UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CÂMPUS DE JABOTICABAL

IDENTIFICAÇÃO DE REGIÕES GENÔMICAS SELECIONADAS DE FORMA DIVERGENTE EM EQUINOS QUARTO DE MILHA DE CORRIDA E TRABALHO

Camila Tângari Meira

Zootecnista

2014

UNIVERSIDADE ESTADUAL PAULISTA - UNESP

CÂMPUS DE JABOTICABAL

IDENTIFICAÇÃO DE REGIÕES GENÔMICAS SELECIONADAS DE FORMA DIVERGENTE EM EQUINOS QUARTO DE MILHA DE CORRIDA E TRABALHO

Camila Tângari Meira

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

2014

Meira, Camila Tângari
M514i Identificação de regiões genômicas selecionadas de forma divergente em equinos Quarto de Milha de corrida e trabalho / Camila Tângari Meira. – – Jaboticabal, 2014
vi, 95 p. ; 28 cm

Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2014
Orientador: Rogério Abdallah Curi
Banca examinadora: Luciana Correia de Almeida Regitano, Mauricio de Alvarenga Mudadu, Henrique Nunes de Oliveira, Josineudson Augusto II de Vasconcelos Silva
Bibliografia

1. Assinatura de seleção. 2. Características quantitativas. 3. Cavalo. 4. Desempenho. 5. Estudo de associação ampla do genoma. 6. SNP I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 636.082:636.1

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

unesp UNIVERSIDADE ESTADUAL PAULISTA

CAMPUS DE JABOTICABAL

FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS DE JABOTICABAL

CERTIFICADO DE APROVAÇÃO

TÍTULO: IDENTIFICAÇÃO DE REGIÕES GENÔMICAS SELECIONADAS DE FORMA DIVERGENTE EM EQUINOS QUARTO DE MILHA DE CORRIDA E TRABALHO

AUTORA: CAMILA TÂNGARI MEIRA

ORIENTADOR: Prof. Dr. ROGERIO ABDALLAH CURI

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

Prof. Dr. ROGERIO ABDALLAH CURI
Departamento de Melhoramento e Nutrição Animal / Faculdade de Medicina Veterinaria e Zootecnia de Botucatu

Profa. Dra. LUCIANA CORREIA DE ALMEIDA REGITANO
EMBRAPA / Embrapa Pecuária Sudeste São Carlos/SP

Prof. Dr. MAURICIO DE ALVARENGA MUDADO MUDADU
EMBRAPA / Embrapa Pecuária Sudeste São Carlos/SP

Prof. Dr. HENRIQUE NUNES DE OLIVEIRA
Departamento de Zootecnia / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal

Prof. Dr. JOSINEUDSON AUGUSTO II DE VASCONCELOS SILVA
Departamento de Melhoramento e Nutrição Animal / Faculdade de Medicina Veterinaria e Zootecnia de Botucatu

Data da realização: 26 de junho de 2014.

DADOS CURRICULARES DO AUTOR

Camila Tângari Meira, filha de Manoel de Araújo Meira e Sonia Tângari, nasceu em Diamantina – MG, em 24 de Julho de 1982. Em 2008, graduou-se no curso de Zootecnia, pela Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM – Diamantina, MG). Em Agosto de 2008, ingressou no curso de Pós-graduação em Zootecnia da UFVJM, sob orientação do professor Idalmo Garcia Pereira, obtendo o grau de mestre em Zootecnia em Julho de 2010. Em Agosto de 2010, iniciou o curso de Doutorado no programa de Pós-graduação em Genética e Melhoramento Animal pela Faculdade de Ciências Agrárias e Veterinárias (UNESP – Jaboticabal, SP) sob orientação do professor Marcilio Dias S. da Mota (*in memorian*) e do professor Rogério A. Curi. Realizou o estágio de Doutorado no exterior na "University of Queensland" – Austrália, de Março à Dezembro de 2013, sob orientação do Dr. Stephen S. Moore

Aos meus amados pais, Manoel e Sônia. Aos meus queridos irmãos e sobrinhos.

Ao meu amor Michel.

Ao saudoso professor Marcilio Dias Silveira da Mota.

DEDICO

AGRADECIMENTOS

À Deus, pela vida, saúde e pelas oportunidades que apareceram no meu caminho.

Ao grande prof. Marcilio Dias Silveira da Mota (*in memorian*) por eu ter tido o prazer da sua convivência, pela orientação, generosidade, amizade, ensinamentos transmitidos e pela grande contribuição deixada ao melhoramento genético de equinos e para o trabalho.

Ao prof. Rogério A. Curi, a quem devo agradecer muito pela orientação, amizade, confiança no meu trabalho, atenção a mim dispensada e pela responsabilidade direta na construção desta Tese.

Ao meu querido pai, Manoel, pelo amor, força, carinho e ajuda.

À minha querida mãe, Sônia, exemplo de força, por todo amor, ajuda e estímulo para seguir em frente no mundo acadêmico.

Ao meu noivo Michel, não só pelo amor, amizade, mas por estar comigo em todos os momentos. Obrigada pela paciência, incentivo e boa disposição em me ajudar nas análises.

Aos meus irmãos (Letícia, Graziela, Viviane, Larissa e Ricardo), sobrinhos (Pietra, Giulia, Sophia, Isabel e Davi) e a toda minha família pela alegria e carinho.

Ao Prof. Henrique Nunes de Oliveira, por não medir esforços em me ajudar nas análises estatísticas deste trabalho.

Ao programa de Pós-graduação em Genética e Melhoramento Animal da FCAV – Jaboticabal e a todos os professores e funcionários.

À CAPES, pela concessão da bolsa de estudo tanto aqui no Brasil quanto com a bolsa para a realização do meu doutorado sanduíche na Austrália (Processo BEX: 13842/12-9) e à FAPESP, pelo financiamento do projeto de pesquisa.

À todos os professores e funcionários do departamento de Melhoramento e Nutrição Animal da Unesp - Botucatu, em especial ao prof. Josineudson Augusto II Vasconcelos Silva e ao prof. Luis Artur Loyola Chardulo, pela amizade e favores e aos funcionários Carlão e à Silene Mamede, pela amizade e caronas.

À família Marques Farah pela amizade, acolhimento, carinho e torcida.

Aos amigos de Botucatu e da pós-graduação de Jaboticabal, em especial à

Graciene Santos, Édina Aguiar, Natália Rincón, Valdecy Cruz, Diogo Garcia, Adam Taiti, Orlando Duitama, Adriana Somavilla, Raphael Bermal e Guilherme Pereira.

Ao Dr. Stephen S. Moore, Dr. Matthew Kelly, Dra. Sigrid Lehnert e a Dra. Emily Piper por me receberem na “University of Queensland” (Austrália), QAAFI e CSIRO.

À Marina Fortes e ao Laércio Porto-Neto (Juca), pela amizade, pelos trabalhos desenvolvidos e acolhimento durante o doutorado sanduíche na Austrália.

Aos amigos da Austrália: Amy Louise, Mr. Russell, Mrs. Flynn, Mrs. Ruth, Greg, Bing, Paula, Mauricio Mudadu, João Paulo A. Rego e Ana Paula Santos; em especial ao McLean McNabb, pela grande amizade, pelo acolhimento, pelos momentos de felicidades e toda ajuda durante a minha estadia.

Aos membros da banca examinadora, Dra. Luciana Correia de A. Regitano, Dr. Mauricio A. Mudadu, Dr. Henrique N. de Oliveira e Dr. Josineudson Agusto II de V. Silva.

SUMÁRIO

Página

IDENTIFICAÇÃO DE REGIÕES GENÔMICAS SELECIONADAS DE FORMA DIVERGENTE EM EQUINOS QUARTO DE MILHA DE CORRIDA E TRABALHO

RESUMO – Caracterizada por sua grande versatilidade em várias modalidades equestres, a raça Quarto de Milha se destaca mundialmente, representando 52% de toda população equina no mundo. Dentro da raça há subdivisão em diferentes segmentos de aptidão, provenientes de distintos objetivos de seleção, consideradas linhagens, entre elas corrida e trabalho. Objetivou-se com este estudo avaliar as diferenças morfológicas e genômicas que ocorrem entre as linhagens de corrida e trabalho como resultado da seleção para diferentes objetivos, a identificação de regiões genômicas que tenham sido alteradas pela seleção na formação da linhagem de corrida em relação à linhagem de trabalho e realizar dois estudos, independentes, de associação ampla do genoma (GWAS) para identificar regiões cromossômicas e genes candidatos posicionais associados com características de desempenho na linhagem de corrida e com características morfométricas na raça como um todo. Para as análises foram utilizados 120 animais de corrida e 64 animais de trabalho de ambos os sexos registrados na Associação Brasileira de Criadores de Cavalo Quarto de Milha. As características morfométricas avaliadas foram: peso, altura à cernelha, comprimentos corporal, da canela, da quartela, da garupa, da cabeça, e do pescoço, perímetros torácico, da canela e do casco e a característica de desempenho índice de velocidade. A análise das diferenças morfológicas foi realizada por meio do procedimento GLM do programa SAS utilizando-se modelo que incluiu os efeitos de sexo e linhagem e a covariável idade à mensuração. Para a análise das diferenças genômicas, 54.602 SNPs foram genotipados utilizando-se o painel de marcadores Illumina Equine SNP50 BeadChip. Os resultados obtidos revelaram mudanças significativas entre as linhagens para todas as características morfométricas avaliadas. Animais de corrida apresentaram maiores pesos, alturas, comprimentos e perímetros corporais em relação aos de trabalho. O número de SNPs informativos e a densidade dos SNPs encontrados no genoma dos animais de corrida e trabalho indicaram que o Equine SNP50 BeadChip pode ser utilizado para diferentes propostas na raça Quarto de Milha. Para identificar as regiões do genoma selecionadas na linhagem de corrida de forma divergente em relação à de trabalho, foram utilizadas as estatísticas homozigose relativa do haplótipo estendido (REHH), uma extensão da análise EHH, e o índice de fixação (F_{ST}). Considerando-se essas regiões do genoma foi feita a anotação de genes a fim de identificar aqueles que possam ter sido importantes ao longo do processo de formação da linhagem de corrida. Foram encontradas 27 regiões do genoma, consideradas assinaturas de seleção, por ambas as estatísticas utilizadas. A anotação dos genes foi feita com base na montagem mais recente da sequência do genoma equino (EquCab2.0), encontrando genes envolvidos com crescimento muscular (8), crescimento esquelético (10), metabolismo energético muscular (15), sistema cardiovascular (14) e sistema nervoso (23), entre os quais *FKTN*, *INSR*, *GYS1*, *CLCN1*, *MYLK*, *SYK*, *ANG*, *CNTFR* e *HTR2B*. A análise de associação (GWAS) para a característica de desempenho, índice de velocidade, foi realizada com 40.787 SNPs e 112 animais da linhagem de corrida, utilizando o programa Qxpak.5 e ajustando um SNP de cada vez. Para a anotação dos genes foi considerada a janela de 100 Kb para ambos os lados de cada SNP significante ($P < 0,0001$). Foram encontrados 8 SNPs significativos associados com essa

característica. Genes, cujos produtos proteicos participam de processos biológicos relacionados às características importantes para a linhagem de corrida, foram: *GRM8*, *GRIK2*, *NEB*, *ANK1 e KAT6A*. O estudo de associação (GWAS) para as características morfométricas peso, comprimento da garupa e comprimento do corpo foi conduzido com 42.058 SNPs e 184 animais de corrida e trabalho, utilizando o programa Qxpak.5 e ajustando um SNP de cada vez para cada característica separadamente. Para a anotação dos genes foi considerada a janela de 100 Kb para ambos os lados de cada SNP significante ($P < 0,0001$). A análise revelou 3 SNPs associados com peso, 8 SNPs associados com comprimento da garupa e 2 SNPs associados com comprimento do corpo. Os genes candidatos posicionais e funcionais que emergiram deste estudo foram: *WWOX* e *AAVPR1A*. Mais estudos devem visar todos estes genes candidatos e tentar validar as associações de SNPs aqui relatadas.

Palavras-chave: assinatura de seleção, características quantitativas, cavalo, desempenho, estudo de associação ampla do genoma, SNP

IDENTIFICATION OF GENOMIC REGIONS DIVERGENTLY SELECTED IN QUARTER HORSE BREED FOR RACING AND CUTTING

ABSTRACT – Characterized by its versatility in several sporting activities, the Quarter Horse breed excels globally, representing 52% of the entire equine population in the world. The Quarter Horse breed is subdivided into different lines according to skills resulting from distinct selection objectives, including cutting and racing horses. The aims of the present study were to investigate the morphological and genomic difference, between cutting and racing lines as a result of selection for different objectives; identification of the genomic regions divergently selected in racing line in relation to cutting line and to perform two genome-wide association studies, independently, to identify chromosomal regions and positional candidate genes associated with performance trait in the racing line and morphometric traits in the breed as a whole. To perform the analyses 120 racing animals and 64 cutting animals of both sexes and registered at the Brazilian Association of Quarter Horse Breeders were used. The evaluated morphometric traits were: weight; height at withers; length of the body, shank, pastern, rump, head, and neck; circumference of the chest, shank, and hoof and speed index performance trait. Morphological differences analysis was performed using a model that included the fixed effects of sex and line, and age at recording as covariate. The GLM procedure of the SAS program was used for statistical analysis. To perform genomic differences analysis, 54,602 SNPs were genotyped by the Illumina Equine SNP50 BeadChip array. The results showed significant changes in the morphometric traits of the animals. Racing animals were heavier and taller and presented greater body lengths and perimeters than cutting horses. The number of informative SNPs and the SNP density found in the genome of cutting and racing animals suggest that the Equine SNP50 BeadChip can be used for different purposes in the Quarter Horse breed. To identify genomic regions divergently selected in the racing line in relation to the cutting line the relative extended haplotype homozygosity (REHH) statistic, an extension of EHH analysis was used as well as the fixation index F_{ST}. Considering these regions of the genome, the annotation of the genes was made to identify those that could have been important during formation of the racing line. Twenty seven regions were considered selection signatures by both statistical methods. The gene annotation was made based on the latest assembly of the horse genome sequence (EquCab2.0) and genes involved in muscle growth (8), skeletal growth (10), muscle energy metabolism (15), cardiovascular system (14) and nervous system (23), including *FKTN*, *INSR*, *GYS1*, *CLCN1*, *MYLK*, *SYK*, *ANG*, *CNTFR* and *HTR2B* were found. Association analysis (GWAS) for speed index trait was performed with 40,787 SNPs and 112 racing animals, using Qxpak.5 software and fitting one SNP at a time. In order to perform gene annotation the window 100Kb upstream and downstream to each significant SNP (*P*-value < 0.0001) was considered. Eight significant SNPs associated with speed index were found. Genes whose protein products participate in biological processes related to important traits in the racing line, within nearest regions from associated SNPs were: *GRM8*, *GRIK2*, *NEB*, *ANK1* and *KAT6A*. The GWAS for the morphometric traits: weight, rump length and body length was conducted with 42,058 SNPs and 184 racing and cutting animals, using Qxpak.5 software and fitting one SNP at a time for each trait separately. For the annotation of genes was considered the window 100 Kb upstream and downstream to each

significant SNP ($P < 0.0001$). The analysis revealed 3 SNPs associated with weight, 8 SNPs associated with rump length and 2 SNPs associated with body length. The positional and functional candidate genes that emerged from this study were: *WWOX* and *AAVPR1A*. Further studies should target all these candidate genes and attempt to validate the SNP associations reported here.

Keywords: selection signatures, quantitative traits, horse, performance, genome-wide association study, SNP

CAPÍTULO 1 – Considerações gerais

Introdução

Raça de grande destaque mundial, representando 52% dos equinos em todo o mundo, o cavalo Quarto de Milha se destaca por sua grande versatilidade em várias modalidades de provas equestres (ABQM, 2014). Dentro da raça Quarto de Milha há subdivisão em diferentes segmentos de aptidão, provenientes de distintos objetivos de seleção, consideradas linhagens, entre as quais corrida e trabalho.

A seleção artificial na raça iniciou-se com os animais de corrida em 1674 (EVANS, 1996), com a busca de indivíduos mais altos, com musculatura mais suave, rápidas explosões de velocidades e com fenótipos mais próximos da raça Puro-Sangue Inglês (SCOTT, 2008). Somente em 1898 os animais de trabalho passaram a ser selecionados para concorrerem em provas de apartação buscando características de instinto “cow sense”, ou seja, a percepção da reação do bovino, corpo robusto e forte, menor e mais compacto (EVANS, 1996; SCOTT, 2008).

Ao longo da domesticação e formação das raças, os animais domésticos foram submetidos a seleção natural e a artificial. 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, uma região genômica contendo uma combinação alélica pode ter se tornado a única ou a mais proeminente na população, situação conhecida como seleção positiva (ANDERSSON; GEORGES, 2004).

A comparação das frequências alélicas entre populações selecionadas e não selecionadas ou entre populações selecionadas para diferentes finalidades fornece indícios das regiões do genoma sujeitas à 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 deve afetar todos os *loci* igualmente e a deriva genética deve afetar os *loci* de forma aleatória, 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, alguns métodos principais têm sido utilizados, entre os quais o índice de fixação F_{ST} (WRIGHT, 1951; COCKERHAM, 1969; WEIR; HILL, 2002), utilizado para a identificação de regiões sujeitas à seleção considerando-se múltiplas populações (SIMIANER et al., 2010), e o teste do haplótipo de longo alcance (LRH), proposto por Sabeti et al. (2002), para a identificação dessas regiões dentro de populações.

Os estudos de associação ampla do genoma (GWAS) utilizando 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 "quatitative trait *loci*" (QTL) para características de importância produtiva e reprodutiva, principalmente em bovinos (MACEACHERN et al., 2009; HAWKEN et al., 2012) e doenças comuns, tais como doenças cardíacas, diabetes, doenças auto-imunes e distúrbios psiquiátricos (VISSCHER et al., 2012). Em equinos, apesar de algumas dificuldades, em função, entre outros, da disponibilidade de animais, alguns estudos tem sido conduzidos em análises de associação com desempenho (BINNS et al., 2010; HILL et al., 2010; TOZAKI et al., 2010), e com características morfométricas (SIGNER-HASLER et al., 2012; TETENS et al., 2013).

O desenvolvimento de painéis (ou chips) para genotipagem com grande número de SNPs distribuídos ao longo dos genomas permite avanços significativos 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.

A partir dos dados fenotípicos e genótipos oriundos da raça Quarto de Milha, o presente trabalho teve como objetivos:

1. Avaliar as diferenças morfológicas e genômicas que ocorrem entre as linhagens de corrida e trabalho como resultado da seleção para diferentes objetivos;

2. Prospecção de regiões genômicas, genes e SNPs que tenham sido alterados pela seleção na formação das linhagens de corrida e de trabalho;

3. Estudo de associação ampla do genoma (GWAS) entre SNPs e

características de desempenho e morfométricas.

Revisão de Literatura

Os equinos no Brasil e no mundo

A história do cavalo iniciou-se há 55 milhões de anos, no baixo eoceno, quando as massas de terra continentais, as faixas montanhosas e os oceanos Atlântico e Índico começaram a se formar. O cavalo, em sua forma mais primitiva, cerca de 50 milhões de anos antes da evolução do homem, era um pequeno mamífero de muitos dedos com cerca de 30 cm de altura aproximadamente, chamado *Hyracotherium*, que acabou desaparecendo há cerca de 40 milhões de anos por não conseguir adaptar-se às constantes mudanças das condições geológicas. Assim, ele foi sucedido pelo *Orohippus* e, subsequentemente, pelo *Epihippus*, animais com estrutura esquelética muito semelhante, mas com dentes progressivamente eficientes. No baixo plioceno surgiu o *Pliohippus*, um animal ungulado, três vezes maior que o *Hyracotherium*, o qual, pelo início da era *Homo sapiens,* evoluiu para o *Equus*, atingindo cerca de 1,32 m de altura (SILVER, 2000).

Como o *Hyracotherium*, o *Equus* parece ter originado na América do Norte. Mas, ao contrário do primeiro, migrou para o Sul, tornando-se o mais antigo cavalo sul americano. Espalhou-se também pela Ásia, Europa e África. Há cerca de 8 mil anos, ele tornou-se extinto nas Américas e os tipos adaptados da Europa, Ásia e África (diferentes espécies de *Equus* surgidas de acordo com o terreno e o clima) tornaram-se os ancestrais exclusivos do cavalo moderno (SILVER, 2000).

Evidências arqueológicas das últimas décadas indicam que a domesticação ocorreu há seis mil anos na região das estepes do sudoeste asiático, onde estão situados atualmente os países Ucrânia e Cazaquistão. Esta ainda é a teoria mais aceita, embora não tenha sido descartada a possibilidade de que a domesticação não tenha ocorrido em pequena área com população restrita, mas sim em ampla área com a domesticação independente de diversas populações de cavalos selvagens (WARMUTH et al., 2012).

Atualmente, existem aproximadamente 58 milhões de cavalos no mundo,

maior parte vivendo na América, Ásia e alguns países da Europa (GLIPHA, 2012). Os Estados Unidos da América é o país com maior número de cabeças (10.250.000), seguido da China (6.802.350), México (6.356.000) e Brasil (5.363.185) (GLIPHA, 2012). Deve-se destacar a redução da população de equinos na Ásia, principalmente na China, de 8.916.154 cabeças em 2000 para 6.802.350 cabeças em 2012 (GLIPHA, 2012), associada à migração interna da população humana, com menor utilização dos equídeos no transporte e agricultura e maior consumo de carne equina (ALMEIDA; SILVA, 2010). Por outro lado, nos Estados Unidos houve aumento expressivo da população de equinos, de 5.240.000 cabeças em 2000 para 10.250.000 cabeças em 2012 (GLIPHA, 2012), em parte devido às restrições legais internas para o abate e exportação de carne de equídeos (ALMEIDA; SILVA, 2010). No Brasil, no período de 2000 à 2012, o número de equinos permaneceu praticamente estável, com uma pequena queda, de 5.831.817 para 5.363.185 (GLIPHA, 2012).

O complexo do agronegócio do cavalo no Brasil é bastante expressivo, movimentando valor econômico superior a R$ 7,5 bilhões ao ano e gerando cerca de 3,2 milhões de empregos diretos e indiretos (LIMA et al., 2006). A utilização de animais na lida com rebanhos bovinos (aproximadamente 5 milhões) vem sendo o segmento 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 a criação de cavalos acompanhar a de 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 segmento 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. 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.

Destacam-se também no agronegócio equino os vários fornecedores de insumos, produtos e serviços para a criação, como medicamentos, rações, selas, acessórios, ferrageamento, entre outros, além do ensino e pesquisa. No complexo agropecuário, o segmento de equinos utilizados em diversas atividades esportivas movimenta valores da ordem de R$ 705 milhões e emprega cerca de 20.500 pessoas, com a participação estimada de 50 mil atletas (LIMA et al., 2006).

Apesar dos dados apresentados por Lima et al. serem do ano de 2006, acredita-se que estes números só tenham aumentado, já que cada vez mais o ramo equestre vêm ganhando notoriedade.

