Analysis of differences between different sheep breeds based on whole-genome resequencing technology

Ting Liu
Inner Mongolia Agricultural University
Wenrui Guo
Inner Mongolia Agricultural University
Taiwu Zhang
Inner Mongolia Agricultural University
Yanni Zhang
Inner Mongolia Agricultural University
Le Yang
Inner Mongolia Agricultural University
Junkang Xie
Inner Mongolia Agricultural University
Letian Kang
Inner Mongolia Agricultural University
Weihao Wang
Inner Mongolia Agricultural University
Lina Sun
Inner Mongolia Agricultural University
Yan Duan (duanyan@imau.edu.cn)
Inner Mongolia Agricultural University
Rina Su
Inner Mongolia Vocational Saihan District

Article

Keywords:

Posted Date: June 21st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1732438/v1

Abstract

The Inner Mongolia Autonomous Region is the region with the highest mutton production in China, but researchers know very little about the genetic evolution of local sheep. In this experiment, whole-genome resequencing technology was used to investigate the genetic distance and single nucleotide polymorphism (SNP) of Chahar sheep, Ujimqin sheep and Xiqi sheep. The results showed that the distribution of SNPs in different regions of the genome was different, thus affecting the function of some genes. Insertion-Deletion (INDEL) also suggested an evolutionary distance between these three types of sheep. The genetic distance between Chahar sheep and the other two types of sheep is relatively far. During the construction of three sheep differential loci, it was found that they could be distinguished by the difference of bases at a certain position on the three chromosomes, which improved the identification of mutton. This suggests that the differences in meat quality between the three types of sheep are not only due to differences in the living environment but also may be influenced by genetic evolution. The research on differential genes can improve the data and theoretical basis for future precise nutrition.

Introduction

Inner Mongolia Autonomous Region is one of the most representative districts with vast grasslands and a unique natural environment in China. Herdsmen have been raising and cultivating sheep in the grasslands for a long time, so there are many excellent breeds of sheep in Inner Mongolia, such as Ujimqin sheep, Chahar sheep and Hulunbeier sheep. Lamb has low fat and cholesterol content, high protein content and tender meat[1,2]. And mutton is an excellent source of essential micronutrients, contributing to physiological and biochemical function in humans[3]. Therefore lamb is more and more popular with consumers. Among them, the Ujumqin sheep produced in the Ujumqin Grassland of Xilin Gol League, Inner Mongolia that has strong adaptability, and minerals (such as calcium, iron and phosphorus)[4–6]. Chahar sheep are mainly distributed in the southern part of Xilingol League, Inner Mongolia. Chahar sheep are a novel breed that has been cultivated in China since the 1990s, producing both meat and wool. The male parent is Merino sheep imported from Germany, and the female parent is a fine wool sheep from Inner Mongolia. It was officially named by the Ministry of Agriculture of China in 2014. This mutton is tender and juicy, with a moderate lipid level, fat but not greasy and marbled meat, rich in nutritional value and delicious taste. Xiqi mutton, a specialty of Xinbarhu Right Banner, Hulunbuir City, Inner Mongolia, is low in moisture, high in protein and fat, rich in umami amino acids such as aspartic acid, and has high nutritional value. Different varieties of sheep have different sensory properties and nutritional value, but the mutton entering the market has a similar appearance, and how to distinguish or select it has become a major difficulty.

Whole-genome resequencing is to sequence the genomes of different individuals or species with known genome sequences, and then perform differential analysis of individuals or groups on this basis[7,8]. Whole-genome resequencing individuals can find a large number of single nucleotide polymorphism sites (SNPs), insertion-deletion sites (InDel, Insertion/Deletion) and other variation information through

sequence alignment. Applying resequencing in population genetics to obtain SNP information and conduct subsequent analysis can more accurately reflect the differences between different individuals or groups at the genome-wide level. Obtain more reliable analysis results and facilitate subsequent sample typing and data mining. Therefore, this study intends to use the whole genome resequencing technology to determine and analyze the gene difference loci of the three breeds of mutton sheep, Ujimqin sheep, Chahar sheep and Xiqi sheep, to provide technical support for scientific and efficient differentiation of these three types of mutton. And the brief flow of this experiment is shown in the figure below (Fig. 1).

Results

Differences in the quality of different varieties of mutton

There were certain differences in the nutritional components of the longissimus dorsi and leg muscles of the three sheep populations, as shown in Table 1. The collagen in the longissimus dorsi and leg muscles of the XQ population was significantly higher than in the other two populations ($P<0.05$). In the longissimus dorsi, the crude fat content of the XQ population was significantly higher than the other two populations ($P<0.05$), but the CHR population had the highest crude fat content in the leg muscles. It can be seen that whether in the longissimus dorsi or the semitendinosus, the muscle water content of the WZMQ population is higher ($P>0.05$). The protein content in the leg muscles of sheep had no significant difference between different sheep ($P>0.05$), while in the longissimus dorsi, the XQ population was significantly lower than the other two populations ($P<0.05$).

Sample whole-genome sequencing quality control and reference comparison

The raw data generated by sequencing was removed from the joints. The proportion of effective reads obtained by removing N bases and low quality is greater than 97%, indicating that the sequencing quality is good and there is a large amount of data available for subsequent analysis. The proportion of bases with quality values greater than or equal to 20 (Q20) in all samples is more than 96%; the proportion of bases with quality values greater than or equal to 30 (Q30) is between 92% and 95%, indicating that the quality of the sample data is excellent. The proportion of the total number of reads aligned to the reference genome for each sample was greater than 98%, indicating that the sequencing data was not disturbed by other data. The sequencing data were aligned to the consensus sequence obtained by the reads clustering, and the statistical results of the alignment rate of the samples are shown in Table 2. Since the reference used is the consensus sequence obtained by reads clustering, there will be some discrepancies in the alignment rate between different samples. In addition, the proportion of the total reads aligned to the reference genome in each sample was greater than 98%, indicating that the sequencing data was not contaminated by other data.

SNP and InDel statistics

SNP Statistical Results

SNP (Single Nucleotide Polymorphisms) refers to the genetic markers formed by the variation of a single nucleotide on the genome, which is numerous in number and rich in polymorphisms[9]. Variations of single nucleotides on the genome include substitutions, deletions and insertions. The ratio of conversion and transversion in the same species is the same. Statistics on the filtered SNPs, the results are as follows Table S1.

The distribution of SNPs in different regions of the genome varies in proportion (Fig. 2), thus affecting the function of certain genes. The effect on intergenic reached 19.01%, but had no effect on transcribed proteins; 38.58% of SNPs affected the introns of the sample genome, forming some intron variants; they had similar effects on the sample transcripts. More than half of the SNPs occurred in non-protein-coding regions, emphasizing that the SNPs that occurred in exons were less likely. For the obtained SNP data, further processing was performed, and filtering was performed according to MAF>0.05 and data integrity>0.8. SNPs with biallelic polymorphisms were retained. The SNPs obtained by the final screening enter into the subsequent analysis.

