Genomic Scans for Selective Sweeps With Four Complementary Statistical Tests on 14 Indigenous Sheep Breeds From Middle East and South Asia

Sirous Eydivandi (sirous.eidivandi@qgg.au.dk)
Department of Animal Science, Behbahan Branch, Islamic Azad University, Behbahan, Iran
Mohammad Osman Karimi
Department of Animal Science, Faculty of Agriculture, Herat University, Herat, Afghanistan
Mahmood Amiri Roudbar
Department of Animal Science, Safiabad-Dezful Agricultural and Natural Resources Research and Education Center, Agricultural Research, Education & Extension Organization (AREEO), Dezful, Iran
Goutam Sahana
Center for Quantitative Genetics and Genomics, Faculty of Technical Sciences, Aarhus University, 8830 Tjele, Denmark

Research Article

Keywords: sheep genome, FST, xp-EHH, Rsb, FLK, selective sweeps

Posted Date: December 15th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-124388/v1

Version of Record: A version of this preprint was published at Scientific Reports on February 2nd, 2021.
See the published version at https://doi.org/10.1038/s41598-021-82625-2.

Genomic scans for selective sweeps with four complementary statistical tests on 14 indigenous sheep breeds from Middle East and South Asia

Running Head: Selection signatures in indigenous sheep breeds

Sirous Eydivandi[1,2], Mohammad Osman Karimi[3], Mahmood Amiri Roudbar [4], and Goutam Sahana[2]

[1] Department of Animal Science, Behbahan Branch, Islamic Azad University, Behbahan, Iran

[2] Center for Quantitative Genetics and Genomics, Faculty of Technical Sciences, Aarhus University, 8830 Tjele, Denmark

[3] Department of Animal Science, Faculty of Agriculture, Herat University, Herat, Afghanistan

[4]Department of Animal Science, Safiabad-Dezful Agricultural and Natural Resources Research and Education Center, Agricultural Research, Education & Extension Organization (AREEO), Dezful, Iran

*Correspondence:

Sirous Eydivandi

sirous.eidivandi@qgg.au.dk

sirouseidivandi@gmail.com

Abstract

The performance and productivity of livestock have consistently improved by natural and artificial selection over the centuries. Both these selections are expected to leave patterns on the genome and lead to changes in allele frequencies, but natural selection has played the major role among indigenous populations. Detecting selective sweeps in livestock may assist in understanding the processes involved in domestication, genome evolution and discovery of genomic regions associated with economically important traits. We investigated selection signals in this study using SNP genotype data of 14 indigenous sheep breeds from Middle East and South Asia, including six breeds from Iran, namely Iranian Balochi, Afshari, Moghani, Qezel, Zel, and Lori-Bakhtiari, three breeds from Afghanistan, namely Afghan Balochi, Arabi, and Gadik, three breeds from India, namely Indian Garole, Changthangi, and Deccani, and two breeds from Bangladesh, namely Bangladeshi Garole and Bangladesh East. The SNP genotype data were generated by the Illumina OvineSNP50 Genotyping BeadChip array. We applied four complementary statistical tests, F_{ST} (fixation index), xp-EHH (cross-population extended haplotype homozygosity), Rsb (extended haplotype homozygosity between-populations), and FLK (the extension of the Lewontin and Krakauer) to detect selective sweeps. Our results not only confirm the previous studies but also provide a suite of novel candidate genes involved in different traits in sheep. On average, F_{ST}, xp-EHH, Rsb, and FLK detected 128, 207, 222, and 252 genomic regions as candidates for selective sweeps, respectively. Furthermore, overlapping candidate genes were detected by these four tests, including; TNIK, DOCK1, SFMBT1, SPO11, USH2A, TYW1B, NELL2, EML5, and IQCE. The first six of these genes, especially TNIK, DOCK1, and SPO11 were enriched by Gene Ontology (GO) and are involved in embryonic development, immune response, and fertility respectively. Knowledge of candidate genomic regions in sheep populations may facilitate the identification and potential exploitation of the underlying genes in sheep breeding.

Keywords: sheep genome, F_{ST}, xp-EHH, Rsb, FLK, selective sweeps

Introduction

Genetic diversity in livestock is important for improving productivity and addressing future challenges, including food security and mitigating climate change (Groeneveld et al. 2010). Diverse agro-ecological conditions have led to the development of more than 80 native sheep breeds in different geographical districts of Iran, India, Afghanistan, and Bangladesh (FAOSTAT 2018; Eydivandi et al. 2020). Sheep play an important role in the livelihood of many rural and nomadic families in these countries (FAOSTAT 2018). The number of sheep (average from 2010 to 2018) were in India (64 million), Iran (42.7 million), Afghanistan (13.6 million), and Bangladesh (2 million).

Study of population structure gives information on anthropogenic activities and historical processes that have influenced recent gene pools and the genetic relationships among breeds (Ju et al. 2019). Population structure among breeds can be studied using principal component analysis (PCA), admixture and phylogenetic analyses. A range of demographic forces and evolutionary trends affects linkage disequilibrium (LD) patterns on the genome (Ardlie et al. 2002). The LD patterns provide good historical information on the population demography.

Natural and artificial selections leave patterns on the genome that result in differences in allele frequencies among populations (Hohenlohe et al. 2010). If the selection pressure is high at the level of an individual locus, the frequency of the selected variant increases. In addition, selection will change the diversity pattern around the selected variant through genetic hitchhiking, known as a selective sweep (Vatsiou et al. 2016). As a result, different genetic variations and various haplotype structures are fixed over time within separated subpopulations, leading to a wide range of farm animal breeds and distinct genetic populations (Ma et al. 2015). Selective sweeps detected in livestock breeds can add to new information about their population history.

Several methods have been developed to scan genome-wide selective sweeps (Oleksyk et al. 2010). Most of the methods are based on: 1- increases in derived allele frequency and decreases in genetic variation near a selective sweep (hitchhiking) within a population, 2- haplotype length and structure measured by extended haplotype homozygosity (EHH) or EHH-derived statistics, and 3- the differentiation of genetic populations measured by F_{ST} (fixation index) or related statistics (Fariello et al. 2013).

To capture any signal in the genome, depending on the number of populations, temporal context scale, and type of selection signatures more than one method is often needed (Hohenlohe *et al.* 2010). Therefore, we implemented four complementary statistical tests, F_{ST} (Fixation index), FLK (the extension of the Lewontin and Krakauer), xp-EHH (cross-population extended haplotype homozygosity), and Rsb (extended haplotype homozygosity between-populations). We studied selection signature in 14 indigenous sheep breeds from Iran, Afghanistan, India and Bangladesh, the four neighboring countries located in the Middle East and South Asia with different ecological conditions and have more than 80 indigenous sheep breeds. These tests can illuminate selection patterns at the genome level of these indigenous sheep breeds, from adaptation to local environment and selection by breeders to improve production.

Materials and Methods

Populations and genotypic data

We employed 50K SNP genotype data on 453 individuals from 14 indigenous sheep breeds located in Iran, India, Bangladesh, and Afghanistan. Unpublished genotype data from three indigenous Iranian sheep breed, Iranian Balochi (IBL), Lori-Bakhtiari (LOR), Zel (ZEL), were used along with publicly available genotype data on another three Iranian sheep breeds, namely Afshari (AFS), Moghani (MOG), and Qezel (QEZ). We included data on three unpublished genotype data of Afghan sheep breeds, Arabi (ARB), Afghan Balochi

(BLO), and Gadik (GDK). From South Asia, we included three Indian sheep breeds, Changthangi (CHA), Indian Garole (GAR), Deccani (IDC), and two Bengal sheep breeds, Bangladeshi Garole (BGA) and Bangladesh East (BGE) (Sempéré et al. 2015). Information on these 14 breeds is summarized in Table 1.

<< Table 1 about here >>

Genotype quality control

OvineSNP50 BeadChip (Illumina, San Diego, CA, USA) was used to genotype animals. The SNP information was taken from the Illumina Oar_v4 assembly, retrieved from SNPChiMp v.3 (Nicolazzi et al. 2014).

The genotype data from different breeds were merged using PLINK (Purcell et al. 2007). We excluded the SNPs located on sex chromosomes and those with unknown chromosomal position. The quality control was performed using PLINK (Purcell *et al.* 2007). SNPs that were genotyped in less than 90% of the animals, had a minor allele frequency (MAF) lower than 1%, or departed from Hardy–Weinberg proportions at a P-value < 10^{-3} were discarded. Furthermore, individuals with more than 10% missing genotypes were removed from the data set. After quality control, we used Beagle V5.0 software to impute sporadic missing genotypes (Browning & Browning 2007). The fcGENE V1.7 software was used to convert the PLINK formatted files to Beagle format and vice versa (Roshyara 2014).

Genetic diversity and population structure

Individual genetic distances for the 14 sheep breeds were represented by a neighbor-joining tree and displayed using VCF-kit V0.1.6 (Cook & Andersen 2017) and FigTree.v1.4.4 (Rambaut 2015).

We performed a principal component analysis (PCA) to investigate the population structure and to check whether samples for a breed came from a homogeneous population. PCA was done for the 14 sheep breeds using the smartpca program, which is part of EIGENSOFT 7.2.1 (Price et al. 2006).

Linkage disequilibrium (mean of r^2) among SNPs was estimated for the breeds using PopLDdecay V1.01 software, and a Perl script was applied to visualize the results (Zhang et al. 2019).

Admixture Analysis

We analyzed ancestry using ADMIXTURE v1.3.0 to infer breed origins and quantify the populations' admixture (Johnston et al. 2015). For a priori defined ancestry component (K), individual ancestry proportions were calculated with ADMIXTURE v1.3.0, which was an assumption of the number of ancestral populations (Alexander et al. 2009). Using 14-fold cross-validation for K values ranging from 2 to 14, admixture analysis was performed. To identify the most likely number of ancestral populations, the lowest 14-fold cross-validation error was applied. Finally, the admixture graphs were visualized using the R package BITE (Milanesi et al. 2017).

Selection Sweep, Gene Annotation, and Functional Analysis

Neighbor-joining tree and PCA analysis divided the sheep populations in three distinct categories, IR (contains AFS, MOG, QEZ, ZEL, LOR breeds), IN (contains BGA, BGE, GAR, CHA, IDC breeds), and AF (contains IBL, ARB, BLO, GDK breeds) (Table 1). Therefore, we compared pairwise these three categories for selective sweeps analysis.

F_{ST}

Fixation index (F_{ST}) analysis is an widely used approach to identify genetic differentiations between populations compared to the within-population polymorphic frequency (Chang et al. 2018). We performed pairwise comparison for a) IR vs IN, b) IR vs AF, and c) IN vs AF to identify genomic regions under increasing differentiation using VCFtools V0.1.15 (Danecek et al. 2011). For each comparison, the mean of F_{ST} value was computed in all 39348 SNPs. Z transformation of the mean of F_{ST} values ($Z(F_{ST})$) was performed using the "scale" command in R software.

FLK

FLK test is an extension of the original Lewontin and Krakauer (LK) statistic (Bonhomme et al. 2010).

It calculates a population differentiation statistic, which includes a kinship matrix representing the relationship between populations (Weigand & Leese 2018).

This test accounts for population structure and differences in the effective population size by modeling the genetic divergence between populations as a result of drift and population division (Bertolini *et al.* 2018).

For FLK analyses, p-values for a) IR vs IN, b) IR vs AF, and c) IN vs AF were computed as explained in the hapFLK software documentation (Fariello et al. 2012). For each comparison, the negative log p-value was calculated using the hapFLK R script (Fariello *et al.* 2012), and the candidate genomic regions under selection were plotted.

xp-EHH and Rsb Methods

Haplotype-based procedures to study genome-wide patterns of divergence studies, in comparison with SNP-based approaches, have an advantage of avoiding ascertainment bias towards common variants in SNP array design (Browning & Weir 2010; Randhawa et al. 2014).

Extended haplotype homozygosity (EHH) detects selection signatures by comparing a high frequency and extended homozygosity based haplotype with other haplotypes at the selected locus (Sabeti et al. 2007). Complete selective sweeps can be approached by using the cross-population EHH (xp-EHH) test, which compares each population regarding corresponding haplotypes to the other populations. The xp-EHH test compares the integrated EHH profiles between two populations at the same SNP (Sabeti *et al.* 2007). The xp-EHH test has a high power to detect selection signatures in small sample sizes, and therefore grouping of genetically similar breeds may help in gaining power (Sabeti et al. 2007; Pickrell et al. 2009).

Rsb test to identify selective sweeps is based on the same idea of estimation of EHH as xp-EHH test. However in contrast to xp-EHH test, it does not require phasing information (Weigand & Leese 2018). Rsb involved the comparison of the EHH patterns of the same allele (referred to as 'iES') between two populations when EHH is compared between alleles at one SNP within a population (Tang et al. 2007).

In this study, for a) IR vs IN, b) IR vs AF, and c) IN vs AF, we used the xp-EHH and Rsb approaches (Sabeti et al. 2006; Tang *et al.* 2007) to determine selected alleles with higher frequency than expected according to their haplotype length to obtain recent and generally segregating selective sweeps. The haplotypes were phased with Beagle (Browning & Browning 2007), and then xp-EHH and Rsb scores were calculated for each haplotype within a population. Haplotype frequencies were computed for 39348 SNPs. For each locus, the xp-EHH and Rsb score were calculated using the rehh package (Gautier & Vitalis 2012) in R and the candidate genomic regions under selection were obtained.

For each test, the genes that were considered as candidates were found within the intervals spanning the candidate genome regions and also overlapping candidate genes among the tests were captured using the Ovis Oar_v4 reference genome assembly in the Ensembl (Zerbino et al. 2017). The candidate genes visualized using Venpainter tool (Lin et al. 2016).

Absolute correlation among four methods used to detect selection sweeps on: a) IR Vs IN, b) IR Vs AF, and c) IN Vs AF sheep breeds were determined using R codes.

The biological enrichment and functional annotation of the genes under selective pressure were defined using Gene Ontology Consortium (http://geneontology.org).

Results

Populations and Genotype Data

After quality control and imputation of missing genotypes from 463 individuals genotype data for 39531 SNPs from 14 sheep breeds (Table 1), 453 individuals and 39348 SNPs remained for analysis.

Population genetic structure and linkage disequilibrium

The Neighbor-joining phylogenetic tree analysis divided the 14 breeds into three main branches, IR, IN, and AF. The IR group included AFS, MOG, QEZ, ZEL, and LOR, in a main branch (Figure 1, blue color), which illustrated close relationships in the blue branch. These five breeds are from mountainous and forest areas with cold and temperate climates of Iran. The AF group has two distinct sub-branches, one for the three Afghan breeds (ARB, BLO, GDK), and the other own for the Iranian IBL breed (Figure 1, red color). The IBL sheep is from a hot dry climate in the south-eastern deserts of Iran, bordering Afghanistan and Pakistan and therefore IBL is geographically closer to Afghan breeds than the other Iranian breeds in this study. The ARB, BLO, and GDK breeds formed a dense sub-branch that indicates their close genetic relationship. The IN branch included BGA, BGE, GAR, IDC, and CHA (Figure 1, green color). In this branch, two Bengal breeds (BGA and BGE) and GAR formed a distinct cluster, and two other Indian sheep breeds were placed in two separate clusters. The GAR and BGA which are both named Garole breed live in West Bengal state

of India and Bangladesh, respectively. Therefore, a close genetic relationship between these two breeds is expected.

<< Figure 1 about here >>

The LD patterns among the IR and IN groups indicated that the mean of correlation coefficient values (r^2) in both groups dropped rapidly at approximately 10 Kb while the AF group showed a slower drop and its r^2 values at 50 Kb was higher than the other groups (Supplementary Figure S2). The average r^2 at 250 Kb for the IR, IN and AF breeds were 0.0351, 0.0230 and 0.0693, respectively.

PCA results (Figure 2) also indicated close relationships within the IR, IN, and AF groups and supported separation into the three broad geographic groups that were identified by the neighbour-joining tree (Figure 1). Although the breeds clustered according to geographic origin, a gradient based on the geographic distance was less pronounced (Figure 2). In addition, the first principal component (PC1), explaining 13.3% of the total genetic variation among breeds, clearly separated the IR and IN breeds from the AF breeds, thus forming two clusters. Along with the PC1 projection spectra, both IBL and Afghan breeds formed the AF group but a large genetic variation are shown between them. Among the AF breeds, IBL is clearly distant from the other breeds and supported the phylogenetic results. The subclusters of MOG, GEZ, and AFS breeds overlapped, indicating a close relationship and possible admixture of these breeds from the same region in north-western Iran. The LOR breed clearly distant from the other IR breeds which show geographic distance between the LOR from the west and south-western of Iran and the other IR breeds from the north and north-western of Iran. The patterns of genetic variation observed for the AFS, MOG, and GEZ breeds suggested a recent admixture between these three Iranian breeds.

PC2, explaining 6.8% of the total genetic variation, separated the Afghan breeds from the other breeds, but it did not clearly show geographic distance between the IR and IN breeds (Supplementary Figure S1). PC3, explaining 3% of the total genetic variation, separated the IR from the IN and also showed close genetic relationship among two Bengal breeds and GAR, while genetic distances among IDC, CHA, and the other IN breeds (Figure 2).

<< Figure 2 about here >>

Admixture analyses were performed with up to 14 ancestral components (K) (Figure 3). Cross-validation (CV) errors were estimated to identify the most likely number of ancestral populations. The lowest CV error was detected for $K = 12$ (Figure 3a). Although at $K = 10$, CV errors had stagnated after a decline, ancestry components up to $K = 10$ separate breeds, and so it was accepted as the optimal value of K (Figure 3b). Admixture results were in general agreement as PCA. At the first ancestry components ($K = 2$) separated the AF breeds, specially the IBL from the IR and IN breeds, and also at $K = 4$, the IR breeds were separated from the IN breeds. Furthermore, based on geographic origin the breeds were divided as follows: $K = 2$: the AF breeds; $K = 4$: the IN breeds; $K = 7$: the IR breeds. At $K = 10$. All breeds except MOG and QEZ and the Afghan breeds (ARB, BLO, and GDK) were clearly characterized by the breed-specific ancestry components. However, our results showed that increasing the number of K above 10 did not yield a clear MOG and QEZ separation. Therefore, four ancestral components characterized the five IR breeds where AFS, ZEL, and LOR were unambiguously identified. The fourth component was shared between MOG and QEZ, which confirms a close genetic relationship, similar to PCA results. There were no differences among the ARB, BLO, and GDK Afghan breeds from K=2 to K=14 which indicated close genetic relation sheep of

them, confirming results from PCA and the neighbor-joining tree. The Bengal breeds (BGA and BGE) separated from *K*=9, but despite expectation, BGA and GAR with the common name and root separated from *K*=5. In general, compared with the other IN breeds, closer genetic relationships were seen between BGA, BGE, and GAR confirming PCA and the neighbor-joining tree analyses.

