Genomic regions associated with performance in racing line of Quarter Horses

G.L. Pereira[a,*], L.A. Chardulo[b], J.A.IIV. Silva[b], R. Faria[a], R.A. Curi[b]

[a] 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
[b] 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

ARTICLE INFO

ABSTRACT

Quarter Horses breed for rancing are able to run short distances faster than any other breed. Although in Brazil the effective number of animals in the racing line is relatively smaller compared to the other lines, its economic importance is substantial. Despite economic and scientific interest in this athletic trait, few efforts have been made to better understand the genetic and physiological mechanisms underlying this trait. The objective of this study was to perform genome-wide association studies (GWAS) in Quarter Horses of the racing line using equine single nucleotide polymorphism (SNP) genotyping chips for prospecting chromosome regions, genes and polymorphisms related to maximum speed index. Genotypic and phenotypic data from 305 horses genotyped with the 54 k or 65 k chip were used. Additional 187 not genotyped animal with SI record were added in the study, for a total of 492 animals and 620 horses in the pedigree used for the GWAS analysis GWAS was performed by the single-step genomic best linear unbiased prediction (ssGBLUP) method using two approaches: ssGWAS1 in which only SNP effects are recalculated at each iteration, and ssGWAS2 in which SNP effects are recalculated from updated genomic estimated breeding values (GEBVs) at each iteration. Twenty-one regions that explained more than 1% of the total genetic variance (gVar) in the maximum speed index were identified by ssGWAS1 and 12 by ssGWAS2. More than 40% of gVar was explained by these regions in ssGWAS1 and about 30% in ssGWAS2. Among chromosomes that explained more than 1% of genetic variance, five were common to both methods (equine autosomal chromosomes 3, 10, 15, 22, 25). The capacity to capture the largest number of small-effect quantitative trait loci was greater for ssGWAS1 than ssGWAS2. A total of 108 genes were identified with the first approach and 59 with the second approach. The present results show that the racing performance of Quarter Horses, given by the maximum speed index, is a quantitative trait and that no major genes exist. In general, the genes identified by enrichment analysis are related to the neuromotor system, bone and tendon development, hormonal responses, muscle contraction, and energy and sterol metabolism. These functions are considered important for the trait studied, corroborating these genes as candidates for racing performance in the Quarter Horse breed.

1. Introduction

Quarter Horses breed for racing are able to run short distances faster than any other breed. In Brazil the effective number of animals in the racing line is relatively small compared to other lines, however its has substantial economic importance. Despite economic and scientific interest in this line and its racing aptitude, few efforts have been made to better understand the genetic mechanisms underlying this trait (Meira et al., 2014a, 2014b; Pereira et al., 2016b, 2016a).

Individual (Gu et al., 2010; Hill et al., 2010b) and genome-wide (Shin et al., 2015) association studies between DNA polymorphisms and the racing performance of horses have been conducted mainly on Thoroughbred horses. However, Pereira et al. (2016b, 2016a) showed that some markers appropriate for the thoroughbred are not suitable for the identification of animals with superior racing performance in the Quarter Horse breed. In an attempt to identify genes and DNA polymorphisms specific for the racing line of Quarter Horses related to speed and performance in horse races, Meira et al. (2014a, 2014b) investigated selection signatures and performed a genome-wide association study (GWAS). However, because of the small number of animals used, the results obtained were not conclusive in determining genetic variants associated with racing performance in the breed.

Genomic studies using single nucleotide polymorphism (SNP) panels can be used for both the detection of quantitative trait loci (QTL) by GWAS and the prediction of genomic estimated breeding values (GEBVs) used in genomic selection (GS) (Goddard and Hayes, 2007; Resende et al., 2012). Conventionally, GWAS seek to find associations between loci and phenotypic traits using hypothesis tests in order to detect statistically significant effects. In order to avoid the occurrence of false-positive results, GWAS can be penalized for high false-negative rates due to the use of very strict cut-off values (Resende et al., 2012). Thus, only markers linked to large-effect QTLs tend to be identified. Genomic selection, which is performed using Bayesian methods and genomic best linear unbiased preditction (GBLUP), is applied to the whole genome too, but simultaneously, and is based on the estimation and prediction of effects instead of hypothesis tests (Resende et al., 2012). The single-step GBLUP method (ssGBLUP) was initially described for use in GS to improve prediction accuracy by simultaneously including phenotypic information from genotyped and non-genotyped individuals in the pedigree (no need to create pseudo-data) however the original ssGBLUP was still limited by its assumption of an infinitesimal model, al adapted it to not assume equal variance for all SNP marker effects (Wang et al., 2012).

The process leads to gains in the detection of genomic regions that are important for the trait of interest by demonstrating the presence of large-effect QTLs or larger number of small-effect QTLs

The objective of this study was to perform GWAS in Quarter Horses of the racing line using equine SNP genotyping chips and the ssGBLUP method for prospecting QTLs and genes that could be related to performance.

2. Material and methods

2.1. Animals, blood collection, and DNA extraction

For this study, 360 Quarter Horses of the racing line, 78 males and 282 females, registered with the Brazilian Association of Quarter Horse Breeders, were used. These animals were born to 83 stallions and 249 mares, with an average offspring/stallion ratio of 4.3 and offspring/mare ratio of 1.4. The animals belonged to 159 breeders and were housed at the Sorocaba Jockey Club (Sorocaba, SP, Brazil) and on 25 farms located in towns of the State of São Paulo, Brazil. Blood samples were collected from 120 of these animals in 2011 and from the remaining 240 animals in 2015 (i.e. in both years, the animals randomly sampled between the farms and the Jockey Club). The collection from and, consequently, the presence of full-sibs in the sample was avoided. All animal procedures were conducted according to Brazilian guidelines on animal welfare (Protocol 157/2014) and were approved by the Ethics Committee on Animal Use (FMVZ, Unesp, Botucatu, SP, Brazil).

DNA was extracted from blood aliquots using the Illustra Blood GenomicPrep Mini Spin Kit (GE Healthcare, USA) according to manufacturer instructions. After extraction, DNA was quantified in a small volume NanoDrop Lite spectrophotometer (Thermo Scientific, USA) and its integrity was analyzed on 0.8% agarose gel. The samples were adjusted to the working concentration necessary for SNP genotyping (40–60 ng/μL) and stored at $-20\,°C$.

2.2. Quality control of genotyping and genotype imputation

The panels used in this study were the Equine SNP50 BeadChip (Illumina, Inc., USA), which contains 54,602 markers (54k), and the Equine SNP70 BeadChip (Illumina, Inc., USA), the commercial successor of the EquineSNP50 BeadChip, which contains 65,157 markers (65 k). The first and second panel share 45,703 markers, i.e., the 54k and 65k panels possess 8899 and 19,454 unshared markers, respectively. Among the 360 horses used in this study, 120 were genotyped in 2011 with the Equine SNP50 BeadChip. The remaining DNA samples (n = 240) were genotyped in 2015 with the Equine SNP70 BeadChip.

Table 1

Number of markers of the Equine SNP50 BeadChip and Equine SNP70 BeadChip and number of SNPs excluded according to each quality control criterion used.

Criteria used for exclusion of samples/SNP	SNP50 (n)	SNP70 (n)
Call rate < 0.9 (samples)	4	7
SNP located on the X chromosome	3 223	3 411
Call rate < 0.9 (SNP)	148	504
Hardy-Weinberg equilibrium – $p < 1 \times 10^{-5}$	4 519	4 638
Remaining samples	116	233
Remaining SNPs	46,712	56,604

The two arrays were read with the HiScan system (Illumina, Inc., USA). Quality control (QC) analysis of the genotyping information for individuals analyzed with one or the other array, 54k (n = 120) and 65k (n = 240), was performed using the snpStats package (Clayton, 2015) of the R v.3.3.1 program (R Core Team, 2016). For individuals, animals with a call rate < 0.9 were excluded from the dataset. In addition, SNPs located on the X chromosome, SNPs with a call rate < 0.9, and SNPs with a p-value $< 1 \times 10^{-5}$ for Hardy-Weinberg equilibrium were excluded (Table 1).

The information used for two-step genotype imputation (i.e., from 54 k to 65 k and from 65 k to 54 k) refers to the number of individuals and SNPs obtained after QC analysis. Genotype imputation was carried out with the FImpute v.2.2 program (Sargolzaei et al., 2014).

For independent two-step imputation, the 65 k panel was used first as the reference set and the 54 k panel as the imputed set. For this purpose, all markers present in the 65 k panel and only shared markers of the SNP50 panel (45 k) were considered. The unshared markers (9 k) were again added after imputation. Thus, the samples genotyped with the low-density panel (54 k) started to have also the exclusive genotypes of the 65 k panel. The same procedure was applied in the second step in which only the panels used as the reference and imputed set were inverted. Thus, the samples genotyped with the 65k array started to have also the exclusive genotypes of the 54 k panel. Next, the two imputed panels were ordered and concatenated, forming a set of genotyped individuals for all markers available in the 54 k and 65 k panels. The QC criterion minor allele frequency (MAF) < 0.05 was applied to the panel of genotypes already imputed in two steps, which was used in the GWAS.

