

# Detecting Positive Selection of Korean Native Goat Populations Using Next-Generation Sequencing

Wonseok Lee<sup>1</sup>, Sojin Ahn<sup>2</sup>, Mengistie Taye<sup>1,3</sup>, Samsun Sung<sup>4</sup>, Hyun-Jeong Lee<sup>5</sup>, Seoae Cho<sup>4</sup>, and Heebal Kim<sup>1,6,\*</sup>

Goats (*Capra hircus*) are one of the oldest species of domesticated animals. Native Korean goats are a particularly interesting group, as they are indigenous to the area and were raised in the Korean peninsula almost 2,000 years ago. Although they have a small body size and produce low volumes of milk and meat, they are quite resistant to lumbar paralysis. Our study aimed to reveal the distinct genetic features and patterns of selection in native Korean goats by comparing the genomes of native Korean goat and crossbred goat populations. We sequenced the whole genome of 15 native Korean goats and 11 crossbred goats using next-generation sequencing (Illumina platform) to compare the genomes of the two populations. We found decreased nucleotide diversity in the native Korean goats compared to the crossbred goats. Genetic structural analysis demonstrated that the native Korean goat and crossbred goat populations shared a common ancestry, but were clearly distinct. Finally, to reveal the native Korean goat's selective sweep region, selective sweep signals were identified in the native Korean goat genome using cross-population extended haplotype homozygosity (XP-EHH) and a cross-population composite likelihood ratio test (XP-CLR). As a result, we were able to identify candidate genes for recent selection, such as the *CCR3* gene, which is related to lumbar paralysis resistance. Combined with future studies and recent goat genome information, this study will contribute to a thorough understanding of the native Korean goat genome.

## INTRODUCTION

Goats (*Capra hircus*) are one of the oldest species of domesticated animals and their domestication started in western Iran about 10,000 years ago. With 599 breeds (Kim et al., 2014) arising in different environmental conditions and selected via both natural selection and artificial selection, goats have developed different traits and are used for their meat, milk, wool, and skin.

Native Korean goats have been raised in the peninsula over the last 2,000 years (Kim et al., 2014). Because native Korean goats produce low volumes of milk and meat, researchers opted to cross them with Saanen goats that are believed to produce greater volumes of both milk and meat. As Saanens are the largest goat breed among the dairy types and are also consistently good milk producers, they have been used most often to form the crossbred goat group. In addition, the Boer goat, which is developed for meat production, has also been used for crossbreeding. These crossbred black goats in Korea not only have a higher milk yield but also a higher growth rate than the native goats. In addition, when fully grown, the crossbred goats are bigger than the native Korean goats (Son, 1999).

Detecting genetic variants related to phenotypic traits is an important issue in livestock genomic research. Due to their large differences in body size, weight, muscle mass, milk production, and coat color (<http://cemendocino.ucanr.edu/files/131469.pdf>, accessed on September 2016), goats are good model animals for genetic studies. However, only a few studies have analyzed the genetic resources of goats. Moreover, studies of native Korean goat populations using whole-genome sequencing have not yet been reported and there have been no studies focusing on selective sweep signatures. Using next-generation sequencing (NGS), it is possible to examine known and unknown single nucleotide polymorphisms (SNPs) in the genome. Selective sweep signatures that have contributed to the domestication process should also be identified. Therefore, our study aimed to reveal the different genetic features of native Korean goats and elucidate the selective sweep signatures that have contributed to phenotypic appearances using whole-genome NGS data at the population level.

## MATERIALS AND METHODS

### Sample preparation and whole-genome sequencing

Whole-blood samples were collected from 11 crossbred goats and 15 native Korean goats in Korea. Blood samples from the

<sup>1</sup>Department of Agricultural Biotechnology, Animal Biotechnology and Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Korea, <sup>2</sup>Department of Natural Science, Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 08826, Korea, <sup>3</sup>Bahir Dar University, college of Agriculture and Environmental Sciences, POBox 79, Bahir Dar, Ethiopia, <sup>4</sup>C&K genomics, Seoul National University Research Park, Seoul 08826, Korea, <sup>5</sup>Division of Animal Genomics and Bioinformatics, National Institute of Animal Science, Rural Development Administration, Suwon 16637, Korea, <sup>6</sup>Institute for Biomedical Sciences, Shinshu University, Nagano, Japan  
\*Correspondence: heebal@snu.ac.kr

Received 8 September, 2016; revised 28 October, 2016; accepted 14 November, 2016; published online 13 December, 2016