<< Figure 3 about here >>

Selective Sweeps Detection

Selective sweeps detection was performed using F_{ST} (Reynolds et al. 1983), FLK (Bonhomme et al. 2010), Rsb (Tang et al. 2007), and xp-EHH (Sabeti et al. 2006). Based on the PCA and the neighbor-joining tree results, these four different tests were conducted for selective sweeps detection on the three pairwise comparisons: a) IR and IN breeds, b) IR and AF breeds, c) IN and AF breeds. The Z-transformation of F_{ST}, $Z(F_{ST})$, values of 39348 SNPs were estimated (Figure 4). For these three pairwise comparison, the maximum of $Z(F_{ST})$ values were 14.524 on chromosome 11 (IR vs IN breeds), 4.744 on chromosome 24 (IR vs AF breeds) and 4.556 located on chromosomes 24 (IN vs AF breeds) (Figure 4). Based on the $Z(F_{ST})$, a total of 131 genes as top 1% candidates for selective sweeps were detected in a) IR and IN breeds, 131 genes in b) IR and AF breeds, and 121 genes in c) IN and AF breeds (Supplementary Table S1). Among these candidate genes, several of them are known for association with economic traits, for example, SLC27A6, ANXA13, ADCY2, HDAC9, TTC8, and WDR70 association with milk traits. HERC2, FTO, TP73, GRM3, KCNIP4, GRM7, and UBR2 related to body weight and growth traits. TMEM132B, TMEM232, and SLC8A3 affected fertility traits. GALNT6, ATP2C1, TMPRSS3, PCDH15, DOCK1, DOCK4, DOCK10, PPA2, CHD3, ITGA4, NFATC1, and ZNF609 involved in the immune system.

The xp-EHH scores were calculated for haplotype frequencies (Figure 5). The top 1% of xp-EHH, considered as selective sweeps, identified 164 genes for a) IR and IN breeds; 236 genes for b) IR and AF breeds; and 221 genes for c) IR and AF breeds (Supplementary Table S2). Many candidate genes found by the xp-EHH method are related to economic traits, such as, OXT, HSPB1, TBX6, GNA12, BMP7, MYH10, TRHDE, IL27, IL4R, and IL21R involved in heat stress, ATP2A1, ATP2B1, LRP12, CD19, MYO18A, PCDH17, BBS9, NFATC2IP, RNF26, RNF139, ZNF572, ZNF655, and ZNF789 associated with immune system, and MEF2C, TRHDE, FAM222B, FAM177A1, and SSC4D influenced body weight and growth traits.

The Rsb scores were calculated for haplotype frequencies (Figure 6). The top 1% of Rsb, considered as selective sweeps, identified 185 genes for a) IR and IN breeds, 249 genes for b) IR and AF breeds, and 233 genes for c) IR and AF breeds (Supplementary Table S3). Many candidate genes specially association with immune response and heat stress were found by Rsb test, such as; ATP2B1, ATP2C1, LRP1B, CXCL1, CD19, DOCK1, DOCK4, UNC5C, ANKRD2, BBS9, NAFTC2IP, RNF139, and ZNF695 in immune system, and IFT22, EIF2A, HSPB1, TBX6, TBX21, GNA12, BMP7, IL16, IL27, IL4R, and IL21R in heat stress. Furthermore, HOXD1, HEXD2, and MTX2 affected the horn traits, and PRLP, TBC1D10B, TMEM151A, TMEM65, TMEM225B, BMPRIB, and BMP7 genes associated with fertility traits were detected as candidate genes using Rsb.

The -log (p-value) values of 39348 SNPs for the FLK test are presented in Figure 7. Based on the -log (p-value), a total of 244 genes as top 1% candidates for selective signals were detected in a) IR and IN breeds; 265 genes in b) IR and AF breeds; and 247genes in c) IN and AF breeds (Supplementary Table S1). Several candidate genes identified using FLK test are related to economic traits, for example, FABP3, SLC27A6, ACP7, ANXA13, HEATR5B, ADCY2, BRD4, BRD8, HDAC9, TTC8, TTC23, WDR7, WDR31, WDR70, and POU6F1 related to milk traits as well as HERC2, FAM169A, FTO, TP73, GRM2, GRM3, and UBR2 for body weight and growth traits. Several candidate genes related to immune system were detected by FLK,

such as, GALNT6, GALNT13, GALNT18, ATP2C1, LRP1B, CXCL14, TMPRSS3, CD34, COL12A1, PCDH15, DOCK1, DOCK4, DOCK10, UNC5B, BBS9, CDH6, CHD3, IRF6, ITGA, LRP1B, NAFATC1, RNF26, ZNF609, and ZNF692.

The F_{ST} and FLK tests with average 128 and 252 genes showed the minimum and maximum captured genes among these four tests. Furthermore, five, six and three concordant genomic regions for a) IR and IN breeds, b) IR and AF breeds, c) IN and AF breeds were identified by F_{ST}, xp-EHH, Rsb, and FLK tests as candidates for selection signals, respectively (Figure 8). These overlapping candidate genes for a) IR and IN breeds include the following genes; Scm-like with four MBT domains protein 1 (SFMBT1) on chromosome 19, plays a role during spermatogenesis. Dedicator of cytokinesis protein 1(DOCK1) on chromosome 22, has an essential role in embryonic development and involved immune response, Neural EGFL like 2 (NELL2) on chromosome three, involved in involved in pubertal development. NCK-interacting protein kinase (TNIK) on chromosome one, the protein encoded by this gene plays important role in embryonic development, especially during the early embryo to blastocyst stages, participates in the regulation of the inflammatory response against infections.

The overlapping candidate genes for b) IR and AF breeds include; Echinoderm microtubule-associated protein-like 5 (EML5) on chromosome seven, may change the assembly dynamics of microtubules to make microtubules are slightly longer but more dynamic and it is possible that Eml5 plays a role during neuronal development in the regulation of cytoskeletal rearrangements, IQ domain-containing protein E (IQCE) on chromosome 24, involved in body development, TRNA-YW Synthesizing Protein 1 Homolog B (TYW1B) on chromosome 24, influenced on the wybutosine biosynthesis pathway. Usherin (USH2A) on chromosome 12, may be involved in the function of synapses and plays an important role in the development and maintenance of cells in the inner ear and retina. SPO11 initiator of meiotic double-stranded breaks (SPO11) on chromosome 13, involved in the production of double-strand breaks (DSB) of DNA and it is specifically

involved in the growth of the testis, maintenance of the male germ line, and maturation of sperm. Three overlapping candidate genes for c) IN and AF breeds were detected; the IQCE, TYW1B, and an unknown gene with Ensemble number (ENSOARG00000025902) which all of these three genes were detected before in b group data (IR and AF).

We also detected overlapping candidate genes for IR Vs IN, IR Vs AF, and IN Vs AF data on the F_{ST}, xp-EHH, Rsb, and FLK tests (Figure 9). For the F_{ST} test PPA2, involved in the immune system, and KCNIP4 plays important role in heart performance and it is related to skeletal muscle growth and also immune response. SYT1, associated with feeding behavior traits such as residual feed intake and TMEFF2, involved in a wide range of traits such as; immune response, milk production and sperm morphology, were detected as overlapping candidate genes for the Rsb test. For the FLK test, PPA2, EML5 genes, which have been found in the previous tests, MGAT5, associate with dry matter intake and NEB, involved in environment adaptation, were detected as overlapping candidate genes on the three different data. We did not find any overlapping candidate genes on the all data by the xp-EHH test.

<< Figure 4 about here >>

<< Figure 5 about here >>

<< Figure 6 about here >>

<< Figure 7 about here >>

<< Figure 8 about here >>

<< Figure 9 about here >>

Biological enrichment analysis of significant biological processes for candidate genes under positive selective pressure revealed 26 Gene Ontology (GO) terms (Table 2). These GO terms reflected protein

function and biosynthetic processes, including the TNIK and DOCK1 genes associated with cytoskeleton organization (GO:0007010) and six other GOs related to the TNIK gene include: regulation of dendrite morphogenesis (GO:0048814), actin cytoskeleton reorganization (GO:0031532), protein localization to plasma membrane (GO:0072659), positive regulation of protein phosphorylation (GO:0001934), protein auto phosphorylation (GO:0046777), and intracellular signal transduction (GO:0035556). Four other GOs associated with DOCK1 include; hematopoietic progenitor cell differentiation (GO:0002244), small GTPase mediated signal transduction (GO:0007264), cell migration (GO:0016477), and positive regulation of GTPase activity (GO:0043547). The SFMBT1 gene associated with negative regulation of transcription (GO:0035556). Seven GOs associated with the spo11 include: reciprocal meiotic recombination (GO:0007131), synaptonemal complex assembly (GO:0007130), male meiosis I (GO:0007141), DNA metabolic process (GO:0006259), ovarian follicle development (GO:0001541), oogenesis (GO:0048477), synapsis (GO:0007129), and spermatid development (GO:0007286). Four GOs related to the USH2A gene include: sensory perception of light stimulus (GO:0050953), photoreceptor cell maintenance (GO:0045494), establishment of protein localization (GO:0045184), and sensory perception of sound (GO:0007605). Finally, tRNA processing (GO:0008033) associated with the TYW1B gene.

<< Table 2 about here >>

Absolute correlation coefficients among these four tests on a) IR vs. IN, b) IR vs. AF, and c) IN vs. AF sheep breeds showed the maximum correlation between F_{ST} and FLK on the all data (average: 0.861) and the minimum correlation between FLK and Rsb on IR vs. IN (0.107) and F_{ST} and Rsb on IR vs. AF and IN vs. AF data (average: 0.021) (Figure 10).

<< Figure 10 about here >>

Discussion

The present study investigates the genetic diversity and selective sweeps of 14 sheep breeds from Iran, Afghanistan, India, and Bangladesh. The selective sweeps were studied using the $F_{ST,}$ FLK, xp-EHH and Rsb statistical methods on the three cluster of breeds (IR, IN, and AF). Our goal in the current study was to search the genomes of these indigenous sheep breeds to highlight specific genetic variants or haplotypes that can be used in developing next-generation productive breeds, better suited to diverse Iran environments, in a comparative scale with Indian, Bengal, and Afghan sheep breeds. Furthermore, the other goal was using four comparable selective sweeps tests to cover all the regions of the genomes and capture maximum candidate genes, as well as review their biological function. The results showed that these breeds' genomes contain multiple regions under selection. These regions contain well-known economic trait- related candidate genes. This could help sheep breeders to: 1) improve adaptation in extant breeds; 2) develop new breeds or crossbreeds that are better adapted to local agro-climatic conditions; 3) launch future research work on the genomes of Iranian, Afghan, Indian, and Bengal sheep, and highlight essential genetic variants or haplo-blocks that can be used in the production of higher productivity and efficiency next-generation breeds, better adapted to various Iranian environments, on a comparative scale with Afghan, Indian, and Bengal sheep breeds.

Genetic Relatedness and Geographic Origin

We demonstrated that the IR sheep breeds are genetically distinct from the breeds of IN and AF. Based on their geographic origins, the studied sheep breeds are well clustered, and we categorized the IR, IN and AF breeds into three phylogeographic clades. Close connections between breeds originating in the same

geographical region have been found, as well as little evidence of migration between breeds from different geographical regions. In fact, phylogenetic analysis showed a close genetic relationship among the IR breeds. These breeds are from cold and temperate climates of Iran. On the other hand, the IN breeds showed a closer relationship among BGA, BGE, and GAR breeds from the eastern region of India and Bangladesh. In contrast, IDC from the western peninsular region and CHA from northern Himalayan part of India formed two distinct sub clusters.

Furthermore, the AF cluster showed an IBL sub-cluster and a compact sub-cluster of three Afghan breeds, indicating a closer relationship among Afghan breeds and their genetic distance from IBL.

These findings are consistent with previous research on sheep (Kijas et al. 2012; Ciani et al. 2015; Deniskova et al. 2019; Eydivandi *et al.* 2020), which showed that individuals were separated by global population structure patterns according to their geographical origin.

In accordance with previous findings (Ciani *et al.* 2015; Barbato et al. 2017; Alberto *et al.* 2018), our PCA results demonstrated that the genetic variation was associated with the separation among sheep breeds from different parts of the world. This was further supported by neighbour-joining tree analysis revealing that the population was split according to geographic origin (IR, IN, and AF). Population structure analyses of the IR, IN, and AF breeds clearly reflected the geographic distribution at PC1 and the separation of northern from southern breeds at PC3.

Admixture and phylogenetic patterns

In accordance with the previous analyses, admixture results confirmed that the first few ancestral breed components (K = 2 to K = 5) were related to the geographic origins. High levels of breed admixture were detected among the Iranian (IR and IBL) breeds, and also among the IN breeds. However, low levels of admixture events among the breeds originating from the different geographical regions were detected. For

example, although the GAR and BGA from India and Bangladesh have a common breed name (Garole), they separated at k=5 ancestral breed components, while BGA and BGE which are known as two different breeds in Bangladesh showed more relationship and they have been separated at k=9 which confirm the effect of geographic origin in breeds admixture. Admixture results confirmed genetic divergence identified through the neighbor-joining and PCA.

Inference based on population neighbour-joining trees based on genome-wide allele frequencies clustered the breeds into three monophyletic clades according to the geographical origin. The deepest population split among the AF breeds separated IBL from the other AF breeds. Among the IN breeds, IDC and CHA showed deeper population splits, in line with geographic clades detected by the PCA and admixture analysis. These results support the previous findings (Ciani *et al.* 2015; Barbato *et al.* 2017; Alberto *et al.* 2018).

Genome-wide selective sweeps

The ability of specific genomic regions to detect selective sweeps depends on the selection of analytical tools appropriate to the biological situation but no single method can detect selective sweeps that are both starting and nearly completed. However, combining several tests increases significantly the power to recognize the region selected (Hohenlohe et al. 2010; Vatsiou et al. 2016). Therefore, we used F_{ST}, FLK, xp-EHH, and Rsb test statistics to detect genome-wide selective sweeps in a) IR and IN breeds, b) IR and AF breeds, c) IN and AF populations. F_{ST} was first implemented to measure the degree of genetic differentiation between populations based on variations in allele frequency (Wright 1949). The genomic variation information is provided by F_{ST} at a locus between the populations compared to within the populations. Therefore, the F_{ST} is an evidence of selection: low F_{ST} values indicate negative or neutral selection, while high F_{ST} values indicate positive local adaptation (Kullo & Ding 2007). The older selection events between populations are expected to be identified by F_{ST} (Ma et al. 2014; Maiorano et al. 2018). The xp-EHH test is an extension of EHH (Sabeti et al. 2007), that incorporates information on the relationship between an allele's frequency and

LD measurements with neighboring alleles. Therefore, this test may provide maximal statistical power and low ascertainment bias sensitivity (Tang et al. 2007). The Rsb test is population comparison test to identify selective sweeps (Tang *et al.* 2007). The test is based on the same idea as the XP-EHH, identifies loci similar to the XP-EHH test under selection, but can be implemented with unphased data (Weigand & Leese 2018). Generally, the xp-EHH and Rsb tests are used to detect recent positive selection within population and between-populations, respectively (Oleksyk *et al.* 2010). The FLK (extended Lewontin and Krakauer test) test is based on the assumption that two new populations are formed by the splitting of a population; calculates a statistic of population differentiation, which incorporates a matrix of kinship describing the relationship between populations (Bonhomme *et al.* 2010; Weigand & Leese 2018). For each SNP, the FLK test calculates a global F_{ST}, but allele frequencies are first rescaled using a matrix of population kinship. This matrix, which is estimated from the genome-wide data observed, measures the amount of genetic drift that can be predicted along all branches of the population tree under neutral evolution (Bonhomme *et al.* 2010). Therefore, the integration of these four complementary statistical tests provides a valuable tool for detecting, with greater confidence, positive selection of genomic regions.

For F_{ST} and FLK, only the top 1% $Z(F_{ST})$ values and the top 1% -log(p-value) were considered, respectively to be representing selective sweeps as recommended in previous studies (Kijas *et al.* 2012; Bertolini *et al.* 2018).

Analyses of selective sweeps were reported for several international sheep populations from several countries, including China (Yuan et al. 2017), Europe (Barbato *et al.* 2017; Purfield et al. 2017), Russia (Yurchenko et al. 2019), Egypt (Kim et al. 2016), Brazil (de Simoni Gouveia et al. 2017), and New Zealand (McRae et al. 2014). Furthermore, several studies of selective sweeps on sheep carried out using different tests, including the FLK and hapFLK (Fariello et al. 2014; Alberto *et al.* 2018), hapFLK, FLK, F_{ST}, and

445 hapF$_{ST}$ (Fariello *et al.* 2013), REHH and xp-EHH (ZHAO et al. 2016), F$_{ST}$ and hapFLK (Yuan et al. 2017),
446 F$_{ST}$ and iHS (Kim *et al.* 2016), F$_{ST}$, Rsb, and iHS (de Simoni Gouveia *et al.* 2017).

447 In our study, using F$_{ST}$, xp-EHH, Rsb, and FLK were detected in average 128, 207, 222, and 252 genomic
448 regions as candidates for selective sweeps, respectively. Although the selected candidate regions are narrow
449 and are distributed across different chromosomes, however for F$_{ST}$ and FLK tests, chromosome 1 showed a
450 low value for b and c comparisons which may indicate the genome of two populations are the same in this
451 region and many genes were expected to be commonly fixed in both populations (Maiorano *et al.* 2018),
452 (Figure 4 and 7). Several of these genes encode economically important traits. Such as genes that have
453 directly or indirectly influenced traits for adaptation to hot arid conditions and heat tolerance (TRHDE,
454 IL4R,IL21R, and SLC4A4), which reported as candidate genes involved in heat tolerance on sheep by
455 Berihulay et al. (2019), and the heat shock protein B1 (HSPB1) gene which expresses both at mRNA and
456 protein levels under heat stress on poultry (Sharma et al. 2020), and was reported on sheep (Armstrong et al.
457 2018), and cattle (Srikanth et al. 2017). All of these candidate genes were detected in b (IR and AF), and c
458 (IN and AF) clusters, where the AF breeds are common (Supplementary Table S1, S2, and S3). This indicates
459 that the AF breeds, which are from a hot dry climate, are more adapted to heat tolerance.

