



A Genome-wide Scan for Selective Sweeps in Racing Horses

Sunjin Moon^{1,a}, Jin Woo Lee^{2,a}, Donghyun Shin¹, Kwang-Yun Shin³, Jun Kim³,
Ik-Young Choi⁴, Jaemin Kim⁵, and Heebal Kim^{1,5,6,*}

¹ Department of Agricultural Biotechnology and Research Institute for Agriculture and Life Sciences,
Seoul National University, Seoul 151-742, Korea

ABSTRACT: Using next-generation sequencing, we conducted a genome-wide scan of selective sweeps associated with selection toward genetic improvement in Thoroughbreds. We investigated potential phenotypic consequence of putative candidate loci by candidate gene association mapping for the finishing time in 240 Thoroughbred horses. We found a significant association with the trait for Ral GAPase alpha 2 (RALGAP2) that regulates a variety of cellular processes of signal trafficking. Neighboring genes around *RALGAP2* included insulinoma-associated 1 (*INSMI*), pallid (*PLDN*), and Ras and Rab interactor 2 (*RIN2*) genes have similar roles in signal trafficking, suggesting that a co-evolving gene cluster located on the chromosome 22 is under strong artificial selection in racehorses. (**Key Words:** Single Nucleotide Polymorphism, Racehorse, Selective Sweep, Quantitative Trait Loci)

INTRODUCTION

The Thoroughbred is best known as a racehorse breed. They have been bred exclusively for structural and functional adaptations that contribute to athletic performance phenotypes (Williamson and Beilharz, 1998) since their establishment about 300 to 400 years ago (Cunningham et al., 2001; Hill et al., 2002). Various traits, such as racing time, rank, annual earnings, and others, are used to measure a horse's racing performance (Ricard, 1998). Previous genome-wide association studies (GWAS) have identified a handful of genetic markers in the

myostatin (*MSTN*) (Grobet et al., 2003) with racing performance (Binns et al., 2010; Hill et al., 2010; Tozaki et al., 2010). In dogs, mutations in *MSTN* were found to be associated with athletic performance (Tozaki et al., 2010; McGivney et al., 2012). A scan for positive selection based on a microsatellite data detected hundreds of domesticated genes in horse (Gu et al., 2009), showing that *MSTN* is not the only determinant of athletic prowess (Lee, 2007). Recently, comprehensive analysis of equine breed diversity suggested that genomic footprints of selection are important candidates of GWAS (Petersen et al., 2013; Park et al., 2014).

Several association studies followed by selective sweep mapping found genetic markers associated with the complex traits in dog (Vaysse et al., 2011) and cattle (Qanbari et al., 2014). Motivated by these approaches, we aimed to use selective sweeps to detect markers associated with artificial selection for racing performance in Thoroughbreds. To this end, we compared genetic diversity between Thoroughbreds and Jeju domestic breed. Jeju breed is a close relative of the ancestral horse population (Nam, 1969; Kim et al., 1999; Jansen et al., 2002). We conducted a genome-wide scan to detect putative candidate loci showing high genetic differentiation between populations as well as low genetic diversity in Thoroughbreds. On obtaining

* Corresponding Author: Heebal Kim. Tel: +82-2-880-4803, Fax: +82-2-883-8812, E-mail: heebal@snu.ac.kr

² Horse Registry, Korea Racing Authority (KRA), Gwacheon 427-711, Korea.

³ Institute for Livestock Promotion, Jeju 690-802, Korea.

⁴ Genome analysis center, National Instrumentation and Environmental Management (NICEM), Seoul National University, Seoul 151-921, Korea.

⁵ Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 151-742, Korea.

⁶ C&K Genomics, Seoul 151-742, Korea.

^a These authors contributed equally to this work.

Submitted Sept. 8, 2014; Revised Nov. 9, 2014; Accepted Feb. 22, 2015

candidate loci under directional selection, we evaluated their potential effects by association study of finishing time in a large sample of Thoroughbreds. We found a large linkage-disequilibrium (LD) block of gene cluster on *Equus caballus* chromosome 22 (ECA22), mainly consisting of signal regulation-related genes, that is subject to artificial selection in the Thoroughbred.

MATERIALS AND METHODS

Ethics statement

Blood samples were taken from horses and ponies by trained veterinarians according to relevant international guidelines as well as national guidelines and under permission from the Korean Racing Authority and from the Jeju Provincial Livestock Institute, Korea.

