IJAS Iranian Journal of Applied Animal Science

Genome-Wide Scanning for Signatures of Selection Revealed Candidate Genes Involved in Pelt Quality and Coat Color in the Karakul Sheep Breed in Compared to other Iranian Breeds

Research Article

A. Mirzapour-Abibagloo[1], N. Hedayat[1*], R. Khalkhali-Evrigh[1], R. Seyedsharifi[1], H. Abdi-Benemar[1], R. Hassanzadeh[2] and A. Tanveer Hussain[3]

[1] Department of Animal Science, Faculty of Agriculture and Natural Resources, University of Mohaghegh Ardabili, Ardabil, Iran
[2] Department of Engineering Mathematics, Faculty of New Technologies, University of Mohaghegh Ardabili, Ardabil, Iran
[3] Department of Biological Sciences, Virtual University of Pakistan, Islamabad, Pakistan

Received on: 31 Aug 2023
Revised on: 10 Dec 2023
Accepted on: 31 Dec 2023
Online Published on: Mar 2024

*Correspondence E-mail: nhedayat@uma.ac.ir

Online version is available on: www.ijas.ir

ABSTRACT

Karakul (KAR) is one of the resistant sheep breeds to harsh desert conditions, which is also known for its excelent lamb pelt quality. This study was performed to identify the signature of selections in the KAR breed using whole-genome sequencing data (WGS) compared with five other Iranian native sheep. Three methods, including population differentiation index (Fst), nucleotide diversity (π), and cross-population extended haplotype homozygosity (XP-EHH) applied to detect the genomic signature of selection. Data analysis leads to identifying 38 shared genes among three methods as positively selected genes for the KAR breed. The most of mentioned genes were associated with coat color (*KIT*, *DVL3*, *YPEL3*, *ERBB4*, *ZNF451*, and *CTSO*), fat and energy metabolism (*GDPD3*, *STARD13*, *ZNF106*, *MAPK3*, *RGS6*, *PHYH*, *AP2M1*, *SPAG9*, *DNAH9*, *NDUFAF6*, and *ARSK*), muscle function (*MYOCD* and *MCTP1*), growth (*CPNE4*), altitude adaptation (*DNAH9* and *SERGEF*), and reproduction (*TBX6*, *PHYH*, *SPAG9*, and *ARSK*). Based on our results, these candidate genes may have a positive effect on the adaptation of the KAR breed to a desert environment.

KEY WORDS Karakul sheep, pigmentation, signature of selection, whole-genome sequencing.

INTRODUCTION

Sheep were domesticated in the Middle East as small ruminants for meat, milk, wool, and fur production (Cilek and Petkova, 2016a). Due to the multipurpose rearing of more sheep breeds, this animal is considered one of the best animals for use in the different regions as an available protein source. Due to its diverse climatic types, Iran is home to various breeds of sheep, including Moghani (MOG), Ghezel (GEZ), Shal (SHA), Afshari (AFS), and Makui (MAK). Most of these breeds are scattered in the west northern regions of the country and mainly raised for meat production. Although the Karakul (KAR) sheep originated in currently Turkmenistan and Uzbekistan countries as a fat-tailed sheep (Näsholm and Eythorsdottir, 2011) it is one of the breeds that is bred in different parts of Iran. In addition to producing meat and suitable wool for carpets, KAR sheep is well known for the high quality of its lambskin, especially at an early age (Safdarian *et al*. 2006) that differentiated it from other native breeds and ecotypes. Considering fatty tail and the origin of KAR sheep, this breed being fatty tail shows very high resistance to harsh

environmental conditions and can survive in extremely cold and hot ambient temperatures (Degen, 2013; Cilek and Petkova, 2016b), and is often bred under a semi-nomadic management system.

Selective signals are regions of the genome that carry significant variants, these regions may be under natural or artificial selection and thus have specific genetic patterns (Qanbari and Simianer, 2014). When particular traits are selected over generations, the frequency of alleles involved in these traits increases in the population and gradually stabilizes (Chen *et al*. 2016). By investigating these fixed patterns, valuable information on the genetic and demographic structure of different species can be found. In recent years, many studies have been performed to identify selection of signatures in different species, including sheep (Chen *et al*. 2021), goat (Guo *et al*. 2018), cattle (Jiang *et al*. 2021), horse (Salek Ardestani *et al*. 2019) and pig (Huang *et al*. 2021). In the present study, whole-genome sequencing (WGS) data were used to identify the selection signatures on the genome of KAR sheep as a breed known for pelt quality. The potential biological function of identified positively selected genes was investigated to better explain the accossiation between these genes and the characteristics of KAR breed. Also, we used WGS data of the other five Iranian sheep breeds to investigate their population structure.

MATERIALS AND METHODS

Data collection and relationship analysis

The genomic data of six Iranian sheep breeds including MOG (n=3), MAK (n=3), GEZ (n=3), SHA (n=3), AFS (n=4), and KAR (n=5) were obtained from the Sequence Read Archive (SRA) at NCBI public database. After downloading the data, the relationship analysis was performed by KING program v.2.1.3 (Manichaikul *et al*. 2010) to inferred pairwise kinships between the samples.

Quality filtration and mapping to reference genome

The FastQC v.0.11.9 software (https://www.bioinformatics.babraham.ac.uk/projects/fastq) was used to perform quality control of the raw sequence reads. Low-quality bases and reads were removed using Trimmomatic software v.0.39 (Bolger *et al*. 2014) based on the following options: LEADING:5, TRAILING:5, SLIDINGWINDOW:5:20, and MINLEN:40. There was no adapter contamination in the studied data. All clean paired-end short reads were mapped to *Ovis aries* reference genome (GCF_016772045.1), using mem command of Burrows-Wheeler Aligner (BWA) v.0.7.17 (Li and Durbin, 2009).

In the next step, all created Sequence Alignment MAP (SAM) files were converted into Binary Alignment MAP (BAM) files using SAMtools (Li *et al*. 2009). The output BAM files were sorted, and then duplicated reads were removed using the MarkDuplicates algorithm in Picard v.2.26.0 program.