460 Many of the genes identified as candidate genes in this study are effective in genetic resistance to disease
461 and immune response. Since genetic resistance against diseases and harsh environmental conditions is one
462 of the important characteristics of indigenous animal breeds, the identification of a large number of genes in
463 this study indicates that the associated genes have been under selection pressure over time due to the natural
464 selection of immune response traits (Scarpa et al. 2003; Mwai et al. 2015). For example, we detected the
465 DOCK family (DOCK1, DOCK4, DOCK10) (Laurin *et al.* 2008; Kunimura *et al.* 2020), ZNF family
466 (ZNF572, ZNF655, ZNF609, ZNF692, and ZNF789) (Cassandri *et al.* 2017; Feng et al. 2017), ATP family
467 (ATP2A1, ATP2B1, and ATP2C1) (Benavides *et al.* 2015), TMEFF2 (Richardson *et al.* 2016), CXCL1

(Atlija *et al.* 2016), PCDH15 (Atlija *et al.* 2016), and (COL12A1, COL15A1, COL27A1) (Atlija *et al.* 2016), candidate genes involved in the immune response (Supplementary Table S1, S2, and S3). Almost all these genes were detected in all three clusters, a (IR and IN), b (IR and AF), and c (IN and AF), which may indicate genetic resistance and high immune response against diseases and harsh environmental conditions in these native breeds.

Calvo *et al.* (2004) showed linkage disequilibrium between FABP3 gene and quantitative trait loci (QTL) for milk fat content trait. Other related milk traits candidate genes included the LRP1B and CNTN4 which previously reported on sheep (Li *et al.* 2020) and cattle (Marete *et al.* 2018). The ITPR2 and SLC27A6 are also two examples of important candidate genes detected by Li *et al*. (2020) on sheep and both have been proposed to be candidate genes for milk and fat production in cattle (Nafikov *et al.* 2013; Chen *et al.* 2018). All of these genes were found in b (IR and AF), and c (IN and AF) clusters, indicating AF breeds may be under selection pressure related to milk traits but it needs further research to conclude. (Supplementary Table S1, S2, and S3).

We found several candidate genes involved in body weight and growth traits specially post-weaning gain in all population clusters, such as the TRHD, UBR2, GRM2, GRM3 (Zhang et al. 2013; Gebreselassie *et al.* 2020).

The TBC1D10B, and BMPR1B are two examples of candidate genes involved in fertility traits which we detected in this study and are consistent with previous studies on sheep and cattle (Höglund *et al.* 2015; Gebreselassie *et al.* 2020).

Furthermore, 11 overlapping candidate genome regions were detected for F_{ST}, Rsb, xp-EHH, and FLK tests on: a) IR vs IN, b) IR vs AF, and c) IN vs AF sheep breeds (Figure 8). The number of overlapping unique

candidate genes are consistent with the previous results using F_{ST}, Rsb, and iHS tests (Cádiz et al. 2020), F_{ST}, xp-EHH, and iHS tests (Maiorano *et al.* 2018), ROH, F_{ST}, and xp-EHH (Ablondi et al. 2019).

Three of them (SFMBT1, NELL2 and SPO11) are involved in fertility traits. The SFMBT1 on chromosome 19 plays a role during spermatogenesis in rat (Altobelli *et al.* 2013), and cattle (dos Santos Silva et al. 2019). NELL2 on chromosome 3 is involved in pubertal development in swine (Nonneman *et al.* 2016). SPO11 on chromosome 13 is involved in the production of double-strand breaks (DSB) of DNA and it is specifically involved in the growth of the testis, maintenance of the male germ line, and maturation of sperm which previously confirmed on sheep (Mandon-Pépin et al. 2003; Mwacharo et al. 2017) and cattle (Nicolini *et al.* 2018).

Two of the genes (DOCK1 and TNIK) play important roles specially in resistance against diseases. DOCK1 located on chromosome 22 is involved in immune response (Laurin *et al.* 2008; Kunimura *et al.* 2020). TNIK on chromosome 1 plays different functions in embryonic development, especially during the early embryo to blastocyst stages, participates in the regulation of the inflammatory response against infections (Blanco *et al.* 2017).

EML5 on chromosome seven plays a role during neuronal development in the regulation of cytoskeletal rearrangements (O'Connor *et al.* 2004). IQCE on chromosome 24 is involved in body development (Umair *et al.* 2017); TYW1B on chromosome 24 influences artery disease and blood pressure in human (Willer et al. 2013); USH2A on chromosome 12 may be involved in the function of synapses and plays an important role in the development and maintenance of cells in the inner ear and retina (Kim *et al.* 2017).

In total, seven unique candidate genes were detected for IR vs. IN, IR vs. AF, and IN vs. AF comparisons by F_{ST}, Rsb, and FLK analysis, but no overlapping candidate gene was found for the xp-EHH method (Figure 9).

PPA2 on chromosome 6 is associated with immune response and disease resistance in cattle (Brym & Kamiński 2017). KCNIP4 gene on chromosome 6 is directly involved in processes related to muscle growth and fat deposit in sheep (Pasandideh *et al.* 2018) and was reported in cattle involved in bovine growth and calcium metabolism (Smith et al. 2019). SYT1 gene on chromosome 3 is associated with feeding behavior traits (Pattaro *et al.* 2010), and TMEFF2 gene on chromosome 2 is involved in a wide range of traits such as; immune response, milk production and, sperm morphology (Richardson *et al.* 2016; Zhang *et al.* 2020). For the FLK test, PPA2, EML5, MGAT5, and NEB genes were detected, which PPA2 and EML5 have been found in the previous tests (Figure 9). MGAT5 gene on chromosome 2 is associate with dry matter intake in cattle (Seabury *et al.* 2017), and NEB gene on chromosome 2 is involved in environmental adaptation. Among 1262 selected genomic regions reported by Yudin & Larkin (2019), only NEB gene was a shared candidate gene among cattle, sheep, mammoth, polar bear, and whale genomes.

GO classifications of the candidate genes were performed to enable a better understanding of their molecular functions. Based on the GO biological process (BP) for a significant threshold ($p \leq 0.05$), we implemented the GO on 11 overlapping candidate genes. Only six genes (TNIK, DOCK1, SFMBT1, SPO11, USH2A, and TYW1B) associated with the 26 GO terms were identified. In total 11 GOs were related to TNIK and DOCK1, which are associated with local adaptation (resistance against diseases). The first GO (GO:0007010) included TNIK and DOCK1 genes associated with cytoskeleton organization, and six other GOs related to TNIK gene showed different biological functions, include; regulation of dendrite morphogenesis (GO:0048814), actin cytoskeleton reorganization (GO:0031532), protein localization to plasma membrane (GO:0072659), positive regulation of protein phosphorylation (GO:0001934), protein auto phosphorylation (GO:0046777), and intracellular signal transduction (GO:0035556).

In confirmation of our results, Nie *et al.* (2020), reported different GO terms associated with the TNIK gene in human (Nie et al. 2020). Four other GOs associated with DOCK1 include; hematopoietic progenitor cell

differentiation (GO:0002244), small GTPase mediated signal transduction (GO:0007264), cell migration (GO:0016477), and positive regulation of GTPase activity (GO:0043547). DOCK family genes play different functions which confirm different biological function of this gene (Laurin *et al.* 2008; Kunimura *et al.* 2020).

Absolute correlation among the F_{ST}, FLK, xp-EHH, and Rsb tests were calculated (Figure 10). The xp-EHH, and Rsb are based on the frequency of extended haplotypes between two populations (Sabeti *et al.* 2006; Tang *et al.* 2007), whereas F_{ST} and FLK are based on allele frequencies (Reynolds *et al.* 1983; Hohenlohe *et al.* 2010). So as expected, maximum correlations were observed between F_{ST} and FLK, as well as between xp-EHH and Rsb. On the other hand, we detected minimum correlations between haplotype based tests (xp-EHH, and Rsb) and allele based tests (F_{ST} and FLK). These findings are consistent with the previous reports (Ma *et al.* 2015; Weigand & Leese 2018).

Conclusions

Our results showed the population structure and selective candidate genomic regions of the 14 indigenous sheep breeds from Middle East and South Asia. This information would be valuable in future study on genetic basis for local adaptation of indigenous breeds. In $F_{ST,}$ FLK, xp-EHH, and Rsb complementary statistical tests, some candidate genomic regions under selective pressure were detected in indigenous sheep breeds and these candidate genomic regions may facilitate identification of the underlying genes and possible exploitation in future sheep breeding.

Data Availability

Genotype data from the sheep breeds (Afshari, Moghani, Qezel, Bangladeshi Garole, Bangladesh East , Indian Garole, Changthangi, and Deccani) are available through the Sheep HapMap project (Sempéré *et al.* 2015). The ZEL, Lori-Bakhtiari, Iranian Balochi, Arabi, Afghan Balochi, and Gadik breeds data are part of

the Iranian national genetic evaluations of economic traits conducted at the Animal Breeding Center of Iran.

Any request for data should be addressed to the corresponding author.

References

1. Ablondi M., Viklund Å., Lindgren G., Eriksson S. & Mikko S. (2019) Signatures of selection in the genome of Swedish warmblood horses selected for sport performance. BMC genomics 20, 717.
2. Alberto F.J., Boyer F., Orozco-terWengel P., Streeter I., Servin B., De Villemereuil P., Benjelloun B., Librado P., Biscarini F. & Colli L. (2018) Convergent genomic signatures of domestication in sheep and goats. Nature Communications 9, 1-9.
3. Alexander D.H., Novembre J. & Lange K. (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome research 19, 1655-64.
4. Altobelli G., Bogdarina I.G., Stupka E., Clark A.J. & Langley-Evans S. (2013) Genome-wide methylation and gene expression changes in newborn rats following maternal protein restriction and reversal by folic acid. PloS one 8, e82989.
5. Ardlie K.G., Kruglyak L. & Seielstad M. (2002) Patterns of linkage disequilibrium in the human genome. Nature Reviews Genetics 3, 299-309.
6. Armstrong E., Ciappesoni G., Iriarte W., Da Silva C., Macedo F., Navajas E., Brito G., San Julián R., Gimeno D. & Postiglioni A. (2018) Novel genetic polymorphisms associated with carcass traits in grazing Texel sheep. Meat Science 145, 202-8.
7. Atlija M., Arranz J.-J., Martinez-Valladares M. & Gutiérrez-Gil B. (2016) Detection and replication of QTL underlying resistance to gastrointestinal nematodes in adult sheep using the ovine 50K SNP array. Genetics Selection Evolution 48, 4.
8. Barbato M., Hailer F., Orozco-terWengel P., Kijas J., Mereu P., Cabras P., Mazza R., Pirastru M. & Bruford M.W. (2017) Genomic signatures of adaptive introgression from European mouflon into domestic sheep. Scientific Reports 7, 1-13.
9. Benavides M.V., Sonstegard T.S., Kemp S., Mugambi J.M., Gibson J.P., Baker R.L., Hanotte O., Marshall K. & Van Tassell C. (2015) Identification of novel loci associated with gastrointestinal parasite resistance in a Red Maasai x Dorper backcross population. PloS one 10, e0122797.
10. Berihulay H., Abied A., He X., Jiang L. & Ma Y. (2019) Adaptation mechanisms of small ruminants to environmental heat stress. Animals 9, 75.
11. Bertolini F., Servin B., Talenti A., Rochat E., Kim E.S., Oget C., Palhière I., Crisà A., Catillo G. & Steri R. (2018) Signatures of selection and environmental adaptation across the goat genome post-domestication. Genetics Selection Evolution 50, 1-24.
12. Blanco F.C., Soria M.A., Klepp L.I. & Bigi F. (2017) ERAP1 and PDE8A are downregulated in cattle protected against bovine tuberculosis. Journal of molecular microbiology and biotechnology 27, 237-45.
13. Bonhomme M., Chevalet C., Servin B., Boitard S., Abdallah J., Blott S. & SanCristobal M. (2010) Detecting selection in population trees: the Lewontin and Krakauer test extended. Genetics 186, 241-62.
14. Browning S.R. & Browning B.L. (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. The American Journal of Human Genetics 81, 1084-97.
15. Browning S.R. & Weir B.S. (2010) Population structure with localized haplotype clusters. Genetics 185, 1337-44.
16. Brym P. & Kamiński S. (2017) Microarray analysis of differential gene expression profiles in blood cells of naturally BLV-infected and uninfected Holstein–Friesian cows. Molecular biology reports 44, 109-27.

17. Cádiz M.I., López M.E., Díaz-Domínguez D., Cáceres G., Yoshida G.M., Gomez-Uchida D. & Yáñez J.M. (2020) Whole genome re-sequencing reveals recent signatures of selection in three strains of farmed Nile tilapia (Oreochromis niloticus). Scientific Reports 10, 1-14.
18. Calvo J., Marcos S., Jurado J. & Serrano M. (2004) Association of the heart fatty acid-binding protein (FABP3) gene with milk traits in Manchega breed sheep. Animal Genetics 35, 347-9.
19. Cassandri M., Smirnov A., Novelli F., Pitolli C., Agostini M., Malewicz M., Melino G. & Raschellà G. (2017) Zinc-finger proteins in health and disease. Cell death discovery 3, 1-12.
20. Ceballos F.C., Joshi P.K., Clark D.W., Ramsay M. & Wilson J.F. (2018) Runs of homozygosity: windows into population history and trait architecture. Nature Reviews Genetics 19, 220.
21. Chang L.-Y., Toghiani S., Ling A., Aggrey S.E. & Rekaya R. (2018) High density marker panels, SNPs prioritizing and accuracy of genomic selection. BMC genetics 19, 4.
22. Chen Z., Yao Y., Ma P., Wang Q. & Pan Y. (2018) Haplotype-based genome-wide association study identifies loci and candidate genes for milk yield in Holsteins. PloS one 13, e0192695.
23. Ciani E., Lasagna E., D'andrea M., Alloggio I., Marroni F., Ceccobelli S., Bermejo J.V.D., Sarti F.M., Kijas J. & Lenstra J.A. (2015) Merino and Merino-derived sheep breeds: a genome-wide intercontinental study. Genetics Selection Evolution 47, 64.
24. Consortium B.H. (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324, 528-32.
25. Cook D.E. & Andersen E.C. (2017) VCF-kit: assorted utilities for the variant call format. Bioinformatics 33, 1581-2.
26. 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. (2011) The variant call format and VCFtools. Bioinformatics 27, 2156-8.
27. de Simoni Gouveia J.J., Paiva S.R., McManus C.M., Caetano A.R., Kijas J.W., Facó O., Azevedo H.C., de Araujo A.M., de Souza C.J.H. & Yamagishi M.E.B. (2017) Genome-wide search for signatures of selection in three major Brazilian locally adapted sheep breeds. Livestock science 197, 36-45.
28. Deniskova T., Dotsev A., Lushihina E., Shakhin A., Kunz E., Medugorac I., Reyer H., Wimmers K., Zhunushev A. & Brem G. (2019) Population structure and genetic diversity of sheep breeds in the Kyrgyz Republic. Frontiers in genetics 10, 1311.
29. dos Santos Silva D.B., Fonseca L.F.S., Pinheiro D.G., Muniz M.M.M., Magalhães A.F.B., Baldi F., Ferro J.A., Chardulo L.A.L. & De Albuquerque L.G. (2019) Prediction of hub genes associated with intramuscular fat content in Nelore cattle. BMC genomics 20, 520.
30. Eydivandi S., Sahana G., Momen M., Moradi M. & Schönherz A. (2020) Genetic diversity in Iranian indigenous sheep vis-à-vis selected exogenous sheep breeds and wild mouflon. Animal Genetics 51, 772-87.
31. Faostat F. (2017) Available online: http://www. fao. org/faostat/en/# data. QC (accessed on January 2018).
32. FAOSTAT F. (2018) Crop statistics.
33. Fariello M.-I., Servin B., Tosser-Klopp G., Rupp R., Moreno C., San Cristobal M., Boitard S. & Consortium I.S.G. (2014) Selection signatures in worldwide sheep populations. PloS one 9, e103813.
34. Fariello M.I., Boitard S., Naya H., SanCristobal M. & Servin B. (2012) Using haplotype differentiation among hierarchically structured populations for the detection of selection signatures. arXiv preprint arXiv:1210.7583.
35. Fariello M.I., Boitard S., Naya H., SanCristobal M. & Servin B. (2013) Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. Genetics 193, 929-41.
36. Feng X., Jiang J., Padhi A., Ning C., Fu J., Wang A., Mrode R. & Liu J.-F. (2017) Characterization of genome-wide segmental duplications reveals a common genomic feature of association with immunity among domestic animals. BMC genomics 18, 293.
37. Gautier M. & Vitalis R. (2012) rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics 28, 1176-7.

