Evolutionary Analyses of Hanwoo (Korean Cattle)-Specific Single-Nucleotide Polymorphisms and Genes Using Whole-Genome Resequencing Data of a Hanwoo Population

Daehwan Lee[1,4], Minah Cho[1,4], Woon-young Hong[1], Dajeong Lim[2], Hyung-Chul Kim[2], Yong-Min Cho[2], Jin-Young Jeong[2], Bong-Hwan Choi[2], Younhee Ko[3], and Jaebum Kim[1,*]

Advances in next generation sequencing (NGS) technologies have enabled population-level studies for many animals to unravel the relationships between genotypic differences and traits of specific populations. The objective of this study was to perform evolutionary analysis of single nucleotide polymorphisms (SNP) in genes of Korean native cattle Hanwoo in comparison to SNP data from four other cattle breeds (Jersey, Simmental, Angus, and Holstein) and four related species (pig, horse, human, and mouse) obtained from public databases through NGS-based resequencing. We analyzed population structures and differentiation levels for the five cattle breeds and estimated species-specific SNPs with their origins and phylogenetic relationships among species. In addition, we identified Hanwoo-specific genes and proteins, and determined distinct changes in protein-protein interactions among five species (cattle, pig, horse, human, mouse) in the STRING network database by additionally considering indirect protein interactions. We found that the Hanwoo population was clearly different from the other four cattle populations. There were Hanwoo-specific genes related to its meat trait. Protein interaction rewiring analysis also confirmed that there were Hanwoo-specific protein-protein interactions that might have contributed to its unique meat quality.

[1]Department of Stem Cell and Regenerative Biology, Konkuk University, Seoul 05029, Korea [2]National Institute of Animal Science, Wanju 55365, Korea, [3]Department of Clinical Genetics, Department of Pediatrics, Yonsei University College of Medicine, Seoul 03722, Korea, [4]These authors contributed equally to this work.
*Correspondence: jbkim@konkuk.ac.kr

INTRODUCTION

Next-generation sequencing (NGS) technologies (Metzker, 2010) have enabled the accumulation of population-scale DNA sequence data. NGS has provided opportunities as well as challenges to many population-based genome projects such as the 1000 genomes project (Genomes Project et al., 2010), the 1000 bull genomes project (Hayes, 2012), the international HapMap project (International HapMap, 2003), and the Drosophila population genomics project (Begun et al., 2007). In addition, various species- and breed-specific studies have been conducted to identify unique genomic features. For example, novel nonsynonymous mutations specific to dogs living at high altitude areas have been identified though sequencing of 60 individual dogs (Gou et al., 2014). Similar study has been conducted for a pig population by sequencing 69 individuals, yielding a set of loci related to genetic adaptation to a high- and low-latitude environments (Ai et al., 2015). In addition, sequencing data of 234 bulls from the 1000 bull genome projects have been used to identify variants and traits associated with milk production level and curly coat (Daetwyler et al., 2014). Gir cattle population has also been analyzed through sequencing 11 individuals, resulting in the finding of a number of loci associated with osmotic stress and heat shock that can influence their adaptation to tropical climates (Liao et al., 2013). Recently, several studies have been performed on Hanwoo cattle breed, which is indigenous and representative cattle breed in Korea. The Hanwoo breed has evolved from the 1960s to the present in Korea with genetic improvement associated with meat traits (Lee et al., 2014). For examples, a comparative study on three cattle breeds (Hanwoo, Black Angus, and Holstein) has been performed to reveal genetic and genomic characteristics specific to the Hanwoo breed (Lee et al., 2013). Using whole-genome sequencing, a similar comparative analysis has been performed to identify variations in economically important traits in three Korean cattle breeds (Hanwoo, Jeju Heugu, and Korean Holstein) (Choi et al., 2014). Moreover, potential selective-sweep regions have been discovered through sequencing 10 Hanwoo and 10 Yanbian cattle individuals (Choi et al., 2015). However, most of these studies have usually focused on the identification of breed-specific variants and traits. Less attention

has been paid to evolutionary and network-level perspective features to explain their uniqueness. Therefore, the objective of this study was to perform evolutionary analysis for Hanwoo cattle breed in the perspective of breed-specific single-nucleotide polymorphisms (SNPs), genes, and proteins through resequencing of Hanwoo cattles and build a protein-protein interaction database. Specifically, we analyzed the population structure and differentiation of five cattle breeds (Hanwoo, Jersey, Simmental, Angus, and Holstein). We identified cattle breed-specific SNPs and their evolutionary origins. In addition, we discovered Hanwoo-specific genes/proteins. Moreover, we investigated how these interactions among Hanwoo-specific proteins might have been rewired during evolution.