InDel Statistical Results

InDel (Insertion-Deletion) refers to the insertion and deletion of small fragments in the sample relative to the reference genome, which may contain one or more bases[10]. According to the position of InDel in the genome, it can be divided into InDel of coding and non-coding region sequence[11]. The occurrence of InDel in the coding sequence is related to the encoded protein function and amino acid site. If one or several bases (not multiples of 3) are inserted or deleted in the DNA coding sequence, the mutation is called frameshift mutation[12]. This kind of mutation will cause all changes in the DNA coding frame downstream of the insertion point or deletion point, and as a result, the amino acid sequence after the mutation point will be changed.

As can be seen from Fig. 3, 38.44% of InDel occurred in the sample transcript, which had an impact on the sample transcription and translation. The InDel occurring in the non-coding region will reduce the efficiency of transcription and the accuracy of splicing. Fig. 3 shows that more than half of the InDel occurs in the non-coding region of the genome, which has little effect on the appearance traits of the sample.

Genetic population structure of the sheep individuals

Principal components analysis (PCA) was performed based on the SNP, and the principal component clustering of all samples was obtained, as shown in Fig. 4a. In the PCA score plot, three principal components (PC1, PC2 and PC3) were extracted to be 7.53%, 2.96% and 2.89%, respectively. From the analysis in Fig. 4a, it can be concluded that WZMQ is tightly clustered, XQ is more dispersed but better than CHR. The PCA results showed that the three kinds of sheep could be completely separated, and there was no crossover phenomenon. Clustering with PC1 and PC2 showed better separation among the three samples, as was clustering with PC2 and PC3. When clustering with PC1 and PC3, it was found that the distance between XQ and WZMQ was relatively close, and CHR was completely separated from them.

The evolutionary history of individuals was inferred with the neighbor-joining (NJ) tree (Fig. 4b). The phylogenetic tree shows that the CHR population samples are clustered together individually. In a population, two loci on the same chromosome will be linked, that is, the genotypes of these two loci in the population are not in a random combination state, which is called "linkage disequilibrium" (LD). LD analysis was performed between different SNPs, and the linkage disequilibrium coefficient (r2) was calculated. It can be seen from the LD-decay decay diagram (Fig. 4c) that the LD of the CHR sample population decays slowly, indicating that the group formation time is short, the linkage exchange between individuals is insufficient, and the kinship relationship is relatively close; the LD decay of the WZMQ group is faster, indicating that this group is formed relatively ancient. The LD-decay decay map once again proves that The distance between CHR and the other two sheep is farther, and the distance between WZMQ and XQ is short. In addition to thisFor SNPs up to 50 kb apart, the average r2 values were equal to 0.093 (CHR), 0.082 (WZMQ), 0.083 (XQ). Further details about the LD analysis using r2 are included in supplementary Table S2. The result indicated that the LD decay tends to be stable when the distance is 100 kb. Therefore, genes located within ±50kb near important SNP sites can be listed as candidate genes in the follow-up study.

This experiment performed unsupervised cluster analysis using ADMIXTURE. In the analysis of population genetic structure, different K values represent the assumption that there are K ancestral groups (K=1~10). The analysis shows that when K=3, the three groups of samples have the best clustering. All three populations in the experiment had relatively uniform genetic components (Fig. 4d).

Fst analysis

The population fixed coefficient Fst reflects the level of population allelic hetero-zygosity. Fst was a basic indicator for traditionally measuring genetic differentiation among populations. Understanding population structure and genetic background in an evolutionary context[13–15]. The corresponding gene loci and their functions were found by analyzing the selected regions with the Fst value in the top 1%. Finally, the differential genes between different groups are found and the molecules are marked.

The three groups of samples were subjected to two-to-two Fst analysis to obtain the following Fig. S1. The highest Fst value between the CHR and WZMQ populations is located on the NC_019470.2 chromosome, indicating that the genetic differentiation between the two populations is relatively large. This region has undergone a strong selection and has more differential loci. The comprehensive analysis of Fig. S1 shows that the CHR population was more genetically differentiated than the other two populations. The Fst value between WZMQ and XQ populations was lower than 0.25. The degree of genetic differentiation in both the WZMQ and XQ populations was small.

Functional annotation of GO and KEGG genes in the screened region

CHR&WZMQ&XQ

The GO database divides gene functions into three parts: cellular component, molecular function, and biological process. The KEGG database is not only a functional annotation of genes themselves but also a database related to pathways. The final result obtained by GO and KEGG is the integrated macro result[16]. To detect candidate genes for wool traits by resequencing Chinese fine-wool sheep, a genome-wide association study (GWAS) was performed to detect candidate genes for eight wool traits. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment results revealed that many important pathways were associated with keratin and cell proliferation and differentiation[17].

Regions were screened with the 99% threshold of Fst value between WZMQ and CHR, and genes contained in all regions were subjected to GO annotation analysis. As a result, 519 genes were enriched into 50 GO entries. More than 80% of the gene numbers in Fig. S2a were enriched in the cytoplasm (nucleus). Among the GO entries, bio-logical processes were less enriched. KEGG pathway analysis was performed on the genes contained in the region (Fig. S2b). Most genes were annotated to pathways related to signaling, infectious diseases and the immune system. Two of these pathways were annotated as being related to the degradation and metabolism of xenobiotics.

Region screening was performed with the 99% threshold of Fst value between XQ and CHR. GO annotation was performed on the genes contained in the region. As a result, 396 genes were enriched in 50 GO items (Fig. S3a), of which the number of genes enriched in 25 biological processes didn't exceed 30%. Among these GO entries, the most enriched genes were cytoplasm. There were some differences in gene GO annotations between CHR and the other two groups of populations (Fig. S1a and Fig. S2a). In terms of molecular function between the CHR population and the other two populations, the genes contained in the different regions were mostly enriched in the combined sub-entry. In the biological process, they were mostly enriched in transcription-translation-related entries. The KEGG pathway analysis of the genes contained in the region showed that (Fig. S2b) about 15% of the genes were enriched in the pathways related to signal transduction; about 9% of the genes were enriched in the pathways related to the immune system. The results showed that the difference region between the two populations of CHR and XQ contained more genes than the samples were related to environmental adaptability. Among them, there were 10 KEGG pathway categories related to body metabolism.