After QC of genotyping and imputation of exclusive genotypes, analyses were carried out to determine the population stratification in the dataset used. For this purpose, genomic inflation (λ_{gc}) and principal component analysis was performed using the Plink 1.9 software (Purcell et al., 2007). The λ_{gc} values were obtained by simple association between the genotypes and uncorrected phenotypes and subsequently between genotypes and corrected phenotypes (EBV) obtained for each animal. The EBVs were obtained with the blupf90 program (Misztal, 2014), which included parental information (relationship matrix) and adjustments for fixed effects following the statistical model to be used for ssGWAS. The substructures were visualized by plotting the first two principal components, PC1 and PC2, using the adegenet package (Jombart and Ahmed, 2011) of the R v.3.3.1 program (R Core Team, 2016).

2.3. Phenotypic information

The performance data of racing Quarter Horses, updated until December 2015, were obtained from the Department of Statistics or from the online database of the Sorocaba Jockey Club (http://www.jcsorocaba.com.br/indice-de-velocidade/) and the Equibase online database (http://www.equibase.com/homehorseplayer.cfm). The performance record given by the maximum speed index (SImax) obtained throughout the competition history of each animal was used. Records of SI were available for most of the individuals used, while only information on SImax was found for others. The mean of SImax was 97.3

(± 8.12) and the minimum and maximum values to SImax was 72 and 120, respectively. Since the mean SI shows a high correlation with SImax (r = 0.8762) (Pereira et al., 2016a), the latter was chosen to avoid the loss of performance data. In addition, to ensure the reliability of SImax, only individuals with three or more races in their competition history were considered for GWAS.

The variance components were estimated by Gibbs sampling for Bayesian computation using the gibbs1f90 and postgibbs1f90 programs (Misztal, 2014). The samples of the variance components were obtained from a chain of 500,000 cycles using a burn-in period of 50,000 cycles and a thinning interval of 50 cycles, totaling 9000 samples. From these samples, the posterior means of the variance components, heritability and posterior standard deviations were calculated for each parameter using the R v.3.3.1 program (R Core Team, 2016). The following statistical model was used: $y = X\beta + Za + e$, where y is the vector of observations; β is the vector of fixed effects (sex, age and year or race referring to the observation of SImax); a is the vector of direct additive genetic effects; X is the incidence matrix; Z is the incidence matrix of direct additive genetic random effects, and e is the vector of residual effects. The relationship matrix contained the information of 620 animals; 492 of these animals had records of SImax. The estimates of genetic and residual variances was 6.809 (± 3.609) and 50.74 (± 4.848), respectively. The heritability estimated was 0.12 (± 0.06).

2.4. Genome-wide association study

To find genomic regions that explain an important part of the total genetic variance in SImax, the effects of SNPs and variances explained by SNP windows were estimated by ssGWAS as described by Wang et al. (2012). The analyses were carried out with the package of blupf90 programs (Misztal et al., 2014; Aguilar et al., 2011).

The statistical model described in item 2.3 was applied to ssGWAS. The variances of a and e are given by:

$$\text{var}\begin{bmatrix} a \\ e \end{bmatrix} = \begin{bmatrix} H\sigma_a^2 & 0 \\ 0 & I\sigma_e^2 \end{bmatrix},$$

where σ_a^2 and σ_e^2 are the total and residual genetic variance, respectively, and H is the combined pedigree-genomic relationship matrix as described by Aguilar et al. (2010). Its inverse is:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix},$$

where A is the numerator of the pedigree matrix for all animals; A_{22} is the numerator of the pedigree matrix for genotyped animals, and G is the genomic relationship matrix. Matrix G, constructed according to VanRaden et al. (2008), assumes the allele frequencies of the population under analysis and adjusts for compatibility with A_{22} (Chen et al., 2011).

The prediction equation of SNP effects that uses the weighted genomic relationship matrix, G^*, proposed by Wang et al. (2014) was given by: $\hat{u} = \lambda\ DZ'G^{*-1}\hat{a}_g = DZ'[ZDZ']^{-1}\hat{a}_g$, where λ is a ratio of variance or a normalization constant, and $\hat{a}_g$ is the GEBV, with $\lambda = \frac{\sigma_u^2}{\sigma_a^2} = \frac{1}{\sum_{i=1}^{M} 2p_i(1-p_i)}$, where M is the number of SNPs; p is the frequency of the second allele of the i-th marker; D is the diagonal of the weight matrix for the variances of SNPs (D = I for GBLUP); Z is the matrix that relates the genotypes to each locus, and G^* is the weighted genomic relationship matrix. Estimates of SNP effects were used to estimate the individual variance of each marker (Zhang et al., 2010), as follows: $\sigma_{u,i}^2 = u_i^2 2p_i(1-p_i)$.

For GWAS, the SNP effects were derived from GEBVs or EBVs of genotyped individuals using the methodology and algorithm described by Wang et al. (2014), as follows: 1) $t = 0, D_{(t)} = I; G_{(t)}^* = ZD_{(t)}Z'\lambda$. 2) Compute $\hat{a}_g$ by ssGBLUP. 3) Calculate $\hat{u}_{(t)} = \lambda D_{(t)}Z'G_{(t)}^{*-1}\hat{a}_g$. 4) Calculate $G_{(t+1)}^* = \hat{u}_{i(t)}^2 2p_i(1-p_i)$ for all i as described by Zhang et al. (2010).

5) Normalization by $D_{(t+1)} = \frac{tr(D_{(0)})}{tr(D_{(t+1)}^*)} D_{(t+1)}^*$. 6) Calculate $G_{(t+1)}^* = ZD_{(t+1)}Z'\ \lambda$. 7) $t = t + 1$. 8) Stop or return to step 2 or step 3. By t representing the number of interactions and i the i-th SNP, a "re-weighted matrix G^* is applied when returning to step 3 (ssGWAS1) in which only the SNP effects are recalculated. By returning to step 2 (ssGWAS2), both the effects of the animal and of the SNP are recalculated. Five iterations were used in the different scenarios (ssGWAS1 and ssGWAS2), as well as SNP windows of 250 kb, 500 kb and 1 Mb.

The percentage of genetic variance explained by the i-th region was calculated as follows:

$$\frac{Var(s_i)}{\sigma_a^2} \times 100\% = \frac{Var(\Sigma_{j=1}\ z_j\hat{u}_j)}{\sigma_a^2} \times 100\%$$

where s_i is the breeding value of the i-th region, which consists of windows of 1 Mb with different numbers of SNPs; σ_a^2 is the total genetic variance; Z_j is the vector of genetic content of the j-th SNP for all individuals, and u_j is the marker effect of the j-th SNP in region i.

2.5. Gene annotation and functional analysis

SNP windows of 1 Mb that explained a proportion of genetic variance (gVar) higher than 1% in GWAS were used for the prospection of genes in the chromosome regions considered to be biologically important and identification of genes for inclusion in the enrichment analysis. The positions and names of each gene were determined based on the latest sequence assembly of the equine genome (EquCab2.0) using the BioMart package (Durinck et al., 2009) of the R software (R Core Team, 2016). Enrichment analysis was performed with the topGO R package (Alexa and Rahnenfuhrer, 2016) using the genes present in SNP windows with gVar higher than 1% (target genes) and a universe of genes located in SNP windows with gVar higher than 0.1%. The standard algorithm of the topGO package (weight01 option) was used for this purpose and Fisher's test was applied at a level of significance of 5%.

3. Results

3.1. Data used and population stratification

Forty-four of the 349 individuals undergoing genotyping QC were excluded because of the lack of SImax records (n = 12) or because there were fewer than three races in their history (n = 32), thus remaining 305 genotyped animals with reliable performance records. In addition, records from 187 Quarter Horses that were not genotyped, but that were present in the relationship matrix, were included in the ssGWAS analyses. These individuals had SImax records and more than three races in their history. Thus, SImax records from 492 animals and pedigree information from 620 horses were used for GWAS.

Two-step genotype imputation considerably increased the information of each panel. After QC analysis, the two panels contained 39,664 shared SNPs, with 7 048 and 16,940 exclusive markers for the 54 k and 65 k arrays, respectively. The final density of the imputed panel was 63,652 SNPS without the MAF filter. The gain in the number of markers in relation to the final imputed panel was 26% for SNP50, exceeding the original number of SNPs, and 12% for SNP70. Application of the MAF filter (MAF < 0.05) resulted in the exclusion of 8 538 SNPs. Thus, the imputed panel contained 55,114 SNPs that were effectively used in the association analysis.