MATERIALS AND METHODS

Ethics statement
The DNA extraction protocol was approved by the Committee on Ethics of Animal Experiments, National Institute of Animal Science, Republic of Korea (Permit Number: NIAS2015-774). Genomic DNAs were extracted from AI bull semen straws or blood samples obtained from the Hanwoo Improvement Center of the National Agricultural Cooperative Federation in Republic of Korea with permission from the owners.

Resequencing of the Hanwoo genomes
We generated whole-genome resequencing data from Hanwoo (N = 126). Hanwoo samples were obtained from the Hanwoo Improvement Center (National Agricultural Cooperative Federation, Republic of Korea). Indexed shotgun paired-end (PE) libraries with average inserts of 500 bp were generated using TruSeq Nano DNA Library Prep Kit (Illumina, USA) following standard Illumina sample-preparation protocol. Briefly, 200 ng of gDNAs were fragmented with Covaris M220 (USA) to obtain median fragment size of ~500 bp. These fragmented DNAs were end repaired followed by A-tailing and ligation to indexed adapter (~125 bp adapter). Gel-based size selection was performed for adapter-ligated DNAs to generate DNAs in the range of 550 to 650 bp. PCR amplification was performed in eight cycles. Size-selected libraries were analyzed with Agilent 2100 Bioanalyzer (Agilent Technologies) to determine the size distribution and determine whether there was adapter contamination. The resulting libraries without adaptor contamination were sequenced on Illumina HiSeq 2500 (2 × 125 bp paired-end sequences) and NextSeq500 (2 × 150 bp paired-end sequences) sequencing platforms.

Sequence mapping and SNP calling
Resequenced data of the 126 Hanwoo genomes and the sequencing data of other four cattle breeds (Jersey, Simmental, Angus, and Holstein; N = 10 for all breeds) collected from the NCBI SRA database were aligned to bovine reference genome assembly (UMD 3.1) using Bowtie2 v2.2.4 with default parameters (Langmead and Salzberg, 2012). SAMtools v1.1 (Li et al., 2009) was used for converting (SAM/BAM), sorting, and indexing alignments. Picard tools v1.125 (http://picard.sourceforge.net) was used to generate quality matrices for mapping and to exclude duplicate reads. Local re-alignment and re-calibration were performed using Genome Analysis Toolkit (GATK; v3.3) framework (McKenna et al., 2010). Initial novel SNP discovery was performed using multi-sample SNP-calling procedure in the GATK package. To reduce false discovery rate, a filtering step was applied based on the GATK best practice guideline as follows: QD < 2.0, MQ < 40.0, FS > 200.0, Haplo-

typeScore > 13.0, MQRankSum < -12.5 and ReadPosRank-Sum < -8.0 (McKenna et al., 2010). SNPs of the other species were obtained from the dbSNP database (Sherry et al., 2001). The sequencing statistics of 126 Hanwoo genomes and the list of SRA data of the four cattle breeds are available in Table S1. The flow of SNP calling and the number of SNPs in each step are shown in Supplementary Fig. S1.

The evoSNPI pipeline
The evoSNPI pipeline has been developed to predict evolutionary origins of SNPs and rewiring information of protein interactions among related species (Cho et al., 2015). Using evoSNPI, we found target species-specific genes/proteins as well as changes in protein interactions among different species from the SNP data. Input for evoSNPI included the following: (i) VCF files containing SNP information for each species obtained by independent SNP calling pipeline, (ii) pairwise whole-genome alignments between a chosen reference and all other species, and (iii) a phylogenetic tree in newick format. First, evoSNPI was used to find SNPs in orthologous positions given pairwise whole-genome alignments using liftover tool from UCSC Genome Browser (Karolchik et al., 2003). Interactivenn (Heberle et al., 2015) was then used to visualize orthologous information. Second, the evolutionary origin of SNPs was inferred based on position information of SNPs, whether those SNPs exist in orthologous regions, and the maximum parsimony algorithm (Takahashi and Nei, 2000). Once the evolutionary origins of SNPs were predicted, the number of SNPs on each branch of a phylogenetic tree was recorded. Third, target species-specific nonsynonymous SNPs and associated genes with those SNPs were identified. Finally, interactions among proteins of the genes in different species were identified from the STRING network database (Szklarczyk et al., 2014), one of the largest database of protein-protein interactions of many species. In this step, Random Walk with Restart (RWR) algorithm (Kim et al., 2008) was applied to the STRING network database to incorporate indirectly linked proteins with the original target species-specific proteins. Specifically, each protein in the STRING network was ranked with a score representing the degree of closeness with the original protein sets. The top 5% of those proteins were used as additional proteins. Edge scores in the STRING network database were normalized (between 0 to 1). These scores were used to quantify the similarity and difference in protein interactions among different species. Orthologous protein information was obtained from OrthoDB which covers 3,027 complete genomes including 61 vertebrate species (Kriventseva et al., 2015; Waterhouse et al., 2013).

