INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
19–0660/2019/22–5–1079–1086
DOI: 10.17957/IJAB/15.1172
http://www.fspublishers.org

Full Length Article

Identification of Divergently Selected Regions in the Thoroughbred and Jeju Pony by using Equine SNP Chip Array

Sangwook Kim†, Byeonghwi Lim† and Kyoungtag Do*
Lab of Equine Science, Department of Animal Biotechnology, Faculty of Biotechnology, Jeju National University, Jeju 63243, Republic of Korea
*For correspondence: challengekt@jejunu.ac.kr; challengekt@gmail.com
†These authors contributed equally to this work

Abstract

Allelic differences that are associated with traits suggest evidence for artificial selections. The sliding window approach can identify distinctly selected genomic regions between two breeds. The purpose of this study was to identify divergently selected regions between the Jeju pony (JP) and the Thoroughbred (THB) horse and to confirm the potential candidate genes. After a high-density single nucleotide polymorphism (SNP) array was analyzed for 50 JP and 50 THB, 38,378 SNPs were selected. Observed heterozygosity, expected heterozygosity and polymorphism information content were calculated to evaluate the genetic diversity of both breeds. All values for the JP were lower than those for the THB. An analysis was performed on 38,099 windows to identify the regions with the highest frequency of alleles between JP and THB. The sliding window average difference values were concentrated on 43 windows and were distributed in 10 regions. Furthermore, 19 candidate genes were detected inside the 10 regions. We found that 12 candidate genes mutually interact, and 4 candidate genes are involved in significant biological pathways. Results from this study suggest that the 10 regions are selection signatures with causative genes and mutations that affect certain traits. © 2019 Friends Science Publishers

Keywords: Candidate gene; Jeju pony; Thoroughbred horse; Selection; SNP

Introduction

Since the first domesticated-horse breeding event around 5,000 years ago, selective breeding has been used to choose horses for use in agriculture, transportation and warfare (Ludwig *et al*., 2009; Outram *et al*., 2009; Lippold *et al*., 2011). Over the past 400 years, the establishment of formal breed registrations and ongoing breed specialization has focused on preserving and improving the traits associated with esthetics and performance.

The thoroughbred (THB) is the most commonly used racehorse and was developed after the 16th century using selective breeding to produce horses with excellent racing capabilities (Thiruvenkadan *et al*., 2009; Bower *et al*., 2012). Meanwhile, the Jeju pony (JP, Natural Monument number 347) is a medium-sized horse and is characterized by robustness and high resilience against diseases. This species is a hybrid of ponies inhabiting Jeju Island, Republic of Korea. Originally, 160 Mongolian horses were brought to the island from Mongolia in 1276. Thereafter, the JP developed by adaptations of conformation to the harsh environment of Jeju Island. To this day, various efforts have been made to preserve this species and maintain its breeding (Nam, 1969; Lim *et al*., 2019).

The development of next generation sequencing (NGS) technology and the applications of whole-genome sequencing have increased the ability to analyze large amounts of genomic data in horses. Genetic factors for various characteristics of horses, such as disease susceptibility, robustness, and athletic performance, have been explored at the level of whole-genome sequencing through high-density single nucleotide polymorphism (SNP) chip applications (Go *et al*., 2011; Meira *et al*., 2013; Petersen *et al*., 2013a). Population genetic analysis techniques have also been used to study horse domestication (Petersen *et al*., 2013b), linkage disequilibrium, and historical effective population sizes (Corbin *et al*., 2010; Do *et al*., 2014).

The use of the sliding-window analysis approach in livestock has allowed for the detection of differently selected genomic regions through the analysis of multiple consecutive SNPs. It has been suggested that this approach identifies the genes that affect production and other phenotypic traits through the comparison of SNP frequencies in specific genomic regions to mark the selected regions of different breeds (Stella *et al*., 2010; Huang *et al*., 2014; Gurgul *et al*., 2015).

Studies on gene networking and biochemical pathways

To cite this paper: Kim, S., B. Lim and K. Do, 2019. Identification of divergently selected regions in the thoroughbred and Jeju Pony by using Equine SNP Chip Array. *Intl. J. Agric. Biol*., 22: 1079–1086

provide useful information for understanding the biomechanisms at the system level (Kitano, 2002). Recent studies have analyzed the networks of candidate genes associated with pathways, such as muscle metabolism, energy metabolism, and immunity in relation to the exercise ability of horses (Park *et al*., 2012; Kim *et al*., 2013; Park *et al*., 2014; Lee *et al*., 2016).

The purpose of this study was to identify genomic regions that have differences in polymorphism frequency between the THB and JP breeds. Sites exhibiting such differences between the two breeds were then compared using an equine high-density single nucleotide polymorphism (SNP) chip array and a sliding-window approach. Moreover, gene networks and pathways were analyzed to identify the potential functions of the identified candidate genes of metabolism.

Materials and Methods

DNA were extracted from hair samples of 50 Jeju pony (JP) and 50 Thoroughbred (THB) horses in Jeju Island. Genotyping for 100 horses was performed using Equine SNP 70K for JP and 50K BeadChips for THB (Geneseek, Lincoln, NE), respectively.

Data filtering and marker screening were carried out for the two breeds to compare frequencies under similar conditions. A total of 65,157 markers were identified for JP and 54,602 for THB but only 63,229 and 54,430 markers were identified, respectively, from the screening that did not correspond to the 90% call rate (JP: 1,928 and THB: 172 markers). Thereafter, 40,700 common markers were selected using overlapped data filtering between the two breeds and 38,378 SNPs were chosen for the analyses by screening the common markers located on the sex chromosomes (2,118 markers) and the common fixed alleles (204 markers).

After the filtering and screening process, a population genetic indices analysis was performed to estimate the within-population genetic diversity. We calculated observed and expected heterozygosity (H_O and H_E) and polymorphism information content (PIC) for the datasets using PLINK software (Purcell *et al*., 2007). The population genetic indices for H_O , H_E , and PIC were calculated according to Nei's methods (Nei and Roychoudhury, 1974). The formulas were as follows:

Expected heterozygosity (H_E)

$$H_E = 1 - \sum_{i=1}^{n} p_i^2,$$

Where n is the total number of alleles and p_i is the frequency of the i^{th} allele in the population and Polymorphism information content (PIC)

$$PIC = 1 - \left(\sum_{i=1}^{n} p_i^2\right) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i^2 p_j^2,$$

Where n is the total number of alleles and p_i and p_j are the frequencies of the i^{th} and j^{th} alleles in the population.