Sample preparation and resequencing horse genomes

Sets of whole-blood samples were collected from 14 Thoroughbred race stallions from the Korean Racing Authority and 2 male Jeju domestic ponies (*Equus caballus*), Korea's National Treasure Number 347, from the Jeju Provincial Livestock Institute, Korea. Blood (10 mL) was drawn from the carotid artery and was treated with heparin to prevent clotting.

A genomic DNA quality check (DNA QC) was carried out using an agarose gel+fluorescence-based quantification check with a standard electrophoresis using a 0.6% agarose gel and 200 ng pulse-field gel loading. Briefly, mate-pair library construction (500 bp fragment), amplification and paired-end sequencing using Illumina HiSeq system (San Diego, CA, USA) was done by National Instrumentation Center for Environmental Management (NICEM), Seoul, Korea; the steps including purifying genomic DNA, generating fragments of less than 800 bp, using blunt-ended fragments with 5'-phosphorylated ends and 3'-dA overhang, ligating adapter-modified ends, purifying the ligation product, and producing the genomic DNA library, and then, sequence data were generated using the Illumina HiSeq system (San Diego, CA, USA). The data set of 14 Thoroughbred racing stallions and 2 'Jeju' lineage have been uploaded to the Sequence Read Archive (SRA) in NCBI (accession number: SRA053569).

Genotype calling from next-generation sequencing data

Pair-end sequence reads were mapped to the reference horse genome (equCab2) using the Burrows-Wheeler Aligner (BWA; version 0.6.1) with the default settings (Li and Durbin, 2009). Three open-source packages were used for downstream processing and variant calling: Picard tools, SAMtools (Anisimova and Kosiol, 2009), and Genome Analysis Toolkit (McKenna et al., 2010). Substitution calls were made with GATK UnifiedGenotyper. All calls with a

Phred-scaled quality score of less than 20 were filtered out. We identified alternative homozygous and heterozygous sites in respect to the reference genome in each sample variants. Then, we used BEAGLE to infer haplotype phase and impute missing alleles for the entire set of Thoroughbred genomes simultaneously (Browning and Browning, 2007).

Data analysis

For a limited number of samples available, especially for Jeju horse, a large number of variants can substitute for sample size when estimating F_{ST} (Willing et al., 2012). Horse genomes were divided into a large window of 50 kb that produced 47,308 bins covering ~87.6% of the whole genome. An unbiased estimation of nucleotide diversity (Nei, 1987) and population differentiation (Nei and Li, 1979) based on the pairwise difference between haplotypes was calculated for each bin by using Arlequin 3.5 (Excoffier and Lischer, 2010). The ratio of nucleotide diversity of the Jeju breed to that of the Thoroughbred was standardized to have a mean of 0 and standard deviation of 1. We adopted an empirical outlier approach to identify potential targets of directional selection by obtaining top 0.1% of highly divergent regions as well as top 0.1% of strong reduction of genetic diversity in Thoroughbreds (Excoffier et al., 2009).

Validation of associated loci on *Equus caballus* chromosome 22

A group of 240 Thoroughbreds that were registered in the Korea Racing Authority (KRA) for breeding as stallions was genotyped for validating selective sweeps using EquineSNP50 BeadChips (Illumina, San Diego, CA, USA) in the NICEM at Seoul National University. After the quality control steps, removing single nucleotide polymorphisms (SNPs) with a percent of missing allele >5%, minor allele frequency <5%, and Hardy-Weinberg equilibrium test $<10^{-3}$, 940 SNPs on ECA22 were obtained for association analysis that was conducted using linear regression in PLINK (Purcell et al., 2007). We employed a genomic control to correct for spurious association due to population stratification. The data set is available upon request.

Total records of 262,326 racing times between 1994 and 2011 in racecourses (Seoul Racecourse, Busan-Gyeongnam Racecourse) of the KRA were obtained. Each sample was assigned an estimated breeding value (EBV), which was defined as a statistical numerical prediction of the relative genetic breeding value and used to rank breeding stock for selection. When a horse has multiple records, the EBV can be used to combine the measurements in all records into one numerical value. To improve the accuracy of the EBV, we simplified the animal model by reducing unnecessary effects (racing year, racing type, and type of weight carried).