38. Gebreselassie G., Berihulay H., Jiang L. & Ma Y. (2020) Review on Genomic Regions and Candidate Genes Associated with Economically Important Production and Reproduction Traits in Sheep (Ovies aries). Animals 10, 33.
39. Groeneveld L., Lenstra J., Eding H., Toro M., Scherf B., Pilling D., Negrini R., Finlay E., Jianlin H. & Groeneveld E. (2010) Genetic diversity in farm animals–a review. Animal Genetics 41, 6-31.
40. Höglund J.K., Guldbrandtsen B., Lund M.S. & Sahana G. (2015) Identification of genomic regions associated with female fertility in Danish Jersey using whole genome sequence data. BMC genetics 16, 60.
41. Hohenlohe P.A., Phillips P.C. & Cresko W.A. (2010) Using population genomics to detect selection in natural populations: key concepts and methodological considerations. International journal of plant sciences 171, 1059-71.
42. Johnston S.E., Bérénos C., Slate J. & Pemberton J.M. (2015) A genomic region containing RNF212 and CPLX1 is associated with sexually-dimorphic recombination rate variation in Soay sheep (Ovis aries). BioRxiv, 024869.
43. Kijas J.W., Lenstra J.A., Hayes B., Boitard S., Neto L.R.P., San Cristobal M., Servin B., McCulloch R., Whan V. & Gietzen K. (2012) Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS biology 10.
44. Kim E.-S., Elbeltagy A., Aboul-Naga A., Rischkowsky B., Sayre B., Mwacharo J.M. & Rothschild M.F. (2016) Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. Heredity 116, 255-64.
45. Kim S.-J., Ka S., Ha J.-W., Kim J., Yoo D., Kim K., Lee H.-K., Lim D., Cho S. & Hanotte O. (2017) Cattle genome-wide analysis reveals genetic signatures in trypanotolerant N'Dama. BMC genomics 18, 371.
46. Kullo I.J. & Ding K. (2007) Patterns of population differentiation of candidate genes for cardiovascular disease. BMC genetics 8, 48.
47. Kunimura K., Uruno T. & Fukui Y. (2020) DOCK family proteins: key players in immune surveillance mechanisms. International immunology 32, 5-15.
48. Laurin M., Fradet N., Blangy A., Hall A., Vuori K. & Côté J.-F. (2008) The atypical Rac activator Dock180 (Dock1) regulates myoblast fusion in vivo. Proceedings of the National Academy of Sciences 105, 15446-51.
49. Li H., Wu X.L., Tait Jr R., Bauck S., Thomas D., Murphy T. & Rosa G. (2020) Genome-wide association study of milk production traits in a crossbred dairy sheep population using three statistical models. Animal Genetics.
50. Ma Y., Ding X., Qanbari S., Weigend S., Zhang Q. & Simianer H. (2015) Properties of different selection signature statistics and a new strategy for combining them. Heredity 115, 426-36.
51. Ma Y., Zhang H., Zhang Q. & Ding X. (2014) Identification of selection footprints on the X chromosome in pig. PloS one 9, e94911.
52. Maiorano A.M., Lourenco D.L., Tsuruta S., Ospina A.M.T., Stafuzza N.B., Masuda Y., Filho A.E.V., Cyrillo J.N.d.S.G., Curi R.A. & Silva J.A.I.d.V. (2018) Assessing genetic architecture and signatures of selection of dual purpose Gir cattle populations using genomic information. PloS one 13, e0200694.
53. Mandon-Pépin B., Oustry-Vaiman A., Vigier B., Piumi F., Cribiu E. & Cotinot C. (2003) Expression profiles and chromosomal localization of genes controlling meiosis and follicular development in the sheep ovary. Biology of reproduction 68, 985-95.
54. Marete A.G., Guldbrandtsen B., Lund M.S., Fritz S., Sahana G. & Boichard D. (2018) A meta-analysis including pre-selected sequence variants associated with seven traits in three French dairy cattle populations. Frontiers in genetics 9, 522.
55. McRae K.M., McEwan J.C., Dodds K.G. & Gemmell N.J. (2014) Signatures of selection in sheep bred for resistance or susceptibility to gastrointestinal nematodes. BMC genomics 15, 637.
56. Milanesi M., Capomaccio S., Vajana E., Bomba L., Garcia J.F., Ajmone-Marsan P. & Colli L. (2017) BITE: an R package for biodiversity analyses. bioRxiv, 181610.

57. Mwacharo J.M., Kim E.-S., Elbeltagy A.R., Aboul-Naga A.M., Rischkowsky B.A. & Rothschild M.F. (2017) Genomic footprints of dryland stress adaptation in Egyptian fat-tail sheep and their divergence from East African and western Asia cohorts. Scientific Reports 7, 1-10.
58. Mwai O., Hanotte O., Kwon Y.-J. & Cho S. (2015) African indigenous cattle: unique genetic resources in a rapidly changing world. Asian-Australasian journal of animal sciences 28, 911.
59. Nafikov R., Schoonmaker J., Korn K., Noack K., Garrick D., Koehler K., Minick-Bormann J., Reecy J., Spurlock D. & Beitz D. (2013) Association of polymorphisms in solute carrier family 27, isoform A6 (SLC27A6) and fatty acid-binding protein-3 and fatty acid-binding protein-4 (FABP3 and FABP4) with fatty acid composition of bovine milk. Journal of dairy science 96, 6007-21.
60. Nicolazzi E.L., Picciolini M., Strozzi F., Schnabel R.D., Lawley C., Pirani A., Brew F. & Stella A. (2014) SNPchiMp: a database to disentangle the SNPchip jungle in bovine livestock. BMC genomics 15, 123.
61. Nicolini P., Amorín R., Han Y. & Peñagaricano F. (2018) Whole-genome scan reveals significant non-additive effects for sire conception rate in Holstein cattle. BMC genetics 19, 1-8.
62. Nie F.-y., Zhang M.-r., Shang S.-s., Zhang Q.-x., Zhang R., Chen P. & Ma J. (2020) Methylome-wide association study of first-episode schizophrenia reveals a hypermethylated CpG site in the promoter region of the TNIK susceptibility gene. Progress in Neuro-Psychopharmacology and Biological Psychiatry, 110081.
63. Nonneman D.J., Schneider J.F., Lents C.A., Wiedmann R.T., Vallet J.L. & Rohrer G.A. (2016) Genome-wide association and identification of candidate genes for age at puberty in swine. BMC genetics 17, 50.
64. O'Connor V., Houtman S.H., De Zeeuw C., Bliss T. & French P. (2004) Eml5, a novel WD40 domain protein expressed in rat brain. Gene 336, 127-37.
65. Oleksyk T.K., Smith M.W. & O'Brien S.J. (2010) Genome-wide scans for footprints of natural selection. Philosophical Transactions of the Royal Society B: Biological Sciences 365, 185-205.
66. Pasandideh M., Rahimi-Mianji G. & Gholizadeh M. (2018) A genome scan for quantitative trait loci affecting average daily gain and Kleiber ratio in Baluchi Sheep. Journal of genetics 97, 493-503.
67. Pattaro C., De Grandi A., Vitart V., Hayward C., Franke A., Aulchenko Y.S., Johansson A., Wild S.H., Melville S.A. & Isaacs A. (2010) A meta-analysis of genome-wide data from five European isolates reveals an association of COL22A1, SYT1, and GABRR2with serum creatinine level. BMC medical genetics 11, 41.
68. Pickrell J.K., Coop G., Novembre J., Kudaravalli S., Li J.Z., Absher D., Srinivasan B.S., Barsh G.S., Myers R.M. & Feldman M.W. (2009) Signals of recent positive selection in a worldwide sample of human populations. Genome research 19, 826-37.
69. Price A.L., Patterson N.J., Plenge R.M., Weinblatt M.E., Shadick N.A. & Reich D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics 38, 904-9.
70. 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. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics 81, 559-75.
71. Purfield D.C., McParland S., Wall E. & Berry D.P. (2017) The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. PloS one 12, e0176780.
72. Rambaut A. (2015) FigTree–Molecular Evolution, phylogenetics and Epidemiology (2007). Available:. Accessed 14.
73. Randhawa I.A.S., Khatkar M.S., Thomson P.C. & Raadsma H.W. (2014) Composite selection signals can localize the trait specific genomic regions in multi-breed populations of cattle and sheep. BMC genetics 15, 34.
74. Reynolds J., Weir B.S. & Cockerham C.C. (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics 105, 767-79.
75. Richardson I.W., Berry D.P., Wiencko H.L., Higgins I.M., More S.J., McClure J., Lynn D.J. & Bradley D.G. (2016) A genome-wide association study for genetic susceptibility to Mycobacterium bovis infection in dairy cattle identifies a susceptibility QTL on chromosome 23. Genetics Selection Evolution 48, 19.
76. Roshyara N.R. (2014) Documentation for Genotype Format Converter: fcGENE.

77. Sabeti P.C., Schaffner S.F., Fry B., Lohmueller J., Varilly P., Shamovsky O., Palma A., Mikkelsen T., Altshuler D. & Lander E. (2006) Positive natural selection in the human lineage. Science 312, 1614-20.
78. Sabeti P.C., Varilly P., Fry B., Lohmueller J., Hostetter E., Cotsapas C., Xie X., Byrne E.H., McCarroll S.A. & Gaudet R. (2007) Genome-wide detection and characterization of positive selection in human populations. Nature 449, 913-8.
79. Scarpa R., Ruto E.S., Kristjanson P., Radeny M., Drucker A.G. & Rege J.E. (2003) Valuing indigenous cattle breeds in Kenya: an empirical comparison of stated and revealed preference value estimates. Ecological Economics 45, 409-26.
80. Seabury C.M., Oldeschulte D.L., Saatchi M., Beever J.E., Decker J.E., Halley Y.A., Bhattarai E.K., Molaei M., Freetly H.C. & Hansen S.L. (2017) Genome-wide association study for feed efficiency and growth traits in US beef cattle. BMC genomics 18, 386.
81. Sempéré G., Moazami-Goudarzi K., Eggen A., Laloë D., Gautier M. & Flori L. (2015) WIDDE: a Web-Interfaced next generation database for genetic diversity exploration, with a first application in cattle. BMC genomics 16, 1-8.
82. Sharma A., Kumar B.S., Dash S., Singh S. & Verma R. (2020) Heat shock protein B1 expression is associated with age at sexual maturity in Rhode Island Red and Punjab Red layers under heat stress. International Journal of Biometeorology, 1-11.
83. Smith J.L., Wilson M.L., Nilson S.M., Rowan T.N., Oldeschulte D.L., Schnabel R.D., Decker J.E. & Seabury C.M. (2019) Genome-wide association and genotype by environment interactions for growth traits in US Gelbvieh cattle. BMC genomics 20, 926.
84. Srikanth K., Kwon A., Lee E. & Chung H. (2017) Characterization of genes and pathways that respond to heat stress in Holstein calves through transcriptome analysis. Cell Stress and Chaperones 22, 29-42.
85. Tang K., Thornton K.R. & Stoneking M. (2007) A new approach for using genome scans to detect recent positive selection in the human genome. PLoS biology 5.
86. Umair M., Shah K., Alhaddad B., Haack T.B., Graf E., Strom T.M., Meitinger T. & Ahmad W. (2017) Exome sequencing revealed a splice site variant in the IQCE gene underlying post-axial polydactyly type A restricted to lower limb. European Journal of Human Genetics 25, 960-5.
87. Vatsiou A.I., Bazin E. & Gaggiotti O.E. (2016) Detection of selective sweeps in structured populations: a comparison of recent methods. Molecular ecology 25, 89-103.
88. Weigand H. & Leese F. (2018) Detecting signatures of positive selection in non-model species using genomic data. Zoological Journal of the Linnean Society 184, 528-83.
89. Willer C.J., Schmidt E.M., Sengupta S., Peloso G.M., Gustafsson S., Kanoni S., Ganna A., Chen J., Buchkovich M.L. & Mora S. (2013) Discovery and refinement of loci associated with lipid levels. Nature genetics 45, 1274.
90. Wright S. (1949) The genetical structure of populations. Annals of eugenics 15, 323-54.
91. Yuan Z., Liu E., Liu Z., Kijas J., Zhu C., Hu S., Ma X., Zhang L., Du L. & Wang H. (2017) Selection signature analysis reveals genes associated with tail type in Chinese indigenous sheep. Animal Genetics 48, 55-66.
92. Yudin N. & Larkin D. (2019) Shared signatures of selection related to adaptation and acclimation in local cattle and sheep breeds from Russia. Russian Journal of Genetics 55, 1008-14.
93. Yurchenko A.A., Deniskova T.E., Yudin N.S., Dotsev A.V., Khamiruev T.N., Selionova M.I., Egorov S.V., Reyer H., Wimmers K. & Brem G. (2019) High-density genotyping reveals signatures of selection related to acclimation and economically important traits in 15 local sheep breeds from Russia. BMC genomics 20, 294.
94. Zerbino D.R., Achuthan P., Akanni W., Amode M.R., Barrell D., Bhai J., Billis K., Cummins C., Gall A. & Girón C.G.J.N.a.r. (2017) Ensembl 2018. 46, D754-D61.
95. Zhang C., Dong S.-S., Xu J.-Y., He W.-M. & Yang T.-L. (2019) PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics 35, 1786-8.

96. Zhang F., Wang Y., Mukiibi R., Chen L., Vinsky M., Plastow G., Basarab J., Stothard P. & Li C. (2020) Genetic architecture of quantitative traits in beef cattle revealed by genome wide association studies of imputed whole genome sequence variants: I: feed efficiency and component traits. BMC genomics 21, 36.
97. Zhang L., Liu J., Zhao F., Ren H., Xu L., Lu J., Zhang S., Zhang X., Wei C. & Lu G. (2013) Genome-wide association studies for growth and meat production traits in sheep. PloS one 8.
98. ZHAO F.-p., WEI C.-h., ZHANG L., LIU J.-s., WANG G.-k., Tao Z. & DU L.-x. (2016) A genome scan of recent positive selection signatures in three sheep populations. Journal of Integrative Agriculture 15, 162-74.

Table 1. Breed names and the corresponding code used throughout the manuscript, the country of origin, sample size, and data source.

Breed	Acronym	Geographic Origin	Category	Sample Size	Data Source
Afshari	AFS	Iran	IR[1]	37	HapMap
Moghani	MOG	Iran	IR	34	HapMap
Qezel	QEZ	Iran	IR	35	HapMap
Zel	ZEL	Iran	IR	44	Unpublished data
Lori-Bakhtiari	LOR	Iran	IR	46	Unpublished data
Iranian Balochi	IBL	Iran	AF[2]	87	Unpublished data
Arabi	ARB	Afghanistan	AF	14	Unpublished data
Afghan Balochi	BLO	Afghanistan	AF	15	Unpublished data
Gadik	GDK	Afghanistan	AF	14	Unpublished data
Bangladeshi Garole	BGA	Bangladesh	IN[3]	24	HapMap
Bangladesh East	BGE	Bangladesh	IN	24	HapMap
Changthangi	CHA	India	IN	29	HapMap
Indian Garole	GAR	India	IN	26	HapMap
Deccani	IDC	India	IN	24	HapMap

1: Contain Iranian sheep breeds exception Iranian Balouchi.

2: Contain Afghan sheep breeds and Iranian Balouchi.

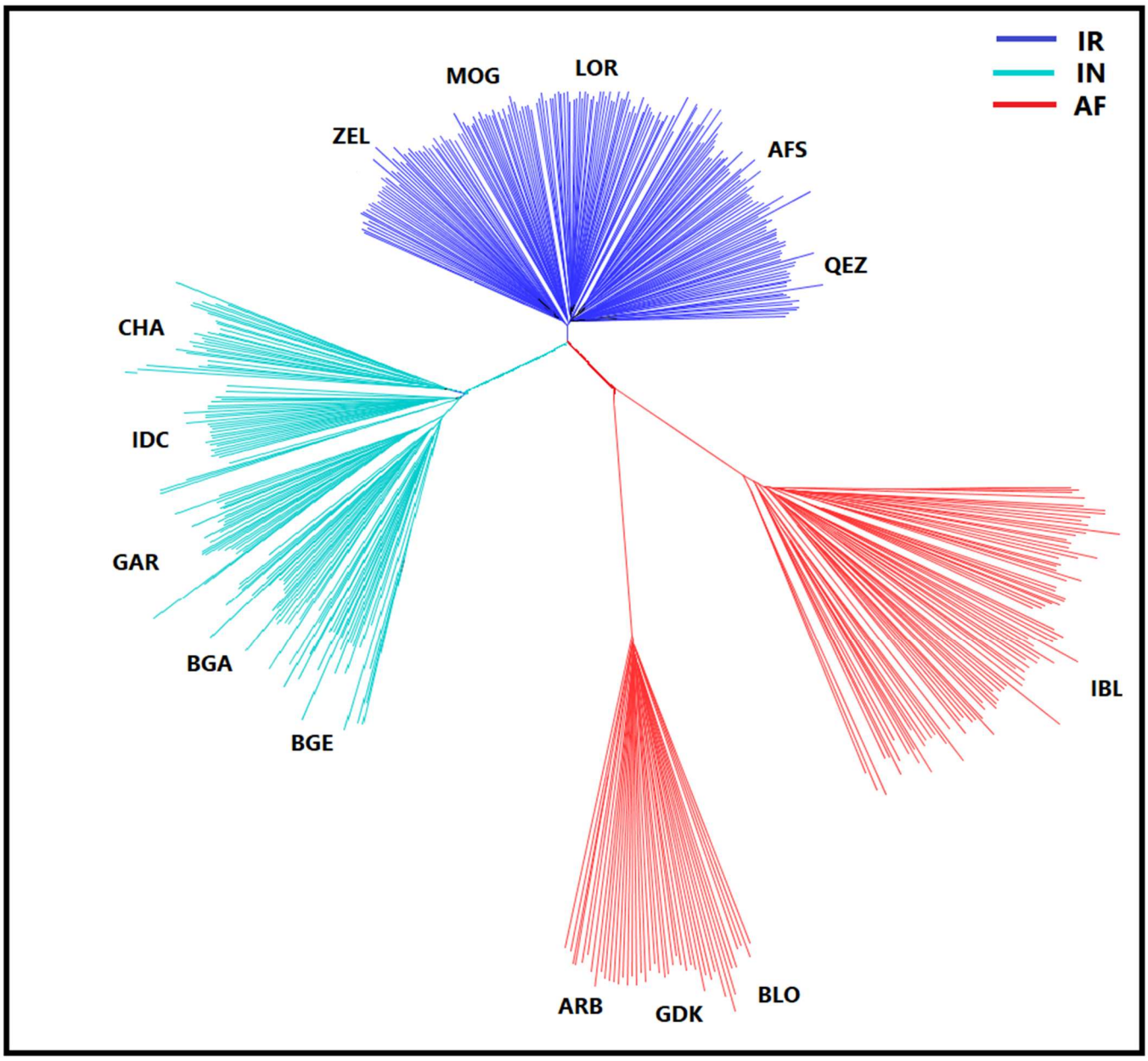
3: Contain Indian and Bengal sheep breeds.

Table 2. Breed biological process of common candidate genes under selective pressure for F_{ST}, Rsb, xp-EHH, and FLK tests on: a) IR Vs IN, b) IR Vs AF, c) IN Vs AF.