WZMQ&XQ

Regions were screened with the 99% threshold of Fst value between WZMQ and XQ. GO annotations were performed on the genes contained in the regions. The results showed that 405 genes were annotated in 50 GO items (Fig. S3a). About 79% of the gene GO annotations were related to the cytoplasm. Then the KEGG pathway analysis of the genes contained in the region showed that 39 pathways were enriched in the Signal transduction category, indicating that most of the genes screened in the region were related to the body's signal transmission. The KEGG pathways related to body diseases are more enriched (Fig. S3b), comparing WZMQ with the other two sheep breeds, it can be found that a large part of the genes screened by region is enriched in the pathway categories related to sheep diseases. There are 58

pathways enriched in body metabolism, of which 12 pathways are enriched in the category of Carbohydrate metabolism.

The GO annotation map of the genes in the regional screening between WZMQ and the other two sheep breeds shows that most of the GO annotations of the genes in the three sheep regions are related to cellular components (Fig. S1a, S2a and S3a). The number of pathways enriched in the immune system and endocrine system of WZMQ, XQ and CHR were significantly higher than in other systems (Fig. S1b, S2b and S3b), In terms of body metabolism, the number of enriched pathways between XQ and WZMQ is lower than that of CHR and WZMQ.

Construction of three kinds of sheep differential loci

Genomes of different individuals were sequenced by whole-genome resequencing of sheep of the known genome sequence. Through sequence alignment, a large number of SNP, InDel and other variation information can be found. In turn, it reflects the differences between different sheep individuals or groups at the genome-wide level. As shown in Table S3, in the five sheep gene fragments corresponding to chromosomes NC_019470.2 and NC_019474.2, the CHR population is completely different from the other two populations in base types. Several differential loci shown in Table S3 can completely distinguish CHR from the other two populations, but cannot distinguish the two populations of WZMQ and XQ. It is necessary to combine more differential sites to achieve the purpose of differentiation.

In the two sheep gene fragments corresponding to chromosome NC_019484.2, the bases of WZMQ and XQ populations are completely different at several sites on this chromosome (Table S4). Finally, the two populations of WZMQ and XQ can be distinguished by these loci. According to Table S3 and Table S4, a schematic diagram like Fig. 5 can be drawn. It can be seen that the three types of sheep can be distinguished by the differences in the three chromosomes.

Discussion

The nutritional composition of animal skeletal muscle can be affected by species, diet, and exercise[18–20]. Both the longissimus dorsi and the semitendinosus in the XQ group had the highest collagen content among the three groups. Previous studies found that collagen content was positively correlated with the shear force of the meat and negatively correlated with tenderness[21]. Shen et al. found that the content of collagen fibers in Ziwuling black goats' muscles was significantly lower than that of Liaoning cashmere goats, and the tenderness of Ziwuling black goats was higher than that of Liaoning cashmere goats[22]. The higher collagen content in Xiqi sheep could be due to the muscle's higher connective tissue content, such as the more developed perimysium. The content of crude fat in the longissimus dorsi muscle of Xiqi sheep was the highest, and the content of crude fat in the thigh muscle of Chahar sheep was the highest. This may be related to its species, feed composition, feeding methods, etc[19,23,24]. Studies have found that intramuscular fat can improve the water retention of meat, which is related to its tenderness and juiciness[25,26]. In addition to intramuscular fat, intramuscular water content also affects meat tenderness

and juiciness[27]. Because of the aforementioned factors affecting meat quality, as well as the data obtained from Table 1, which showed that the longissimus dorsi muscle of the WZMQ population was lower in fat and higher in water and protein content, indicating that the meat quality of the WZMQ sheep was better than that of the other two sheep breeds in this part. In the leg muscles, the meat quality of XQ sheep is relatively better.

By understanding the diversity of single nucleotide base morphology, scholars can analyze the genome structure of species and the evolutionary history of different sheep genes. SNP appear most frequently on CG sequences, and most of them are C-T conversions, because C in CG is often methylated, and becomes thymine (T) after spontaneous deamination. According to the position of the single nucleotide in the gene, it can be divided into gene non-coding region SNP, intergenic region SNP and gene coding region SNP[28]. There are fewer SNP (coding SNP, cSNP) located in the coding regions of genes because within exons, their mutation rate is only 1/5 of the surrounding sequences. From the perspective of the impact on the genetic traits of organisms, cSNP can be divided into two types: synonymous cSNP (synonymous cSNP), where the change of the coding sequence caused by the SNP has no effect on the amino acid sequence of the protein being translated. The other is non-synonymous cSNP (non-synonymous cSNP), which means that changes in the base sequence can change the translated protein sequence, thereby affecting the function of the protein[29,30]. Such changes are often the direct cause of changes in biological traits. Non-synonymous cSNPs account for about half of all cSNPs[31]. By analyzing the SNP loci on the genomes of three different breeds of sheep. According to the analysis process of GATK, the mutation sites were analyzed to obtain the possible SNP information of each sample, and SNPeff was used to annotate the structure of the mutation sites.

Through the cluster analysis of the three groups of sheep, we can see that XQ and other breeds of sheep were farther from each other, while CHR and WZMQ were relatively close, and some samples appear outliers, which may be due to the hybridization of their parents. This may be because the Chahar sheep have been artificially bred by crossbreeding in the past few years, and individual variances are relatively considerable. Part of the reason for the relatively close kinship of the CHR and WZMQ groups may be that the two groups are geographically close and have similar living environments, so the environment is less selective for the two groups (Fig. 3A). Most of the genetic material of individuals in the XQ and CHR populations came from different ancestors, indicating that the two are far from each other again. The genetic material sources of some sample individuals are crossed, and the possible reason is that their parents have a hybrid generation, which causes them to contain some genetic material of other populations. The closer the disequilibrium coefficient was to 1, the higher the degree of linkage between the two loci. As the distance between SNPs increases, linkage disequilibrium decays very strongly and reaches equilibrium after a certain distance. The faster the LD decays, the lower the probability of linkage between SNPs with the same distance. It is generally believed that the species with fast LD decay is older, and the population linkage exchange is more sufficient; on the contrary, it means that the species or population is newly formed, and the relationship between individuals is relatively close.

The F statistic is affected by different factors, such as mutation, genetic drift, inbreeding, selection, etc. And studies have found that FST affects wool quality traits, and its SNPs 2 and 4 may be useful markers for marker-assisted selection and sheep breeding[32]. Under neutral evolutionary conditions, the size of the F statistic was primarily governed by factors such as genetic drift and migration. If a certain allele in the population was adaptively selected, then its high frequency will increase the level of differentiation among populations, and the selected region will have a larger Fst value. However, there were many screening areas higher than the threshold of 1%. Finally, WZMQ and XQ sheep meat products could be distinguished by analyzing multi-locus differences.

Further analysis of the KEGG pathway annotation of genes contained in different regions between CHR and the other two populations revealed differences in body metabolism. In terms of Carbohydrate metabolism, Amino acid metabolism, etc. The number of enriched pathways was significantly different between the CHR and the other two groups, which may ultimately account for the difference in meat quality. Continue to study according to this analysis pathway, to find the genes that determine the difference in meat quality between populations. The metabolic sub-categories with the highest number of enriched pathways differed between pairs, which may ultimately influence the difference between WZMQ and the other two sheep's apparent traits. Pathway enrichment was similar between WZMQ and the two sheep, with no significant difference.