The repeated-records animal model used to estimate the genetic parameters (breeding value):

$$Y = Xb + W_1v + W_2p_e + Za + e$$

where Y = the vector of observations, b = the vector of fixed effects, v = the vector of random effects, p_e = the vector of permanent environmental effect: common environment, a = the vector of individual additive genetic effect, and e = the vector of residual error. X , W_1 , W_2 , and Z are design matrices for b , v , p_e , and a , respectively. All parameters were estimated using the ASREML program (Gilmour et al., 2009) facilitated by the derivative-free restricted maximum likelihood method for a single-trait animal model. Using this model, we calculated the EBV for the phenotype used in this association study.

To minimize “over-correction” problem, we defined LD blocks from $R^2 > 0.7$ that resulted $n = 672$, in total number of blocks+inter-block SNPs. We used $1/n$ as a cutoff for significance that has less violation of the assumption of independence (Duggal et al., 2008).

RESULTS

Sequence variations in horses

Overall, 228,696,430 sequencing paired-end reads of length 75 bp, totaling 22 Gbp, were generated for each sample. By applying a genotype-calling pipeline (see Methods for details) to them, the next-generation sequencing yielded an average of 2,586,724 variable sites

per sample with a 17.1-fold read depth per base. A total of 2,709,922 SNPs segregating in Thoroughbred and Jeju breed were identified after imputation. These SNPs overlapped with 67.74% of SNPs on the EquineSNP50 BeadChips. Nucleotide diversity (π) in Thoroughbred and Jeju breed was similar across the genome. However, the mean value of nucleotide diversity (π) of the Jeju breed was slightly higher (10%) than that of the Thoroughbred.

Signatures of selection associated with racehorse domestication

High F_{ST} is indicative of diversifying selection. The low ratio of nucleotide diversity in Jeju breed to that in Thoroughbreds (π_{Jeju}/π_{Tb}) results from reduction of diversity by strong artificial selection in the Thoroughbreds. For their complex demographic history (Orlando et al., 2013), we used the outlier approach to identify candidate loci showing low π score as well as high population differentiation (F_{ST}) (Figure 1). We obtained twelve outliers as putative signatures of directional selection in the Thoroughbreds (Table 1).

Two loci on ECA22 showing strongest selection signals were involved in actin-based cell motility and synapse contents. One of them is Ral GAPase alpha 2 (RALGAPA2; Figure 2A), showing a clear pattern of reduced nucleotide sequence diversity up- and down-stream of focal sites. The RALGAPA2 produces multifunctional proteins that regulate a variety of cellular processes, including cell signal trafficking (Thomas et al., 2003). Another region showing a peak of F_{ST} value was located in a 3' end of synapse

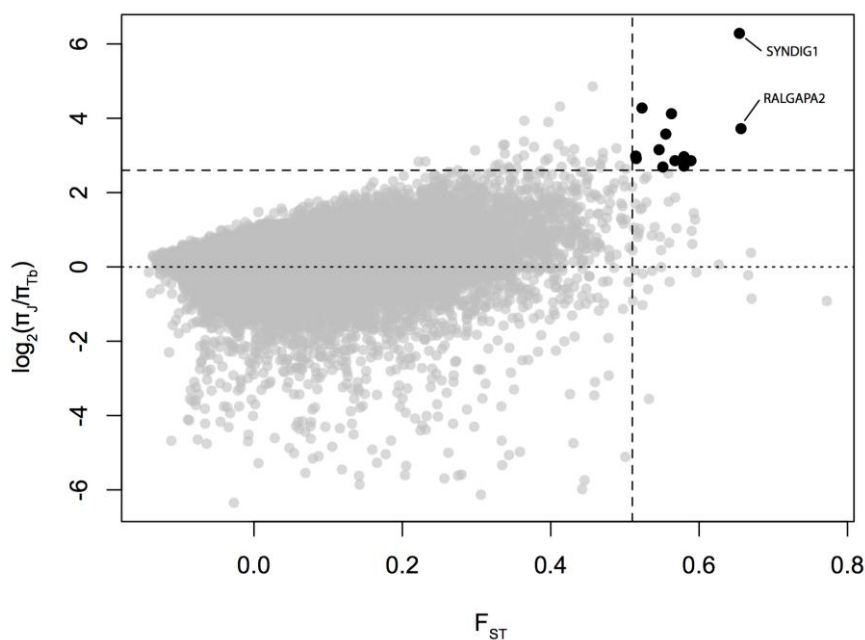


Figure 1. Putative signatures of selective sweeps in the Thoroughbred population. Black dots depict potential candidates of top 0.1% (vertical dashed line) of F_{ST} as well as top 0.1% (horizontal dashed line) of the ratio of π in Thoroughbreds to that in Jeju breed. Dotted line depicts π in Thoroughbred being equal to that in Jeju breed.