Group	Biological Process		Genes
	cytoskeleton organization	GO:0007010	DOCK1,TNIK
	regulation of dendrite morphogenesis	GO:0048814	TNIK
	actin cytoskeleton reorganization	GO:0031532	TNIK
	hematopoietic progenitor cell differentiation	GO:0002244	DOCK1
	small GTPase mediated signal transduction	GO:0007264	DOCK1
	protein localization to plasma membrane	GO:0072659	TNIK
	cell migration	GO:0016477	DOCK1
	positive regulation of protein phosphorylation	GO:0001934	TNIK
	protein auto phosphorylation	GO:0046777	TNIK
	positive regulation of GTPase activity	GO:0043547	DOCK1
	intracellular signal transduction	GO:0035556	TNIK
a	negative regulation of transcription	GO:0045892	SFMBT1
	reciprocal meiotic recombination	GO:0007131	SPO11
	sensory perception of light stimulus	GO:0050953	USH2A
	synaptonemal complex assembly	GO:0007130	SPO11
	male meiosis I	GO:0007141	SPO11
	DNA metabolic process	GO:0006259	SPO11
	ovarian follicle development	GO:0001541	SPO11
	oogenesis	GO:0048477	SPO11
	synapsis	GO:0007129	SPO11
	photoreceptor cell maintenance	GO:0045494	USH2A
	establishment of protein localization	GO:0045184	USH2A
	tRNA processing	GO:0008033	TYW1B
	spermatid development	GO:0007286	SPO11
b	sensory perception of sound	GO:0007605	USH2A
c	tRNA processing	GO:0008033	TYW1B

Figures:

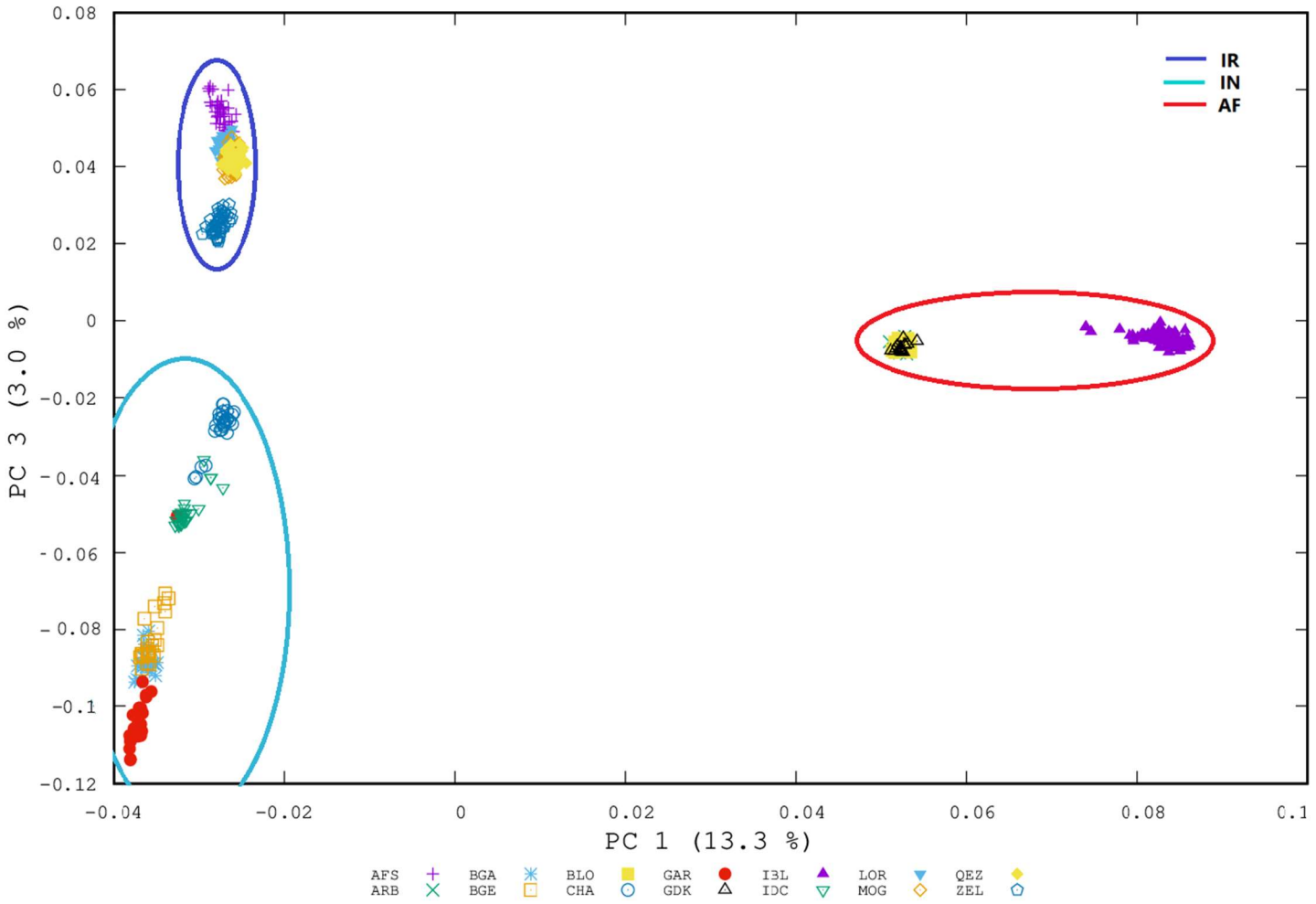
Figure 1. Neighbor-joining phylogenetic tree for 14 sheep breeds based on autosomal SNPs. For breed abbreviations, see Table 1.



811

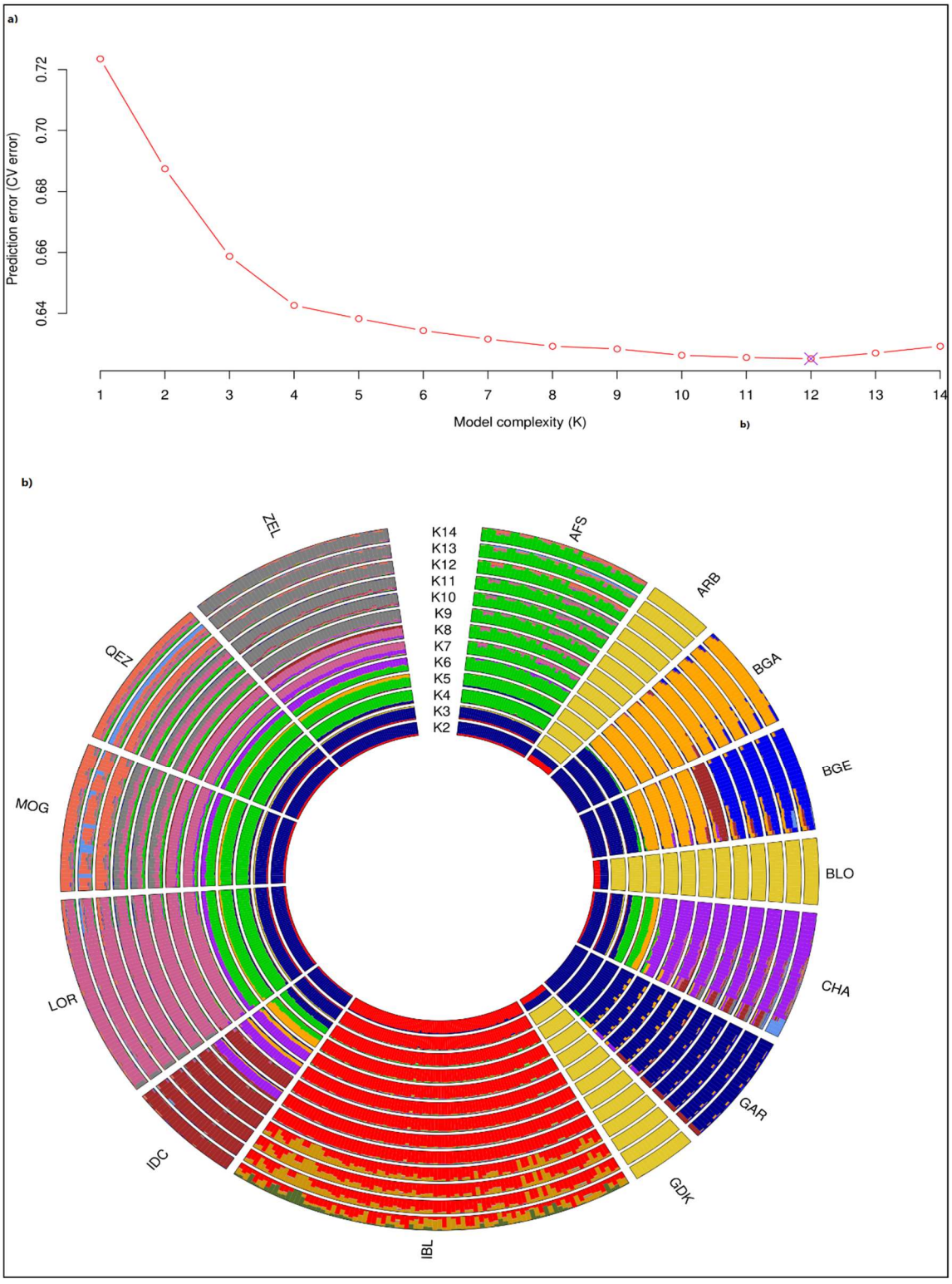
812

Figure 2. Principal components analysis (PC 1 and PC 3) of among 14 sheep breeds based on autosomes. For breed abbreviations, see Table 1.



821 **Figure 3.** Prediction error (a) and circle Admixture from K2 to K14 (b) plotted, respectively. For breed

822 abbreviations, see Table 1.



823

Figure 4. The distribution of absolute Z(F_{ST}) values on 26 sheep autosomes: a) IR and IN breeds (The horizontal blue line, Z(F_{ST}) ≥ 3.93), b) IR and AF breeds (The horizontal blue line, Z(F_{ST}) ≥ 3.18), c) IN and AF breeds (The horizontal blue line, Z(FST) ≥ 3.08). The data points above the horizontal line (blue line) are top 1% Z(FST) values. FST: Fixation index. For breed abbreviations, see Table 1.

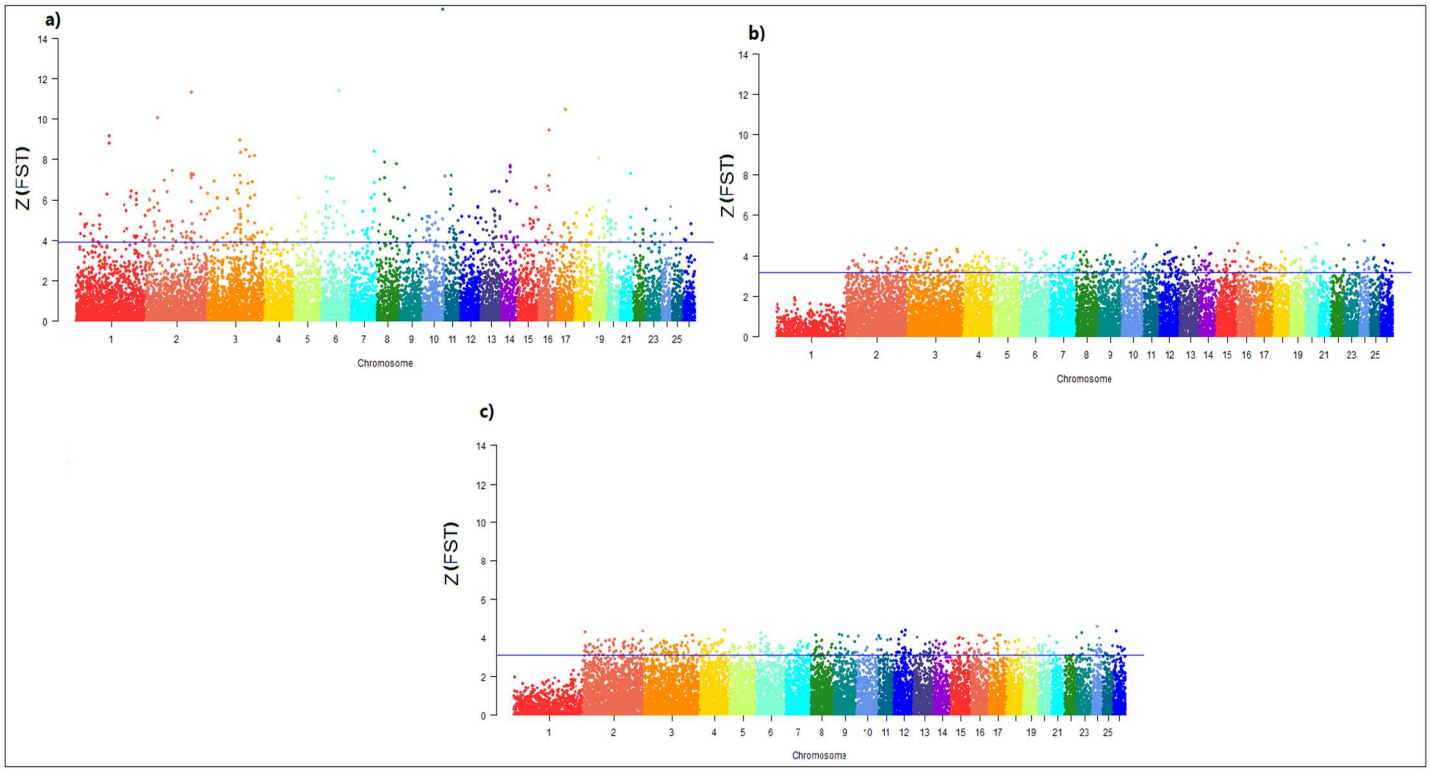


Figure 5. Genomic distribution of standardized cross-population extended haplotype homozygosity (xp-EHH) scores on 26 sheep autosomes pairwise: a) IR and IN breeds, b) IR and AF breeds, c) IN and AF breeds. For breed abbreviations, see Table 1.

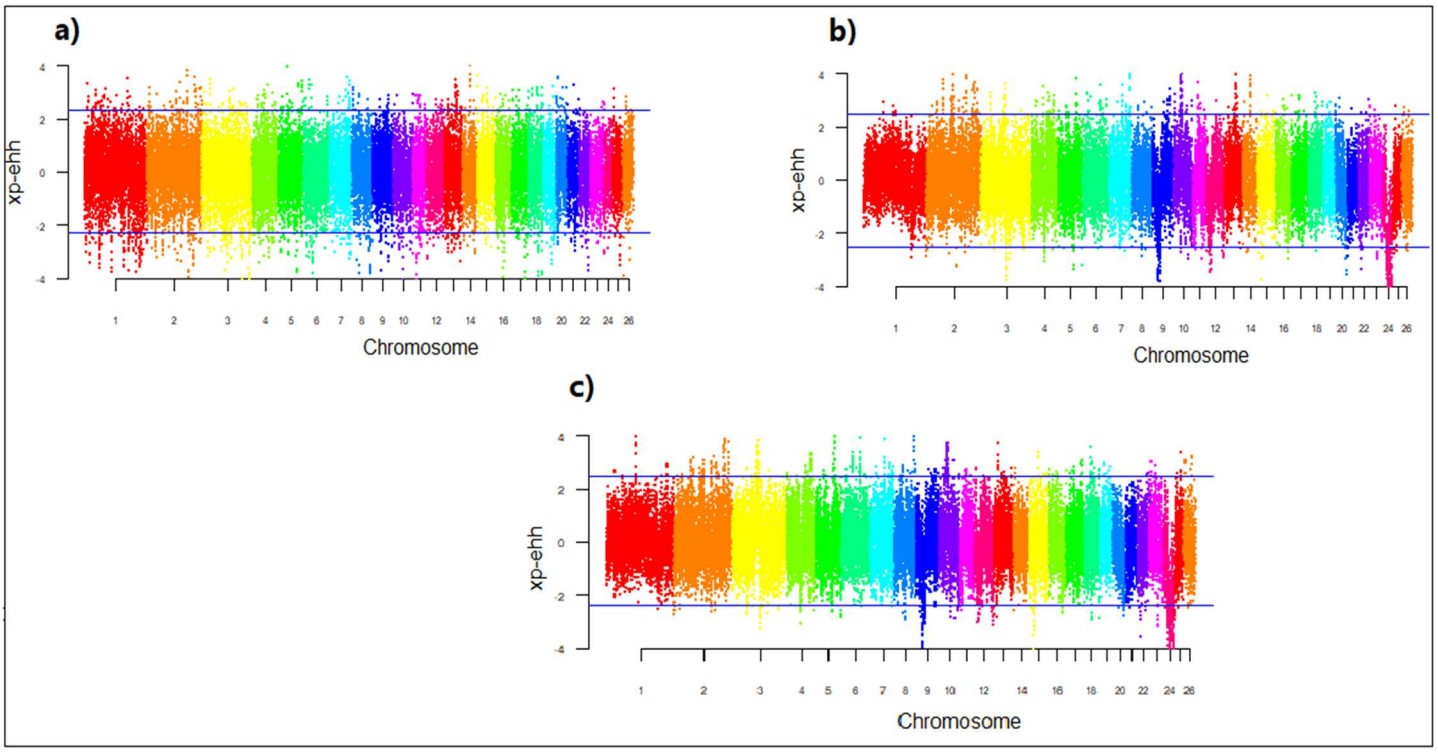


Figure 6. Genomic distribution of standardized haplotype differentiation (Rsb) scores on 26 sheep autosomes pairwise: a) IR and IN breeds, b) IR and AF breeds, c) IN and AF breeds. For breed abbreviations, see Table 1.

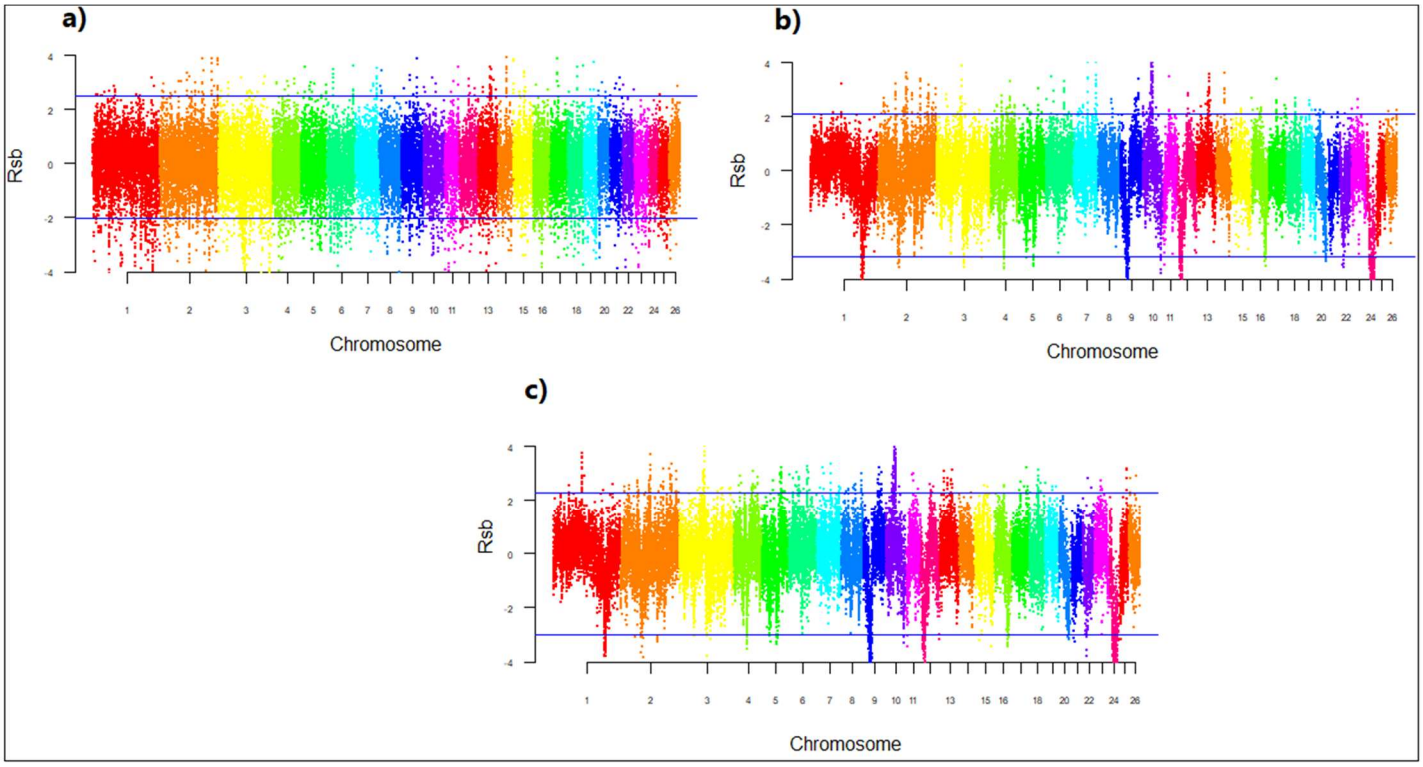


Figure 7. Genomic distribution of single marker statistic (FLK) scores on 26 sheep autosomes pairwise: a) IR and IN breeds, b) IR and AF breeds, c) IN and AF breeds. For breed abbreviations, see Table 1.