Variant calling

Prior to variant calling, by "RealignerTargetCreator" and "IndelRealigner" modules of Genome analysis toolkit (GATK) v.3.7-0 (McKenna *et al*. 2010), a local realignment around InDels was performed on mark duplicated BAM files. Also, we used GATK to call the variants for all individuals by the HaplotypeCaller algorithm in "–ERC GVCF" mode, and all the samples were genotyped jointly by the GenotypeGVCFs module to generate a multi-sample VCF file. Due to the purpose of the study, only single nucleotide polymorphisms (SNPs) were needed for the next steps of the analysis, so they were selected by option "SelectVariants" of the GATK program. GATK, BCFtools, and VCFtools (Danecek *et al*. 2011) were used to filtrate identified SNPs. SNPs with QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0 || SOR > 3.0 were discarded. We also filtered out those SNPs with a cluster windows size of 10bp and a cluster size of 3 SNPs. Finally, we used VCFtools to discard SNPs that did not meet the following criteria: --minDP 5 --maf 0.05 --min-alleles 2 --max-alleles 2 --max-missing 0.8 --hwe 0.001. It should be noted that, sex chromosomes and unplaced contigs were excluded from the final VCF file. The SnpEff v.5.0e program (Cingolani *et al*. 2012) was performed to annotate final SNPs.

Population structure analysis

The genetic relationship between six sheep breeds was investigated using principal component analysis (PCA) by PLINK1.9 (Purcell *et al*. 2007), and then both components were visualized by R software. The identified SNPs were pruned using the "indep-pairwise 100 50 0.1" option in PLINK1.9 before PCA analysis. This command, removes one of the pairs of SNPs with a square correlation greater than 0.1 in windows with 100 and a step size of 50 SNPs. A neighbor-joining (NJ) tree was constructed using an identity-by-state distance matrix generated by PLINK1.9 and this tree was visualized using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree). Admixture analysis was performed using the ADMIXTURE v1.3.0 program (Alexander *et al*. 2009) with K values ranging from 2 to 10. Based on cross-validation error (CV), K= 2 had the lowest CV and was the best. The decay of linkage disequilibrium (LD) was calculated by PopLDdecay (Zhang *et al*. 2019).

Detection of selection signatures

Three methods of population differentiation index (Fst), nucleotide diversity (π) and cross-population extended haplotype homozygosity (XP-EHH) (Sabeti *et al*. 2007) were used to the identification of genomic regions under positive selection in the KAR sheep. In all three methods, the used sliding widow and step size were 50kb and 25kb, respectively. Estimation of Fst (inter-population) and π (intra-population) in each window were performed using VCFtools, and the XP-EHH score for each SNP was calculated by Selscan v.2.0 (Szpiech and Hernandez, 2014).

Windows containing < 10 SNPs were discarded to reduce the probability of identifying false positive selected regions. XP-EHH scores were normalized using the norm program distributed along with Selscan. The Beagle v.5.2 program (Browning *et al*. 2021) were performed with argument burnin=5 and iterations=20 for haplotype phasing. In the present study and identifying signatures of selection for KAR breed, we employed five native sheep breeds (AFS, MAK, MOG, GEZ, and SHA) as reference population and KAR breed as a target population. We transformed Fst values to ZFst:

$$ZFst = (Fst - \mu Fst) / \sigma Fst$$

Where:
μFst and σFst: mean and standard deviation of Fst values in all windows.

Also, π-ratio (π(Other-Five-Breeds)/π(KAR)) were calculated for each window. The average normalized XP-EHH score in each window was calculated using an in-house python script. The genomic windows in the top 1% percentile of ZFst, π-ratio and XP-EHH values were considered the signature of selection in the KAR breed genome.

Annotation and functional enrichment analysis

Genes in selected windows as a putative signature of selection were annotated using BEDtools v.2.27.1 (Quinlan and Hall, 2010) and GTF (gene transfer format) file related to the sheep reference genome (ARS-UI_Ramb_v2). Then, functional enrichment analysis of candidate genes was performed using "g:Profiler" (Reimand *et al*. 2016) to obtain significantly enriched GO and KEGG pathway terms. The calculated P-value for each term was corrected by the Benjamini–Hochberg false discovery rate (FDR) method and enriched terms were considered statistically significant at p-adjusted < 0.05.

RESULTS AND DISCUSSION

An estimated kinship coefficient range >0.354 and 0.177-0.354 corresponds to duplicate/twin and first-degree relationships respectively (Manichaikul *et al*. 2010). Based on relationship analysis, no duplicated or first-degree related samples were identified in the current study. The highest calculated kinship coefficient was between the two samples MAK2 and MOG2 (0.131), however, considering the mentioned ranges of the kinship coefficients, even these two samples are not closely related. About 4.57 billion paired-end reads were obtained for 21 individuals of six indigenous Iranian sheep breeds, and 4.32 billion clean reads remained after trimming. The average depth was 17.45x per individual, ranging from 11.55x to 22.03x. The average sequence coverage for MOG, MAK, GEZ, SHA, AFS, and KAR breeds were 17.33x, 17.14x, 17.73x, 18.02x, 15.36x, and 19.12x, respectively (Table 1). A number of 22,874,242 SNPs were identified and 9,911,105 SNPs were common among the six breeds (Figure 1). The number of SNPs identified for MOG, MAK, GEZ, SHA, AFS, and KAR were 17.19, 17.44, 17.23, 17.17, 18.35, and 19.25 million, respectively. The Ti/Tv ratio was 2.44 for multi-sample VCF files, indicating the excellent quality of the detected SNPs.

The fat-tail KAR sheep, which evolved in harsh desert environmental conditions, is known for unique skin quality and grey color among other Iranian native sheep breeds. This study was performed to explore the genetic variants of KAR sheep and five native breeds (GEZ, AFS, MOG, SHA, and MAK) using whole-genome sequencing data. Also, a report on population structure for all breeds and candidate selective sweeps for KAR sheep was presented. The PCA was performed to show the genetic relationship among the six breeds. The PCA results revealed that the first component of the PCA plot separated the MAK and MOG breeds from the others (Figure 2a). Also, based on these results, it was found that the KAR breed is clearly in a separate group, suggesting that this breed has unique features. To investigate the phylogenetic relationships of these six breeds, the NJ tree was constructed (Figure 2b). The NJ tree showed that except for MOG and MAK, which displayed close genetic relationships, all individuals were classified into different clusters based on their breed background. To further investigate the relationships between populations, ADMIXTURE analysis was performed based on the Bayesian model with K= 2 to 8. According to the admixture results, GEZ and AFS, MAK and MOG, and KAR and SHA showed a similar pattern in the K= 2 (Figure 3a).

