Livestock Science 174 (2015) 1–9

Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci

Prospection of genomic regions divergently selected in cutting line of Quarter Horses in relation to racing line

Natalia Andrea R. Beltrán [a], Camila T. Meira [a], Henrique N. de Oliveira [a], Guilherme L. Pereira [a], Josineudson Augusto II V. Silva [b], Marcílio Dias S. da Mota [b], Rogério A. Curi [b,*]

[a] Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal, São Paulo 14884-900, Brazil
[b] Faculdade de Medicina Veterinária e Zootecnia, Universidade Estadual Paulista, Botucatu, São Paulo 18618-970, Brazil

ARTICLE INFO

Article history:
Received 21 January 2014
Received in revised form
5 November 2014
Accepted 18 January 2015

Keywords:
Horse
Fixation index
Extended haplotype homozygosity
Selection signature
SNP array

ABSTRACT

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

1. Introduction

The Quarter Horse is one of the most versatile horses in the world. Today, the American Quarter Horse Association (AQHA) has registered more than 5 million horses in 75 countries (ABQM, 2012). In Brazil, about 358,000 Quarter Horses are registered with the Brazilian Association of Quarter Horse Breeders (ABQM), which have a significant impact on national agribusiness (ABQM, 2012). Quarter Horses have developed into different lines as a result of different selection objectives (Evans, 1996), including cutting and racing lines. The cutting line is destined for functional tests, exploring skills such as agility, obedience and cow sense, characteristics that are important for cattle management in the field. The racing line is characterized by its great sprinting speed over short distances on flat tracks.

Recent technological advances in the area of molecular genetics have culminated in the development of so-called high-density single nucleotide polymorphism (SNP) genotyping chips, which permit analysis of the genetic structure of different populations of various domestic animal species (Edea et al., 2013; Petersen et al., 2014), estimation of the degree of genetic diversity between populations (McCue et al., 2012; Petersen et al., 2013) and, recently, the identification and localization of genomic regions undergoing positive selection (Akey et al., 2010; Stella et al., 2010; Qanbari et al., 2011). There are two basic methods used to identify changes caused by positive selection across the genome: extended haplotype homozygosity (EHH) statistic (Sabeti et al., 2002), which is used to identify these regions within populations,

* Corresponding author. Tel.: +55 14 38802984; fax: +55 14 38117187.
E-mail address: rogcuri@fmvz.unesp.br (R.A. Curi).

http://dx.doi.org/10.1016/j.livsci.2015.01.011
1871-1413/

and the fixation index (F_{ST}) (Cockerham, 1969; Weir and Hill, 2002; Wright, 1951), which is used to identify regions under selection considering multiple populations (Simianer et al., 2010). The basic concept of the EHH test proposed by Sabeti et al. (2002) is that the frequency of a new mutation under positive selection pressure will increase rapidly, carrying with it neighboring alleles in linkage disequilibrium (LD), an event known as selective sweep (Maynard Smith and Haigh, 1974). Since selection carries an allele in a specific high-frequency haplotype faster than an allele broken up by recombination, high-frequency haplotypes will be longer than expected under neutral evolution (Simianer et al., 2010). In this respect, long haplotypes showing an elevated frequency and homozygosity may indicate regions of the genome that have been selected recently. F_{ST} statistic measures the relationship between pairs of alleles of polymorphisms within a subpopulation relative to the total population (Cockerham, 1969; Weir and Hill, 2002; Wright, 1951) and can be interpreted as a measure of dispersion of allele frequencies among groups relative to the variation expected in the population from which these groups derived.

The objective of the present study was to identify genomic regions divergently selected in cutting line of Quarter Horses in relation to racing line using high-density SNP genotyping arrays and the relative extended haplotype homozygosity (REHH) test, an extension of EHH analysis, and F_{ST} as statistical methods. These genomic regions were then used for gene annotation in order to identify those that might have been important during formation of the cutting line and that could be used in the future for the development of tools to select the best animals.

2. Materials and methods

2.1. Animals

A total of 188 Quarter Horses of both sexes, born between 1995 and 2009, raised in the State of São Paulo, Brazil, and registered with ABQM, were used. Of these, 68 horses were from the cutting line and 120 from the racing line. Animals of the cutting line (26 males and 42 females) were progeny of 44 stallions and 64 mares corresponding to an average of 1.5 offspring/stallion and 1.06 offspring/mare. These animals belonged to 42 breeders and were raised on three farms in the State of São Paulo. Animals of the racing line (18 males and 102 females) were progeny of 48 stallions and 107 mares, corresponding to an average of 2.5 offspring/stallion and 1.1 offspring/mare. These animals belonged to 57 breeders and were raised on five farms in the State of São Paulo. The presence of full siblings in the two lines was avoided. All animal procedures were conducted in accordance with the Brazilian legislation on animal well-being.

2.2. Genotyping of SNPs

Genomic DNA was extracted from blood samples of Quarter Horses using the Illustra Blood GenomicPrep Mini Spin kit (GE Healthcare, USA) according to manufacturer instructions. DNA integrity was analyzed on 0.8% agarose gel and DNA was quantified in a NanoDrop ND-2000 spectrophotometer (Thermo Scientific, USA). The DNA concentration in the samples was adjusted to 40–60 ng/μL. SNPs 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.

2.3. Quality control of genotype data

Quality control of genotype data was performed for individuals and SNPs using the Genome Studio 2011.1 program (Illumina, Inc., USA). For individuals, the call rate, heterozygosity and gender label were determined. Animals with a call rate < 0.95, heterozygosity with ± 3 standard deviations of the mean, and gender-labeling errors were excluded from the sample. In addition, concordance of four replicates and concordance of kinship relations (allele sharing) between four stallion/offspring pairs and three stallion/mare/offspring sets were evaluated.

For quality control filtering of SNPs considering all individuals and each line, markers located on the X chromosome and those with low genotyping quality (cluster separation < 0.3), a call frequency < 0.9, minor allele frequency (MAF) < 0.05 (including fixed alleles), and a P-value $< 1 \times 10^{-3}$ for Hardy–Weinberg equilibrium were excluded.

2.4. Analysis within the cutting line

EHH can be defined as the probability that two randomly chosen core haplotypes are homozygous for the entire interval from the core region to a certain locus (Sabeti et al., 2002). However, since recombination rates can vary substantially across different regions of the genome, a higher EHH value may be obtained due to low recombination rates in a certain region and not necessarily because of recent positive selection. In this respect, relative extended haplotype homozygosity (REHH) statistic corrects the EHH value observed in a core haplotype for the average level of EHH values of all significant core haplotypes on the same chromosome.

After the process of SNP filtering, only the dataset of the cutting line was used for the EHH test. The FastPHASE software with a standard parameter configuration was used to reconstruct haplotypes and to infer the linkage phases of SNPs on each chromosome. This software implements methods to estimate missing genotypes and to reconstruct haplotypes based on SNP data from unrelated individuals using a cluster-based model for haplotypes (Scheet and Stephens, 2006). The Sweep 1.1 program (Sabeti et al., 2002), which implements REHH statistic, was used to detect putative selection signatures in the genome of the cutting line. First, core regions were identified using the algorithm suggested by Gabriel et al. (2002), which defines a pair of SNPs to be in strong LD if the upper 95% confidence limit of D' is between 0.98 and 0.7. 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. It was decided that core regions should have a minimum of 3 SNPs and a maximum of 20.

The EHH test was applied at a distance of about 1 Mb on both sides from the core region. To determine the significance of REHH values, i.e., EHH values corrected for each chromosome, the Sweep 1.1 program divided the haplotypes into 20