SNP annotation and functional analysis
ANNOVAR v 2015JUN17 (Wang et al., 2010) and SnpEff v4.1 (Cingolani et al., 2012) with Ensembl gene annotation database (UMD3.1) were used to annotate SNPs of the five cattle breeds. Hanwoo-specific nonsynonymous SNPs and genes in the Hanwoo breed were identified by comparing Hanwoo to other breeds. Hanwoo-specific genes were then analyzed to find overrepresented biological functions using panther website (Mi et al., 2005). Enriched biological functions associated with Hanwoo-specific genes (Bonferroni-corrected p-value < 0.05) were reported.

Population structure analysis
The VCF files from randomly selected ten Hanwoo breeds and the other four cattle breeds were generated from the SNP calling step, merged by VCFtools v 0.1.13 (Danecek et al., 2011),

and converted to PLINK format file (.ped and .map) using PLINK v1.90b (Purcell et al., 2007). Additional filtering was carried out with the PLINK tool using the following parameters: -geno 0.01 --maf 0.05 --hwe 0.000001. Principal component analysis (PCA) was applied with GCTA v1.24.4 (Yang et al., 2011). It was performed with the following two steps: (i) calculation of genetic relationship matrix (GRM) with parameters of "--make-grm", and (ii) estimation of the first four principal components with parameters of "--pca 4". The R package was used to generate the PCA plot. Population structure was inferred with ADMIXTURE v1.3.0 (Alexander et al., 2009) and visualized with CLUMPAK (Kopelman et al., 2015).

Population differentiation analysis

To identify regions of population differentiation among Hanwoo, Jersey, Simmental, Angus, and Holstein, the mean Z-transformed Fst values [Z(Fst)] were calculated for 100 kbp non-overlapping genomic windows in all chromosomes from the VCF files used in the population structure analysis with VCFtools v 0.1.13 (Danecek et al., 2011). Gene-level analysis was performed in genomic windows with extremely high Z-transformed Fst value (> 5) by identifying enriched Gene Ontology (GO) terms in those regions using the getBM function in the biomaRt R package (Durinck et al., 2005). In this analysis, copy number variable regions were collected from literature (Bickhart et al., 2012; Choi et al., 2013; 2016), and genes within those regions were not used. The manhattan plot of Z(Fst) values were generated with the qqman R package (Li et al., 2015).

RESULTS

SNP annotation

A total of 16,361,482, 7,313,386, 8,180,573, 7,085,527 and 8,125,851 SNPs were identified from Hanwoo, Jersey, Simmental, Angus, and Holstein, respectively (Materials and Methods; Table 1). Among them, 14,551,596 (88.94%), 7,283,202 (99.59%), 8,159,778 (99.75%), 7,064,818 (99.71%) and 8,097,083 (99.65%) SNPs of Hanwoo, Jersey, Simmental, Angus, and Holstein, respectively, were reported in the dbSNP database (version 146). The transition-to-transversion ratio (Ti/Tv) was also calculated to evaluate SNP quality. The Ti/Tv ratios for Hanwoo, Jersey, Simmental, Angus, and Holstein were 2.29, 2.25, 2.24, 2.22 and 2.24, respectively. To identify SNPs explaining phenotypic differences in each cattle breed, we annotated all SNPs with 19 functional categories, such as synonymous, nonsynonymous, intron, and untranslated regions (Supplementary Table S2). The majority of SNPs were founded in the intergenic (72% in Hanwoo and 73% in other four cattle breeds) and intron (27% in Hanwoo and 26% in other four breeds) regions. Only a small fraction of SNPs (1.2, 1.1, 1.0, 1,1 and 1,1% in Hanwoo, Jersey, Simmental, Angus, and Holstein, respectively) were detected in genic regions including exonic, splice site, and untranslated regions (Supplementary Table S2).

Population analysis of the five cattle breeds

After filtering out SNPs with various population statistics such as minor allele frequency, genotype rate, and Hardy-Weinberg equilibrium (Materials and Methods), a total of 1,826,768 SNPs from ten individuals of each cattle breeds were used to analyze population structure. In this analysis, randomly selected ten Hanwoo individuals were used to reduce a sample size bias. We first used principal component analysis (PCA) to identify the relationships among the five cattle populations using SNP data. As shown in Fig. 1A, the Hanwoo population was distinctly
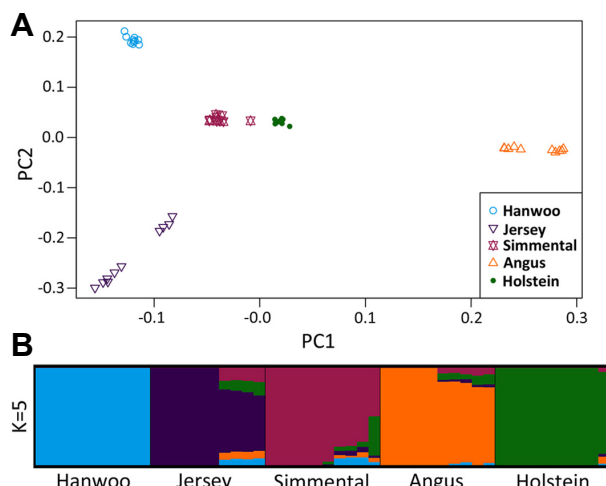
Fig. 1. Population structure analysis for five cattle populations (Hanwoo, Jersey, Simmental, Angus, and Holstein). (A) The principal component analysis plot of five cattle populations with the first two components. (B) Population structures obtained from the number of clusters K at 5. Each individual is represented with a vertical line. It is partitioned into K colored segments. The length of each segment represents a relative membership to different cluster. Black vertical lines separate the five major cattle populations.