Based on the calculated LD for the breeds, MOG and KAR showed the highest and lowest LD, respectively (Figure 3b). By definition, LD is the close and non-random association of several alleles in one locus of the genome that can be influenced by demographic and evolutionary forces and change their patterns (Ardlie *et al*. 2002).

Table 1 Summary information of obtained genomic reads for each breed

Breeds[1]	Sample size	Total reads	Cleaned reads	Cleaned reads (%)	Coverage
MAK	3	627,761,198	590,964,980	94.15	17.14
MOG	3	657,400,728	620,659,296	94.41	17.33
GEZ	3	636,055,925	602,084,813	94.66	17.73
SHA	3	653,269,967	614,026,538	93.66	18.02
AFS	4	807,484,671	766,548,576	94.90	15.36
KAR	5	1,183,671,380	1,128,862,995	95.39	19.12

MAK: Makui; MOG: Moghani; GEZ: Gezel; SHA: Shal; AFS: Afshari and KAR: Karakul.

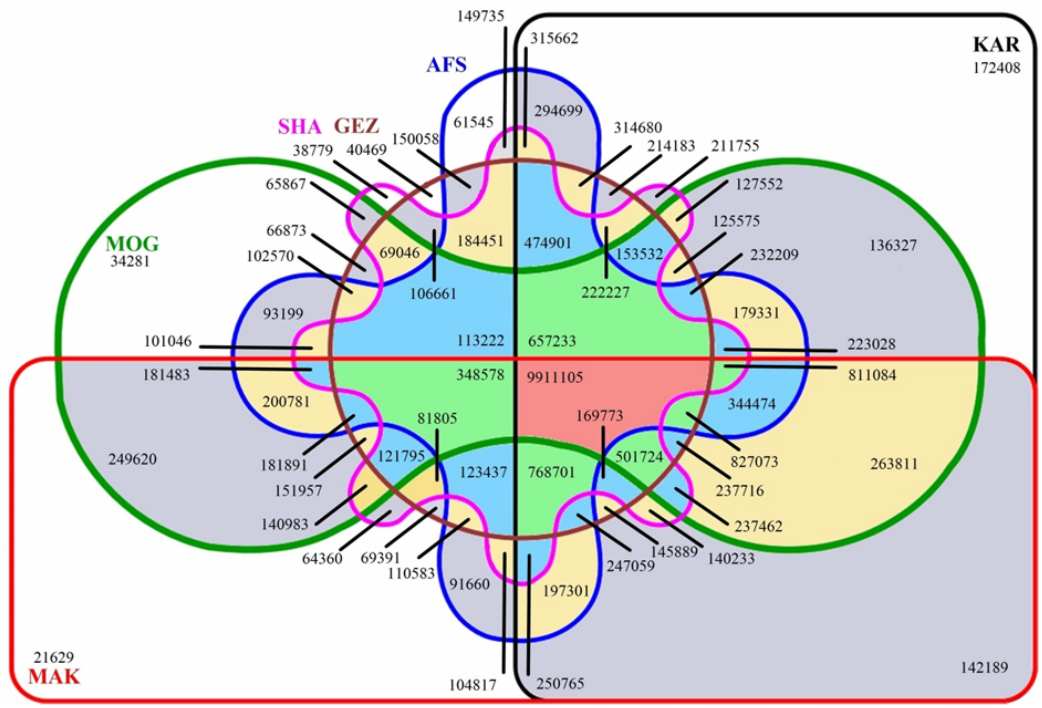

Figure 1 Overlapping and breed specific identified SNPs in Iranian indigenous sheep breeds

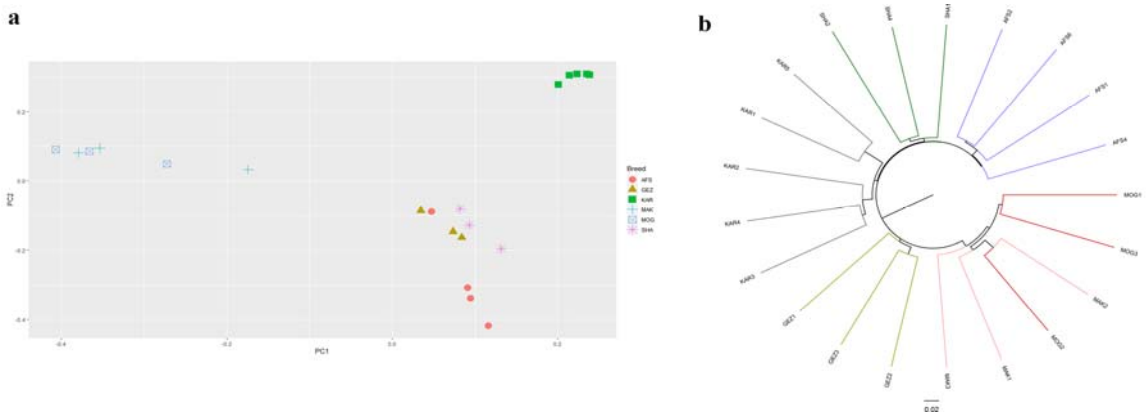

Figure 2 Genetic relationship among the studied six breeds
(a) represents PCA plot for the first two PCs, (b) shows phylogenetic tree of different sheep breeds based on NJ method

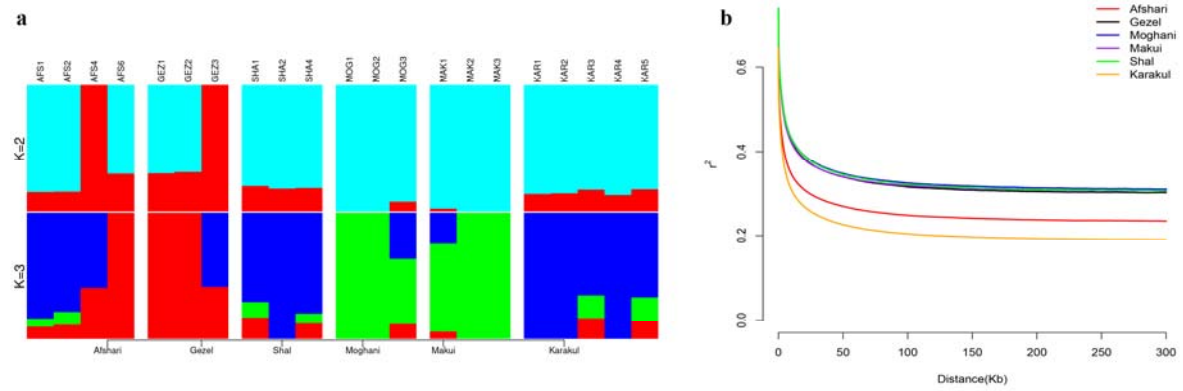

Figure 3 Results of population structure analysis (a) and LD decay (b) for studied six breeds