Identification of the genomic regions with distinct selection signatures were measured as follows: 1) For each of the SNPs, the absolute difference between the allelic frequencies of JP and THB was calculated, 2) after ordering the SNPs on each autosome, 10 consecutive SNPs were defined as components of a sliding window and 3) for each window, the average difference in allelic frequency of JP and THB was calculated and called the sliding window average difference (SWAD). The SWAD were not calculated on each end of all autosomes, so the number of the calculated SWAD was 38,378–31×9 = 38,099 (Hosokawa *et al*., 2012).

A fraction of the identified genomic regions (with the highest SWAD value > 0.499) was scanned to confirm the presence of the protein coding genes with the most varying genomic regions between the breeds using Ensembl Genome Browser (Genome assembly: Equ Cab 2). The molecular functions of the encoded proteins were analyzed using the PANTHER (Protein Analysis Through Evolutionary Relationships) Classification System (http://www.pantherdb.org/) (Thomas *et al*., 2003). Gene network and pathway analysis provides important insights into the genetic architecture of complex polygenic traits. Therefore, we carried out a gene network and pathway analysis with the identified candidate genes using the GeneMANIA tool plugin for Cytoscape software 3.6 (Shannon *et al*., 2003).

Results

The common average inter-marker distance of 38,378 SNPs from both JP and THB was determined to be 58.181kb (± 74.778) using a high-density SNP chip array. The *Equus caballus* chromosome 24 (ECA 24) (52.140 kb ± 63.502) had the shortest average inter-marker distance in both JP and THB, while ECA 26 (82.972 kb ± 107.215) was the most remote. The average minor allelic frequency (MAF) of JP and THB was 0.198 ± 0.150 and 0.231 ± 0.155, respectively, and this showed that the allelic frequency of THB was higher than that of JP (Table 1). ECA 9 (0.215 ± 0.154) had the highest average MAF, whereas, ECA 28 (0.186 ± 0.148) had the lowest average MAF in the JP. The average MAF of ECA 26 (0.248 ± 0.160) was the highest, and ECA 1 (0.215 ± 0.154) was the lowest in the THB. Additionally, the average difference in allelic frequencies was 0.267 ± 0.251 for total ECA. ECA 30 had the highest average difference in allelic frequencies (0.296 ± 0.267), whereas, ECA 3 had the lowest (0.224 ± 0.228).

The genetic index values, for both JP and THB, for observed heterozygosity (H_O), expected heterozygosity (H_E), and polymorphism information contents (PIC) are provided in Table 2. The mean values of H_O, H_E, and PIC in the JP were 0.279 ± 0.173, 0.273 ± 0.167, and 0.222 ± 0.124,

Table 1: Distribution of analyzed SNP markers, observed MAF and average difference of allelic frequency between Jeju pony (JP) and Thoroughbred horse (THB) for individual locus by each chromosome

ECA	SNP (n)	Inter-marker distance (± SD), kb	MAF (± SD)		Average difference (± SD)*
			JP	THB	
1	2,922	63.535 (85.279)	0.197 (0.148)	0.215 (0.154)	0.290 (0.262)
2	2,141	56.275 (70.034)	0.198 (0.150)	0.228 (0.149)	0.262 (0.248)
3	1,870	63.829 (92.390)	0.205 (0.153)	0.227 (0.148)	0.224 (0.228)
4	1,864	58.265 (71.802)	0.198 (0.147)	0.240 (0.162)	0.286 (0.257)
5	1,728	57.665 (74.178)	0.197 (0.155)	0.239 (0.166)	0.263 (0.250)
6	1,483	54.776 (82.412)	0.197 (0.148)	0.220 (0.152)	0.282 (0.262)
7	1,642	59.972 (88.440)	0.200 (0.148)	0.239 (0.158)	0.277 (0.250)
8	1,675	55.922 (70.514)	0.208 (0.151)	0.232 (0.153)	0.233 (0.236)
9	1,462	57.108 (66.396)	0.215 (0.154)	0.218 (0.156)	0.259 (0.248)
10	1,451	57.868 (71.188)	0.197 (0.151)	0.234 (0.158)	0.269 (0.254)
11	1,127	54.407 (67.833)	0.201 (0.148)	0.238 (0.156)	0.270 (0.243)
12	516	63.689 (81.865)	0.199 (0.144)	0.236 (0.160)	0.283 (0.279)
13	701	60.598 (77.002)	0.195 (0.149)	0.232 (0.153)	0.241 (0.247)
14	1,706	54.766 (69.274)	0.198 (0.146)	0.228 (0.157)	0.283 (0.255)
15	1,603	56.872 (70.230)	0.198 (0.149)	0.234 (0.156)	0.253 (0.243)
16	1,532	57.016 (73.503)	0.193 (0.145)	0.224 (0.148)	0.254 (0.241)
17	1,452	55.554 (70.151)	0.193 (0.151)	0.241 (0.151)	0.268 (0.245)
18	1,386	59.522 (68.341)	0.194 (0.150)	0.229 (0.160)	0.267 (0.251)
19	1,105	54.219 (65.948)	0.188 (0.147)	0.233 (0.153)	0.268 (0.258)
20	1,111	57.224 (69.184)	0.197 (0.154)	0.234 (0.151)	0.242 (0.239)
21	1,067	53.931 (68.165)	0.193 (0.146)	0.236 (0.157)	0.292 (0.261)
22	907	54.887 (61.950)	0.199 (0.149)	0.231 (0.153)	0.285 (0.248)
23	975	56.111 (69.816)	0.206 (0.149)	0.244 (0.149)	0.260 (0.237)
24	887	52.140 (63.502)	0.207 (0.152)	0.229 (0.150)	0.238 (0.252)
25	664	59.144 (69.410)	0.187 (0.150)	0.237 (0.166)	0.287 (0.258)
26	502	82.972 (107.215)	0.197 (0.145)	0.248 (0.160)	0.267 (0.248)
27	646	61.558 (80.524)	0.200 (0.153)	0.238 (0.159)	0.255 (0.243)
28	819	56.102 (64.356)	0.186 (0.148)	0.230 (0.149)	0.268 (0.259)
29	553	60.065 (78.151)	0.190 (0.149)	0.236 (0.159)	0.293 (0.275)
30	504	59.618 (72.213)	0.201 (0.149)	0.224 (0.154)	0.296 (0.267)
31	377	66.020 (72.248)	0.203 (0.157)	0.247 (0.157)	0.274 (0.257)
All	38,378	58.181 (74.778)	0.198 (0.150)	0.231 (0.155)	0.267 (0.251)