A raça Quarto de Milha

A raça Quarto de Milha foi a primeira a ser introduzida na América do Norte, a partir do século XVII, com a introdução de equinos de origem árabe e turca, trazidos por colonizadores europeus. O 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 às diversas necessidades (ABQM, 2014). Em 1940, fundou-se nos Estados Unidos da América, a “American Quarter Horse Association” (AQHA), primeira envolvendo a raça. Atualmente, é considerada a maior associação de criadores do mundo, com cerca de 400 mil sócios e mais de 5 milhões de cavalos registrados, distribuídos em 43 países, representando 52% dos equinos em todo o mundo (dados até 31/12/2011, ABQM, 2014). Em 1969, foi fundada a Associação Brasileira de Criadores de Cavalo Quarto de Milha (ABQM), com um plantel composto por 424.000 animais registrados, com 79.700 criadores, proprietários e associados cadastrados, espalhados por todos os estados brasileiros (dados até janeiro/2014 ABQM, 2014).

No Brasil, a criação da raça apresenta impacto relevante no agronegócio nacional do cavalo, visto que estão avaliados em aproximadamente US$ 600

milhões, ocupando área de 593,2 mil hectares de propriedades rurais, estimadas em aproximadamente US$ 827,3 milhões (dados até janeiro/2014 ABQM, 2014).

A seleção nesta raça para diferentes propostas (EVANS, 1996), levou à formação de linhagens, entre as quais: corrida, conformação e a de trabalho. A linhagem de trabalho destina-se às provas de caráter funcional, explorando habilidades como agilidade e obediência, características consideradas de grande importância no manejo do gado a campo. A linhagem de conformação enfatiza a morfologia do padrão racial. Já a linhagem de corrida explora a aptidão dos animais quanto à velocidade em pistas retas e de curta distância. De acordo com Meira et al. (2013) existem diferenças significativas entre as linhagens de corrida e de trabalho da raça Quarto de Milha com relação às características morfológicas de peso, altura à cernelha, comprimentos corporal, da canela, da quartela, da garupa, da cabeça, e do pescoço e perímetros torácico, da canela e do casco. Os autores observaram que animais de corrida apresentaram maiores pesos, alturas, comprimentos e perímetros corporais em relação aos de trabalho.

Apesar do efetivo de animais ser relativamente menor na linhagem de corrida do 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).

Um atributo de seleção utilizado pelos criadores do cavalo Quarto de Milha de corrida é a pontuação conhecida como Índice de Velocidade (IV). Este é obtido durante a campanha de um animal com o intuito de classificar o seu desempenho em diferentes condições (distâncias, hipódromo, clima, país) (EVANS, 1996). Cada hipódromo tem sua própria tabela de IV, que é elaborada a partir da média das três vitórias mais rápidas (três melhores tempos) para cada um dos três últimos anos consecutivos, em cada distância, sendo que o valor da média desses nove tempos equivalerá ao IV igual a 100 (JCS, 2014). Os pontos de IV são inteiros e variam de acordo com o tempo, ao nível de centésimos de segundo, seguindo ajustes em acordo com a distância percorrida.

A tabela de IV (Tabela 1) faz a conversão do tempo em pontos do IV com

ajustes pelas distâncias. Como exemplo, nas distâncias de 365 metros (m), 402m e 503m, a cada 4 centésimos de segundo, a mais ou a menos, que um animal obtém, em relação ao tempo que representa o índice de velocidade igual a 100, diminui-se ou acresce-se um ponto neste índice. Assim, ao se considerar que a média das nove vitórias (IV =100) foi de 22 segundos para os 402m, o animal cujo tempo se situe entre 22,01 e 22,04 terá IV igual a 99, se o tempo estiver entre 22,05 e 22,08 o IV será 98, e assim por diante. Por outro lado, se o tempo estiver entre 21,96 e 21,99 será acrescido em um ponto, obtendo o IV 101, e os que tiverem tempo entre 21,92 e 21,95 seus IV serão de 102 pontos, e assim sucessivamente a cada 4 centésimos de segundo a menos. Para distâncias menores a variação do IV ocorrerá em uma menor variação do tempo, em 320m se alternará a cada 3 e 4 centésimos de segundo, iniciando com 3 centésimos de segundo, enquanto que aos 301m e 275m será de 3 centésimos de segundo e os 228m será a cada 2 centésimos de segundo. Essa tabela é válida para animais que correm carregando um peso mínimo de 53 Kg, e para aqueles com peso inferior, deve-se acrescentar 5 centésimos de segundo ao seu tempo, para cada quilo a menos, antes de consultar a tabela.

Tabela 1. Esquema da variação para pontuação do índice de velocidade (IV), de acordo com a distância (metros), tendo como ponto de partida os tempos referentes ao IV igual a 100

	Centésimos de segundo			
	4	3 e 4*	3	2
Distância (m)	365	320	275	228
	402		301	
	503			

*alternados, iniciando-se com 3 centésimos

Polimorfismos de nucleotídeos únicos (SNPs)

Marcador molecular é toda e qualquer variação oriunda de um gene expresso ou de um segmento específico de DNA (correspondente a regiões expressas ou não do genoma). 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 também definido como marcador genético (FERREIRA; GRATTAPAGLIA, 1998). Do ponto de vista molecular ocorrem três tipos principais de variações na molécula de DNA, as regiões repetitivas (minissatélites e microssatélites), as inserções e deleções (Indels) e as alterações de uma única base (polimorfismos de nucleotídeos únicos ou “single nucleotide polymorphism” - 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 pares de bases e são denominadas SNPs.

As substituições mais frequentes observadas no DNA envolvem bases nitrogenadas de mesma característica estrutural, ou seja, são trocas entre duas purinas (Adenina/Guanina ou G/A) ou duas pirimidinas (Citosina/Timina ou T/C) e são denominadas transições. As transversões são substituições de uma purina por uma pirimidina ou o contrário. Essas alterações podem ser provocadas por erros de incorporação de bases durante a replicação do DNA ou em outros casos, são causadas por agentes ambientais. Caso essas mutações ocorram em células germinativas, sejam transmitidas às gerações seguintes e se fixem 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 proteica, 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 proteica, eles podem modificar a estrutura e a estabilidade do RNA mensageiro, e, consequentemente afetar a quantidade de proteína produzida. Esta também pode ser afetada quando ocorrem alterações nas regiões não traduzidas do RNA mensageiro (5’ UTR e 3’ UTR). Além disso, polimorfismos gênicos podem promover 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; 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 introns (micro-RNAs e “long noncoding” RNAs) estão envolvidos em diferentes processos biológicos tais como os controles transcricional e pós-transcricional da expressão gênica (NAKAYA et al., 2007; MERCER; MATTICK, 2013).

Estudos em humanos, e em espécies de interesse zootécnico, mostraram a ocorrência de milhões de 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 que haja uma distribuição homogênea de SNPs pelo genoma (CAETANO, 2009).

Chips de genotipagem de SNPs de alta densidade

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, teoricamente, 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).

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 1.162.753 SNPs que ocorrem entre diversas raças (WADE et al., 2009). O genoma equino é composto por 31 cromossomos autossômicos e 1 cromossomo sexual com tamanho de aproximadamente 2,7 bilhões de pares de bases (WADE et al., 2009).

Desenhado 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. Este "array" contém 54.602 SNPs extraídos a partir da base de dados EquCab2.0 SNP "Collection" e apresenta densidade média, nos 31 cromossomos autossomos, de um SNP por 43,1 kb.

O "BeadChip" vêm servindo com enorme sucesso, principalmente, à identificação de regiões genômicas ou genes relacionados à importantes doenças que acometem determinadas raças equinas (COOK et al., 2009; EBERTH et al., 2009; BROOKS et al., 2010; KOMM, 2010) e com características de desempenho, como mostram os estudos de Binns et al., (2010) e Hill et al., (2010), os quais identificaram regiões cromossômicas que abrigam variantes associadas com distância ótima em corridas de Puro-Sangue Inglês.

Índice de fixação F_{ST}

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 o índice de fixação F_{ST} que fornece uma estimativa do quanto da variabilidade genética é dividida entre populações, ao invés de dentro de populações (WRIGHT, 1951; COCKERHAM, 1969). Esta estatística assume que variadas forças seletivas favorecem diferentes variantes gênicas em diversas regiões do genoma, sendo amplamente utilizada para detectar assinaturas de seleção entre populações ao comparar os seus valores entre os *loci* (WRIGHT, 1951; COCKERHAM, 1969; WEIR; HILL, 2002), determinando assim, se diferenças nas frequências alélicas

dos SNPs entre raças e/ou linhagens seriam decorrentes de seleção recente. Segundo Holsinger; Weir (2009), F_{ST} está diretamente relacionado com a variância na frequência alélica entre populações e, inversamente relacionado ao grau de semelhança entre os indivíduos dentro de populações. Dessa forma, se o valor de F_{ST} é pequeno, isto significa que as frequências alélicas entre populações (dentro de cada população) são similares, porém, se o valor é grande, significa que as frequências alélicas são diferentes.

Segundo Akey et al. (2002), o exame da variação nas frequências alélicas dos SNPs entre as populações, a qual pode ser quantificada pela estatística F_{ST}, é uma estratégia promissora para a detecção de assinaturas de seleção.

De acordo com Holsinger; Weir (2009) existem 3 diferentes abordagens amplamente utilizadas para estimar F_{ST}: método de momentos, método da máxima verossimilhança e métodos bayesianos.

O método de momentos é baseado na análise de variância (ANAVA) das frequências alélicas. ANAVA é um método estatístico que testa se a média de dois ou mais grupos são iguais podendo, assim, ser utilizado para avaliar o grau de diferenciação entre as populações. Pela ANAVA, calcula-se o quadrado médio esperado entre as populações (isto é, a variação das frequências alélicas da amostra em torno da frequência do alelo médio ao longo de toda a população) e o quadrado médio esperado dentro de populações (isto é, a heterozigosidade dentro das populações quando os genótipos estão em proporções do equilíbrio “Hardy-Weinberg”), calculados sobre todas as amostras. Estes valores esperados são então equiparados aos quadrados médios observados, calculados à partir de uma amostra, e o conjunto de equações é resolvido para os componentes de variância correspondentes.

Em contraste ao método anterior, os métodos da máxima verossimilhança e o bayesiano requerem a especificação da distribuição de probabilidade da qual a amostra foi retirada. Uma vez que esta distribuição for especificada, pode-se calcular a verossimilhança, que é obtida encontrando-se os valores dos parâmetros desconhecidos que maximizam a sua probabilidade. Na maioria dos casos, as estimativas de máxima verossimilhança são tendenciosas. No entanto, elas normalmente têm uma menor variância e desviam menos do parâmetro

populacional desconhecido do que as estimativas do método de momentos. O método bayesiano compartilha muita das vantagens associadas ao método anterior, pois usa a mesma probabilidade para relacionar os dados aos parâmetros desconhecidos. No entanto, se difere das estimativas de máxima verossimilhança, porque a probabilidade é modificada colocando-se distribuições a priori sobre os parâmetros desconhecidos, e as estimativas são baseadas na distribuição a posteriori, a qual é proporcional ao produto da probabilidade e às distribuições anteriores. As estimativas são obtidas através de métodos computacionais, como a cadeia de Markov e o método de Monte Carlo (métodos MCMC) obtidas até mesmo em modelos complexos com milhares de parâmetros, para o qual a maximização numérica da probabilidade seria difícil ou impossível.

Teste do haplótipo de longo alcance (LRH)

O teste do haplótipo de longo alcance ou “long-range haplotype test” (LRH), identifica regiões do genoma submetidas à seleção positiva por meio da análise da estrutura de haplótipos em indivíduos aleatórios de uma população, na relação entre frequências alélicas e na extensão do desequilíbrio de ligação (LD) em torno dela, utilizando a estatística Homozigose do Haplótipo Estendido (EHH). Esta análise se baseia na estimação do grau de homozigose entre regiões cromossômicas que carregam um haplótipo comum, sendo, portanto, definida 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. (SABETI et al., 2002). O EHH de um haplótipo núcleo testado (t) é matematicamente calculado como:

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

onde c_t é o número de amostras de um haplótipo núcleo particula, e_{ti} é o número de amostras de um haplótipo estendido e s é o número de haplótipos

estendidos.

Como taxas de recombinação podem variar consideravelmente ao longo de diferentes regiões do genoma, existe a possibilidade de que um maior valor obtido EHH possa ser devido às baixas taxas de recombinação em determinada região e não necessariamente à seleção positiva recente. Assim, a estatística homozigose relativa do haplótipo estendido (REHH), corrige o valor EHH observado em um haplótipo núcleo para o nível médio dos valores EHH para todos os haplótipos núcleos relevantes do mesmo cromossomo, corrigindo, assim, as variações nas taxas de recombinação.

Para calcular o REHH, primeiro calcula-se o $\overline{EHH}$, que é definido como o declínio do EHH em todos os outros haplótipos núcleos combinados e, *n* é número de diferentes haplótipos núcleos:

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

Em seguida, o REHH é definido por:

$$EHH_t/\overline{EHH}$$

Estudo de Associação Ampla do Genoma (GWAS)

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; GIBBS et al., 2009; KIJAS et al., 2009), estimar o grau de diversidade dentro e divergência genética entre populações (ZENGER et al., 2006), determinar a relação entre a perda de alelos e o aumento da endogamia devido à seleção (MUIR et al., 2008), e, 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).

Segundo Sahana et al. (2010) o principal propósito de um estudo de associação ampla do genoma ou “genome-wide associations studies” (GWAS), é

identificar regiões do cromossomo que abrigam genes que contribuem para a variação fenotípica de uma característica, servindo, posteriormente, como regiões putativas de QTL para mais estudos. Os GWAS baseiam-se no princípio do desequilíbrio de ligação (LD) no nível da população, o qual é a associação não aleatória entre alelos em *loci* diferentes. Ele é criado por forças evolutivas tais como mutação, deriva e seleção e é dividido por recombinação (VISSCHER et al., 2012).

Os estudos de associação têm emergido como uma poderosa ferramenta para revelar as bases genéticas de doenças e de características quantitativas (COLLINS et al., 1998).

Os GWAS, utilizando 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. Lee et al. (2013) utilizaram painel de aproximadamente 50.000 SNPs para identificar “major” *loci* significativamente associados com o peso de carcaça, e os seus efeitos, a fim de proporcionar uma maior compreensão da arquitetura genética da característica em bovinos Hanwoo. Os autores identificaram uma grande região do cromossomo bovino 14 variando de 23 Mb para 25 Mb associado com o peso de carcaça em Hanwoo, demonstrando que a característica foi afetada por um grande QTL de efeito maior e por muitos SNPs de pequenos efeitos.

Apesar da maior dificuldade de aplicação de ferramentas moleculares em estudos visando a seleção e o melhoramento genético de equinos em função, entre outros, da disponibilidade de animais, pesquisas ao redor do mundo vem analisando grande quantidade de genes e buscando a identificação de marcadores potencialmente associados às características de interesse em cavalos tais como: característica de desempenho em corridas de Puro-sangue Inglês (BINNS et al., 2010; HILL et al., 2010; TOZAKI et al., 2010), características morfométricas (SIGNER-HASLER et al., 2012; TETENS et al., 2013), cor da pelagem (REISSMANN et al., 2007); doenças (TRYON et al., 2007; YOUNG et al., 2007); resistência a doenças (BROWN et al., 2006; RIOS et al., 2007); reprodução e fertilidade (HAMMANN et al., 2007; GIESECKE et al., 2009); e comportamento e temperamento (MOMOZAWA et al., 2006).

Referências

ABQM. **Associação brasileira dos criadores de cavalos Quarto de Milha**. Disponível em: < http://www.abqm.com.br/item1.htm>. Acesso em: 15 jan. 2014

ALMEIDA, F. Q.; SILVA, V. P. Progresso científico em equideocultura na 1a década do século XXI. **Revista Brasileira de Zootecnia**, v. 39, p. 119-129, 2010.

AKEY, J. M.; ZHANG, G.; ZHANG, K.; JIN, L.; SHRIVER, M. D. Interrogating a high density SNP map for signatures of natural selection. **Genome Research**, v. 12, p. 1805-1814, 2002.

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**, v. 10:178, 2009. Disponível em: <http://www.biomedcentral.com/1471-2164/10/178>.

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

BROOKS, S. A.; GABRESKI, N.; MILLER, D.; 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**, v. 6, p. 1-7, 2010.

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

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

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

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

COLLINS, F. S.; BROOKS, L. D.; CHAKRAVARTI, A. A DNA Polymorphism Discovery Resource for Research on Human Genetic Variation. **Genome Research**, v. 8, p. 1229–1231, 1998.

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**, v. 29, n. 5, p. 315-316, 2009.

EBERTH, J.; SWERCZAK, T.; BAILEY, E. Investigation of Dwarfism Among Miniature Horses using the Illumina Horse SNP50 Bead Chip. **Journal of Equine Veterinary Science**, v. 29, n. 5, p. 315, 2009.

EVANS, J. W. **Horses: a guide to selection, care and enjoyment**. Freeman and Company: New York, 1996. p. 455-457, 472-474.

FERREIRA, M. E.; GRATTAPALIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. Embrapa-Cenargen: Brasília, 1998. p. 15-67.

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

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

GLIPHA. **Global Livestock Production and Health Atlas**. Disponível em: <http://www.fao.org/ag/againfo/home/en/news_archive/AGA_in_action/glipha.html>. Acesso em: 15 abr. 2014

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

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

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

HAWKEN, R. J.; ZHANG, Y. D.; FORTES, M. R. S.; COLLIS, E.; BARRIS, W. C.; CORBET, N. J.; WILLIANS, P. J.; FORDYCE, G.; HOLROYD, J. R.; WALKLEY, W.; BARENDSE, W.; JOHNSTON, D. J.; PRAYAGA, K. C.; TIER, B.; REVERTER, A.; LEHNERT, S. A. Genome-wide association studies of female reproduction in tropically adapted beef cattle. **Journal of Animal Science**, v. 90, p. 1398-1410, 2012.

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

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

JCS. **Jockey Club de Sorocaba**. Disponível em: <http://www.jockeyclubdesorocaba.com.br/guia.php>. Acesso em: 20 mar. 2014.

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

KOMM, K. **Genome wide association analysis for osteochondrosis in Hanoverian warmblood horses using an equine SNP assay. In: Fine mapping of quantitative trait loci (QTL) for osteochondrosis in Hanoverian warmblood horses**. 2010. 188f. Tese (Doutorado em Medicina Veterinária) - University of Veterinary Medicine Hannover, Hannover, 2010.

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

LEE, S. H.; CHOI, B. H.; LIM, D.; GONDRO, C.; CHO, Y. M.; DANG, C. G.; SHARMA, A.; JANG, G. W.; LEE, K. T.; YOON, D.; LEE, H. K.; HEUM, S.; YANG, B. S.; KANG, H. S.; HONG, S. K. Genome-wide association study identifies major loci for carcass weight on BTA14 in Hanwoo (Korean cattle). **PloS One**, v. 8, n. 10, e74677, 2013.

LIMA, R. A. S.; SHIROTA, R.; BARROS, G. S. C. **Estudo do Complexo do Agronegócio Cavalo no Brasil**. Piracicaba: ESALQ/USP: CEPEA, 2006. 250p.

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

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

MERCER, T. R.; MATTICK, J. S. Structure and function of long noncoding RNAs in epigenetic regulation. **Nature Structural & Molecular Biology**, v. 20, p. 300-307, 2013.

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

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. **Journal of Equine Veterinary Science**, v. 13, p. 244-249, 2013.

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. **Journal of Veterinary Medical Science**, v. 68, p. 619-621, 2006.

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

NAKAYA, H. L.; AMARAL, P. P.; LOURO, R.; LOPES, A.; FACHEL, A. A.; MOREIRA, Y. B.; EL-JUNDI, T. A.; SILVA, A. M.; REIS, E. M.; VERJOVSKI-ALMEIDA, S. 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, n. 3, 2007.

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

REISSMANN, M.; BIERWOLF, J.; BROCKMANN, G. A. Two SNPs in the SILV gene are associated in silver coat colour in ponies. **Animal Genetics**, v. 38, 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**, v. 8, n. 313, 2007.

SABETI, P. C.; REICH, D. E.; HIGGINS, J. M.; LEVINE, H. Z. P.; 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.; COOPERK, R.; KWIATKOWSKI, D.; WARD, R.; LANDER, E. S. Detecting recent positive selection in the human genome from haplotype structure. **Nature**, v. 419, p. 832–837, 2002.

SAHANA, G.; GULDBRANDTSEN, B.; BENDIXEN, C.; LUND, M. S. Genome-wide association mapping for female fertility traits in Danish and Swedish Holstein cattle. **Animal Genetics**, v. 41, p. 579–588, 2010.

SCOTT, M. Musculoskeletal Injuries in Nonracing Quarter Horses. Veterinary Clinics. **Equine Practice**, v. 24, p. 133-152, 2008.

SIGNER-HASLER, H.; FLURY, C.; HAASE, B.; BURGER, D.; SIMIANER, H.; LEEB, T.; RIEDER, S. A Genome-wide association study reveals loci influencing height and other conformation traits in horses. **PloS One**, v. 7, n. 5, e37282, 2012.

SILVER, C. **Tudo sobre cavalos: um guia mundial de 200 raças**. São Paulo: Martins Fontes, 2000. 231p.

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. Disponível em: < http://www.wcgalp2010.org/topics.php>. Acesso em: 25 nov. 2013.

TETENS, J.; WIDMANN, P.; KÜHN, C.; THALLER, G. A genome-wide association study indicates LCORL/NCAPG as a candidate locus for withers height in German Warmblood horses. **Animal Genetics**, v. 44, p. 467–471, 2013.

TOZAKI, T.; MIYAKE, T.; KAKOI, H.; GAWAHARA, H.; SUGITA, S.; HASEGAWA, T.; ISHIDA, N.; HIROTA, K.; NAKANO, Y. A genome-wide association study for racing performances in Thoroughbreds clarifies a candidate region near the MSTN gene. **Animal Genetics**, v. 41, supl. 2, p. 28–35, 2010.

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**, v. 90, p. 93-102, 2007.

VISSCHER, P. M.; BROWN, M. A.; McCARTHY, M. I.; YANG, J. Five Years of GWAS Discovery. **The American Journal of Human Genetics**, v. 90, p. 7–24, 2012.

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

WARMUTH, V.; ERIKSSON, A.; BOWER, M. A.; BARKER, G.; BARRETT, E.; HANKS, B. K.; LI, S.; LOMITASHVILI, D.; OCHIR-GORYAEVA, M.; SIZONOV, G. V.; SOYONOV, V.; MANICA, A. Reconstructing the origin and spread of horse domestication in the Eurasian steppe. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, p. 8202–8206, 2012.

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

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

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

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

CAPÍTULO 2 – Morphological and genomic differences between cutting and racing lines of Quarter Horse

Camila Tangari Meira[a], Rogério Abdallah Curi[b], Josineudson Augusto II Vasconcelos Silva[b], Marcio José Monteiro Corrêa[b], Henrique Nunes de Oliveira[a], Marcilio Dias Silveira da Mota[b]

[a] Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal, SP 14884-900, Brazil.
[b] Faculdade de Medicina Veterinária e Zootecnia, Universidade Estadual Paulista, Botucatu, SP 18618-000, Brazil.