Principal component analysis, which used 305 individuals, revealed substructures in the population studied (Fig. 1). However, these substructures were not clearly distinct. The λ_{gc} estimates obtained in this study were 1.11 when uncorrected phenotypes were used and 1.01 when corrected phenotypes (EBVs) were used.
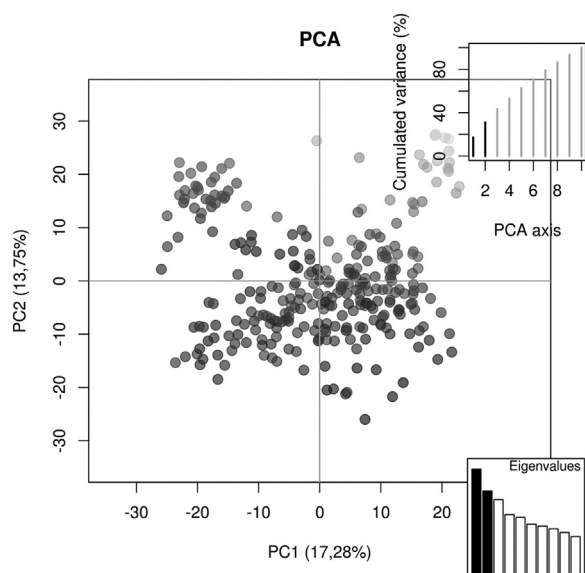
Fig. 1. Graphical representation of the first two principal components (PC) and the percentage of variance explained by each component in parenthesis. The upper right inset shows the % cumulated variance explained by PC1 and PC2 in relation to the first 10 PCs. The lower right inset gives the eigenvalues of the 10 PCs, highlighting PC1 and PC2.

3.2. Genome-wide association

Two approaches were used for GWAS, ssGWAS1 and ssGWAS2, because each approach uses different procedures to obtain SNP effects and, consequently, explained variances, resulting in different capacities to capture large- or small-effect QTLs. Thus, the combined analysis of the two scenarios may lead to a greater gain in information about regions and genes related to the trait of interest. The results were obtained for both methods taking into consideration the third of the five iterations performed since regions of small effect became strongly penalized by the weights after the third iteration, while the opposite occurred for large-effect regions. The mean number of SNPs observed in windows of 250 kb, 500 kb and 1 000 kb, respectively, was 8.8, 13.1 and 25.9 in ssGWAS1 and 9.6, 17 and 28.5 in ssGWAS2

In ssGWAS1 in which only SNP effects were recalculated at each iteration, few alterations in the number of associated regions were observed between the different window sizes (250 kb, 500 kb, and 1 Mb) (Fig. 2). Furthermore, apparently, some regions of small effect were benefitted in larger windows (equine chromosome autosome - ECA 1, 8 and 29). On the other hand, in ssGWAS2 in which the SNP effects were obtained from re-estimated GEBVs at each iteration, alterations were observed in the number of associated regions between the different window sizes (250 kb, 500 kb, and 1 Mb) (Fig. 2). In this respect, regions of small effect were more strongly penalized by the weights in larger SNP windows, while the opposite occurred in large-effect regions.

Considering SNP windows of 1 Mb, 21 regions that explained more than 1% of the total genetic variance in SImax were identified by ssGWAS1 (Table 2) and 12 by ssGWAS2 (Table 3). More than 40% of gVar was explained by these regions in ssGWAS1 and about 30% in ssGWAS2. Among the chromosomes harboring regions that explained more than 1% of genetic variance, five were common to both methods (ECA 3, 10, 15, 22, and 25), five were exclusive to ssGWAS1 (ECA 1, 2, 8, 11, and 16), and four were exclusive to ssGWAS2 (ECA 5, 9, 20, and 30). Taking the two methods into consideration, important regions were described on 14 chromosomes. With respect to the associated regions, a single portion of ECA 16 (SS1_REG18) explained 4.2% of gVar in ssGWAS1 (Table 2), while in ssGWAS2 region SS2_REG12 of ECA 30 explained 7.3% (Table 3). ECA 15 was the chromosome that harbored most regions of importance identified by the two methods,

explaining the highest proportion of gVar (10.67%) in ssGWAS1 when the five most important regions were summed (SS1_REG11 to SS1_REG15) and 6.7% in ssGWAS2 when the three most important regions were summed (SS2_REG05 to SS2_REG07). Considering the results obtained with the two methods, 33 regions of importance were identified, with four regions of ssGWAS1 coinciding with three regions of ssGWAS2 (SS1_REG08 and SS1_REG09 with SS2_REG04; SS1_REG19 with SS2_REG09, and SS1_REG20 with SS2_REG11).

3.3. Gene annotation and functional analysis

In ssGWAS1, 110 genes were found in 19 of the 21 associated regions (Table 4). In ssGWAS2, 59 genes were described in 10 of the 12 SNP windows (Table 5). Eighty-three exclusive genes were identified by the ssGWAS1 approach, 32 were identified by the ssGWAS2 approach, and 27 genes were common to both methods. These 27 genes are located on ECA 10, 22, and 25. Considering the two methods simultaneously, 142 genes were prospected. It is worth noting that approximately 50% of the genes found in ssGWAS2 were also identified in ssGWAS1.

Enrichment analysis revealed four significant terms in ssGWAS1 and 15 in ssGWAS2 (Table 6). In general, both methods (ssGWAS1 and ssGWAS2) did not show highly significant enrichments as observed for traits controlled by large-effect genes. However, the ordering and association of genes with biological processes led to the systematic investigation of functions and pathways.

4. Discussion

In population analysis, the application of mixed models was efficient in correcting bias caused by stratification. Mixed models can be used to model the structure of a population and family, as well as the relationship of non-genotyped or dead individuals (Price et al., 2010). Studying stratification correction methods for nine quantitative traits in humans, Kang et al. (2010) showed that mixed models are more efficient in correcting genomic inflation than other methods. However, principal components need to be included in the model when a high value for genomic inflation (λ_{gc}) is observed in a dataset with a significant degree of relatedness (Price et al., 2010).

In relation to the association analysis, the capacity to capture small-effect QTLs was greater in ssGWAS1 than in ssGWAS2. Thus, as observed by Wang et al. (2012) using simulated data, the first method (ssGWAS1) was efficient in detecting a larger number of regions of small effect compared to the second method (ssGWAS2) which, in turn, was more efficient in capturing a smaller number of large-effect QTLs.

The large number of associated regions identified by the two methods (ssGWAS1 and ssGWAS2) demonstrates the quantitative nature of SImax and, consequently, of racing performance in the Quarter Horse breed, i.e., this trait is influenced by a large number of small-effect genes. Furthermore, low heritability found for SImax indicates an important influence of the environment on the trait, such as the training to which the animal was submitted. Or even, the lack of information about systematic environmental factors (e.g., running, paddock, jockey weight), which would increase the estimate of the residual variance component (not explained by the model). Therefore, given the complexity of the SImax trait, influenced by many genes and environmental factors, a more complex analysis such as a genomic selection (GS) would be more appropriate to stalions and mares selection. Wherein SNPs of interest (i.e. explained a considerable proportion of genetic variance of the SImax) could be used before long to improve the estimates and accuracy of the genetic and genomic values of the individuals considered for selection.