MAF, minor allelic frequencies; ECA, equine chromosome; SD, standard deviation
JP, Jeju pony; THB, Thoroughbred
*Average differences of allelic frequencies between JP and THB for individual locus

respectively. The mean values of H_O, H_E and PIC in the THB were 0.312 ± 0.175, 0.307 ± 0.170 and 0.246 ± 0.126, respectively. The JP showed lower mean values for the three categories (H_O, H_E and PIC) than that of the THB. ECA 9 H_E was the highest in the JP (0.290 ± 0.167) and ECA 28 H_E had the lowest value (0.259 ± 0.169). Meanwhile, ECA 23 had the highest value (0.324 ± 0.159), and ECA 1 had the lowest value (0.290 ± 0.173) in the THB.

The 38,099 10-SNP windows were analyzed to identify the genomic region that showed the highest divergence of allelic frequency between JP and THB. There were 368 windows analyzed for ECA 31 and 2,913 for ECA 1. The mean size of windows was estimated at 527.6 kb (± 223.3). The ECA 24 (465.1 kb ± 171.7) was the smallest and ECA 26 (743.3 kb ± 287.8) was the largest (Table 3). The mean sliding window average difference (SWAD) of all windows was estimated at 0.141 (± 0.074) and had a range between 0.003 and 0.743. The mean SWAD of each ECA ranged between 0.109 and 0.172. This study only considered SWAD values greater than 0.499 to identify the regions with the highest SWAD for a single window (Fig. 1). The selection of 43 (top 0.1%) 10-SNP windows was allowed and included the highest SWAD value (0.743) observed in ECA 17. Among them, 16 windows were assigned to ECA 17, 11 windows were assigned to ECA 28, eight windows were assigned to ECA 6, and the other eight windows were assigned to ECA 1, 14, 30 and 31, respectively. The overlapping windows were merged into 10 genomic regions located on 7 different autosomes. The genomic coordinates of the regions and their sizes are shown in Table 4.

The 10 genomic regions that potentially possessed the selection signatures contained the 19 protein coding genes or overlapped with them (Table 4). Ten genes were identified in ECA 17 and had the highest SWAD value and 4 genes were identified in ECA 28 in the assigned regions. The 19 candidate genes were detected through functional annotation as the PANTHER system analyzed the gene ontology (GO) terms from three categories: biology process: 55%, cellular components: 28%, and molecular functions: 17%. Additionally, 19 genes were recognized by the system and accounted for 32% in protein binding, 26% in cellular processes, 16% in metabolic processes, and the remainder was in biological regulation, cell signaling, transporter activity, enzyme regulator activity, response to a stimulus, immune response and others (data not shown).

Table 2: Average genetic indexes in Jeju pony (JP) and Thoroughbred (THB) for individual locus by each chromosome

