



OPEN ACCESS

EDITED BY

Ran Di,
Chinese Academy of Agricultural
Sciences, China

REVIEWED BY

He Shen,
University at Buffalo, United States
Lyu Fenghua,
China Agricultural University, China

*CORRESPONDENCE

Xiaoxue Zhang
zhangxx@gsau.edu.cn

SPECIALTY SECTION

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Veterinary Science

RECEIVED 01 September 2022

ACCEPTED 30 September 2022

PUBLISHED 18 October 2022

CITATION

Li X, Yuan L, Wang W, Zhang D, Zhao Y,
Chen J, Xu D, Zhao L, Li F and Zhang X
(2022) Whole genome re-sequencing
reveals artificial and natural selection
for milk traits in East Friesian sheep.
Front. Vet. Sci. 9:1034211.
doi: 10.3389/fvets.2022.1034211

COPYRIGHT

© 2022 Li, Yuan, Wang, Zhang, Zhao,
Chen, Xu, Zhao, Li and Zhang. This is
an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction
in other forums is permitted, provided
the original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Whole genome re-sequencing reveals artificial and natural selection for milk traits in East Friesian sheep

Xiaolong Li¹, Lv Feng Yuan², Weimin Wang³, Deyin Zhang³, Yuan Zhao³, Jiangbo Chen¹, Dan Xu¹, Liming Zhao¹, Fadi Li³ and Xiaoxue Zhang^{1*}

¹College of Animal Science and Technology, Gansu Agricultural University, Lanzhou, China,

²Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou, China,

³The State Key Laboratory of Grassland Agro-ecosystems, College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, China

The East Friesian sheep is one of the important high-yielding dairy sheep breeds, but still little is known about their genetic and genomic variation during domestication. Therefore, we analyzed the genomic data of 46 sheep with the aim of identifying candidate genes that are closely related to milk production traits. Our genomic data consisted of 20 East Friesian sheep and 26 Asian Mouflon wild sheep. Finally, a total of 32590241 SNPs were identified, of which 0.61% (198277) SNPs were located in exonic regions. After further screening, 122 shared genomic regions in the top 1% of F_{ST} and top 1% of Nucleotide diversity ratio were obtained. After genome annotation, these 122 candidate genomic regions were found to contain a total of 184 candidate genes. Finally, the results of KEGG enrichment analysis showed four significantly enriched pathways ($P < 0.05$): beta-Alanine metabolism (*SMOX*, *HIBCH*), Pathways in cancer (*GLI2*, *AR*, *TXNRD3*, *TRAF3*, *FGF16*), Non-homologous end-joining (*MRE11*), Epstein-Barr virus infection (*TRAF3*, *PSMD13*, *SIN3A*). Finally, we identified four important KEGG enrichment pathways and 10 candidate genes that are closely related to milk production in East Friesian sheep. These results provide valuable candidate genes for the study of milk production traits in East Friesian sheep and lay an important foundation for the study of milk production traits.

KEYWORDS

sheep, whole-genome resequencing, F_{ST} , PI, sheep milk

Introduction

In recent years, the rapid development of next-generation sequencing (NGS) technology has provided important techniques and tools for the analysis of genetic traits. This development provides new opportunities to study economically important traits in animals (1–3). Recently, a number of population-related resequencing data studies

have emerged that aim to identify important candidate genes for specific populations and specific traits. For example, For example, the identification of candidate genes and genomic intervals for heat-adapted traits in rainbow trout based on whole-genome resequencing data (4), the identification of candidate genes for litter size in goats (5), the characterization of lumbar vertebrae in sheep (6), and the study of candidate genes for adaptive capacity of yak populations to extreme environments (7). As well as a large number of studies that have used whole-genome resequencing data to explore population genomic features and candidate genes in different species, such as pigs (8), dogs (9), chickens (10), and ducks (11).

The East Friesian sheep (EFS) is one of the most productive dairy sheep breeds, providing dairy products and by-products needed for daily human life. The East Friesian sheep lactation lasts about 230 days and produces 500–700 kg of milk in a single lactation (12). East Friesian sheep have been imported to many countries and are often used to improve native sheep in various countries due to their good milk and meat production performance (13). Although there have been reports on the use of East Friesian sheep as a research subject. For example, studies on wild and domesticated sheep (14), the effect of prolactin, β -lactoglobulin and kappa-casein genotypes on milk production in East Friesian sheep (15), studies on milk composition of different goat breeds (16), etc. However, little is known about the genetic and genomic variation associated with the East Friesian sheep population. To identify candidate genes for milk production traits in the East Friesian population more clearly, we used resequencing data from 20 East Friesian and 26 Asian Mouflon wild sheep (WS) to identify candidate genomic regions and candidate genes associated with milk production traits. The data included our own sequenced genomic data of 10 East Friesian sheep and genomic data of 10 East Friesian and 26 Asian Mouflon wild sheep from the NCBI SRA database (<https://www.ncbi.nlm.nih.gov/sra>). During the analysis, we used methods such as, phylogenetic tree analysis, population genetic structure analysis, principal component analysis (PCA), F_{ST} , nucleotide diversity (π) and Kyoto Encyclopedia of Genes and Genomes (KEGG).

Materials and methods

Ethics statement

All experiments in this study were carried out in accordance with the approved guidelines from the Regulation of the Standing Committee of Gansu People's Congress. All experimental protocols, including the sample collection protocol, were approved by the Ethics Committee of Gansu Agriculture University (China) under permission no. DK-005.

Samples and resequencing

We selected 10 East Friesian sheep (EFS) from Beijing Aoxin Animal Husbandry Co., LTD., China, for whole-genome resequencing (Supplementary Table S1) (17). The 5 ml blood sample was collected from the jugular vein of sheep, and the DNA was extracted by easy pure blood genomic DNA kit (TransGen biotech, Beijing, China), and then sequenced on Illumina nova seq 6000 (Illumina, San Diego, ca, United States). We also collected the resequencing data of 10 EFS and 26 Asian mouflon wild sheep (WS), which came from NCBI SRA database (<https://www.ncbi.nlm.nih.gov/sra>; Supplementary Table S3). Therefore, the resequencing data of 46 sheep were used for subsequent analysis.

