
7	38560106 .. 40676848	8	5	2
7	81128673 .. 83162788	14	9	2
7	82918229 .. 84994456	7	5	1
8	89823053 .. 91916645	4	1	0
10	58555790 .. 60651974	24	9	1
10	78182445 .. 80214536	29	6	1
10	81698103 .. 83824983	18	8	0
11	52186121 .. 54420235	21	12	5
13	23101510 .. 25257087	28	12	1
13	23759779 .. 25771201	10	5	1
14	18566294 .. 20644479	17	3	0
14	66970297 .. 68999643	8	4	2
16	22357570 .. 24397019	4	3	0
16	25857799 .. 27923109	6	2	1
17	9529730 .. 11551399	18	8	1
17	20662862 .. 23155099	23	7	1
21	34619933 .. 36629116	6	0	0
22	37783802 .. 39866466	21	12	3
22	40052708 .. 42093213	7	1	0
24	1169652 .. 3174810	16	4	1
25	7372124 .. 9681268	12	5	2
26	14938915 .. 17115722	10	0	0
27	17241350 .. 19574484	4	2	0
28	43187632 .. 45306702	2	1	0
30	3679502 .. 5825839	7	3	0
31	6212711 .. 8351288	14	4	1
Total		519	205	45

Os genes relacionados com atividade motora e resposta à atividade (*SLC18A2*, *BBS4*, *HEXA*, *MYO94*, *SGCE*, *BSX*, *APLP2*, *FIG4*, *MYH1*, *MYH2* e *FEZF2*), visão e percepção visual (*ALDH1A3*, *MYO94*, *TEAD1*, *EYA4*, *MYH13*,

HMGB1 e *SALL4*), audição e percepção auditiva (*HEXA*, *EYA4* e *OTO4*) e função cognitiva e aprendizado (*SGCE*, *BSX* e *APLP2*) merecem atenção pelo fato de que, em confluência, podem estar relacionados ao *cow sense* ou habilidade de apartação. Ao longo da história da indústria do gado bovino na América do Norte, sempre houve interesse na produção de cavalos com *cow sense* superior (ELLERSIECK *et al.*, 1985).

Cow sense, ou habilidade de trabalhar com o gado bovino, pode ser medida pela capacidade do cavalo em cercar o gado e apartar do rebanho um animal escolhido (HINTZ, 1980), com pouca assistência do cavaleiro (ELLERSIECK *et al.*, 1985). Acredita-se que o cavalo deva ter a habilidade de perceber e antecipar os movimentos do bovino para ser um bom apartador (WAGONER, 1978).

A medição dessa característica, entretanto, é considerada difícil. Cada cavalo é julgado por período de dois minutos e meio e, durante este tempo, usualmente aparta dois ou três bezerros do rebanho. A pontuação recebida em competições é uma possível medida de desempenho, porém a sua amplitude de variação é estreita. Logo, a variabilidade entre animais é pequena, o que pode explicar a escassa existência de pesquisas em genética do *cow sense* (HINTZ, 1980). Além disso, existe a dificuldade de se obter informações de um grande número de animais (ELLERSIECK *et al.*, 1985).

Neste sentido, Kieffer (1968) estimou em 0,04 a herdabilidade da habilidade de apartação pela análise de 92 meios irmãos, filhos de 34 garanhões. Por outro lado, Lasley (1976) sugeriu, sem o suporte de dados publicados, que a herdabilidade do *cow sense* deve variar de moderada a alta. De acordo com Hintz (1980), essas informações deixam claro que os aspectos genéticos do *cow sense* não foram

estudados o suficiente para serem incluídos em programas de seleção. Em pesquisa mais consistente com 3.399 cavalos, filhos de 327 garanhões, Ellersieck *et al.* (1985) estimaram a herdabilidade da habilidade de apartação em $0,19 \pm 0,05$.

Considerando que a eficiência das técnicas tradicionais de seleção é bastante limitada em relação as características de baixa herdabilidade, de difícil mensuração, que não se expressam no indivíduo ou de expressão tardia, a utilização de ferramentas moleculares pode trazer grande contribuição para o melhoramento genético destas. Polimorfismos nos genes destacados no presente trabalho, entre os quais os do tipo SNP (Tabelas 5 e 6) podem ser determinantes para variações nas suas expressões qualitativa ou quantitativa e, conseqüentemente, para variações em características importantes necessárias aos cavalos da linhagem de trabalho da raça Quarto de Milha. Estes marcadores, mapeados em genes localizados em assinaturas de seleção, poderão ser utilizados em futuros estudos de associação em cavalos Quarto de Milha de trabalho objetivando o desenvolvimento de ferramentas que aumentem a eficiência de seleção. Nesses estudos, SNPs gênicos responsáveis por troca de aminoácidos na cadeia polipeptídica, localizados em regiões regulatórias e em sítios de ligação de microRNAs, ou relacionados a outras modificações funcionais deverão ser priorizados.