ECA	SNP (n)	*Ho* (± SD)		*He* (± SD)		PIC (± SD)	
		JP	THB	JP	THB	JP	THB
1	2,922	0.275 (0.169)	0.291 (0.175)	0.273 (0.166)	0.290 (0.173)	0.222 (0.123)	0.233 (0.129)
2	2,141	0.282 (0.176)	0.310 (0.169)	0.273 (0.168)	0.307 (0.165)	0.222 (0.124)	0.247 (0.122)
3	1,870	0.286 (0.173)	0.311 (0.165)	0.280 (0.167)	0.307 (0.161)	0.226 (0.123)	0.247 (0.119)
4	1,864	0.282 (0.171)	0.316 (0.183)	0.275 (0.164)	0.312 (0.178)	0.224 (0.121)	0.248 (0.132)
5	1,728	0.280 (0.183)	0.310 (0.179)	0.269 (0.173)	0.309 (0.177)	0.218 (0.128)	0.246 (0.130)
6	1,483	0.275 (0.171)	0.300 (0.173)	0.273 (0.166)	0.297 (0.170)	0.222 (0.123)	0.239 (0.127)
7	1,642	0.281 (0.173)	0.321 (0.178)	0.275 (0.166)	0.314 (0.172)	0.224 (0.123)	0.250 (0.127)
8	1,675	0.286 (0.170)	0.313 (0.172)	0.284 (0.167)	0.310 (0.167)	0.230 (0.123)	0.248 (0.123)
9	1,462	0.298 (0.175)	0.295 (0.174)	0.290 (0.167)	0.292 (0.170)	0.234 (0.123)	0.235 (0.126)
10	1,451	0.277 (0.175)	0.317 (0.180)	0.271 (0.167)	0.309 (0.173)	0.220 (0.124)	0.246 (0.128)
11	1,127	0.277 (0.167)	0.323 (0.177)	0.277 (0.167)	0.314 (0.169)	0.225 (0.123)	0.251 (0.124)
12	516	0.285 (0.172)	0.313 (0.182)	0.278 (0.164)	0.309 (0.177)	0.226 (0.122)	0.246 (0.132)
13	701	0.271 (0.170)	0.311 (0.174)	0.269 (0.166)	0.309 (0.169)	0.219 (0.123)	0.247 (0.125)
14	1,706	0.280 (0.172)	0.306 (0.176)	0.275 (0.165)	0.303 (0.173)	0.223 (0.123)	0.242 (0.129)
15	1,603	0.282 (0.175)	0.317 (0.177)	0.273 (0.167)	0.310 (0.170)	0.222 (0.124)	0.248 (0.126)
16	1,532	0.273 (0.168)	0.310 (0.166)	0.270 (0.164)	0.304 (0.162)	0.220 (0.122)	0.245 (0.120)
17	1,452	0.272 (0.174)	0.326 (0.174)	0.266 (0.168)	0.320 (0.167)	0.216 (0.125)	0.255 (0.124)
18	1,386	0.275 (0.175)	0.308 (0.184)	0.268 (0.167)	0.302 (0.177)	0.218 (0.124)	0.241 (0.132)
19	1,105	0.269 (0.174)	0.316 (0.174)	0.262 (0.166)	0.311 (0.168)	0.214 (0.123)	0.248 (0.125)
20	1,111	0.277 (0.180)	0.313 (0.167)	0.269 (0.171)	0.313 (0.165)	0.218 (0.127)	0.250 (0.122)
21	1,067	0.275 (0.170)	0.314 (0.180)	0.269 (0.162)	0.311 (0.174)	0.220 (0.119)	0.247 (0.131)
22	907	0.279 (0.171)	0.316 (0.176)	0.275 (0.166)	0.308 (0.168)	0.223 (0.122)	0.247 (0.125)
23	975	0.289 (0.173)	0.328 (0.163)	0.282 (0.166)	0.324 (0.159)	0.229 (0.123)	0.259 (0.117)
24	887	0.284 (0.170)	0.312 (0.170)	0.282 (0.167)	0.308 (0.163)	0.228 (0.123)	0.247 (0.121)
25	664	0.268 (0.180)	0.316 (0.188)	0.260 (0.170)	0.307 (0.180)	0.212 (0.127)	0.244 (0.134)
26	502	0.278 (0.171)	0.327 (0.181)	0.275 (0.166)	0.322 (0.173)	0.223 (0.123)	0.255 (0.129)
27	646	0.280 (0.178)	0.317 (0.179)	0.273 (0.171)	0.312 (0.172)	0.221 (0.127)	0.249 (0.127)
28	819	0.263 (0.173)	0.309 (0.166)	0.259 (0.169)	0.310 (0.164)	0.211 (0.126)	0.248 (0.121)
29	553	0.270 (0.176)	0.314 (0.182)	0.263 (0.169)	0.310 (0.177)	0.214 (0.126)	0.246 (0.132)
30	504	0.281 (0.172)	0.304 (0.175)	0.277 (0.167)	0.300 (0.170)	0.225 (0.124)	0.241 (0.127)
31	377	0.282 (0.178)	0.329 (0.172)	0.274 (0.169)	0.323 (0.168)	0.222 (0.124)	0.257 (0.124)
All	38,378	0.279 (0.173)	0.312 (0.175)	0.273 (0.167)	0.307 (0.170)	0.222 (0.124)	0.246 (0.126)

[1] Observed heterozygosity
[2] Expected heterozygosity
[3] Polymorphism information contents
ECA, equine chromosome
JP, Jeju pony; THB, Thoroughbred

Table 3: The number of sliding windows analyzed, their size and mean sliding window average difference (SWAD) in the allele frequencies between Jeju pony (JP) and Thoroughbred (THB)

ECA	No. of 10-SNP windows analyzed	Mean window size (kb)	Mean window size SD (kb)	SWAD	SWAD SD
1	2,913	571.8	317.0	0.159	0.081
2	2,132	506.8	212.2	0.139	0.069
3	1,861	575.2	380.3	0.113	0.060
4	1,855	525.1	215.8	0.153	0.072
5	1,719	519.3	229.6	0.142	0.070
6	1,474	487.0	199.8	0.153	0.086
7	1,633	539.6	333.9	0.150	0.078
8	1,666	503.0	197.1	0.120	0.066
9	1,453	512.5	203.2	0.136	0.071
10	1,442	520.7	202.6	0.137	0.067
11	1,118	490.8	189.1	0.151	0.072
12	507	575.7	275.5	0.128	0.068
13	692	546.0	257.3	0.117	0.059
14	1,697	490.7	200.7	0.152	0.075
15	1,594	512.6	231.3	0.130	0.061
16	1,523	513.1	229.0	0.137	0.069
17	1,443	501.2	209.9	0.155	0.090
18	1,377	537.2	204.9	0.143	0.074
19	1,096	486.7	198.2	0.137	0.065
20	1,102	515.0	187.9	0.122	0.052
21	1,058	484.0	178.6	0.149	0.070
22	898	494.8	160.9	0.172	0.089
23	966	504.2	220.2	0.145	0.079
24	878	465.1	171.7	0.109	0.054
25	655	530.2	192.4	0.157	0.079
26	493	743.3	287.8	0.143	0.069
27	637	545.6	227.4	0.128	0.056
28	810	502.8	177.3	0.144	0.096
29	544	536.8	211.0	0.139	0.059
30	495	529.5	203.0	0.161	0.080
31	368	590.5	216.5	0.145	0.072
All	38,099	527.6	223.3	0.141	0.074

ECA, equine chromosome; SD, standard deviation

Furthermore, the GeneMANIA analysis tool was used to search for pathways based on the GO term categories through functional annotation for each of the 14 candidate genes, including *CAB39L*, *SETDB2*, *PHF11*, *RCBTB1*, *ARL11*, *EBPL*, *KPNA3*, *SPRYD7*, *TRIM13*, *KCNRG*, *TMTC3*, *CEP290*, *C12orf29*, and *C12orf50*, located on Chromosomes 17 and 28, which had the highest SWAD values in this study. This study also investigated the direct gene networks among the genes. As a result, we found that 12 candidate genes, excluding *ARL11* and *C12orf50*, performed gene networks. The 4 genes, *CAB39L*, *KPNA*, *PHF11*, and *CEP290*, were reported to be involved in significant biological pathways (Fig. 2).