Studies have shown that when establishing a genotyping method, selecting a few loci for determination may increase the chance of confounding samples due to the combined effect of alleles[33]. The increase in the number of selected sites can improve the resolution rate, but when the number of selected sites reaches a certain number, it will increase the consumption of manpower and material resources. According to the previous article, these three sheep populations have parental hybridization or artificial introduction due to geographical location and other reasons. It appears that a certain population is less genetically differentiated from other populations and has fewer differential loci so that it cannot be completely distinguished by one differential locus. Therefore, this topic will be distinguished by the exclusion method. Through the comparison of the corresponding gene fragments of each chromosome in different populations, SNP loci that can be used to distinguish different populations are found.

In conclusion, the quality of longissimus dorsi muscle of WZMQ is better, and the quality of leg muscle of XQ is better. The overall degree of genetic differentiation between the CHR population and the other two populations was large from the genetic evolution analysis. And the mutton of the three types of sheep could be differentiated by differences in bases on their chromosomes.

Materials

Three sheep breeds in Inner Mongolia Autonomous Region were selected: 15 Ujumqin sheep (WZMQ), 16 Chahar sheep (CHR) and 15 Xiqi sheep (XQ), and the muscle tissues were collected after slaughtering and stored at -80 ℃ immediately.

Three kinds of sheep longissimus dorsi muscle and semitendinosus qualities

The longissimus dorsi muscle (LD) between the 12th and 13th rib and semitendinosus (ST) located on the inside of the sheep's femur was used to evaluate meat quality traits. The contents of moisture, protein, and fat in the LD and ST of sheep have been analyzed following the AOAC procedure (AOAC, 2015). Moisture in the LD and ST of sheep was determined by drying at 105℃ overnight in a GRX-9053A thermoelectric thermostat drying box (Shanghai bluepard instruments Co., Ltd., Shanghai, China). Crude protein in the LD and ST of sheep was measured by the Kjeldahl method with Kjeltec 8200 (FOSSInc., Hillerød, Denmark), as referred by the national standard (GB 5009.5-2016). Fat in LD and ST of sheep was extracted in an SXT-02 apparatus (Shanghai HongJi Instrument Co., Ltd., Shanghai, China) using petroleum ether. Total collagen (TC) measurement was based on the colorimetric determination of hydroxyproline (Hyp). Samples of CHR, WZMQ and XQ muscles, were thawed and trimmed of fat, fascia and visible connective tissue and were used for TC determination. Meat proteins were hydrolyzed in an acid medium (sulfuric acid) and heated so that residues of hydroxyproline which are released are oxidized by the action of chloramine T. Pyrrole derivatives are generated and, after the addition of p dimethylaminobenzaldehyde, they resulted in a colored compound.

Three kinds of sheep whole-genome sequencing

Firstly, the genomic DNA of three groups of samples, CHR, WZMQ and XQ, was extracted, and the qualified DNA samples were randomly broken into small fragments under a Covaris crusher. The entire library was prepared through the steps of end repair, adding ployA tails, adding sequencing adapters, purification, and PCR amplification. After passing the library inspection, high-throughput sequencing was performed by Hangzhou Lianchuan Biotechnology Co., Ltd.

After the sequencing is completed, the quality control of the sequencing data is performed, and the low-quality sequences and adapter sequences are removed to obtain CleanData. The obtained CleanData data were compared with the reference genome by BWA software, SNP (single nucleotide polymorphism) and InDel (insertion and deletion) were detected by GATK software, and the Detected variant sites were subjected to mass filtering.

Construction of three sheep differential loci

During evolutionary analysis, the population structure of the sample will be analyzed based on SNP data, and the analysis content includes phylogenetic tree and principal component analysis. Subsequent Fst calculations were also performed based on the SNP data and looked at differences between groups in different segments of the genome. Through the comparison of the corresponding gene fragments of each chromosome in different populations, SNP loci that can be used to distinguish different populations are found.

Declarations

Acknowledgment

This research was supported by Department of Finance of Inner Mongolia Autonomous Region, Department of Science and Technology of Inner Mongolia Autonomous Region, grant number “KCBJ2018068” and Natural Science Foundation of Inner Mongolia, grant number “2021MS03012”.

Author contributions statement

T.L., W.G., and Y.D. conducted project administration. T.L., and T.Z. conducted visualization. T.L., and T.Z. wrote the original draft. T.L., T.Z., J.X., L.K., and Y.D. conducted the review and editing. T.L., T.Z., and Y.Z. conducted the formal analysis. W.G., W.W., and L.S. conducted supervision. W.G. conducted methodology. W.G., T.Z., L.Y., J.X., and L.K. operated software. R.S., and Y.Z. conducted investigation. Y.D. acquired resources. Y.D. funding acquisition. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Availability of data

Raw sequence data of three sheep breeds from this study were deposited in the NCBI (https://www.ncbi.nlm.nih.gov/) SRA database. BioProject accession numbers is PRJNA827983.

Ethics approval and consent to participate

All animal management and experimental procedures for this study were approved by the Institutional Animal Care and Use Committee of Inner Mongolia Agricultural University (Approval number: NND2021072) and were carried out according to the guidelines for animal experiments of the National Institute of Animal Health, China (GB 14925-2010). All experimental protocols followed ARRIVE guidelines.

References

1. Li, D., Zhang, H., Ma, L., Tao, Y. & Liu Jun and Liu, D. Effects of ficin, high pressure and their combination on quality attributes of post-rigor tan mutton. LWT-FOOD SCIENCE AND TECHNOLOGY 137, (2021).
2. Zhang, Y., Sun, Y. & Song, H. Variation in Volatile Flavor Compounds of Cooked Mutton Meatballs during Storage. FOODS 10, (2021).
3. Zhang, Q., Que, M., Li, W., Gao, S. & Tan Xin and Bu, D. Gangba sheep in the Tibetan plateau: Validating their unique meat quality and grazing factor analysis. JOURNAL OF ENVIRONMENTAL SCIENCES 101, 117–122 (2021).
4. Siqin, Q. et al. Relationships among muscle fiber type composition, fiber diameter and MRF gene expression in different skeletal muscles of naturally grazing Wuzhumuqin sheep during postnatal

development. ANIMAL SCIENCE JOURNAL 88, 2033–2043 (2017).