separated from the other four cattle populations. Interestingly, individuals of the Jersey and Angus population were relatively more dispersed.

Next, we further analyzed the population structure of the five cattle populations using ADMIXTURE to estimate individual ancestry and admixture proportions (Materials and Method). Population structure plots for the number of clusters K from two to seven were drawn (Supplementary Fig. S2). Assuming that they were five ancestral populations (Fig. 1B), the Hanwoo population was clearly differentiated from the other populations. Although K was increased to seven, the Hanwoo population was still clustered as one distinct group and showed clear separation from the other four populations (Supplementary Fig. S2).

To identify the regions associated with population differentiation in the five cattle populations, we calculated the mean of Z-transformed Fst [Z(Fst)] values from SNPs in 100 kbp non-overlapping genomic regions (Materials and Methods). As shown in Fig. 2, 62 significant regions [Z(Fst) > 5] were identified as regions to explain population differentiation, with a total of 86 genes including 2,390 SNPs across all chromosomes. Some highly scored differentiation-regions included the PSAT1 gene associated with metabolic process (GO: 0008152), the BLCAP gene related to protein binding (GO: 0005515), and the FBLIM1 gene with functions in the regulation of protein localization (GO:0032880), mitochondrial inner membrane (GO:0005743), and filamin binding (GO:0031005), respectively. We also compared genes in the 62 significant regions to known cattle trait-associated genes (Kawahara-Miki et al., 2011), and found the SLC43A3 gene with transmembrane transport (GO:005585) function and the LEPR gene with leptin realted functions (GO:0033210, GO0038021 and 0044321) that have clear association with the meat trait of cattle.

Evolutionary analysis of SNP and protein interactions

Recently, we have developed a pipeline called evoSNPI

Table 1. Statistics of SNPs identified from Hanwoo, Jersey, Simmental, Angus, and Holstein cattle breeds

Cattle breeds	No. of SNPs	Found in dbSNP[a]	Ti/Tv ratio[b]
Hanwoo	16,361,482	14,551,596 (88.94%)	2.29 (0.00001)
Jersey	7,313,386	7,283,202 (99.59%)	2.25 (0.00001)
Simmental	8,180,573	8,159,778 (99.75%)	2.24 (0.00018)
Angus	7,085,527	7,064,818 (99.71%)	2.22 (0.00000)
Holstein	8,125,851	8,097,083 (99.65%)	2.24 (0.00013)

[a] The number of SNPs found in dbSNP database (version 146). Fractions are in parentheses.
[b] The ratio of the number of transitions to the number of transversions. Standard deviations are in parentheses.
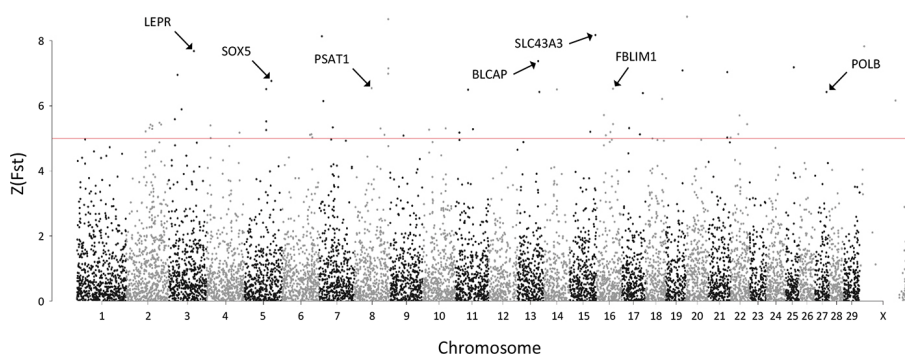

Fig. 2. Manhattan plot of Z-transformed Fst [Z(Fst)] among the five cattle populations (Hanwoo, Simmental, Jersey, Angus, and Holstein). The Fst values were calculated for each 100-kbp window on autosomal and X chromosomes. Red line denotes a threshold of Z(Fst) at 5.