Discussion

This equine genome sequencing study confirmed the identification of many SNPs and the high-density SNP panels made by arranging these SNPs serve as a highly effective tool for genome analysis. This technique has been widely adopted in numerous studies. For example, because of the continuous development of NGS technology (Meira *et al*., 2013; Frischknecht *et al*., 2016), the use of genome-wide association study techniques are increasing to explore the genetic factors on several traits of horses, such as exercise ability, robustness and disease susceptibility in the whole genome. Additionally, studies on other animals have shown approaches capable of detecting genomic regions influenced by artificial selection by calculating the allelic frequency instead of directly examining the association between traits of diversely selected breed groups (Prasad *et al*., 2008; Hayes *et al*., 2009; Wilkinson *et al*., 2013). Recently, studies have identified the genomic regions and the candidate genes that are expected to reflect different artificial selections by comparing allelic frequency of different horse breeds through statistical analysis (McCue *et al*., 2012; Metzger *et al*., 2015). This approach suggests that a breed can potentially express the genomic regions that can add the selection pressures that influence the traits that distinguishes between breeds, which can then be used to pinpoint the candidate genes.

In this study, JP and THB, which are two genetically distinct horse breeds, were used to identify genomic regions that represent the highest divergence in allelic frequency. We analyzed 38,378 SNPs in both breeds, and the JP breed had lower average genetic index values (H_O, H_E and PIC) than those of the THB. The measures of PIC were obtained using the values of H_E and represented the polymorphisms on each gene locus (Botstein *et al*., 1980). Moreover, the JP breed had a lower observed MAF and average difference of allelic frequency than that of the counterpart. The JP breed is predicted to have higher homozygosity than that of the THB. The THB breed largely accounts for the horse industry. A closed stud book and race capability-oriented strict selection was performed using inbreeding, dramatically narrowing the genetic diversity in this population (Cunningham *et al*., 2001). Similarly, the JP is expected to show extremely low genetic diversity with consideration to the geographical features of Jeju Island where this species has long resided. Additional research is necessary to examine the genetic problems involved in preserving the JP.

Similar to the results of previous studies, our study identified 19 genes in the chromosomal regions of the 10 discovered candidate genes. A previous study that used DNA-resequencing technology with 10 JPs and 10 THBs revealed that the ECA 6 (31,153,575–31,314,522 bp), ECA 14 (42,345,298–42,376,306 bp), ECA 17 (20,728,652–21,789,288 bp), and ECA 30 (24,994,783–25,103,174 bp) regions were identical to the analyzed regions (Kim *et al*., 2013). *TSPAN9*, *TSPAN11*, *FSTL4*, and *KPNA3* were determined to be strongly and negatively selected genes for evolution. Conversely, *CAB39L* and *CRB1* were determined to be strongly and positively selected genes. The *KCNRG* gene was determined to represent a specific allele frequency change in the THB.

Additionally, the ECA 17 (20,728,652–21,789,288 bp) and ECA 28 (14,306,397–14,663,060 bp) regions were determined to be the highest SWAD (0.74266 and 0.72249, respectively). There were 12 genes that conducted the networks among the 14 genes in both regions and 4—*CAB39L*, *KPNA*, *PHF11* and *CEP290*—were observed to engage in important biological pathways.

First, the *CAB39L* (calcium binding protein 39-like) gene was involved in energy mechanism, growth factor, disease, and immunity mechanisms, such as insulin receptor signaling by Type 1 Insulin-like Growth Factor 1 Receptor (IGF1R), regulation of AMPK activity *via* LKB1, energy dependent regulation of mTOR by LKB1-AMPK, and mTOR signaling. As mentioned above, *CAB39L* was determined to be a strong and positively selected gene for horse evolution. Protein CAB39L is reported as MO25beta and encodes the *CAB39L* and is represented in the nervous system and sensory organs in mice. It facilitates phosphorylation such as serine/threonine kinase 11/liver kinase (B1STK11/LKB1) that activates AMP-activated protein kinase (AMPK) and functions as a gene network, thereby regulating feeding (Hawley *et al*., 2003; Minokoshi *et al*., 2004). In addition, this gene was reported to influence feed intake, growth, and body weight in chickens and pigs (Zhang *et al*., 2011; Yuan *et al*., 2015; Yang *et al*., 2016).

Second, KPNA3 (karyopherin alpha 3) was found to participate in an antiviral mechanism by IFN-stimulated genes and the ISG15 antiviral mechanism pathway and was determined to be a strong and negatively selected gene for horse evolution. KPNA3 also plays a central role in intracellular trafficking and secretion. It mediates the transport of molecules between the nucleus and the cytoplasm in eukaryotic cells and is mediated by the nuclear pore complex (NPC), which consists of 60–100 proteins. Small molecules (up to 70 kDa) can pass through the nuclear pore by nonselective diffusion while larger molecules are transported by an active process.

Table 4: Summary of 10 candidate chromosomal regions showing the highest sliding window average difference (SWAD)

ECA	Highest SWAD [a]	Start position [b]	End position [b]	Region size (kb)	No. of encompass or overlapped genes in the region (Ensembl)
1	0.52007	45,290,499	45,314,694	24.2	0
6	0.56629	23,136,288	23,288,105	151.8	1: COPS8
	0.55252	31,153,575	31,314,522	160.9	2: TSPAN9*, TSPAN11*
14	0.51905	42,345,298	42,376,306	31.0	1: FSTL4*
17	0.74266	20,728,652	21,789,288	1060.6	10: CAB39L*, SETDB2, PHF11, RCBTB1, ARL11, EBPL, KPNA3*, SPRYD7, TRIM13, KCNRG*
	0.50793	23,171,546	23,173,987	2.4	0
	0.53750	64,015,784	64,031,892	16.1	0
28	0.72249	14,306,397	14,663,060	356.7	4: TMTC3, CEP290, C12orf29, C12orf50
30	0.52234	24,994,783	25,103,174	108.4	1: CRB1*
31	0.51202	4,692,129	4,708,150	16.0	0

[a] Highest value of sliding window average difference (SWAD) observed in each candidate region
[b] Chromosomal position was indicated as the map position from the most centromeric marker to the most distal one in each region based on horse genome assembly (Equ Cab 2.0)
* Marked included candidate genes by previously reported (Kim *et al*., 2013)
ECA, equine chromosome