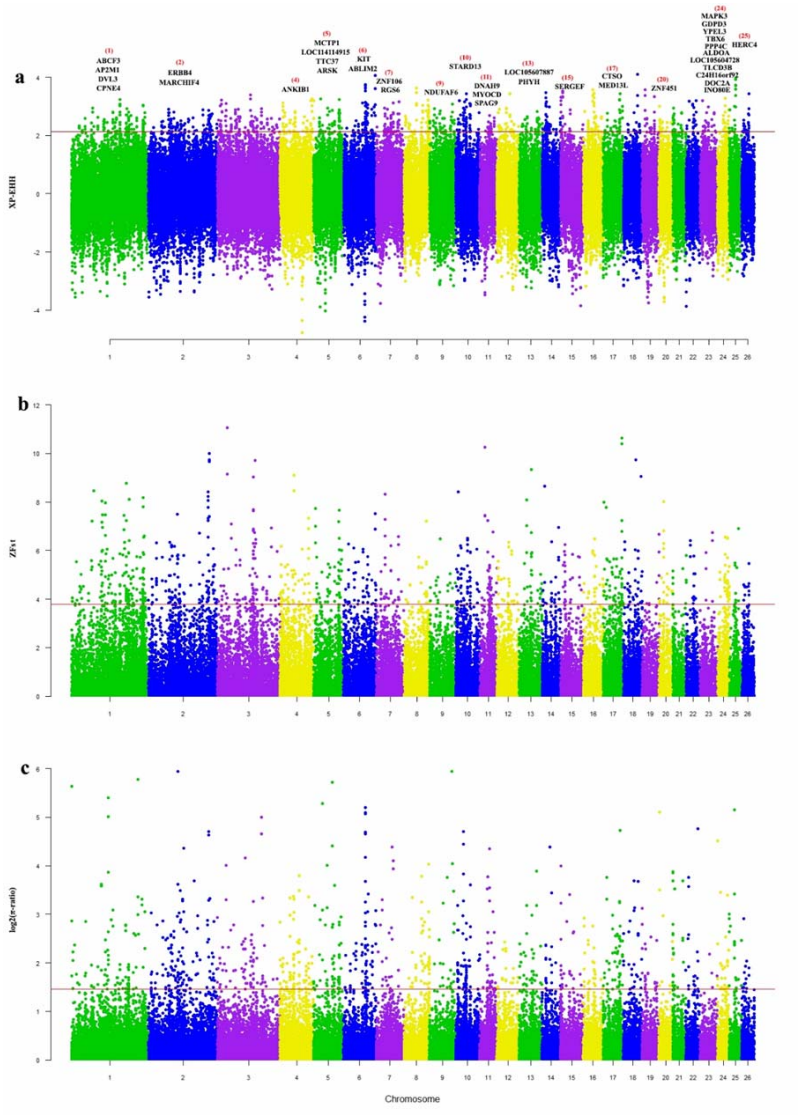

Figure 4 Manhattan plot of the genome-wide distribution of XP-EHH, ZFst and log2 (π-ratio) between KAR and five other Iranian sheep breeds. The red line denotes a threshold of XP-EHH > 2.14, ZFst > 3.79 and log2 (π-ratio) > 1.46. The gene symbols inserted in the figure were identified as positively selected genes using both methods. The red numbers in parentheses represent the number of chromosomes. (a), (b) and (c) represent the results of *XPEHH*, ZFst and log2 (π-ratio), respectively

LD levels in KAR sheep were the lowest among other breeds, possibly indicating higher genetic diversity and larger effective population size for this breed.

Three main approaches, including Fst, XP-EHH, and π were used to identify possible selective signals affecting the unique characteristics of KAR sheep. After Z transformation of Fst for each genomic window, a total of 991 genomic windows containing 641 protein-coding genes were identified as selective sweep based on a defined threshold for ZFst (>3.79). By the XP-EHH method and based on a defined threshold (>2.14), 1003 genomic windows containing 397 protein-coding genes were found as positively selected genes. In the π method, the top 1% of calculated π-ratio (>2.76) were selected as signatures of selection, and based on this threshold, 985 genomic windows containing 646 protein-coding genes were detected for downstream analysis (Figure 4). Finally, the 38 protein-coding genes identified by all three methods were considered as positively selected genes in the KAR breed (Figure 5). Also, 206 genes were found as candidate genes by at least two methods.

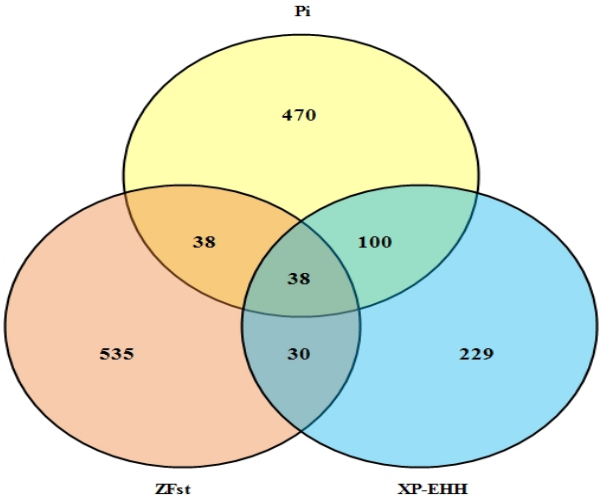

Figure 5 Specific and shared genes obtained by three methods including ZFst, XP-EHH and log2 (π-ratio)

The GO analysis was performed to obtain known biological functions for a different list of identified genes. For 38 shared genes among all three methods, we found three enriched terms in the biological process (BP) category, including "positive regulation of MAPK cascade" (GO: 0043410), "stem cell differentiation" (GO: 0048863), and "positive regulation of smooth muscle cell differentiation" (GO: 0051152). Also, this analysis was performed on 206 genes that were candidates in at least two methods, and in this regard, 105 GO terms were obtained as significant terms. By performing GO analysis on candidate genes identified using Fst, XP-EHH, and π methods, a total of 248, 52, and 226 enriched GO terms were obtained, respectively.