Table 1. Potential domesticated genes associated with strong signatures in racing horse population

Chr.	Bin start	Bin end	Gene	Description	F_{ST}	π_{jeju}	π_{tb}	$\log_2(\pi_{jeju}/\pi_{tb})^*$
22	4,550,000	4,600,000	<i>RALGAPA2</i>	Ral GTPase-activating protein subunit alpha-2	0.66	77.17	5.85	3.72
22	1,100,000	1,150,000	<i>SYNDIG1</i>	Synapse differentiation-inducing gene protein 1	0.65	27.83	0.36	6.28
18	67,850,000	67,900,000	<i>AK130051</i>	Unknown	0.59	48.67	6.71	2.86
8	2,300,000	2,350,000	<i>P2RX6</i>	P2X purinoceptor 6 isoform 1	0.58	203.17	25.99	2.97
9	18,300,000	18,350,000	<i>TET2/COP55</i> <i>/PPP1R42</i>	Tet oncogene family member 2	0.58	87.83	13.40	2.71
4	16,700,000	16,750,000	<i>ADCY1</i>	Adenylate cyclase type 1	0.57	19.33	2.67	2.86
4	16,150,000	16,200,000	<i>ADCY1</i>	Adenylate cyclase type 1	0.56	28.33	1.63	4.12
1	15,500,000	15,550,000	<i>PLIP</i>	Pancreatic lipase	0.56	27.83	2.34	3.57
6	81,700,000	81,750,000	<i>IRAK3</i>	Interleukin-1 receptor-associated kinase 3	0.55	31.50	4.89	2.69
1	22,700,000	22,750,000	<i>JL635247</i>	Unknown	0.52	116.83	6.03	4.28
22	4,600,000	4,650,000	<i>RALGAPA2</i>	Ral GTPase-activating protein subunit alpha-2	0.52	55.33	7.33	2.92
1	15,550,000	15,600,000	<i>PLIP</i>	Pancreatic lipase	0.51	17.33	2.19	2.98

differentiation induced gene 1 (*SynDIG1*, Figure 2B). The *SynDIG1* encodes a conserved type II transmembrane protein that may play an essential role in the central regulation of excitatory synaptic strength (Kalashnikova et al., 2010). This region overlapped with one of Thoroughbred-specific putative signatures of selection that have been identified previously in large-scale population genetics study using SNP arrays (Petersen et al., 2013).

Intense artificial selection contributing to racing performance on chromosome 22

We further investigated the role of those candidates on ECA22 by conducting an association study to evaluate association of SNPs within them to racing performance in Thoroughbred. We estimated effects of SNPs on EBV of the

finishing time for independent groups of 240 Thoroughbreds. With a significance threshold after multiple tests being 0.0015, the identified significant SNPs were significantly associated with racing performance. Half of them were positioned in genes (Table 2), including *RALGAPA2*. Furthermore, we found a long LD block at *RALGAPA2* (Figure 3), reflecting strong selection has been acting on the region to have high level of haplotype homozygosity. However, the other half distributed in intergenic regions, suggesting that, even though significant SNPs identified by association studies can be used as genetic markers for breeding value prediction and breeding purpose, it may be difficult to link these markers to the genetic basis underlying the quantitative traits. Thus, there results provide strong evidence that *RALGAPA2* plays