Lower percentages of explained gVar than those obtained here have been reported in many of the studies using the ssGWAS method and involving different species, with maximum values of approximately 6% considering windows of 10–20 SNPs in most cases (Wang et al., 2014;
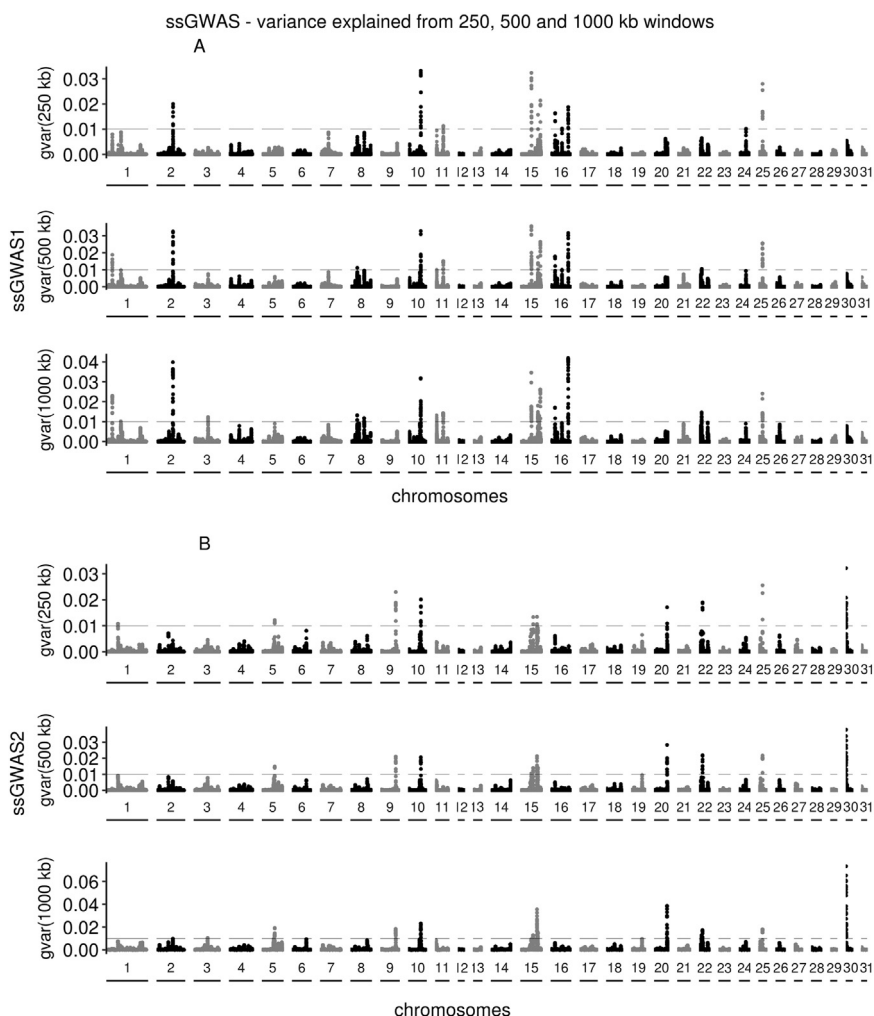
Fig. 2. Manhattan plot of the proportion of total genetic variance (gVar) explained by each SNP window (250 kb, 500 kb, and 1000 kb) in ssGWAS1 (A) and ssGWAS2 (B). The proportion of genetic variance explained by each region is plotted on the y-axis. The chromosomes are given on the x-axis. The dotted line indicates the threshold of 1%.

Table 2

Chromosome location of regions of importance for SImax (gVar > 1%) in the sample of racing Quarter Horses studied, number of SNPs and variance explained by each window in ssGWAS1 considering SNP windows of 1 Mb.

Region	Chr[1]	Position of each window (start to finish) in bp	SNPs	Genetic variance (%)
SS1_REG18[2]	16	77,697,272 − 78,681,644	31	4.19
SS1_REG03[2]	2	69,370,047 − 70,369,774	35	3.98
SS1_REG11[2]	15	45,572,062 − 46,563,225	26	3.46
SS1_REG07[2]	10	57,990,860 − 58,961,242	25	3.19
SS1_REG15[2]	15	87,766,362 − 88,753,736	29	2.62
SS1_REG20	25	17,757,118 − 18,713,444	23	2.41
SS1_REG01	1	20,721,039 − 21,689,628	17	2.30
SS1_REG13	15	76,296,526 − 77,289,510	20	1.80
SS1_REG16	16	16,542,440 − 17,474,810	23	1.71
SS1_REG08	10	58,997,563 − 59,976,633	29	1.68
SS1_REG12	15	46,601,694 − 47,575,761	23	1.50
SS1_REG19	22	10,868,934 − 11,861,413	24	1.47
SS1_REG10	11	35,523,097 − 36,481,549	29	1.43
SS1_REG05	8	30,526,194 − 31,504,074	21	1.32
SS1_REG09	11	3,345,863 − 4,341,072	24	1.31
SS1_REG14	15	82,855,154 − 83,838,250	21	1.29
SS1_REG04	3	65,380,573 − 66,369,357	20	1.23
SS1_REG21	25	18,768,518 − 19,722,839	24	1.20
SS1_REG06	8	62,566,792 − 63,559,486	34	1.18
SS1_REG17	16	17,542,693 − 18,503,480	30	1.17
SS1_REG02	1	61,826,124 − 62,818,065	36	1.01

[1]chromosome; [2] regions common to ssGWAS1 and ssGWAS2.

Table 3

Chromosome location of regions of importance for SImax (gVar > 1%) in the sample of racing Quarter Horses studied, number of SNPs and variance explained by each window in ssGWAS2 considering SNP windows of 1 Mb.

Region	Chr[1]	Position of each window (start to finish) in bp	SNPs	Genetic variance (%)
SS2_REG12[2]	30	138,095 − 1,097,585	34	7.34
SS2_REG08[2]	20	59,429,342 − 60,422,986	24	3.86
SS2_REG06[2]	15	72,126,768 − 73,121,731	35	3.56
SS2_REG04[2]	10	58,735,380 − 59,710,571	34	2.32
SS2_REG02[2]	5	62,937,896 − 63,920,127	27	1.92
SS2_REG03	9	66,879,753 − 67,826,611	24	1.84
SS2_REG11	25	18,073,334 − 19,043,039	26	1.80
SS2_REG10	22	14,927,558 − 15,876,595	25	1.72
SS2_REG07	15	75,795,552 − 76,782,973	30	1.55
SS2_REG05	15	53,973,327 − 54,889,258	28	1.28
SS2_REG09	22	10,600,456 − 11,589,137	28	1.28
SS2_REG01	3	63,351,099 − 64,348,556	27	1.03

[1] Chromosome; [2] regions common to ssGWAS1 and ssGWAS2.

Magalhães et al., 2016; Silva et al., 2017). However, Lemos et al. (2016) reported a window of 10 SNPs to explain approximately 12% of total genetic variance of the trait studied in Nellore cattle. Some authors have used 150 SNPs per window, which explained a maximum of 5% of total genetic variance of the trait (Silva et al., 2017). It should be noted that the genotyping panels used in these studies on cattle and broiler chickens ranged from 60k to high density (~770k) and that the

Table 4

Genes prospected in the chromosome regions of importance for SImax (gVar > 1%) in the sample of racing Quarter Horses studied using the ssGWAS1 method.

Region	Chr[a]	Position of each window (start to finish) in bp	Genes (HGCN symbol)	Genes (n)
SS1_REG01	1	20,721,039 − 21,689,628	DUSP5, ADD3, XPNPEP1	3
SS1_REG02	1	61,826,124 − 62,818,065	ADK, KAT6B, DUSP13, SAMD8,COMTD1, VDAC2	6
SS1_REG03	2	69,370,047 − 70,369,774	MSMO1, KLHL2, TMEM192	3
SS1_REG04	3	65,380,573 − 66,369,357	–	0
SS1_REG05	8	30,526,194 − 31,504,074	EPB41L3	1
SS1_REG06	8	62,566,792 − 63,559,486	RIT2, SYT4	2
SS1_REG07[b]	10	57,990,860 − 58,961,242	NR2E1, SNX3, LACE1, FOXO3, ARMC2[c], SESN1[c], CEP57L1[c]	7
SS1_REG08[b]	10	58,997,563 − 59,976,633	PPIL6[c], SMPD2[c], MICAL1[c], ZBTB24[c], FIG. 4[c], WASF1[c], CDC40, METTL24, DDO	9
SS1_REG09	11	3,345,863 − 04,341,072	RBFOX3, ENGASE, C1QTNF1, CANT1, LGALS3BP, TIMP2, USP36, CYTH1, DNAH17, PGS1, SOCS3, TMEM235, AFMID	13
SS1_REG10	11	35,523,097 − 36,481,549	CA4, ZNHIT3, MYO19, PIGW, GGNBP2, DHRS11, MRM1, LHX1, AATF, ACACA,	10
SS1_REG11	15	45,572,062 − 46,563,225	SPTBN1, C2orf73, ACYP2	3
SS1_REG12	15	46,601,694 − 47,575,761	PSME4, ERLEC1, CHAC2	3
SS1_REG13	15	76,296,526 − 77,289,510	KCNS3, MSGN1, GEN1, SMC6, VSNL1	5
SS1_REG14	15	82,855,154 − 83,838,250	NOL10, ODC1, HPCAL1, RRM2, KLF11, GRHL1, TAF1B, YWHAQ	8
SS1_REG15	15	87,766,362 − 88,753,736	ALLC, COLEC11, RPS7, ADI1, TRAPPC12	5
SS1_REG16	16	16,542,440 − 17,474,810	PDZRN3, PPP4R2, GXYLT2,	3
SS1_REG17	16	17,542,693 − 18,503,480	SHQ1, RYBP	2
SS1_REG18	16	77,697,272 − 78,681,644	–	0
SS1_REG19[b]	22	10,868,934 − 11,861,413	RYBP, TASP1[c], ISM1[c], SPTLC3[c]	4
SS1_REG20[b]	25	17,757,118 − 18,713,444	ZFP37, SLC31A2[c], FKBP15[c], SLC31A1[c], CDC26[c], PRPF4[c], RNF183[c], WDR31[c], BSPRY[c], HDHD3[c], ALAD[c], RGS3[c]	12
SS1_REG21	25	18,768,518 − 19,722,839	ZNF618[c], AMBP[c], KIF12[c], COL27A1[c], AKNA, WHRN, ATP6V1G1, TMEM268, TNFSF15, TNFSF8, TNC	11

[a] Chromosome.
[b] Regions common to ssGWAS1 and ssGWAS2.
[c] Genes common to ssGWAS1 and ssGWAS2.