Abstract

To investigate morphological and genomic differences between cutting and racing lines of Quarter Horses, 120 racing and 68 cutting animals of both sexes, registered at the Brazilian Association of Quarter Horse Breeders, were used. Blood samples were collected, and the following physical traits were measured: weight; height at withers; body length; length of the shank, pastern, rump, head, and neck; and chest, shank, and hoof circumference. For analysis of genomic differences, 54,602 single-nucleotide polymorphisms (SNPs) were genotyped using the Equine SNP50 BeadChip, and the quality of individual and SNP genotype data were evaluated. The fixation index, F_{ST}, was used to identify genome regions that were altered in the lines by selection. The results showed significant differences between the lines in all physical traits. Quality control led to the exclusion of four cutting animals with a call rate of <0.95. After filtering, 12,544, 13,815, and 13,370 SNPs were excluded for the whole population (n = 184), the 120 racing animals, and the 64 cutting animals, respectively. The number of informative polymorphisms detected in each line and in the whole population indicated that the Equine SNP50 BeadChip can be used in genetic studies of Quarter Horses. The fixation index, F_{ST}, identified 2,558 genome regions that may have been modified by divergent selection.

Keywords: physical traits, horse, genome-wide association study, SNPs

1. Introduction

As a breed of global importance, corresponding to 52% of all horses, Quarter Horses are important because of their great versatility in different equestrian events [1]. Quarter Horses were developed in North America in the 17th century from Arabian and Turkish horses brought by European settlers. The major development of this breed occurred during westward expansion when pioneers needed robust and versatile horses, fit for the saddle and for traction, in view of the difficulty to keep a varied stock of animals to satisfy diverse necessities [1].

The Quarter Horse breed is subdivided into different lines according to skills resulting from distinct selection objectives, including cutting and racing horses. The cutting line is destined for functional tests, exploring skills such as agility and obedience which are important for cattle management in the field. The racing line is characterized by great sprinting speed over short distances on straight tracks. The cutting type is shorter and more compact and has muscular hindquarters, whereas the racing type is taller, has longer legs and a less prominent musculature.

The simultaneous study of thousands of DNA polymorphisms spread across the genome, known as genome-wide association analysis, has permitted the study of different populations of various domestic animal species [2-4], as well as the estimation of genetic divergence within and between populations [5]. During the process of domestication and breed formation, domestic animals were subjected to natural and artificial selection. These selection pressures led to an increase in the frequency of some mutations in specific regions of the genome, which generated more adapted animals or provided them with favorable characteristics to meet human needs. At the same time, the frequency of other polymorphisms decreased or they were completely eliminated [6]. In this respect, the comparison of allele frequencies between selected and unselected populations or between populations selected for different objectives provides insights into the regions of the genome that have been modified by selection.

According to Chowdhary and Raudsepp [7] one of the highlights from the analysis of the horse genome are its complete sequencing from a Thoroughbred animal (EquCab2.0) and, from this, the identification of 1,162,753 single-nucleotide

polymorphisms (SNPs) in different breeds [8]. Designed to identify SNPs and genes that contribute to traits of interest in the major horse breeds raised today in the world, the Equine SNP50 BeadChip developed by Illumina, Inc. (San Diego, California, USA), represents a powerful platform for genetic studies of this species, permitting researchers to perform a variety of experiments that require the genotyping of DNA polymorphisms.

In view of these considerations, the aims of the present study were to investigate the genomic differences, using the Equine SNP50 BeadChip, and morphological differences between cutting and racing lines of Quarter Horses as a result of selection for different objectives.

2. Materials and Methods

2.1. Animals and Phenotypic Data

One hundred eighty-eight Quarter Horses of both sexes born between 1985 and 2009, including 120 racing horses and 68 cutting horses, registered at the Brazilian Association of Quarter Horse Breeders, were studied. All experimental procedures were conducted in accordance with the Brazilian legislation on animal welfare.

The following physical traits were measured according to Torres and Jardim [9]: weight, height at withers, body length, length of the shank, pastern, rump, head and neck, and chest, shank and hoof circumference. The measurements were performed by the same person with a tape measure and measuring stick, always on the right side of the animal, with the horse standing with front and rear legs perpendicular to the ground. For genotyping, a 5-mL sample of whole blood was collected from each animal by puncture of the left jugular vein in the neck region into vacuum tubes containing 7.5 mg ethylenediaminetetraacetic acid.

The animals of the racing line (18 male and 102 female horses), born to 48 stallions and 107 mares, belonged to five farms in the countryside of the State of São Paulo, Brazil. The animals of the cutting line (26 male and 42 female horses), born to 44 stallions and 64 mares, belonged to three other farms in the countryside of the

State of São Paulo. In both lines, full-sibs were avoided.

2.2. Genotyping of SNPs

Genomic DNA was extracted from the blood samples of Quarter Horses using the Illustra Blood Genomicprep Mini Spin kit (GE Healthcare, Little Chalfont, Buckinghamshire, UK) according to manufacturer instructions. DNA integrity was analyzed by 0.8% agarose gel electrophoresis, and DNA was quantified with a NanoDrop ND-2000 spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts,USA). The DNA concentration in the samples was adjusted to 40-60 ng/μL.

SNPs were genotyped on the HiScan system (Illumina Inc.) using the Illumina Equine SNP50 BeadChip at the Faculty of Agricultural and Veterinary Sciences, Unesp, Jaboticabal, São Paulo, Brazil. The chip contains 54,602 SNPs uniformly distributed across the entire genome of 15 horse breeds. The SNPs are distributed across the 31 autosomes and X chromosome. The mean interval between SNPs is 43,200 bp. This content is derived from the EquCab2.0 SNP Collection compiled by the Broad Institute's Equine Genome Sequencing Project, which identified > 940,000 SNPs in Arabian, Andaluz, Akhal-Teke, Islandesa, Standardbred, Thoroughbred, and Quarter Horses.

2.3. Analysis of Morphological Differences and Differences in Inbreeding between Cutting and Racing Lines

Morphological differences between the two Quarter Horse lines (cutting and racing) were evaluated using a model that included the effects of sex and line, and age at recording as covariate. The general linear model (GLM) procedure of the Statistical Analysis System v.9.1 program (SAS Institute Inc., Cary, North Carolina, USA) [10] was used for statistical analysis, and means were compared by the Tukey test at a level of significance of 5%.

The inbreeding coefficient was calculated for each animal of the two lines based on pedigree records using the Relax2 program (MTT Agrifood Research

Finland, Biometrical Genetics, Jokioinen, Finland) [11]. The relationship matrix contained 762 animals, with a depth of ancestry of four generations. The average inbreeding of consanguineous animals and of all animals of each line was estimated using the coefficients of inbreeding.

2.4. Quality of Genotype Data

The quality of individual and SNP genotype data was investigated using the Genome Studio program, version 2011.1 (Illumina Inc). For individuals, call rate, heterozygosity and gender estimation were determined. Animals with a call rate < 0.95, heterozygosity of ± 3 standard deviations from the mean, and errors in gender estimation were excluded from the sample. In addition, agreement between four replicates and parentage concordance (allele sharing) between four stallion/progeny and three stallion/mare/progeny pairs were evaluated.

With respect to the quality of SNP genotypes in the whole population and in each line, SNPs located on the X chromosome were excluded (filtered). SNPs with low genotyping quality (cluster separation <0.3), a call frequency <0.9, a minor allele frequency (MAF) <0.05 (including fixed alleles), and a Hardy-Weinberg p-value <.001 were also excluded.

2.5. Analysis of Genomic Differences between Cutting and Racing Lines

The fixation index, F_{ST}, was used to identify genome regions in the cutting and racing lines that have been modified by selection [12-14]. The $\theta=F_{ST}$ parameter was estimated for all SNPs that passed quality control (considering all individuals) using the Bayesian method proposed by Gianola et al. [15]. This method is a two-step procedure that eventually leads to clusters of θ values. The first step uses a simple Bayesian structure for removing samples from the posterior distribution of θ-parameters without constructing Markov chains. This step assigns a weakly informative prior of allele frequencies and does not make any assumptions about evolutionary models. The second step considers samples from this posterior distribution as “data” and fits a sequence of finite mixture models in order to identify

clusters of θ-statistics.

3. Results and Discussion

Selecting Quarter Horses for different objectives (racing or cutting) promoted significant changes in the physical traits of the animals. Table 1 compares the fitted means of the characteristics analyzed between the two lines. A significant effect of line ($P < .0001$) was observed for all traits. As expected, racing animals presented higher weight, height, body length and body circumference than cutting animals. Thus, the phenotypic differences observed between the two lines have a possible genetic background. However, it is necessary to considerer that phenotypes are also influenced by several environmental factors, such as feeding, training and other. Although there are no other morphological studies characterizing Quarter Horse lines, the present investigation is important because it demonstrated a close relationship between physical characteristics and the function for which the animal is used. The different abilities of the breed, such as running, jumping, reining and barrel racing, have resulted in an appropriate biotype for each modality and differences in performance exist owing to morphological adequacy or inadequacy.

Among the 188 animals studied, 12 racing animals and one cutting animal were consanguineous. The mean coefficient of inbreeding (F) was 0.0028 or 0.28% for the whole population. In contrast, the mean coefficient of inbreeding was 0.0364 or 3.64% (0%-7.81%) for the 12 inbred animals of the racing line and 0.0036 (or 0.36%) for all racing animals. These coefficients were higher than those obtained for the only inbred animal of the cutting line (0.0078 or 0.78%) and for all cutting animals (0.0001 or 0.01%), suggesting a higher frequency of inbreeding in the racing line. Coefficients of 0.8% [16] to 1% [17] have been reported for Thoroughbreds, and of 10.3% for a population of Lipizzan mares [18]. In this research, as well as in the study by McCue et al. [19], the coefficients of inbreeding of Quarter Horses were found to be low when compared with other breeds, indicating that Quarter Horse breeders tend to avoid mating between closely related animals.

Quality control of individual genotype data led to the exclusion of four cutting animals from the sample because of a call rate <0.95 (95%). The remaining 184

animals (120 racing and 64 cutting animals) presented a mean call rate of 0.9929 ± 0.0054 (range: 0.9503 to 0.9979). Mean heterozygosity estimated for all individuals was 0.3468 ± 0.0148 (range: 0.3049 to 0.3788). The samples genotyped in duplicate showed agreement ≥99.8% (0.9979 to 0.9990). Similarly, parentage concordance between stallion/progeny and stallion/mare/progeny pairs was very high (0.9985 to 0.9995). Although some individuals were excluded because of a low genotyping rate, taken together, the present results indicate the lack of DNA contamination between samples and a generally successful hybridization between DNA and the chip.

Considering the whole population (n = 184) that passed quality control, 12,544 SNPs were excluded by the filtering process and 42,058 remained for further analysis. With respect to the 120 racing and 64 cutting animals, after filtering, 40,787 (13,815 excluded) and 41,232 (13,370 excluded) SNPs remained for analysis, respectively. Table 2 shows the number of excluded SNPs and the reason for exclusion for the whole population and for animals of the racing and cutting lines.

According to Wiggans et al. [20] and Ziegler et al. [21], inconsistently genotyped SNPs or those that do not contribute to the accuracy of genetic evaluations should be excluded in order to reduce computational effort and the number of false results, as well as to improve precision of the estimates of the remaining polymorphisms. The numbers of excluded SNPs owing to an MAF <0.05 were similar in the racing line (8,925), cutting line (7,958), and whole population (7,587). Schröder et al. [22] reported a similar number in a study of Quantitative Trait Loci (QTL) for show-jumping in Hanoverian Warmblood horses (7,875 excluded SNPs). In contrast, Lykkjen et al. [23] excluded 13,265 SNPs with an MAF <0.05 in a study identifying QTLs associated with osteochondrosis in Norwegian Standardbred trotters. In the present study, the number of informative SNPs in the two lines (racing animals: 40,787; cutting animals: 41,232) and in the whole population (42,058) were lower than the 47,699 SNPs found in Quarter Horses by McCue et al. [19]. This divergence of results is probably due to different criteria adopted in the filtering process of the genotyped SNPs used in their research, particularly the exclusion of polymorphisms with MAF <0.01. In contrast, the present results were similar to those reported in studies that investigated other horse breeds using the Equine SNP50 BeadChip, such as Thoroughbreds (40,977) [24], Standardbreds (41,170) [23],

Hanoverian horses (43,441) [22], and French trotters (41,249) [25]. Despite the difference in the number of informative SNPs between the racing and cutting lines, autosomal coverage, although higher in the latter, did not vary considerably.

Comparison of the mean MAF of the genotyped SNPs revealed no significant difference between the racing (0.2267 ± 0.1543) and cutting lines (0.2256 ± 0.1496) (*F* test, P >.05). Similar values have been reported by Corbin et al. [26] and McCue et al. [19] who studied linkage disequilibrium (LD) in Thoroughbred (0.30 ± 0.12) and Quarter horses (0.232), respectively. However, the frequency of polymorphisms with an MAF <0.05 was higher in racing horses (17.87%) than in cutting horses (15.96%). Among these, the frequency of fixed (monomorphic) polymorphisms and those with rare alleles were 6.79% and 11.08% in racing animals and 5.11% and 10.85% in cutting animals, respectively. In addition, the number of informative SNPs (MAF ≥0.05) was higher in the cutting line (84.04%) than in the racing line (82.13%). Table 3 shows the number of SNPs according to MAF class. The higher frequency of polymorphisms with an MAF <0.05 in racing Quarter Horses might be explained by the more common practice of consanguineous mating and higher selection intensity in this line. Polymorphic SNPs that are associated with one population but not with the other, which were more frequent in the cutting line, might be used as informative SNPs for a line in particular. In contrast, highly informative SNPs (MAF = 0.4-0.5), which were more frequent in the racing line, are useful for individual identification by DNA testing (paternity and forensic tests).

Table 4 shows the distribution and density of informative SNPs (MAF ≥0.05) across the 31 autosomes in the sample of Quarter Horses studied. Considering the whole population, the largest number of informative SNPs was detected on chromosome 1 (3,502), and the smallest number, on chromosome 31 (492). With respect to genome coverage, the density of SNPs was 18.72 SNPs/million base pairs (Mb) or one SNP at an interval of 53,419 bp. This density varied between chromosomes, ranging from 16.82 SNPs/Mb (ECA12: average of one SNP at an interval of 59,453 bp) to 19.91 SNPs/Mb (ECA16: average of SNP at an interval of 52,356 bp). In the racing line, chromosome 1 contained the largest (3,381) and chromosome 31 contained the smallest number of informative SNPs (486). The density was 18.15 SNPs/Mb or one SNP at an interval of 55,096 bp, ranging from

16.09 SNPs/Mb (ECA 12: average of one SNP at an interval of 62,150 bp) to 19.44 SNPs/Mb (ECA31: average of one SNP at an interval of 51,440 bp). In the cutting line, ECA1 contained the largest number of informative SNPs (3,433) and ECA31 the smallest number (496). The lowest SNP density was observed for chromosome 12, with 16.67 SNPs/Mb (average of one SNP at an interval of 59,988 bp) and the highest density for chromosome 31, with 19.84 SNPs/Mb (average of one SNP at an interval of 50,403 bp), corresponding to a mean interval of 54,496 bp between SNPs distributed across the entire autosomal genome or to a density of 18.35 SNPs/Mb.

Equine chromosomes differ significantly in length, with the longest chromosome (ECA1) containing 186 Mb and the shortest chromosome (ECA31) containing 25 Mb [27]. Therefore, the higher frequency of informative polymorphisms on larger chromosomes and the lower frequency on smaller chromosomes observed in the two lines and in the whole Quarter Horse population indicate that the filtering process did not affect the uniformity of distribution of SNPs across the genome.

The F_{ST} mean estimated for the 42,058 SNPs in the two lines was 0.0342 ± 0.0403, with θ ranging from 0.0025 to 0.3556. This value was substantially lower than that reported by Gu et al. [28], who used microsatellite markers to identify genome regions that distinguish Thoroughbreds from non-Thoroughbred breeds and found an overall mean F_{ST} of 0.12. The low F_{ST} value found in the present study was expected, as fixation indices tend to be lower when lines instead of breeds are studied because of a theoretically lower genetic distance. Qanbari et al. [29], using F_{ST} to quantify genetic differentiation in *Bos taurus* cattle, found lower values within milk-producing (Holstein and Brown Swiss; 0.057 ± 0.076) and meat-producing breeds (Angus and Piedmontese; 0.022 ± 0.041) than between these breeds (mean of 0.278 ± 0.324).

The distribution of posterior means of the θ values (F_{ST}) in the two lines permitted classification of the 42,058 SNPs (loci) into seven clusters. According to Gianola et al. [15], the expectation is that these clusters are representative of different processes that occur in the populations, such as balancing, directional selection and neutrality. Cluster 4, which consisted of 2,558 loci, presented the highest values of θ. The conditional probability of membership to this cluster was 1 for 271 loci. Fifteen of these loci, mapped to chromosomes 1, 2, 4, 5, 8, 9, 10, 16, 21 and 30, presented the highest values of θ (0.3013 to 0.3556), indicating genome

regions that were more likely to have been subjected to divergent selection between lines.

Recently, genomic studies in the Quarter Horse got prominent. In the first sequencing of a horse genome by next-generation sequencing and the first genomic sequence of an individual Quarter Horse mare, Doan et al. [30] identified 3.1 milion SNPs, 193,000 insertion/deletion polymorphisms (InDels), and 282 copy number variants (CNV) in relation to the reference Thoroughbred genome. Additionally, these researches genotyped this Quarter Horse for gene mutations of known diseases and for gene variants associated with economically important traits, including racing performance. The four genes related to racing performance in Thoroughbred horses, *MSTN* (myostatin) [24,31-33], *PDK4* (piruvate dehydrogenase kinase, isozyme 4) [28,34], *CKM* (creatine kinase, muscle) and *COX4I2* (cytochrome c oxidase, subunit 4, isoform 2) [35], are located in equine chromosomes 18, 4, 10 e 22, respectively. These chromosomes are shown in this research as having 77 of 271 genomic regions (with conditional probability equal to one) selected divergently between racing and cutting lines. However, considering that the LD decrease in the Quarter Horse occurs between loci that are located at distances longer than 50-100 kb [19], none of the 77 regions close enough (in LD) to the gene polymorphisms associated with racing performance. Considering that speed is also a desirable trait in cutting animals, these results suggest the possibility that the same alleles of these gene polymorphisms have been selected in both lines of Quarter Horse.

4. Conclusions

Divergent selection of cutting and racing Quarter Horses promoted significant changes in the physical and genomic characteristics of these lines. The number of informative SNPs and SNP density found in the genome of cutting and racing Quarter Horses suggest that the Equine SNP50 BeadChip can be used for different purposes in the breed, such as genetic structure analysis, estimation of genetic divergence within and between populations, and identification of QTL and genome regions subjected to selection.

Acknowledgements

The authors thank the following horse farms for providing the Quarter Horses: Haras Fazenda Bela, Capela do Alto/SP; Haras Vista Verde, Boituva/SP; Central de Reprodução Rancho das Américas, Porto Feliz/SP; Haras Carrera, Avaré/SP; Haras São Jorge, Tatuí/SP; Xandi Ranch, Bauru/SP; Haras Villa San José, Porto Feliz/SP, and Haras Rancho Emanuel, Porto Feliz/SP.

The present research was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo and fellowships were granted by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

References

[1] ABQM. Associação brasileira dos criadores de cavalos Quarto de Milha. Available at: http://www.abqm.com.br. Accessed December 12, 2011.

[2] Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, et al. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. PLoS One 2009;4:1-13.

[3] McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppieters W, et al. An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. BMC Genet 2008;9:1-9.

[4] Gibbs RA, Taylor JF, Van Tassell CP. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 2009;324:528-32.

[5] Zenger KR, Khatkar MS, Cavanagh JAL, Hawken RJ, Raadsma HW. Genome-wide genetic diversity of Holstein Friesian cattle reveals new insights into Australian and global population variability, including impact of selection. Anim Genet 2006;38:7-14.

[6] Andersson L, Georges M. Domestic-animal genomics: Deciphering the genetics of complex traits. Nat Rev Genet 2004;5:202-12.

[7] Chowdhary BP, Raudsepp T. The Horse Derby: racing from map to whole genome sequence. Chrom Res 2008;16:109-27.

[8] Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, et al. Genome sequence, comparative analysis, and population genetics of the domestic horse. Science 2009;326:865-67.

[9] Torres AP, Jardim WR. Criação do cavalo e de outros eqüinos. São Paulo: Livraria Nobel; 1992.

[10] SAS. Statistical Analysis System, 2004 SAS/STAT User's Guide, Version 9.1, Volumes 1-7. SAS Institute Inc., Cary, NC, USA.

[11] Strandén I, Vuori K. Relax2: pedigree analysis program. Proceedings of the 8th World Congress on Genetics Applied to Livestock Production; 2006 Aug 13-18; Belo Horizonte, MG. Brasil; 2006.

[12] Wright S. The genetical structure of populations. Ann Eugenics 1951;15:323-54.

[13] Cockerham CC. Variance of gene frequencies. Evolution 1969;23:72-84.

[14] Weir BS, Hill WG. Estimating F-statistics. Annu Rev Genet 2002;36:721–50.

[15] Gianola D, Simianer H, Qanbari S. A Two-step method for detecting selection signatures using genetic markers. Genet Res 2010;92:141-55.

[16] Abrahão AR, Mota MDS, Oliveira HN, Madureira AP. Endogamia em éguas da raça Puro-Sangue Inglês. Anais IV Simpósio Nacional de Melhoramento Animal; 2002; Campo Grande, Brasil.

[17] Taveira RZ, Mota MDS, Oliveira HN, Abrahão AR, Madureira AP. Endogamia em cavalos da raça Puro-Sangue Inglês. 47° Congresso Nacional de Genética. A genética no século XXI 2001.

[18] Curik I, Zechner P, Sölkner J, Achmann R, Bodo I, Dovc P, et al. Inbreeding, microsatellite heterozygosity, and morphological traits in Lipizzan horses. J Hered 2003;94:125–32.

[19] McCue ME, Bannasch DL, Petersen JL, Gurr J, Bailey E, Binns MM, et al. A high density SNP array for the domestic horse and extant perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. PLoS Genet 2012;8:1-14.

[20] Wiggans GR, Sonstegard TS, VanRaden PM, Matukumalli LK, Schnabel RD, Taylor JF, et al. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. J Dairy Sci 2009;92:3431-36.

[21] Ziegler A. Genome-Wide Association Studies: Quality Control and Population-Based Measures. Genet Epidemiol 2009;33(Suppl 1):S45–S50.

[22] Schröder W, Klostermann A, Stock KF, Distl O. A genome-wide association study for quantitative trait loci of show-jumping in Hanoverian Warmblood horses. Anim Genet 2011;1-9.