5. Li, S. et al. Whole-genome resequencing of Ujumqin sheep to investigate the determinants of the multi-vertebral trait. GENOME 61, 653–661 (2018).
6. Jin, Y., Zhang, X., Zhang, J., Zhang, Q. & Tana. Comparison of Three Feeding Regimens on Blood Fatty Acids Metabolites of Wujumqin Sheep in Inner Mongolia. ANIMALS 11, (2021).
7. Gao, J., Xu, G. & Xu, P. Whole-genome resequencing of three Coilia nasus population reveals genetic variations in genes related to immune, vision, migration, and osmoregulation. BMC GENOMICS 22, (2021).
8. Bhati, M., Kadri, N. K., Crysnanto, D. & Pausch, H. Assessing genomic diversity and signatures of selection in Original Braunvieh cattle using whole-genome sequencing data. BMC GENOMICS 21, (2020).
9. Zhang, F. et al. Genome-Wide SNPs and InDels Characteristics of Three Chinese Cattle Breeds. ANIMALS 9, (2019).
10. Jiang, J. et al. Whole-Genome Resequencing of Holstein Bulls for Indel Discovery and Identification of Genes Associated with Milk Composition Traits in Dairy Cattle. PLOS ONE 11, (2016).
11. Stafuzza, N. B. et al. Single nucleotide variants and indels identified from whole-genome resequencing of Gyr, Girolando, and Holstein cattle breeds. JOURNAL OF ANIMAL SCIENCE 95, 80–81 (2017).
12. Ochoa, A. & Storey, J. D. Estimating F-ST and kinship for arbitrary population structures. PLOS GENETICS 17, (2021).
13. Kitada, S., Nakamichi, R. & Kishino, H. Understanding population structure in an evolutionary context: population-specific F-ST and pairwise F-ST. G3-GENES GENOMES GENETICS 11, (2021).
14. Ghoreishifar, S. M. et al. Shared Ancestry and Signatures of Recent Selection in Gotland Sheep. GENES 12, (2021).
15. Eydivandi, S., Roudbar, M. A., Ardestani, S. S., Momen, M. & Sahana, G. A selection signatures study among Middle Eastern and European sheep breeds. JOURNAL OF ANIMAL BREEDING AND GENETICS 138, 574–588 (2021).
16. Chen, L. et al. Prediction and analysis of essential genes using the enrichments of gene ontology and KEGG pathways. PLOS ONE 12, (2017).
17. Zhao, H. et al. Genome-wide association studies detects candidate genes for wool traits by re-sequencing in Chinese fine-wool sheep. BMC GENOMICS 22, (2021).
18. Zhang, M. et al. Effects of physical exercise on muscle metabolism and meat quality characteristics of Mongolian sheep. FOOD SCIENCE & NUTRITION 10, 1494–1509 (2022).
19. Kawecka, A., Sikora, J., Gasior, R. & Puchala Michal and Wojtycza, K. Comparison of carcass and meat quality traits of the native Polish Heath lambs and the Carpathian kids. JOURNAL OF APPLIED ANIMAL RESEARCH 50, 109–117 (2022).

20. Rybarczyk, A., Boguslawska-Was, E. & Pilarczyk, B. Carcass and Pork Quality and Gut Environment of Pigs Fed a Diet Supplemented with the Bokashi Probiotic. ANIMALS 11, (2021).
21. Li, X., Ha, M., Warner, R. D. & Dunshea, F. R. Meta-analysis of the relationship between collagen characteristics and meat tenderness. MEAT SCIENCE 185, (2022).
22. Shen, J. et al. Comparative Transcriptome Profile Analysis of Longissimus dorsi Muscle Tissues From Two Goat Breeds With Different Meat Production Performance Using RNA-Seq. FRONTIERS IN GENETICS 11, (2021).
23. Natalello, A. et al. Effect of different levels of organic zinc supplementation on pork quality. MEAT SCIENCE 186, (2022).
24. Wang, B. et al. Effects of feeding regimens on meat quality, fatty acid composition and metabolism as related to gene expression in Chinese Sunit sheep. SMALL RUMINANT RESEARCH 169, 127–133 (2018).
25. Swiatkiewicz, M., Olszewska, A., Grela, E. R. & Tyra, M. The Effect of Replacement of Soybean Meal with Corn Dried Distillers Grains with Solubles (cDDGS) and Differentiation of Dietary Fat Sources on Pig Meat Quality and Fatty Acid Profile. ANIMALS 11, (2021).
26. Nogalski, Z., Pogorzelska-Przybylek, P., Sobczuk-Szul, M. & Purwin, C. The effect of carcase conformation and fat cover scores (EUROP system) on the quality of meat from young bulls. ITALIAN JOURNAL OF ANIMAL SCIENCE 18, 615–620 (2019).
27. Kokoszynski, D. et al. Carcass characteristics and selected meat quality traits from commercial broiler chickens of different origin. ANIMAL SCIENCE JOURNAL 93, (2022).
28. Cortaga, C. Q., Lachica, J. A. P., Lantican, D. v. & Ocampo, E. T. M. Genome-wide SNP and InDel analysis of three Philippine mango species inferred from whole-genome sequencing. Journal of Genetic Engineering and Biotechnology 20, (2022).
29. Kumar, S. et al. Whole genome SNP identification and validation in Cucumis melo L. cultivars using genome resequencing approach. INDIAN JOURNAL OF GENETICS AND PLANT BREEDING 78, 478–486 (2018).
30. Thimmegowda, G. C. et al. Whole genome resequencing of tobacco (Nicotiana tabacum L.) genotypes and high-throughput SNP discovery. MOLECULAR BREEDING 38, (2018).
31. Wankhede, D. P., Aravind, J. & Mishra, S. P. Identification of Genic SNPs from ESTs and Effect of Non-synonymous SNP on Proteins in Pigeonpea. Proceedings of the National Academy of Sciences India Section B - Biological Sciences 89, 595–603 (2019).
32. Ma, G.-W. et al. Polymorphisms of FST gene and their association with wool quality traits in Chinese Merino sheep. PLOS ONE 12, (2017).
33. Adomako-Ankomah, Y., Wier, G. M., Borges, A. L., Wand, H. E. & Boyle, J. P. Differential locus expansion distinguishes Toxoplasmatinae species and closely related strains of Toxoplasma gondii. mBio 5, (2014).

Tables

Table 1. Basic nutrient of the longissimus dorsi and semitendinosus of different sheep breeds (% . Please note: Different letters in the same line indicate a significant difference (P<0.05), and the same letter indicates an insignificant difference (P>0.05).