Table 5

Genes prospected in the chromosome regions of importance for SImax (gVar > 1%) in the sample of racing Quarter Horses studied using the ssGWAS2 method.

Region	Chr*	Position of each window (start to finish) in bp	Genes (HGCN symbol)	Genes (n)
SS2_REG01	3	63,351,099 − 64,348,556	GC, DCK, MOB1B, GRSF1, RUFY3, UTP3, ENAM	7
SS2_REG02	5	62,937,896 − 63,920,127	COL11A1	1
SS2_REG03	9	66,879,753 − 67,826,611	ZNF572, SQLE, KIAA0196, SMCE2, TRIB1	5
SS2_REG04[a]	10	58,735,380 − 59,710,571	ARMC2[b], SESN1[b], CEP57L1[b], PPIL6, SMPD2[b], MICAL1[b], ZBTB24[b], FIG. 4[b], WASF1[b]	9
SS2_REG05	15	53,973,327 − 54,889,258	SLC3A1, PREPL, PPM1B, LRPPRC, ABCG8, ABCG5, DYNC2LI1, PLEKHH2	8
SS2_REG06	15	72,126,768 − 73,121,731	KLHL29	1
SS2_REG07	15	75,795,552 − 76,782,973	–	
SS2_REG08	20	59,429,342 − 60,422,986	–	
SS2_REG09[a]	22	10,600,456 − 11,589,137	TASP1[b], ISM1[b], SPTLC3[b]	3
SS2_REG10	22	14,927,558 − 15,876,595	PLCB1, TMX4, HAO1	3
SS2_REG11[a]	25	18,073,334 − 19,043,039	SLC31A2[b], FKBP15[b], SLC31A1[b], CDC26[b],PRPF4[b], RNF183[b],WDR31[b], BSPRY[b],HDHD3[b], ALAD[b],RGS3[b],ZNF618[b],AMBP[b],KIF12[b],COL27A1[b]	15
SS2_REG12	30	138,095 − 1,097,585	HNRNPU, COX20, DESI2, C1orf101, ADSS, C1orf100, ZBTB18	7

[1]Chromosome.
[a] Regions common to ssGWAS1 and ssGWAS2.
[b] Genes common to ssGWAS1 and ssGWAS2.

windows were chosen based on the number of SNPs and not on size in base pairs, a fact leading to wide variation in windows size. Thus, the average size of most windows was similar to those used here. Using simulated data (two chromosome of 100 Mb each and 3 000 SNPs), Wang et al. (2012) studied the precision of the ssGBLUP method to estimate the effect of QTLs based on the correlation between QTL effects and the sum of *m* adjacent SNP marker effects, where *m* ranged from 1 to 40. The authors observed higher correlations for windows containing 8–16 SNPs (average size of 0.5–1 Mb, respectively), with a strong decrease for windows of 40 SNPs or more (average size > 2.5 Mb).

Most of the genome studies on Quarter Horses did not analyze separately categories with different abilities, i.e., lines existing within the breed. Consequently, studies using only racehorses are sparse, especially those analyzing performance traits. In this respect, i.e., considering the breed as a whole, genome studies have addressed mainly population aspects (Petersen et al., 2013; Meira et al., 2014c; McCue et al., 2012), morphometric characteristics (Meira et al., 2014a), and hereditary diseases (Go et al., 2011; McQueen et al., 2014). However, some of these studies reported important, although indirect, information that could contribute to the understanding of the genetic mechanisms related to racing performance in Quarter Horses. Meira et al. (2014b) identified genes in selection signatures whose functions could cause variations in racing performance in the Quarter Horse breed. These authors found selection signatures on 19 chromosomes, seven of them coinciding with those found in the present study (ECA 1, 2, 3, 5, 10, 11, and 25). However, there were no regions of selection signatures overlapping with those identified here by GWAS. Four SNP windows (gVar > 1%) detected on ECA 5, 10 and 25 were approximately 5 Mb distant from the regions of selection signatures described by these authors.

On ECA 18, in a region of approximately 5.5 Mb in which the

Table 6

Identification (GO.ID) and description (Terms) of Gene Ontology (G) terms significantly enriched after ssGWAS1 and ssGWAS2 in the racing line of Quarter Horses.

Method	GO.ID	Genes	Regions	Terms	Ann.[a]	Sign.[b]	Exp.[c]	*p-value*
ssGWAS1	GO:0006878	*SLC31A2, SLC31A1*	SS1_REG20	*cellular copper ion homeostasis*	2	2	0.17	0.0069
	GO:0000724	*GEN1, SMC6, PPP4R2*	SS1_REG13, SS1_REG16	*double-strand break repair via homologou...*	6	3	0.5	0.0363
	GO:0007528	*PDZRN3, TNC*	SS1_REG16, SS1_REG21	*neuromuscular junction development*	4	2	0.33	0.0369
	GO:0072659	*SPTBN1, PIGW, EPB41L3*	SS1_REG11, SS1_REG10, SS1_REG05	*protein localization to plasma membrane*	18	3	1.5	0.0371
ssGWAS2	GO:0045796	*ABCG8, ABCG5*	SS2_REG05	*negative regulation of intestinal choles...*	2	2	0.07	0.0013
	GO:0006878	*SLC31A1, SLC31A2*	SS2_REG11	*cellular copper ion homeostasis*	2	2	0.07	0.0013
	GO:0007588	*ABCG8, ABCG5*	SS2_REG05	*excretion*	4	2	0.14	0.0073
	GO:0033344	*ABCG8, ABCG5*	SS2_REG05	*cholesterol efflux*	5	2	0.18	0.0119
	GO:0035434	*SLC31A1, SLC31A2*	SS2_REG11	*copper ion transmembrane transport*	2	2	0.07	0.0354
	GO:0030042	*PLEKHH2, MICAL1*	SS2_REG05, SS2_REG04	*actin filament depolymerization*	6	2	0.22	0.0355
	GO:0031145	*CDC26*	SS2_REG11	*anaphase-promoting complex-dependent cat...*	1	1	0.04	0.0361
	GO:0070979	*CDC26*	SS2_REG11	*protein K11-linked ubiquitination*	1	1	0.04	0.0361
	GO:0035989	*COL11A1*	SS2_REG02	*tendon development*	1	1	0.04	0.0361
	GO:0003431	*COL27A1*	SS2_REG11	*growth plate cartilage chondrocyte devel...*	1	1	0.04	0.0361
	GO:0033617	*COX20*	SS2_REG12	*mitochondrial respiratory chain complex ...*	1	1	0.04	0.0361
	GO:0009157	*DCK*	SS2_REG01	*deoxyribonucleoside monophosphate biosyn...*	1	1	0.04	0.0361
	GO:0036305	*ENAM*	SS2_REG01	*ameloblast differentiation*	1	1	0.04	0.0361
	GO:0070175	*ENAM*	SS2_REG01	*positive regulation of enamel mineraliza...*	1	1	0.04	0.0361
	GO:0001561	*HAO1*	SS2_REG10	*fatty acid alpha-oxidation*	1	1	0.04	0.0361

[a] Genes annotated for the term described.
[b] Significant genes within the annotated genes.
[c] Expected.

myostatin gene (*MSTN*) is found, a selection signature was described in Paint Horses and Quarter Horses (racing and cutting line), and was described polymorphisms related to muscle fiber type (Petersen et al., 2013). The data generated in this study showed that SNPs in this region were fixed and were therefore excluded by QC analysis, in agreement with the results reported by Pereira et al. (2016b) who demonstrated that allele C of SNP g.66493737 C > T of the *MSTN* gene is fixed in the racing line of the Quarter Horse breed. Similarly, this allele was found at high frequency in Thoroughbred horses participating in sprint races (Hill et al., 2010a; Tozaki et al., 2010). In addition to the strong pressure of selection on this region of ECA 18 in the racing line of Quarter Horses, these findings show the genetic influence of Thoroughbreds with sprint aptitude on this category.