[23] Lykkjen S, Dolvik NI, McCue ME, Rendahl AK, Mickelson JR, Roed KH. Genome-wide association analysis of osteochondrosis of the tibiotarsal joint in Norwegian Standardbred trotters. Anim Genet 2010;41:111–20.

[24] Hill EW, McGivney BA, Gu J, Whiston R, MacHugh DE. 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 2010b;11:1-10.

[25] Teyssèdre S, Dupuis MC, Guérin G, Schibler L, Denoix JM, Elsen JM, et al. Genome-wide association studies for osteochondrosis in French Trotters. J Anim Sci 2012;90:45-53.

[26] Corbin LJ, Blott SC, Swinburne JE, Vaudin M, Bishop SC, Woolliams JA. Linkage disequilibrium and historical effective population size in the Thoroughbred horse. Anim Genet 2010;41:8–15.

[27] NCBI. National Center for Biotechnology Information. Available at: www.ncbi.nlm.nih.gov. Accessed November 7, 2011.

[28] Gu J, Orr N, Park SD, Katz LM, Sulimova G, MacHugh DE, et al. A genome scan for positive selection in Thoroughbred horses. PLoS One 2009;4:1-17.

[29] Qanbari S, Gianola D, Hayes B, Schenkel F, Miller S, Moore S, et al. Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. BMC Genomics 2011;12:1-12.

[30] Doan R, Cohen ND, Sawyer J, Ghaffari N, Johnson CD, Dindot SV. Whole-genome sequencing and genetic variant analysis of a Quarter Horse mare. BMC Genomics 2012;13:1-11.

[31] Hill EM, Gu J, Eivers SS, Fonseca RG, McGiveny BA, Govindarajan P, et al. A Sequence Polymorphism in MSTN Predicts Sprinting Ability and Racing Stamina in Thoroughbred Horses. PLoS One 2010a;5:1-6.

[32] Binns MM, Boehler DA, Lambert DH. Identification of the myostatin locus (MSTN) as having a major effect on optimum racing distance in the Thoroughbred horse in the USA. Anim Genet 2010;41:154-58.

[33] Tozaki T, Miyake T, Kakoi H, Gawahara H, Sugita S, Hasegawa T, et al. A genome-wide association study for racing performances in Thoroughbreds clarifies a candidate region near the MSTN gene. Anim Genet 2010;41:28-35.

[34] Hill EM, Gu J, McGiveny BA, MacHugh DE. Targets of selection in the Thoroughbred genome contain exercise-relevant gene SNPs associated with elite racecourse performance. Anim Genet 2010c;41:56-63.

[35] Gu J, MacHugh DE, McGiveny BA, Park SDE, Katz LM, Hill EM. 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 Vet J 2010;42:569-75.

Table 1. Least square means and standard errors of physical traits in racing and cutting Quarter Horses

Traits	Racing line	Cutting line
Body length (m)	1.8081±0.0057[a]	1.6343±0.0087[b]
Shank length (cm)	27.065±0.0014[a]	24.478±0.0014[b]
Pastern length (cm)	13.382±0.0009[a]	12.231±0.0009[b]
Rump length (cm)	62.032±0.0032[a]	55.173±0.0028[b]
Head length (cm)	62.170±0.0020[a]	56.057±0.0021[b]
Neck length (cm)	79.772±0.0034[a]	73.333±0.0049[b]
Weight (kg)	538.97±3.7671[a]	450.69±5.6902[b]
Height at withers (m)	1.5592±0.0040[a]	1.4660±0.0045[b]
Chest circumference (m)	1.9669±0.0127[a]	1.8020±0.0094[b]
Shank circumference (cm)	20.521±0.0008[a]	18.812±0.0009[b]
Hoof circumference (cm)	43.013±0.0018[a]	39.751±0.0027[b]

[a,b] Means in the same row followed by different superscript letters differ significantly at the 5% level (Tukey test).

Table 2. Number of SNPs excluded by quality control of genotype data considering the whole population, racing line, and cutting line of Quarter Horses and reason for exclusion

Category/Reason for Exclusion	Whole Population (n)	Racing Line (n)	Cutting Line (n)
Genotyped SNPs	54,602	54,602	54,602
Located on the X chromosome	2,539	2,539	2,539
Cluster separation < 0.3	1,972	1,972	1,972
Call frequency < 0.9	148	143	227
Fixed or with minor allele frequency < 0.05	7,587	8,925	7,958
Hardy-Weinberg equilibrium - $P < .001$	298	236	674
SNPs selected for genetic analysis	42,058	40,787	41,232

Table 3. Number and percentage of SNPs divided into four classes of minor allele frequency obtained for the racing and cutting lines of Quarter Horses and for the whole population

	Minor allele frequency			
Line	<0.05	0.05 – 0.5	0.05 – 0.4	0.4 – 0.5
Racing	8,925 (17.87%)	41,023 (82.13%)	31,580 (63.22%)	9,443 (18.91%)
Cutting	7,958 (15.96%)	41,906 (84.04%)	33,084 (66.35%)	8,822 (17.69%)
Whole population	7,587 (15.19%)	42,356 (84.81%)	32,985 (66.05%)	9,371 (18.76%)

Table 4. Distribution and density of informative SNPs (minor allele frequency ≥0.05) across the autosomes of racing and cutting Quarter Horses and of the whole population

Chr. – size (Mb)	Informative SNPs – racing line	Informative SNPs – cutting line	Informative SNPs – whole population	Density – racing line (SNPs/Mb)	Density – cutting line (SNP/Mb)	Density – whole population (SNP/Mb)
1 - 186	3,381	3,433	3,502	18.18	18.46	18.83
2 - 121	2,238	2,197	2,289	18.50	18.16	18.92
3 - 119	2,213	2,213	2,261	18.60	18.60	19.00
4 - 109	1,980	2,053	2,059	18.17	18.83	18.89
5 - 100	1,773	1,794	1,826	17.73	17.94	18.26
6 - 85	1,474	1,451	1,510	17.34	17.07	17.76
7 - 99	1,792	1,782	1,834	18.10	18.00	18.53
8 - 94	1,720	1,722	1,767	18.30	18.32	18.80
9 - 84	1,529	1,555	1,578	18.20	18.51	18.79
10 - 84	1,478	1,484	1,527	17.60	17.67	18.18
11 - 61	1,102	1,128	1,134	18.07	18.49	18.59
12 - 33	531	550	555	16.09	16.67	16.82
13 - 43	739	754	776	17.19	17.53	18.05
14 - 94	1,703	1,777	1,764	18.12	18.90	18.77
15 - 92	1,759	1,703	1,767	19.12	18.51	19.21
16 - 87	1,675	1,676	1,732	19.25	19.26	19.91
17 - 81	1,492	1,511	1,546	18.42	18.65	19.09
18 - 83	1,405	1,462	1,455	16.93	17.61	17.53
19 - 60	1,151	1,144	1,166	19.18	19.07	19.43
20 - 64	1,184	1,208	1,225	18.50	18.88	19.14
21 - 58	1,044	1,052	1,087	18.00	18.14	18.74
22 - 50	904	969	953	18.08	19.38	19.06
23 - 56	975	986	1,016	17.41	17.61	18.14
24 - 47	885	882	908	18.83	18.77	19.32
25 - 40	703	740	738	17.58	18.50	18.45
26 - 42	765	765	792	18.21	18.21	18.86
27 - 40	676	719	721	16.90	17.98	18.03
28 - 46	860	847	885	18.70	18.41	19.24
29 - 34	599	616	614	17.62	18.12	18.06
30 - 30	571	563	579	19.03	18.77	19.30
31 - 25	486	496	492	19.44	19.84	19.68

CAPÍTULO 3 – Prospection of genomic regions divergently selected in racing line of Quarter Horse in relation to cutting line

C.T. Meira[1], R.A. Curi[2], M.M. Farah[1], H.N. de Oliveira[1], N.A.R. Béltran[1], J.A.V. Silva[2], M.D.S. da Mota[2]

[1] *Department of Animal Science, College of Agriculture and Veterinary Science, São Paulo State University (UNESP), access route Paulo Donato Castellane, 14884-900, Jaboticabal, São Paulo, Brazil.*

[2] *Department of Animal Breeding and Nutrition, College of Veterinary and Animal Science, São Paulo State University (UNESP), Rubião Junior District, 18618-970, Botucatu, São Paulo, Brazil.*

Abstract

Selection of Quarter Horses for different purposes has led to the formation of lines, including racing and cutting horses. The objective of this study was to identify genomic regions divergently selected in racing line of Quarter Horses in relation to cutting line applying relative extended haplotype homozygosity (REHH) analysis, an extension of EHH analysis, and the fixation index (F_{ST}) statistic. A total of 188 horses of both sexes, born between 1985 and 2009 and registered at the Brazilian Association of Quarter Horse Breeders (ABQM), including 120 of the racing line and 68 of the cutting line, were genotyped using SNP arrays. On the basis of 27 genomic regions identified as selection signatures by REHH and F_{ST} statistics, functional annotations of genes were made in order to identify those that could have been important during formation of the racing line and that could be used subsequently for the development of selection tools. Genes involved in muscle growth (n = 8), skeletal growth (n = 10), muscle energy metabolism (n = 15), cardiovascular system (n = 14), and nervous system (n = 23) were identified, including the *FKTN, INSR*, *GYS1*, *CLCN1*, *MYLK*, *SYK*, *ANG*, *CNTFR* and *HTR2B*.

Keywords: equine, fixation index, extended haplotype homozygosity, selection signature, SNP array

Implications

The Quarter horse breed is renowned for its functional versatility, presenting abilities in several sporting activities. Among these activities, horse racing has the highest economic importance to the equine industry. Racing generates income, in the form of prizes and betting, but it is also a high cost activity to train and maintain racing horses. The search for recent signatures of selection in the racing Quarter Horse, artificial or natural, allowed the identification of genes and DNA polymorphisms that might be useful as markers associated to racing performance. Association and validation studies will be required to confirm the relation of these markers with racing performance.

Introduction

Quarter Horses were the first breed to be developed in America and emerged in the United States in the 17th century. Selection of this breed for different purposes (Evans, 1996) led to the formation of lines, including racing and cutting horses. The racing line explores the sprinting ability of the animals over short distances, whereas the cutting line is used in functional tests, exploring abilities such as agility, obedience and cow sense which are important for the management of bovines on cattle ranches.

During domestication and breed formation, domestic animals have undergone natural and artificial selection. As a consequence, a genomic region containing multiple alleles or haplotypes could have become the only or most prominent region in the population, a situation known as positive selection. The identification of regions of the genome undergoing selection is of major interest and can be done exclusively by analysis of genomic patterns without the need for phenotypic information. Once identified and mapped, it is possible to annotate the biological function of these genomic regions and, consequently, to better understand how selection acts on complex traits of interest.

The main methods used for the identification of genomic regions undergoing positive selection include the long-range haplotype (**LRH**) test for the identification of these regions within populations, and the fixation index ($\mathbf{F_{ST}}$) which considers multiple populations. The LRH test, which detects recent positive selection by analyzing

haplotype structure in random individuals of a population and relies on the relationship between the frequency of an allele and the extent of linkage disequilibrium (**LD**) between that allele and the loci that surround it, uses extended haplotype homozygosity (**EHH**) statistic. This analysis is based on contrasts between an extended core haplotype exhibiting high frequency and homozygosity and other core haplotypes at the same locus (Sabeti *et al.*, 2002). On the other hand, the F_{ST} measures genetic differentiation to identify signatures of selection. This statistical method assumes that variable selective forces favor different gene variants in different regions of the genome. As a consequence, differences in allele frequencies between populations may be more extreme in regions that harbor these variants (Qanbari *et al.*, 2011). The method screens patterns of variation over many loci and considers those located in the tail of the empirical distribution of F_{ST} as outliers (Akey *et al.*, 2002), with higher values indicating divergent selection and lower values indicating balancing selection.

In view of the above considerations, the objective of the present study was to identify genomic regions divergently selected in the racing line of Quarter Horses in relation to the cutting line using single nucleotide polymorphism (**SNP**) genotypic arrays. For this purpose, relative extended haplotype homozygosity (**REHH**) statistic, an extension of EHH analysis, was applied only to the racing line and the F_{ST} was used for the analysis of racing and cutting animals. Annotation of genes underlying these genomic regions was made to identify those that could have been important during formation of the racing line and that could be used for the development of tools to increase selection efficiency.

Materials and Methods

Animals, SNP genotyping and data preparation

Whole blood samples (5 mL) were collected of 188 Quarter Horses of both sexes and registered at the Brazilian Association of Quarter Horse Breeders (ABQM). Of these, 120 animals belonged to the racing line and 68 to the cutting line. Animals of the racing line, including 18 males and 102 females born between 1985 and 2007 to 48 stallions and 107 mares, were from five properties in the State of São Paulo, Brazil. Animals of the cutting line, including 26 males and 42 females born between

1991 and 2009 to 44 stallions and 64 mares, were from three properties of the State of São Paulo. The presence of full siblings was avoided in the two lines. All animal procedures used in the experiment were in accordance with the Brazilian legislation for animal welfare.

Genomic DNA was extracted from blood cells using the Illustra Blood GenomicPrep Mini Spin kit (GE Healthcare, Little Chalfont, Buckinghamshire, UK). The DNA concentration in the samples was adjusted to 40-60 ng/μL. Genotyping was performed with the Equine SNP50 BeadChip (Illumina, Inc., USA) using the HiScan system (Illumina, Inc., USA).

Quality control of the genotype data for individuals (188) and SNPs (54,602) was performed using the Genome Studio 2011.1 program (Illumina, Inc., USA). Quality control of genotype data for individuals led to the exclusion of four animals of the cutting line because they presented a call rate of less than 0.95 (95%). With respect to SNPs, those located on the X chromosome with low genotyping accuracy (cluster separation <0.3; call frequency <0.9; *p*-value <1 x 10^{-3} for Hardy-Weinberg equilibrium) and a minor allele frequency <0.05 were removed. Thus, the final dataset consisted of 184 individuals (120 racing and 64 cutting) and 42,058 and 40,787 autosomal SNPs from the whole sample of animals and from the sample of animals of the racing line, respectively.

Analysis within the racing line

After quality control filtering of SNPs, only the dataset of the racing line, i.e., 40,787 SNPs, was used for application of the LRH test. The FastPHASE software, configured with standard parameters, was used for haplotype reconstruction and inference on the linkage phases of SNPs on each *Equus caballus* autosomes chromosomes (**ECA**).

Since recombination rates can vary markedly across different regions of the genome, it is possible that higher EHH values are due to low recombination rates in a given region and not necessarily to recent positive selection. In this respect, REHH statistic corrects the observed EHH value at one core haplotype for the average level of EHH values for all relevant core haplotypes of the same chromosome. The Sweep 1.1 software program (Sabeti *et al.*, 2002) was used for the detection of putative

selection signatures in the core regions as defined by REHH analysis. This program implements the algorithm proposed by Gabriel *et al*. (2002), which defines a pair of SNPs to be in strong LD if the upper 95% confidence interval of D' is between 0.7 and 0.98. It was specified that each core region should contain a minimum of 3 and a maximum of 20 SNPs. According to Qanbari *et al.* (2010), core regions are defined as regions of interest in the genome characterized by strong LD between SNPs and that involve a set of core haplotypes. The EHH test was applied at a distance of 1 cM on both sides of the core region and to determine the significance of REHH values, i.e., EHH values corrected for each ECA, the Sweep 1.1 program divides the haplotypes into 20 classes of frequency and compares the REHH value for each haplotype with other equally frequent haplotypes. The REHH values were log-transformed (-Log10) to achieve a normal distribution. Next, probability distributions, means, standard deviations and *P* values were obtained and also transformed to -Log10 (*P*). Therefore, core haplotypes with extreme *P* values in the distribution, i.e., -Log10 (*P*) higher than 2 ($P<0.01$), were considered to be significant. However, according to Glick *et al*. (2012), REHH can serve as evidence of recent selection only for haplotypes present at relatively high frequency in the population. As a consequence, only core haplotypes with a frequency ≥ 0.25 were considered to be relevant, since inferior values could be indicative of recombination events (Sabeti *et al*., 2002).

Analysis between the racing and cutting lines

Parameter $\theta=F_{ST}$ was estimated for the final set of SNPs that passed the quality control criteria (considering all individuals; 42,058 SNPs) using the Bayesian method proposed by Gianola *et al.* (2010). This method is a two-step procedure that provides clusters of θ values. The first step consists of a simple Bayesian structure to remove samples from the posterior distribution of θ-parameters without constructing Markov chains. This step, which uses Jeffrey's rule (Bernardo and Smith, 1994) assigns a weakly informative prior of beta distribution (1/2,1/2) to allele frequencies and does not make assumptions about evolutionary models. The second step of the procedure consists of clustering a set of estimates of posterior means of $\theta=F_{ST}$ from multilocus analysis into data-driven groups, considering that θ values derived from

the same distribution or from different distributions resulting from heterogeneity of adjacent stochastic processes. The cluster structure was given by fitting a sequence of finite mixture models of posterior means of $\theta=F_{ST}$ values for each locus, clustering loci according to their similarity in θ values. The models were compared using Akaike's information criterion (**AIC**) (Akaike, 1974), which is a measure of the relative quality of a statistical model for a given set of data and those with lower AIC values were preferred. The parameters of the model were estimated by the maximum likelihood method using the expectation-maximization algorithm implemented in the FlexMix package (Leisch, 2004) of the R Project. If the values of *θ* arise from different evolutionary or artificial processes, one would expect to observe a mixture of distributions that lead to clusters representative of the types of mechanisms operating; there may be various clusters harboring SNPs that underwent different processes of selection. For each SNP of each cluster, the conditional probability that *θ* belongs to this cluster was calculated. The SNPs of a cluster with higher *θ* values and a higher probability of belonging to this cluster probably reflect footprints of positive selection. Therefore, SNPs with high *θ* values and a conditional probability of belonging to the cluster ≥95% were defined as selection signatures.

Gene annotation

For annotation of genes in genomic regions selected divergently in racing line of Quarter Horses in relation to cutting line, core regions with relevant haplotypes (i.e., regions with significant REHH values at $P<0.01$ and a haplotype frequency ≥ 0.25) extended 1 Mb to both sides, which coincided with loci (SNPs) that belonged to a cluster with high $\theta=F_{ST}$ values and with a conditional probability of belonging to this cluster ≥ 95%, were selected. The genes corresponding to the genomic regions containing signatures of positive selection by the two statistical methods (REHH and F_{ST}) and criteria used were identified based on the latest assembly of the horse genome sequence (EquCab2.0) using the NCBI Map Viewer tool (National Center for Biotechnology Information, 2013). On the basis of the list of genes identified, a second list was generated that contained only genes with known function. These genes were annotated regarding the biological processes in which they participated using the Gene Ontology database and considering the species *Homo sapiens*.

Genes whose protein products participate in biological processes related to important traits in the racing line of Quarter Horses were highlighted.

Results and Discussion

Analysis within the racing line

A summary of informative SNPs and core regions established after REHH analysis, using Sweep 1.1 software program (Sabeti *et al*., 2002), were provided at Table 1. This table also shows the proportion of ECA length covered by core regions in relation to the total length of each ECA, as well as the number of SNPs forming core regions in relation to the total number of informative SNPs.

For the 4,690 core regions identified, 34,258 EHH tests (mean of 7.31 tests per core region) were performed. Table 2 shows the LRH statistics, including the outliers for each ECA. Among the 12,340 tests applied to core haplotypes with a frequency ≥0.25, 70 regions presented peak outliers at a threshold of 0.01, indicating that these regions might be under recent selection. These regions were present in larger numbers on ECA 2, 3, 10 and 17. According to Myles *et al.* (2008), alleles that are subject to selection present higher EHH (REHH) values than alleles of similar frequency that have not undergone selection. Other haplotypes present on the chromosomes in which selection signatures were identified had a shorter extent of homozygosity, demonstrating high recombination during the formation of the line. Therefore, the long runs of homozygosity observed in these regions are unlikely to be due to low recombination rates, but rather reflect selection pressures.

Qanbari *et al.* (2010) used SNPs of the Illumina BovineSNP50 BeadChip and the REHH test to evaluate LD and the decay of haplotype homozygosity in some candidate regions harboring major genes related to milk production and quality in Holstein-Friesian cattle. In general, the authors found high values of LD and slow decay of haplotype homozygosity, indicating the validity of the approach for the identification of selection signatures. Glick *et al.* (2012) used the LRH test together with direct analysis of changes in haplotype frequencies over time to provide evidence of selection in a population of Holstein-Israeli cattle. The positive correlation of 0.34 between the two approaches demonstrated that the LRH test was able to detect haplotypes that suffered significant and recent changes in their frequency,

indicating signatures of selection. Using EHH statistics to identify genome-wide selection signatures in Holstein-Friesian cattle, Qanbari *et al.* (2010) found 161 regions (P<0.01) that showed signs of recent positive selection. The larger number of regions indicative of positive selection identified by these authors, who used a SNP chip of the same density and the same significance criteria for the REHH test as in the present study, suggests more diverse (for different economically important traits) and intense selection in dairy cattle when compared to Quarter Horse racing, one of the most recently domesticated large animal species (Aberle and Distl, 2004). According to Glick *et al.* (2012), an increase of selection intensity should result in changes in the number of loci that respond to selection and faster changes in haplotype frequencies which, in turn, increase REHH values.

For better visualization of the distribution of putative selection signatures across the genome, the transformed *P* values of the REHH test were plotted against the position of the haplotypes on the ECA (Figure 1). It can be noticed that the signatures were not distributed uniformly across the genome, with the absence of significant selection signatures on ECA 9, 12, 22, 30, and 31.

Analysis between the racing and cutting lines

The mean value of parameter $\theta=F_{ST}$ (± standard deviation) estimated for the 42,058 informative SNPs, considering both lines, was 0.0342 (± 0.0403), with θ ranging from 0.0025 to 0.3556. The global F_{ST} value obtained by Petersen *et al. (*2014), who used genome-wide SNP data from Quarter Horses represented by six performance groups (halter, western pleasure, reining, working cow, cutting and racing) to examine genetic diversity within the breed, was 0.035. With the exception of the working cow and cutting horses, pairwise F_{ST} values show significant population structure among the six performance groups with the highest divergence observed between the cutting and racing (0.074). It is possible to speculate that this observed difference, 0.074 versus 0.0342 (present data), is due to a possible stronger selection in these lines of Quarter Horses in the United States than in Brazil.

The AIC favored the finite mixture model with seven clusters for the 42,058 loci (SNPs), since this model with seven clusters obtained the lowest AIC value. According to Gianola *et al.* (2010), the expectation is that these clusters are

representative of different processes that occurred in the populations, such as positive and/or negative directional selection, balancing selection, and neutrality. In the absence of directional selection, all F_{ST} values are expected to derive from the same stochastic evolutionary process and therefore share the same distribution. If, however, the differentiation of a breed or line is not only based on genetic drift, but directional selection is active and operates differently in different genomic regions, a higher degree of differentiation is expected for some regions when compared to the average differentiation caused by genetic drift. In the present study, cluster 4 comprising 2,558 loci presented higher θ values (large differentiation), reflecting potential selection signatures. Of these, 1,333 and 271 loci showed a conditional probability of belonging to this cluster ≥ 95% and 100%, respectively (Table 3). The θ value for the 1,333 loci ranged from 0.1383 to 0.3556, with the 15 highest values (0.3013-0.3556) being found on ECA 1, 2, 4, 5, 8, 9, 10, 16, 21, and 30.