Tabela 5. Genes relacionados com processos biológicos dos sistemas muscular, esquelético, cardiovascular, respiratório, nervoso, atividade motora, visão, audição, função cognitiva, e metabolismo energético muscular, localizados nos cromossomos 1, 4 e 7 próximos as regiões de assinaturas de seleção do genoma de equinos da linhagem de trabalho de trabalho da raça Quarto de Milha, e o número de SNPs identificado nesses genes.

Gene	ECA	Posição (pb)	Processo biológico	SNPs*
SLC18A2	1	14932045..14966719	Metabolismo energético muscular, neurotransmissão, atividade motora	4
ENO4	1	15270780..15295862	Metabolismo energético muscular	29
ATRNL1	1	16112000..16849091	Metabolismo energético muscular, sistema respiratório, comportamento	148
ALDH1A3	1	106035802..106071061	Visão, desenvolvimento facial	1
CHSY1	1	106296983..106368248	Metabolismo energético muscular, sistema esquelético	14
PCSK6	1	106430693..106561014	Desenvolvimento do sistema nervoso, simetria corporal	39
TRPM1	1	108196684..108300518	Percepção visual	13
NPTN	1	119901817..119972398	Desenvolvimento do sistema nervoso, comportamento	18
BBS4	1	120712712..120774697	Des. sistema nervoso, comportamento, atividade motora, percepção auditiva	2
HEXA	1	120969386..120993746	Percepção auditiva, sistema nervoso, atividade motora	5
PKM2	1	121089527..121116345	Metabolismo energético muscular	0
MYO9A	1	121236684..121475426	Percepção visual, atividade motora	14
PLXNA4	4	86335276..86693923	Desenvolvimento do sistema nervoso	176
COL1A2	4	37977963..38013850	Sistema esquelético	4
SGCE	4	38161314..38228239	Sistema muscular, atividade motora, função cognitiva	1
PPP1R9A	4	38360475..38666865	Desenvolvimento do sistema nervoso, transmissão sináptica	40
PDK4	4	38965071..38975327	Metabolismo energético muscular	5
SLC25A13	4	39417697..39585284	Metabolismo energético muscular	106
BSX	7	30186324..30190061	Atividade motora, função cognitiva	3
APLP2	7	38596775..38634442	Des. sistema nervoso, atividade motora, aprendizado espacial, plasticidade sináptica	22
NTM	7	40226037..40610559	Desenvolvimento do sistema nervoso	38
TEAD1	7	81601510..81765984	Visão	23
PTH	7	82252928..82255142	Metabolismo energético muscular, sistema esquelético	0
PDE3B	7	83181012..83364177	Metabolismo energético muscular, sistema cardiovascular	27

*Fonte: NCBI (2013)

Tabela 6. Genes relacionados com processos biológicos dos sistemas muscular, esquelético, cardiovascular, respiratório, nervoso, atividade motora, visão, audição, função cognitiva, e metabolismo energético muscular, localizados nos cromossomos 10, 11, 13, 14, 16, 17, 22, 24, 25, e 31 próximos às regiões de assinaturas de seleção do genoma de equinos da linhagem de trabalho da raça Quarto de Milha, e o número de SNPs identificado nesses genes.

Gene	ECA	Posição (pb)	Processo biológico	SNPs*
FIG4	10	59365115..59480950	Desenvolvimento do sistema nervoso, atividade motora	98
EYA4	10	79540734..79613388	Percepção visual, percepção auditiva	5
GAS7	11	52512007..52593442	Desenvolvimento do sistema nervoso	72
MYH13	11	52722115..52842701	Contração muscular, movimentação ocular reflexiva e voluntária	28
MYH8	11	52864805..52891800	Contração muscular	31
MYH1	11	52973970..52998897	Atividade motora	1
MYH2	11	53001789..53027479	Contração muscular, resposta à atividade	6
OTOA	13	24671087..24737424	Percepção auditiva	63
EEF2K	13	25616499..25668398	Metabolismo energético muscular	3
PAM	14	67200169..67348050	Des. do sistema nervoso, sistema cardiovascular, desenvolvimento de membros	22
ST8SIA4	14	68852965..68942786	Desenvolvimento do sistema nervoso	51
FEZF2	16	26692920..26696781	Desenvolvimento do sistema nervoso, atividade motora	2
HMGB1	17	9783952..9788942	Visão, sistema respiratório, desenvolvimento do sistema nervoso	4
CAB39L	17	21745684..21817545	Metabolismo energético muscular	14
PTPN1	22	38418762..38602458	Metabolismo energético muscular	134
KCNG1	22	38921079..38926986	Neurotransmissão	14
SALL4	22	39550030..39566556	Visão, desenvolvimento de membros	5
BMP4	24	2378786..2385792	Sistema respiratório, sistema cardiovascular, sistema esquelético	1
ALDOB	25	7767982..7776978	Metabolismo energético muscular	8
GRIN3A	25	7908402..8054143	Desenvolvimento do sistema nervoso, neurotransmissão	199
RPS6KA2	31	7099023..7282593	Sistema cardiovascular, transmissão sináptica	150