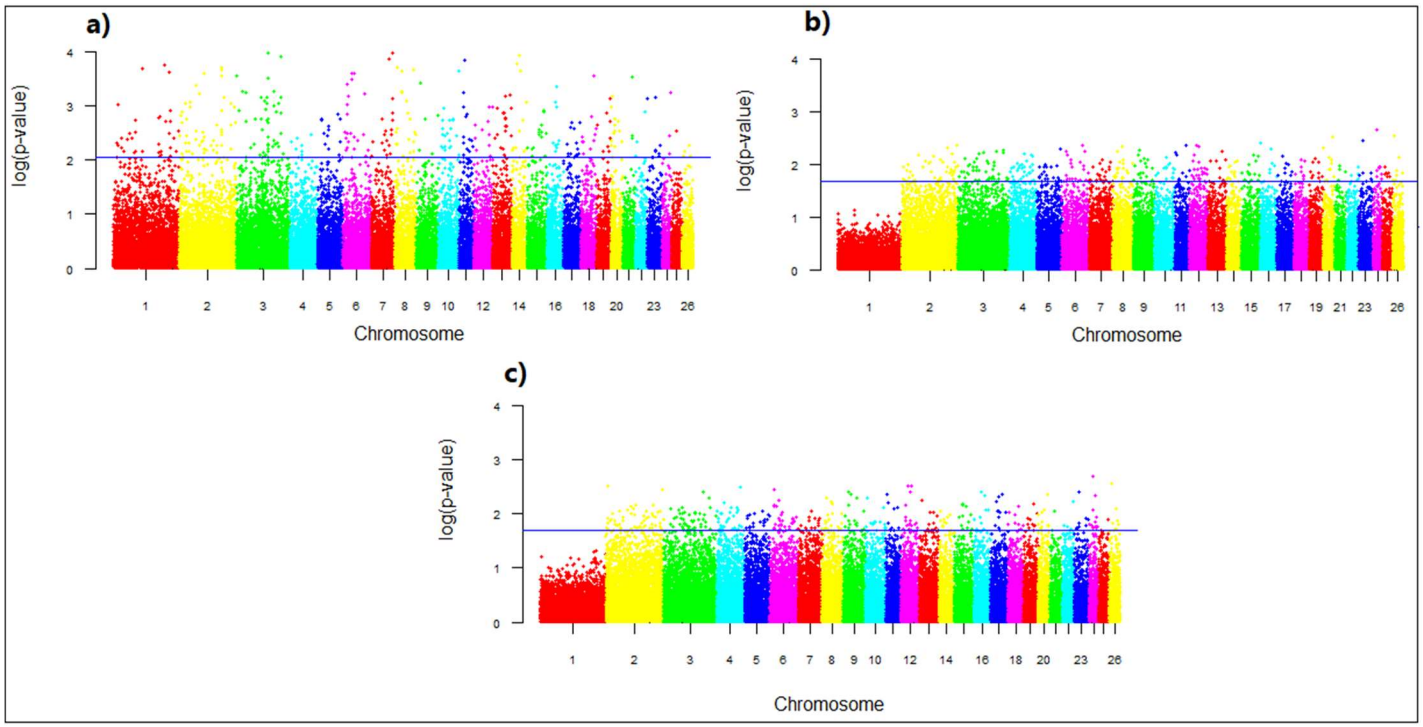


Figure 8. Venn diagram showing the unique and shared candidate genes for FST, Rsb, xp-EHH, and FLK tests on: a) IR Vs IN, b) IR Vs AF, and c) IN Vs AF sheep breeds. For breed abbreviations, see Table 1.

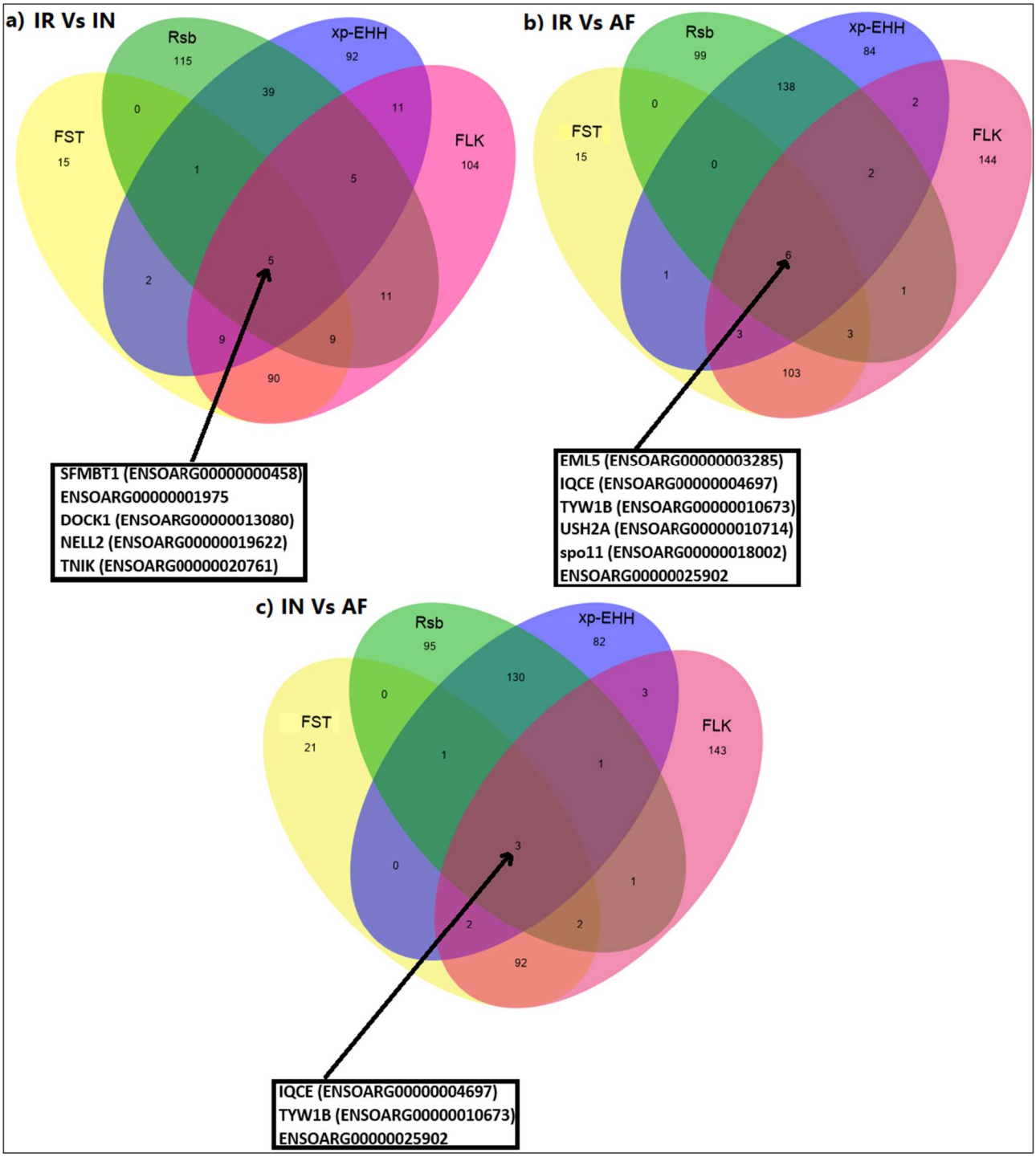


Figure 9. Venn diagram showing the unique and shared candidate genes for IR Vs IN, IR Vs AF, and IN Vs AF data on: a) FST, b) xp-EHH, c) Rsb, and d) FLK tests. For breed abbreviations, see Table 1.

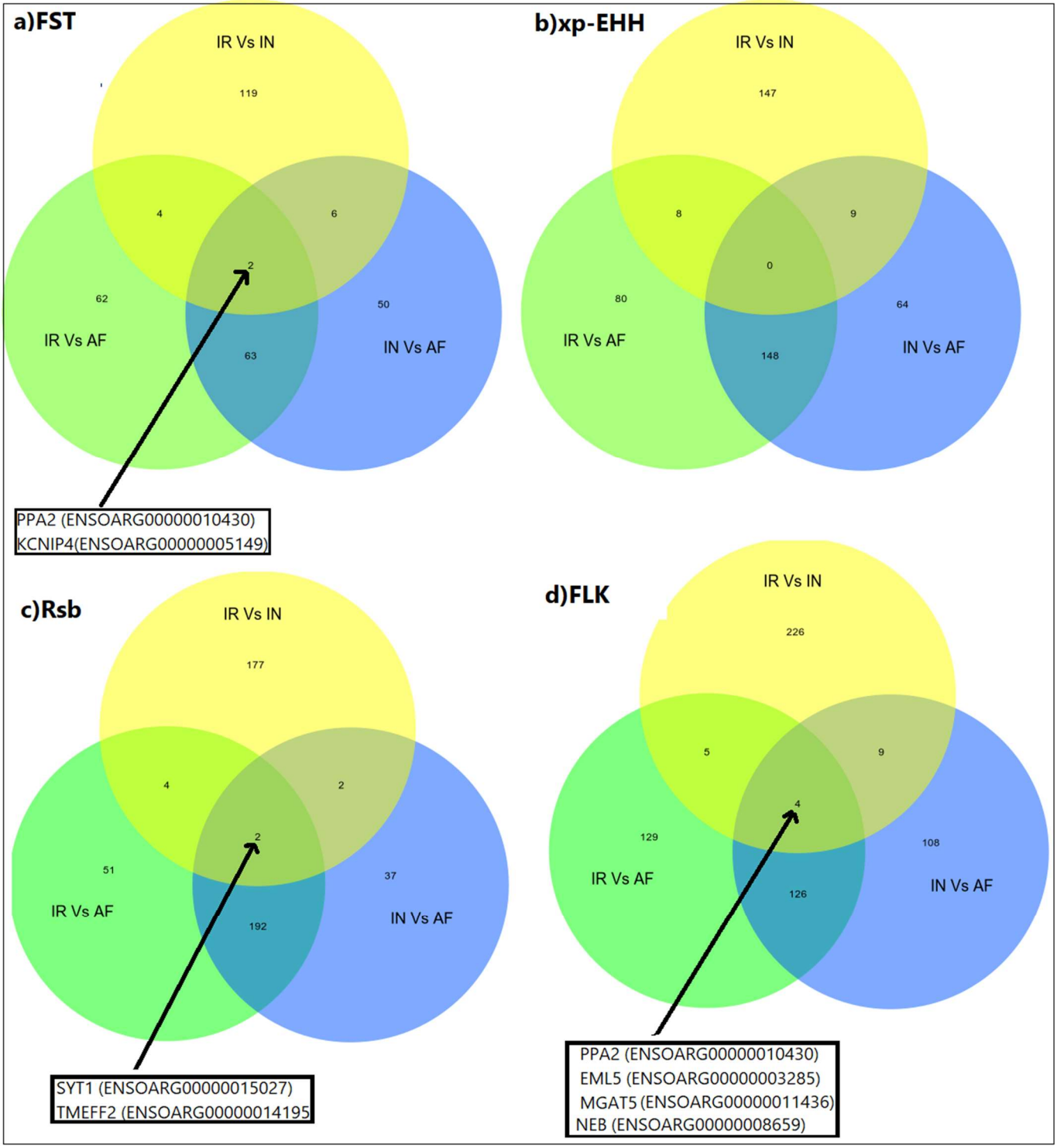
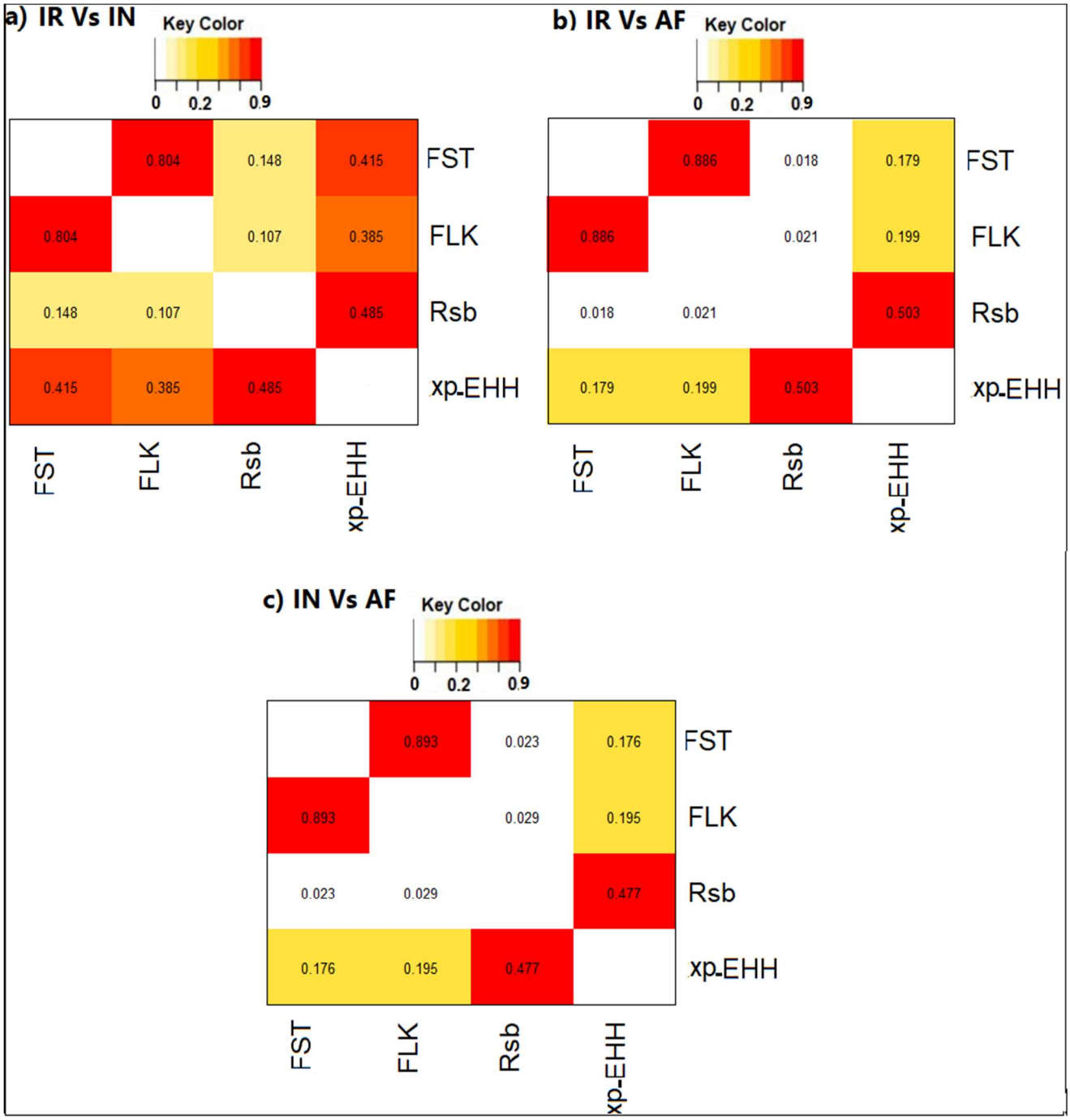


Figure 10. Absolute correlation among different methods used to detect selectve sweeps on: a) IR Vs IN, b) IR Vs AF, and C) IN Vs AF sheep breeds. For breed abbreviations, see Table 1.



