

# Analysis of cross-population differentiation between Thoroughbred and Jeju horses

Wonseok Lee<sup>1</sup>, Kyung-Do Park<sup>2</sup>, Mengistie Taye<sup>1,3</sup>, Chul Lee<sup>4</sup>, Heebal Kim<sup>1,4,5</sup>, Hak-Kyo Lee<sup>2</sup>, and Donghyun Shin<sup>2,\*</sup>

\* **Corresponding Author:** Donghyun Shin  
**Tel:** +82-63-270-4748, **Fax:** +82-63-270-2614,  
**E-mail:** sdh1214@gmail.com

<sup>1</sup> Department of Agricultural Biotechnology, Animal Biotechnology, and Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Korea

<sup>2</sup> Department of Animal Biotechnology, College of Agricultural and Life Sciences, Chonbuk National University, Jeonju 54896, Korea

<sup>3</sup> College of Agriculture and Environmental Sciences, Bahir Dar University, PO Box 79, Bahir Dar, Ethiopia

<sup>4</sup> Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 08826, Korea

<sup>5</sup> Institute for Biomedical Sciences, Shinshu University, Nagano 8304, Japan

**ORCID**

Wonseok Lee

<https://orcid.org/0000-0002-5103-3334>

Kyung-Do Park

<https://orcid.org/0000-0002-1945-6708>

Mengistie Taye

<https://orcid.org/0000-0001-6795-9943>

Chul Lee

<https://orcid.org/0000-0002-9403-0774>

Heebal Kim

<https://orcid.org/0000-0003-3064-1303>

Hak-Kyo Lee

<https://orcid.org/0000-0001-5387-4885>

Donghyun Shin

<https://orcid.org/0000-0002-0819-0553>

Submitted Jun 14, 2017; Revised Sept 16, 2017;  
Accepted Nov 18, 2017

**Objective:** This study was intended to identify genes positively selected in Thoroughbred horses (THBs) that potentially contribute to their running performances.

**Methods:** The genomes of THB and Jeju horses (JH, Korean native horse) were compared to identify genes positively selected in THB. We performed cross-population extended haplotype homozygosity (XP-EHH) and cross-population composite likelihood ratio test (XP-CLR) statistical methods for our analysis using whole genome resequencing data of 14 THB and 6 JH.

**Results:** We identified 98 (XP-EHH) and 200 (XP-CLR) genes that are under positive selection in THB. Gene enrichment analysis identified 72 gene ontology biological process (GO BP) terms. The genes and GO BP terms explained some of THB's characteristics such as immunity, energy metabolism and eye size and function related to running performances. GO BP terms that play key roles in several cell signaling mechanisms, which affected ocular size and visual functions were identified. GO BP term Eye photoreceptor cell differentiation is among the terms annotated presumed to affect eye size.

**Conclusion:** Our analysis revealed some positively selected candidate genes in THB related to their racing performances. The genes detected are related to the immunity, ocular size and function, and energy metabolism.

**Keywords:** Eye; Positive Selection; Thoroughbred; XP-CLR; XP-EHH

## INTRODUCTION

Horses were domesticated 6,000 years ago in the Eurasian steppe [1]. Domestication and artificial selection strongly affected differentiation of horse breeds to increase horse capacity related to racing or packing type. Especially, Thoroughbreds (THB) became an outstanding horse breed for racing preferable to any other horse breed. The athletic performance of THB has come from the intense selection that resulted in different anatomical and physiological characteristics [2]. Among the physiological characteristics, typical of THB are large muscle mass to body weight ratio, high skeletal muscle mitochondrial density and oxidative enzyme activity, and considerable intramuscular stores of energy substrates [2]. The anatomical characteristics of THB are their long legs and a lean body [3]. In addition to these characteristics, THB have larger eyes compared to their relatives [4] which might contribute to their running performances. Ocular size is hypothesized to have an effect on running speed in animals (Supplementary Figure S1). According to Leuckart's Law [5], animals capable of achieving fast speeds require large eyes to enhance visual acuity and avoid collisions with obstacles in their environment. This law is an empirical law in zoology and applied to vertebrate animals [5].

Selective sweep is among the major factors which can increase genetic differentiation between two populations and causes allele frequency spectra to depart from expectation under neutrality [6]. Most methods of identifying evidence of positive selection are based on the decay of linkage disequilibrium and distortion in the variation of allele frequency spectra [7]. Using heterozygosity statistics, Gu et al [1] reported the positive selection of candidate athletic-performance gene regions that are responsible for fatty acid oxidation, increased insulin sensitivity and muscle strength in thoroughbred horses. In another study, Park et al [8] identified positively selected genes related to exercise response in horses using cross-population extended haplotype homozygosity (XP-EHH) method. In this study, we used XP-EHH [9] and cross-population composite likelihood ratio method (XP-CLR) [6] methods to test for signatures of selective sweeps in THB. XP-EHH calculates haplotype decay separately for each group using the EHH [9] and XP-CLR is a likelihood method for detecting selective sweeps using jointly modeling the multilocus allele frequency differentiation between the two groups [7]. XP-CLR provides higher power than other approaches to detect selective sweeps and good localization of the selected allele. Additionally, it has been reported that XP-CLR is much more robust to ascertainment bias in SNP discovery than methods based on the allele frequency spectrum [6].

Here, using XP-EHH and XP-CLR population statistics, we compared THB and Jeju horse (JH) populations to identify positive selection sweep regions in THB. JH is a Korean native breed in Jeju Island located far south of the Korean peninsula. They are hardy with a small to medium body size [10]. JH are general breeds that have been raised for several purposes as riding, racing, and meat, and not intensively selected for a special purpose [11].