*Fonte: NCBI (2013)

7. CONCLUSÕES

A utilização de painéis de genotipagem de SNPs de alta densidade e das estatísticas REHH e F_{ST} permitiu a identificação de regiões do genoma equino selecionadas na linhagem de trabalho da raça Quarto de Milha, de forma divergente em relação à de corrida. Ao realizar-se a anotação de genes das regiões de assinatura detectadas por ambas as estatísticas identificou-se, pela primeira vez, aqueles que possam ter sido importantes ao longo do processo de formação da linhagem de trabalho – características genéticas e fisiológicas de interesse humano – incluindo os possivelmente relacionados ao *cow sense*.

REFERÊNCIAS

ABQM. Associação Brasileira dos Criadores de Cavalos Quarto de Milha. Disponível em: <<http://www.abqm.com.br>>. Acesso em: 21 dez. 2012.

ANDERSSON, L.; GEORGES, M. Domestic-animal genomics: Deciphering the genetics of complex traits. **Nature reviews Genetics**, v.5, p.202-212, 2004.

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**, Londres, v. 10, n.178, p. 1-15, 2009.

BENNETT, S. Solexa Ltda. **Pharmacogenomics**, London, v. 5, n. 4, p. 433-438, 2004.

BERNARDO, J. M.; SMITH, A. F. M. **Bayesian theory**. Chichester:Wiley, 1994.

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

BROOKS, S. A. N.; GABRESKI, D.; MILLER, A.; BRISBIN, A.; BROWN, H. E.; STREETER, C.; MEZEY, J.; COOK, D.; ANTCZAK, D. F. Whole-genome SNP association in the horse: identification of a deletion in Myosin Va responsible for Lavender Foal Syndrome. **PLOS Genetics**, Cambridge, v. 6, e1000909, p. 1-7, 2010. Disponível em: <<http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1000909>>. Acesso em: 07 nov. 2013.

BROWN, J. J.; OLLIER, W. E.; THOMSON, W.; MATHEUS, J. B.; CARTER, S. D.; BINNS, M.; PINCHBECK, G.; CLEGG, P. D. TNF-alpha SNP haplotype frequencies in equidae. **Tissue Antigens**, Malden, v. 67, n. 5, p. 377-382, 2006.

BRUNBERG, E.; ANDERSSON, L.; COTHRAN, G. A.; SANDBERG, K.; MIKKO, S.; LINDGREN, G. Missense mutation in PMEL17 is associated with the Silver coat color in the horse. **BMC Genetics**, Londres, v. 7, n. 46, 2006. Disponível em: <<http://dx.doi.org/10.1186/1471-2156-7-46>>.

CAETANO, A. R. Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. **Revista Brasileira de Zootecnia**, Viçosa, v. 38, supl., p. 64-71, 2009.

CHOWDHARY, B. P.; RAUDSEPP, T. The horse genome derby: racing from map to whole genome sequence. **Chromosome Research**, Dordrecht, v. 16, n. 1, p. 109-127, 2008.

COCKERHAM, C. C. Variance of gene frequencies. **Evolution**, Oxford, v. 23, n. 1, p. 72-84, 1969.

COOK, D.; GALLAGHER, P.; BAILEY, E. Illumina equine SNP50 bead chip investigation of adolescent idiopathic lordosis among american Saddlebred horses. **Journal of Equine Veterinary Science**, Amsterdam, v. 29, n. 5, p. 315-316, 2009.

COOK, D.; GALLAGHER, P.; BAILEY, E. Genetics of swayback in American Saddlebred horses. **Animal Genetics**, Chichester, v. 41, suppl. 2, p. 64–71, 2010.

CORRÊA, M. J. M.; MOTA, M. D. S. Genetic evaluation of performance traits in Brazilian Quarter horse. **Journal of Applied Genetics**, Heidelberg, v. 48, n. 2, p. 145-151, 2007.

DAETWYLER, H. D.; SCHENKEL, F. S.; SARGOLZAEI, M.; ROBINSON, J. A. B. A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. **Journal of Dairy Science**, Champaign, v.91, n. 8, p. 3225, 2008

DUPUIS, M. C.; ZHANG, Z.; DRUET, T.; DENOIX, J. M.; CHARLIER, C.; LEKEUX, P.; GEORGES, M. Results of a haplotype-based GWAS for recurrent laryngeal neuropathy in the horse. **Mammalian Genome**, New York, v. 22, n. 9-10, p. 613-620, 2011.