(Cho et al., 2015) to predict the evolutionary origin of SNPs and the rewiring of protein-protein interactions among different species (Materials and Methods). We applied the evoSNPI pipeline to the Hanwoo SNP data as well as SNP data from the other cattle breeds (*Bos taurus*; Jersey, Simmental, Angus, and Holstein), pig (*Sus scrofa*), horse (*Equus caballus*), and outgroup species human (*Homo sapiens*) and mouse (*Mus musculus*). We used previously created SNPs for five cattle breeds (Table 1) and 55,377,259, 4,991,883, 139,444,739 and 67,869,085 SNPs for pig, horse, human and mouse, respectively. Among the total of SNPs from all species, 39,158,606 SNPs were located in orthologous positions in three species (pig, horse, and five cattle breeds). Interestingly, the Hanwoo breed had relatively larger number of unique SNPs (7,423,131 SNPs, 45.37%) compared to the other cattle breeds such as Simmental (805,788 SNPs; 9.85%), Jersey (532,278 SNPs; 7.28%), Angus (496,436; 7.00%) and Holstein (634,322; 7.81%). Only 93 SNPs were shared by all five species including the five cattle breeds. Fig. 3 shows the number of SNPs originated from each branch. A total of 9,182,484 SNPs were found after the speciation of the Hanwoo breed from the other four cattle breeds. A total of 6,924,567 SNPs were generated earlier than the speciation of the Hanwoo breed from the other four cattle breeds, while a total of 18,799,180 SNPs in pig and 1,643,951 SNPs in horse were found after speciation from the other species.

Based on the identified SNPs in orthologous positions, we extracted Hanwoo-specific genes including nonsynonymous SNPs only in the Hanwoo breed. As a results, we found 1,509 Hanwoo-specific genes corresponding to 1,646 proteins identified in the Ensembl bioMart Database (Kinsella et al., 2011). To explain the specificity of Hanwoo, we extended the gene-level analysis to rewiring analysis for protein-protein interactions. The "rewiring" concept is a widely used term in systems biology to indicate the changes of interactions among proteins (or genes), and the systems-level characteristics of Hanwoo compared

with other related species can be obtained from this analysis. From the initial Hanwoo-specific genes/proteins (1,509/1,646), we first ran the Random Walk with Restart algorithm, and found closely associated additional proteins based on the STRING network database (Supplementary Table S3). We also compared the extended protein set (a total of 2,592 proteins) to known cattle trait-associated genes. As a result, 76 of the 2,592 extended proteins were cattle trait-associated ones, including
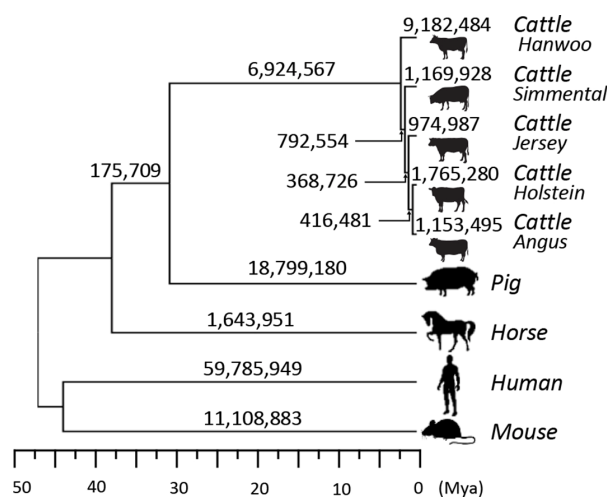
Fig. 3. Evolutionary analysis of SNPs. A phylogenetic tree of five species (with five cattle breeds) with the number SNPs originated from each branch. The divergence time was obtained from the TimeTree website (Hedges et al., 2015). For the branch lengths among the five cattle breeds, arbitrary small lengths were used.

MYOD1, MYH3, and PYGM (Supplementary Table S4). Majority (63 out of 76) of known cattle trait-associated genes had association with meat quality. Eleven genes were associated with milk produc tion, while eleven genes were related to growth (Supplementary Table S4). Next, we compared protein-protein interactions of the 19 Hanwoo-specific proteins with orthologous information in OrthoDB among the five species (Fig. 4). After converting edge-scores to be between 0 and 1, the STRING network database was used to report similarity or difference in protein-protein interactions among species. The degree of network rewiring of 15 protein-protein interaction pairs of extended Hanwoo-specific proteins in the five species, which have edge-score difference 0.2 or higher between cattle and the other two species, is shown in Fig. 4. For example, there was no interaction between EGR3 and FGF2 in cattle breeds, although this interaction was observed in other species (0.36, 0.60, 0.36, and 0.53 in pig, horse, mouse, and human respectively). In contrast, there were many exclusive interactions only in the cattle breeds, including ABL and FGF2, GPX3 and CAT, PTN and FGF2. Results for all other protein pairs are summarized in Table S5. Examples of rewired protein interactions among cattle, pig, and horse are shown in Fig. 5. There was unique interaction between ABL and ANAS and between MYOD1 and IL2 in cattle. These unique interactions were not observed in pig or horse. In contrast, the interactions between PBX2 and CREBBP in pig or horse was not observed in the cattle network. In Fig. 5, red coloured genes are known to be associated with a meat, growth or milk trait in cattle (Kawahara-Miki et al., 2011).