**Keywords:** genomic comparison, native Korean goat, NGS, population analysis, XP-CLR, XP-EHH

native Korean goats were collected from the Animal Genetic Resources Station, National Institute of Animal Science, Rural Development Administration in Korea. For the crossbred goats, we obtained the samples from a Korean small black goat farm. Blood (10 ml) was drawn from the carotid artery and treated with heparin to prevent clotting. DNA was isolated from the whole blood solution using a G-DEXTMIIb Genome DNA Extraction Kit (iNtRoN Biotechnology, Korea) according to the manufacturer's protocol. We randomly sheared 3 µg of genomic DNA using the Covaris System to generate inserts of about 300 bp. The fragments of sheared DNA were end-repaired, A-tailed, adaptor-ligated, and amplified using a TruSeq DNA Sample Prep Kit (Illumina, USA). Paired-end sequencing of the goat genomes to about ten-fold coverage was conducted at NICEM (National Instrumentation Center for Environment Management, Seoul, Korea) using the Illumina HiSeq2000 platform with the TruSeq SBS Kit vs-HS (Illumina). All of the short-read data have been deposited at the Short Read Archive (SRA) under accession lot SRA160379.

### Sequencing alignment and variant calling

Approximately 220,000,000-230,000,000 paired-end reads were mapped to the reference goat genome (Dong et al., 2013) from the Goat Genome Database Web site (<http://goat.kiz.ac.cn/>) using Bowtie 2-2.1.0 and default settings (Langmead and Salzberg, 2012). Open-source packages were used for downstream processing and variant calling for SNPs and INDELS, including Picard tools 1.94 (<http://picard.sourceforge.net>), SAMtools 0.1.19 (Li et al., 2009), VCFtools 4.0 (Danecek et al., 2011), and the Genome Analysis Toolkit (GATK) 2.6.4 (McKenna et al., 2010). A Read Group was added and duplicate reads were removed. All of the mate-pair information between each read was confirmed using the module MarkDuplicates, and FixMateInformation of Picard tools. SAMtools was used to index the resulting bam format files and calculate the mapped read length with the flagstat option (Li et al., 2009). The modules RealignerTargetCreator, IndelRealigner, UnifiedGenotyper, SelectVariants, and VariantFiltration in GATK were used for realignment and variant calling. VCFtools was used for handling the VCF format files (Danecek et al., 2011). Substitution calls were made using the GATK UnifiedGenotyper module (McKenna et al., 2010). SNPs and INDELS called with a PhRed-scaled quality score of < 30 were filtered out. Variants were removed based on MQ0 (mapping quality score) > 3, (MQ0/read depth) > 0.1 (10%), quality depth < 5, and FS (PhRed-scaled P-value using Fisher's exact test to detect strand bias) > 200. After filtering, the SNPs were filtered out again by removing those within 10 bp of INDELS. For the last SNP filtering step, only biallele-type SNPs were considered to be real SNPs. For the haplotype information on each chromosome, we used BEAGLE (Browning and Browning, 2007) to infer the haplotype phase and impute missing alleles simultaneously for the entire set in the goat population.

### Statistical analysis

To identify and compare highly variable regions, the distribution of variants and nucleotide diversity information was obtained based on the goat reference genome (Dong et al., 2013) from the Goat Genome Database website (<http://goat.kiz.ac.cn/>). VCFtools was used to count the number of variants and the nucleotide diversity in each 1-Mb bin region (Danecek et al., 2011).

### Variant annotation

To identify SNPs that contribute to differences in phenotypes

between native Korean goats and crossbred goats, VCF format files of candidate SNPs were separately concatenated using VCFtools (Danecek et al., 2011). Variant regions were then annotated using SNPEff (Cingolani et al., 2012) with a GFF format file (Dong et al., 2013) from the Goat Genome Database website (<http://goat.kiz.ac.cn/>). We filtered out variants that were not population-specific, and we searched for breed-specific non-synonymous genes. To identify genes that have distinct allele types (between Korean native goats and crossbred goats), we used the protein IDs of the Ensembl Genome Browser [<http://www.ensembl.org>; (Hubbard et al., 2002)] in an annotated file.

### Principal component analysis

We performed principal component analysis (PCA) (Jackson, 2005) to examine population differentiation between native Korean goats and crossbred goats using genotype data from 15 native Korean goats and 11 crossbred goats. We used GCTAtool (Browning and Browning, 2007), which implements PCA in EIGENSTRAT (Price et al., 2006), to estimate eigenvectors. VCFtools (Danecek et al., 2011) and PLINK (Purcell et al., 2007) were used to prepare input data sets for GCTAtool (Yang et al., 2011).

### Population structure analysis

We used STRUCTURE software (Evanno et al., 2005) to identify the genetic proportions of each goat individual from ancestral populations. STRUCTURE software uses Bayesian algorithms to detect the true number of clusters, also referred to as K (the number of ancestral populations). BEAGLE was used to generate input files for running STRUCTURE using single nucleotide variants (SNVs) in every 50 SNVs located in only intergenic regions to avoid bias from evolutionary gene acceleration. We used 100,000 iterations with 2,000 burn-in iterations for each analysis from K = 2-4.