In this study, a selective sweep test was performed by three approaches, including Fst, XP-EHH, and π, and finally, 38 shared candidate genes among all three methods were obtained as positively selected genes in the KAR breed. Most of these genes are involved in skin color (*KIT*, *DVL3*, *YPEL3*, *ERBB4*, *ZNF451*, and *CTSO*), muscle function (*MYOCD* and *MCTP1*), growth (*CPNE4*), altitude adaptation (*DNAH9* and *SERGEF*), fat and energy metabolism (*GDPD3*, *STARD13*, *ZNF106*, *MAPK3*, *RGS6*, *PHYH*, *AP2M1*, *SPAG9*, *DNAH9*, *NDUFAF6*, and *ARSK*), and reproduction (*TBX6*, *PHYH*, *SPAG9*, and *ARSK*).

Studies in pigs have shown that the KIT Proto-Oncogene, Receptor Tyrosine Kinase (KIT) gene is one of the genes involved in pigmentation. Mutations in the KIT encode mast/stem cell growth factor receptor (MGF) that cause differences in the coat color of domestic pigs (Pielberg *et al*. 2002). Huang *et al*. (2021) stated that KIT is an important gene associated with the skin color phenotype. In some research on cattle (Illa *et al*. 2021), KIT was reported to be one of the genes involved in the formation of coat color patterns and facial pigmentation. Studies in goats (Kumar *et al*. 2018; Nazari-Ghadikolaei *et al*. 2018) and horses (Haase *et al*. 2009) have reported similar roles for the KIT gene. Also, yippee like 3 (*YPEL3*) and mediator complex subunit 13L (*Med13L*) genes are two essential factors that play a crucial role in the wnt signaling pathway. The role of YPEL3 is to limit the expression of this pathway (Zhang *et al*. 2016). Since the wnt signaling pathway is probably involved in regulating the pigmentation process in the skin (Nigenda-Morales *et al*. 2018), it seems that these genes influence the pigmentation of the skin in KAR sheep by controlling the mentioned pathway. The positive selection of coat color-related genes in KAR sheep can be viewed from two perspectives, including artificial selection (breeders' trend to colored sheep), as well as natural selection (evolution in harsh desert conditions). The pigmentation process in the skin is probably one of the most important mechanisms to deal with the sun's ultraviolet radiation (López and Alonso, 2014). Since the KAR sheep evolved in the desert condition, the positive selection of these genes has probably been one of the natural selection strategies to adapt this breed to the desert, although further studies are needed to substantiate this claim. The Erb-b2 receptor tyrosine kinase 4 (*ERBB4*), zinc finger protein 451 (*ZNF451*), and cathepsin O (*CTSO*) are other candidate genes influencing skin characteristics that have been reported as selection signals in various studies on cattle, pig, and mice (Gobeil *et al*. 2008; Huang *et al*. 2019b; Bertolini *et al*. 2021). While the ZNF451 causes wrinkles on facial pigskin, the CTSO gene works in a type of malignant skin cancer called mela-

noma in human. Another skin-related gene identified by both ZFst and XP-EHH methods was Collagen Type I Alpha 1 Chain (*COL1A1*). This type of collagen is highly present in the dermis and connective tissues. Evidence suggests that this gene plays a vital role in the healing of heat-damaged skin tissue (Liu *et al*. 2016).

Also, UV radiation can cause skin aging by reducing the expression of type I collagen (Murai *et al*. 2018). Hence this gene seems to be vital for a breed that has evolved in desert conditions and is prone to damage from the high level of UV radiation.

The myocardin (*MYOCD*) and serum response factor (*SRF*) are the most important genes that affect the differentiation and development of cardiac and smooth muscle (Pérot *et al*. 2009). Among the three methods used to identify candidate genes, SRF was detected by the Fst method as the positively selected gene for KAR sheep. This gene appears to play a vital role in the development and stability of the cutaneous epithelium (Verdoni *et al*. 2010). In studies on broilers (Pampouille *et al*. 2018), *MYOCD* was one of the candidate genes involved in pathways related to muscle structure and development, as well as body weight. Multiple C2 and transmembrane domain containing 1 (*MCTP1*) is another active gene in the musculoskeletal system and its expression has been reported to be high in the skeletal muscles of rats (Shin *et al*. 2005). In studies on cattle and chicken, the association of the *CPNE4* (copine 4) gene with growth (Illa *et al*. 2021) and glycogen metabolism (Liu *et al*. 2020) has been reported.

The glycerophosphodiester phosphodiesterase domain containing 3 (*GDPD3*) gene encodes an enzyme that produces 1-acyl-LysoPA and alkyl-LysoPA. Overexpression of the *GDPD3* gene accelerates the production of lysophosphatidic acid resulting in these mechanisms increasing the uptake of fatty acid (Key *et al*. 2020). This gene was a significant candidate in fat metabolism and degradation of phospholipids in other genomic studies (Anthérieu *et al*. 2011; Wang *et al*. 2017). The StAR related lipid transfer domain containing 13 (*STARD13*) gene have a hydrophobic binding site with compounds such as glycerolipids, sphingolipids, and sterols (Alpy and Tomasetto, 2014) and play a salient role in lipid metabolism. *STARD13* is probably associated with the storage and secretion of hepatic lipids in rats (Soffientini *et al*. 2015). In studies on chickens (Sun *et al*. 2015) and sheep (Manzari *et al*. 2019), *STARD13* was selected as a candidate gene for fat and energy metabolism. Also, the results of some studies (Sevillano *et al*. 2018; Xu *et al*. 2020) revealed that *STARD13* and *ZNF106* might be influential in the backfat thickness of the pigs. The mitogen-activated protein kinase 3 (*MAPK3*) gene, along with regulator of G protein signaling 6 (*RGS6*), has recently been shown to play a role in the insulin signaling pathway and lipid metabolism, suggesting that *MAPK3* is probably one of the genes involved in energy metabolism (Lomas-Soria *et al*. 2018; Huang *et al*. 2019a). Phytanoyl-CoA 2-hydroxylase (*PHYH*) is a protein-coding gene involved in the alpha-oxidation process. This enzyme plays a crucial role in converting phytanoyl-CoA to 2-hydroxyphytanoyl-CoA, and there is probably a link between *PHYH* and the differentiation of adipose cells (Schluter *et al*. 2002).