## MATERIALS AND METHODS

### Samples and ethics statement

Blood samples were collected from THB and JH horses by trained veterinarians according to relevant international as well as national guidelines and under permission from the Guide for the Care and Use of Laboratory Animals of Pusan National University. All experimental procedures used in this study were approved by the Institutional Animal Care and Use Committee of the Pusan National University (PNU-2013-0417).

### Pre-processing of DNA resequencing data

Whole-blood samples (10 mL) were collected from 14 THB and 6 JH. Sequence data of these 20 samples were generated using the Illumina HiSeq2000 platform. The DNA sequencing data has been submitted to the NCBI Sequence Read Archive (SRA) database with accession numbers (SRS345323 to SRS345338 and SRS346577 to SRS346580) [8].

Then, we carried out a base sequence quality check using the fastQC (ver 0.10) software (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>). We removed the potential adapter sequence using Trimmomatic-0.32. Paired-end sequence reads were mapped to the reference *Equus caballus* (ver 2.66) genome using Bowtie2 [12] with the default setting. The overall alignment rate of reads to the reference sequence was 94.58% (91.24% to 98.76%) with an average read depth of 15.87× (12.13× to 22.26×). On average across the whole samples, the reads covered 97.66% (97.53% to 97.77%) of the genome. For downstream processing and variant calling, we used open-source software packages (Supplementary Figure S2). Using Picard (ver 1.56: <http://broadinstitute.github.io/picard/>), potential polymerase chain reaction duplicates were filtered. Then, we used SAMtools (ver 0.1.18) [13] to make index files for reference and bam files. After preparation of these files, Genome Analysis ToolKit 1.4 (GATK) was used to perform local realignment of reads to correct misalignments due to the presence of indels (Realigner Target Creator and In del Realigner arguments) [14]. The Unified Genotyper and Select Variants arguments of GATK were used for calling candidate single nucleotide polymorphisms (SNPs). To filter variants and remove possible false positives, option “VariantFiltration” was adopted with the following command options: i) All SNPs with a phred-scaled quality score of less than 30 and with MQ0 (mapping quality zero); ii) Total count across all samples of mapping quality zero reads >4 were filtered; iii) Quality depth (unfiltered depth of non-reference samples; low scores are indicative of false positives and artifacts) less than 5 were filtered; and iv) SNPs with FS (phred-scaled p value using Fisher’s exact test) >200 were filtered as FS represents variation on either the forward or the reverse strand, which are implied of false-positive calls. After this, it remained with ~12.9 million autosomal SNPs. These SNPs were phased and imputed using BEAGLE Version 3.3.2 [15].

### Population structure analysis

For principal component analysis (PCA), we used the genome-wide complex trait analysis (GCTA) [16] to estimate the eigenvectors incorporating genotype data from THB and JH. Structure admixture analysis between the two breeds was performed. We limited the genotype data to a random subset of approximately 0.1% of total SNPs using PLINK (-thin option) [17,18] and conducted the STRUCTURE (ver 2.3.4) with 2 options: the “admixture model” and  $K = 2$ . Then we used ancestry graphs implemented Treemix 1.12 [19] to show the historical relationship between these two populations, using -m flag option in this study to infer migration events with 1,000 replicated bootstraps.

### Selective sweep analysis and gene annotation

We performed two analyses to detect positive selection sig-

natures in THB population. Whole SNP sets were used from both THB and JH for the analysis. Initially, the XP-EHH that measures cross-population extended haplotype homozygosity was used to identify positive selection regions. The calculation for XP-EHH was performed using the software xpehh [9]; <http://hgdp.uchicago.edu/Software/>). We assumed that genetic distance was equal to physical distance. These log ratios (unstandardized XP-EHH) were standardized to have a mean of zero and a variance of one. Then, we split the genome into non-overlapping segments of 50 kb and computed the maximum XP-EHH score in each segment. Top 1% regions with high XP-EHH values were considered strong signals in the THB population.

Next, the XP-CLR (ver 1.0) test for detecting selective sweep regions that involve jointly modeling the multilocus allele frequency between two populations [6]. Whole SNP sets were used from both THB and JH for the analysis. The parameters used were as follows: Non-overlapping sliding windows of 50 kb, the maximum number of SNPs within each window as 400, and correlation level from which the SNPs contribution to XP-CLR result was down-weighted 0.95 following Lee et al [20]. The regions with the XP-CLR values in the top 1% using XP-CLR score were designated candidate sweeps. Significant genomic regions identified from XP-EHH and XP-CLR were annotated to nearby genes (Equ cab 2). Genes that are located (partially or completely) in the window regions were presumed as candidate genes [20].

Database for Annotation, visualization, and integrated discovery tool was used for annotation and pathway analyses. In addition, using these positively selected genes, ClueGO plugin of Cytoscape was used to cluster by gene ontology and visualized them [21].

## RESULTS AND DISCUSSION

### DNA re-sequencing

From the re-sequencing of DNA from 14 THB and 6 JH whole genome, we obtained sequencing approximately  $15.87\times$  coverages on average, with a total of approximately 39 billion bp in 40 million reads per sample. Sequence reads of each sample were aligned with an overall alignment rate of 94.58% of the whole genome area (Supplementary Table S1). We finally obtained a total of  $\sim 12.9$  million autosomal SNPs used for sweep analysis (Supplementary Table S2).