### Deciphering candidate genes under positive selection

We used the cross-population extended haplotype homozygosity (XP-EHH) method (Sabeti et al., 2007) to test for signatures of selective sweeps, which calculates haplotype decay separately for each group using the EHH (Sabeti et al., 2002). The test integrates these values with respect to genetic distance and calculates a log ratio iHH for the native Korean goats and the crossbred goats. The calculation for XP-EHH was performed using the software xpehh [(Sabeti et al., 2007); <http://hgdp.uchicago.edu/Software/>]. We assumed that genetic distance was equal to physical distance. These log ratios (unstandardized XP-EHH) were standardized to have a mean of zero and a variance of one, and these XP-EHH scores or z-scores were assigned P-values assuming a normal distribution, with the exception of tails that diverged from the null expectation. The top 100 loci based on P-value (Benjamini and Hochberg, 1995) were considered strong signals. Since XP-EHH is not sensitive to allele frequencies, there was no need to stratify data into frequency bins before determining significance. Next, we performed the cross-population composite likelihood ratio test (XP-CLR) using XP-CLR software with non-overlapping windows of 50 kb (Chen et al., 2010). The XP-CLR used two models: the Brownian motion model was used to model genetic drift under neutrality and a deterministic model was used to approximate the effect of a selective sweep on SNPs in the vicinity (Chen et al., 2010). The XP-CLR uses allele frequency differentiation between populations. Compared to the allele frequency spectrum-based methods, the XP-CLR is much

**Table 1.** Major candidate genes obtained from XP-CLR and XP-EHH analysis

Candidate genes	Chromosome	Start	End	Max XP-CLR	XP-CLR <i>P</i> -value	Max XP-EHH	XP-EHH <i>P</i> -value	Description
<i>CCR3</i>	22	52849231	52882844	23.73496	0.005	-	-	Receptor for a C-C type chemokine
<i>HM13</i>	13	59025564	59064557	21.42064	0.01	-	-	Minor histocompatibility antigen H13
<i>IGSF10</i>	1	114342294	114369614	18.15857	0.01	-	-	Immunoglobulin superfamily member 10
<i>ROBO1</i>	1	24953553	26069449	24.10493	0.01	-	-	Roundabout homolog 1
<i>ROBO 2</i>	1	22148573	22808722	19.23946	0.05	-	-	Roundabout homolog 2
<i>CLNK</i>	6	102650000	102700000	-	-	3.61921	0.0074	Cytokine-dependent hematopoietic cell linker
<i>NTM</i>	29	33380373	34343966	28.250062	0.01	3.68597	0.0061	Neurotrimin
<i>MYO5A</i>	4	54037221	54172686	15.234912	-	-	-	Myosin, Heavy Polypeptide Kinase

**Table 2.** *CCR3* region sequence information of the 11 loci

CHR <sup>1</sup>	BP <sup>2</sup>	Alt <sup>3</sup>	Native <sup>4</sup>	Cross <sup>5</sup>	Ref <sup>6</sup>	CHISQ <sup>7</sup>	P <sup>8</sup>	OR <sup>9</sup>
22	52854331	T	2	21	G	31.92	1.61E-08	0.01667
22	52859337	G	2	21	C	31.92	1.61E-08	0.01667
22	52860819	A	2	21	T	31.92	1.61E-08	0.01667
22	52850476	G	23	5	A	28.72	8.37E-08	38.33
22	52870519	G	21	5	A	23.01	1.61E-06	21
22	52872604	T	21	5	A	23.01	1.61E-06	21
22	52873011	A	21	5	G	23.01	1.61E-06	21
22	52880722	C	21	5	T	23.01	1.61E-06	21
22	52854052	G	20	5	C	20.46	6.08E-06	16.67
22	52851203	C	19	5	T	18.1	2.10E-05	13.57
22	52851723	T	19	5	G	18.1	2.10E-05	13.57

CHR<sup>1</sup> = Chromosome Number; BP<sup>2</sup> = SNP physical location; Ref<sup>3</sup> = Reference allele code; Alt<sup>4</sup> = Alternate allele code; Native<sup>5</sup> = Korean native goats alternative allele frequency; Cross<sup>6</sup> = Cross breed goats alternative allele frequency; CHISQ<sup>7</sup> = chi-square test statistic (1df); P<sup>8</sup> = p-value; OR<sup>9</sup> = Odds ratio

more robust when ascertaining bias in SNP discovery (Chen et al., 2010). We designated windows with an XP-CLR value in the top 1% of the empirical distribution. Genes located in the regions under significant selection were annotated and as well as those undergoing XP-EHH analysis, the lowest top 100 loci based on P-value were regarded as selective sweep candidates. Candidates identified from both the XP-EHH and XP-CLR results were annotated to the closest genes that spanned (partially or completely) the window regions using CHIR\_1.0 (Dong et al., 2013). These were defined as candidate genes. DAVID (Database for Annotation, Visualization, and Integrated Discovery) was used for gene ontology and pathway analyses (Dennis Jr et al., 2003). Finally, to identify *CCR3* sequence structure, gene association studies were performed using PLINK (Purcell et al., 2007).