DISCUSSION

In this study, 126 individuals of Korean native cattle Hanwoo were subjected to whole-genome resequencing using high-throughput next-generation sequencing technologies, and compared with the genomes of four other cattle breeds, Jersey, Simmental, Angus, and Holstein in terms of SNPs. The four cattle breeds were selected because they are all used for resources of meat or milk production as Hanwoo is the most important meat resource in Korea. In addition, these four cattle breeds have been widely used in recent studies for Hanwoo (Choi et al., 2014; Daetwyler et al., 2014; Ramey et al., 2013; Stothard et al., 2011) and their phylogenetic relationship was recently investigated (Decker et al., 2009).

We conducted population structure and differentiation analyses using SNPs to explain population genetic similarity and difference among the five cattle breeds. The Hanwoo population was clearly separated from the other four cattle populations, which represents that the Hanwoo breed might have a unique set of SNPs comparing to other cattle breeds, and such unique SNPs can explain the phenotypic differences of Hanwoo such as mean quality. The population structure of the five cattle populations yielded similar results to a previously reported phylogenomics study on cattle breeds (Decker et al., 2009). We discovered several candidate regions covering highly differentiated SNPs among the five cattle populations. From the GO enrichment analysis for the genes in these regions, metal ion binding, protein localization, mitochondrial inner membrane, and filamin binding functions were identified as enriched biological functions. Among them, the metal ion binding function is closely related to skeletal muscle responsible for the meat quality (Jeremiah et al., 2003).

We also performed network-based evolutionary analyses by using the evoSNPI pipeline and found perturbed changes in
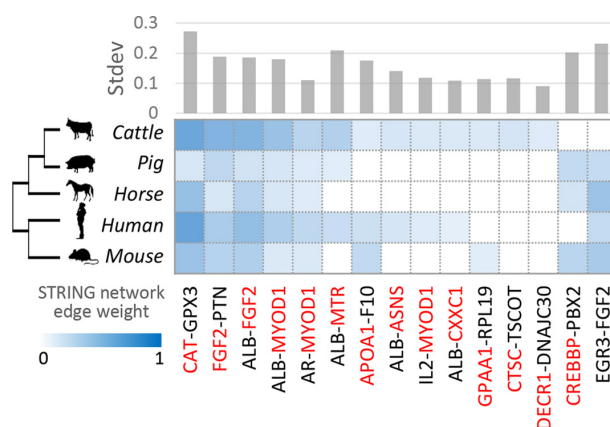
Fig. 4. Evolutionary analysis of protein interactions. Differences of protein interactions among five species (cattle, pig, horse, human, and mouse) in the STRING network were shown only for 15 protein pairs selected by normalized edge scores in the five cattle breeds with 0.2 or more difference in normalized edge scores between cattle breeds and other species. The bottom panel shows normalized STRING network interaction scores of each protein pair in different species (Methods and Materials). The upper panel indicates standard deviation of interaction scores of all five species. The names of cattle trait-associated genes are shown in red color.
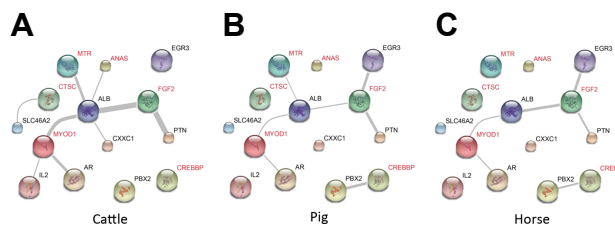
Fig. 5. Examples of protein-protein interactions rewired in three species. Interactions focused on cattle trait-associated proteins such as MYOD1 and FGF2 in three species obtained from the STRING network are shown. The names of cattle trait-associated genes are shown in red color.

protein-protein interactions related with Hanwoo-specific genes comparing to other species (cattle, pig, horse, human, and mouse). There were Hanwoo-specific genes that have nonsynonymous SNPs not present in other four cattle breeds and other four species. Most of them are associated with the meat trait of Hanwoo. Rewired protein-protein interaction analysis among different species also identified Hanwoo breed-specific protein-protein interactions exclusively present only in the Hanwoo breed network such as interactions between EGR3 and FGF2, between ANAS and ALB, between CTSC and SLC46A2, between MYOD1 and IL2, between ALB and CXXC1, and between CREBBP and PBX2. In addition, the interaction between ALB and MTR was present both in the Hanwoo and pig networks, but the interaction in the Hanwoo network was more strong. Among them, three proteins, MTR, FGF2, and EGR3, are related with metabolism-related functions, which are known as an critical factor in making marbled meat in cattle (Lim et al., 2015). Therefore, this analysis confirmed that there were Hanwoo-specific protein interactions that

might have contributed to its unique meat quality. This analysis enables the investigation of additional genes (and proteins) interacting with original breed-specific genes (and proteins) discovered by only using direct genetic differences, and the identification of systems-level features and their evolutionary changes relevant to phenotypic differences.

Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).

REFERENCES

Ai, H., Fang, X., Yang, B., Huang, Z., Chen, H., Mao, L., Zhang, F., Zhang, L., Cui, L., He, W., et al. (2015). Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. Nat. Genet. *47*, 217-225.

Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655-1664.

Begun, D.J., Holloway, A.K., Stevens, K., Hillier, L.W., Poh, Y.P., Hahn, M.W., Nista, P.M., Jones, C.D., Kern, A.D., Dewey, C.N., et al. (2007). Population genomics: whole-genome analysis of polymorphism and divergence in Drosophila simulans. PLoS Biol. *5*, e310.

Bickhart, D.M., Hou, Y., Schroeder, S.G., Alkan, C., Cardone, M.F., Matukumalli, L.K., Song, J., Schnabel, R.D., Ventura, M., Taylor, J.F., et al. (2012). Copy number variation of individual cattle genomes using next-generation sequencing. Genome Res. *22*, 778-790.

Cho, M., Lee, D., Hong, W.Y., Lee, J., and Kim, J. (2015). evoSNPI: a pipeline for the evolutionary analysis of the origin of single nucleotide polymorphisms and the change of protein interactions. In Proceedings of the 6th Computational Systems-Biology and Bioinformatics (CSBio2015) (Bankok, Tailand), pp. 17-21.

Choi, J.W., Lee, K.T., Liao, X., Stothard, P., An, H.S., Ahn, S., Lee, S., Lee, S.Y., Moore, S.S., and Kim, T.H. (2013). Genome-wide copy number variation in Hanwoo, Black Angus, and Holstein cattle. Mamm Genome *24*, 151-163.

Choi, J.W., Liao, X., Stothard, P., Chung, W.H., Jeon, H.J., Miller, S.P., Choi, S.Y., Lee, J.K., Yang, B., Lee, K.T., et al. (2014). Whole-genome analyses of Korean native and Holstein cattle breeds by massively parallel sequencing. PLoS One *9*, e101127.

Choi, J.W., Choi, B.H., Lee, S.H., Lee, S.S., Kim, H.C., Yu, D., Chung, W.H., Lee, K.T., Chai, H.H., Cho, Y.M., et al. (2015). Whole-genome resequencing analysis of Hanwoo and Yanbian cattle to identify genome-wide SNPs and signatures of selection. Mol. Cells *38*, 466-473.

Choi, J.W., Chung, W.H., Lim, K.S., Lim, W.J., Choi, B.H., Lee, S.H., Kim, H.C., Lee, S.S., Cho, E.S., Lee, K.T., et al. (2016). Copy number variations in Hanwoo and Yanbian cattle genomes using the massively parallel sequencing data. Gene *589*, 36-42.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) *6*, 80-92.

Daetwyler, H.D., Capitan, A., Pausch, H., Stothard, P., van Binsbergen, R., Brondum, R.F., Liao, X., Djari, A., Rodriguez, S.C., Grohs, C., et al. (2014). Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat. Genet. *46*, 858-865.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156-2158.

Decker, J.E., Pires, J.C., Conant, G.C., McKay, S.D., Heaton, M.P., Chen, K., Cooper, A., Vilkki, J., Seabury, C.M., Caetano, A.R., et al. (2009). Resolving the evolution of extant and extinct

ruminants with high-throughput phylogenomics. Proc. Natl. Acad. Sci. USA *106*, 18644-18649.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics *21*, 3439-3440.

Genomes Project, C., Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061-1073.

Gou, X., Wang, Z., Li, N., Qiu, F., Xu, Z., Yan, D., Yang, S., Jia, J., Kong, X., Wei, Z., et al. (2014). Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. Genome Res. *24*, 1308-1315.

Hayes, B. (2012). 1000 bull genomes consortium project. In Plant and Animal Genome XX Conference (January 14-18, 2012) (Plant and Animal Genome).

Heberle, H., Meirelles, G.V., da Silva, F.R., Telles, G.P., and Minghim, R. (2015). InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. BMC Bioinformatics *16*, 169.

Hedges, S.B., Marin, J., Suleski, M., Paymer, M., and Kumar, S. (2015). Tree of life reveals clock-like speciation and diversification. Mol. Biol. Evol. *32*, 835-845.