868

869

Figures

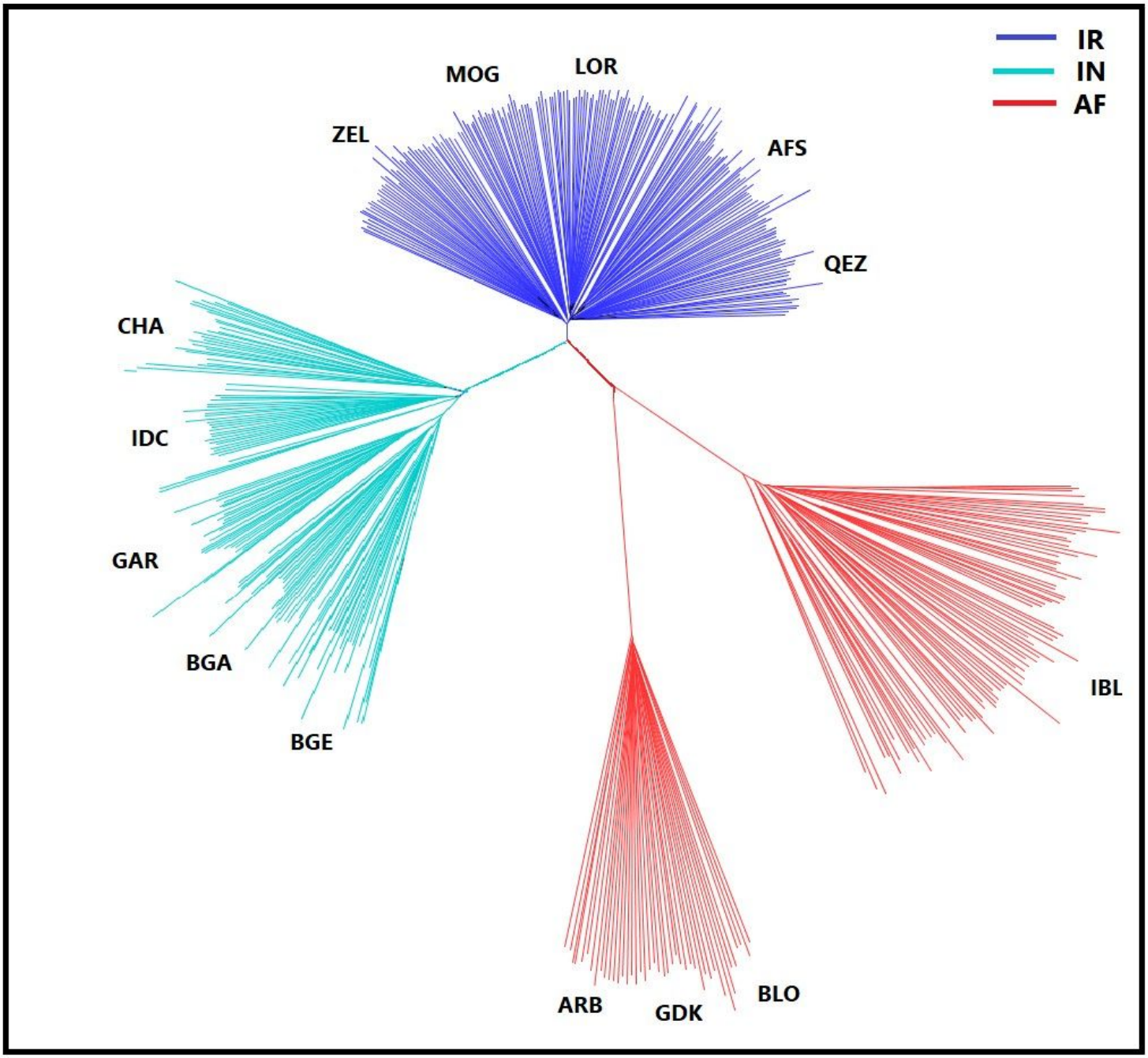


Figure 1

Neighbor-joining phylogenetic tree for 14 sheep breeds based on autosomal SNPs. For breed abbreviations, see Table 1.

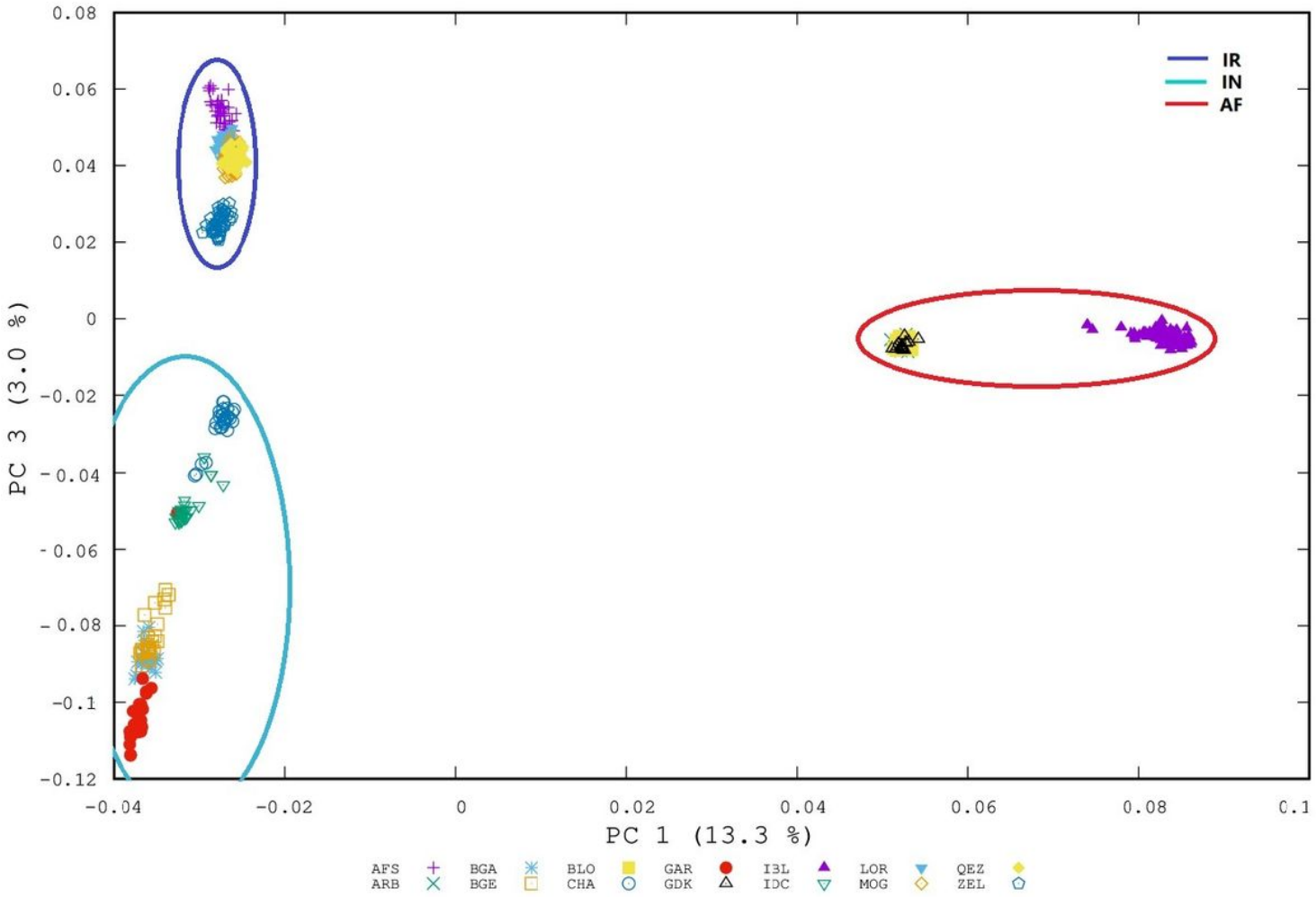


Figure 2

Principal components analysis (PC 1 and PC 3) of among 14 sheep breeds based on autosomes. For breed abbreviations, see Table 1.

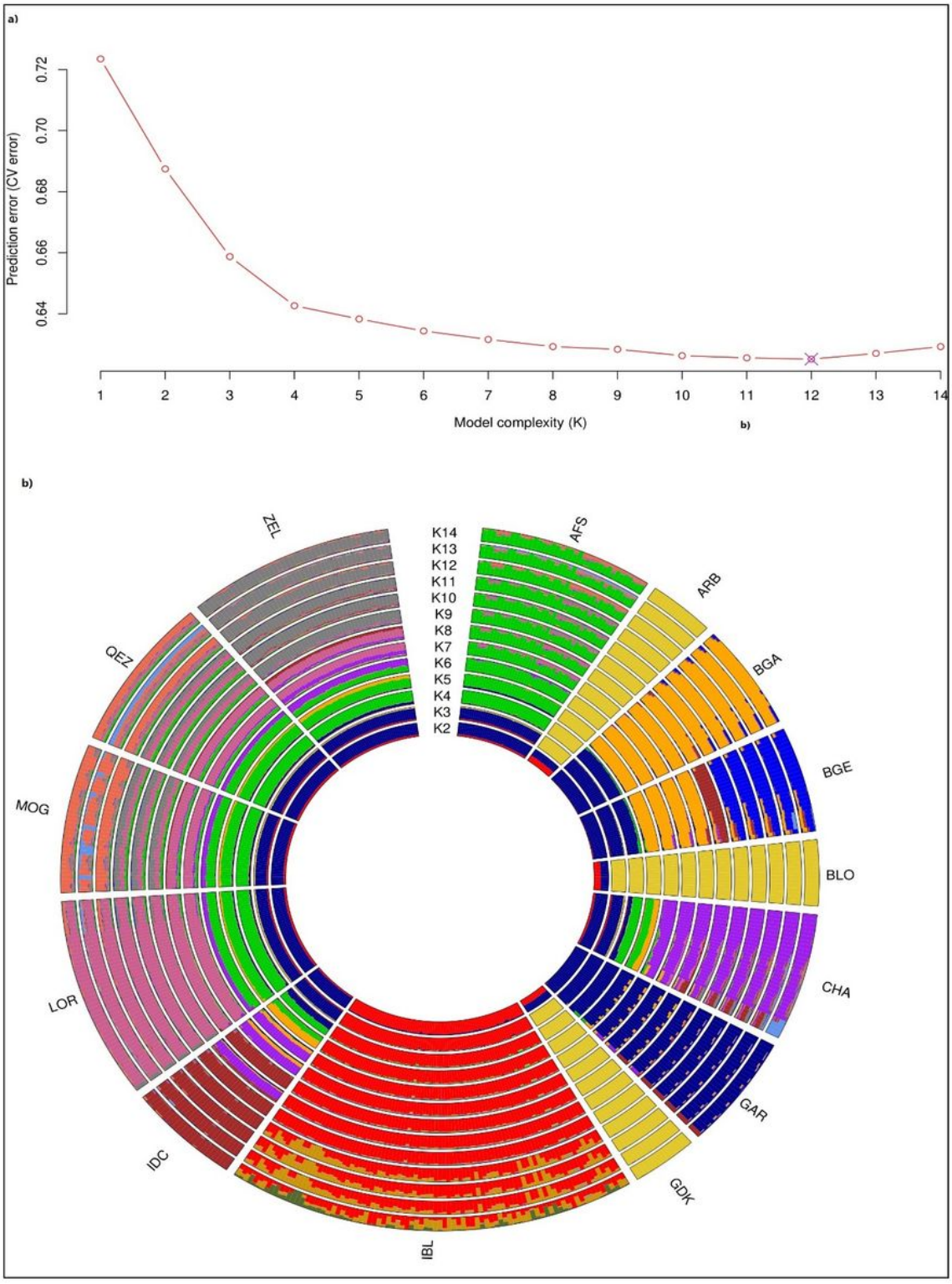


Figure 3

Prediction error (a) and circle Admixture from K2 to K14 (b) plotted, respectively. For breed abbreviations, see Table 1.

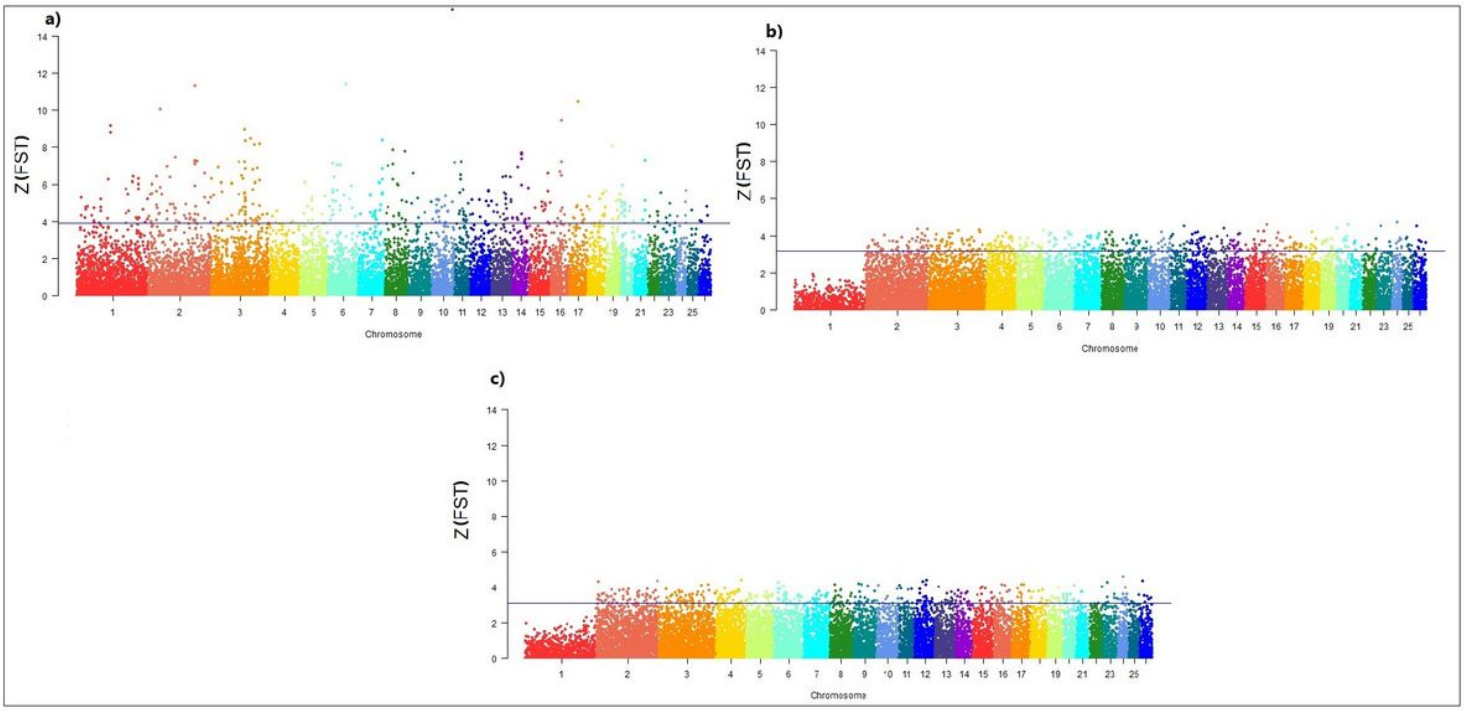


Figure 4

The distribution of absolute Z(FST) values on 26 sheep autosomes: a) IR and IN breeds (The horizontal blue line, Z(FST) ≥ 3.93), b) IR and AF breeds (The horizontal blue line, Z(FST) ≥ 3.18), c) IN and AF breeds (The horizontal blue line, Z(FST) ≥ 3.08). The data points above the horizontal line (blue line) are top 1% Z(FST) values. FST: Fixation index. For breed abbreviations, see Table 1.

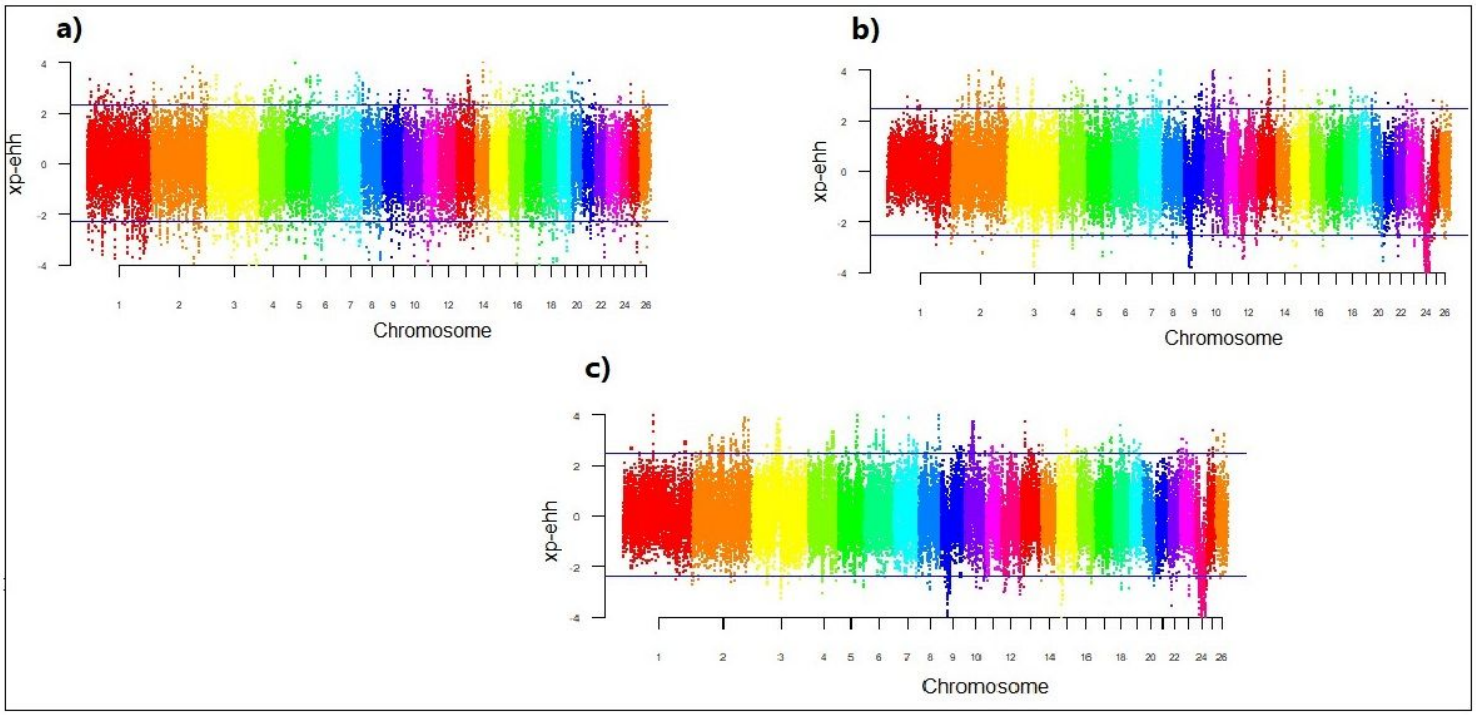


Figure 5

Genomic distribution of standardized cross-population extended haplotype homozygosity (xp-EHH) scores on 26 sheep autosomes pairwise: a) IR and IN breeds, b) IR and AF breeds, c) IN and AF breeds. For breed abbreviations, see Table 1.