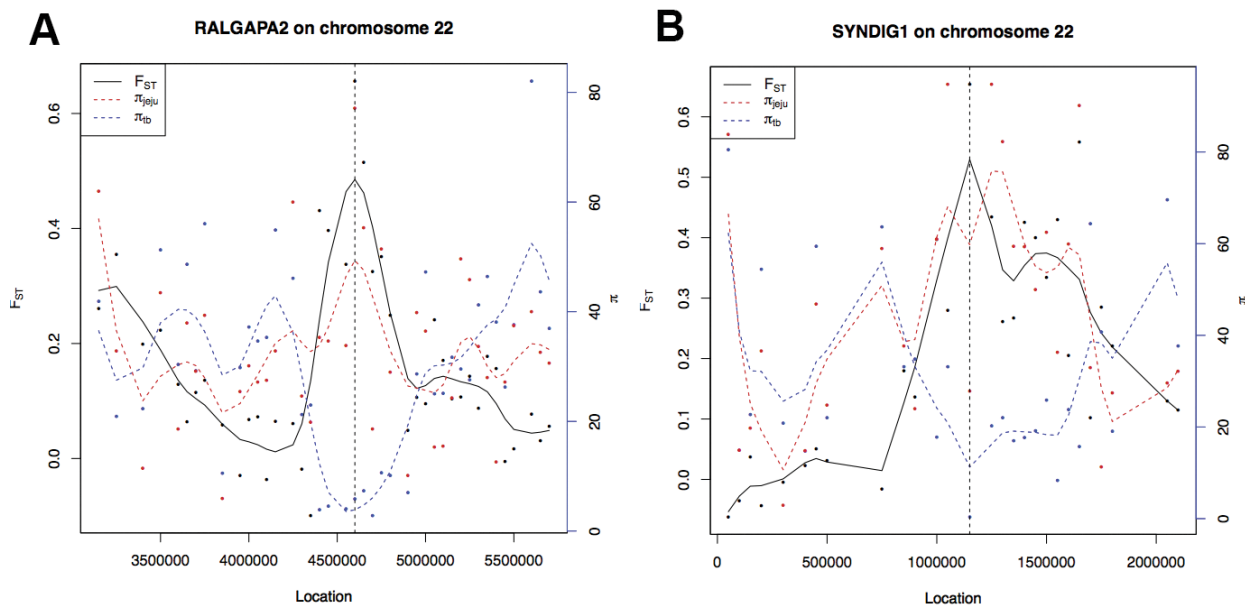


Figure 2. Genome regions with strongest signatures of selective sweep. The vertical line depicts the focal site proximal to *RALGAPA2* (A) and *SYNDIG1* (B), where the F_{ST} value (black line) is high and the π in the Thoroughbred (blue dashed line) is relatively lower than that in Jeju (red dashed line) in regions. *RALGAPA2*, Ral GTPase alpha 2; *SYNDIG1*, synapse differentiation induced gene 1.

Table 2. Significant SNPs associated with racing performance

SNP ID	Chr	Position	Major allele	Minor allele	p-value	MAF	SNP effect	Gene*	Description
BIEC2_577013	22	4632335	G	A	0.001084	0.16	0.04011	<i>RALGAPA2</i>	Ral GTPase activating protein, alpha subunit 2
BIEC2_579415	22	8075458	G	A	0.0005322	0.12	0.04527	<i>KIF16B</i>	Kinesin family member 16B
BIEC2_580028	22	8877380	G	A	0.0002055	0.50	0.0524	<i>MACROD2</i>	MACRO domain containing 2
BIEC2_580291	22	9113291	A	G	0.0002192	0.26	0.05192	<i>MACROD2</i>	MACRO domain containing 2
BIEC2_580790	22	11114068	G	A	0.000647	0.34	0.04417	<i>SPTLC3</i>	Serine palmitoyltransferase, long chain base subunit 3
BIEC2_600400	22	44418264	A	G	0.0004451	0.27	0.0468	<i>AKI28005</i>	
BIEC2_600504	22	44665253	A	C	0.0008664	0.17	0.04162	<i>PPP4R1L</i>	Protein phosphatase 4, regulatory subunit 1-like
BIEC2_600563	22	45058230	A	C	3.13E-05	0.10	0.06649	<i>STX16</i>	Syntaxin 16
BIEC2_600572	22	45095728	G	A	3.13E-05	0.10	0.06649	<i>STX16</i>	Syntaxin 16
BIEC2_600582	22	45181810	G	A	3.13E-05	0.10	0.06649	<i>GNAL</i>	Guanine nucleotide binding protein (G protein), alpha activating activity polypeptide

SNP, single nucleotide polymorphism; MAF, minor allele frequency.

* Genes within 5 kb up- and down-stream of SNPs were annotated.