Alignment and SNP calling

The main data processing process is as follows: (1) Use Fastp (version: 0.12.4) software to filter the raw data with parameters: `fastp --in1 A_1.fastq --in2 A_2.fastq --out1 A_clean_1.fastq --out2 A_clean_2.fastq -h A_fastp.html -j A_fastp.json` (18). (2) Use BWA (Burrows-Wheeler Aligner, version: 0.7.15) software to match the filtered raw data to the reference genome (Oar_rambouillet_v1.0) with the parameters: `bwa mem -R '@RG/tID:A/tSM:A' sheep1.0.fa A_clean_1.fastq A_clean_2.fastq > A.sam` (19). (3) Use samtools (version: 1.10) software to first convert sam files into bam files, and then sort them by chromosome and locus with the following parameters: `samtools view -b A.sam > A.bam; samtools sort -O BAM A.bam > A.sorted.bam` (20). (4) Use samtools software to establish the index of the first Bam file, and then mark duplicates, parameters: `samtools index A.sorted.bam; gatk MarkDuplicates -I A.sorted.bam -O A.dedup.bam -M A.mark_dup_metrics.txt; samtools index A.dedup.bam; samtools sheep1.0.fa; gatk CreateSequenceDictionary -R sheep1.0.fa`. (5) Use GATK software to generate the gvcf file for each sample with the following parameters: `gatk HaplotypeCaller -R sheep1.0.fa -I A.dedup.bam -O A.raw.gvcf -ERC GVCF -ploidy 2 -contamination 0 -G StandardAnnotation -G StandardHCAnnotation -G AS_StandardAnnotation -GQB 10 -GQB 20 -GQB 30 -GQB 40 -GQB 50 -GQB 60 -GQB 70 -GQB 80 -GQB 90` (21). (6) Use GATK software to merge all sample gvcf files into one gvcf file, parameter: `gatk CombineGVCFs -R sheep1.0.fa -O combine_variants.raw.gvcf --variant A.raw.gvcf --variant B.raw.gvcf.....` (7) Use GATK software to generate the original vcf file from the gvcf file with the following parameters: `gatk GenotypeGVCFs -R sheep1.0.fa -O combine_variants.raw.vcf --variant combine_variants.raw.gvcf`. (8) Use GATK software to generate SNP-only vcf files, parameter: `gatk SelectVariants -R sheep1.0.fa -O combine_SNP.raw.vcf --variant combine_variants.raw.vcf --select-type-to-include SNP`. (9) Use Vcftools software to filter

vcf files, parameter: --remove-indels --min-allsles 2 --max-alleles 2 --minDP 4 --maxaDP 100 --minGQ 10 --minQ 30 --max-missing 0.6 --maf 0.05 (22). Finally, the ANNOVAR program (23) was used to annotate the filtered SNPs based on the sheep reference genome (Oar_rambouillet_v1.0), and finally the SNPs are divided into: exonic regions (variant overlaps a coding), intronic regions (variant overlaps an intron), upstream egions, downstream regions, and intergenic regions (variant is in the intergenic region).

Phylogenetic and principal component analysis

A total of 32590241 high-quality SNPs were used for principal component analysis and phylogenetic tree construction. Use VCFtools (version: 0.1.16) to generate PLINK format files (22). Then, the PLINK (version: 1.9) program was used to generate principal component analysis (--out PCA_file) and phylogenetic tree (--distance --out tree_file) files (24). The phylogenetic tree uses the neighbor.exe subroutine in the phylip-3.698 program to generate a FigTree format file (<https://evolution.genetics.washington.edu/phylip.html>). Finally, FigTree was used to visualize the result file(<http://tree.bio.ed.ac.uk/software/figtree/>).

Nucleotide diversity and selective sweep

We used the VCFtools program to analyze nucleotide diversity (θ_π) and fixation index (F_{ST} ; (25)). By comparing the differentiation index and nucleotide diversity of the EFS population and the WS population, the selected genomic intervals of the EFS population related to lactation traits were screened. We used VCFtools to calculate the F_{ST} value and θ_π value of the genome interval, the parameters:--fst-window-size 150000 --fst-window-step 75000, and calculate the θ_π ratio ($\theta_{\pi WS} / \theta_{\pi EFS}$) (26). For the presentation of the theta π ratio, we use the $y = \log_2^x$ function to transform it and finally display it in the picture. In the two analysis methods of F_{ST} and θ_π ratio, the threshold line of each individual analysis is top 1%. Finally, visualization is performed using the R language packages ggplot2 (27) and ggExtra (28).

Candidate interval and function annotation analysis

In order to find the genomic interval closely related to the lactation traits of the EFS population, we screened the interval shared by the top1% of the F_{ST} top1% and the θ_π ratio top1% of the genomic interval as candidate regions. ANNOVAR was used to annotate the SNPs in these regions to genes (23),

and then Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was performed on these candidate genes using KOBAS (29).

Results

Genomic variation

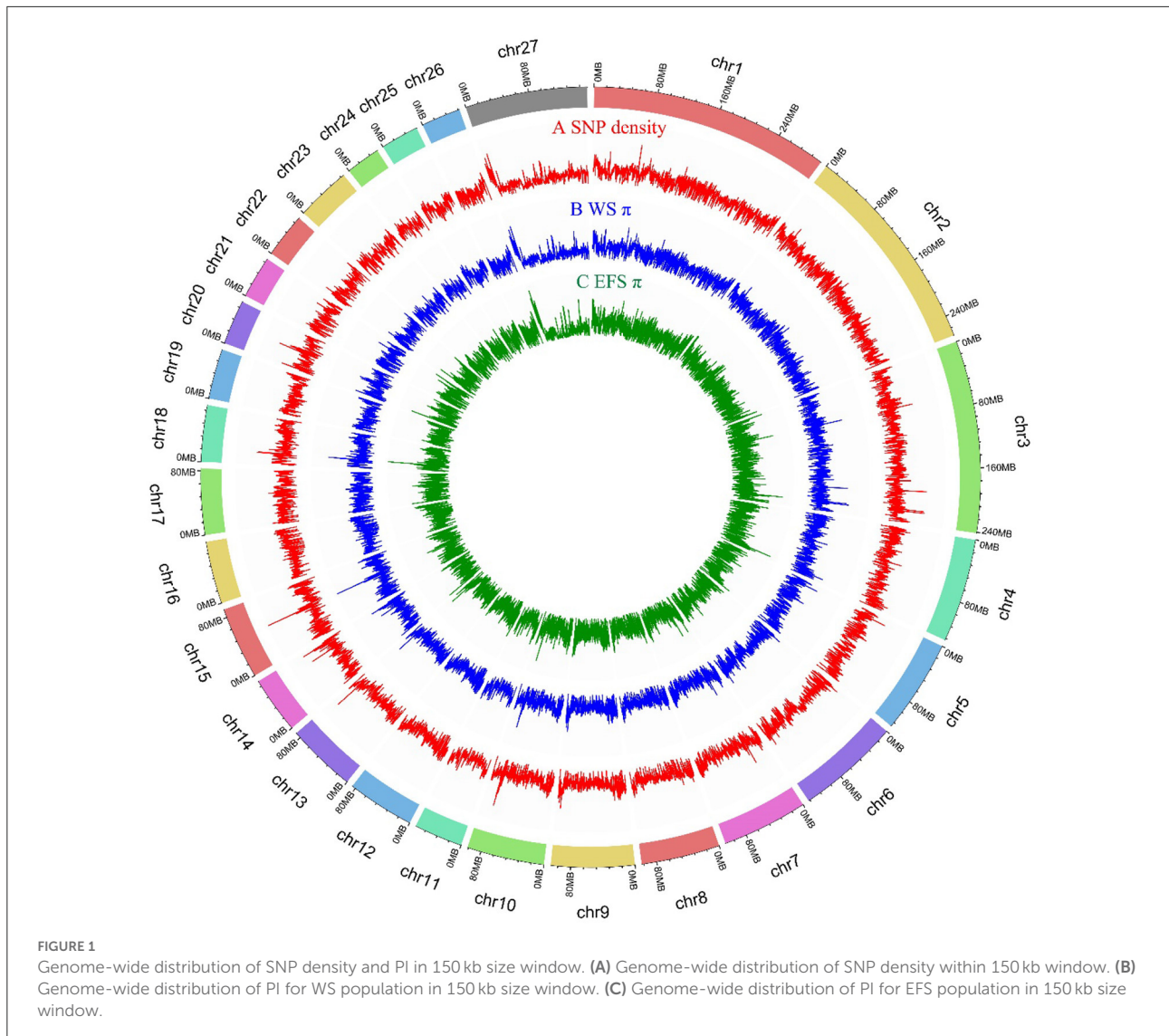
We generated whole genome data for 10 EFS using Illumina NovaSeq 6000 instrument, and the statistical results revealed a total of 285.96 Gb of raw data, with an average sequencing depth of 10.6x (Supplementary Table S2). The data used in this study also included whole-genome sequencing data of 26 WS and 10 EFS from the NCBI database. Thus, the study used resequencing data from 46 sheep with a total data volume of 1.56 Tb. After rigorous matching and filtering, a total of 32590241 nucleotide polymorphisms were identified (Supplementary Table S4), of which 0.61% of SNP mutations were located in exonic regions ($N = 198227$), and the remaining SNP loci were mainly distributed in intergenic and intronic regions (94.41% of (Supplementary Table S5). The genome-wide 150 kb SNPs and PIs are depicted in Figure 1.