EBERTH, J.; SWERCZAK, T.; BAILEY, E. Investigation of dwarfism among miniature horses using the Illumina Horse SNP50 bead chip. **Journal of Equine Veterinary Science**, Amsterdam, v. 29, n. 5, p. 315, 2009.

ELLERSIECK, M. R.; LOCK, W. E.; VOGT, D. W. Aipperspach R Genetic evaluation of cutting scores in horses. **Equine Veterinary Science**, Hoboken, v. 5, p. 287-289, 1985.

EVANS, J. W. **Horses**: a guide to selection, care and enjoyment. 3rd. ed. New York: Freeman and Company, 1996. 896 p.

FAO. Food Agriculture Organization. 2006. Disponível em: <www.fao.org>. Acesso em:

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

FLORI, L.; FRITZ, S.; JAFFREZIC, F.; BOUSSAHA, M.; GUT, I.; HEATH, S.; FOULLEY, J-L.; GAUTHIER, M. The genome response to artificial selection: a case study in dairy cattle. **PLoS One**, v. 4, n. 8, p. e 6595, 2009. Disponível em: <<http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0006595> >. Acesso em: 08 nov. 2013

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

<<http://pods.dasnr.okstate.edu/docushare/dsweb/Get/Document-5466/ANSI-3983web.pdf>>. Acesso em: Acesso em: 08 nov. 2013.

FULLERTON, S. M.; BARTOSZEWICZ, A.; YBAZETA, G.; HORIKAWA, Y.; BELL, G. I.; KIDD, K. K.; COX, N. J.; HUDSON, R. R.; DI RIENZO, A. Geographic and haplotype structure of candidate type 2 diabetes susceptibility variants at the calpain-10 locus. **The American Journal of Human Genetics**, Cambridge, v. 70, n. 5, p.1096–1106, 2002.

GABRIEL, S. B.; SCHAFFNER, S. F.; NGUYEN, H.; The structure of haplotype blocks in the human genome. **Science**, Washington, v. 296, n. 5576, p. 2225-2229, 2002.

GHILPA. **The global livestock production and health atlas**. Disponível em: <<http://kids.fao.org/glipha/>>. Acesso em: 20 nov. 2011.

GIANOLA, D.; SIMIANER, H.; QANBARI, S. A. two-step method for detecting selection signatures using genetic markers. **Genetic Research**, Cambridge, v. 92, n. 2, p. 141-155, 2010.

GIBBS, R.A.; TAYLOR, J. F.; VAN TASSELL, *et al.* Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. **Science**, Washington, v. 324, n. 5926, p. 528-532, 2009.

GIESECKE, K.; HAMANN, H.; STOCK, K.F.; WOHLKE, A.; SIEME, H.; DISTL, O. Evaluation of SPATA1-associated markers for stallion fertility. **Animal Genetics**, Chichester, v. 40, n. 4, p. 359–365, 2009.

GILAD, Y.; ROSENBERG, S.; PRZEWORSKI, M.; LANCET, D.; SKORECKI, K. Evidence for positive selection and population structure at the human MAO-A gene. **Proceedings of the National Academy of Sciences**, Kirskiville, v. 99, n. 2, p. 862–867, 2002.

GLICK, G.; SHIRAK, A.; ULIEL, S.; ZERON, Y.; EZRA, E.; SEROUSSI, E.; RON, M.; WELLER, J. I. Signatures of contemporary selection in the Israeli Holstein dairy cattle. **Animal Genetics**, Chichester v. 43, suppl. 1, p. 45-55, 2012.

GU, J.; MACHUGH, D. E.; MCGIVENY, B. A.; MCGIVNEY, B. A.; PARQUE, S. D.; KATZ, L. M.; COLINA, E. W. Association of sequence variants in CKM (creatine kinase, muscle) and COX4I2 (cytochrome c oxidase, subunit 4, isoform 2) genes with racing performance in Thoroughbred horses. **Equine Veterinary Journal**, Amsterdam, v. 42, n. 38, p. 569-575, 2010.

GUIMARÃES, P. E. M.; COSTA, M. C. R. SNPs: sutis diferenças de um código. **Biociência**, Brasília, v. 26, p. 24-27, 2002.

HALEY, S. Livestock QTLs - bringing home the bacon? **Trends in Genetics**, Maryland Heights, v.11, n. 12, p. 488-492, 1995.

HAMANN, H.; JUDE, R.; SIEME, H.; SIEME, H.; MERTENS, U.; TOPFER-PETERSEN, E.; DISTL, O.; LEEB, T. A polymorphism within the equine CRISP3 gene is associated with stallion fertility in Hanoverian warmblood horses. **Animal Genetics**, Chichester, v. 38, n. 3, p. 259-264, 2007.