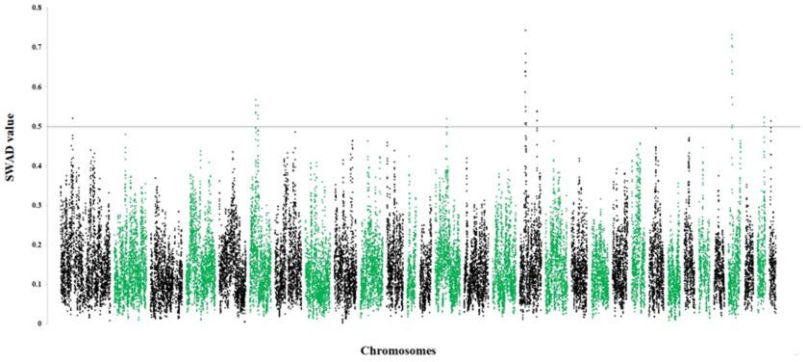

Fig. 1: Sliding window average differences (SWAD) on all Chromosomes. Solid line indicates the top 0.1 percentile threshold; 0.499 respectively

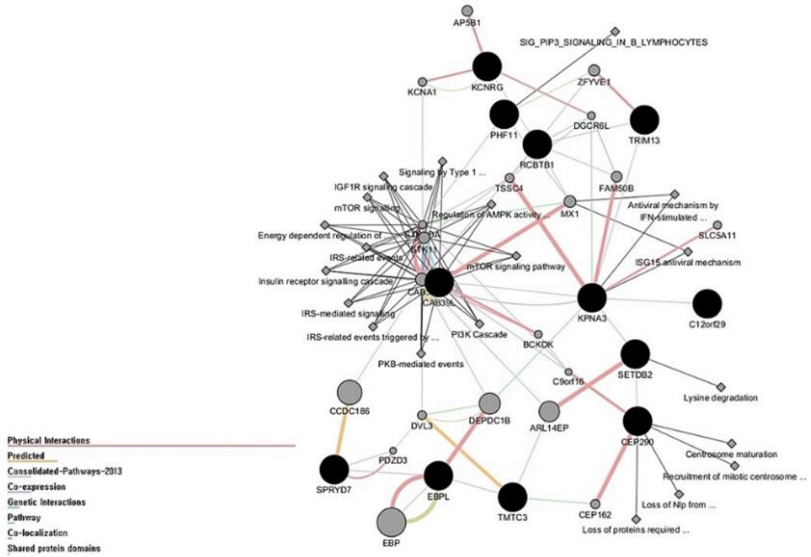

Fig. 2: Gene network and pathway analysis of the 12 candidate genes and their regulatory relationships. Black marked genes and pathway indicates candidate genes from gene network and pathway analysis

The protein encoded by this gene belongs to the important alpha family and is involved in nuclear protein import (Takeda *et al*., 1997). Genetic functions of KPNA3 include the muscle contraction/extension of cells. A previous study showed that the expression patterns of skeletal muscle transcriptome, KPNA3, had significant down-regulated

effects during horse post-training when compared to that during pre-training (McGivney *et al*., 2010). Moreover, this gene was found to regulate the slow expression of fiber in skeletal muscle contraction during pig adulthood (Siengdee *et al*., 2013).

Lastly, PHF11 (PHD Finger Protein 11) and CEP290 (centrosomal protein 290 kDa) are involved in the immune response. PHF11 was reported to be involved in pathways of PIP3 signaling in B lymphocytes. CEP290 was involved in pathways of centrosome maturation, recruitment of mitotic centrosome proteins and complexes, loss of proteins required for interphase microtubule organization from the centrosome, and the ISG15 antiviral mechanism. PHF11 was known to play a pivotal role in the modulation of expression of the Th1-specific cytokines, interferon-gamma, and interleukin-2 and CEP290 was associated with several pathologies that play a role in the centrosome and cilia development (Zhang *et al*., 2003; Coppieters *et al*., 2010).

Conclusion

Our study showed divergently selected genomic regions which may be associated with different biological and phenotypical features between THB and JP horse breeds. Identified 10 genomic regions may affect both selection signatures and some changes, and detected 4 candidate genes (CAB39L, KPNA, PHF11 and CEP290) in those regions potentially involved in growth, exercise ability, epidemics, and immunity traits. Further studies are necessary to validate 4 candidate genes.

Acknowledgement

This work was supported by a grant from the Next-Generation BioGreen 21 Program (Project No. PJ01332901), Rural Development Administration, Republic of Korea.

References

Botstein, D., R.L. White, M. Skolnick and R.W. Davis, 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Amer. J. Hum. Genet*., 32: 314–331

Bower, M.A., B.A. McGivney, M.G. Campana, J. Gu, L.S. Andersson, E. Barrett, C.R. Davis, S. Mikko, F. Stock, V. Voronkova, D.G. Bradley, A.G. Fahey, G. Lindgren, D.E. MacHugh, G. Sulimova and E.W. Hill, 2012. The genetic origin and history of speed in the Thoroughbred racehorse. *Nat. Commun*., 3: 1–8

Coppieters, F., S. Lefever, B.P. Leroy and E.D. Baere, 2010. CEP290, a gene with many faces: Mutation overview and presentation of CEP290base. *Hum. Mutat.*, 31: 1097–1108

Corbin, L.J., S.C. Blott, J.E. Swinburne, M. Vaudin, S.C. Bishop and J.A. Woolliams, 2010. Linkage disequilibrium and historical effective population size in the thoroughbred horse. *Anim. Genet*., 41: 8–15

Cunningham, E.P., J.J. Dooley, R.K. Splan and D.G. Bradley, 2001. Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to thoroughbred horses. *Anim. Genet*., 32: 360–364

Do, K.T., J.H. Lee, H.K. Lee, J. Kim and K.D. Park, 2014. Estimation of effective population size using single-nucleotide polymorphism (SNP) data in Jeju horse. *J. Anim. Sci. Tecchnol*., 56: 28

Frischknecht, M., H. Signer-Hasler, T. Leeb, S. Rieder and M. Neuditschko, 2016. Genome-wide association studies based on sequence-derived genotypes reveal new QTL associated with conformation and performance traits in the Franches–Montagnes horse breed. *Anim. Genet*., 47: 227–229

Go, Y.Y., E. Bailey, D.G. Cook, S.J. Coleman, J.N. MacLeod, K.C. Chen, P.J. Timoney and U.B.R. Balasuriya, 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