Phylogenetic analysis and nucleotide diversity

Principal component analysis (PCA) was performed using the SNP dataset obtained above. The results showed that the first PCA axis clearly separated the EFS population from the WS population (Figure 2A). We then investigated the phylogenetic relationships between the EFS population and the WS population (Figure 2B). The results showed that the distance-based Neighbor-Joining tree was consistent with the results of the PCA analysis, both dividing the populations into the EFS and WS groups. Finally, similar results emerged in the analysis of the population genetic structure (Figure 2C). After calculating nucleotide diversity within a 150-kb window (step size 75 kb) for both groups (Supplementary Tables S6, S7), the EFS group (0.001250) showed lower nucleotide diversity than the WS group (0.003242) (Figures 2D,E).

Genome-wide selection scans

We calculated the fixed differentiation index (F_{ST} ; Supplementary Table S8) and the ratio of WS population to EFS nucleotide diversity θ_π ($\theta_{\pi WS} / \theta_{\pi EFS}$) within a 150-kb window (step size 75 kb) (Supplementary Table S9), aiming to screen for positive selection windows associated with lactation traits. The top 1% window of each test performed was used as a candidate window (Figure 3). To further search



for possible positive selection windows, we screened the shared windows between the top 1% of the F_{ST} and the top 1% of the $\theta\pi$ ratio as the final candidate windows. Finally, a total of 122 shared genomic regions were obtained (Figure 4, Supplementary Table S10). By annotation with ANNOVAR software, we obtained a total of 184 candidate genes from these 122 regions (Supplementary Table S11).