### Population structure

We performed Structure analysis in a randomly sampled subset of 12,855 SNPs ( $\sim 0.1\%$  of the total autosomal SNPs in this study) to understand the admixture level between the breeds considered (Figure 1a) that showed clear differences. This was supported by PCA (Figure 1b) which infers global patterns of genetic structure without breed membership as unsupervised

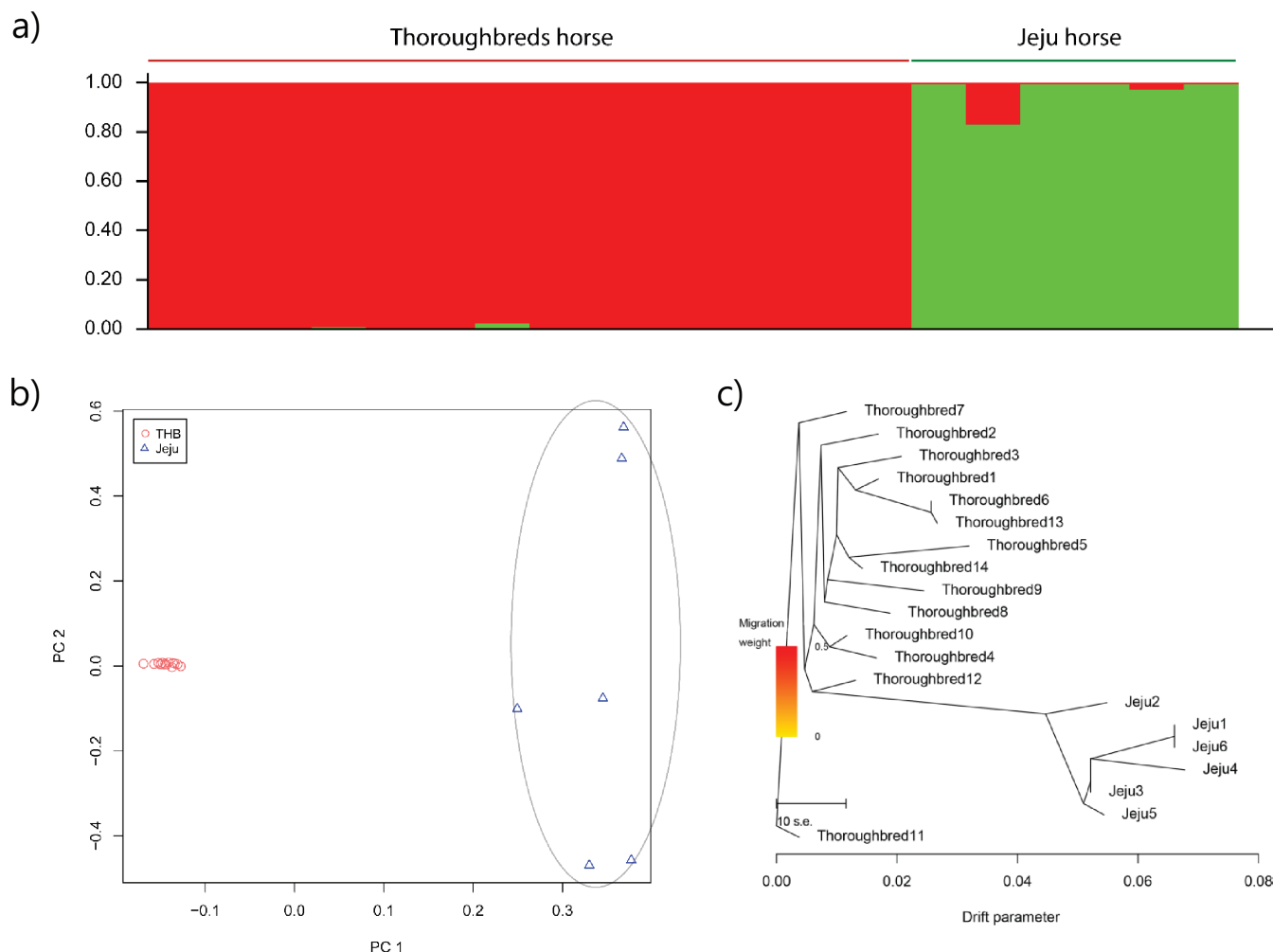
analysis. The largest principal component (PC1), positioned THB apart from JH explaining 17.8% of the variation. In addition, we performed the Treemix 1.12 analysis to infer the migration events of THB and JH. However, we didn't find any potential migration events between the two breeds (Figure 1c). Given this information, we suggest that they are clearly divided into two groups for downstream analysis. This suggests that THB and JH have evolved separately in different places. These results are consistent with a study using micro-satellite markers [22].

### Putative positive selection signals in THB horses

We used the XP-EHH method [9] to find genes under positive selection in THB, which calculates haplotype decay separately for each group using the EHH [23]. In addition, we calculated the XP-CLR statistics between THB and JH breeds. This statistic searches for the selective sweep on SNPs in the vicinity of the selected allele, using Brownian motion to model genetic drift under neutrality through allele frequency differentiation between populations [6]. The Manhattan plot of the  $-\log_{10}$  transformed XP-EHH and XP-CLR score p-values is presented in Figure 2a and 2b, respectively. Using the top 1% outlier regions, a total of 288 genes were detected using XP-EHH (98 genes) and XP-CLR (200 genes) population statistics (Supplementary Table S3, S4).