## RESULTS

### Genetic variations in native Korean and crossbred goats

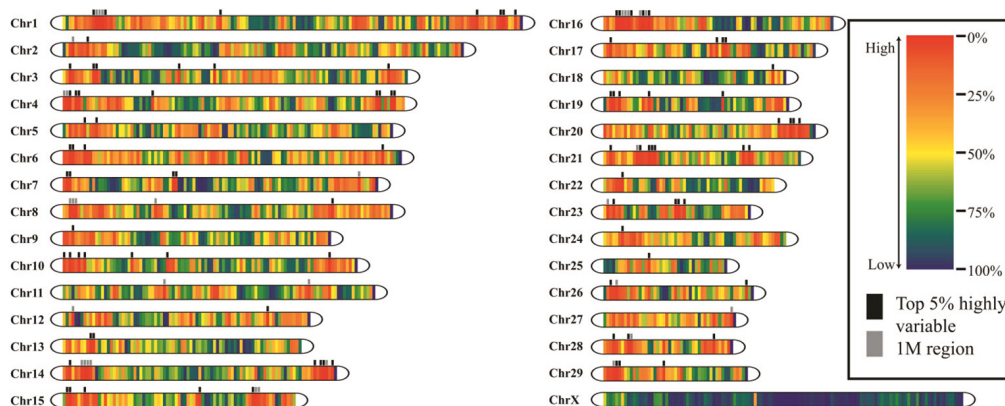
We generated NGS pair-end reads to about ten-fold coverage for 15 native Korean goats and 11 crossbred goats using Illumina HighSeq2000 to obtain goat re-sequencing data. For each individual goat, over 92% of all the reads [excluding pos-

sible polymerase chain reaction (PCR) duplicates] were successfully aligned to the reference goat genome [domestic goat, *Capra hircus*, 2n = 60, predicted size = 2.66 Gb; (Dong et al., 2013)] (Supplementary Table 1). The mapped reads covered an average of 94.77% of the reference genome (Supplementary Table 2), and read coverage of the reference genome was calculated using BEDTools (Quinlan and Hall, 2010).

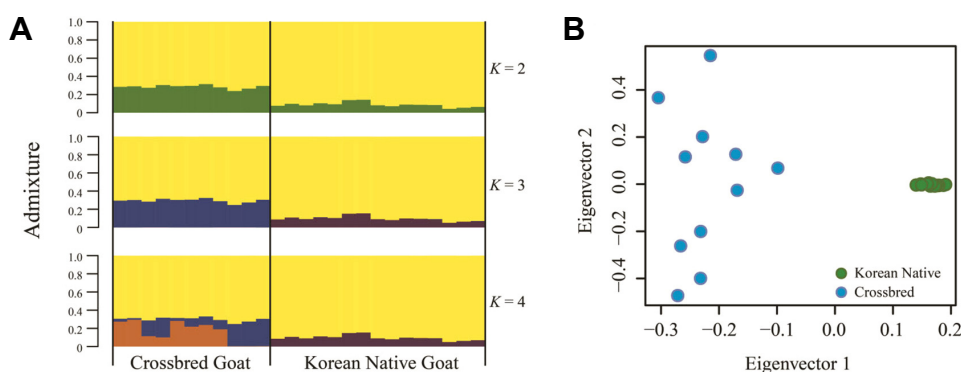
In total, 22,759,033 SNVs and 2,450,921 INDELs were identified, and 26.6% of these SNVs and 26.8% of the INDELs were located in genic regions (Supplementary Table 3), while 73.4% of the SNPs and 73.2% of the INDELs were located in intergenic regions. Within the genic regions, 845,203 SNVs and 108,653 INDELs were detected in only the native Korean goat population and 1,661,071 SNVs and 108,653 INDELs were detected in only the crossbred goat population (Supplementary Table 4). The average nucleotide change rate was 1 per 110 base pairs. Distribution of the SNVs in the goat genome showed that the terminal part of the chromosome was highly variable, and that the X chromosome showed less variation than the autosomes (Fig. 1).