The role of *PHYH* in fat deposition, obesity, and fatty acid oxidation process has been reported in previous studies (Wang *et al*. 2022). Several other genes that we identified as positively selected genes for KAR sheep were associated with fat deposition, lipid, and energy metabolism, including *AP2M1* (Yi *et al*. 2014), *SPAG9* (Zhang *et al*. 2014), *DNAH9* (Wu *et al*. 2016), *NDUFAF6* (Howard *et al*. 2015), and *ARSK* (Guo *et al*. 2014).

The T-Box Transcription Factor 6 (*TBX6*) gene, from the T-Box transcription factor family, plays a vital role in the patterning and specification of somites during embryonic development (Chapman *et al*. 2003). In a study on mice, it was reported that this gene might play an important role in mammalian embryonic development (Yasuhiko *et al*. 2017). Expression of *TBX6* in the early stages of embryonic formation is essential for tail shaping (Showell *et al*. 2004). In addition, in a study on sheep, it was found that *TBX6* is associated with heat stress (Eydivandi *et al*. 2021). We found that three of the genes identified in this study are associated with fertility and sperm motility, including *PHYH* (Oliver *et al*. 2020), *ARSK* (van Son *et al*. 2020), and *SPAG9* (Gao *et al*. 2019).

CONCLUSION

The WGS data of six Iranian native sheep breeds were used to detect the genome-wide signature of selection in the KAR breed. Genomic regions under positive selection that harbor some important genes were identified using three methods. Some of the identified genes in the present study were associated with coat color, reproduction, and energy metabolism, which may have been involved in adapting the KAR breed to desert conditions. Identification of selection signatures and genes involved in pelt quality can be very decisive in the future of pelt producing sheep breeds. Designing breeding strategies using the results of omics studies can be effective in rearing of sheep breeds with maximum pelt quality. It seems that breeding in order to maintain the quality of the pelt until older ages can give the farmers the possibility to produce two products (pelt and meat) instead of one product (pelt).

ACKNOWLEDGEMENT

We thank University of Mohaghegh Ardabili for providing the necessary conditions to conduct present study.

REFERENCES

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

Alpy F. and Tomasetto C. (2014). START ships lipids across interorganelle space. *Biochimie*. **96,** 85-95.

Anthérieu S., Rogue A., Fromenty B., Guillouzo A. and Robin M.A. (2011). Induction of vesicular steatosis by amiodarone and tetracycline is associated with up-regulation of lipogenic genes in heparg cells. *Hepatology*. **53,** 1895-1905.

Ardlie K.G., Kruglyak L. and Seielstad M. (2002). Patterns of linkage disequilibrium in the human genome. *Nat. Rev. Genet*. **3,** 299-309.

Bertolini F., Moscatelli G., Schiavo G., Bovo S., Ribani A., Ballan M., Bonacini M., Prandi M., Dall'Olio S. and Fontanesi L. (2021). Signatures of selection are present in the genome of two close autochthonous cattle breeds raised in the North of Italy and mainly distinguished for their coat colours. *J. Anim. Breed Genet*. **3,** 307-319.

Bolger A.M., Lohse M. and Usadel B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. **30,** 2114-2120.

Browning B.L., Tian X., Zhou Y. and Browning S.R. (2021). Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet*. **108,** 1880-1890.

Chapman D.L., Cooper-Morgan A., Harrelson Z. and Papaioannou V.E. (2003). Critical role for Tbx6 in mesoderm specification in the mouse embryo. *Mech. Dev*. **120,** 837-847.

Chen M., Pan D., Ren H., Fu J., Li J., Su G., Wang A., Jiang L., Zhang Q. and Liu J.F. (2016). Identification of selective sweeps reveals divergent selection between Chinese Holstein and Simmental cattle populations. *Genet. Sel. Evol*. **48,** 1-12.

Chen Z.H., Xu Y.X., Xie X.L., Wang D.F., Aguilar-Gómez D., Liu G.J., Li X., Esmailizadeh A., Rezaei V., Kantanen J. and Ammosov I. (2021). Whole-genome sequence analysis unveils different origins of European and Asiatic mouflon and domestication-related genes in sheep. *Commun. Biol*. **4,** 1307-1314.

Çilek S. and Petkova M. (2016a). Phenotypic correlations between some body measurements and prediction of body weight of Malya sheep. *Bulgarian J. Agric. Sci*. **22,** 99-105.

Çilek S. and Petkova M. (2016b). Impact of age and gender on head measurements of Malya sheep. *Bulgarian J. Agric. Sci*. **22,** 106-109.

Cingolani P., Platts A., Wang L.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*. **6,** 80-92.

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. and McVean G. (2011). The variant call format and VCFtools. *Bioinformatics*. **27,** 2156-2158.

Degen A.A. (2013). Karakul sheep production in Kazakhstan: an efficient collective enterprise under the state farm (sovkhoz) system and its collapse with the break-up of the Soviet Union. *World Rev. Entrepren., Manag. Sustain. Dev*. **9,** 1-9.

Eydivandi S., Roudbar M.A., Karimi M.O. and Sahana G. (2021). Genomic scans for selective sweeps through haplotype homozygosity and allelic fixation in 14 indigenous sheep breeds from Middle East and South Asia. *Sci. Rep*. **11,** 2834-2841.

Gao N., Chen Y., Liu X., Zhao Y., Zhu L., Liu A., Jiang W., Peng X., Zhang C., Tang Z. and Li X. (2019). Weighted single-step GWAS identified candidate genes associated with semen traits in a Duroc boar population. *BMC Genom*. **20,** 1-10.