Annotation of genes in genomic regions divergently selected in the racing line in relation to cutting line

To identify regions of the genome divergently selected in the racing line in relation to cutting line, 70 core regions with relevant haplotypes obtained by REHH analysis, which coincided with 1,333 loci (SNPs) identified by F_{ST} analysis, were used. Thus, 27 regions distributed on ECA 1, 2, 3, 4, 5, 6, 7, 10, 11, 13, 17, 18, 19, 21, 23, 24, 25, 26, and 27 were considered (Table 4). According to Walsh (2010), investigators have applied different tests as enrichment procedure to correct for demographic events, with regions presenting significant results based on a series of statistics being strong evidence of selection signatures.

Using the coinciding regions, the equine genome sequence (EquCab2.0) was searched for functional annotation of genes in order to identify those that might have been important during formation of the racing line and are therefore related to athletic phenotypes aimed at speed and performance. According to Evans (2007), the evaluation of the roles of several body systems as causes of limited performance in horses is reasonable and should include the pulmonary, cardiovascular and neuromuscular systems, in addition to anaerobic and aerobic mechanisms of energy supply. In this respect, of a list of 552 genes identified, 196 were genes with known

function which were annotated with the biological processes in which they participate. Genes involved in muscle growth (8), skeletal growth (10), muscle energy metabolism (15), cardiovascular system (14) and nervous system (23) were identified. Taken together, 42 genes were candidates since their protein products participate in biological processes related to important traits in the racing line (Tables 5, 6). Of the 552 genes identified, 49 had SNPs from a core region. Of these, 17 showed known functions and three are on the list of candidates (*GAB1*, *INSR* and *ADCY5*).

Although the breed was formally recognized in 1940 with the establishment of the American Quarter Horse Association (AQHA) in United State of America, according to Evans (1996) the first recorded races of these animals were held in 1674. Individuals used for racing are typically taller and leaner for short bursts of speed and with phenotypes similar to those of Thoroughbreds (Scott, 2008). Cutting animals began to be selected only in 1898 for participation in cutting events, looking for cow sense, i.e., the perception of the reaction of cattle. This type was smaller, more compact, robust, and strong (Evans, 1996; Scott, 2008). In agreement with these data, Meira *et al.* (2013) identified significant differences between racing and cutting Quarter Horses in terms of morphological traits. The authors observed that racing animals were heavier and taller and presented greater body lengths and perimeters than cutting horses.

The temperament of a horse determines its success in a specific sport discipline or at a given type of work (Mills, 1998), since it is the combination of appropriate temperament and physical traits that forms a winning athlete. Therefore, throughout history individuals have searched for horses with certain temperamental characteristics suitable for their intended use (Mills, 1998) and, to increase the prevalence of preferred temperamental characteristics in the population, the trait is commonly included as a breeding objective. These efforts resulted in the creation of breeds and lines that differ in temperament. Energetic horse breeds, such as racing animals, are more reactive, whereas cold blood breeds, such as most pulling and cutting horses, are quiet and calmer. Quarter Horses vary widely in conformation and temperamental characteristics. The slim and hot-tempered type, which often has a Thoroughbred influence, is appropriate for racing. These characteristics are not

suitable for cutting horses which should be obedient, easier to control (especially during lateral movements), serene and much more compact, i.e., smaller withers height and greater shoulder width (B. Langlois, personal communication).

Although different physical and behavioral attributes would be necessary for different types of competitive horses, the success in competitions mainly depends on the metabolic capacity of the animal to convert chemical into mechanical energy. Components of these energy processes include supply of glucose and oxygen by the cardiovascular system and the rate, efficiency and interaction of aerobic and anaerobic metabolism in muscles. According to Freeman (2013), horses perform different types of physical exercise, ranging from predominantly aerobic to predominantly anaerobic. Quarter Horse races are mainly anaerobic, whereas some arena performance classes, such as cutting and reining, modalities performed by the cutting line of Quarter Horses, intercalate short bouts of anaerobic exercise with longer periods of aerobic activity (Freeman, 2013).

The superiority of the cardiovascular system of Thoroughbred horses, which are skilled runners, is related to the fact that the size of the heart per unit body weight is proportionally greater when compared to other large mammals (Gunn, 1989). Analysis of the pedigrees of Quarter Horses shows that the influence of Thoroughbreds is so marked that it permits to follow the so-called “large-heart gene”. The main objective of the integration of the cardiovascular and respiratory systems is the transfer of oxygen from the upper airways to skeletal muscle mitochondria and the removal of residual products, such as carbon dioxide, in the opposite direction. Therefore, the adequate maintenance of capillary vessels that transport blood to and from the muscle is of the utmost importance for performance. According to Henckel (1983), physical exercise can induce angiogenesis (capillarization) in the muscles of horses.

Among the genes identified near selection signatures in the racing line of Quarter Horses that are related to the biological processes of growth, muscle energy metabolism and nervous and cardiovascular systems, the following genes will be discussed: *FKTN, INSR*, *GYS1*, *CLCN1*, *MYLK*, *SYK*, *ANG*, *CNTFR* and *HTR2B*. According to NCBI (2013), the number of SNPs mapped in these genes is 25, 49, 5, 14, 148, 12, 6, 11 and 1, respectively, considering the specie *Equus caballus* (Tables

5, 6). Of these genes, *ANG*, *CNTFR*, *GYS1* and *MYLK* are considered candidates for performance and health-related fitness phenotypes in humans (Bray *et al.*, 2009).

Many of the genes highlighted and discussed below are implied in diseases or disorders in humans and horses, however, it should be noted that different nucleotide changes occurring in these genes may result in milder phenotypes that do not cause physical incapacity, or even phenotypes favorable to the development of higher athletic potential. Thus, these putative different gene forms would be under selective pressures.

In humans, mutations in *FKTN* (fukutin*)* are associated with altered glycosylation of α-dystroglycan, causing Fukuyama congenital muscular dystrophy. This condition is characterized by delayed motor development, muscle weakness, loss of muscle mass and hypotonia, and is associated with central nervous system anomalies and ocular disorders (Murakami *et al.*, 2006).

Mutations in the *INSR* (insulin receptor) human gene cause a spectrum of hereditary insulin resistance syndromes that can lead, for example, to growth retardation (Kim *et al.*, 2012). According to Suagee *et al.* (2011), insulin resistance in horses is associated with obesity, an increased risk of laminitis, osteochondrosis and rhabdomyolysis, in addition to defective glucose transport in skeletal muscle and adipose tissue. The most important implication of insulin resistance in horses is laminitis, a disease affecting the locomotor system of the animal, with consequent low performance in any type of activity.

Mutations in the *GYS1* (glycogen synthase 1) are associated with polysaccharide storage myopathy (PSSM), a disease characterized by high concentrations of glycogen and glucose-6-phosphate in skeletal muscle and abnormal accumulation of amylase-resistant complex polysaccharide (Aleman, 2008). Animals diagnosed with PSSM are reluctant to exercise and present poor performance, rhabdomyolysis, weakness, gait abnormalities, and muscle atrophy and stiffness (Aleman, 2008; Barrey, 2010). A nucleotide polymorphism in *GYS1*, identified by McCue *et al.* (2008), causes the substitution of a guanine with adenine and changes the amino acid arginine to histidine (Arg309His) in affected Quarter Horses. Stanley *et al.* (2009) investigated whether this mutation in *GYS1* occurs in horses from the United Kingdom with histopathological evidence of PSSM and

rhabdomyolysis. Among 47 horses with rhabdomyolysis or poor performance, 21% carried the mutation in *GYS1* and 25% of the breeds with this mutation were Quarter Horses, Appaloosa or related breeds. According to McCoy *et al*. (2014) the region surrounding the Arg309His mutation exhibits several signatures of selection, including high-derived allele frequency and lack of haplotype diversity surrounding the mutation. However, to these authors, the persistence of the *GYS1* mutation in the Quarter Horse population cannot be attributed to selection but is likely the result of founder effect and a recent population expansion. In the present study, it is also possible to speculate that the potential for this *locus* has high REHH and divergent F_{ST} due to drift is high given the two sample design.

In humans, mutations in *CLCN1* (chloride channel, voltage-sensitive 1) cause two forms of muscle diseases with different patterns of inheritance (NCBI, 2013), which are characterized by slow muscle relaxation after voluntary contraction or electrical stimulation, stiffness, and muscle weakness. The main function of chloride channels is to stabilize the cells' electrical charge, permitting normal muscle contraction. In a case study of a New Forest pony, Wijinberg *et al.* (2012) reported a missense mutation in the *CLCN1* gene that was associated with a phenotype of hereditary recessive myotonia. According to Nollet and Deprez (2005), horses with this syndrome show gait abnormalities and stiffness after birth and these symptoms get worse after a period of rest.

In a study of *MYLK* (myosin light chain kinase) genotypes in 157 predominantly Caucasian men and women, Clarkson *et al.* (2005) demonstrated an association between individuals homozygous for the rare alleles 49T and 3788A and an increase of creatine kinase and myoglobin when compared to the wild-type genotypes. In addition, elevated concentrations of creatine kinase and myoglobin were strongly associated with the T allele of the *MYLK* gene. The A allele was associated with a loss of post-exercise strength, with both polymorphisms probably increasing the risk of rhabdomyolysis and acute kidney failure. In view of these findings, the *MYLK* gene has been indicated as a functional candidate gene for physical performance in both humans (Bray *et al.*, 2009) and horses (Schröder *et al*., 2011).

The *SYK* (spleen tyrosine kinase) has been implicated in different functions,

including the morphogenesis of endothelial cells and maintenance of vascular integrity *in vivo* (Yanagi *et al.*, 2001). Recently, Kazerounian *et al.* (2011) confirmed the participation of *SYK* in angiogenesis in human cell cultures. Missense variants in the *ANG* gene (angiogenin, ribonuclease, RNase A family, 5), an angiogenic factor that is expressed in ventral horn neurons and up-regulated by hypoxia, were detected in patients with amyotrophic lateral sclerosis from Ireland/Scotland, North America, Italy, France, Holland, and Germany (Fernandez-Santiago *et al.*, 2009). This disease causes muscle weakness accompanied by hardening of the muscles (sclerosis), first on one side of the body (lateral), muscle atrophy (amyotrophic), cramps, tremor, abnormal reflexes, spasms, and loss of sensation.

SNPs in *CNTFR* (ciliary neurotrophic factor receptor) may be associated with variations in muscle strength, as well as early onset of eating disorders (NCBI, 2013). De Mars *et al.* (2007) examined multiple polymorphisms at both the *CNTF* and the *CNTF* receptor (*CNTFR*) loci in a study of 493 middle-aged and older men and women with measures of knee flexor and extensor strength. T-allele carriers of the C-1703T *CNTFR* polymorphism exhibited higher strength levels for multiple measures compared with CC homozygotes, including all knee flexor torque values. In middle-aged women, A-allele carriers at the T1069A *CNTFR* locus exhibited lower concentric knee flexor peak torque at multiple speeds and isometric torque at 120 compared with T/T homozygotes.

The *HTR2B* (5-hydroxytryptamine (serotonin) receptor 2B, G protein-coupled) gene encodes one of the several different 5-hydroxytryptamine (serotonin) receptors, which belong to the G protein-coupled receptor 1 family. Serotonin receptors mediate many of the central and peripheral physiological functions of serotonin, including the regulation of cardiovascular function and impulsive behavior (NCBI, 2013), i.e., an action executed without delay, reflection, voluntary direction or obvious control in response to a stimulus. Family- and population-based analysis in humans of a minor allele, which blocks the expression of the protein, and knockout studies in rats suggest a relationship between this gene and impulsivity (Bevilacqua *et al.*, 2010).

Conclusions

The combined use of high-density SNP genotyping arrays and REHH and F_{ST}

statistics permitted the identification of regions in the horse genome that were divergently selected in racing Quarter Horses in relation to cutting. Gene annotation in these regions was used to identify genes that could have been important during formation of the racing line and that are related to racing performance. These genes should be used in the future in association studies on Quarter Horse racing for the development of tools to increase selection efficiency. In these studies, gene SNPs responsible for amino acid exchange, located at microRNA binding sites or related to other functional modifications will be privileged.

Acknowledgements

The authors thank the horse farms for providing the Quarter Horses and the financial support of Fapesp and Capes. They also thank the invaluable contribution of the late Professor Dr. Marcilio Dias Silveira da Mota.

References

Aberle KS and Distl O 2004. Domestication of the horse: results based on microsatellite and mitochondrial DNA markers. Archiv Tierzucht 47, 517-535.

Akaike H 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716-723.

Akey JM, Zhang G, Zhang K, Jin L and Shriver MD 2002. Interrogating a high density SNP map for signatures of natural selection. Genome Research 12, 1805-1814.

Aleman M 2008. A review of equine muscle disorders. Neuromuscular Disorder 18, 277-287.

Barrey E 2010. Genetics and genomics in equine exercise physiology: an overview of the new applications of molecular biology as positive and negative markers of performance and health. Equine Veterinary Journal 42, 561-568.

Bernardo JM and Smith AFM 1994. Bayesian theory. Wiley, Chichester, UK.

Bevilacqua L, Doly S, Kaprio J, Yuan Q, Tikkanen R, Paunio T, Zhou Z, Wedenoja J, Maroteaux L, Diaz S, Belmer A, Hodgkinson CA, Dell'Osso L, Suvisaari J, Coccaro E, Rose RJ, Peltonen L, Virkkunen M and Goldman D 2010. A population-specific HTR2B stop codon predisposes to severe impulsivity. Nature 468, 1061-1066.

Bray MS, Hagberg JM, Pérusse L, Rankinen T, Roth SM, Wolfarth B and Bouchard C 2009. The human gene map for performance and health-related fitness phenotypes: the 2006-2007 update. Medicine and Science in Sports and Exercise 41, 35-73.

Clarkson PM, Hoffman EP, Zambraski E, Gordish-Dressman H, Kearns A, Hubal M, Harmon B and Devaney JM 2005. ACTN3 and MLCK genotype associations with exertional muscle damage. Journal of Applied Physiology 99, 564-569.

De Mars G, Windelinckx A, Beunen G, Delecluse C, Lefevre J and Thomis MA 2007. Polymorphisms in the CNTF and CNTF receptor genes are associated with muscle strength in men and women. Journal of Applied Physiology 102, 1824-1831.

Evans JW 1996. Horses: a guide to selection, care and enjoyment. Freeman and Company, New York, USA.

Evans DL 2007. Physiology of equine performance and associated tests of function. Equine Veterinary Journal 39, 373-383.

Fernandez-Santiago R, Hoenig S, Lichtner P, Sperfeld AD, Sharma M, Sperfeld AD, Sharma M, Berg D, Weichenrieder O, Illig T, Eger K, Meyer T, Anneser J, Münch C, Zierz S, Gasser T and Ludolph A 2009. Identification of novel angiogenin (ANG) gene missense variants in German patients with amyotrophic lateral sclerosis. Journal of Neurology 256, 1337-1342.

Freeman DW 2013. Physical conditioning of horses. Oklahoma state University: Division of Agricultural Sciences and Natural Resources. ANSI-3983, 1-6. Retrieved on 02 June 2013, from http://pods.dasnr.okstate.edu/docushare/dsweb/View/Collection-305

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ and Altshuler D 2002. The structure of haplotype blocks in the human genome. Science 296, 2225-2229.

Gianola D, Simianer H and Qanbari S 2010. A two-step method for detecting selection signatures using genetic markers. Genetics Research Cambridge 92,141-155.

Glick G, Shirak A, Uliel S, Zeron Y, Ezra E, Seroussi E, Ron M and Weller JI 2012. Signatures of contemporary selection in the Israeli Holstein dairy cattle. Animal Genetics 43,45-55.

Gunn HM 1989. Heart weight and running ability. Journal of Anatomy 167, 225-233.

Henckel P 1983. Training and growth induced changes in the middle gluteal muscle of young Standardbred trotters. Equine Veterinary Journal 14, 134-140.

Kazerounian S, Duquette M, Reyes MA, Lawler JT, Song K, Perruzzi C, Primo L, Khosravi-Far R, Bussolino F, Rabinovitz I and Lawler J 2011. Priming of the vascular endothelial growth factor signaling pathway by thrombospondin-1, CD36, and spleen tyrosine kinase. Blood 117, 4658-4666.

Kim D, Cho SY, Yeau SH, Park SW, Sohn YB, Kwon MJ, Kim JY, Ki CS and Jin DK 2012. Two novel insulin receptor gene mutations in a patient with Rabson-Mendenhall syndrome: the first Korean case confirmed by biochemical, and molecular evidence. Journal of Korean Medical Science 27, 565-568.

Leisch F 2004. FlexMix: A general framework for finite mixture models and latent class regression in R. Journal of Statistical Software 11, 1-18.

McCoy AM, Schaefer R, Petersen JL, Morrell PL, Slamka MA, Mickelson JR, Valberg SJ and McCue ME 2014. Evidence of positive selection for a glycogen synthase (GYS1) mutation in domestic horse populations. Journal of Heredity 105, 163-172.

McCue ME, Valberg SJ, Miller MB, Wade C, DiMauro S, Akman HO and Mickelson JR 2008. Glycogen synthase (GYS1) mutation causes a novel skeletal muscle glycogenosis. Genomics 91, 458-466.

Meira CT, Curi RA, Silva JAIIV, Corrêa MJM, Oliveira HN and Mota MDS 2013. Morphological and genomic differences between cutting and racing lines of Quarter Horses. Journal of Equine Veterinary Science 13, 244-249.

Mills DS 1998. Personality and individual differences in the horse, their significance, use and measurement. Equine Veterinary Journal Supplement 27, 10-13.

Murakami T, Hayashi YK, Noguchi S, Ogawa M, Nonaka I, Tanabe Y, Ogino M, Takada F, Eriguchi M, Kotooka N, Campbell KP, Osawa M and Nishino I 2006. Fukutin Gene mutations cause dilated cardiomyopathy with minimal muscle weakness. Annals of Neurology 60, 597-602.

Myles S, Tang K, Somel M, Green RE, Kelso J and Stoneking M 2008. Identification and analysis of genomic regions with large between-population differentiation in humans. Annals of Human Genetics 72, 99-110.

NCBI 2013. National Center for Biotechnology Information. Retrieved on 01 May 2013, from http://www.ncbi.nlm.nih.gov

Nollet H and Deprez P 2005. Hereditary skeletal muscle diseases in the horse: a review. The Veterinary Quartely 27, 65-75.

Petersen JL, Mickelson JR, Cleary KD and McCue ME 2014. The American Quarter Horse: population structure and relationship to the Thoroughbred. Journal of Heredity, 105(2): 148-162.

Qanbari S, Pimentel ECG, Tetens J, Thaller G, Lichtner P, Sharifi AR and Simianer H 2010. A genome-wide scan for signatures of recent selection in Holstein cattle. Animal Genetics 41, 377-389.

Qanbari S, Gianola D, Hayes B, Schenkel F, Miller S, Moore S, Thaller G and Simianer H 2011. Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. BMC Genomics 12, 318.

Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooperk R, Kwiatkowski D, Ward R and Lander ES 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419, 832–837.

Schröder W, Klostermann A and Distl O 2011. Candidate genes for physical performance in the horse. The Veterinary Journal 190, 39-48.

Scott M 2008. Musculoskeletal injuries in nonracing Quarter Horses. Veterinary Clinics Equine Practice 24, 133-152.

Stanley RL, McCue ME, Valberg SJ, Mickelson JR, Mayhew IG, Mcgowan C, Hahn CN, Patterson-Kane JC and Piercy RJ 2009. A glycogen synthase 1 mutation associated with equine polysaccharide storage myopathy and exertional rhabdomyolysis occurs in a variety of UK breeds. Equine Veterinary Journal 41, 597-601.

Suagee JK, Corl BA, Wearn JG, Crisman MV, Hulver MW, Geor RJ and McCutcheon LJ 2011. Effects of the insulin-sensitizing drug pioglitazone and lipopolysaccharide administration on insulin sensitivity in horses. Journal of Veterinary Internal Medicine 25, 356-364.

Walsh B 2010. Tests for molecular signatures of selection. In: Institute on Statistical Genetics, August 2010, Beijing, China.

Wijnberg ID, Owczarek-Lipska M, Sacchetto R, Mascarello F, Pascoli F, Grunberg W, van der Kolk JH and Drogemuller C 2012. A missense mutation in the skeletal muscle chloride channel 1 (CLCN1) as candidate causal mutation for congenital myotonia in a New Forest pony. Neuromuscular Disorders 22, 361-367.