Segment	Section	Longissimus dorsi		
	Type	CHR	WZMQ	XQ
Longissimus dorsi	Collagen	1.13±0.19^{a}	1.22±0.13^{a}	1.83±0.07^{b}
	Crude fat	5.63±0.24^{a}	4.06±0.94^{a}	10.16±0.87^{b}
	Water content	72.68±0.42^{a}	75.64±0.17^{b}	71.31±0.80^{a}
	Protein	20.02±0.39^{b}	20.10±0.95^{b}	18.08±0.83^{a}
Semitendinosus	Collagen	1.28±0.12^{ab}	0.96±0.21^{a}	1.53±0.16^{b}
	Crude fat	6.50±0.29^{c}	3.05±0.55^{a}	5.38±0.61^{b}
	Water content	72.74±0.24^{a}	74.81±0.42^{b}	73.44±0.89^{a}
	Protein	18.88±0.59^{a}	20.02±0.25^{a}	20.01±0.97^{a}

Table 2. Sequencing accusation results

Breed	Raw Reads	Valid Reads	Valid (%)	Q20 (%)	Q30 (%)	Mapped (%)	Mean Coverage
CHR	218607445	215357112	97.73	96.93	92.01	98.42	11.32
WZMQ	222366292	222359073	99.99	98.04	94.25	98.70	11.67
XQ	193845116	193820334	99.99	97.23	92.83	98.22	9.64

Figures

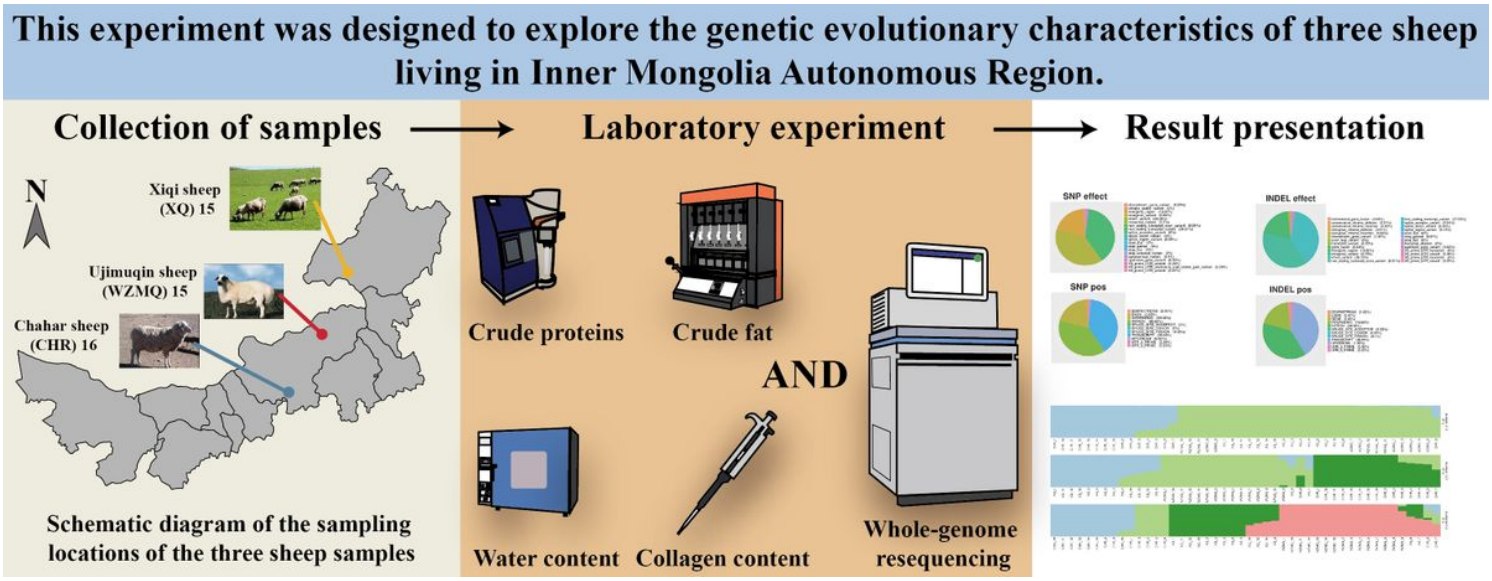

Figure 1

Brief flow chart of this work

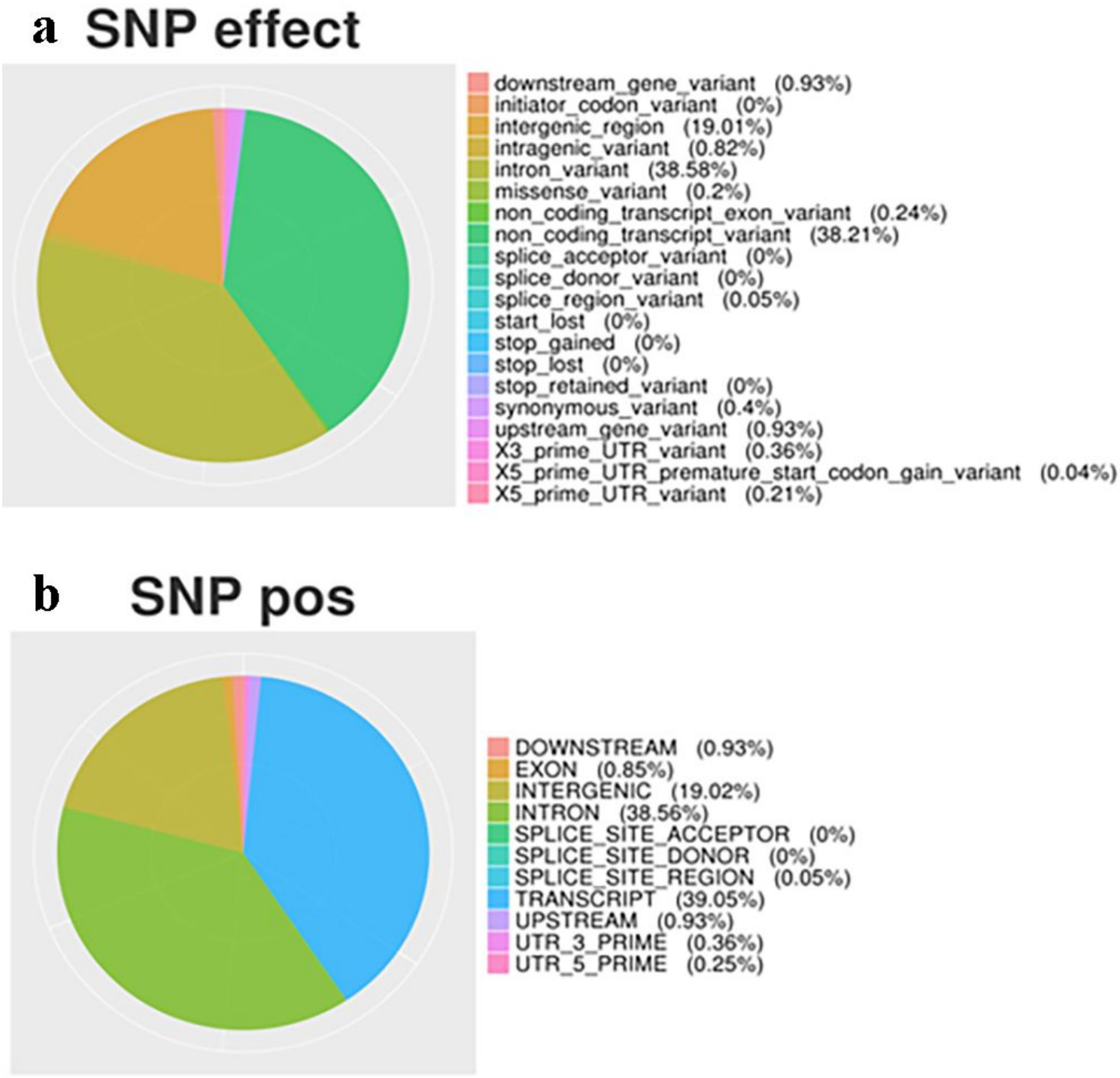

Figure 2

SNP diagram. (a) The distribution of the proportion of SNPs with different functions; (b) The proportion of SNPs in different regions of the genome.