A comparison between THB and JH was appropriate because these populations have been bred under different environments for a long time. We calculated the XP-EHH values and XP-CLR scores as the window statistic of a total 44,826 and 44,844 genetic regions, respectively. By dividing the genome into a non-overlapping segment of 50 kb, we compared the genomic regions across populations and defined those genetic regions on whole horse genome. Empirical distributions using total regions can be constructed due to whole genome sequencing data. XP-EHH scores of 44,826 genetic regions and XP-CLR scores of 44,844 genetic regions showed normal distribution as expected (Supplementary Figure S3, S4). In this analysis, we used the outlier approach in distribution to detect a significant selective region [24]. We defined the top 1 percent of the XP-EHH and the XP-CLR score as a significant selective region and identified 448 significant genetic regions each which were a selective region in THB compared to JH. We identified 98 genes (XP-EHH) and 200 genes (XP-CLR) in 116 and 164 (with annotation) of total 448 significant regions.

We thought that regions with outlier XP-EHH and XP-CLR score provided several important pieces of evidence of THB domestication and selection. We constructed a biological network using Gene Ontology analysis which resulted in 72 gene ontology biological process terms (Figure 3). Then, the BP terms were grouped into 20 categories based on genes involved in which we focused on those supporting THB's characteris-

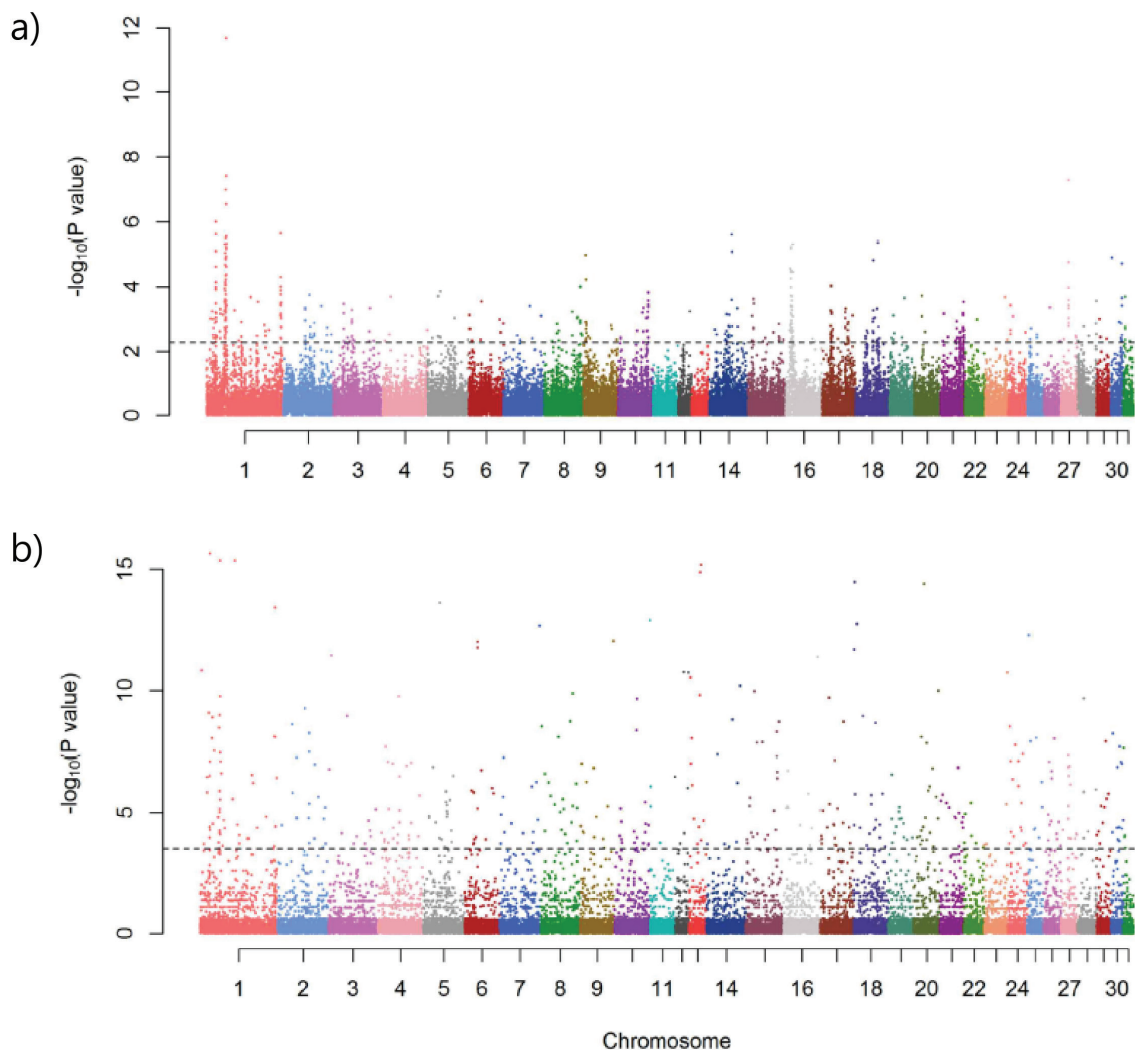


**Figure 1.** Population stratification of Thoroughbred and Jeju Horses (a) Population structure, (b) principal component analysis (PCA) plot, and (c) Treemix analysis. (a) Each segment represents the proportion of a horse individual genome from ancestral populations. Different colored segments in individuals assume that part of the genome originated from different ancestral populations. This figure shows the genetic structure of horse breeds when we assume that the number of ancestral populations of horse is 2. (b) Red circles are individuals in Thoroughbred horses, and blue triangles are individuals in Jeju horses. The horizontal axis indicates eigenvector 1, and the vertical axis indicates eigenvector 2. Values of eigenvectors were estimated using genome-wide complex trait analysis (GCTA) tool. (c) The result of TreeMix shows pattern of population splits and mixture between the two horse breeds. The drift parameter means proportional to  $N_e$  generations, where  $N_e$  is the effective population size. The scale bar shows ten times the average standard error of the estimated entries in the sample covariance matrix.

tics (Supplementary Table S5). We hypothesize that the BP terms enriched that are related to immune function, ocular size and visual function, and energy metabolism might contribute to the THB's superior racing performances (Table 1).