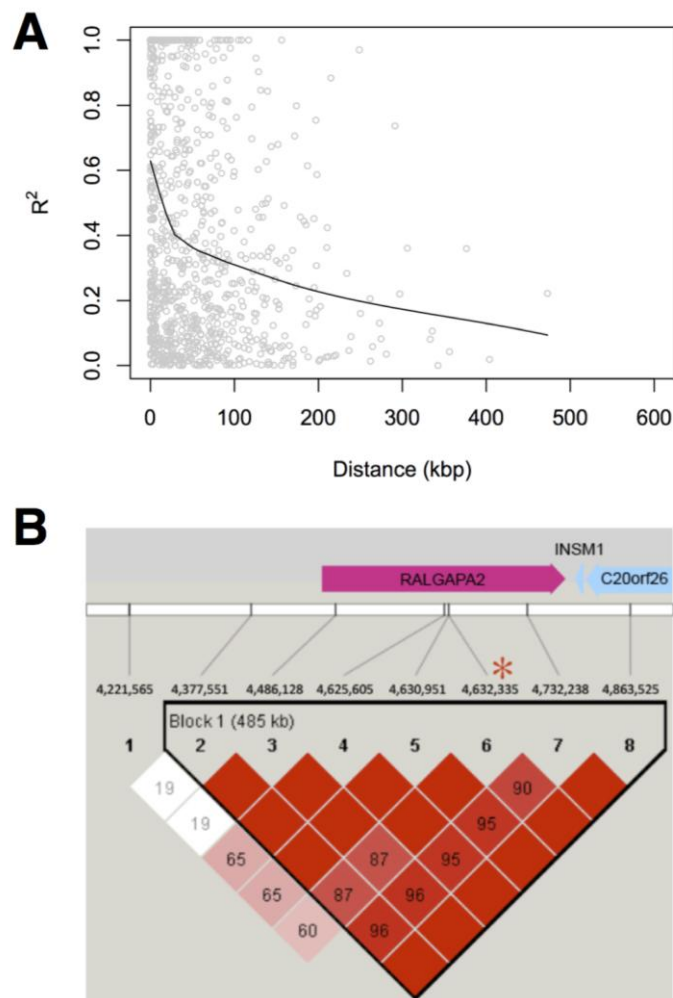


Figure 3. Linkage disequilibrium around potential candidate genes. The degree of linkage-disequilibrium (LD) decrease as a function of distance (A). Red asterisk depicts the position of genetic maker significantly affecting racing performance on horse chromosome 22 (B).

critical role in improving racing performance in horses.

DISCUSSION

The population differentiation index has the power to detect selection on standing variation as well as on newly selected sites (Innan and Kim, 2008). The relative level of genetic variability around the focal locus is a measure of selection pressure. We identified putative signals of strong artificial selection in Thoroughbred genomes by combining both tests.

Previous gene-centered studies have focused on genetic mechanism related to muscle contents and energy metabolism for racing performance in Thoroughbreds. For example, *MSTN* is known to be associated with double-muscling phenotype (Kambadur et al., 1997; McPherron and Lee, 1997; Riquet et al., 1997). This region has a low F_{ST} index (0.08) with low π values in both the Jeju breed (20.17) and Thoroughbred (13.40), suggesting that *MSTN* could have an essential role in the phenotypes of muscle, and selection has been acting on this before the domestication of horse.

Our results provide strong evidence of the genetic basis for selective breeding for racing performance in horses. We found a large gene cluster around *RALGAPA2* in the Thoroughbred genomes. The *RalGAPs*, $\alpha 1$, and $\alpha 2$, are critical for efficient termination of *Ral* activation induced by extracellular stimuli that recombinant *RalGAP1* accelerates the guanosine triphosphate hydrolysis rate of *RalA* by 280,000-fold (Suzuki et al., 2000). Genes within the cluster have a similar function. For example, like *RALGAPA2*, pallid (*PLDN*), and Ras and Rab interactor 2 (*RIN2*) genes encode protein for cellular vehicles or membrane trafficking. 5'-3' Exoribonuclease 2 (*XRN2*), *RIN2*, crooked neck pre-mRNA splicing factor 1 (*CRNKL1*), and *PLDN* genes are associated with cellular component organization. Insulinoma-associated 1 (*INSM1*) also plays an essential role in neuronal phenotype in vertebrate hindbrain (Jacob et al., 2009). Therefore, this cluster appears to have an important molecular role in controlling the function of signal trafficking, reflecting artificial selection processes driving them into a tight gene cluster for efficacy of the associated function.