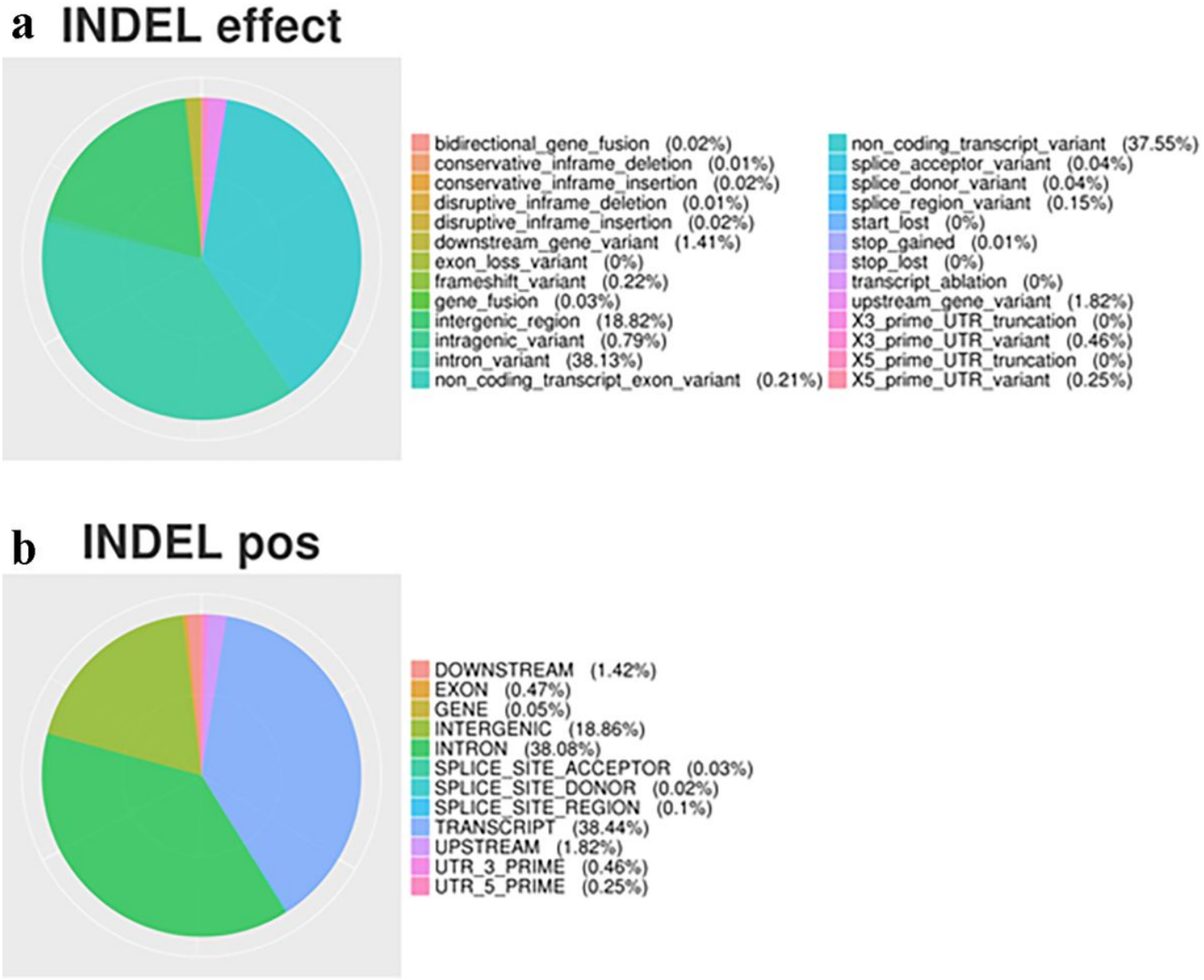

Figure 3

INDEL diagram. (a) The distribution of INDEL ratios for different functions; (b) The INDEL ratios in different regions of the genome.