HAMBLIN, M. T.; THOMPSON, E. E.; DI RIENZO, A. Complex signatures of natural selection at the Duffy blood group locus. **The American Journal of Human Genetics**, Cambridge, v. 70, n. 2, p. 369–383, 2002.

HANSEN, M.; KNORR, C.; HALL, A. J.; BROAD, T. E.; BRENIG, B. Sequence analysis of the equine SLC26A2 gene locus on chromosome 14q15-q21. **Cytogenetic Genome Research**, Basel, v. 118, n. 1, p. 55-62, 2007.

HANZAWA, K.; LEAR, T. L.; PIUMI, F.; BAILEY, I. Mapping of equine potassium chloride co-transporter (SLC12A4) and amino acid transporter (SLC7A10) and preliminary studies on associations between SNPs from SLC12A4, SLC7A10 and SLC7A9 and osmotic fragility of erythrocytes. **Animal Genetics**, Chichester v. 33, n. 6, p.455-459, 2002.

HAYES, B. J.; CHAMBERLAIN, A. ; GODDARD, M. E. Use of linkage markers in linkage disequilibrium with QTL in breeding programs. **Proc. 8 th World. Congr. Genet. Appl. Livest. Prod.** Belo Horizonte, Brazil, Vol. pp. (2006).

HAYES, B. J.; LIEN, S.; NIELSEN, H.; OLSEN, H. G.; BERG, P.; MACEACHERN, S.; POTTER, S.; MEUWISSEN, T. H. E. The origin of selection signatures on bovine chromosome 6. **Animal Genetics**, Chichester, v.39, n. 2, p. 105-111, 2008.

HAYES, B. J.; CHAMBERLAIN, A. J.;MACEACHERN S.; SAVIN, K. *et al.* A genome map of divergent artificial selection between *Bos taurus* dairy cattle and *Bos Taurus* beef cattle. **Animal Genetics**, v. 40, 176-84, 2009.

HILL, E. W.; MCGIVNEY, B. A.; GU, J.; WHISTON, R.; MACHUGH, 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**, London, v. 11, p. 552, 2010a.

HILL, E. M.; GU, J.; MCGIVENY, B. A.; MACHUGH, D. E. Targets of selection in the Thoroughbred genome contain exercise-relevant gene SNPs associated with elite racecourse performance. **Animal Genetics**, Chichester, v. 41, suppl. 2, p. 56-63, 2010b.

HILL, E. M.; GU, J.; EIVERS, S. S.; FONSECA, R. G.; MCGIVNEY, B. A.; KATZ, L. M.; MACHUGH, D. GOVINDARAJAN, P.; ORR, N. A sequence polymorphism in MSTN predicts sprinting ability and racing stamina in thoroughbred horses. **PLoS One**,v. 5, p. 1-6, 2010. Disponível em: <<http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0008645>>. Acesso em: 20 out. 2013.

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

HOLLOX, E. J.; POULTER, M.; SWALLOW, D. M. Lactase haplotype diversity in the old world. **The American Journal of Human Genetics**, Cambridge, v. 68, n. 1, p. 160–172, 2001.

HOLSINGER, K. E.; WEIR, B. S. Genetics in geographically structured populations: defining, estimating and interpreting FST. **Nature Reviews Genetics**, London, v. 10, p. 639-650, 2009.

JEFFREYS, H. An invariant form for the prior probability in estimation problems. **Proceedings of the Royal Society of London Series A: mathematical and physical sciences**, London, v. 186, n. 1007, p. 453–461, 1946.

KAUPE, B.; WINTER, A.; FRIES, R.; ERHARDT, G. DGAT1 polymorphism in *Bos indicus* and *Bos taurus* cattle breeds. **Journal of Dairy Research**, Cambridge, v. 71, n. 2, p. 182–187, 2004.

KIEFFER, N. M. Heritability of cutting in horses. In: HORSE SHORT COURSE, 1968, Texas. **Proceedings** Texas: College Station, 1968. p. 46.

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.; McEVAN, J.; COCKETT, N. Genome wide survey of SNP variation reveals the genetic structure of sheep breeds. **PLoS One**, v. 4, p. 3, p. e.4668, 2009.

KWOK, P. Y.; ZHIJIE, G. U. Single nucleotide polymorphism libraries: why and how are we building them?. **Molecular Medicine Today**, Oxford, v. 5, n. 12, p. 538–543, 1999.

LASLEY, J. F. **Genetic principles in horse breeding**. Houston: Cordovan Press, 1976. 124 p.

LEISCH, F. FlexMix: a general framework for finite mixture models and latent class regression in R. **Journal of Statistical Software**, Los Angeles, v. 11, n. 8, p. 1-18, 2004.

LEWONTIN, R. C.; KRAKAUER, J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. **Genetics**, Bethesda, v. 74, n. 1, p. 175–195, 1973.