Yanagi S, Inatome R, Ding J, Kitaguchi H, Tybulewicz VL and Yamamura H 2001. SYK expression in endothelial cells and their morphologic defects in embryonic SYK-deficient mice. Blood 98, 2869-2871

Table 1 *Descriptive statistics of informative SNPs and core regions (CR) in Quarter Horses of the racing line*

ECA	SNPs	ECA length (Mbp)	Mean Distance[1] (Kb)	N° CR	Mean_CR length (Kb)	Coverage CR length (Kb)	Max_CR length (Kb)	CR_length/ECA length[2]	CR_SNPs[3]	MaxCR SNPs	CR_SNPs/ SNPs
1	3,381	186	55.01	386	196.05 ± 187.80	75,673.59	1,384.86	0.41	2,078	20	0.61
2	2,238	121	54.07	272	166.82 ± 158.51	45,373.91	996.69	0.37	1,363	17	0.60
3	2,213	119	53.77	239	170.53 ± 172.84	40,756.54	1,109.77	0.34	1,230	20	0.55
4	1,980	109	55.05	189	284.00 ± 316.06	53,676.28	1,872.05	0.49	1,261	20	0.63
5	1,773	100	56.40	206	157.25 ± 148.88	32,394.50	895.29	0.32	946	15	0.53
6	1,474	85	57.67	142	174.65 ± 153.99	24,800.10	718.27	0.29	743	17	0.50
7	1,792	99	55.25	209	216.99 ± 281.59	45,352.16	2,320.76	0.45	1,181	20	0.65
8	1,720	94	54.65	168	168.06 ± 168.76	28,234.22	1,218.97	0.30	851	18	0.49
9	1,529	84	54.94	161	169.63 ± 161.96	27,311.56	904.51	0.32	807	20	0.52
10	1,478	84	56.83	173	193.18 ± 236.59	33,420.51	1,638.15	0.39	950	20	0.64
11	1,102	61	55.35	106	193.25 ± 204.10	20,485.46	1,186.57	0.33	567	20	0.51
12	531	33	62.15	77	145.36 ± 143.59	11,192.90	837.75	0.33	318	12	0.59
13	739	43	58.19	66	162.92 ± 184.08	10,753.08	1,019.94	0.25	307	20	0.41
14	1,703	94	55.20	159	205.54 ± 213.08	32,681.53	1,310.51	0.34	928	19	0.54
15	1,759	92	52.30	168	174.36 ± 168.53	29,293.43	998.51	0.31	894	20	0.50
16	1,675	87	51.94	184	173.35 ± 157.43	31,898.20	708.80	0.36	985	20	0.58
17	1,492	81	54.29	181	169.00 ± 151.73	30,590.31	775.31	0.37	978	17	0.65
18	1,405	83	59.07	155	198.32 ± 189.94	30,740.52	976.12	0.37	848	18	0.60
19	1,151	60	52.13	127	185.08 ± 169.85	23,505.77	732.58	0.39	686	16	0.59
20	1,184	64	54.05	162	167.37 ± 143.36	27,114.52	704.75	0.42	810	20	0.68
21	1,044	58	55.56	97	186.44 ± 169.39	18,085.55	1,049.08	0.31	517	13	0.49
22	904	50	55.31	88	184.31 ± 163.29	16,219.57	801.06	0.32	470	13	0.51
23	975	56	57.44	157	172.11 ± 167.04	27,021.58	751.46	0.48	744	15	0.76
24	885	47	53.11	131	190.48 ± 193.68	24,954.06	1,176.21	0.53	677	18	0.76
25	703	40	56.90	97	209.53 ± 186.03	20,325.08	785.39	0.50	538	17	0.76
26	765	42	54.90	116	182.45 ± 163.84	21,164.73	629.26	0.50	595	20	0.77
27	676	40	59.17	93	189.04 ± 180.17	17,581.52	791.13	0.43	465	14	0.68
28	860	46	53.49	129	162.49 ± 123.14	20,962.29	608.65	0.45	638	19	0.74
29	599	34	56.76	98	154.41 ± 160.86	15,132.61	1,075.87	0.44	465	17	0.77
30	571	30	52.54	81	150.88 ± 126.54	12,221.33	638.36	0.40	408	12	0.71
31	486	25	51.44	73	119.78 ± 102.54	8,744.14	469.67	0.34	324	13	0.66
Total	40,787	2,247		4,690		857,661.70			24,572		
Mean			55.32		179.80 ± 175.79			0.39			0.62

[1]Mean distance between SNPs; [2]The proportion of total core region lengths on ECA length; [3]Number of SNPs forming core regions

Table 2 *Statistical results of the application of the long-range haplotype (LRH) test to the genome of Quarter Horses of the racing line*

ECA	Haplotypes of the core region with *P*-value<0.01	-Log10 *(P)* REHH
1	2	2.014; 2.225
2	6	2.041; 2.061; 2.025; 2.816; 2.155; 2.085
3	6	2.108; 2.295; 2.038; 2.745; 2.071; 2.493
4	2	2.090; 2.100
5	1	2.017
6	3	2.012; 2.194; 2.457
7	1	2.321
8	1	2.274
9	0	-
10	6	2.011; 2.029; 2.082; 2.318; 2.383; 2.643
11	2	2.353; 2.583
12	0	-
13	2	2.280; 2.299
14	1	2.371
15	1	2.173
16	3	2.085; 2.163; 2.570
17	5	2.161; 2.303; 2.306; 2.493; 2.512
18	2	2.096; 2.306
19	3	2.023; 2.091; 2.214
20	1	2.159
21	2	2.001; 2.596
22	0	-
23	3	2.060; 2.354; 2.614
24	3	2.049; 2.107; 2.220
25	3	2.001; 2.084; 2.355
26	3	2.028; 2.173; 2.212
27	3	2.016; 2.331; 2.487
28	3	2.014; 2.950; 3.266
29	2	2.130; 2.188
30	0	-
31	0	-
Total	70	-

Table 3 *Number of SNPs identified by F_{ST} analysis with probabilities of 50 to 100%, ≥95% and 100% of belonging to cluster 4, clustered according to their similarity in θ values*

ECA	n° SNPs (probability of 50 - 100%)	n° SNPs (probability ≥95%)	n° SNPs (probability of 100%)
1	343	203	49
2	85	31	3
3	63	29	5
4	144	91	22
5	141	78	18
6	82	40	8
7	120	53	9
8	93	55	11
9	79	38	7
10	222	133	35
11	102	48	3
12	26	9	3
13	35	19	2
14	113	58	13
15	45	22	2
16	52	24	3
17	71	27	5
18	110	55	8
19	41	16	1
20	46	24	4
21	91	40	11
22	114	64	12
23	33	13	0
24	46	15	1
25	74	45	15
26	38	26	6
27	20	8	0
28	63	28	1
29	13	4	2
30	45	33	12
31	8	4	0
Total	2558	1333	271

Table 4 *Genomic regions identified simultaneously by REHH and F_{ST} analysis in Quarter Horse racing, total number of genes identified in these regions, genes with known function, and possible candidate genes related to performance in the line*

ECA	Regions identified simultaneously by REHH and F_{ST} analysis (Mb)	Total of genes identified in the region	Genes with known function	Highlighted genes
1	155681241-157729078	30	4	1
2	3500445-5712047	27	7	0
2	87453500- 89546885	8	5	2
3	1135502-3148623	11	6	1
4	96192056-98315650	64	2	1
5	71331029-73597669	28	11	1
6	17372689-19443677	26	9	2
7	3852528-5885046	30	20	2
10	17198826-19258896	30	20	2
10	65461613-67583618	14	4	0
10	66332691-68390785	7	3	0
11	13217035-15365385	25	10	2
11	16263849-18610427	30	14	4
13	30034068-32120503	16	4	1
17	29268897-31272574	5	2	0
17	68569943-70974288	11	4	1
18	30519448-32573824	9	5	1
19	16749723-18751024	15	4	3
19	35081387-37245740	22	10	4
21	32208101-34356967	6	2	0
23	50273860-52336613	27	6	2
24	28594194-30860185	1	1	0
25	11521403-13721665	9	4	1
25	31977643-34015996	30	12	4
25	34107712-36261119	30	15	5
26	38221791-40393545	30	9	1
27	14932202-16989870	11	3	1
Total		552	196	42

Table 5 *Genes related to biological processes of the muscular, skeletal, cardiovascular and nervous systems and muscle energy metabolism located on Equus caballus autosomes chromosomes (ECA) 1, 2, 3, 4, 5, 6, 7, 10, 11, 13, 17, and 18 near selection signatures in the genome of Quarter Horses of the racing line and the number of SNPs identified in these genes*

Gene	ECA	Position (bp)	Biological process	SNPs*
ANG	1	157659668-157660108	Cardiovascular system	6
GAB1	2	88519893-88561558	Muscle energy metabolism, cardiovascular system	28
HHIP	2	87708863-87791726	Skeletal growth, muscle energy metabolism, nervous system	40
ZNF423	3	2578026-2838645	Nervous system	300
CLCN1	4	96354931-96383582	Muscle energy metabolism, nervous system	14
GLMN	5	72486695-72524803	Muscle growth, cardiovascular system	2
CAB39	6	18205843-18248940	Muscle energy metabolism	9
HTR2B	6	18495039-18508624	Cardiovascular system, nervous system	1
INSR	7	4881372-4988912	Muscle growth, skeletal growth, muscle energy metabolism, cardiovascular system	49
TRIP10	7	4307807-4315960	Muscle growth	5
GYS1	10	18932024-18946464	Muscle energy metabolism, cardiovascular system	5
TEAD2	10	19197526-19209520	Cardiovascular system	1
BPTF	11	14626924-14773051	Nervous system	71
MAPT	11	17585299-17628465	Nervous system	35
NSF	11	17909543-18051100	Nervous system	62
PRKCA	11	13529948-13812554	Skeletal growth, muscle energy metabolism, cardiovascular system, nervous system	131
WNT3	11	18057065-18102930	Skeletal growth, nervous system	20
WNT9B	11	18142485-18146383	Skeletal growth, nervous system	4
MKL2	13	30738393-30850452	Muscle growth	30
ZIC2	17	68992933-69001557	Nervous system	0
ACVR2A	18	30446696-30531592	Skeletal growth	7

* Source: NCBI (2013)

Table 6 *Genes related to biological processes of the muscular, skeletal, cardiovascular and nervous systems and muscle energy metabolism located on Equus caballus autosomes chromosomes (ECA) 19, 23, 25, 26, and 27 near selection signatures in the genome of Quarter Horses of the racing line and the number of SNPs identified in these genes*

Gene	ECA	Position (bp)	Biological process	SNPs*
ACTL6A	19	18615458-18639332	Muscle growth, skeletal growth, nervous system	0
ADCY5	19	36106052-36255077	Muscle energy metabolism, nervous system	56
GNB4	19	18473338-18503623	Muscle energy metabolism	2
KALRN	19	35092792-35480732	Nervous system	194
KCNMB3	19	18344870-18361079	Nervous system	0
MYLK	19	35876148-35957070	Muscle energy metabolism, cardiovascular system	148
SEMA5B	19	36529003-36605958	Nervous system	11
CNTFR	23	50484862-50514602	Skeletal growth, nervous system	11
SYK	23	51818352-51873030	Cardiovascular system	12
ABL1	25	33445465-33548338	Muscle growth	13
ASS1	25	33202800-33251824	Muscle growth, skeletal growth, muscle energy metabolism	20
BARHL1	25	34940813-34947928	Nervous system	4
DBH	25	35806027-35829206	Muscle energy metabolism, cardiovascular system, nervous system	5
FKTN	25	11644336-11698798	Muscle growth, nervous system	25
NUP188	25	32056416-32101114	Muscle energy metabolism	37
NUP214	25	33730288-33827417	Muscle energy metabolism	61
RAPGEF1	25	34102495-34214220	Nervous system	128
TSC1	25	35212315-35258127	Muscle energy metabolism, cardiovascular system, nervous system	35
WDR5	25	36214721-36231761	Skeletal growth	15
SIK1	26	38743732-38752729	Cardiovascular system	5
DLC1	27	16635814-17033049	Cardiovascular system, nervous system	227

* Source: NCBI (2013)

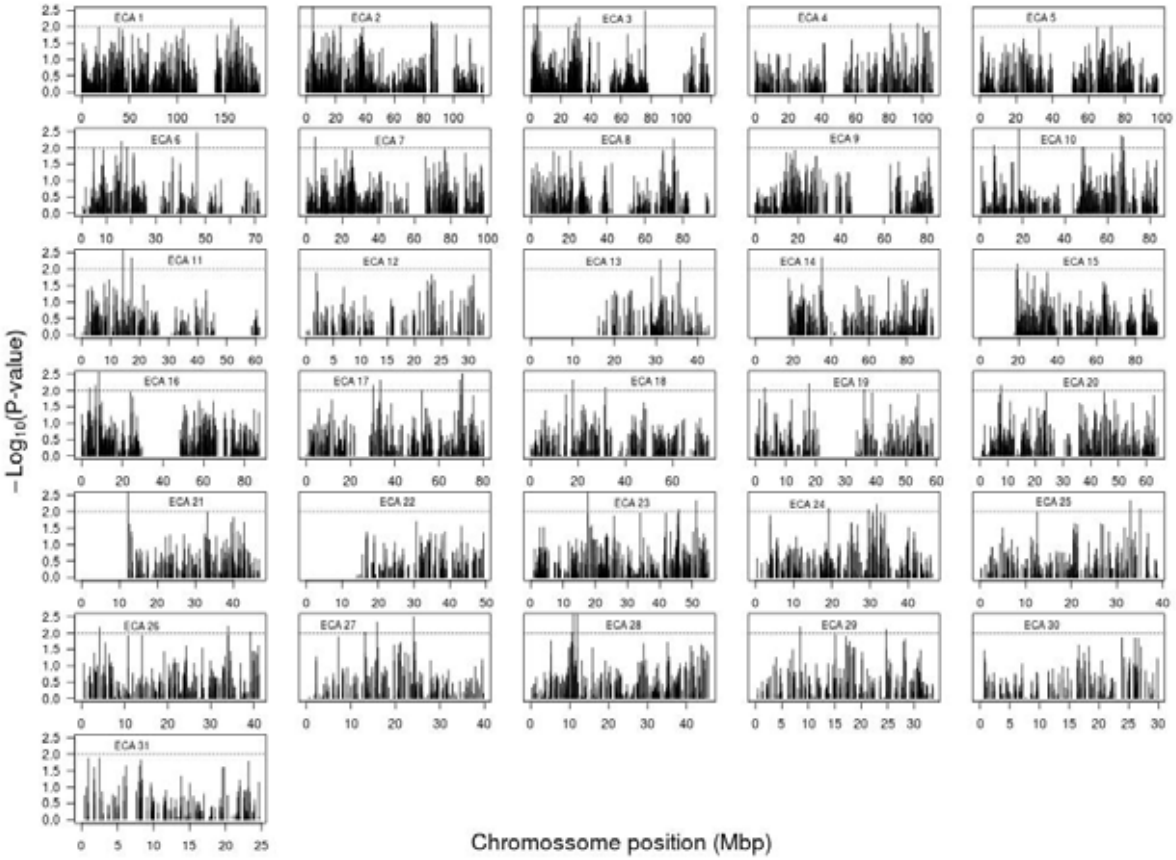

Figure 1 Map of putative selection signatures in the genome of Quarter Horses of the racing line identified by REHH analysis. Only haplotypes with a frequency > 25% were considered. Dashed lines indicate $P<0.01$.

CAPÍTULO 4 – Speed index in the racing Quarter Horse: a genome wide association study

C. T. Meira[1], M. R. S Fortes[2], M. M. Farah[1], M. Kelly[2], L. R. Porto Neto[3], G. L. Pereira[1], S. Moore[2], R. A. Curi[4], M. D. S. Mota[4]

[1]Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal, São Paulo 14884-900, Brazil.
[2]Queensland Alliance for Agriculture and Food Innovation, Centre for Animal Science, The University of Queensland, Brisbane, Queensland 4072, Australia.
[3]CSIRO Food Futures Flagship and Animal, Food and Health Sciences, Queensland Bioscience Precinct, Brisbane, QLD, 4067, Australia.
[4]Faculdade de Medicina Veterinária e Zootecnia, Universidade Estadual Paulista, Botucatu, São Paulo 18618-000, Brazil.

Abstract

The racing line of Quarter Horses is characterized by great sprinting speed over short distances on straight tracks. To perform selection on racing horses, the speed index (SI) and conformation traits are often used. A genome-wide association study (GWAS) on 112 racing Quarter Horses was performed for the SI trait and markers and genes associated were reported. The GWAS was carried out using the Qxpak.5 software and the genotyping data obtained from the Equine SNP50 BeadChip. A total of eight significant SNPs ($P < .0001$; Q-value of 0.25) distributed on ECA 2, 4, 10, 18 and 27 were found and served as markers for genomic regions mined for candidate genes associated with SI. For candidate gene annotation was considered 100Kb windows upstream and downstream to each important SNP. The highlighted genes were *GRM8, GRIK2, NEB, ANK1* and *KAT6A*, since their function could be related to racing performance. Future studies should consider a validation study with an independent population and sequencing of these candidate genes should be done to identify causal SNPs.

Keywords: *Equus caballus*, single nucleotide polymorphism, performance, quantitative trait, racehorse

1. Introduction

Quarter horses originated from animals brought from Arabia and Turkey to North America by Spanish explorers and traders, becoming the first breed developed in Americas. Artificial selection of the breed started with racing animals in 1647 [1], with the aim of selecting taller and leaner individuals for short bursts of speed and with phenotypes similar to those of Thoroughbreds [2]. The racing line of Quarter Horses is characterized by great sprinting speed over short distances on straight tracks. A selection attribute utilized by breeders of the racing Quarter Horse is speed index (SI), which is based on racing times achieved by the very best runners at each racetrack. The aim of the SI is to classify the performance of animals and allow comparisons between horses under different conditions, e.g. various distances, racetracks, climate, and countries [1]. It is a measure of easy interpretation by the breeders and available in the race records. A continued development of the racing Quarter Horse is being by Brazilian breeders, aided by SI annotation.

The availability of an equine single nucleotide polymorphism (SNP) array made possible the execution of many genome wide association studies (GWAS). In many species, GWAS based on SNP arrays have been used to detect changes caused by genetic selection and identified quantitative trait loci (QTL). In humans, GWAS helped to elucidate genetic diseases [3]. In cattle, GWAS successfully targeted productive and reproductive traits [4,5]. In horses, GWAS are less frequent and the lower numbers of animals with relevant phenotypes can partially explain this low frequency. Nonetheless, GWAS is now a proven and established approach for genetic studies in mammalian species.

Despite difficulties, some studies applying genomic approaches to performance traits in horses have been conducted. Current studies have identify SNPs on *Equus caballus* autosomes chromosomes (ECA) 18 within and proximal to the myostatin gene (*MSTN*) that is associated with racing performance in Thoroughbred horses [6-8]. According to Binns et al. [6], the identification of SNPs linked to the myostatin loci, which were associated with athletic ability in Thoroughbred represent the first gene found to influence athletic ability in horses. Gu et al. [9] suggested that sequence variation in two other genes, *COX4I2* and *CKM*, also contributes to Thoroughbred performance observed on the racetrack. Hill et al.

[10] identified a variant in the genomic sequence for *PDK4* gene strongly associated ($P < .0005$) with elite racing Thoroughbred performance. Recently, Andersson et al. [11] showed that a premature stop codon in the *DMRT3* gene has a major effect on the pattern of locomotion in horses with possible implications on racing performance.

Due to the long generation interval, among others, genetic improvement in horses is relatively slow and costly. The application of genetic markers and genomic selection schemes to improve physical performance of horses appears, therefore, to hold great potential. So, the objective of this study was to perform GWAS to identify DNA polymorphism (SNPs), chromosomal regions and positional candidate genes associated with performance trait (SI) in the racing Quarter Horse breed.

2. Materials and Methods

2.1. Animals and Phenotypes Data

All animal procedures used in the experiment were in accordance with the Brazilian legislation for animal welfare.

Whole blood samples (5 mL) were collected by puncture of the external jugular vein in the neck region of 112 racing Quarter Horses. This resource consisted of samples from 17 males and 95 females, which were born to 48 stallions and 101 mares, from five properties located in São Paulo state, Brazil. Horses were born between 1985 and 2007 and registered at the Brazilian Association of Quarter Horse Breeders (ABQM).

Speed index (SI) points were provided by the Sorocaba Jockey Club, São Paulo state, Brazil. A horse receives a SI number every time it races at a Brazilian and American Association Quarter Horse recognized track. Because each racetrack is different, race times vary from track to track, and the SI system was developed as a way to compare horses and races run at different tracks. Every year each racetrack creates its own SI table, which is derived from the average of three wins (top three times) for each of the last three consecutive years in each distance, and the average value of those nine times is equivalent to an SI of 100, creating a scale. Therefore, SI points are integers and vary with time, the level of hundredths of a second, and are

adjusted by the distance traveled in the race. Details about SI calculation were described previously [12]. The maximum SI (the best score of the animal obtained during his entire life) was used for the analysis. This option was due to the lack of SI information for all years that a particular animal ran and because the maximum SI is highly correlated with the average SI available in the database (r = 0.8762).

2.2. Genotyping Assay and Quality Control

Genomic DNA was extracted from white blood cells using the Illustra Blood GenomicPrep Mini Spin kit (GE Healthcare, USA) according to manufacturer instructions. DNA integrity was analyzed using 0.8% agarose gel electrophoresis. DNA concentration and purity were determined in a NanoDrop ND-2000 spectrophotometer (Thermo Scientific, USA). Concentrations of DNA were adjusted to 40-60 ng/μL and samples were genotyped with the Equine SNP50 BeadChip (Illumina, Inc., USA) using the HiScan system (Illumina, Inc., USA) at the Faculty of Agricultural and Veterinary Sciences, UNESP, Jaboticabal, São Paulo, Brazil. This array contains 54,602 SNPs selected from the EquCab2.0 SNP Collection database, with a mean density of one SNP per 43.2 kb.

Genome Studio 2011.1 software (Illumina, Inc., USA) was used for SNP calling and quality control of the genotype data for individuals and SNPs. For individuals, call rate, heterozygosity, and gender estimation were determined as part of quality control analysis. No individual was excluded. In addition, agreement between four replicates and parentage concordance (allele sharing) between four stallion/progeny pairs and three stallion/mare/progeny trios were evaluated. With respect to SNPs, those with low genotyping accuracy (cluster separation <0.3; call frequency <0.9; *p*-value <1 x 10^{-3} for Hardy-Weinberg equilibrium), with a minor allele frequency (MAF) <0.05 and located at chromosome X were removed from further analysis. Application of these filters resulted in the exclusion of 13,815 SNPs from the dataset. Thus, the final dataset consisted of 40,787 autosomal SNPs.

2.3. Genome-Wide Association Study

Association analyses were performed with 40,787 SNPs for SI using Qxpak.5 [13] and fitting one SNP at a time. Qxpak.5 relies on the theory of mixed models, performing a likelihood ratio test with each SNP in turn, testing the model with SNP *versus* the model without the SNP, against a chi-square distribution with one degree of freedom. Heritability estimates and the effect of each SNP were estimated using the mixed model (Equation 1) by Qxpak.5 program, which included sex (2 levels), stud (5 levels) and racing distance (9 levels) as fixed effect and animal and SNP as a random effect.

$$y_{ij} = X\beta + Zu + S_k s_{jk} + e_{ij} \text{ (Eq. 1)},$$

where y_{ij} represents the vector of phenotypic observations from the *i*-th horse (*i* = 1 to 112) at the *j*-th trait (*j* = SI); X is the incidence matrix relating fixed effects in *β* with observations in y_{ij}; Z is the incidence matrix relating random additive polygenic effects in *u* with observations in y_{ij}; S_k is the vector of genotypes for the *k*-th SNP across all animals; s_{jk} represents the additive effect of the *k*-th SNP on the *j*-th trait, and e_{ij} is the vector of random residual effects.

The genetic relationship between animals was estimated from genotypes, used to build a genomic relationship matrix (G), using the same methodology describe by VanRanden [14] (Equation 2):

$$G = \frac{(M-P)(M-P)'}{2\Sigma_{j=1}^{m} p_j(1-p_j)} \text{ (Eq. 2)},$$

where M is an allele-sharing matrix with m columns (m=40,787 SNPs) and n rows (n=112 individuals), and P is a matrix containing the observed frequency of the second allele (p_j), expressed as $2p_j$. M_{ij} was 0 if the genotype of individual *i* for SNP *j* was homozygous for the first allele, was 1 if heterozygous, or 2 if the genotype was the other homozygous state.

To deal with the issue of multiple-testing, Q-value was calculated with the

package for R (Version 2.10) [15]. The Q-value provides a measure of each feature's significance, i.e. of each SNP in case this study, automatically taking into account the fact that thousands are simultaneously being tested. Therefore, the Q-value is an extension of a quantity called the ‘‘false discovery rate’’[15].

The percentage of the phenotypic variance accounted by each SNP was estimated according to Equation 3:

$$\%V_i = 100\left(\frac{2\,p_i q_i \hat{a}_i^2}{\sigma_e^2+\sigma_a^2}\right) \text{(Eq. 3)},$$

where $p_i$and q_i are the allele frequencies for the *i*-th SNP estimated across the entire population, a_i is the estimated additive effect of the *i*-th SNP on SI, and σ_e^2 and σ_a^2 are the REML estimate of the residual and the additive variances for the trait, respectively.