The BP term categories “Negative regulation of intracellular transport of viral material” is related to immunity. The genes involved in this term (bone marrow stromal cell antigen 2 [BST2] and tripartite motif containing 5 [TRIM5]) are associated with negative regulation of intracellular transport of viral material term, referring to any process that stops, prevents or reduces the frequency, rate or extent of intracellular transport of viral material. BST2 is associated with growth and development of B-cells. It is an interferon inducible transmembrane protein that provides innate immune response activity by inhibiting members of the retrovirus, filovirus, arenavirus,

and herpesvirus families [25]. Equine tetherin orthologues without dual tyrosine motif could potentially activate the nuclear factor kappa B subunit 1 (NF- $\kappa$ B) signaling. NF- $\kappa$ B plays a key role in regulating the immune response to infection [26]. TRIM5 gene encodes a member of the tripartite motif (TRIM) family that include three zinc-binding domains, a ring, a B-box type 1 and a B-box type 2, and a coiled-coil region. The protein forms homo-oligomers via the coiled-coil region and localizes to cytoplasmic bodies. It appears to function as an E3 ubiquitin-ligase and ubiquitinates itself to regulate its subcellular localization. It may play a role in retroviral restriction. Multiple alternatively spliced transcript variants encoding different isoforms have been described for this gene [27]. Another immune-related gene (RIN3) which plays a role in the maturation of phagosomes that engulf pathogens have been previously found