In fact, the Thoroughbred breed played an important role in the formation of the racing line of Quarter Horses (American Quarter Horse Association, 2016). Thus, the results of genetic studies involving the English breed may be similar to the racing line of the North American breed. In a genomic association study on Thoroughbred horses, Shin et al. (2015) described important genes associated with muscle physiology on seven chromosomes (ECA 5, 8, 16, 20, 21, 28, and 30); five of these chromosomes (ECA 5, 8, 16, 20, and 30) coincided with those found here. The authors described five linked genes on ECA 16, four of them present in a region comprising 14–18 Mb, the same region where SS1_REG16 and SS1_REG17 were located. On ECA 30, the position of importance identified by these authors was 1.5 Mb downstream of SS2_REG12. Although the results obtained by Shin et al. (2015) are relevant, most of the polymorphisms described in molecular genetics and genomic studies on the racing performance of Thoroughbreds (Tozaki et al., 2010; Binns et al., 2010; Hill et al., 2010a, 2010b; Gu et al., 2010) did not have the same effect in the racing line of the Quarter Horse breed (Pereira et al., 2016b, 2016a). According to the latter authors, this fact might be explained by the differences in the races disputed by the two breeds (shorter distance by Quarter Horses and longer distance by Thoroughbred horses), in which significant genetic alterations in the latter would improve the anaerobic muscle energy metabolism that predominates in long-distance races. However, alterations in other genes related to exercise physiology could contribute to important metabolic adaptations and be shared by the two breeds, as demonstrated by the comparison of the results obtained here and those reported for Thoroughbred horses (Shin et al., 2015).

As mentioned earlier, SImax was characterized as a quantitative

trait in this study and it is therefore expected that a large number of genetic alterations with small effects contribute to differences in the metabolism, anatomy and behavior of animals of the racing line of Quarter Horses. In this respect, the excessive variety of GO terms and associated genes impairs enrichment, a fact described for polygenic traits, especially when ssGWAS1 was used in which a larger number of important regions (gVar > 1%) with small effects were found.

Among the genes found by enrichment in the common regions, *PDZRN3*, *PPP4R2*, *GXYLT2* and *SHQ1* of ECA 16 have been identified by Shin et al. (2015) in a study on Thoroughbred horses. Particularly important are the *PDZRN3* and *PPP4R2* genes that were enriched (Table 6) and are related to the terms formation, differentiation and maintenance of neuromotor functions. In humans, the protein product of *PPP4R2* plays an important role in a motor disorder characterized by the progressive loss of motor neurons and muscular paralysis (spinal muscular atrophy) through its regulatory interaction with the survival of motor neuron (SMN) complex. Reduced levels of SMN protein cause this neuropathy (Bosio et al., 2012). *PDZRN3* participates in embryo morphogenesis, acting on an important pathway that is involved in the planar orientation and highly branched organization of vascular plexuses (Sewduth et al., 2014).

Among the most enriched terms are those related to the *ABCG5*, *ABCG8*, *SLC31A1* and *SLC31A2* genes. The *SLC31A1* and *SLC31A2* genes were significantly enriched in ssGWAS1 and ssGWAS2. These genes encode the CTR1 and CTR2 proteins that control intra- and extracellular copper (Cu^{2+}) fluxes (Wee et al., 2013). The biological role of copper includes its incorporation into copper enzymes for oxidative phosphorylation (cytochrome c oxidase), cellular antioxidant activity (superoxide dismutase), connective tissue formation (lysyl oxidase), neurological function (dopamine p-hydroxylase), and iron metabolism (ceruloplasmin) (Festa and Thiele, 2011). Genetic diseases of copper excess or deficiency are related to the inability to control the copper balance, which has recently been associated with neurodegenerative disorders (Kim et al., 2008).

Using ssGWAS1, a smaller number of genes associated with significant terms were observed. The *GEN1* and *SMC6* genes are basically related to the process of homologous chromosome recombination, being involved in Holliday junction (HJ) and DNA double-strand breaks. The mechanism of HJ is highly conserved from bacteria to humans (Chan and West, 2015), in which four-way DNA intermediates, are formed during mitotic and meiotic recombination, and their

efficient resolution is essential for proper chromosome segregation. In 2008, *GEN1* was identified as the gene encoding a human HJ resolvase with a key a role in this recombinational repair pathway (Ip et al., 2008). In additional, this gene has been associated to cancer susceptibility in some studies (Forbes et al., 2008; Turnbull et al., 2010). The *SPTBN1*, *PIGW* and *EPB41L3* genes are associated with the guidance and localization of proteins in the plasma membrane. *SPTBN1* is a member of the family of beta-spectrin genes (non-erythrocytic). Spectrins are a superfamily of F-actin crosslinking proteins that compose the actin cytoskeletal network underlying the plasma membrane and that play a role in the determination of cell shape, arrangement of transmembrane proteins, and organelle organization (Jiang et al., 2010).

Considering the genes significantly enriched in ssGWAS2, *ABCG5* and *ABCG8*, located on ECA 15, showed the lowest *p* values. These genes encode ATP-binding cassette (ABC) half-transporters G5 and G8 (ABCG5 and ABCG8), which limit intestinal absorption and promote biliary excretion of neutral sterols (Graf et al., 2003). The ABC transporter superfamily contains the largest family of transporter genes. Proteins ABCG5 and ABCG8 translocate a variety of substrates, including sugars, amino acids, metal ions, peptides and proteins, as well as a large number of hydrophobic compounds and metabolites, across extra- and intracellular membranes (Dean et al., 2001). These genes are associated with sitosterolemia in humans, a rare hereditary disease in which fats derived from plant-based foods accumulate in blood and tissues, causing arteriosclerosis, abnormal red blood cells and fat deposits in tendons (Lee et al., 2001).

The *PLEKHH2* and *MICAL1* genes have been associated with actin filament depolymerization in humans. However, most studies on *PLEKHH2* have investigated the gene in nephropathy (Greene et al., 2008). On the other hand, clearer evidence is available that the protein product of *MICAL1* participates in the depolymerization process of F-actin, which is related to many important cellular processes, including muscle contraction, cell motility, cell division and cytokinesis, vesicle and organelle movement, cell signaling, and the establishment and maintenance of cell junctions and shape (Dominguez and Holmes, 2011).

Particularly important are the *COL11A1* and *COL27A1* genes, which are related to the terms tendon development and growth plate cartilage chondrocyte growth and development. The *COL11A1* gene encodes one of the two alpha chains of collagen type XI (Zhidkova et al., 1995). Dysfunctions of this gene can cause Stickler syndrome and Marshall syndrome in humans, which are characterized by distinctive facial anomalies, eye problems, hearing loss, and joint problems (Stickler et al., 1965; Annunen et al., 1999). The *COL27A1* gene encodes a member of the fibrillar collagen family and is involved in cartilage calcification and in the transition from cartilage to bone. In humans, dysfunctions of this gene cause Steel syndrome, a condition characterized by cranial bone deformities and scoliosis (Gonzaga-Jauregui et al., 2015).

The *COX20* gene is located in the SS2_REG12 region, which alone explained the highest percentage of variance in SImax found in this study. This gene was described in yeast and its human ortholog is called *FAM36A*. The product of this gene is involved in the incorporation of subunits into the cytochrome c oxidase complex (COX) of the respiratory chain. A missense mutation in the *FAM36A* gene has been associated with ataxia (loss of coordination) and muscle hypotony in humans. The interaction of FAM36A with COX2 has been emphasized considering that in affected patients COX1 accumulates and COX2 is absent in the assembly of the COX complex, supporting that modified FAM36A impairs the initial step during incorporation of the COX2 subunit (Szklarczyk et al., 2013). The interaction of COX4 isoforms (1 and 2) has been associated with better or poorer efficiency of cellular respiration in situations of hypoxia (Fukuda et al., 2007). In the study of Gu et al. (2010), a weak but significant association was observed between the intronic SNP g.22684390 C > T of the equine *COX4I2* gene (ECA 22) and racing performance history in Thoroughbred horses.

However, this polymorphism did not have the same effect in Quarter Horses of the racing line, possibly because muscle metabolism is predominantly anaerobic in sprint races (Pereira et al., 2016a).

The speed trait in different horse breeds that participate in races of different distances and intensities is characterized by important physiological adaptations, especially those related to the aerobic and anaerobic muscle metabolism (Nielsen et al., 2006). However, the present study on Quarter Horses found a large number of genes related to the neuromotor system, bone and tendon development, hormonal responses, muscle contraction, and energy and sterol metabolism. Meira et al. (2014b) described many genes in genomic regions possibly selected for performance in the racing line, even though they used more superficial classification levels of biological processes (e.g., related to the nervous system). In contrast, studying Thoroughbred horses, Gu et al. (2010) indicated genomic regions responsible for biological functions such as insulin signaling, energy metabolism, and steroid metabolic processes to be important for the differentiation of performance in the breed. It should be noted that several enriched genes such as *DCK*, *ENAM*, *HAO1*, *PIGW*, and *EPB41L3* do not have well-defined functions, even when model organisms are considered, and could therefore not be discussed in detail. Furthermore, the terms for biological processes and functions of the genes described here have more clearly been associated with metabolism and physiology in humans and little or no information about the functions of these genes in biological processes is available for horses. However, humans and horses are highly similar in terms of exercise physiology (Lindinger, 1999; Capomaccio et al., 2011), a fact highlighting the importance of the genes prospected in this study as candidates for racing performance in the Quarter Horse breed.