LIMA, R. A. S.; SHIROTA, R.; BARROS, G. S. C. **Estudo do complexo do agronegócio cavalo no Brasil**. CEPEA – Centro de Estudos Avançados em Economia Aplicada. Piracicaba, ESALQ/USP, 2006. Disponível em: <<http://pt.scribd.com/doc/65305362>>. Acesso em: 15 nov. 2011.

LYKKJEN, S.; DOLVIK, N. I.; MCCUE, M. E.; MICKELSON, J. R.; ROED, K. H. Genome-wide association analysis of osteochondrosis of the tibiotarsal joint in Norwegian Standardbred trotters. **Animal Genetics**, Chichester, v. 41, suppl. 2, p. 111–120, 2010.

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

MAEDA, M.; MURAYAMA, N.; ISHII, H.; URYU, N.; OTA, M.; TSUJI, K.; INOKO, H. A simple and rapid method for HLA-DQA1 genotyping by digestion of PCR amplified DNA with allele specific restriction endonucleases. **Tissue Antigens**, Malden, v. 34, n. 5, p. 290-298, 1989.

MARGULIES, M.; EGHOLM, M.; ALTMAN, W. E. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. **Nature**, London, v. 437, n. 7057, p. 376-380, 2005.

MAYNARD.; SMITH, J.; HAIGH, J. The hitch-hiking effect of a favorable gene. **Genetics Research**, Cambridge, v. 23, n. 1, p. 23–35, 1974.

MCGIVNEY, B.; GU, J.; EIVERS, S.; KATZ, L.; HILL, E. Population and functional genomics investigations of performance associations in thoroughbred horses. In: DOROTHY RUSSELL HAVEMEYER INTERNATIONAL EQUINE GENOME MAPPING WORKSHOP, 7., 2007, Tahoe City, CA.

MCKAY, S. D.; SCHNABEL, R. D.; MURDOCH, B. M.; MATUKUMALLI, L. L. An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. **BMC Genetics**, London, v. 9, p. 37, 2008.

MEIRA, C. T.; CURI, R. A.; SILVA, J. A. II. V.; CORRÊA, M. J. M.; OLIVEIRA, H. N.; MOTA, M. D. S. Morphological and genomic differences between cutting and racing lines of Quarter Horses. **The Journal of Equine Veterinary Science**, Amsterdam, v. 13, n. 4, p. 244-249, 2013.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome wide dense marker maps. **Genetics**, Bethesda, v. 157, n. 4, p. 1819-1829, 2001.

MILLS, D. S. Personality and individual differences in the horse, their significance, use and measurement. **The Journal of Equine Veterinary Science**, Amsterdam, v. 27, p. 10-13, 1998.

MOMOZAWA, Y.; TAKEUCHI, Y.; TOZAKI, T.; KIKUSUI, T.; HASEGAWA, T.; RAUDSEPP, T.; CHOWDHARY, B. P.; KUSUNOSE, R.; MORI, Y. Sequence, detection of polymorphisms and radiation hybrid mapping of the equine catechol-O-methyltransferase gene. **Animal Genetics**, Chichester v. 36, n. 2, p. 190, 2005.

MOMOZAWA, Y.; TAKEUCHI, Y.; TOZAKI, T.; KIKUSUI, T.; HASEGAWA, T.; RAUDSEPP, T.; CHOWDHARY, B. P.; KUSUNOSE, R.; MORI, Y. Polymorphism identification, RH mapping, and association analysis with the anxiety trait of the equine serotonin transporter (SLC6A4) gene. **The Journal of Equine Veterinary Science**, Amsterdam, v. 68, n. 6, p. 619-621, 2006.

MOODY, D. E.; POMP, D.; NEWMAN, S.; MacNEIL, M. D. Characterization of DNA polymorphisms in three populations of Hereford cattle and their associations with growth and maternal EPD in line 1 Herefords. **The Journal of Equine Veterinary Science**, Amsterdam, v. 74, n. 8, p. 1784-1793, 1996.

MORADI, M. H.; NEJATI-JAVAREMI, A.; MORADI-SHAHRBABA, M.; DODDS, K. G.; MCEWAN, J. C. Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. **BMC Genetics**, London, v. 13, p. 2-15, 2012.

MOTA, M. D. S.; ABRAHÃO, A. R. Environmental factors affecting time in Quarter Horse races. **Archivos de Zootecnia da Universidade de Córdoba**, Córdoba, v. 53, n. 201, p. 95-98, 2004.

MUIR, W. M.; WONG, G. K.; ZHANG, Y.; WANG, J.; GROENEN, M. A. M.; CROOIJMANS, R. P. M. A.; MEGENS, H. J.; ZANG, H.; OKIMOTO, R.; VEREIJKEN, A.; JUNGRIUS, A.; ALBERT, G. A. A.; LAWLEY, C. T.; DELANY, M. E.; MACEACHERN, S.; CHENG, H. H. Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. **PNAS**, Washington, v. 105, n. 45, 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. da; 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, article. R-43, 2007.