### Nucleotide diversity of the Korean native goat population

We examined the nucleotide diversity of native Korean goats



**Fig. 1.** Distribution of SNVs in the Korean native goat genome. Each segment represents a 1-Mb binning region of chromosomes, which were colored according to the number of SNVs in the region. The 1-Mb binning region with red color showed more variations, and the binning region with blue color showed less variation. The top 5% of highly variable regions contained a dot on the chromosome bar. A black dot represents the 1-Mb binning region including the genic region, and the gray dot represents the 1-Mb binning region including the intergenic region.



**Fig. 2.** Population structure of goat breeds assessed using STRUCTURE and Result of Principal Component Analysis. (A) Each segment represents the proportion of a goat individual genome from ancestral populations. Different colored segments in individuals assume that part of the genome originated from different ancestral populations. This figure shows the genetic structure of goat breeds when we assume that the number of ancestral population of goats is 2-4. (B) Green circles are individuals in Korean native goats, and light blue circles are individuals in crossbred goats. The horizontal axis indicates eigenvector 1, and the vertical axis indicates eigenvector 2. Values of eigenvalues were estimated using GCTA tool.

and crossbred goats using VCFtools (Danecek et al., 2011). We also counted the number of SNVs and integrated them for each 1-Mb bin region of the genome using VCF format files, which contained variant information on 26 goats based on the reference goat genome (Dong et al., 2013). The overall distributions of native Korean goats and crossbred goats in terms of nucleotide diversity differed. Native Korean goats generally showed lower nucleotide diversity (total average nucleotide diversity = 0.0007) than the crossbred goats (total average nucleotide diversity = 0.0010) and the SNV density of the native Korean goats (total average SNV density = 6662.83) was also lower than the crossbred goats (total average SNV density = 7701.61) in the same genomic regions (Supplementary Figs. 1 and 2).

**Population structure of Korean native and crossbred goats**  
According to the PCA results (Jackson, 2005), the native Korean goats were clearly distinct from the crossbred goats (Fig. 2).

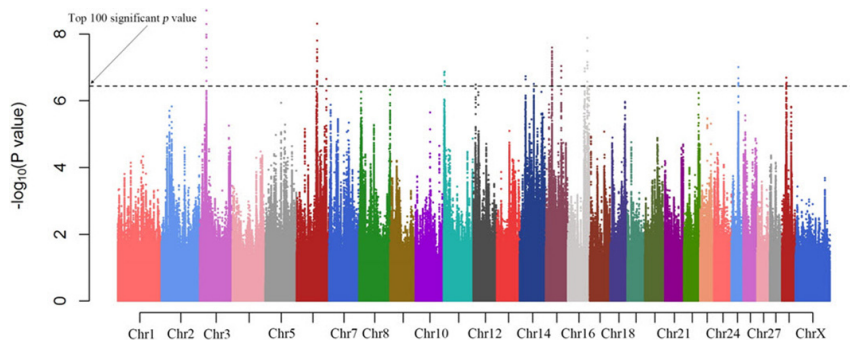
We then examined the genetic structures of native Korean and crossbred goat populations through admixture analysis using STRUCTURE software (Evanno et al., 2005). We observed that the native Korean and crossbred goats shared a majority of their ancestral states and that the crossbred goats originated from the crossbreeding of native Korean native goats with other breeds (Fig. 2).

#### Highly variable native Korean goat genomic regions

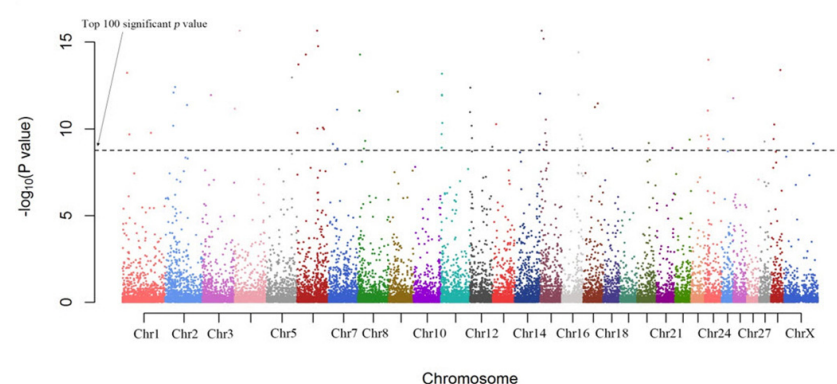
We took the top 5% of the highly variable regions among all the chromosomal regions in each population as significant. We referenced the Goat Genome Database (Dong et al., 2013) and the Ensembl Genome Database (Hubbard et al., 2002) to identify gene locations and annotation information (Supplementary Table 5). We then performed gene ontology (GO)-term analysis of gene sets in highly variable regions that we identified to clarify the differences in the gene distribution of native Korean goat variable regions. The DAVID analysis tool (Dennis Jr et al.,



## A. XP-EHH



## B. XP-CLR



**Fig. 3.** Manhattan plot shows the (A) XP-EHH and (B) XP-CLR between Korean native goat and crossbred goat populations. The vertical axis indicates  $-\log_{10}(\text{p-value})$  of XP-EHH and XP-CLR values

2003) was used for analysis of GO-terms and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Huang et al., 2009). As a result, the variable regions were found to be significantly enriched with genes involved in sensory perception in terms of the Biological Process GO-term and Olfactory transduction in the KEGG pathways (Supplementary Table 6).