5. Conclusions

So far, few genomic studies had as a key objective to understand the type of genetic control and/or to identify genomic regions and genes associated with performance in the racing line of Quarter Horses. The ssGBLUP method was used for the first time in this type of study involving this breed, permitting the prospection of regions that act on SImax. Taken together, the two approaches used in this study (ssGWAS1 and ssGWAS2) permitted to identify 33 SNP windows of 1 Mb on 14 chromosomes, which explained more than 1% of the total genetic variance of the trait. A total of 142 genes were identified in these regions. As expected, SImax is controlled by a large number of small-effect genes. In general, the *PDZRN3*, *PPP4R2*, *SLC31A1*, *SLC31A2*, *GEN1*, *SMC6*, *ABCG5*, *ABCG8*, *PLEKHH2*, *MICAL1*, *COL11A1*, *COL27A1* and *COX20* genes, which were highlighted by enrichment analysis, are related to the neuromotor system, bone and tendon development, hormonal responses, muscle contraction, and energy and sterol metabolism. These functions are considered important for athletic performance, corroborating these genes as candidates for racing performance in the Quarter Horse breed.

Conflicts of interest

The authors declare no conflicts of interest.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.livsci.2018.02.015.

References

Aguilar, I., Misztal, I., Johnson, D.L., Legarra, A., Tsuruta, S., Lawlor, T.J., 2010. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93, 743–752.

Aguilar, I., Misztal, I., Tsuruta, S., Wiggans, G.R., Lawlor, T.J., 2011. Multiple trait genomic evaluation of conception rate in Holsteins. J. Dairy Sci. 94, 2621–2624.

Alexa, A., Rahnenfuhrer, J., 2016. topGO: Enrichment Analysis for Gene Ontology. R package version 2.26.0. 2016. Retrieved on 8 January, from ⟨https://bioconductor.org/packages/release/bioc/html/topGO.html⟩.

American Quarter horse association, 2016. Breeding a Thoroughbred to a Quarter Horse. ⟨https://www.aqha.com/daily/at-your-service/2016/october/breeding-a-thoroughbred-to-a-quarter-horse/⟩ (Accessed 18 December 2016).

Annunen, S., Korkko, J., Czarny, M., Warman, Ml, Brunner, H.G., Kaariainen, H., Mulliken, J.B., Tranebjaerg, L., Brooks, D.G., Cox, G.F., Cruysberg, J.R., Curtis, M.A., Davenport, S.L., Friedrich, C.A., Kaitila, I., Krawczynski, M.R., Latos-Bielenska, A., Mukai, S., Olsen, B.R., Shinno, N., Somer, M., Vikkula, M., Zlotogora, J., Prockop, D.J., Ala-Kokko, L., 1999. Splicing mutations of 54-bp exons in the COL11A1 gene cause Marshall syndrome but other mutations cause overlapping Marshall/Stickler phenotypes. Am. J. Hum. Genet. 65, 974–983.

Binns, M.M., Boehler, D.A., Lambert, D.H., 2010. Identification of the myostatin locus (MSTN) as having a major effect on optimum racing distance in the Thoroughbred horse in the USA. Anim. Genet 41, 154–158.

Bosio, Y., Berto, G., Camera, P., Bianchi, F., Ambrogio, C., Claus, P., Di Cunto, F., 2012. PPP4R2 regulates neuronal cell differentiation and survival, functionally cooperating with SMN. Eur. J. Cell. Biol. 91, 662–674.

Capomaccio, S., Cappelli, K., Spinsanti, G., Mencarelli, M., Muscettola, M., Felicetti, M., Bonifazi, M., 2011. Athletic humans and horses: comparative analysis of interleukin-6 (IL-6) and IL-6 receptor (IL-6R) expression in peripheral blood mononuclear cells in trained and untrained subjects at rest. BMC Physiol. 11.

Chan, Y.W., West, S., 2015. GEN1 promotes Holliday junction resolution by a coordinated nick and counter-nick mechanism. Nucleic Acids Res. 43, 10882–10892.

Chen, C.Y., Misztal, I., Aguilar, I., Tsuruta, S., Meuwissen, T.H.E., Aggrey, S.E., Wing, T., Muir, W.M., 2011. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: an example using broiler chickens. J. Anim. Sci. 89, 23–28.

Clayton, D., 2015. snpStats: SnpMatrix and XSnpMatrix classes and methods. R package version 1.24.0. Retrieved on 12 November 2016, from ⟨http://bioconductor.org/packages/release/bioc/html/snpStats.html⟩.

Dean, M., Rzhetsky, A., Allikmets, R., 2001. The human ATP-binding cassette (ABC) transporter superfamily. Gen. Res 11, 1156–1166.

Dominguez, R., Holmes, K.C., 2011. Actin structure and function. Ann. Rev. Biophys. 40, 169–186.

Durinck, S., Spellman, P., Birney, E., Huber, W., 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4, 1184–1191.

Festa, R.A., Thiele, D.J., 2011. Copper: an essential metal in biology. Curr. Biol. 21, 877–883.

Forbes, S.A., Bhamra, G., Bamford, S., Dawson, E., Kok, C., Clements, J., Menzies, A., Teague, J.W., Futreal, P.A., Stratton, M.R., 2008. The Catalogue of Somatic Mutations in Cancer (COSMIC). Curr. Protoc. Hum. Genet.

Fukuda, R., Zhang, H., Kim, J.W., Shimoda, L., Dang, C.V., Semenza, G.L., 2007. HIF-1 regulates cytochrome oxidase subunits to optimize efficiency of respiration in hypoxic cells. Cell 129, 111–122.

Go, Y.Y., Bailey, E., Cook, D.G., Coleman, S.J., Macleod, J.N., Chen, K.C., Timoney, P.J., Balasuriya, U.B.R., 2011. Genome-wide association study among four horse breeds identifies a common haplotype associated with in vitro CD3+ T cell susceptibility/resistance to equine arteritis virus infection. J. Virol. 85, 13174–13184.

Goddard, M.E., Hayes, B.J., 2007. Genomic selection. J. Anim. Breed. Genet 124, 323–330.

Gonzaga-Jauregui, C., Gamble, C.N., Yuan, B., 2015. Mutations in COL27A1 cause steel syndrome and suggest a founder mutation effect in the Puerto Rican population. Eur. J. Hum. Genet. 23, 342–346.

Graf, G.A., Yu, L., Li, W.P., Gerard, R., Tuma, P.L., Cohen, J.C., Hobbs, H.H., 2003. ABCG5 and ABCG8 are obligate heterodimers for protein trafficking and biliary cholesterol excretion. J. Biol. Chem. 278, 48275–48282.

Greene, C.N., Keong, L.M., Cordovado, S.K., Mueller, P.W., 2008. Sequence variants in the PLEKHH2 region are associated with diabetic nephropathy in the GoKinD study population. Hum. Genet. 124, 255–262.

Gu, J., Machugh, D.E., Mcgiveny, B.A., Park, S.D.E., Katz, L.M., Hill, E.M., 2010. 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. 42, 569–575.

Hill, E.M., Gu, J., Eivers, S.S., Fonseca, R.G., Mcgiveny, B.A., Govindarajan, P., 2010a. A sequence polymorphism in MSTN predicts sprinting ability and racing stamina in thoroughbred horses. PLoS One 5.

Hill, E.M., Gu, J., Mcgiveny, B.A., MACHUGH, D.E., 2010b. Targets of selection in the thoroughbred genome contain exercise-relevant gene SNPs associated with elite racecourse performance. Anim. Genet 41, 56–63.

Hill, E.W., Mcgivney, B.A., Gu, J., Whiston, R., Machugh, D.E., 2010b. A genomewide 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 Genom. 11.

Jiang, X., Gillen, S., Esposito, I., Giese, N.A., Michalski, C.W., Friess, H., Kleeff, J., 2010. Reduced expression of the membrane skeleton protein beta1-spectrin (SPTBN1) is associated with worsened prognosis in pancreatic cancer. Histol. Histopathol. 25, 1497–1506.

Ip, S.C.Y., Rass, U., Blanco, M.G., Flynn, H.R., Skehel, J.M., West, S.C., 2008. Identification of Holliday junction resolvases from humans and yeast. Nature 456, 11–20.

Jombart, T., Ahmed, I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics.

Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.Y., Freimer, N.B., Sabatti, C., Eskin, E., 2010. Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. 42, 348–354.

Kim, B.-E., Nevitt, T., Thiele, D.J., 2008. Mechanisms for copper acquisition, distribution and regulation. Nat. Chem. Biol. 4, 176–185.