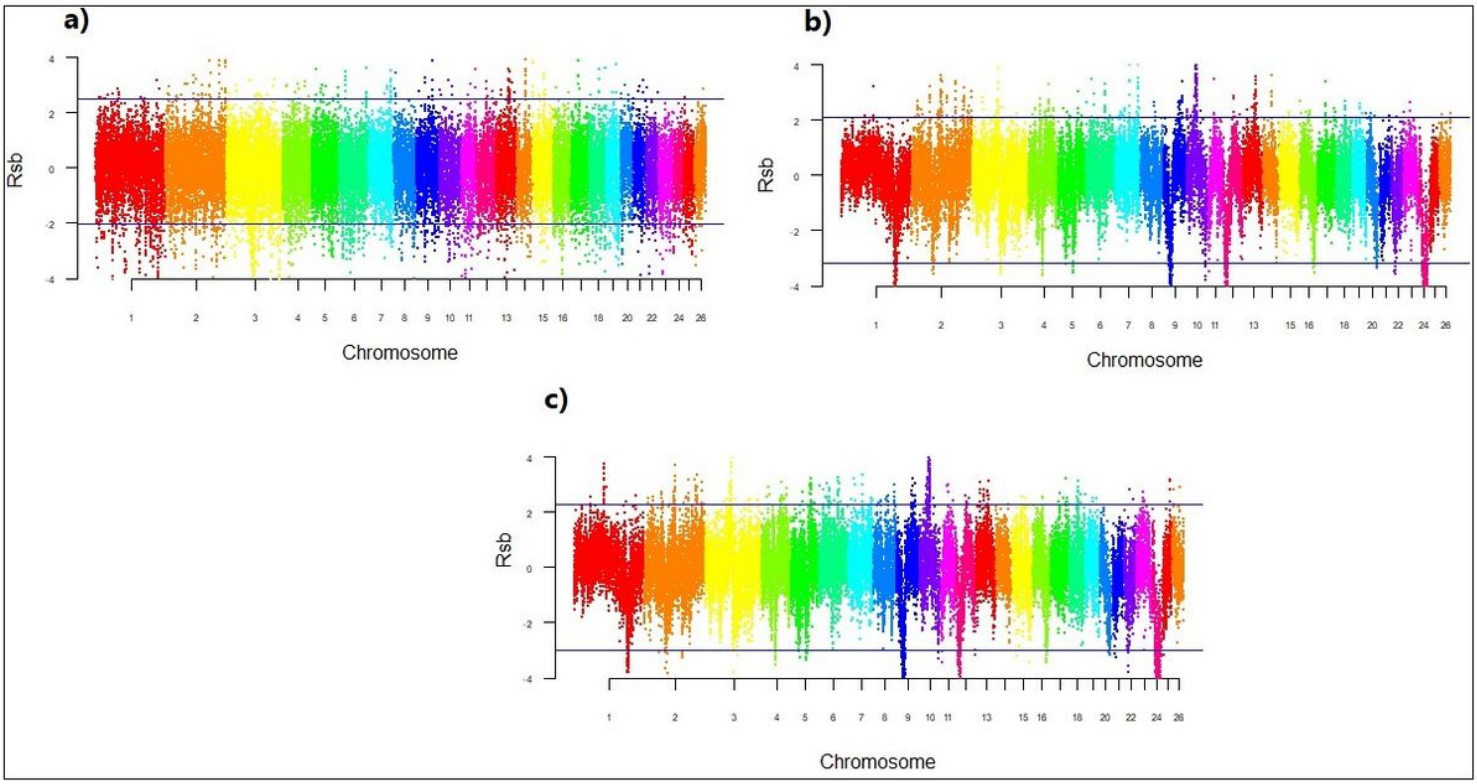


Figure 6

Genomic distribution of standardized haplotype differentiation (Rsb) scores on 26 sheep autosomes pairwise: a) IR and IN breeds, b) IR and AF breeds, c) IN and AF breeds. For breed abbreviations, see Table 1.

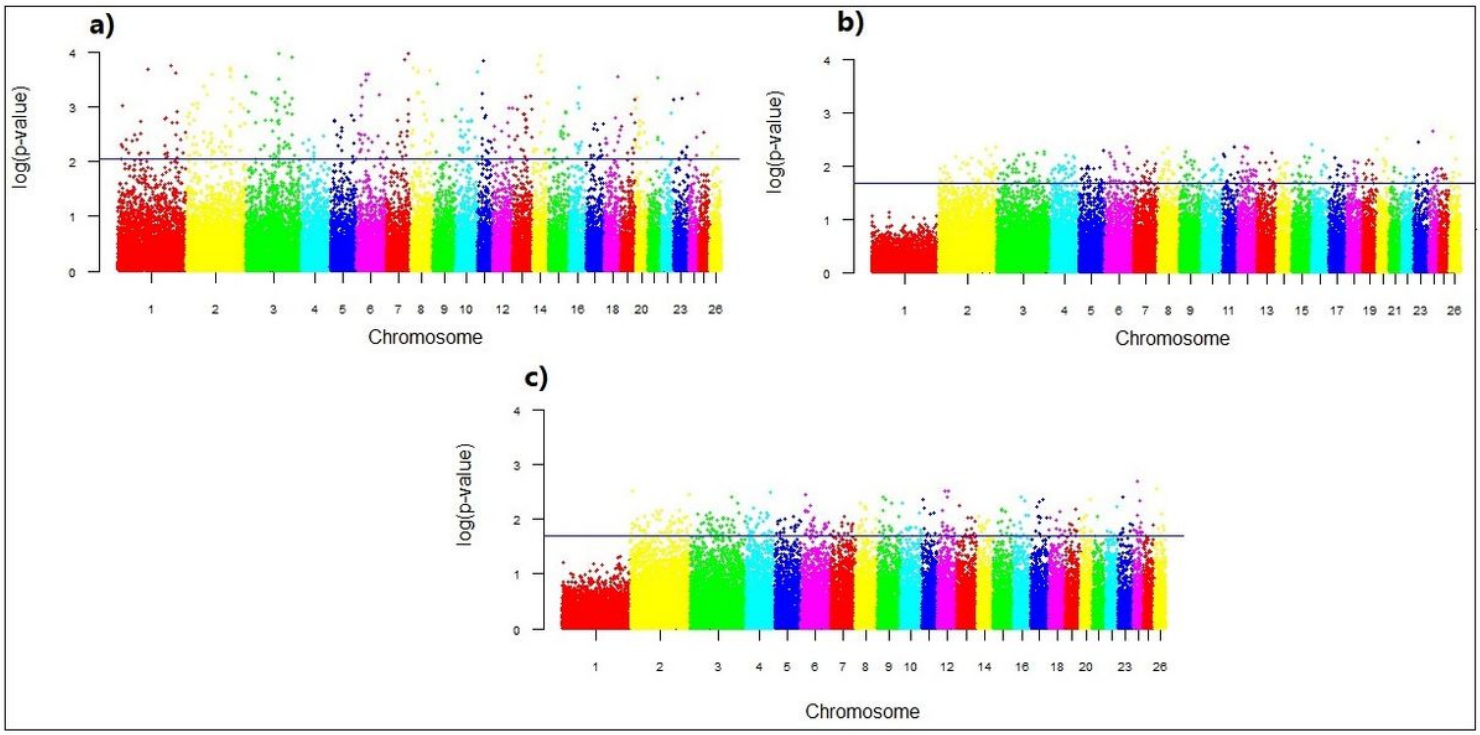


Figure 7

Genomic distribution of single marker statistic (FLK) scores on 26 sheep autosomes pairwise: a) IR and IN breeds, b) IR and AF breeds, c) IN and AF breeds. For breed abbreviations, see Table 1.

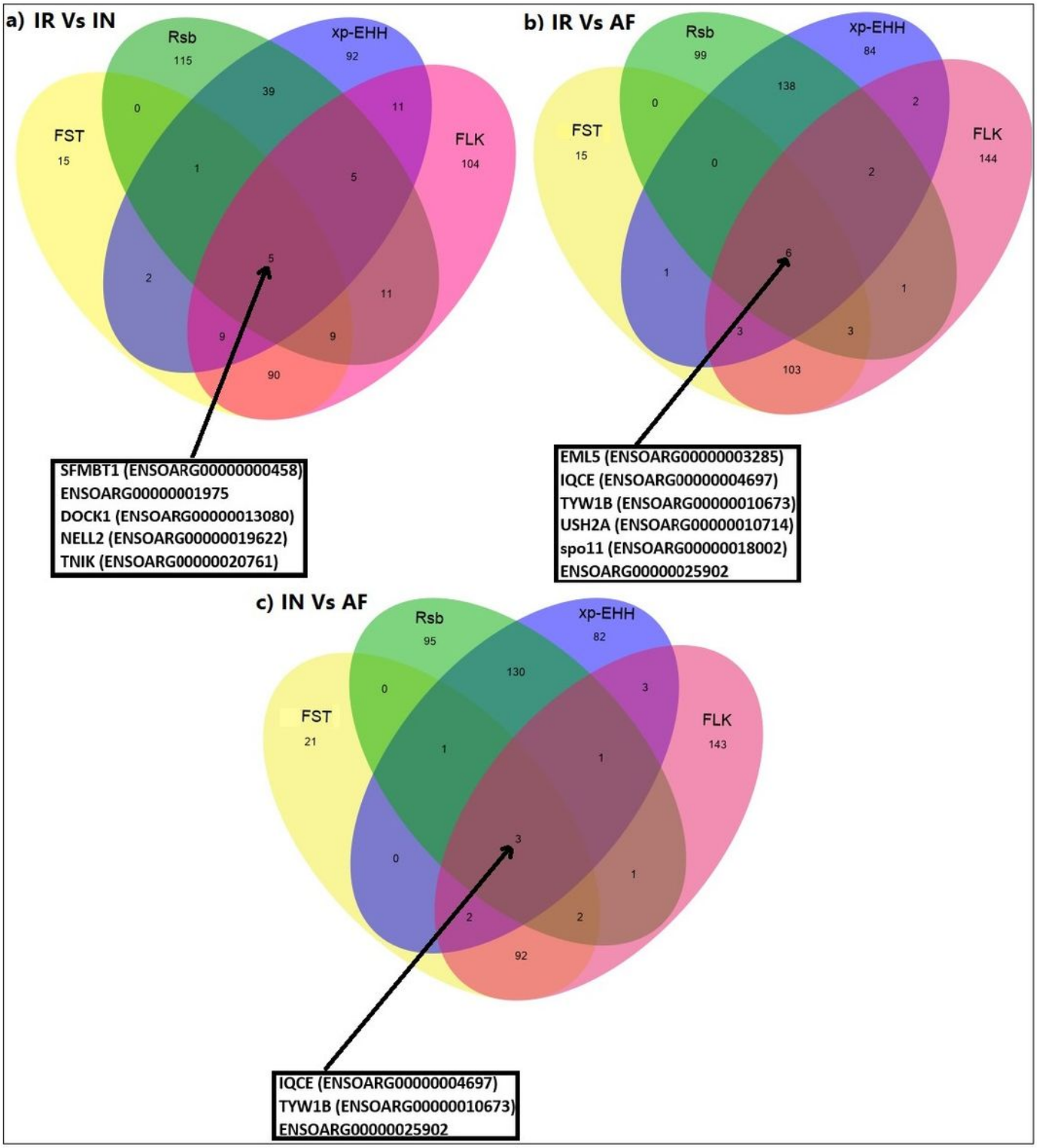


Figure 8

Venn diagram showing the unique and shared candidate genes for FST, Rsb, xp-EHH, and FLK tests on: a) IR Vs IN, b) IR Vs AF, and c) IN Vs AF sheep breeds. For breed abbreviations, see Table 1.

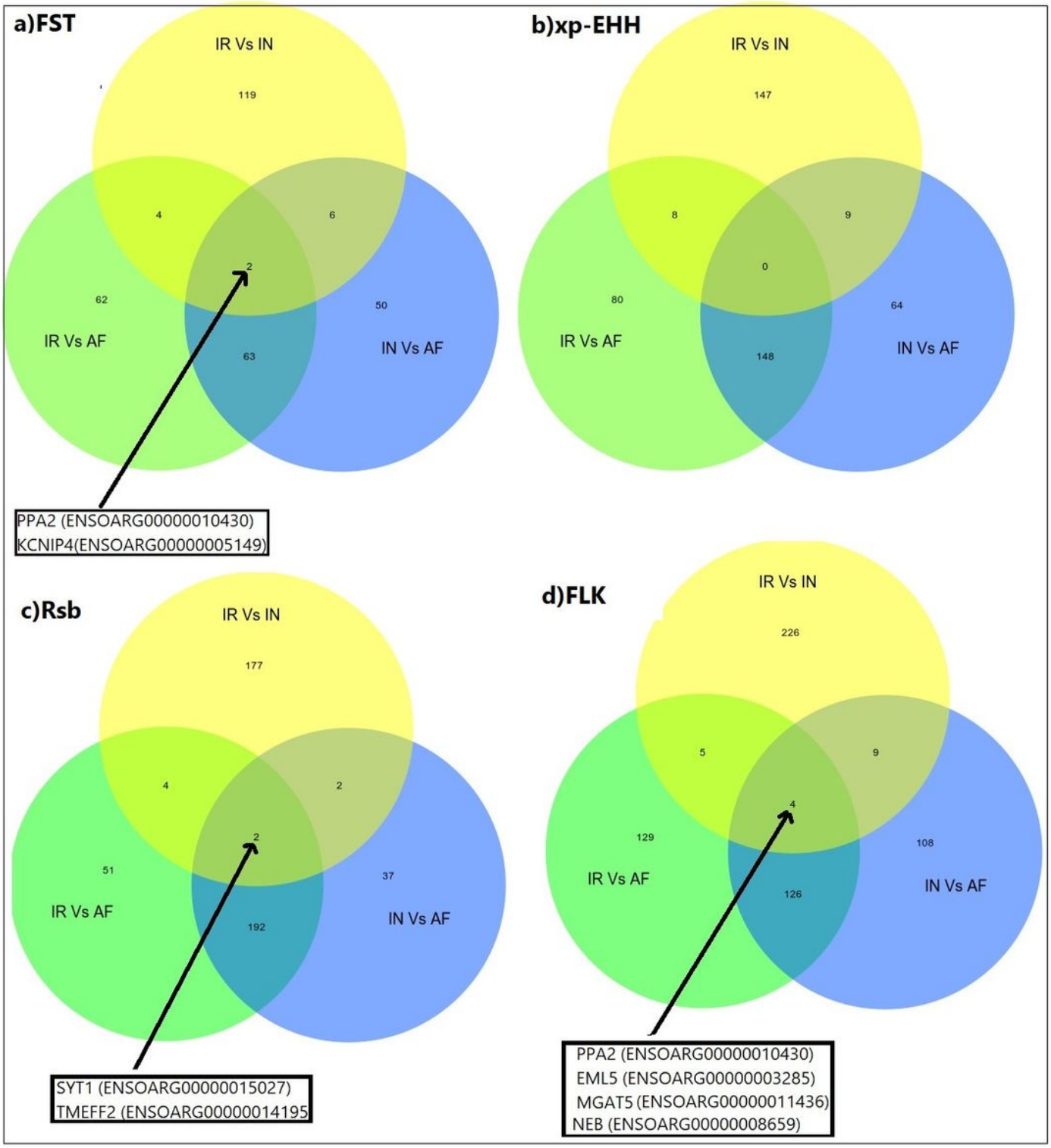


Figure 9

Venn diagram showing the unique and shared candidate genes for IR Vs IN, IR Vs AF, and IN Vs AF data on: a) FST, b) xp-EHH, c) Rsb, and d) FLK tests. For breed abbreviations, see Table 1.

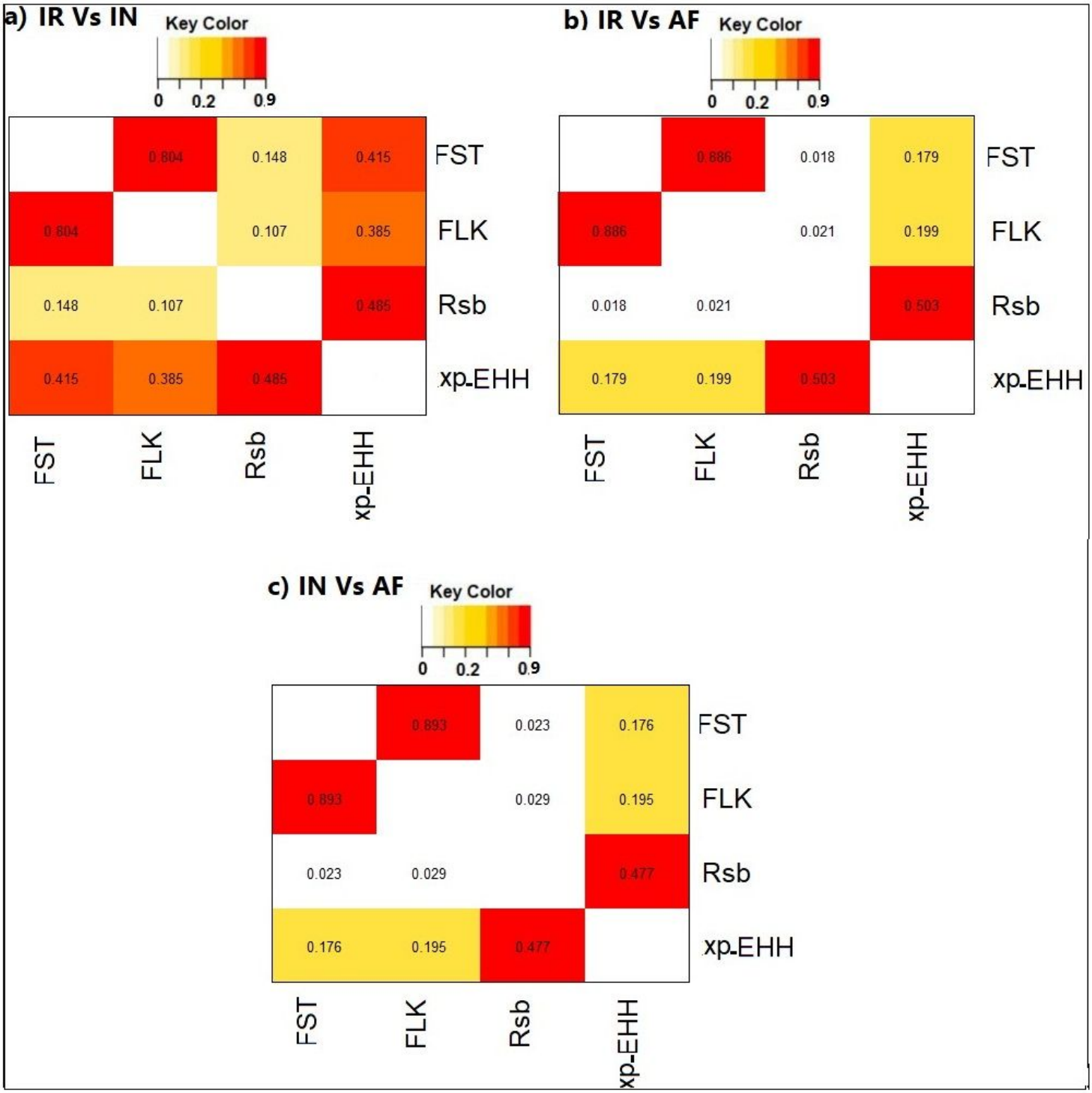


Figure 10

Absolute correlation among different methods used to detect selectve sweeps on: a) IR Vs IN, b) IR Vs AF, and C) IN Vs AF sheep breeds. For breed abbreviations, see Table 1.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryDataset.pdf
- Tables.docx