### Non-synonymous SNP study of native Korean goats

Among 16,570,906 SNV sites, we identified 76 significant non-synonymous SNP sites of native Korean goats using generated re-sequenced data and SNPeff Software. We identified the chymosin (*CYM*) and collagen, type Xi, alpha 2 (*COL11A2*) genes which may contribute to growth and body size differences between the two populations (Supplementary Table 7).

### Putative selective sweep signatures observed in the native Korean population

We estimated the XP-EHH values (Sabeti et al., 2007) to detect selective sweep regions and performed a pairwise test of the native Korean and crossbred goat populations (Fig. 3). The genome was split into non-overlapping segments of 50 kb and we computed the maximum XP-EHH score in each segment. To define the empirical P-value, the genomic windows were binned in increments of 500 SNPs according to methods used elsewhere (Lee et al., 2014). The regions with P-values less than 0.01 (1%) were considered strong signals of positive selection in the native goat population. Based on the P-value from the XP-EHH test, we identified 82 loci under positive selection. Significant loci were annotated to the closest genes and we found 64 genes (Supplementary Table 8).

In addition, we used an XP-CLR (Chen et al., 2010) to detect selective sweep regions between the two populations (Fig. 3). Using the top 100 XP-CLR score regions, 161 significant genes were identified in the native Korean population. Based on the 161 genes, we performed GO-term analysis using the DAVID analysis tool (Dennis Jr et al., 2003). GO-term cell adhesion (FDR < 0.05) was among the most enriched functional categories that might be related to lumbar paralysis. In addition, neuron development was enriched in the GO-term (FDR < 0.05) (Supplementary Table 9). Among the genes identified in native Korean goats, we chose *CCR3*, which is involved in lumbar paralysis, and calculated the frequencies of SNPs between the native and crossbred goats using the PLINK program. As a result, we found 11 loci that are related to *CCR3* sequence differences (Table 2).

## DISCUSSION

### Goat genomes confirm breeding history

Native Korean goat populations and some outbred goat lines were used to form crossbred goat populations; this population was based on native Korean goats but was meant to improve on the inferior traits of these same goats. From whole-genome sequencing data, we observed a reduction in nucleotide diversity in native Korean goats compared with the crossbred goats; this might be an indication of inbreeding in native Korean goats (Supplementary Fig. 1). In support of our findings, other scholars (Odashima et al., 2006) have shown that native Korean goats have lower genetic variability compared with other Asian goat populations.

The results of our admixture analysis using STRUCTURE (Evanno et al., 2005) are presented in Fig. 2. The proportion of imported alleles increased during crossbreeding, while a significant portion of the alleles became indigenous allele majorities in the native Korean goat population. In the crossbred goat population, nucleotide diversity in the genome may have increased due to the consistent introduction of new alleles, unlike in the native Korean goat population.

### Genes involved in highly variable regions

Recombination rates are known to affect nucleotide diversity (Nachman, 2001) and the terminal regions of chromosomes are known to be prone to recombination events. Based on the distribution of variation in the goat genome (Fig. 1), the terminal regions of chromosomes showed more variation than the other regions, and the top 5% of highly variable regions were enriched with genes involved in olfactory sensors and neurological systems. From the results of the non-synonymous SNP analysis in our study, we identified the enrichment of amino acid substitution in genes that were related to olfactory sensors.

Genes involved in olfactory systems may have had more opportunities to gain variants in nucleotide sequences. Olfactory receptors interact with odorant molecules in the nose to initiate a neuronal response that triggers the perception of a smell. Odor molecules in the environment are detected by olfactory receptors. For animals, olfactory receptors are essential to finding nutritious food and in their ability to avoid eating toxic substances, avoid predators, and identify suitable mating partners and their offspring (Mombaerts, 2004; Niimura, 2009). Native Korean goats were developed under feed shortage conditions and were as a result forced to graze in fields, freely or confined, and tend to overgraze shrubs and the bark of trees. The digestibility of these plants is low. The positive selection of olfactory genes might therefore have arisen as an adaptation to these conditions. Saanen goats, on the other hand, represent commercial breeds that have been artificially selected for intensive production (Choi et al., 2006; Son, 1999).