Gurgul, A., K. Pawlina, M. Frys-Żurek and M. Bugno-Poniewierska, 2015. Identification of differential selection traces in two Polish cattle breeds. *Anim. Sci. J*., 86: 17–24

Hawley, S.A., J. Boudeau, J.L. Reid, K.J. Mustard, L. Udd, T.P. Mäkelä, D.R. Alessi and D.G. Hardie, 2003. Complexes between the LKB1 tumor suppressor, STRADalpha/beta and MO25alpha/beta are upstream kinases in the AMP-activated protein kinase cascade. *J. Biol*., 2: 28

Hayes, B.J., A.J. Chamberlain, S. Maceachern, K. Savin, H. McPartlan, I. MacLeod, L. Sethuraman and M.E. Goddard, 2009. A genome map of divergent artificial selection between *Bos taurus* dairy cattle and *Bos taurus* beef cattle. *Anim. Genet*., 40: 176–184

Hosokawa, D., A. Ishii, K. Yamaji, S. Sasazaki, K. Oyama and H. Mannen, 2012. Identification of divergently selected regions between Japanese Black and Holstein cattle using bovine 50 k SNP array. *Anim. Sci. J*., 83: 7–13

Huang, J., Y. Zhao, W. Shiraigol, B. Li, D. Bai, W. Ye, D. Daidiikhuu, L. Yang, B. Jin, Q. Zhao, Y. Gao, J. Wu, W. Bao, A. Li, Y. Zhang, H. Han, H. Bai, Y. Bao, L. Zhao, Z. Zhai, W. Zhao, Z. Sun, Y. Zhang, H. Meng and M. Dugarjaviin, 2014. Analysis of horse genomes provides insight into the diversification and adaptive evolution of karyotype. *Sci. Rep*., 4: 1–8

Kim, H., T. Lee, W.C. Park, J.W. Lee, J. Kim, B.Y. Lee, H. Ahn, S. Moon, S. Cho, K.T. Do, H.S. Kim, H.K. Lee, C.K. Lee, H.S. Kong, Y.M. Yang, J. Park, H.M. Kim, B.C. Kim, S. Hwang, J. Bhak, D. Burt, K.D. Park, B.W. Cho and H. Kim, 2013. Peeling back the evolutionary layers of molecular mechanisms responsive to exercise-stress in the skeletal muscle of the racing horse. *DNA Res*., 20: 287–298

Kitano, H., 2002. Systems biology: a brief overview. *Science*, 295: 1662–1664

Lee, J.H., K.D. Song, J.M. Kim, H.K. Leem and K.D. Park, 2016. Identification of genes with nonsynonymous SNP in Jeju horse by whole-genome resequencing reveals a functional role for immune response. *J. Anim. Sci*., 94: 895–901

Lim, B., S. Kim, D. Kang, K.D. Park and K. Do, 2019. Research Trends and Implications for Preservation and Improvement of Jeju Ponies (Review). *Ann. Anim. Resour. Sci*., 30: 34–48

Lippold, S., N.J. Matzke, M. Reissmann and M. Hofreiter, 2011. Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. *BMC Evol. Biol*., 11: 328

Ludwig, A., M. Pruvost, M. Reissmann, N. Benecke, G.A. Brockmann, P. Castaños, M. Cieslak, S. Lipplod, L. Llorente, A.S. Malaspinas, M. Slatkin and M. Hofreiter, 2009. Coat color variation at the beginning of horse domestication. *Science*, 324: 1–4

McCue, M.E., D.L. Bannasch, J.L. Petersen, J. Gurr, E. Bailey, M.M. Binns, O. Distl, G. Guérin, T. Hasegawa, E.W. Hill, T. Leeb, G. Lindgren, M.C.T. Penedo, K.H. Røed, O.A. Ryder, J.E. Swinburne, T. Tozaki, S.J. Valberg, M. Vaudin, K. Lindblad-Toh, C.M. Wade and J.R. Mickelson, 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: 1–14

McGivney, B.A., P.A. McGettigan, J.A. Browne, A.C. Evans, R.G. Fonseca, B.J. Loftus, A. Lohan, D.E. MacHugh, B.A. Murphy, L.M. Katz and E.W. Hill, 2010. Characterization of the equine skeletal muscle transcriptome identifies novel functional responses to exercise training. *BMC Genomics*, 11: 398

Meira, C.T., M.R.S. Fortes, M.M. Farah, L.R. Porto-Neto, R.A. Curi, S.S. Moore and M.D.S. Mota, 2013. A genome-wide association study for height at withers in racing quarter horse. *Proc. Assoc. Adv. Anim. Breed. Genet*., 20: 420–423

Metzger, J., M. Karwath, R. Tonda, S. Beltran, L. Águeda, M. Gut, I.G. Gut and O. Distl, 2015. Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC Genom*., 16: 764

Minokoshi, Y., T. Alquier, N. Furukawa, Y.B. Kim, A. Lee, B. Xue, J. Mu, F. Foufelle, P. Ferré, M.J. Birnbaum, B.J. Stuck and B.B. Kahn, 2004. AMP-kinase regulates food intake by responding to hormonal and nutrient signals in the hypothalamus. *Nature*, 428: 569–574

Nam, D.Y., 1969. *Horse production in Cheju during Lee dynasty*, 4th edition. Korea History Research Society. Republic of Korea

Nei, M. and A.K. Roychoudhury, 1974. Sampling variances of heterozygosity and genetic distance. *Genetics*, 76: 379–390

Outram, A.K., N.A. Stear, R. Bendrey, S. Olsen, A. Kasparov, V. Zaibert, N. Thorpe and R.P. Evershed, 2009. The earliest horse harnessing and milking. *Science*, 323: 1332–1335

Park, K.D., J. Park, J. Ko, B.C. Kim, H.S. Kim, K. Ahn, K.T. Do, H. Choi, H.M. Kim, S. Song, S. Lee, S. Jho, H.S. Kong, Y.M. Yang, B.H. Jhun, C. Kim, T.H. Kim, S. Hwang, J. Bhak, H.K. Lee and B.W. Cho, 2012. Whole transcriptome analyses of six thoroughbred horses before and after exercise using RNA-Seq. *BMC Genomics*, 13: 473