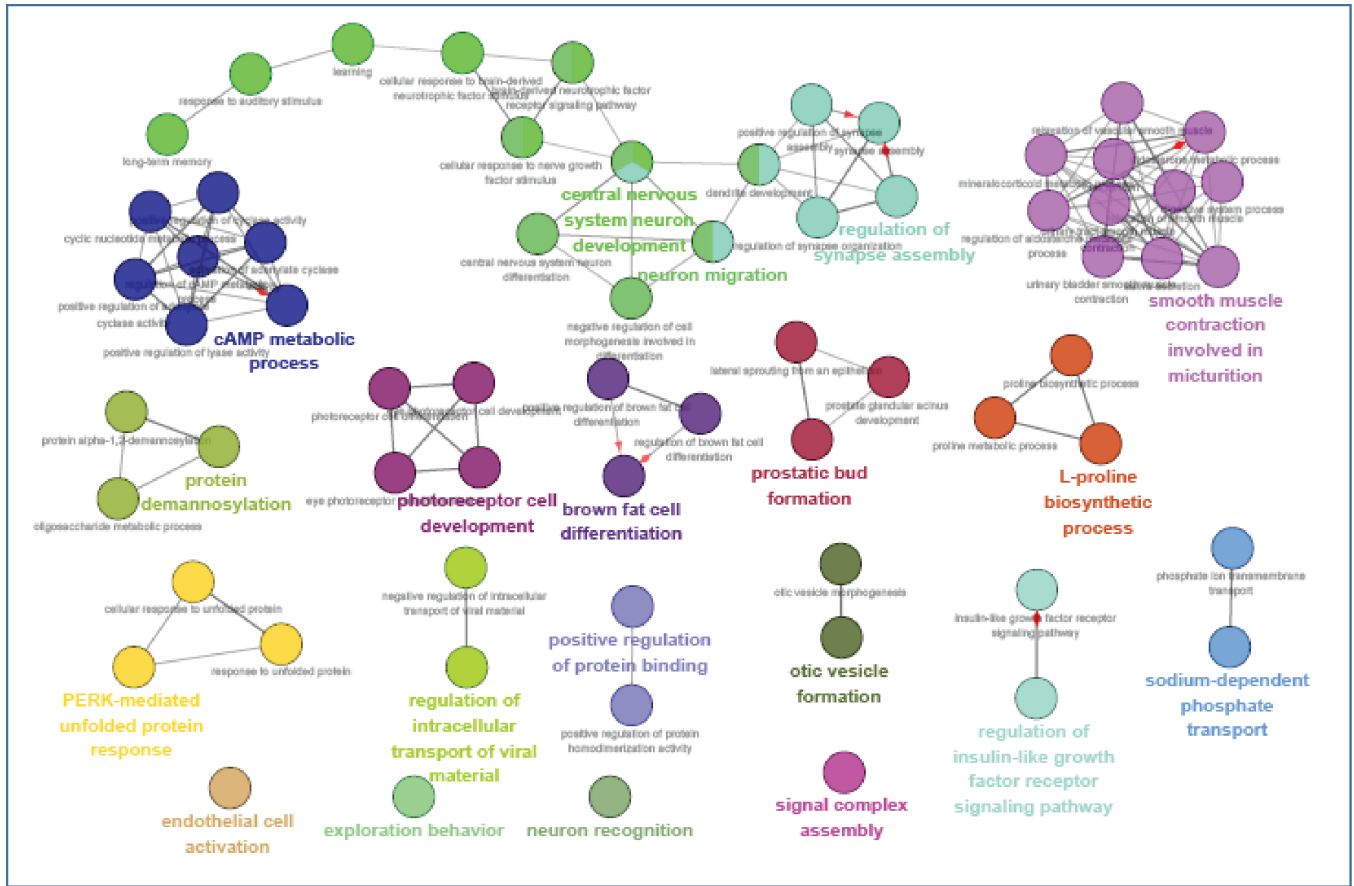


**Figure 2.** Manhattan Plot of  $-\log_{10}$  transformed (a) XP-EHH values, and (b) XP-CLR score p-values of Thoroughbred horses as compared to Jeju horses. The y-axis indicates  $-\log_{10}$  (p-value) of XP-EHH and XP-CLR values and the x-axis is the chromosomal position.

under positive selection in THB associated with racing performance [28].

THB are the fastest runners among the horse breeds used in the horse racing industry. Quite a few researchers studied why they run faster [1,5,28]. Here, we report genes and BP terms that potentially contribute for the superior racing performances of THB. THB has been selected for structural and functional adaptations that contribute to its fast running performance [28]. We identified that photoreceptor cell development and otic vesicle morphogenesis BP terms were enriched in relation to eye and ear development, respectively (Table 1). Genes related to eye photoreceptor cell differentiation include centrosomal protein 290, G protein subunit gamma transducin 1, crumbs 1, cell polarity complex component (*CRB1*), olfactomedin 3, and neurotrophic receptor tyrosine kinase 2 (*NTRK2*). We inferred that strong selection of eye photoreceptor cell differentiation can directly affect increment of ocular size which

leads to increased horse eyesight in the view of biological evolution at intra-species level. In vertebrate animals, ocular characteristic is influenced by many factors including body or head size, diet, and activity pattern. Heard-Booth and his colleague stressed that maximum locomotive speed plays a key role in determining ocular shape in mammals [5]. Leuckart's Law describes the relationship between a measure of axial eye diameter and maximum speed [5]. It has been reported that absolute ocular diameter is significantly correlated to maximum running speed in mammals [2]. This law also proposed that animals capable of achieving fast running speed require large eyes to enhance visual acuity and avoid collisions with environmental obstacles. The relationship between maximum running speed and eye size in a diverse sample of mammals proved this law [5]. Additionally, there were two more GO terms which supported directly or indirectly positive selection of ocular size and function in THB in this study; dendrite de-



**Figure 3.** Biological network using genes related to selective regions in Thoroughbreds. Gene ontology biological process (GO BP) network analysis of biological processes in Thoroughbreds and Jeju horses. GO terms visualized by ClueGo plugin of Cytoscape. Nodes are represented by a circle and imply that two GO terms share genes from the considered gene lists. The size of the circle corresponds to the number of genes related to the GO term. Edges are connections between GO groups defined by 50% genes in common.

velopment, and regulation of synapse assembly. Several genes (protein kinase, CGMP-dependent, type 1 [*PRKG1*], rap guanine nucleotide exchange factor 2 [*RAPGEF2*], RAB17, member RAS oncogene family [*RAB17*], neural EGFL like 1 [*NELL1*], DCC netrin 1 receptor [*DCC*], NCK adaptor protein 2 [*NCK2*], ghrelin and obestatin prepropeptide [*GHRL*], adhesion G protein-coupled receptor B3 [*ADGRB3*], and *NTRK2*) were identified that trigger dendrite development [29-31]. When light reaches retina after traveling through cornea and lens, ganglion cells take electronic signal through dendrite and send this signal down to the optic nerve. EPH receptor A5, *RAB17*, SH3 and multiple ankyrin repeat domains 2, and *GHRL* are related to regulation of synapse assembly [32-34]. Based on this knowledge related to optic nerve, we reasoned that eyesight is closely related to synapse because the retina has several neuron layers and communication among these several neuron layers is very important in eye function. Adenylate cyclase 1, involved in the regulatory processes in the central nervous system that play a role in memory and learning, have been found to be under selection in racehorse populations [28].

The BP term brown fat cell differentiation, defined as the

process in which a relatively unspecialized cells acquire specialized features of a brown adipocyte, is an animal connective tissue cell involved in adaptive thermogenesis [35]. Brown adipose tissue differs from white adipose tissue in the way they expend energy [1]. The type, intensity, and duration of exercise determine the amount of form of fuel used (carbohydrate vs free fatty acid) that, aerobic activities (long duration, low intensity) use more free fatty acids as fuel than anaerobic activities (short duration, high intensity), which use more glucose. However, the horse is almost always using both types to some degree, at the same time. As activity level (e.g. running speed) increases, oxygen consumption rises to meet increased demand for ATP production. Brown fat has more mitochondria than other cells. When the body needs to use energy, it uses ATP. ATP is mainly produced in the mitochondria of cells. When brown fat is activated, it creates a protein called uncoupling protein 1, which prevents ATP production from mitochondria. Instead of generating ATP, heat energy is generated to increase body temperature. The effect of fat supplementation of horse diet on horse performances has been reported. Genes including peroxisomal biogenesis factor 11

**Table 1.** Genes in gene ontology terms related to eye in selective regions in Thoroughbred horse (false discovery rate<0.05)