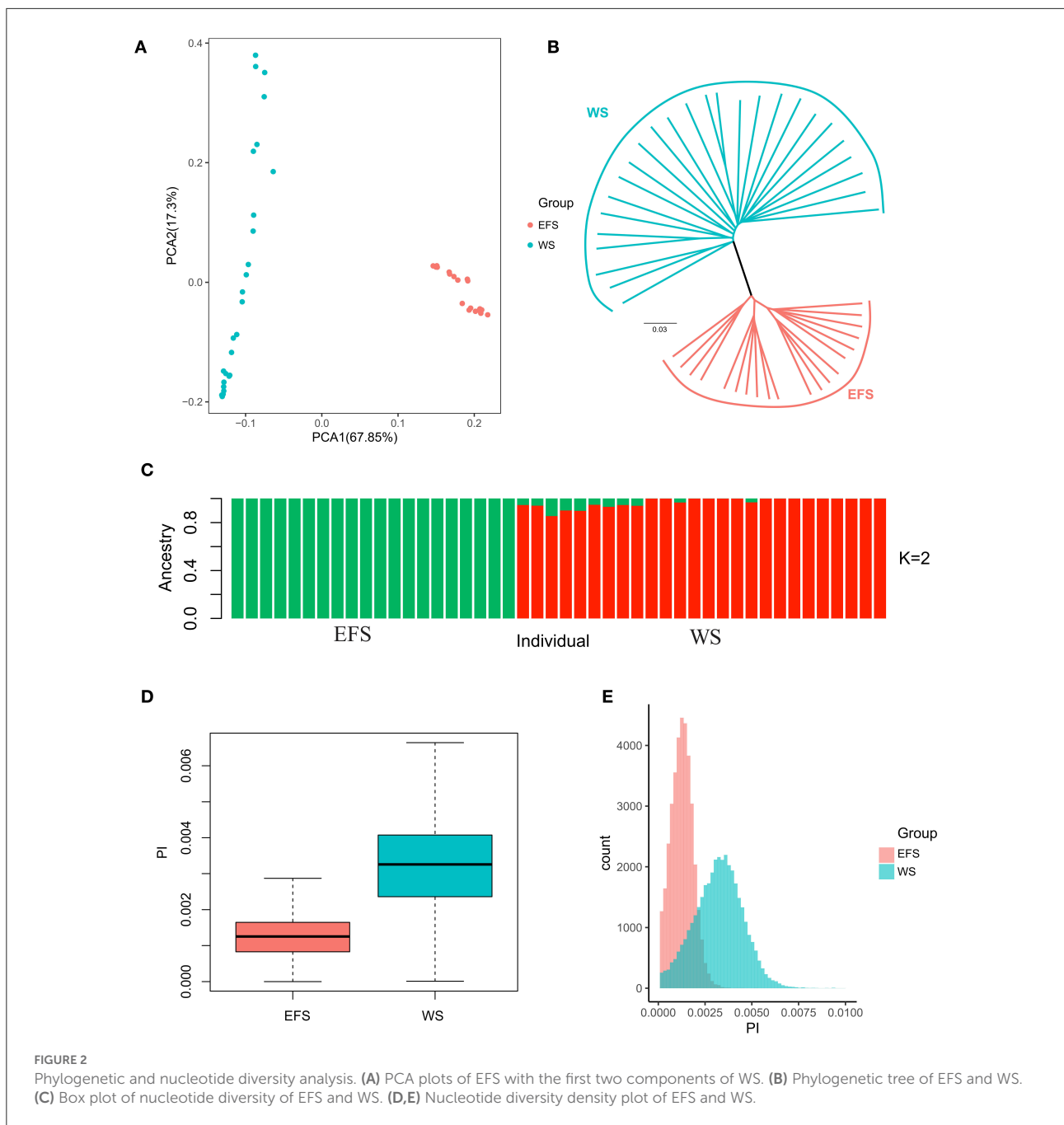
Candidate genes and KEGG

The SNPs in the 122 candidate windows obtained above were annotated to genes using ANNOVAR, and a total of 184 candidate genes were obtained (Supplementary Table S11). Included among these candidate genes are those screened for domestication-related candidates in previous reports

(Figure 5, Supplementary Table S12). Then, these 184 genes were subjected to KEGG pathway enrichment analysis (Supplementary Table S13), and the top 20 enrichment pathways are shown in (Figure 6, Supplementary Table S14). Finally, four significant KEGG enrichment pathways were identified ($P < 0.05$, Table 1): beta-Alanine metabolism (*SMOX*, *HIBCH*), Pathways in cancer (*GLI2*, *AR*, *TXNRD3*, *TRAF3*, *FGF16*), Non-homologous end-joining (*MRE11*), Epstein-Barr virus infection (*TRAF3*, *PSMD13*, *SIN3A*).

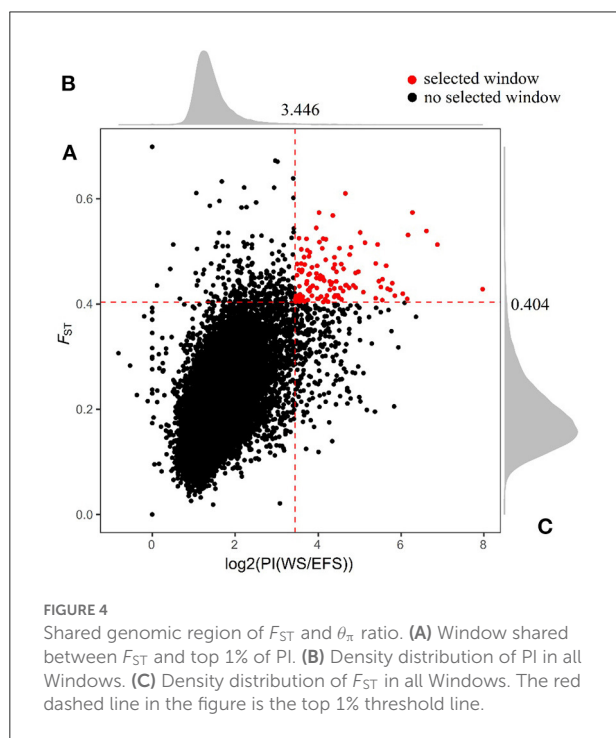
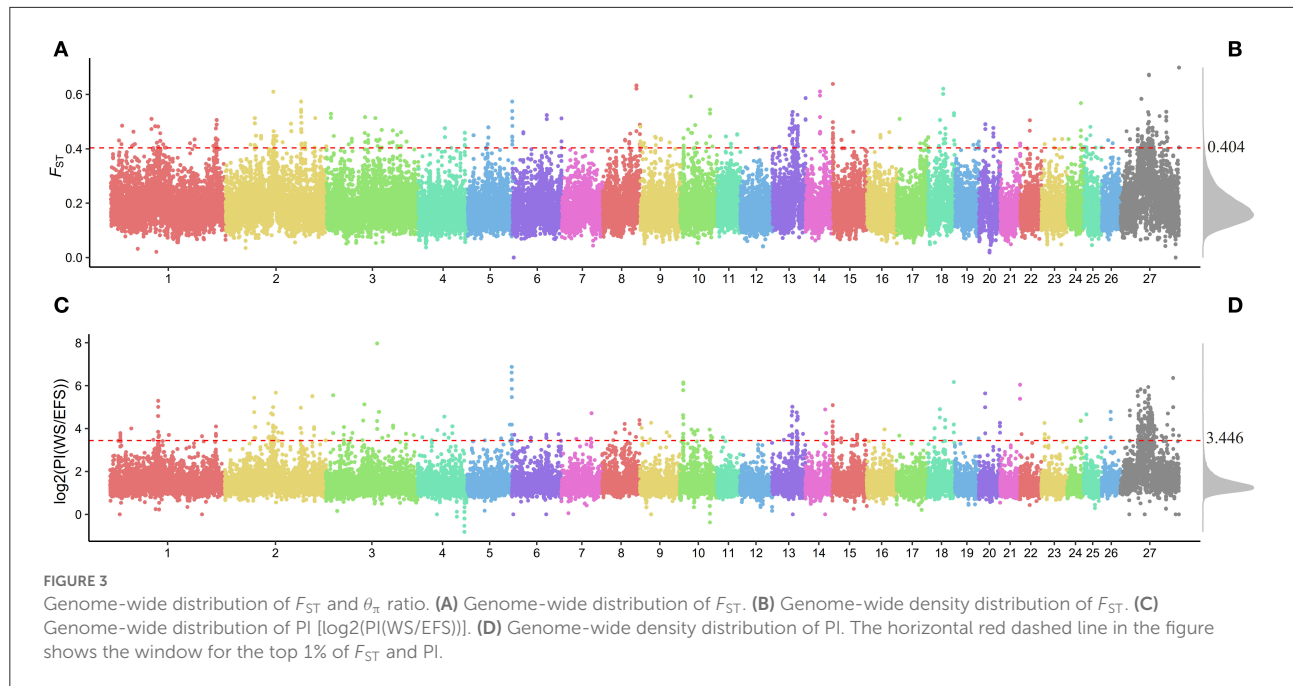
Discussion

In this study, we performed whole-genome resequencing of 10 East Friesian sheep (EFS) and used whole-genome resequencing data of 10 East Friesian sheep (EFS) and 26 Asian mouflon wild sheep (WS) from the NCBI SRA database to



develop subsequent analyses. We used a method similar to the previous study (30), compared the sequencing data to the sheep reference genome, and used various filtering methods (deep filtering, deletion rate and minimum allele frequency, etc.) to call the individual sheep SNP, to ensure that high-quality SNPs data are finally obtained. We identified a total of 32590241 high-quality SNPs from the use of genomic data, of which 0.61% (198277) SNPs are located in the exon region, which helps to reveal the genomic characteristics of the WS population and the EFS population.