Native Korean goat populations are also distinct from crossbred goat populations based on non-synonymous SNP patterns of the *CYM* and *COL11A2* genes. *CYM* is an enzyme involved in milk ingestion of young ruminant animals (Emmons and Lister, 1976), and *COL11A2* is involved in osteochondrodysplasias (Vikkula et al., 1995) and Stickler syndrome (Sirkko-Osadsa et al., 1998). These genes may affect the early period of the goat life cycle and body development, and may contribute to differences in the growth of young kids in these two populations.

### Selective sweep regions in native Korean goat populations

Native Korean goats are resistant to lumbar paralysis, a condition that has a severe effect on goats of exotic origin introduced into the peninsula (Son, 1999). Lumbar paralysis is a common disease in ungulate mammals such as goats, sheep, or cattle, and it is transmitted through mosquitos that carry filarial parasites called *Setaria digitata*. This parasite invades the central nervous system in places like the brain or spinal cord causing a disorder in the hind legs (Son, 1999). The positive selection of the *CCR3* gene may have an essential role in the native Korean goat's resistance to this disorder. This gene encodes a chemokine receptor that is expressed in eosinophils and in  $T_H1$  and  $T_H2$  cells (Sallusto et al., 1997) and contributes to an immune response by mobilizing these immune cells when there is a parasitic infection. A previous study reported that antigens of the adult *S. digitata* induced a type of  $T_H1$  and  $T_H2$  cytokine

response (Dalai et al., 1998). In our study, we found SNP frequency differences between the breeds, which might be the reason for the superior resistance to lumbar paralysis via an immune system response (Son, 1999). Additionally, three more genes (*CLNK*, *HM13*, and *IGSF10*) were identified as positively selected in relation to immune response (Dybkaer et al., 2007; Leo and Schraven, 2001; Severino et al., 2014). In addition to immune related genes, we detected neurologically significant genes. The *ROBO1* gene, known for axon guidance and neuronal cell migration, is expressed in mesenchymal stem cells, which differentiate into neuronal cells and are involved in neuroprotective effects (Crigler et al., 2006; Uccelli et al., 2011; Wright et al., 2011). Interestingly, some other genes (*ROBO2* and *NTM*, for instance) are also involved in the development of the neural system (Hivert et al., 2002; Yu et al., 2012). These genes may play a role in recovering damaged nerve cells, and therefore, may contribute to lumbar paralysis resistance. As a result of this study, we have a better understanding of the selective sweep signatures and genetic features of native Korean goat populations, and in turn, we have a better understanding of the physiology of this important domesticated species.

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

### ACKNOWLEDGMENTS

This work was carried out with the support of the Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ01040603), Rural Development Administration, Republic of Korea.

### REFERENCES

- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Statistical Society. Series B (Methodological)*, 289-300.
- Browning, S.R., and Browning, B.L. (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.
- Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. *Genome Res.* 20, 393-402.
- Choi, S., Choy, Y., Kim, Y., and Hur, S. (2006). Effects of feeding browses on growth and meat quality of Korean Black Goats. *Small Ruminant Res.* 65, 193-199.
- Cingolani, P., Platts, A., 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* 6, 80-92.
- Crigler, L., Robey, R.C., Asawachaicharn, A., Gaupp, D., and Phinney, D.G. (2006). Human mesenchymal stem cell subpopulations express a variety of neuro-regulatory molecules and promote neuronal cell survival and neurogenesis. *Exp. Neurol.* 198, 54-64.
- Dalai, S.K., Das, D., and Kar, S.K. (1998). *Setaria digitata* Adult 14- to 20-kDa antigens induce differential  $T_H1/T_H2$  cytokine responses in the lymphocytes of endemic normals and asymptomatic microfilariae carriers in Bancroftian Filariasis. *J. Clin. Immunol.* 18, 114-123.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., and Sherry, S.T. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156-2158.
- Dennis Jr, G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 4, P3.

- Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., Tosser-Klopp, G., Wang, J., Yang, S., and Liang, J. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat. Biotechnol.* *31*, 135-141.
- Dybkaer, K., Iqbal, J., Zhou, G., Geng, H., Xiao, L., Schmitz, A., d'Amore, F., and Chan, W.C. (2007). Genome wide transcriptional analysis of resting and IL2 activated human natural killer cells: gene expression signatures indicative of novel molecular signaling pathways. *BMC Genomics* *8*, 1.
- Emmons, D., and Lister, E. (1976). Quality of protein in milk replacers for young calves. I. Factors affecting in vitro curd formation by rennet (chymosin, rennin) from reconstituted skim milk powder. *Canadian J. Animal Sci.* *56*, 317-325.
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* *14*, 2611-2620.
- Hivert, B., Liu, Z., Chuang, C.-Y., Doherty, P., and Sundaresan, V. (2002). Robo1 and Robo2 are homophilic binding molecules that promote axonal growth. *Mol. Cell. Neurosci.* *21*, 534-545.
- Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* *37*, 1-13.
- Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., and Down, T. (2002). The Ensembl genome database project. *Nucleic Acids Res.* *30*, 38-41.
- Jackson, J.E. (2005). A user's guide to principal components (Wiley.com).
- Kim, S.W., Park, S.B., Kim, M.J., Kim, D.H. and Yim, D.-G. (2014). Effects of different levels of concentrate in the diet on physicochemical traits of Korean native black goat meats. *Korean J. Food Sci. Animal Res.* *34*, 457.
- Lee, H.-J., Kim, J., Lee, T., Son, J.K., Yoon, H.-B., Baek, K.-S., Jeong, J.Y., Cho, Y.-M., Lee, K.-T., and Yang, B.-C. (2014). Deciphering the genetic blueprint behind Holstein milk proteins and production. *Genome Biol. Evol.* *6*, 1366-1374.
- Leo, A. and Schraven, B. (2001). Adapters in lymphocyte signalling. *Curr. Opin. Immunol.* *13*, 307-316.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* *25*, 2078-2079.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kerynitsky, A., Garimella, K., Altshuler, D., Gabriel, S., and Daly, M. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* *20*, 1297-1303.
- Mombaerts, P. (2004). Genes and ligands for odorant, vomeronasal and taste receptors. *Nat. Rev. Neurosci.* *5*, 263-278.
- Nachman, M.W. (2001). Single nucleotide polymorphisms and recombination rate in humans. *Trend Genet.* *17*, 481-485.
- Niimura, Y. (2009). Evolutionary dynamics of olfactory receptor genes in chordates: interaction between environments and genomic contents. *Hum. Genomics* *4*, 107-118.
- Odahara, S., Chung, H., Choi, S., Yu, S., Sasazaki, S., Mannen, H., Park, C., and Lee, J. (2006). Mitochondrial DNA diversity of Korean native goats. *Asian Australasian J. Animal Sci.* *19*, 482.
- Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* *38*, 904-909.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., and Daly, M.J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* *81*, 559-575.
- Quinlan, A.R. and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* *26*, 841-842.
- Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., and McDonald, G.J. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* *419*, 832-837.
- Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., and Gaudet, R. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature* *449*, 913-918.
- Sallusto, F., Mackay, C.R., and Lanzavecchia, A. (1997). Selective expression of the eotaxin receptor CCR3 by human T helper 2 cells. *Science* *277*, 2005-2007.
- Severino, P., Silva, E., Baggio-Zappia, G.L., Brunialti, M.K.C., Nucci, L.A., Rigato Jr. O., da Silva, I.D.C.G., Machado, F.R., and Salomao, R. (2014). Patterns of gene expression in peripheral blood mononuclear cells and outcomes from patients with sepsis secondary to community acquired pneumonia. *PLoS One* *9*, e91886.
- Sirko-Osadsa, D.A., Murray, M.A., Scott, J.A., Lavery, M.A., Warman, M.L., and Robin, N.H. (1998). Stickler syndrome without eye involvement is caused by mutations in COL11A2, the gene encoding the  $\alpha 2(XI)$  chain of type XI collagen. *J. Pediatr.* *132*, 368-371.
- Son, Y. (1999). Production and uses of Korean native Black goat. *Small Ruminant Res.* *34*, 303-308.
- Uccelli, A., Benvenuto, F., Laroni, A., and Giunti, D. (2011). Neuroprotective features of mesenchymal stem cells. *Best practice & research Clin. Haematol.* *24*, 59-64.
- Vikkula, M., Madman, E., Lui, V.C., Zhidkova, N.I., Tiller, G.E., Goldring, M.B., van Beersum, S.E., de Waal Malefijt, M.C., van den Hoogen, F.H. and Ropers, H.-H. (1995). Autosomal dominant and recessive osteochondrodysplasias associated with the COL11A2 locus. *Cell* *80*, 431-437.
- Wright, K.T., Masri, W.E., Osman, A., Chowdhury, J., and Johnson, W.E. (2011). Concise review: bone marrow for the treatment of spinal cord injury: mechanisms and clinical applications. *Stem Cells* *29*, 169-178.
- 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.
- Yu, B., Qian, T., Wang, Y., Zhou, S., Ding, G., Ding, F., and Gu, X. (2012). miR-182 inhibits Schwann cell proliferation and migration by targeting FGF9 and NTM, respectively at an early stage following sciatic nerve injury. *Nucleic Acids Res.* *40*, 10356-10365.