International HapMap, C. (2003). The international HapMap project. Nature *426*, 789-796.

Jeremiah, L.E., Dugan, M.E.R., Aalhus, J.L., and Gibson, L.L. (2003). Assessment of the relationship between chemical components and palatability of major beef muscles and muscle groups. Meat Sci. *65*, 1013-1019.

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T.S., Hinrichs, A., Lu, Y.T., Roskin, K.M., Schwartz, M., Sugnet, C.W., Thomas, D.J., et al. (2003). The UCSC Genome Browser Database. Nucleic Acids Res. *31*, 51-54.

Kawahara-Miki, R., Tsuda, K., Shiwa, Y., Arai-Kichise, Y., Matsumoto, T., Kanesaki, Y., Oda, S., Ebihara, S., Yajima, S., Yoshikawa, H., et al. (2011). Whole-genome resequencing shows numerous genes with nonsynonymous SNPs in the Japanese native cattle Kuchinoshima-Ushi. BMC Genomics *12*, 103.

Kim, T.H., Lee, K.M., and Lee, S.U.(2008). Generative image segmentation using randon walks with estart. In 10th European Conference on Computer vision, P.T. David Forsyth, Andrew Zisserman, ed. (Marseille, France, Springer Berlin Heidelberg), pp. 264-275.

Kinsella, R.J., Kahari, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., et al. (2011). Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database (Oxford) *2011*, bar030.

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., and Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol. Ecol. Resour. *15*, 1179-1191.

Kriventseva, E.V., Tegenfeldt, F., Petty, T.J., Waterhouse, R.M., Simao, F.A., Pozdnyakov, I.A., Ioannidis, P., and Zdobnov, E.M. (2015). OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. Nucleic Acids Res. *43*, D250-256.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357-359.

Lee, K.T., Chung, W.H., Lee, S.Y., Choi, J.W., Kim, J., Lim, D., Lee, S., Jang, G.W., Kim, B., Choy, Y.H., et al. (2013). Whole-genome resequencing of Hanwoo (Korean cattle) and insight into regions of homozygosity. BMC Genomics *14*, 519.

Lee, S.H., Park, B.H., Sharma, A., Dang, C.G., Lee, S.S., Choi, T.J., Choy, Y.H., Kim, H.C., Jeon, K.J., Kim, S.D., et al. (2014). Hanwoo cattle: origin, domestication, breeding strategies and genomic selection. J. Anim. Sci. Technol. *56*, 2.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, Z., Yang, C., Jin, B., Yu, M., Liu, K., Sun, M., and Zhan, M. (2015). Enabling big geoscience data analytics with a cloud-based, MapReduce-enabled and service-oriented workflow framework. PLoS One *10*, e0116781.

Liao, X., Peng, F., Forni, S., McLaren, D., Plastow, G., and Stothard, P. (2013). Whole genome sequencing of Gir cattle for identifying polymorphisms and loci under selection. Genome / National Research Council Canada = Genome / Conseil national de recherches Canada *56*, 592-598.

Lim, D., Chai, H.H., Lee, S.H., Cho, Y.M., Choi, J.W., and Kim, N.K. (2015). Gene expression patterns associated with peroxisome proliferator-activated receptor (PPAR) signaling in the longissimus dorsi of Hanwoo (Korean Cattle). Asian-Australas J. Anim. Sci. *28*, 1075-1083.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297-1303.

Metzker, M.L. (2010). Sequencing technologies - the next generation. Nat. Rev. Genet. *11*, 31-46.

Mi, H., Lazareva-Ulitsky, B., Loo, R., Kejariwal, A., Vandergriff, J., Rabkin, S., Guo, N., Muruganujan, A., Doremieux, O., Campbell, M.J., et al. (2005). The PANTHER database of protein families, subfamilies, functions and pathways. Nucleic Acids Res. *33*, D284-288.

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

Ramey, H.R., Decker, J.E., McKay, S.D., Rolf, M.M., Schnabel, R.D., and Taylor, J.F. (2013). Detection of selective sweeps in cattle using genome-wide SNP data. BMC Genomics *14*, 1-18.

Sherry, S.T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. *29*, 308-311.

Stothard, P., Choi, J.W., Basu, U., Sumner-Thomson, J.M., Meng, Y., Liao, X., and Moore, S.S. (2011). Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. BMC Genomics *12*, 559.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2014). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447-D452.

Takahashi, K., and Nei, M. (2000). Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. Mol. Biol. Evol. *17*, 1251-1258.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164.

Waterhouse, R.M., Tegenfeldt, F., Li, J., Zdobnov, E.M., and Kriventseva, E.V. (2013). OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. Nucleic Acids Res. *41*, D358-365.

Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am J. Hum. Genet. *88*, 76-82.