Lee, M.-H., Lu, K., Hazard, S., Yu, H., Shulenin, S., Hidaka, H., Kojima, H., Allikmets, R., Sakuma, N., Pegoraro, R., Srivastava, A.K., Salen, G., Dean, M., Patel, S.B., 2001. Identification of a gene, ABCG5, important in the regulation of dietary cholesterol absorption. Nat. Genet. 27, 79–83.

Lemos, M.V.A., Chiaia, H.L.J., Berton, M.P., Feitosa, F.L.B., Aboujaoud, C., Camargo, G.M.F., Pereira, A.S.C., Albuquerque, L.G., Ferrinho, A.M., Mueller, L.F., Mazalli, M.R., Furlan, J.J.M., Carvalheiro, R., Gordo, D.M., Tonussi, R., Espigolan, R., Silva, R.M.O., De Oliveira, H.N., Duckett, S., Aguilar, I., Baldi, F., 2016. Genome-wide association between single nucleotide polymorphisms with beef fatty acid profile in Nellore cattle using the single step procedure. BMC Genom. 17, 1471–2164.

Lindinger, M.I., 1999. Exercise in the heat: thermoregulatory limitations to performance in humans and horses. Can. J. Appl. Physiol. 24, 152–163.

Magalhães, A.F.B., De Camargo, G.M.F., Fernandes, G.A., Junior, Gordo, D.G.M., Tonussi, R.L., 2016. Genome association study of meat quality traits in nellore cattle. PLoS ONE 11, 6.

McCue, M.E., Bannasch, D.L., Petersen, J.L., Gurr, J., Bailey, E., Binns, M.M., Distl, O., Guerin, G., Hasegawa, T., Hill, E.W., Leeb, T., Lindgren, G., Penedo, M.C., Roed, K.H., Ryder, O.A., Swinburne, J.E., Tozaki, T., Valberg, S.J., Vaudin, M., Lindblad-Toh, K., Wade, C.M., Mickelson, J.R., 2012. A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. PLoS Genet. 8.

McQueen, C.M., Doan, R., Dindot, S.V., Bourquin, J.R., Zlatev, Z.Z., 2014. Identification of genomic loci associated with rhodococcus equi susceptibility in foals. PLoS One 9, 6.

Meira, C.T., Farah, M.M., Fortes, M.R.S., Moore, S.S., Pereira, G.L., Silva, J.A.I.I.V., Mota, M.D.S., Curi, R.A., 2014a. A genome-wide association study for morphometric traits in quarter horse. J. Equine Vet. Sci. 34, 1028–1031.

Meira, C.T., Curi, R.A., Farah, M.M., Oliveira, H.N., Beltran, N.A.R., Silva, J.A.I.I.V., Mota, M.D.S., 2014b. Prospection of genomic regions divergently selected in racing line of Quarter Horses in relation to cutting line. Animal 8, 1754–1764.

Meira, C.T., Fortes, M.R.S., Farah, M.M., Porto Neto, L., Kelly, M., Moore, S.S., Pereira, G.L., Chardulo, L.A.L., Curi, R.A., 2014c. Speed index in the racing Quarter Horse: a genome-wide association study. J. Equine Vet. Sci. 34, 1263–1268.

Misztal, I., Tsuruta, S., Lourenço, D., Aguilar, I., Legara, A., Vitezica, Z., 2014. Manual for BLUPF90 Family of Programs. University of Georgia, Athens, GA, USA.

Nielsen, B.D., Turner, K.K., Ventura, B.A., Woodward, A.D., O'Connor, C.I., 2006. Racing speeds of quarter horses, thoroughbreds and Arabians. Equine Vet. J. 36, 128–132.

Pereira, G.L., De Matteis, R., Meira, C.T., Regitano, L.C.A., Silva, J.A.I.I.V., Chardulo, L.A.L., Curi, R.A., 2016a. Comparison of sequence variants in the PDK4 and COX4I2 genes between racing and cutting lines of quarter horses and associations with the speed index. J. Equine Vet. Sci. 39, 1–6.

Pereira, G.L., De Matteis, R., Regitano, L.C.A., Chardulo, L.A.L., Curi, R.A., 2016b. MSTN, CKM, and DMRT3 gene variants in different lines of quarter horses. J. Equine Vet. Sci. 39, 33–37.

Petersen, J.L., Mickelson, J.R., Cothran, E.G., Andersson, L.S., Axelsson, J., 2013. Genetic diversity in the modern horse illustrated from genome-wide SNP data. PLoS One 8, 1.

Price, A.L., Zaitlen, N.A., Reich, D., Patterson, N., 2010. New approaches to population stratification in genome-wide association studies. Nat. Rev. Genet. 11, 459–463.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575.

R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Resende, M.D.V., Silva, F.F., Lopes, O.S., Azevedo, C.F., 2012. Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada (RRM) e Estatística Espacial. 7Federal University of Viçosa, Viçosa, MG, Brasil.

Sargolzaei, M., Chesnais, J.P., Schenkel, F.S., 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genom. 15, 478.

Sewduth, R.N., Jaspard-Vinassa, B., Peghaire, C., Guillabert, A., Franzl, N., Larrieu-Lahargue, F., 2014. The ubiquitin ligase PDZRN3 is required for vascular morphogenesis through Wnt/planar cell polarity signalling. Nat. Commun. 5.

Shin, D.-H., Lee, J.W., Park, J.-E., Choi, I.-Y., Oh, H.-S., Kim, H.J., Kim, H., 2015. Multiple genes related to muscle identified through a joint analysis of a two-stage genome-wide association study for racing performance of 1,156 thoroughbreds. Asian Austral J. Anim. 28, 771–781.

Silva, R.M.O., Stafuzza, N.B., Fragomeni, B.O., Camargo, G.M.F., Ceacero, T.M., Cyrillo, J.N.S.G., Baldi, F., Boligon, A.A., Mercadante, M.E.Z., Lourenco, D.L., Misztal, I., De Albuquerque, L.G., 2017. Genome-wide association study for carcass traits in an experimental nelore cattle population. PLoS ONE 12.

Stickler, G.B., Belau, P.G., Farrell, F.J., 1965. Hereditary progressive arthro-ophthalmopathy. Mayo Clin. Proc. 40, 433–455.

Szklarczyk, R., Wanschers, B.F., Nijtmans, L.G., Rodenburg, R.J., Zschocke, J., Dikow, N.,

Van Den Brand, M.A., Hendriks-Franssen, M.G., Gilissen, C., Veltman, J.A., Nooteboom, M., Koopman, W.J., Willems, P.H., Smeitink, J.A., Huynen, M.A., Van Den Heuvel, L.P., 2013. A mutation in the FAM36A gene, the human ortholog of COX20, impairs cytochrome c oxidase assembly and is associated with ataxia and muscle hypotonia. Hum. Mol. Genet. 22, 656–667.

Tozaki, T., Miyake, T., Kakoi, H., Gawahara, H., Sugita, S., Hasegawa, T., Ishida, N., Hirota, K., Nakano, Y., 2010. A genome-wide association study for racing performances in Thoroughbreds clarifies a candidate region near the MSTN gene. Anim. Genet 41, 28–35.

Turnbull, C., Hines, S., Renwick, A., Hughes, D., Pernet, D., Elliott, A., Seal, S., Warren-Perry, M., Gareth, E.D., Eccles, D., Stratton, M.R., Rahman, N., 2010. Mutation and association analysis of GEN1 in breast cancer susceptibility. J. Breast Cancer Res. Treat. 124, 1573–7217.

Vanraden, P.M., 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91, 4414–4423.

Wang, C., Ma, P., Zhang, Z., Ding, X., Liu, J., Fu, W., Weng, Z., Zhang, Q., 2012. Comparison of five methods for genomic breeding value estimation for the common dataset of the 15th QTL-MAS Workshop. BMC Proceedings, 6, 1–5.

Wang, H., Misztal, I., Aguilar, I., Legarra, A., Fernando, R.L., Vitezica, Z., Okimoto, R., Wing, T., Hawken, R., Muir, W.M., 2014. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chicken. Genet. Res. 94, 73–83.

Wee, N.K.Y., Weinstein, D.C., Fraser, S.T., Assinder, S.J., 2013. The mammalian copper transporters CTR1 and CTR2 and their roles in development and disease. Int. J. Biochem. Cell. Biol. 45, 960–963.

Zhang, Z., Liu, J., Ding, X., Bijma, P., De Koning, D.J., Zhang, Q., 2010. Best linear unbiased prediction of genomic breeding values using a trait-specific marker derived relationship matrix. PLoS One 5.

Zhidkova, N.I., Justice, S.K., Mayne, R., 1995. Alternative mRNA processing occurs in the variable region of the pro-α1(XI) and pro-α2(XI) collagen chains. J. Biol. Chem. 270, 9486–9493.