NCBI. National Center for Biotechnology Information. Disponível em: <<http://www.ncbi.nlm.nih.gov>>. Acesso em: 27 maio 2013

NCHA. National Cutting Horse Association. Disponível em: <<http://www.nchacutting.com>> Acesso em: 13 nov. 2013.

O'BRIEN, S. J.; GRAVES, J. A. M. Report of the committee on comparative gene mapping. **Cytogenetics and Cell Genetics**, Basel, v. 58, p. 1124-1151, 1991.

ORITA, M., IWAHANA, H., KANAZAWA, H.; HAYASHI, K.; SEKIYA, T. Detection of polymorphisms of human DNA by gel electrophoresis as singlestrand conformation polymorphisms. **Proceedings of the National Academy of Sciences USA**, Washington, v. 86, n. 8, p. 2766-2770, 1989.

PRASAD, A.; SCHNABEL, R. D.; MCKAY, S. D.; MURDOCH, M. 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**, Chichester, v. 39, n. 6, p. 597-605, 2008.

QANBARI, S.; PIMENTEL, E. C. G.; TETENS, J.; THALLER, G.; LICHTNER, P.; SHARIFI, A. R.; SIMIANER, H. A genome-wide scan for signatures of recent selection in Holstein cattle. **Animal Genetics**, Chichester v. 41, n. 4, p. 377-389, 2010.

QANBARI, S.; GIANOLA, D.; HAYES, B.; SCHENKEL, F.; MILLER, S.; MOORE, S.; THALLER, G.; SIMIANER, H. Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. **BMC Genomics**, London, v. 12, p. 31, 2011.

RANA, B. K.; HEWETT-EMMETT, D.; JIN, L.; CHANG, B. H.; SAMBUUGHIN, N.; LIN, M.; WATKINS, S.; BAMSHAD, M.; JORDE, L. B.; RAMSAY, M.; JENKINS, T.; LI, W. H. High polymorphism at the human melanocortin 1 receptor locus. **Genetics**, Bethesda, v. 151, n. 4, p. 1547–1457, 1999.

REEBEN, M.; KOHO, N. M.; RAEKALLIO, M.; HYYPPA, S.; POSO, A. R. MCT1 and CD147 gene polymorphisms in standardbred horses. **Equine Veterinary Journal**, Hoboken, v. 38, s36, p. 322-325, 2006.

REISSMANN, M.; BIERWOLF, J.; BROCKMANN, G. A. Two SNPs in the SILV gene are associated with silver coat colour in ponies. **Animal Genetics**, Chichester, v. 38, n. 1, p. 1-6, 2007.

RIOS, J. J.; PERELYGIN, A. A.; LONG, M. T.; LEAR, T. L.; ZHARKIKH, A. A.; BRINTON, M. A.; ADELSON, D. L. Characterization of the equine 2-5 oligoadenylate synthetase 1 (OAS1) and ribonuclease L (RNASEL) innate immunity genes. **BMC Genomics**, London, v. 8, p. 313, 2007.

ROCHA, J. L.; SANDERS, J. O.; TAYLOR, J. F. Genetic markers to manipulate QTL: The additive illusion. In: ANNUAL MEETING OF THE AMERICAN SOCIETY OF ANIMAL SCIENCE, 72., 1994, Minneapolis. **Proceedings** p. 250.

ROSENGREN, PIELBERG, G., GOLOVKO, A., SUNDSTROM, E., *et al.* Positional identification of the grey coat color mutation in horse. In: DOROTHY RUSSELL

HAVEMEYER INTERNATIONAL EQUINE GENOME MAPPING WORKSHOP, 7., 2007, Tahoe City, CA.

SABETI, P. C.; REICH, D. E.; HIGGINS, J. M.; LEVINE, H. Z.; RICHTER, D. J.; SCHAFFNER, S. F.; GABRIEL, S. B.; PLATKO, J. V.; PATTERSON, N. J.; McDONALD, G. J.; ACKERMAN, H. C.; CAMPBELL, S. J.; ALTSHULER, D.; COOPER, R.; KWIATKOWSKI, D.; WARD, R.; LANDER, E. S. Detecting recent positive selection in the human genome from haplotype structure. **Nature**, London, v. 419, n. 6909, p. 832–837, 2002.

SAIKI, R. K.; BUGAWAN, T. L.; HORN, G. T.; MULLIS, K. B.; ERLICH, H. A. Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. **Nature**, London, v. 324, n. 6093, p. 163-166, 1986.

SCHEET, P.; STEPHENS, M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. **The American Journal of Human Genetics**, Cambridge, v. 78, n. 4, p. 629-644, 2006.