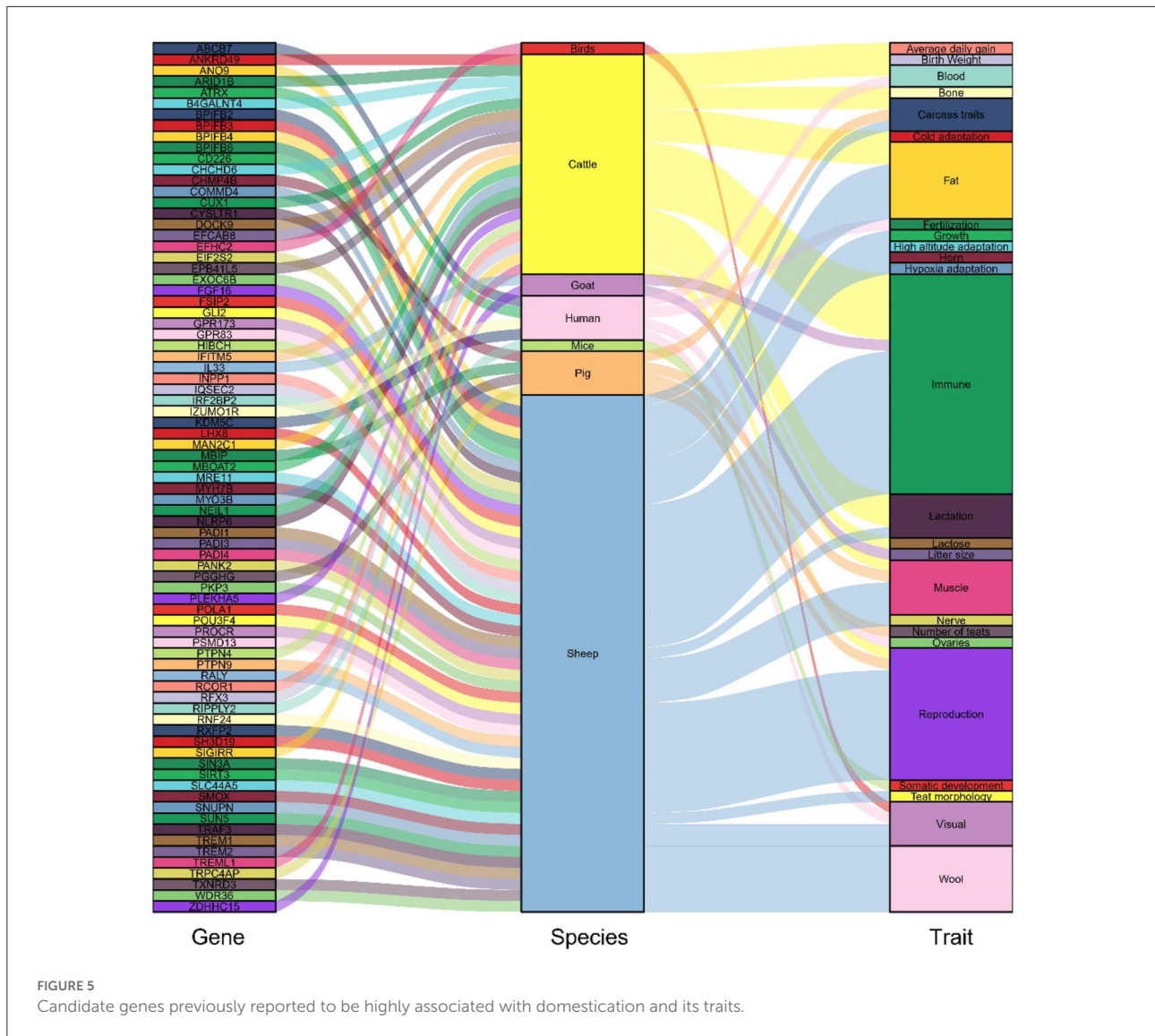
The population genetic structure of EFS and WS was first analyzed, and the statistical results revealed that in the PCA analysis, PCA1 clearly divided the EFS and WS populations into 2 Clusters (Figure 2A). Immediately afterwards, in the phylogenetic tree (Figures 2B,C), the EFS and WS populations showed a similar classification pattern as in the principal component analysis, and this clustering pattern also appeared in previous studies (14). The nucleotide diversity of the two populations was found to be lower in the EFS population than in the WS population (Figures 2D,E), with an average nucleotide



diversity (θ_{π}) of 0.001250 in the EFS population and 0.003221 in the WS, which is consistent with previous studies that wild sheep have higher nucleotide diversity than other domesticated sheep breeds (14). This may be related to the complexity of the living environment of the Asiatic mouflon sheep, for example, the Asiatic mouflon sheep is one of the ancestors of the domesticated

sheep, forced to forage, avoid natural enemies, adaptability of the environment and other pressures, the genome structure is diversified, and it produces a variety of proteins, only in this way can the Asiatic mouflon survive in the cruel environment of the wild, so the Asiatic mouflon shows a higher nucleotide diversity compared to the East Friesian sheep.

After screening the top 1% of genomic regions for F_{ST} (Figure 3A) and Pi ratio (Figure 3B), we then screened the genomic regions shared in these two methods and finally obtained 122 shared genomic regions (Figure 4). Among these 122 regions, we obtained a total of 184 candidate genes by annotation (Supplementary Table S11). Among these genes, we identified candidate genes that have been reported to be associated with economically important traits in animals (Figure 5, Supplementary Table S12). For example, in the reported studies in sheep, most of the genes were identified as important candidates associated with traits such as immunity, reproduction, vision, lipid synthesis, lactation, muscle differentiation, bone, blood, tail fat, wool and carcass. As well as in reported studies in cattle, pigs, goats, mice and birds, candidate genes associated with economically important traits such as daily weight gain, immunity, lactation, primordial weight, lactose synthesis, fat production, lambing numbers, muscle development, blood and cold environment adaptations also appeared in the results of this study. This demonstrates that the domestication and artificial selection of East Friesian dairy sheep is accompanied by simultaneous selection for multiple traits, and that the gradual increase in milk production is accompanied by simultaneous increases in traits such as immunity, neurodevelopment and metabolic levels.