In this study, based on association analysis within the context of population genomics study, we suggest that the novel candidate genes for racing performance can potentially illuminate previously unsuspected biological mechanisms. Therefore, the principal challenge of multifactorial traits may lay not only in detecting quantitative trait loci, but also in identifying targets of selection. Further functional studies are needed to investigate the downstream targets of domesticated genes affecting racing performance.

ACKNOWLEDGMENTS

This work was supported by KRA for the project of Thoroughbred horse (No. 0569-20110008). We thank Profs. Rasmus Nielsen (UC, Berkeley) and Yuseob Kim (Ewha Womans University) for critical reading and comments on the manuscript.

REFERENCES

- Anisimova, M. and C. Kosiol. 2009. Investigating protein-coding sequence evolution with probabilistic codon substitution models. *Mol. Biol. Evol.* 26:255-271.
- Binns, M., D. A. Boehler, and D. H. Lambert. 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.
- Browning, S. R. and B. L. Browning. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81:1084-1097.
- 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.
- Duggal, P., E. M. Gillanders, T. N. Holmes, and J. E. Bailey-Wilson. 2008. Establishing an adjusted p-value threshold to control the family-wide type I error in genome wide association studies. *BMC Genomics* 9:516.
- Excoffier, L., T. Hofer, and M. Foll. 2009. Detecting loci under selection in a hierarchically structured population. *Heredity* 103:285-298.
- Excoffier, L. and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10:564-567.
- Gilmour, A. R., B. J. Gogel, B. R. Cullis, R. Thompson, D. Butler, M. Cherry, D. Collins, G. Dutkowski, S. A. Harding, and K. Haskard. 2009. ASReml User Guide Release 3.0. VSN International Ltd., UK. <http://www.vsnl.co.uk> 275.
- Grobet, L., D. Pirottin, F. Farnir, D. Poncelet, L. J. Royo, B. Brouwers, E. Christians, D. Desmecht, F. Coignoul, R. Kahn, and M. Georges. 2003. Modulating skeletal muscle mass by postnatal, muscle-specific inactivation of the myostatin gene. *Genesis* 35:227-238.
- Gu, J., N. Orr, S. D. Park, L. M. Katz, G. Sulimova, D. E. MacHugh, and E. W. Hill. 2009. A genome scan for positive selection in thoroughbred horses. *PLoS One* 4(6):e5767-e5767.
- Hill, E. W., D. G. Bradley, M. Al Barody, O. Ertugrul, R. Splan, I. Zakharov, and E. P. Cunningham. 2002. History and integrity of thoroughbred dam lines revealed in equine mtDNA variation. *Anim. Genet.* 33:287-294.
- Hill, E. W., J. Gu, S. S. Eivers, R. G. Fonseca, B. A. McGivney, P. Govindarajan, N. Orr, L. M. Katz, and D. MacHugh. 2010. A sequence polymorphism in *MSTN* predicts sprinting ability and racing stamina in Thoroughbred horses. *PLoS One* 5(1):e8645.
- Innan, H. and Y. Kim. 2008. Detecting local adaptation using the joint sampling of polymorphism data in the parental and