Gene ontology biological process	Genes in selective region	Chr.	XP-EHH value	XP-CLR scores
Dendrite development	<i>PRKG1</i>	1	5.060	18.250
	<i>RAPGEF2</i>	2	-	22.628
	<i>RAB17</i>	6	-	16.497
	<i>NELL1</i>	7	5.758	20.734
	<i>DCC</i>	8	-	24.395
	<i>NCK2</i>	15	-	16.142
	<i>GHRL</i>	16	-	18.943
	<i>ADGRB3</i>	20	-	26.040
	<i>NTRK2</i>	23	-	16.068
Photoreceptor cell development	<i>GNGT1</i>	4	-	16.560
	<i>OLFM3</i>	5	-	19.519
	<i>NTRK2</i>	23	-	16.068
	<i>CEP290</i>	28	-	20.069
	<i>CRB1</i>	30	5.596	-
Regulation of synapse assembly	<i>PRKG1</i>	1	-	18.250
	<i>RAPGEF2</i>	2	-	22.628
	<i>EPHA5</i>	3	-	16.749
	<i>LRRN3</i>	4	-	21.775
	<i>RAB17</i>	6	-	16.497
	<i>NELL1</i>	7	-	20.734
	<i>PTK2</i>	9	5.203	-
	<i>SHANK2</i>	12	-	20.333
	<i>SPOCK1</i>	14	5.341	-
	<i>NCK2</i>	15	-	16.142
	<i>GHRL</i>	16	-	18.943
	<i>CHL1</i>	16	5.482	-
	<i>SLC4A10</i>	18	-	19.221
	<i>NTRK2</i>	23	-	16.068

XP-EHH, cross-population extended haplotype homozygosity; XP-CLR, cross-population composite likelihood ratio test); *PRKG1*, protein kinase, CGMP-dependent, type I; *RAPGEF2*, rap guanine nucleotide exchange factor 2; *RAB17*, RAB17, member RAS oncogene family; *NELL1*, neural EGFL like 1; *DCC*, DCC netrin 1 receptor; *NCK2*, NCK adaptor protein 2; *GHRL*, ghrelin and obestatin prepropeptide; *ADGRB3*, adhesion G protein-coupled receptor B3; *NTRK2*, neurotrophic receptor tyrosine kinase 2; *GNGT1*, G protein subunit gamma transducin 1; *OLFM3*, olfactomedin 3; *CEP290*, centrosomal protein 290; *CRB1*, crumbs 1, cell polarity complex component; *EPHA5*, EPH receptor A5; *LRRN3*, leucine rich repeat neuronal 3; *PTK2*, protein tyrosine kinase 2; *SHANK2*, SH3 and multiple ankyrin repeat domains 2; *SPOCK1*, SPARC (osteonectin), cwcv and kazal like domains proteoglycan 1; *CHL1*, cell adhesion molecule L1 like; *SLC4A10*, solute carrier family 4 member 10.

alpha, laminin subunit alpha 4, zinc finger protein 516, and PR/SET domain 16 are related to brown fat cell differentiation. The positive selection of genes involved in brown adipose tissue differentiation has been previously identified in THB [1].

Insulin receptor signaling pathway is another pathway enriched which control critical energy functions such as glucose and lipid metabolism. It has been found previously to be enriched in THB horses [1] in relation to racing performance. It has also a role in the differentiation of brown adipocytes [36].

Through QTL analysis, we identified six QTL regions that overlapped to genes in selective regions of THB (Table 2). Ceramide synthase 6 (QTL chr18:48212639\_48319679) is well-known racing distance associated gene and *ADGRB3* (QTL chr20:60009473\_60987311) is closely related to recurrent uveitis disease of the eye. Recurrent uveitis is an acute, non-granulomatous inflammation of the uveal tract of the eye, occurring commonly in horses of all types of breeds universally [37].

THB are the epitomes of variation under domestication, yet much of the evolutionary processes underlying the genetics of this diversity are poorly understood. So we tried to detect novel selective regions which were not reported, previously. We attained novel selective regions using XP-CLR analysis which helped us to observe the relationship between THB and JH in a different angle. These results can be used to characterize functional variants and explore the specificity of the Thoroughbred breed.

**Limitations of the present study**

The possibility of obtaining false positive results is common in such kind of study. Therefore, gene expression analysis, and/or candidate gene approach experimental procedures are required to validate the candidate genes.

**CONCLUSION**

We explored the whole genome and detected several positively selected genes involved in different biological and cellular functions affecting THB horses’ characteristics. The genes

**Table 2.** QTL overlapped with selective regions in Thoroughbreds compared to Jeju horses

Gene name	Chr	Gene begin	Gene end	QTL ID <sup>1)</sup>	Related traits
<i>SEC61G</i>	4	24,372,218	24,380,086	qtl_4_24010915_24868953	Insect bite hypersensitivity (29305)
<i>CERS6</i>	18	48,176,453	48,465,399	qtl_18_48212639_48319679	Racing distance (32133)
<i>ADGRB3</i>	20	60,514,643	61,156,448	qtl_20_60009473_60987311	Recurrent uveitis (29387)
<i>SLC17A1</i>	20	23,713,124	23,741,030	qtl_20_23723503_23816767	Equine sarcoids (28919)
<i>SLC17A3</i>	20	23,767,516	23,814,310	qtl_20_23723503_23816767	Equine sarcoids (28919)
<i>GOLGA1</i>	25	28,825,836	28,868,824	qtl_25_24227654_30109054	Equine sarcoids (28921)

QTL, quantitative trait locus; *SEC61G*, Sec61 translocon gamma subunit; *CERS6*, ceramide synthase 6; *ADGRB3*, adhesion G protein-coupled receptor B3; *SLC17A1*, solute carrier family 17 member 1; *GOLGA1*, golgin A1.