Candidate genes previously reported to be associated with production traits appeared in the results of this study. For example, candidate genes related to lactation traits: *DOCK9*, *ARID1B*, *B4GALNT4*, *RCOR1*, *SH3D19*. The *DOCK9* gene was identified in a study on the impact of liver expression data on future lactation performance, showing that the *DOCK9* gene was differentially expressed and predicted to have an impact on the computational model of lactation performance (31). As well, *ARID1B* was identified as a significant differentially expressed gene in two groups of Holstein mammary glands with very high and very low percentages of milk protein and fat (32). The candidate genes *B4GALNT4* (33) involved in lactose synthesis, *RCOR1* (34) related to milk yield and *SH3D19* related to lactation process in sheep were also identified (35). In addition, our results contain a large number of candidate genes related to production traits, such as reproductive trait candidates: *FSIP2*

(17), *CD226* (36), *SLC44A5* (37), *GPR83* (38), *SUN5* (39), *GPR173* (40), *LHX8* (41), *SIRT3* (42), *FGF16* (43), *POU3F4* (44), these genes affect reproductive traits by influencing days of estrus, luteal gland, participating in reproductive regulation, and hypothalamic-pituitary-gonadal axis regulation. For example, the candidate genes for wool traits: *RALY* (45), *GLI2* (46), *TXNRD3* (47), *EIF2S2* (45), *IRF2BP2* (48), etc.

Finally, the results of the KEGG enrichment analysis of genes in 122 regions showed four significantly enriched pathways ($p < 0.05$, Figure 6): beta-Alanine metabolism (*SMOX*, *HIBCH*), pathways in cancer (*GLI2*, *AR*, *TXNRD3*, *TRAF3*, *FGF16*), non-homologous end-joining (*MRE11*), Epstein-Barr virus infection (*TRAF3*, *PSMD13*, *SIN3A*). Therefore, we predict that these genes play an important role in lactation. For example, the *GLI2* gene was identified as a candidate gene for the Chronic subclinical mastitis trait in dairy cows, a disease that, when

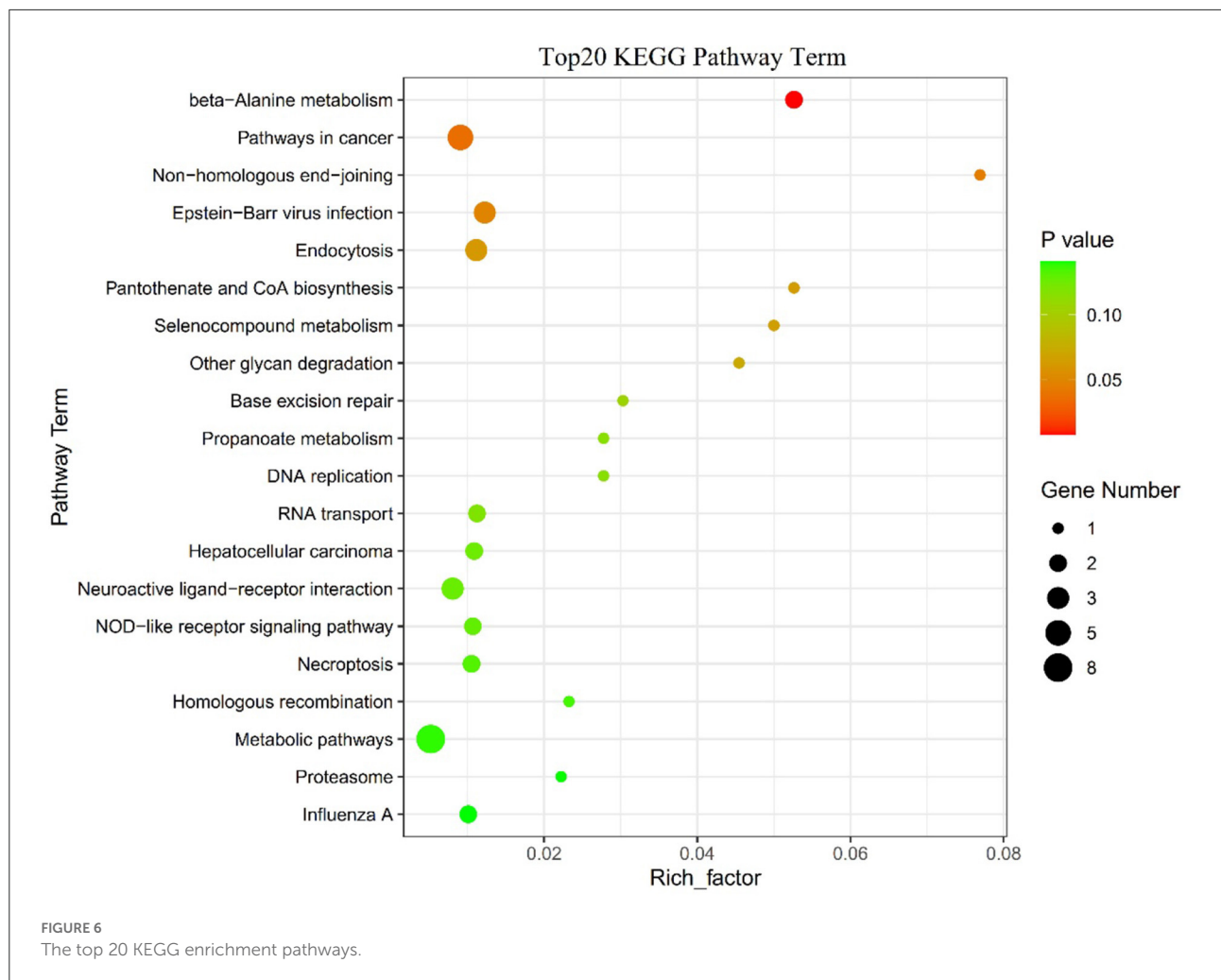


TABLE 1 Four significantly enriched KEGG pathways.

Related pathways	P-value	Related genes
beta-Alanine metabolism	0.00774	<i>SMOX, HIBCH</i>
Pathways in cancer	0.03702	<i>GLI2, AR, TXNRD3, TRAF3, FGF16</i>
Non-homologous end-joining	0.04524	<i>MRE11</i>
Epstein-Barr virus infection	0.04902	<i>TRAF3, PSMD13, SIN3A</i>

it occurs, can reduce milk production or lead to culling of cows (49). The *TRAF3* gene is essential for immune function of mammary epithelial cells in cows (50). The *FGF16* gene was identified as differentially expressed in endometrial glandular, luminal, and stromal cells of cows with persistent inflammation and recovery from postpartum endometritis, demonstrating that the *FGF16* gene is a potential candidate gene to promote the transition from an inflammatory to a healthy state in postpartum dairy cows (51). As well, in a genome-wide association analysis

of dairy cows, *SIN3A* was identified as a candidate gene for individual laboratory cheese yield and protein percentage traits in milk traits (52). Although some of these genes have been reported in dairy cattle studies, the biological functions of these candidate genes during milk production in sheep still need further validation.