2.4. Gene Annotation

According to McCue et al. [16] the linkage disequilibrium (LD) of the Quarter Horse decline within the first 50 - 100 Kb. Therefore, the window considered for gene annotation was 100Kb upstream and downstream to each significant SNP ($P < .0001$). Positions of SNPs and genes were according to NCBI Reference Assembly [17], based on the latest assembly of the horse genome sequence (EquCab2.0). These genes were annotated using the Gene Ontology database and considering the species *Homo sapiens*.

3. Results and Discussion

Quarter horse breed is renowned for its functional versatility, presenting abilities in several sporting activities, like cutting, reining, lasso, western riding and racing. Of these activities, racing performance has the highest economic importance to the equine industry.

According to Gaffney and Cunningham [18], athletic phenotypes, as racing, are influenced by many environmental factors, as environment itself, management, and training; however, it has long been accepted that there are underlying genetic

factors that influence a horse's athletic performance capabilities. In this study, the SI trait mean (standard deviation) was of 98 (± 9.37), with addictive and residual variances equal 15.20 and 43.66, respectively. The estimated heritability obtained in this population for this trait was 0.25, indicating that the racing performance in terms of SI is heritable. This value was greater than that reported by Corrêa and Mota [19], which found heritability varying from of 0.14 to 0.19 over a range of distances (301, 365 and 402 meters). These different results can be due to the low number of horses of the present study compared to the large number of observations (7775) of SI and due the different effects included in the model used by Corrêa and Mota [19]. However, according to these authors, the low heritability value is not a definite impediment to genetic progress of this particular trait because SI can be measured several times in a horse's career, allowing increase the selection precision and respective response. In a practical sense, SI is the tool available to measure and compare racing potential and so it must be used to inform selection of superior horses.

We performed a GWAS for SI trait using a mixed-model, fitting genomic relationship data, fixed effects and SNP alleles. The genomic relationship matrix (G) was calculated to obtain more accurate kinship between the animals, using identical by state (IBS) information.

A total of eight significant SNPs ($P < .0001$) distributed on ECA 2, 4, 10, 18 and 27 were found and served as markers for genomic regions mined for candidate genes associated with SI. Estimated effects and proportion of phenotypic variance explained were reported for each of the eight SNPs (Table 1). This *P*-value corresponds to Q-value of 0.25 after correction for multiple testing. These associations found are indicative of possible QTL given the limited sample size and the fact that each of these SNP accounted for an important proportion of the phenotypic variance, which ranged from 14.37 to 25.47% (Table 1). The distribution of the *P* values result of the GWAS performed is shown as a Manhattan plot in Figure1 and the quantile-quantile-plot (Q-Q) in Figure 2.

Among significant results, 4 SNPs – rs69498497 (ECA4), rs68879046 (ECA10), rs69127140 (ECA18) and rs69517755 (ECA27) – are interesting findings, since genes within of window of 100 Kb, can be functionally related to athletic

performance. These genes are *GRM8, GRIK2, NEB, ANK1* and *KAT6A* with active roles in nervous system, muscle development and skeletal growth. The SNPs rs69498497 (ECA4), rs69127140 (ECA18) and rs69517755 (ECA27) are positioned within of the genes *GRM8*, *NEB* and *ANK1*, respectively. Although these three SNPs are positioned in non-coding regions, they can be affecting the expression of these genes or be in LD with causal SNPs.

In our study, SNP associations on ECA 4 and ECA 10 were observed implicating the genes: *GRM8* (glutamate receptor, metabotropic 8) and *GRIK2* (glutamate receptor, ionotropic, kainate 2), respectively. Both genes encoding glutamate receptors. The *GRM8* gene belongs to group III of the metabotropic glutamate receptors that are a family of G protein-coupled receptors. The *GRIK2* gene belongs to the kainate family of ionotropic glutamate receptors, which are composed of four subunits and function as ligand activated ion channels. Both metabotropic and ionotropic receptors are activated by L-glutamate, which is the major excitatory neurotransmitter in the central nervous system [17]. Glutamate receptors mediate the majority of excitatory neurotransmission in the mammalian brain and are activated in a variety of normal neurophysiologic processes. According to Carobrez [20] experiments conducted over the last few years have proven the glutamate involvement in neural development, synaptic plasticity, learning, memory and others. Behavioural and learning processes in the horse are likely to influence not only equine athletic success but also the usefulness of the horse as a domesticated species [21]. According to Marinier and Alexander [22] the athletic activities performed by horses, such as racing, show jumping, dressage and carriage work require specialized training and good learning ability on the part of the horse.

On ECA 18, other SNP association was found and the *NEB* (nebulin) emerged as a candidate gene. This gene plays a pivotal role in skeletal muscle contractility by specifying thin filament length and function [23]. Mutations in this gene are a frequent cause of nemaline myopathy, the most common non-dystrophic congenital myopathy characterized by muscle weakness and the presence of rod shaped structures in the muscle fibers [24]. A study that measured nebulin content in muscle before, and 24 h after, exercise in humans, suggest that nebulin may be damaged after exercise [25]. The authors concluded that the loss of these proteins may have important

implications for the mechanisms regulating the adaptive response of skeletal muscle resistance to exercise. Given the biology of nebulin, it is possible to speculate that variations in the *NEB* gene impact racing ability by having an effect on muscle contractility and resistance to exercise. In a previous study, a different region of ECA 18 was implicated. Hill et al. [7] identified a genetic variant associated with the *MSTN* gene on ECA 18, which could be used as a genetic marker for prediction of best racing distance in Thoroughbreds. Tozaki et al. [8] reported that SNP significance in the genomic regions neighboring the *MSTN* gene suggest that a gene in this region of ECA18 or the *MSTN* gene itself was one of the major genes affecting the racing performance of Japanese Thoroughbred racehorses.

Other candidate gene emerging from this GWAS is *ANK1*. The *ANK1* (ankyrin 1, erythrocytic) belongs to a family of proteins that link the integral membrane proteins to the underlying spectrin-actin cytoskeleton, which in muscle forms a highly complex network of inter-myofibril connections as well as links between myofibrils and the sarcolemma and the sarcoplasmic reticulum [26,27]. A role for ankyrins in the localization of proteins participating in the regulation of intracellular calcium homeostasis in striated muscles was initially proposed by Tuvia et al. [28]. A group of striated muscle–specific isoforms of the *ANK1* gene (ANK1.5, ANK1.6, and ANK1.7) are selectively localized on the sarcoplasmic reticulum membrane, with which they are associated through a hydrophobic sequence located at their NH_2-terminal region [29]. The sarcoplasmic reticulum, which is specialized in regulation of intracellular homeostasis, storage, release, and reuptake of calcium, maintains a highly organized relationship of the relaxation–contraction cycle of striated muscles [30]. According to Tee and Peppelenbosch [31] the different stages of muscle development and their phenotypic reaction to strain and exercise are under the control of different ankyrin repeat domain containing proteins, and accordingly their expression at different stages of muscle development seems to be tightly regulated. Due to ankyrin role as structural proteins forming a fundamental component of the cytoskeleton and your importance for muscle contraction and physiology [31], *ANK1* could impact on race performance.

Other gene deserves to be highlighted: *KAT6A* (K(lysine) acetyltransferase 6A). This gene may act as a transcriptional coactivator for *RUNX2* (runt-related

transcription factor 2), which is a member of the RUNX family of transcription factors and encodes a nuclear protein with a Runt DNA-binding domain [17]. The *RUNX2* is essential for skeletal development and plays crucial roles in the processes of both intramembranous and endochondrial ossification, which includes chondrocyte maturation, vascular invasion into the cartilage, and bone formation by osteoblasts [32]. *RUNX2* is the main regulator of chondrocyte hypertrophic differentiation, and an abnormality of chondrocyte development and maturation can lead to altered endochondral ossification, i.e. to condition of osteochondrosis [33]. According to Distl [34] diseases of the locomotor system are the most important causes of reduced performance and premature retirement of horses. Racehorses start training very young and the stress on immature bones and joints can cause breakdown and unsoundness [35]. Therefore, a racehorse must have good conformation to have speed over shorts distances and to stay sound under the stress of training and running at top speed. According to Hill et al. [7], a major characteristic contributing to the ability of a Thoroughbred to perform well in short distances (sprint races) is the extent and maturity of the skeletal musculature.

Finally, it is believed that *COX4I2* and *PDK4* genes associated with Thoroughbred racing performance [9,10], were not detected in the sample of animals studied here possibly because they are related to muscle energy aerobic metabolism, which is typical in Thoroughbred long distance racing. On the other hand, the Quarter Horses racing are of short distance and the muscle energy metabolism is predominantly anaerobic.

4. Conclusions

This study allowed the identification of DNA polymorphism (SNPs), chromosomal regions and positional candidate genes associated with performance trait (SI) in the racing Quarter Horse. Genes identified as novel candidates for racing ability in this study – *GRM8, GRIK2, NEB, ANK1* and *KAT6A* – are within and in the vicinity of SNP associated to SI. Importantly, these genes have biological function that could affect those horse attributes that form a superior animal athlete. After genotyping of independent populations for validation of the SNP associations

reported here, sequencing of these candidate genes should be done to identify causal SNPs.

Acknowledgments

The authors thank the horse farms for providing the Quarter Horses and also thank the invaluable contribution of the late Professor Dr. Marcilio Dias Silveira da Mota.

The present research was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo and fellowships were granted by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Bolsista da CAPES - Proc.n°13842-12-9).

References

[1] Evans JW. Horses: a guide to selection, care and enjoyment. New York: Freeman and Company;1996.

[2] Scott M. Musculoskeletal Injuries in Nonracing Quarter Horses. Veterinary Clinics of North America: Equine Pract 2008;24:133-52.

[3] Xu M, Bi Y, Cui B, Hong J, Wang W, Ning G. The New Perspectives on Genetic Studies of Type 2 Diabetes and Thyroid Diseases. Curr Genomics 2013;14:33-48.

[4] Porto-Neto LR, Jonsson NN, D'Occhio MJ, Barendse W. Molecular genetic approaches for identifying the basis of variation in resistance to tick infestation in cattle. Vet Parasitol 2011;180:165-72.

[5] Hawken RJ, Zhang YD, Fortes MRS, Collis E, Barris WC, Corbet NJ, et al. Genome-wide association studies of female reproduction in tropically adapted beef cattle. J Anim Sci 2012;90:1398-410.

[6] Binns MM, Boehler DA, Lambert DH. Identification of the myostatin locus (MSTN) as having a major effect on optimum racing distance in the Thoroughbred horse in the USA. Anim Genet 2010;41:154-58.

[7] Hill EW, McGivney BA, Gu J, Whiston R, MacHugh DE. 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 2010a;11:552.

[8] Tozaki T, Miyake T, Kakoi H, Gawahara H, Sugita S, Hasegawa T, et al. A genome-wide association study for racing performances in Thoroughbreds clarifies a candidate region near the MSTN gene. Anim Genet 2010;41:28-35.

[9] Gu J, Machugh DE, McGivney BA, Park SDE, Katz LM, Hill EW. 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 Vet J 2010;42:569-75.

[10] Hill EW, Gu J, McGivney BA, MacHugh DE. Targets of selection in the Thoroughbred genome contain exercise-relevant gene SNPs associated with elite racecourse performance. Anim Genet 2010b;41:56-63.

[11] Andersson LS, Larhammar M, Memic F, Wootz H, Schwochow D, Rubin C-J, et al. Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. Nature 2012;488:642-46.

[12] Corrêa MJM. Avaliação genético-quantitativa de características de desempenho em cavalos da raça Quarto de Milha [Quantitative genetic evaluation of performance traits in Quarter Horse]. Botucatu (Brazil): MSc, Dissertation, College of Veterinary and Animal Science São Paulo State University; 2005.

[13] Perez-Enciso M, Misztal I. Qxpak.5: Old mixed model solutions for new genomics problems. BMC Bioinformatics 2011;12:202.

[14] VanRaden PM. Efficient methods to compute genomic predictions. Journal of Dairy Science 2008;91:4414-23.

[15] Storey JD, Tibshirani R. Statistical significance for genome wide studies. Proc Natl Acad Sci U S A 2003;100:9440-5.

[16] McCue ME, Bannasch DL, Petersen JL, Gurr J, Bailey E, Binns MM, et al. A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. PLoS Genetics 2012;8(1):e1002451.

[17] NCBI 2014. National Center for Biotechnology Information. Available at: http://www.ncbi.nlm.nih.gov. Accessed April 15, 2014.

[18] Gaffney B, Cunningham EP. Estimation of genetic trend in racing performance of thoroughbred horses. Nature 1988;332:722-24.

[19] Corrêa MJM, Mota MDS. Genetic evaluation of performance traits in Brazilian Quarter Horse. J Appl Genet 2007;48:145-51.

[20] Carobrez AP. Transmissão pelo glutamato como alvo molecular na ansiedade. Rev Bras Psiquiatr 2003;25:52-8.

[21] Murphy J, Arkins S. Equine learning behavior. Behav Processes 2007;76:1-13.

[22] Marinier SL, Alexander AJ. The use of a maze in testing learning and memory in horses. Appl Anim Behav Sci 1994;39:177-82.

[23] Ottenheijm CAC, Buck D, Winter JM, Ferrara C, Piroddi N, Tesi C, et al. Deleting exon 55 from the nebulin gene induces severe muscle weakness in a mouse model for nemaline myopathy. Brain 2013;136:1718-31.

[24] North KN, Laing NG, Wallgren-Pettersson C. Nemaline myopathy: current concepts. J Med Genet 1997;34:705-13.

[25] Trappe TA, Carrithers JA, White F, Lambert CP, Evans WE, Dennis RA. Titin and nebulin content in human skeletal muscle following eccentric resistance exercise. Muscle Nerve 2002;25:289-92.

[26] Nelson WJ, Lazarides E. Goblin (ankyrin) in striated muscle: Identification of the potential membrane receptor for erythroid spectrin in muscle cells. Proc Natl Acad Sci U S A 1984;81:3292-96.

[27] Bagnato P, Barone V, Giacomello E, Rossi D, Sorrentino V. Binding of an ankyrin-1 isoform to obscurin suggests a molecular link between the sarcoplasmic reticulum and myofibrils in striated muscles. J Cell Biol 2003;160:245-53.

[28] Tuvia S, Buhusi M, Davis L, Reedy M, Bennett V. Ankyrin-B Is required for intracellular sorting of structurally diverse Ca2+homeostasis proteins. J Cell Biol 1999;147:995-1007.

[29] Zhou D, Birkenmeier CS, Williams MW, Sharp JJ, Barker JE, Bloch RJ. Small, membrane-bound, alternatively spliced forms of ankyrin 1 associated with the sarcoplasmatic reticulum of mammalian skeletal muscle. J Cell Biol 1997;136:621-31.

[30] Sorrentino V. Molecular determinants of the structural and functional organization of the sarcoplasmic reticulum. Biochim Biophys Acta 2004;1742:113-8.

[31] Tee JM, Peppelenbosch MP. Anchoring skeletal muscle development and disease: the role of ankyrin repeat domain containing proteins in muscle physiology. Crit Rev Biochem Mol Biol 2010;45:318-30.

[32] Komori T. Runx2, a multifunctional transcription factor in skeletal development. J Cell Biochem 2002;87:1-8.

[33] Jeffcott LB, Henson FMD. Studies on growth cartilage in the horse and their application to aetiopathogenesis of dyschondroplasia (Osteochondrosis). Vet J 1998;156:177-92.

[34] Distl O. Genomics of skeletal disorders. In: Chowdhary BP, editor. Equine Genomics, Texas: Texas A&M University; 2013, p.187-98.

[35] Thomas HS. The Horse Conformation Handbook. North Adams: Storey Publishing;2005.

Table 1. Minor allele frequency (MAF), effect, proportions of the phenotypic variance explained (%V) and genes for single-nucleotide polymorphisms (SNPs) associated (*P*<0.0001) with speed index

SNP	ECA: Position (Mb)	MAF	Effect	%V	Genes
rs68593122	2: 97634986	0.20	6.120	20.08	-
rs68593141	2: 97642160	0.20	5.841	17.98	-
rs69498497	4: 82402787	0.09	-9.603	25.47	*GRM8, MIR592*
rs68879046	10: 52188059	0.39	4.232	14.45	*GRIK2*
rs69127140	18: 33671391	0.18	5.351	14.37	*NEB, RIF1*
rs69134578	18: 36917870	0.22	5.444	17.57	-
rs69194990	18: 39128343	0.16	-6.356	18.09	*CCDC148*
rs69517755	27: 3522223	0.24	5.278	17.53	*ANK1, KAT6A*

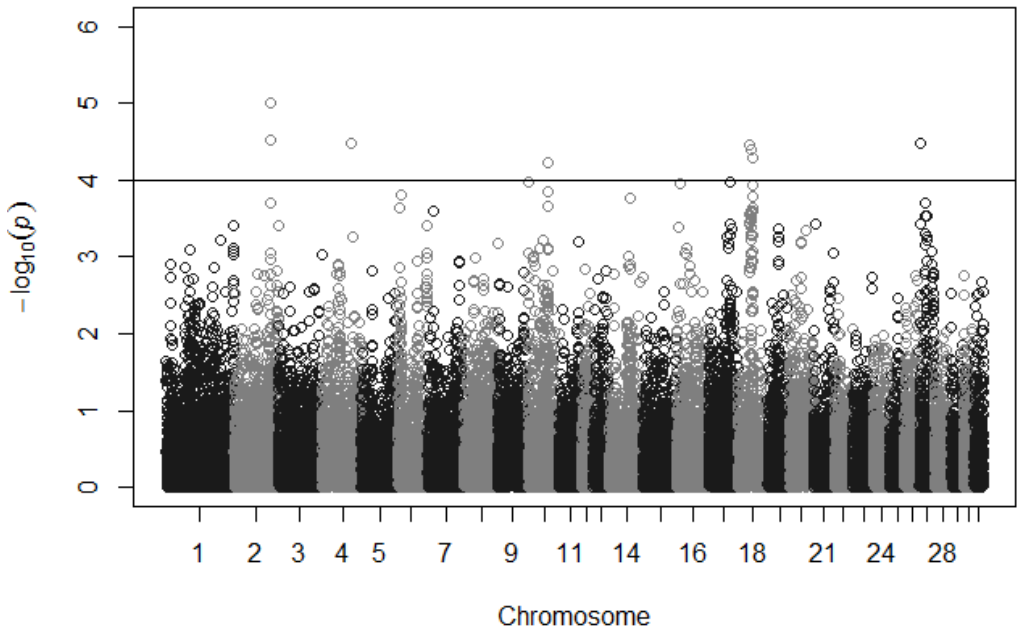

Fig. 1. Manhattan plot of all SNPs for speed index. The log inverse *P*-values estimated for each polymorphism is plotted in the y-axis. Chromosome number is plotted in the x-axis. Horizontal line indicates the threshold $P < .0001$.

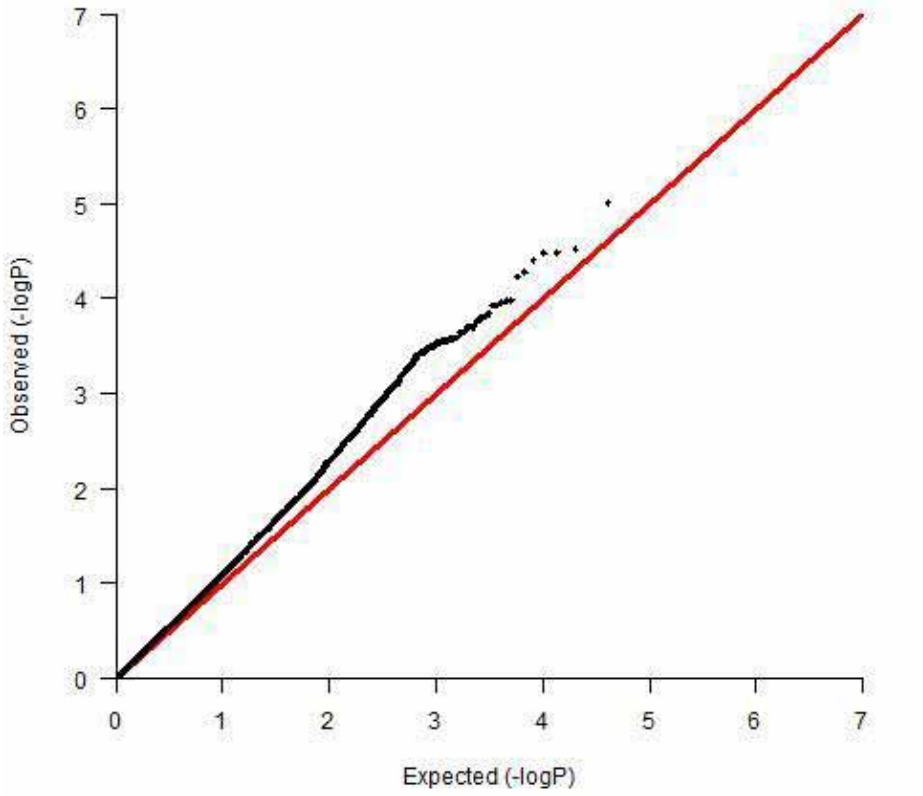

Fig. 2. A quantile–quantile (QQ) plot for speed index displays the relationship between the experimentally observed *P* values (vertical axis) to the expected *P* values of a null distribution (horizontal axis).

CAPÍTULO 5 – A genome wide association study for morphometric traits in Quarter Horse

Camila Tangari Meira[a], Michel Marques Farah[a], Marina R.S Fortes[b], Stephen S. Moore[b], Guilherme L. Pereira[a], Josineudson Augusto II V. Silva[c], Marcílio Dias S. da Mota[c], Rogério Abdallah Curi[c]

[a] Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal, SP 14884-900, Brazil.
[b] Queensland Alliance for Agriculture and Food Innovation, Centre for Animal Science, The University of Queensland, Brisbane, QLD, Australia.
[c] Faculdade de Medicina Veterinária e Zootecnia, Universidade Estadual Paulista, Botucatu, SP 18618-000, Brazil.

Abstract

A genome-wide association study for morphometric traits was conducted in 184 Quarter Horses, 120 from a racing population and 64 from a cutting population, which were genotyped using the Illumina EquineSNP50 chip. Association analysis was performed with 42,058 single-nucleotide polymorphisms (SNPs) (after quality control) using Qxpak.5 software. The following traits were measured: weight, rump and body length. These morphometric traits are important for the best performance in race and cutting events. For weight, three SNPs associated ($P < .0001$) were found on chromosomes (*Equus caballus* autosomes [ECA]) 2 and 3. For rump length, eight SNPs associated ($P < .0001$) were found on ECA 2, 3, 6, 7, 9, 21, and 26. On ECA 3 and ECA 8, two SNPs were associated ($P < .0001$) with body length. So, a total of 13 important chromosomal regions were identified with Q-values of 0.53 (SNPs for W), 0.40 (SNPs for RL) and 0.99 (SNPs for BL). Positional and functional candidate genes emerging from this study were *WWOX* and *AAVPR1A*. Further studies are required to confirm these associations in other populations.