Gobeil S., Zhu X., Doillon C.J. and Green M.R. (2008). A genome-wide shRNA screen identifies GAS1 as a novel melanoma metastasis suppressor gene. *Genes Dev*. **22,** 2932-2940.

Guo B., Kongsuwan K., Greenwood P.L., Zhou G., Zhang W. and Dalrymple B.P. (2014). A gene expression estimator of intramuscular fat percentage for use in both cattle and sheep. *J. Anim. Sci. Biotechnol*. **5,** 1-12.

Guo J., Tao H., Li P., Li L.I., Zhong T., Wang L., Ma J., Chen X., Song T. and Zhang H. (2018). Whole-genome sequencing reveals selection signatures associated with important traits in six goat breeds. *Sci. Rep*. **8,** 10405-10412.

Haase B., Brooks S.A., Tozaki T., Burger D., Poncet P.A., Rieder S., Hasegawa T., Penedo C. and Leeb T. (2009). Seven novel KIT mutations in horses with white coat colour phenotypes. *Anim. Genet*. **40,** 623-629.

Howard J.T., Jiao S., Tiezzi F., Huang Y., Gray K.A. and Maltecca C. (2015). Genome-wide association study on legendre random regression coefficients for the growth and feed intake trajectory on *Duroc boars*. *BMC Genet*. **16,** 1-11.

Huang H., Cao J., Hanif Q., Wang Y., Yu Y., Zhang S. and Zhang Y. (2019a). Genome-wide association study identifies energy metabolism genes for resistance to ketosis in Chinese Holstein cattle. *Anim. Genet*. **50,** 376-380.

Huang M., Zhang H., Wu Z.P., Wang X.P., Li D.S., Liu S.J., Zheng S.M., Yang L.J., Liu B.B., Li G.X. and Jiang Y.C. (2021). Whole-genome resequencing reveals genetic structure and introgression in Pudong White pigs. *Animal*. **15,** 100354-100365.

Huang T., Zhang M., Yan G., Huang X., Chen H., Zhou L., Deng W., Zhang Z., Qiu H., Ai H. and Huang L. (2019b). Genome-wide association and evolutionary analyses reveal the formation of swine facial wrinkles in Chinese Erhualian pigs. *Aging*. **11,** 4672-4687.

Illa S.K., Mukherjee S., Nath S. and Mukherjee A. (2021). Genome-wide scanning for signatures of selection revealed the putative genomic regions and candidate genes controlling milk composition and coat color traits in Sahiwal cattle. *Front. Genet*. **12,** 699422-699428.

Jiang L., Kon T., Chen C., Ichikawa R., Zheng Q., Pei L., Takemura I., Nsobi L.H., Tabata H., Pan H. and Omori Y. (2021). Whole-genome sequencing of endangered Zhoushan cattle suggests its origin and the association of MC1R with black coat colour. *Sci. Rep*. **11,** 17359-17368.

Key C.C., Bishop A.C., Wang X., Zhao Q., Chen G.Y., Quinn M.A., Zhu X., Zhang Q. and Parks J.S. (2020). Human GDPD3 overexpression promotes liver steatosis by increasing lysophosphatidic acid production and fatty acid uptake. *J. Lipid Res*. **61,** 1075-1086.

Kumar C., Song S., Dewani P., Kumar M., Parkash O., Ma Y., Malhi K.K., Yang N., Mwacharo J.M., He X. and Jiang L. (2018). Population structure, genetic diversity and selection signatures within seven indigenous Pakistani goat populations. *Anim. Genet*. **49,** 592-604.

Li H. and Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. **25,** 1754-1760.

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.

Liu J., Luo C., Yin Z., Li P., Wang S., Chen J., He Q. and Zhou J. (2016). Downregulation of let-7b promotes COL1A1 and COL1A2 expression in dermis and skin fibroblasts during heat wound repair. *Mol. Med. Rep*. **13,** 2683-2688.

Liu X., Liu L., Wang J., Cui H., Chu H., Bi H., Zhao G. and Wen J. (2020). Genome-wide association study of muscle glycogen in jingxing yellow chicken. *Genes*. **11,** 497-505.

Lomas-Soria C., Reyes-Castro L.A., Rodríguez-González G.L., Ibáñez C.A., Bautista C.J., Cox L.A., Nathanielsz P.W. and Zambrano E. (2018). Maternal obesity has sex-dependent effects on insulin, glucose and lipid metabolism and the liver transcriptome in young adult rat offspring. *J. Physiol*. **596,** 4611-4628.

López S. and Alonso S. (2014). Encyclopedia of Life Sciences (ELS). John Wiley and Sons, Chichester, United Kingdom.

Manichaikul A., Mychaleckyj J.C., Rich S.S., Daly K., Sale M. and Chen W.M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*. **26,** 2867-2873.

Manzari Z., Mehrabani-Yeganeh H., Nejati-Javaremi A., Moradi M.H. and Gholizadeh M. (2019). Detecting selection signatures in three Iranian sheep breeds. *Anim. Genet*. **50,** 298-302.

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

Murai M., Tsuji G., Hashimoto-Hachiya A., Kawakami Y., Furue M. and Mitoma C. (2018). An endogenous tryptophan photoproduct, FICZ, is potentially involved in photo-aging by reducing TGF-β-regulated collagen homeostasis. *J. Dermatol. Sci*. **89,** 19-26.

Näsholm A. and Eythorsdottir E. (2011). Characteristics and utilization of sheep pelts. *Small Rumin. Res*. **101,** 182-187.

Nazari-Ghadikolaei A., Mehrabani-Yeganeh H., Miarei-Aashtiani S.R., Staiger E.A., Rashidi A. and Huson H.J. (2018). Genome-wide association studies identify candidate genes for coat color and mohair traits in the Iranian Markhoz Goat. *Front. Genet*. **9,** 105-112.

Nigenda-Morales S.F., Hu Y., Beasley J.C., Ruiz-Piña H.A., Valenzuela-Galván D. and Wayne R.K. (2018). Transcriptomic analysis of skin pigmentation variation in the Virginia opossum (*Didelphis virginiana*). *Mol. Ecol*. **27,** 2680-2697.

Oliver K.F., Geary T.W., Kiser J.N., Galliou J.M., Van Emon M.L., Seabury C.M., Spencer T.E. and Neibergs H.L. (2020). Loci associated with conception rate in crossbred beef heifers. *PLoS One*. **15,** e0230422.