Conclusions

This is an analysis of genome-wide data from the East Friesian dairy sheep breed. We identified four important KEGG enrichment pathways and 10 candidate genes that are closely associated with milk yield in East Friesian sheep: beta-Alanine metabolism (*SMOX, HIBCH*), pathways in cancer (*GLI2, AR, TXNRD3, TRAF3, FGF16*), non-homologous end-joining (*MRE11*), Epstein-Barr virus infection (*TRAF3, PSMD13, SIN3A*). Some of these genes were identified as candidates for milk yield traits in dairy cattle studies, but the biological functions that these candidate genes play in milk production in

sheep still need further validation. In conclusion, these results provide valuable candidate genes for the study of milk yield traits in East Friesian sheep and lay an important foundation for the study of milk yield traits.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/>, SAMN22870744 - SAMN22870753 (17).

Ethics statement

The animal study was reviewed and approved by Ethics Committee of Gansu Agriculture University (China) under permission no. DK-005. Written informed consent was obtained from the owners for the participation of their animals in this study.

Author contributions

FDL, XXZ and WMW designed the experiments. XLL, XXZ, LFY and WMW analyzed the data. XLL wrote the manuscript. DYZ, YZ, JBC, DX and LMZ contributed to sample collection and prepared biological samples. XXZ, WMW and XLL revised the manuscript. All authors read and approved the final manuscript.

References

1. Wu D, Liang Z, Yan T, Xu Y, Xuan L, Tang J, et al. Whole-Genome resequencing of a worldwide collection of rapeseed accessions reveals the genetic basis of ecotype divergence. *Mol Plant*. (2019) 12:30–43. doi: 10.1016/j.molp.2018.11.007
2. Liu X, Zhang Y, Li Y, Pan J, Wang D, Chen W, et al. *Epas1* gain-of-function mutation contributes to high-altitude adaptation in tibetan horses. *Mol Biol Evol*. (2019) 36:2591–603. doi: 10.1093/molbev/msz158
3. Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, et al. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol*. (2012) 10:e1001258. doi: 10.1371/journal.pbio.1001258
4. Chen Z, Narum SR. Whole genome resequencing reveals genomic regions associated with thermal adaptation in redband trout. *Mol Ecol*. (2021) 30:162–74. doi: 10.1111/mec.15717
5. Wang K, Liu X, Qi T, Hui Y, Yan H, Qu L, et al. Whole-genome sequencing to identify candidate genes for litter size and to uncover the variant function in goats (*capra hircus*). *Genomics*. (2021) 113:142–50. doi: 10.1016/j.ygeno.2020.11.024
6. Li C, Liu K, Dai J, Li X, Liu X, Ni W, et al. Whole-genome resequencing to investigate the determinants of the multi-lumbar vertebrae trait in sheep. *Gene*. (2021) 30:146020. doi: 10.1016/j.gene.2021.146020
7. Lan D, Ji W, Xiong X, Liang Q, Yao W, Mipam T-D, et al. Population genome of the newly discovered jinchuan yak to understand its adaptive evolution in extreme environments and generation mechanism of the multirib trait. *Integr Zool*. (2021) 16:685–95. doi: 10.1111/1749-4877.12484

Funding

This work was supported by the Earmarked Found for National Key R&D Program of China (2021YFD1300901), National Natural Science Foundation of China (31960653) and National Joint Research on improved breeds of Livestock and Poultry [grant no.19210365].

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fvets.2022.1034211/full#supplementary-material>

8. Li W-T, Zhang M-M, Li Q-G, Tang H, Zhang L-F, Wang K-J, et al. Whole-genome resequencing reveals candidate mutations for pig prolificacy. *Proc Biol Sci*. (2017) 284:2437. doi: 10.1098/rspb.2017.2437
9. Amiri Ghanatsaman Z, Wang G-D, Asadollahpour Nanaei H, Asadi Fozi M, Peng M-S, Esmailzadeh A, et al. Whole genome resequencing of the iranian native dogs and wolves to unravel variome during dog domestication. *BMC Genomics*. (2020) 21:207. doi: 10.1186/s12864-020-6619-8
10. Lawal RA, Al-Atiyat RM, Aljumaah RS, Silva P, Mwacharo JM, Hanotte O. Whole-genome resequencing of red junglefowl and indigenous village chicken reveal new insights on the genome dynamics of the species. *Front Genet*. (2018) 9:264. doi: 10.3389/fgene.2018.00264
11. Zhang Z, Jia Y, Almeida P, Mank JE, van Tuinen M, Wang Q, et al. Whole-Genome Resequencing Reveals Signatures of Selection and Timing of Duck Domestication. *Gigascience*. (2018) 7:giy027. doi: 10.1093/gigascience/giy027
12. Nguyen QV, Le VH, Nguyen DV, Malau-Aduli BS, Nichols PD, Malau-Aduli AE. Supplementing grazing dairy ewes with plant-derived oil and rumen-protected epa+ dha pellets enhances health-beneficial N- 3 long-chain polyunsaturated fatty acids in sheep milk. *Eur J Lipid Sci Technol*. (2018) 120:1700256. doi: 10.1002/ejlt.2017.00256
13. Kominakis A, Hager-Theodorides A, Saridakis A, Antonakos G, Tsiamis G. Genome-wide population structure and evolutionary history of the frizarta dairy sheep1. *Animal*. (2017) 11:1680–8. doi: 10.1017/S1751731117000428