<sup>1)</sup> QTL ID was made in this study as followed: qtl+ “chromosome” + “qtl begin” + “qtl end”.

identified in relation to THB characteristics are involved in immunity and eye size, and function that might contribute for THB's superior racing performances. These results provide a basis for further research on the genomic characteristics of THB.

## CONFLICT OF INTEREST

We certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

## ACKNOWLEDGMENTS

This work was supported by a grant from the Next-Generation BioGreen 21 Program (PJ0131512018), Rural Development Administration, Republic of Korea.

## REFERENCES

- Gu J, Orr N, Park SD, et al. A genome scan for positive selection in thoroughbred horses. *PLoS one* 2009;4:e5767.
- Hinchcliff KW, Kaneps AJ, and Geor RJ. *Equine exercise physiology: the science of exercise in the athletic horse*. Melbourne, Australia: Elsevier Health Sciences; 2008.
- Montgomery ES. *The Thoroughbred*. London, UK: Thomas Yoseloff Ltd.; 1971.
- Howland HC, Merola S, Basarab JR. The allometry and scaling of the size of vertebrate eyes. *Vision Res* 2004;44:2043-65.
- Heard-Booth AN, Kirk EC. The influence of maximum running speed on eye size: a test of Leuckart's Law in mammals. *Anat Rec* 2012;295:1053-62.
- Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps. *Genome Res* 2010;20:393-402.
- Ma Y, Zhang H, Zhang Q, Ding X. Identification of selection footprints on the X chromosome in pig. *PLoS One* 2014;9:e94911.
- Park W, Kim J, Kim H, et al. Investigation of de novo unique differentially expressed genes related to evolution in exercise. *PLoS One* 2014;9:e91418.
- Sabeti PC, Varilly P, Fry B, et al. Genome-wide detection and characterization of positive selection in human populations. *Nature* 2007;449:913-8.
- Kim K, Yang YH, Lee SS, et al. Phylogenetic relationships of Cheju horses to other horse breeds as determined by mtDNA D-loop sequence polymorphism. *Anim Genet* 1999;30:102-8.
- Chang-Yeon C, Sung-Heum Y, Byung-Wook C, and Gil-Jae C. Genetic characterization and polymorphisms for parentage testing of the Jeju horse using 20 microsatellite loci. *J Vet Med Sci* 2008;70:1111-5.
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357-9.
- Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25:2078-9.
- McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297-303.
- Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* 2007;81:1084-97.
- Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; 88:76-82.
- Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904-9.
- Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559-75.
- Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* 2012;8:e1002967.
- Lee H-J, Kim J, Lee T, et al. Deciphering the genetic blueprint behind Holstein milk proteins and production. *Genome Biol Evol* 2014;6:1366-74.
- Bindea G, Mlecnik B, Hackl H, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 2009;25:1091-3.
- Cho G-J. Genetic relationship and characteristics using microsatellite DNA loci in horse breeds. *J Life Sci* 2007;17:699-705.
- Ricard A, Bruns E, Cunningham E. *Genetics of performance traits. The genetics of the horse*. Wallingford, UK: CABI Publishing; 2000. pp. 411-538.
- Kelley JL, Madeoy J, Calhoun JC, Swanson W, Akey JM. Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome Res* 2006;16:980-9.
- Evans DT, Serra-Moreno R, Singh RK, Guatelli JC. BST-2/tetherin: a new component of the innate immune response to enveloped viruses. *Trends Microbiol* 2010;18:388-96.
- Yin X, Guo M, Gu Q, et al. Antiviral potency and functional analysis of tetherin orthologues encoded by horse and donkey. *Virology* 2014;11:151.
- O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 2016;44: D733-D45.
- Moon S, Lee JW, Shin D, et al. A genome-wide scan for selective sweeps in racing horses. *Asian-Australas J Anim Sci* 2015;28: 1525-31.
- Jan Y-N, Jan LY. The control of dendrite development. *Neuron* 2003;40:229-42.
- Ohshima T. Neuronal migration and protein kinases. *Front Neurosci* 2014;8:458.



31. Quach TT, Wilson SM, Rogemond V, et al. Mapping CRMP3 domains involved in dendrite morphogenesis and voltage-gated calcium channel regulation. *J Cell Sci* 2013;126:4262-73.
32. Dalva MB, Takasu MA, Lin MZ, et al. EphB receptors interact with NMDA receptors and regulate excitatory synapse formation. *Cell* 2000;103:945-56.
33. Zerial M, McBride H. Rab proteins as membrane organizers. *Nat Rev Mol Cell Biol* 2001;2:107-17.
34. Waites CL, Craig AM, Garner CC. Mechanisms of vertebrate synaptogenesis. *Annu Rev Neurosci* 2005;28:251-74.
35. Puigserver P, Spiegelman BM. Peroxisome proliferator-activated receptor- $\gamma$  coactivator 1 $\alpha$  (PGC-1 $\alpha$ ): transcriptional coactivator and metabolic regulator. *Endocr Rev* 2003;24:78-90.
36. Sharma A, Huard C, Vernochet C, et al. Brown fat determination and development from muscle precursor cells by novel action of bone morphogenetic protein 6. *PloS One* 2014;9:e92608.
37. Laurie GW, Olsakovsky LA, Conway BP, et al. Dry eye and designer ophthalmics. *Optom Vis Sci* 2008;85:643-52.