Keywords: *Equus caballus*, genetics, quantitative traits, single-nucleotide polymorphism

1. Introduction

Breed of great economic expression, specially due to its versatility in different equestrian events, the Quarter Horse is subdivided into different populations according to skills resulting from distinct selection objectives. The racing population explores the sprinting ability of the animals over shorts distances, whereas the cutting population is used in functional tests, exploring abilities such as agility, obedience and cow sense.

The morphology of horses presents functional relationship with the performance of the animals in their various activities, such as racing, reining, barrel racing, cutting, etc. Many times the differences in performance exist owing to adequacy or inadequacy morphometric traits. According to Meira et al. [1], racing horses are heavier and taller and present greater body lengths and perimeters than cutting horses.

The availability of an equine SNP array gave rise to genome wide association studies (GWAS). Recent studies applying genomic approaches reported to conformation traits have been conducted. Signer-Hasler et al. [2], Tetens et al. [3] and Meira et al. [4] reported the identification of quantitative trait loci (QTL) for height at withers on *Equus caballus* autosomes (ECA) 3 and ECA 9 in Franches-Montagnes horses, on ECA 3 in Germam Warmblood horses and on ECA 9 in Quarter Horse, respectively. Studies with body size also had been conducted. Makvandi-Nejad et al. [5] identify four loci on ECA 3, 6, 9 and 11 explaining the vast majority of variation in horse size. Metzger et al. [6] investigated associations of polymorphisms of the candidate gene *LCORL* with the development of body size in horses. In this study we aimed to perform GWAS to identify chromosomal regions and positional candidate genes associated with other important morphometric traits in Quarter Horse.

2. Materials and Methods

For the analyses, a total of 188 Quarter Horses of both sexes born between 1985 and 2009, including 120 racing horses and 68 cutting horses, registered at the Brazilian Association of Quarter Horse Breeders (ABQM), were studied. All

experimental procedures were conducted in accordance with the Brazilian legislation on animal welfare.

The following physical traits were measured according to Torres and Jardim [7]: weight (W), rump length (RL) and body length (BL). The measurements were performed by the same person with a tape measure and measuring stick, always on the right side of the animal, with the horse standing with front and rear legs perpendicular to the ground. For genotyping, a 5-mL sample of whole blood was collected from each animal by puncture of the left jugular vein in the neck region into vacuum tubes containing 7.5 mg ethylenediaminetetraacetic acid.

The racing horses (18 males and 102 females), born to 48 stallions and 107 mares, belonged to five farms in the countryside of the State of São Paulo, Brazil. The cutting horses (26 males and 42 females), born to 44 stallions and 64 mares, belonged to three other farms in the countryside of the State of São Paulo. In both populations, full sibs were avoided.

Genomic DNA was extracted from the blood samples of Quarter Horses using the Illustra Blood Genomicprep Mini Spin kit (GE Healthcare, Little Chalfont, Buckinghamshire, UK) according to manufacturer instructions. DNA integrity was analyzed by 0.8% agarose gel electrophoresis and DNA was quantified with a NanoDrop ND-2000 spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The DNA concentration in the samples was adjusted to 40-60 ng/μL.

SNPs were genotyped on the HiScan system (Illumina Inc.) using the Illumina Equine SNP50 BeadChip at the Faculty of Agricultural and Veterinary Sciences, UNESP, Jaboticabal, São Paulo, Brazil. The chip contains 54,602 SNPs uniformly distributed across the entire genome of 15 horse breeds. The SNPs are distributed across the 31 autosomes and X chromosome. The mean interval between SNPs is 43,200 bp. This content is derived from the EquCab2.0 SNP Collection compiled by the Broad Institute's Equine Genome Sequencing Project, which identified more than 940,000 SNPs in Arabian, Andalusian, Akhal-Teke, Icelandic Pony, Standardbred, Thoroughbred, and Quarter Horses.

The quality of individual and SNP genotype data was investigated using the Genome Studio program, version 2011.1 (Illumina Inc.). For individuals, call rate,

heterozygosity, and gender estimation were determined. Animals with a call rate < 0.95, heterozygosity of ± 3 standard deviations from the mean, and errors in gender estimation were excluded from the sample. In addition, agreement between four replicates and parentage concordance (allele sharing) between four stallion/progeny and three stallion/mare/progeny pairs were evaluated. With respect to the quality of SNP genotypes in the whole population, SNPs located on the X chromosome, with low genotyping quality (cluster separation < 0.3), a call frequency < 0.9, a minor allele frequency (MAF) < 0.05 (including fixed alleles), and a Hardy-Weinberg *P*< 0.001 were excluded. Quality control led to the exclusion of four cutting animals from the sample and 12,544 SNPs. Thus 42,058 autosomal SNPs for 184 animals were used in the analysis.

Genome wide association studies (GWAS) were performed with the 42,058 SNPs for each of the three traits (W, RL and BL) separately using Qxpak.5 and fitting one SNP at a time [8]. Qxpak.5 relies on the well-known theory of mixed models. It performs a likelihood ratio test with every SNP in turn, testing the model with the SNP *versus* the model without the SNP, against a chi-square distribution with 1 degree of freedom.

Heritability estimates and the effect of each SNP were estimated using the mixed model (Equation 1), which included sex (2 levels), population (2 levels) and stud (8 levels) as fixed effects, age at the time of trait measurement as a linear covariate and animal and SNP as a random effect.

$$y_{ij} = X\beta + Zu + S_k s_{jk} + e_{ij} \quad \text{(Eq.1)},$$

where y_{ij} represents the vector of phenotypic observations from the *i*-th horse (*i* = 1 to 184) at the *j*-th trait (*j* = W, RL and BL, respectively); X is the incidence matrix relating fixed effects in β with observations in y_{ij}; Z is the incidence matrix relating random additive polygenic effects in *u* with observations in y_{ij}; S_k is the vector of genotypes for the *k*-th SNP across all animals; s_{jk} represents the additive effect of the *k*-th SNP on the *j*-th trait, and e_{ij} is the vector of random residual effects.

The genetic relationship between animals was estimated from genotypes, used to build a genomic relationship matrix (G), using the same methodology

describe by VanRanden [9] (Equation 2):

$$G = \frac{(M-P)(M-P)'}{2\sum_{j=1}^{m} p_j(1-p_j)} \quad \text{(Eq. 2)},$$

where M is an allele-sharing matrix with m columns (m=42,058 SNPs) and n rows (n=184 individuals), and P is a matrix containing the observed frequency of the second allele (p_j), expressed as $2p_j$. M_{ij} was 0 if the genotype of individual i for SNP j was homozygous for the first allele, was 1 if heterozygous, or 2 if the genotype was the other homozygous state.

To account for multiple tests, Q-value was calculated with the package for R (Version 2.10) [10]. The percentage of the genetic variance accounted by the each SNP was estimated according to Equation 3:

$$\%V_i = 100\left(\frac{2p_i q_i \hat{\alpha}_i^2}{\sigma_g^2}\right) \quad \text{(Eq. 3)},$$

where p_i and q_i are the allele frequencies for the *i*-th SNP estimated across the entire population, α_i is the estimated additive effect of the *i*-th SNP on the trait in question, and σ_g^2 is the restricted maximum likelihood (REML) estimate of the (poly)genetic variance for the trait.

According to McCue et al. [11] the linkage disequilibrium (LD) of the Quarter Horse decline within the first 50 - 100 Kb. Therefore, the window considered for gene identification was 100Kb upstream and downstream to each significant SNP (*P* < .0001 for any of the traits). Genes with biologically interesting findings were highlighted. Positions of SNPs and genes were according to NCBI Reference Assembly [12], based on the latest assembly of the horse genome sequence (EquCab2.0).

3. Results and Discussion

Descriptive statistics of the traits are reported in Table 1. Conformation traits are reported in the literature to be of moderate to high heritability [13]. In this study,

the heritabilities estimated ranged from 0.17 to 0.51, indicating additive genetic variability and the potential for improvement through genetic selection of these morphometric traits in Quarter Horse.

We performed a GWAS for conformation traits (W, RL, BL) using a mixed-model, fitting genomic relationship data, fixed effects and SNP alleles. The genomic relationship matrix was used to obtain more accurate relationship between animals, using identical by state (IBS) information. Fitting a G matrix allowed to estimate relationships between animals of the 2 different populations that were unrelated according to pedigree records.

A total of 13 significant SNP associations ($P < .0001$) were found for the traits, being three SNPs on ECA 2 and 3 associated with weight, eight SNPs on ECA 2, 3, 6, 7, 9, 21 and 26 associated with rump length and two SNPs on ECA 3 and 8 associated with body length (Fig. 1). This *P*-value corresponds to Q-values of 0.53 (SNPs for W), 0.40 (SNPs for RL) and 0.99 (SNPs for BL). These association values are indicative of a possible true association, given the small sample size and the fact that each of these SNP accounted for an important proportion of the genetic variance, which ranged from 17.8 to 54.4% for all the traits (Table 2). Among significant results, the SNP rs68488737 (ECA3) that was associated with W and RL and the SNP rs68714893 (ECA6) associated with RL point to interesting candidate genes: *WWOX* and *AVPR1A*. These genes are biologically interesting findings, because they can be functionally related to important traits in the Quarter Horses as detailed below. The SNP rs68488737 (ECA3) is positioned in non-coding region of the *WWOX*, however it can affect the gene expression or be in LD with causal SNPs.

In our study, two SNP associations at ECA 3 were observed implicating that the gene *WWOX* (WW domain containing oxidoreductase) was associated with W and RL. This gene, according to Aqeilan et al. [14] contributes to bone formation through regulation of *RUNX2* activity in osteoblasts, which encodes a nuclear protein essential for osteoblastic differentiation and skeletal morphogenesis [12]. In a previous study, a different region of ECA 3 was implicated. Tetens et al. [3] found a single major QTL for height at withers explaining ~18% of the phenotypic variance mapped to ECA3 in German Warmblood horses, indicating the *LCORL*/*NCAPG* locus as a strong candidate underlying this QTL. Differences in studied breeds and traits

may account for these different results. A fine mapping exercise of these regions of ECA3 would contribute to resolving this QTL.

The association with RL on ECA6 implicated a new candidate gene: *AVPR1A* (arginine vasopressin receptor 1A). The protein encoded by *AVPR1A* acts as receptor for arginine vasopressin (AVP). Vasopressin is a powerful vasoconstrictor and an important component in the control of blood pressure during exercise in horses [15]. According to McKeever et al. [16], exercise causes an increase in plasma concentration AVP that is correlated with the duration and intensity of the exercise in horses and man. What would be the direct effect of *AVPR1A* in RL is difficult to speculate. From the above described gene function, it is logical to assume that *AVPR1A* could affect exercise performance. Further investigation of the association reported here could provide insight for the molecular links between RL and exercise performance.

In conclusion, genomic regions on ECA 2, 3, 6, 7, 8, 9, 21 and 26 were associated with morphometric traits in Quarter Horse. Genes mapped to these regions and thus positional candidates for weight, rump and body length emerged from this GWAS. However, most of these genes have no obvious function related to the studied traits. Positional and functional candidate genes from this study are *WWOX* and *AVPR1A*. Further studies are required to confirm these SNP associations and candidate genes in other populations and horse breeds.

Acknowledgements

The authors thank the horse farms for providing the Quarter Horses and the financial support of São Paulo Research Foundation (Fapesp) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes). They also thank the invaluable contribution of the late Professor Dr. Marcilio Dias Silveira da Mota.

References

[1] Meira CT, Curi RA, Silva JAV, Corrêa MJM, Oliveira HN, Mota MDS. Morphological and genomic differences between cutting and racing lines of Quarter Horses. J Equine Vet Sci 2013;33:244-49.

[2] Signer-Hasler H, Flury C, Haase B, Burger D, Simianer H, Leeb T, et al. A genome-wide association study reveals loci influencing height and other conformation traits in horses. PLoS One 2012;7(5):e37282.

[3] Tetens J, Widmann P, Kühn C, Thaller G. A genome-wide association study indicates *LCORL*/*NCAPG* as a candidate locus for withers height in German Warmblood horses. Anim Genet 2013;44:467–71.

[4] Meira CT, Fortes MRS, Farah MM, Porto-Neto LR, Curi RA, Moore SS, et al. A genome-wide association study for height at withers in racing quarter horse. Proc Assoc Advmt Anim Breed Genet 2013;20:420-23.

[5] Makvandi-Nejad S, Hoffman GE, Allen JJ, Chu E, Gu E, Chandler AM, et al. Four loci explain 83% of size variation in the horse. PLoS One 2012;7(7):e39929.

[6] Metzger J, Schrimpf R, Philipp U, Distl O. Expression levels of *LCORL* are associated with body size in horses. PLoS One 2013;8(2):e56497.

[7] Torres AP, Jardim WR. Criação do cavalo e de outros equinos. 3rd ed. São Paulo: Nobel; 1992.

[8] Perez-Enciso M, Misztal I. Qxpak.5: Old mixed model solutions for new genomics problems. BMC Bioinformatics 2011;12:202.

[9] VanRaden PM. Efficient methods to compute genomic predictions. J Dairy Sci 2008;91:4414-23.

[10] Storey JD, Tibshirani R. Statistical significance for genome wide studies. Proc Natl Acad Sci USA 2003;100:9440-5.

[11] McCue ME, Bannasch DL, Petersen JL, Gurr J, Bailey E, Binns MM, et al. A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. PLoS Genet 2012;8(1):e1002451.

[12] NCBI. National Center for Biotechnology Information. Available at: http://www.ncbi.nlm.nih.gov. Accessed February 7, 2014.

[13] Zamborlini LC, Bergmann JAG, Pereira CS, Fonseca AG, Carneiro ASR. Estudo genético quantitativo das medidas lineares da raça Mangalarga Marchador –I. Estimativas dos fatores de ambiente e parâmetros genéticos. Rev Bras Ci Vet 1996;3:33-7.

[14] Aqeilan RI, Hassan MQ, Bruin A, Hagan JP, Volinia S, Palumbo T, et al. The WWOX tumor suppressor is essential for postnatal survival and normal bone metabolism. J Biol Chem 2008;283:21629-39.

[15] McKeever KH, Hinchcliff KW. Neuroendocrine control of blood volume, blood pressure and cardiovascular function in horses. Equine Vet J 1995;18:77-81.

[16] McKeever KH, Hinchcliff KW, Schmall LM, Lamb DR, Muir WW. Changes in plasma renin activity, aldosterone, and vasopressin, during incremental exercise in horses. Am J Vet Res1992;53:1290-3.

Table 1. Summary statistics and heritability (h^2) estimates for weight (kilogram), rump length (centimeter) and body length (meter) of the Quarter Horse.

Parameter	Weight (Kg)	Rump length (cm)	Body length (m)
N (animals)	184	184	184
Mean (± STD) (racing animals)	538.97 (± 42.05)	62.03 (±0.03)	1.80 (± 0.06)
Mean (± STD) (cutting animals)	450.69 (± 46.59)	55.17 (±0.02)	1.63 (± 0.07)
σ^2_g	823.2840	0.0004	0.0006
h^2	0.4975	0.5164	0.1710

STD, standard deviation; σ^2_g, restricted maximum likelihood (REML) estimate of the (poly)genetic variance for the trait

Table 2. Minor allele frequency (MAF), effect (Ef), proportion of the additive genetic variance explained (%V) and genes for single-nucleotide polymorphisms (SNPs) associated ($P < .0001$) with weight, rump length and body length of the Quarter Horse.

SNP	ECA: Position (Mb)	MAF	Weight		Rump length		Body length		Genes
			Ef	%V	Ef	%V	Ef	%V	
rs68626382	2:33.5	0.17	24.30	19.9	-	-	-	-	*EIF4G3, HP1BP3, SH2D5, KIF17, LOC102148467, DDOST, PINK1, LOC102148800, CDA,LOC102148657*
rs68488737	3:26.7	0.36	-19.60	21.7	-	-	-	-	*WWOX*
rs68559987	3:29.1	0.42	-17.70	18.6	-	-	-	-	*BCMO1, GAN, LOC102150591, CMIP*
rs68614112	2:103.7	0.06	-	-	-0.03	22.9	-	-	*ANKRD50*
rs68562545	2:98.9	0.10	-	-	0.02	21.6	-	-	-
rs68488737	3:26.7	0.36	-	-	-0.01	19.7	-	-	*WWOX*
rs68714893	6:79.4	0.38	-	-	0.01	19.9	-	-	*AVPR1A*
rs68762127	7:34.1	0.09	-	-	-0.02	24.1	-	-	*LOC100072341, LOC100630076, LOC100630107, LOC102148118*
rs68802762	9:18.5	0.24	-	-	0.01	17.8	-	-	*MCMDC2, SGK3, LOC102147850, C9H8orf44, VCPIP1*
rs69315059	21:4.1	0.26	-	-	-0.01	19.1	-	-	*LOC102148385, LOC100054881, LOC102148426*
rs69402054	26:29.9	0.42	-	-	-0.01	20.5	-	-	*IFNAR2, IL10RB, LOC102150829, IFNAR1*
rs68625621	3:15.0	0.40	-	-	-	-	0.02	54.4	-
rs68795337	8:32.8	0.36	-	-	-	-	-0.02	52.3	*PTPRM*

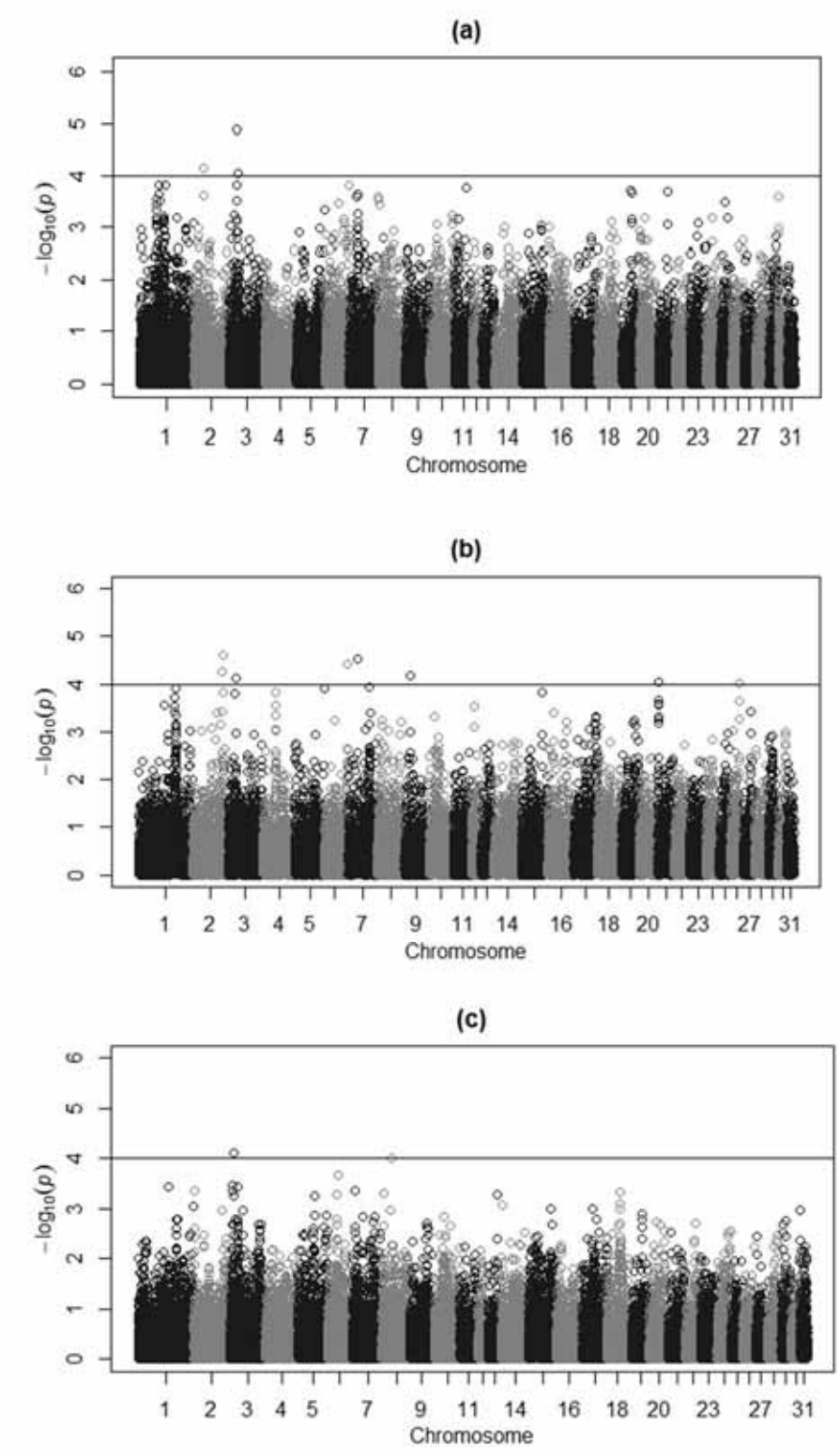

Fig.1. Manhattan plot of *P* value for (a) weight, (b) rump length and (c) body length. The log inverse *P* values estimated for each polymorphism is plotted in the y-axis. Chromosome number is plotted in the x-axis. Horizontal line indicates the threshold $P < .0001$

CAPÍTULO 6 – Considerações finais

Nos trabalhos aqui apresentados foram identificadas diferenças morfológicas e genômicas entre as linhagens de corrida e trabalho da raça Quarto de Milha, regiões do genoma selecionadas de forma divergente nos animais de corrida em relação aos de trabalho, conhecidas como assinaturas de seleção, e regiões cromossômicas, genes e SNPs candidatos associados à característica de desempenho (índice de velocidade ou “speed index”) na linhagem de corrida e com características morfométricas (peso, comprimento da garupa e comprimento do corpo) na raça como um todo.

Trata-se de estudos iniciais que envolvem modernas técnicas de genética molecular aplicadas à raça Quarto de Milha do Brasil e que, futuramente, deverão trazer grande contribuição para a avaliação genética dos animais e para o melhoramento da raça.

Os SNPs encontrados pelo estudo de ampla associação do genoma (GWAS) com a característica índice de velocidade, localizados com base no EquCab2.0 nos cromossomos (Chr) 2: 97634986 e 97642160; Chr 4: 82402787; Chr 10: 52188059; Chr 18: 33671391; 36917870; 39128343 e Chr 27: 35222223 estão próximos às regiões de assinaturas de seleção identificadas pela aplicação simultânea das metodologias REHH e F_{ST} localizadas nos Chr 2: 87453500-89546885; Chr 4: 96192056-983156500; Chr 10: 65461613-67583618; Chr 18: 30519448-32573824 e Chr 27: 14932202-16989870.

Dessa forma, estas regiões do genoma da linhagem de corrida da raça Quarto de Milha, com seus genes e polimorfismos, tornam-se candidatas para futuros estudos de associação entre genótipos e desempenho.