SCHRÖDER, W.; KLOSTERMANN, A.; STOCK, K. F.; DISTL, O. A genome-wide association study for quantitative trait loci of show-jumping in Hanoverian Warmblood horses. **Animal Genetics**, Chichester, 43, n. 4, p. 392-400, 2012.

SIMIANER, H.; QANBARI, S.; GIANOLA, D. Detection of selection signatures within and between cattle populations. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 9., 2010, Leipzig. **Proceedings**. Leipzig: WCGALP, 2010.

SOLBERG, O. D.; JACKSON, K. A.; MILLON, L. V.; STOTT, J. L.; VANDENPLAS, M. L.; MOORE, J. N.; WATSON, J. L. Genomic characterization of equine interleukin-4 receptor alpha-chain (IL4R). **Veterinary Immunology and Immunopathology**, Amsterdam, v. 97, n. 3-4, p. 187-194, 2004.

SOLBERG, T. R.; SONESSON, A. K.; WOOLLIAMS, J. A.; MEUWISSEN, T. H. E. Genomic selection using different marker types and densities. **Journal of Animal Science**, Champaign, v. 86, p. 2447-2454, 2008.

SONSTEGARD, T. S. L.; MA, J. B.; COLE, G. R.; WIGGANS, B. A.; CROOKER, C. P.; VAN TASSEL, B. D.; MARIANI, B. D.; DA, Y. Genome signature of artificial selection for high milk yield in holstein cattle. In: ANIMAL GENETICS INTERNATIONAL CONFERENCE, 17., 2009, San Diego. **Proceedings**. p. 10-14.

STELLA, A.; AJMONE-MARSAN, P.; LAZZARI, B.; BOETTCHER, P. Identification of selection signatures in cattle breeds selected for dairy production. **Genetics**, Bethesda, v. 185, n. 4, p. 1451-1461, 2010.

STRELAU, J. **Temperament personality activity**. London: Academic Press, 1983. 375.

TEYSSEDE, S.; DUPUIS, M. C.; GUERIN, G.; SCHIBLER, L.; DENOIX, J. M.; RICARD, A. Genome-wide association studies for osteochondrosis in french trotters. **Journal Animal Science**, Champaign, v. 90, n. 1, p. 45-53, 2012.

TRYON, R. C.; WHITE, S. D.; BANNASCH, D. L. Homozygosity mapping approach identifies a missense mutation in equine cyclophilin B (PPIB) associated with HERDA in the american quarter horse. **Genomics**, Maryland Heights, v. 90, n. 1, p. 93-102, 2007.

VALOUEV, A.; ICHIKAWA, J.; TONTHAT, T.; STUART, J.; RANADE, S.; PECKHAM, H.; ZENG, K.; MALEK, J. A.; COSTA, G.; MCKERNAN, J. K.; SIDOW, A.; FOGO, A.; KOHNSON, S. M. A high resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. **Genome Research**, Woodbury, v. 18, n. 7, p. 1051-1063, 2008.

VILELA, L. C. V.; MOTA, M. D. S.; OLIVEIRA, H. N. Genetic parameters of racing performance traits of quarter horses in Brazil. **Journal of Animal Breeding and Genetics**, Berlin, v. 119, p. 229-134, 2002.

VISSER, E. K.; VAN REENEN, C. G.; HOPSTER, H.; SCHILDER, M. B. H.; KNAAP, J. H.; BARNEVELD, U. M.; BLOKHUIS, H. J. Quantifying aspects of YOUNG horses temperament: consistency of behavioural variables. **Applied Animal Behaviour Science**, Amsterdam, v. 74, n. 4, p. 241-258, 2001.

WAGONER, D. M. **Equine genetics and selection procedures**. Dallas: Equine Research Publications, 1978. 560 p.

WALSH, B. **Tests for molecular signatures of selection**. Beijing: Institute on Statistical Genetics, 2010. 17 p.

WEIR, B. S.; HILL, W. G. Estimating F-statistics. **Annual Review Genetics**, Palo Alto, v. 36, p. 721-750, 2002.

WOMACK, J. E. The goals and status of the bovine gene map. **Journal of Dairy Science**, New York, v. 76, p.1199-1203, 1993.

WRIGHT, S. The genetical structure of populations. **Annals Human Genetics**, Chichester, v. 15, n. 1, p. 323-354, 1951.

YOUNG, A. E.; BOWER, L. P.; AFFOLTER, V. K.; DE COOCK, H. E.; FERRARO, G. L.; BANNASCH, D. L. Evaluation of FOXC2 as a candidate gene for chronic progressive lymphedema in draft horses. **The Veterinary Journal**, London, v. 174, n. 2, p. 397-399, 2007.

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**, Chichester, v. 38, n. 1, p. 7-14, 2007.