Pampouille E., Berri C., Boitard S., Hennequet-Antier C., Beauclercq S.A., Godet E., Praud C., Jégo Y. and Le Bihan-Duval E. (2018). Mapping QTL for white striping in relation to breast muscle yield and meat quality traits in broiler chickens. *BMC Genomics*. **19,** 1-14.

Pérot G., Derré J., Coindre J.M., Tirode F., Lucchesi C., Mariani O., Gibault L., Guillou L., Terrier P. and Aurias A. (2009). Strong smooth muscle differentiation is dependent on myocardin gene amplification in most human retroperitoneal leiomyosarcomas. *Cancer Res*. **69,** 2269-2278.

Pielberg G., Olsson C., Syvänen A.C. and Andersson L. (2002). Unexpectedly high allelic diversity at the KIT locus causing dominant white color in the domestic pig. *Genetics*. **160,** 305-311.

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. and Sham P.C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet*. **81,** 559-575.

Qanbari S. and Simianer H. (2014). Mapping signatures of positive selection in the genome of livestock. *Livest. Sci*. **166,** 133-143.

Quinlan A.R. and Hall I.M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*. **26,** 841-842.

Reimand J., Arak T., Adler P., Kolberg L., Reisberg S., Peterson H. and Vilo J. (2016). g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res*. **44,** 83-89.

Sabeti P.C., Varilly P., Fry B., Lohmueller J., Hostetter E., Cotsapas C., Xie X., Byrne E.H., McCarroll S.A., Gaudet R. and Schaffner S.F. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*. **449,** 913-918.

Safdarian M., Kafi M. and Hashemi M. (2006). Reproductive performance of Karakul ewes following different oestrous synchronisation treatments outside the natural breeding season. *South African J. Anim. Sci*. **36,** 229-234.

Salek Ardestani S., Aminafshar M., Zandi Baghche M., Banabazi M.H., Sargolzaei M. and Miar Y. (2019). Whole-genome signatures of selection in sport horses revealed selection footprints related to musculoskeletal system development processes. *Animals*. **10,** 53-67.

Schluter A., Giralt M., Iglesias R. and Villarroya F. (2002). Phytanic acid, but not pristanic acid, mediates the positive effects of phytol derivatives on brown adipocyte differentiation. *FEBS Lett*. **517,** 83-86.

Sevillano C.A., Ten Napel J., Guimarães S.E., Silva F.F. and

Calus M.P. (2018). Effects of alleles in crossbred pigs estimated for genomic prediction depend on their breed-of-origin. *BMC Genomics*. **19,** 1-15.

Shin O.H., Han W., Wang Y. and Südhof T.C. (2005). Evolutionarily conserved multiple C2 domain proteins with two transmembrane regions (MCTPs) and unusual Ca2 binding properties. *J. Biol. Chem*. **280,** 1641-1651.

Showell C., Binder O. and Conlon F.L. (2004). T-box genes in early embryogenesis. *Dev. Dyn*. **229,** 201-218.

Sun C., Lu J., Yi G., Yuan J., Duan Z., Qu L., Xu G., Wang K. and Yang N. (2015). Promising loci and genes for yolk and ovary weight in chickens revealed by a genome-wide association study. *PLoS One*. **10,** e0137145.

Szpiech Z.A. and Hernandez R.D. (2014). Selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol*. **31,** 2824-2827.

van Son M., Tremoen N.H., Gaustad A.H., Våge D.I., Zeremichael T.T., Myromslien F.D. and Grindflek E. (2020). Transcriptome profiling of porcine testis tissue reveals genes related to sperm hyperactive motility. *BMC Vet. Res*. **16,** 1-18.

Verdoni A.M., Ikeda S. and Ikeda A. (2010). Serum response factor is essential for the proper development of skin epithelium. *Mamm. Genome*. **21,** 64-76.

Wang S., Yang C., Pan C., Feng X., Lei Z., Huang J., Wei X. and Ma Y. (2022). Identification of key genes and functional enrichment pathways involved in fat deposition in Xinyang buffalo by WGCNA. *Gene*. **818,** 1-12.

Wang Y., Ma C., Sun Y., Li Y., Kang L. and Jiang Y. (2017). Dynamic transcriptome and DNA methylome analyses on longissimus dorsi to identify genes underlying intramuscular fat content in pigs. *BMC Genomics*. **18,** 1-18.

Wu H., Liu Y.H., Wang G.D., Yang C.T., Otecko N.O., Liu F., Wu S.F., Wang L., Yu L. and Zhang Y.P. (2016). Identifying molecular signatures of hypoxia adaptation from sex chromosomes: A case for Tibetan Mastiff based on analyses of X chromosome. *Sci. Rep.* **6,** 35004-35012.

Xu P., Ni L., Tao Y., Ma Z., Hu T., Zhao X., Yu Z., Lu C., Zhao X. and Ren, J. (2020). Genome-wide association study for growth and fatness traits in Chinese Sujiang pigs. *Anim. Genet*. **51,** 314-318.

Yasuhiko Y., Hirabayashi Y. and Ono R. (2017). LTRs of Endogenous Retroviruses as a Source of Tbx6 Binding Sites. *Front. Chem*. **5,** 34-45.

Yi G., Qu L., Liu J., Yan Y., Xu G. and Yang N. (2014). Genome-wide patterns of copy number variation in the diversified chicken genomes using next-generation sequencing. *BMC Genomics*. **15,** 1-16.

Zhang C., Dong S.S., Xu J.Y., He W.M. and Yang T.L. (2019). PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*. **35,** 1786-1788.

Zhang H., Du Z.Q., Dong J.Q., Wang H.X., Shi H.Y., Wang N., Wang S.Z. and Li H. (2014). Detection of genome-wide copy number variations in two chicken lines divergently selected for abdominal fat content. *BMC Genomics*. **15,** 517-525.

Zhang J., Wen X., Ren X.Y., Li Y.Q., Tang X.R., Wang Y.Q., He Q.M., Yang X.J., Sun Y., Liu N. and Ma J. (2016). YPEL3 suppresses epithelial–mesenchymal transition and metastasis of nasopharyngeal carcinoma cells through the Wnt/β-catenin signaling pathway. *J. Exp. Clin. Cancer Res*. **35,** 109-118.