Park, W., J. Kim, H.J. Kim, J. Choi, J.W. Park, H.W. Cho, B.W. Kim, M.H. Park, T.S. Shin, S.K. Cho, J.K. Park, H. Kim, J.Y. Hwang, C.K. Lee, H.K. Lee, S. Cho and B.W. Choi, 2014. Investigation of *de novo* unique differentially expressed genes related to evolution in exercise response during domestication in thoroughbred race horses. *PLoS One*, 9: 1–12

Petersen, J.L., J.R. Mickelson, A.K. Rendahl, S.J. Valberg, L.S. Andersson, J. Axelsson, E. Bailey, D. Bannasch, M.M. Binns, A.S. Borges, P. Brama, A.D.C. Machado, S. Capomaccio, K. Cappelli, E.G. Cothran, O. Distl, L. Fox-Clipsham, K.T. Graves, G. Guérin, B. Haase, T. Hasegawa, K. Hemmann, E.W. Hill, T. Leeb, G. Lindgren, H. Lohi, M.S. Lopes, B.A. McGivney, S. Mikko, N. Orr, M.C.T. Penedo, R.J. Piercy, M. Raekallio, S. Rieder, K.H. Røed, J. Swinburne, T. Tozaki, M. Vaudin, C.M. Wade and M.E. McCue, 2013a. Genome-wide analysis reveals selection for important traits in domestic horse breeds. *PLoS Genet*., 9: 1–17

Petersen, J.L., J.R. Mickelson, E.G. Cothran, L.S. Andersson, J. Axelsson, E. Bailey, D. Bannasch, M.M. Binns, A.S. Borges, P. Brama, A.D.C. Machado, O. Distl, M. Felicetti, L. Fox-Clipsham, K.T. Graves, G. Guérin, B. Haase, T. Hasegawa, K. Hemmann, E.W. Hill, T. Leeb, G. Lindgren, H. Lohi, M.S. Lopes, B.A. McGivney, S. Mikko, N. Orr, M.C.T. Penedo, R.J. Piercy, M. Raekallio, S. Rieder, K.H. Røed, M. Silvestrelli, J. Swinburne, T. Tozaki, M. Vaudin, C.M. Wade and M.E. McCue, 2013b. Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLoS One*, 8: 1–15

Prasad, A., R.D. Schnabel, S.D. McKay, B. Murdoch, P. Stothard, D. Kolbehdari, Z. Wang, J.F. Taylor and S.S. Moore, 2008. Linkage disequilibrium and signatures of selection on chromosomes 19 and 29 in beef and dairy cattle. *Anim. Genet*., 39: 597–605

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

Shannon, P., A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genom. Res*., 13: 2498–2504

Siengdee, P., N. Trakooljul, E. Murani, M. Schwerin, K. Wimmers and S. Ponsuksili, 2013. Transcriptional profiling and miRNA-dependent regulatory network analysis of longissimus dorsi muscle during prenatal and adult stages in two distinct pig breeds. *Anim. Genet*., 44: 398–407

Stella, A., P. Ajmone-Marsan, B. Lazzari and P. Boettcher, 2010. Identification of selection signatures in cattle breeds selected for dairy production. *Genetics*, 185: 1451–1461

Takeda, S., T. Fujiwara, F. Shimizu, A. Kawai, K. Shinomiya, S. Okuno, K. Ozaki, T. Katagiri, Y. Shimada, M. Nagata, T. Watanabe, A. Takaichi, Y. Kuga, M. Suzuki, H. Hishigaki, E. Takahashi, S. Shin, Y. Nakamura and Y. Hirai, 1997. Isolation and mapping of karyopherin α3 (KPNA3), a human gene that is highly homologous to genes encoding Xenopus importin, yeast SRP1 and human RCH1. *Cytogenet. Genomics Res*., 76: 87–93

Thiruvenkadan, A.K., N. Kandasamy and S. Panneerselvam, 2009. Inheritance of racing performance of Thoroughbred horses. *Livest. Sci*., 121: 308–326

Thomas, P.D., M.J. Campbell, A. Kejariwal, H. Mi, B. Karlak, R. Daverman, K. Diemer, A. Muruganujan and A. Narechania, 2003. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res*., 13: 2129–2141

Wilkinson, S., Z.H. Lu, H.J. Megens, A.L. Archibald, C. Haley, I.J. Jackson, M.A.M. Groenen, R.P.M.A. Crooijmans, R. Ogden and P. Wiener, 2013. Signatures of diversifying selection in European pig breeds. *PLoS Genet*., 9: 1–15

Yang, Y., R. Zhou, Y. Mu, X. Hou, Z. Tang and K. Li, 2016. Genome-wide analysis of DNA methylation in obese, lean and miniature pig breeds. *Sci. Rep*., 6: 1–12

Yuan, J., K. Wang, G. Yi, M. Ma, T. Dou, C. Sun, L.J. Qu, M. Shen, L. Qu and N. Yang, 2015. Genome-wide association studies for feed intake and efficiency in two laying periods of chickens. *Genet. Select. Evol*., 47: 82

Zhang, H., S.H. Liu, Q. Zhang, Y.D. Zhang, S.Z. Wang, Q.G. Wang, Y.X. Wang, Z.Q. Tang and H. Li, 2011. Fine-mapping of quantitative trait loci for body weight and bone traits and positional cloning of the RB1 gene in chicken. *J. Anim. Breed. Genet*., 128: 366–375

Zhang, Y., N.I. Leaves, G.G. Anderson, C.P. Ponting, J. Broxholme, R. Holt, P. Edser, S. Bhattacharyya, A. Dunham, I.M. Adcock, L. Pulleyn, P.J. Barnes, J.I. Harper, G. Abecasis, L. Cardon, M. White, J. Burton, L. Matthews, R. Mott, M. Ross, R. Cox, M.F. Moffatt and W.O.C.M. Cookson, 2003. Positional cloning of a quantitative trait locus on chromosome 13q14 that influences immunoglobulin E levels and asthma. *Nat. Genet*., 34: 181–186

[Received 19 Apr 2019; Accept 25 Jun 2019; Published (online) 10 Nov 2019]