- derived populations. *Genetics* 179:1713-1720.
- Jacob, J., R. Storm, D. S. Castro, C. Milton, P. Pla, F. Guillemot, C. Birchmeier, and J. Briscoe. 2009. *Insm1* (IA-1) is an essential component of the regulatory network that specifies monoaminergic neuronal phenotypes in the vertebrate hindbrain. *Development* 136:2477-2485.
- Jansen, T., P. Forster, M. A. Levine, H. Oelke, M. Hurles, C. Renfrew, J. Weber, and K. Olek. 2002. Mitochondrial DNA and the origins of the domestic horse. *Proc. Natl. Acad. Sci. USA* 99:10905-10910.
- Kalashnikova, E., R. A. Lorca, I. Kaur, G. A. Barisone, B. Li, T. Ishimaru, J. S. Trimmer, D. P. Mohapatra, and E. D'áz. 2010. SynDIG1: An activity-regulated, AMPA-receptor-interacting transmembrane protein that regulates excitatory synapse development. *Neuron* 65:80-93.
- Kambadur, R., M. Sharma, T. P. L. Smith, and J. J. Bass. 1997. Mutations in myostatin (*GDF8*) in double-musled Belgian Blue and Piedmontese cattle. *Genome Res.* 7:910-915.
- Kim, K. I., Y. H. Yang, S. S. Lee, C. Park, R. Ma, J. L. Bouzat, and H. A. Lewin. 1999. Phylogenetic relationships of Cheju horses to other horse breeds as determined by mtDNA D-loop sequence polymorphism. *Anim. Genet.* 30:102-108.
- Lee, S. J. 2007. Sprinting without myostatin: A genetic determinant of athletic prowess. *Trends Genet.* 23:475-477.
- Li, H. and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754-1760.
- Lindgren, G., N. Backström, J. Swinburne, L. Hellborg, A. Einarsson, K. Sandberg, G. Cothran, C. Vilà, M. Binns, and H. Ellegren. 2004. Limited number of patrilineages in horse domestication. *Nat. Genet.* 36:335-336.
- McGivney, B. A., J. A. Browne, R. G. Fonseca, L. M. Katz, D. E. MacHugh, R. Whiston, and E. W. Hill. 2012. *MSTN* genotypes in Thoroughbred horses influence skeletal muscle gene expression and racetrack performance. *Anim. Genet.* 43:810-812.
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernysky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297-1303.
- McPherron, A. C. and S. J. Lee. 1997. Double muscling in cattle due to mutations in the myostatin gene. *Proc. Natl. Acad. Sci. USA* 94:12457-12461.
- Nam, D. Y. 1969. Horse production in Cheju during Lee dynasty. *Studies on Korean History* 4:131-131.
- Nei, M. 1987. *Molecular evolutionary genetics*. Columbia Univ. Press, New York, NY, USA.
- Nei, M. and W. H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci.* 76:5269-5273.
- Orlando, L., A. Ginolhac, G. Zhang, D. Froese, A. Albrechtsen, M. Stiller, M. Schubert, E. Cappellini, B. Petersen, and I. Moltke et al. 2013. Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499:74-78.
- Park, K. D., H. Kim, J. Y. Hwang, C. K. Lee, K. T. Do, H. S. Kim, Y. M. Yang, Y. J. Kwon, J. Kim, H. J. Kim, K. D. Song, J. D. Oh, H. Kim, B. W. Cho, S. Cho, and H. K. Lee. 2014. Copy number deletion has little impact on gene expression levels in racehorses. *Asian Australas. J. Anim. Sci.* 27:1345-1354.
- Petersen, J. L., J. R. Mickelson, E. G. Cothran, L. S. Andersson, J. Axelsson, E. Bailey, D. Bannasch, M. M. Binns, A. S. Borges, and P. Brama et al. 2013. Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLoS One* 8(1):e54997.
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559-575.
- Qanbari, S., H. Pausch, S. Jansen, M. Somel, T. M. Strom, R. Fries, R. Nielsen, and H. Simianer. 2014. Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genet.* 10(2):e1004148.
- Riquetl, I., A. Schoeberlein, S. Dunnerz, F. Ménessier, and I. Massabanda. 1997. A deletion in the bovine myostatin gene causes the double-musled phenotype in cattle. *Nat. Genet.* 17:71-71.
- Suzuki, J., Y. Yamazaki, L. Guang, Y. Kaziro, and H. Koide. 2000. Involvement of Ras and Ral in chemotactic migration of skeletal myoblasts. *Mol. Cell. Biol.* 20:4658-4665.
- Thomas, P. D., M. J. Campbell, A. Kejarawal, 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.
- Tozaki, T., T. Miyake, H. Kakoi, H. Gawahara, S. Sugita, T. Hasegawa, N. Ishida, K. Hirota, and Y. Nakano. 2010. A genome-wide association study for racing performances in Thoroughbreds clarifies a candidate region near the *MSTN* gene. *Anim. Genet.* 41:28-35.
- Vaysse, A., A. Ratnakumar, T. Derrien, E. Axelsson, G. Rosengren Pielberg, S. Sigurdsson, T. Fall, E. H. Seppala, M. S. Hansen, and C. T. Lawley et al. 2011. Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genet.* 7(10):e1002316.
- Williamson, S. A. and R. G. Beilharz. 1998. The inheritance of speed, stamina and other racing performance characters in the Australian Thoroughbred. *J. Anim. Breed. Genet.* 115:1-16.
- Willing, E. M., C. Dreyer, and C. van Oosterhout. 2012. Estimates of genetic differentiation measured by F_{ST} do not necessarily require large sample sizes when using many SNP markers. *PLoS One* 7(8):e42649.