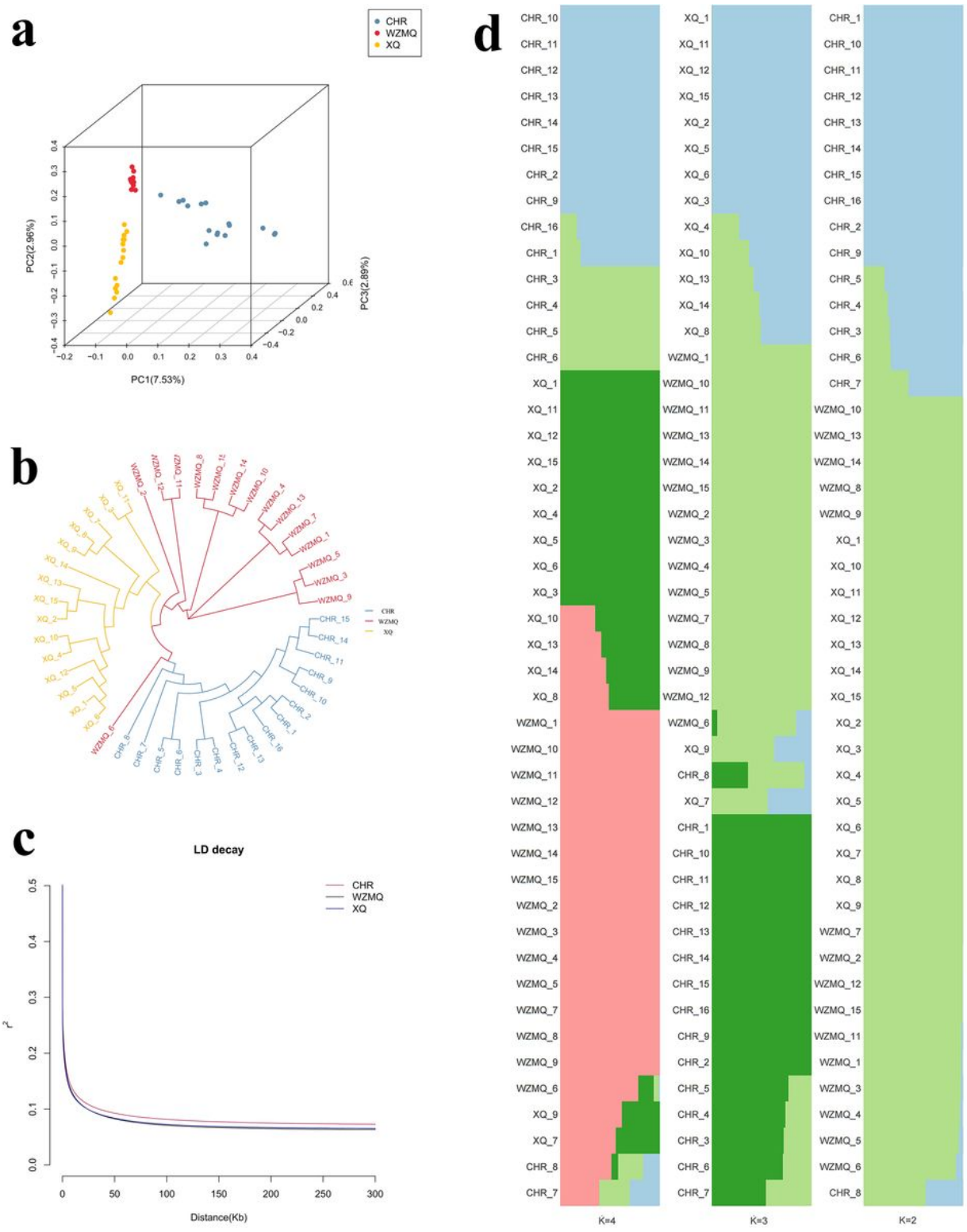

Figure 4

Pictures of the evolution and development of three kinds of sheep. (a) PCA plot (PC1, PC2 and PC3) showing the genetic structure of the 46 sheep individuals. The degrees of explained variance are given in parentheses; (b) Neighbor-joining tree of the 46 sheep individuals; (c) Decay map of linkage disequilibrium measured by r^2 with marker distances; (d) Admixture plot (K=2, 3, 4) for the 46 sheep individuals. Each individual is shown as a vertical bar.

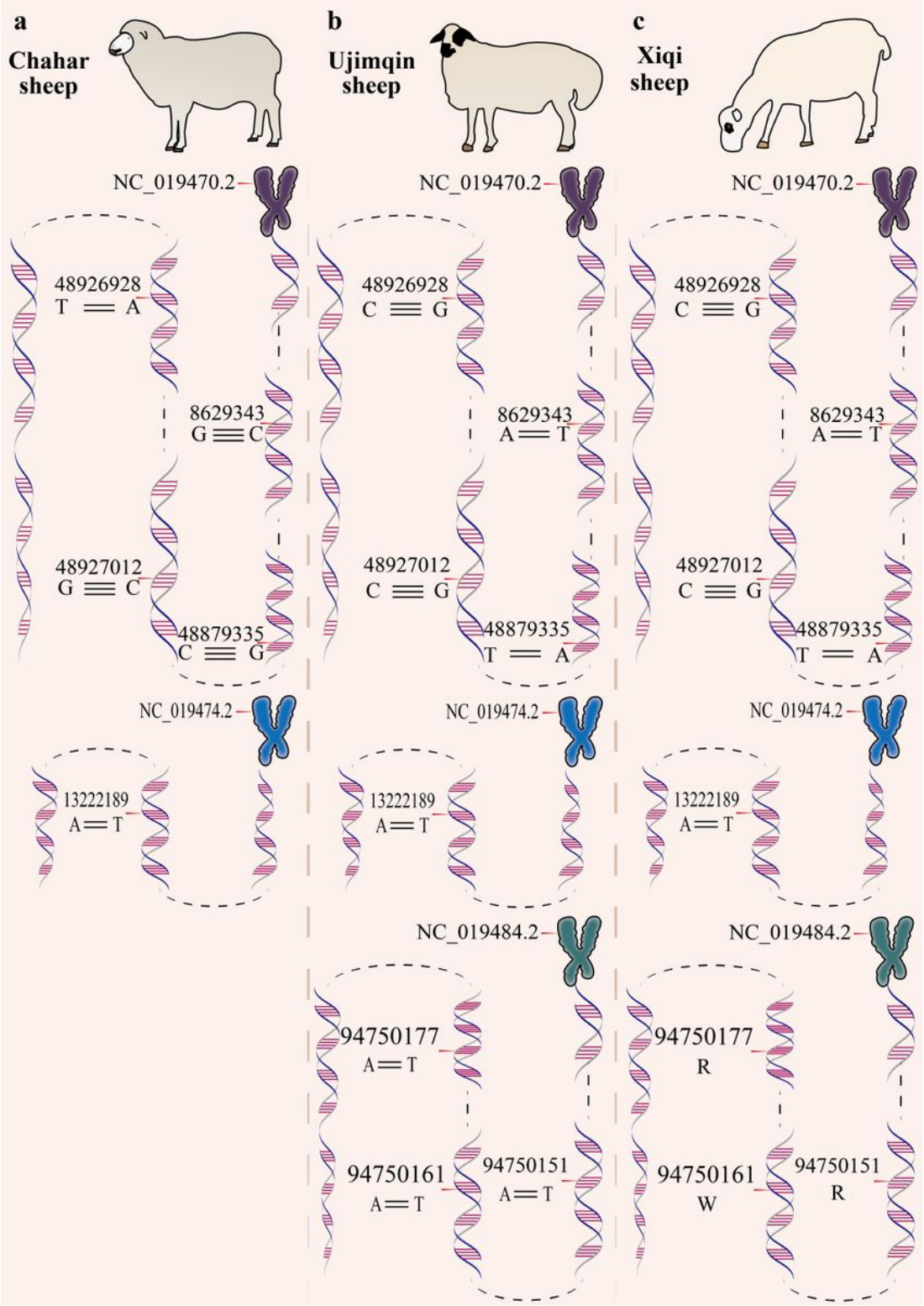

Figure 5

Chromosomal loci differences of three sheep. (a) Base Differences on Chromosomes in Chahar sheep; (b) Base Differences on Chromosomes in Ujimqin sheep; (c) Base Differences on Chromosomes in Xiqi sheep. Because there is no difference between Ujimqin sheep and Xiqi sheep on NC_019470.2 and NC_019474.2 chromosomes, their difference bases on NC_019484.2 chromosomes were detected.

Supplementary Files

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

- SupplementaryTableS1.xlsx
- SupplementaryTableS2.xlsx
- SupplementaryTableS3.xlsx
- SupplementaryTableS4.xlsx
- Supplementaryfigures.docx