14. Li X, Yang J, Shen M, Xie X-L, Liu G-J, Xu Y-X, et al. Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nat Commun.* (2020) 11:2815. doi: 10.1038/s41467-020-16485-1
15. Staiger EA, Thonney ML, Buchanan JW, Rogers ER, Oltenacu PA, Mateescu RG. Effect of prolactin, beta-lactoglobulin, and kappa-casein genotype on milk yield in east friesland sheep. *J Dairy Sci.* (2010) 93:1736–42. doi: 10.3168/jds.2009-2630
16. Kessler EC, Bruckmaier RM, Gross JJ. Immunoglobulin G content and colostrum composition of different goat and sheep breeds in Switzerland and Germany. *J Dairy Sci.* (2019) 102:5542–9. doi: 10.3168/jds.2018-16235
17. Zhang D-Y, Zhang X-X, Li F-D, Yuan L-F, Li X-L, Zhang Y-K, et al. Whole-genome resequencing reveals molecular imprints of anthropogenic and natural selection in wild and domesticated sheep. *Zool Res.* (2022) 43:1–11. doi: 10.24272/j.issn.2095-8137.2022.124
18. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one fastq preprocessor. *Bioinformatics.* (2018) 34:i884–i90. doi: 10.1093/bioinformatics/bty560
19. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics.* (2009) 25:1754–60. doi: 10.1093/bioinformatics/btp324
20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and samtools. *Bioinformatics.* (2009) 25:2078–9. doi: 10.1093/bioinformatics/btp352
21. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* (2010) 20:1297–303. doi: 10.1101/gr.107524.110
22. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and vcfutils. *Bioinformatics.* (2011) 27:2156–9. doi: 10.1093/bioinformatics/btr330
23. Wang K, Li M, Hakonarson H. Annovar: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* (2010) 38:e164. doi: 10.1093/nar/gkq603
24. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. Plink: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* (2007) 81:559–75. doi: 10.1086/519795
25. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution.* (1984) 38:1358–70. doi: 10.1111/j.1558-5646.1984.tb05657.x
26. Yang J, Li W-R, Lv F-H, He S-G, Tian S-L, Peng W-F, et al. Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Mol Biol Evol.* (2016) 33:2576–92. doi: 10.1093/molbev/msw129
27. Wickham H. Data Analysis. In: *Ggplot2*. Springer (2016). p. 189–201.
28. Attali D, Baker C. *Ggextra: Add Marginal Histograms to 'Ggplot2', and More 'Ggplot2' enhancements*. R package version. (2019).
29. Bu D, Luo H, Huo P, Wang Z, Zhang S, He Z, et al. Kobas-I: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res.* (2021) 49:W317–W25. doi: 10.1093/nar/gkab447
30. Li X, Wu Q, Zhang X, Li C, Zhang D, Li G, et al. Whole-Genome Resequencing to Study Brucellosis Susceptibility in Sheep. *Front Genet.* (2021) 12:653927. doi: 10.3389/fgene.2021.653927
31. Doelman J. *Hepatic Gene Expression Profiling to Predict Future Lactation Performance in Dairy Cattle*. University of Guelph. (2011).
32. Cui X, Hou Y, Yang S, Xie Y, Zhang S, Zhang Y, et al. Transcriptional profiling of mammary gland in holstein cows with extremely different milk protein and fat percentage using RNA sequencing. *BMC Genomics.* (2014) 15:1–15. doi: 10.1186/1471-2164-15-226
33. van Leeuwen SS, Te Poele EM, Chatziioannou AC, Benjamins E, Haandrikman A, Dijkhuizen L. Goat milk oligosaccharides: their diversity, quantity, and functional properties in comparison to human milk oligosaccharides. *J Agric Food Chem.* (2020) 68:13469–85. doi: 10.1021/acs.jafc.0c03766
34. Buaban S, Lengnudum K, Boonkum W, Phakdeedindan P. Genome-wide association study on milk production and somatic cell score for thai dairy cattle using weighted single-step approach with random regression test-day model. *J Dairy Sci.* (2022) 105:468–94. doi: 10.3168/jds.2020-19826
35. Farhadian M, Rafat SA, Hasanpur K, Ebrahimie E. Transcriptome signature of the lactation process, identified by meta-analysis of microarray and RNA-Seq data. *J Biotechnol Comput Biol Bionanotechnol.* (2018) 99:75659. doi: 10.5114/bta.2018.75659
36. Lakhssassi K, Lahoz B, Sarto P, Iguácel LP, Folch J, Alabart JL, et al. Genome-Wide association study demonstrates the role played by the Cd226 gene in rasa aragonesa sheep reproductive seasonality. *Animals.* (2021) 11:1171. doi: 10.3390/ani11041171
37. Pokharel K, Peippo J, Weldenogodguad M, Honkatukia M, Li M-H, Kantanen J. Gene expression profiling of corpus luteum reveals the importance of immune system during early pregnancy in domestic sheep. *bioRxiv.* (2020):673558. doi: 10.1101/673558
38. Li X, Lin B, Zhang X, Shen X, Ouyang H, Wu Z, et al. Comparative transcriptomics in the hypothalamic-pituitary-gonad axis of mammals and poultry. *Genomics.* (2022) 114:110396. doi: 10.1016/j.ygeno.2022.110396
39. Tian Y, Sun P, Liu WX, Shan LY, Hu YT, Fan HT, et al. Single-Cell RNA sequencing of the mongolia sheep testis reveals a conserved and divergent transcriptome landscape of mammalian spermatogenesis. *FASEB J.* (2022) 36:e22348. doi: 10.1096/fj.202200152R
40. Dolebo AT, Khayatzaheh N, Melesse A, Wragg D, Rekik M, Haile A, et al. Genome-wide scans identify known and novel regions associated with prolificacy and reproduction traits in a sub-saharan african indigenous sheep (*Ovis aries*). *Mammalian Genome.* (2019) 30:339–52. doi: 10.1007/s00335-019-09820-5
41. Li J, Xu H, Liu X, Xu H, Cai Y, Lan X. Insight into the possible formation mechanism of the intersex phenotype of lanzhou fat-tailed sheep using whole-genome resequencing. *Animals.* (2020) 10:944. doi: 10.3390/ani10060944
42. Silpa M, Naicy T, Aravindakshan T, Radhika G, Boswell A, Mini M. Sirtuin3 (Sirt3) gene molecular characterization and SNP detection in prolific and low prolific goat breeds. *Theriogenology.* (2018) 122:47–52. doi: 10.1016/j.theriogenology.2018.09.008
43. Du X, He X, Liu Q, Di R, Liu Q, Chu M. Comparative Transcriptomics Reveals the Key lncRNA and mRNA of Sunite Sheep Adrenal Gland Affecting Seasonal Reproduction. *Front Vet Sci.* (2022) 9:816241. doi: 10.3389/fvets.2022.816241
44. Lai F-N, Zhai H-L, Cheng M, Ma J-Y, Cheng S-F, Ge W, et al. Whole-genome scanning for the litter size trait associated genes and SNPs under selection in dairy goat (*Capra hircus*). *Sci Rep.* (2016) 6:1–12. doi: 10.1038/srep38096
45. Megdiche S, Mastrangelo S, Ben Hamouda M, Lenstra JA, Ciani E, A. Combined multi-cohort approach reveals novel and known genome-wide selection signatures for wool traits in merino and merino-derived sheep breeds. *Front Genet.* (2019) 10:1025. doi: 10.3389/fgene.2019.01025
46. Zhang R, Li Y, Jia K, Xu X, Li Y, Zhao Y, et al. Crosstalk between androgen and Wnt/B-catenin leads to changes of wool density in Fgf5-knockout sheep. *Cell Death Dis.* (2020) 11:1–12. doi: 10.1038/s41419-020-2622-x
47. Shen J, Wang Y, Bai M, Fan Y, Wang Z, Bai W. Novel circRNAs from cashmere goats: discovery, integrated regulatory network, and their putative roles in the regeneration and growth of secondary hair follicles. *Czech J Animal Sci.* (2022) 67:237–51. doi: 10.17221/179/2021-CJAS
48. Demars J, Cano M, Drouilhet L, Plisson-Petit F, Bardou P, Fabre S, et al. Genome-Wide identification of the mutation underlying fleece variation and discriminating ancestral hairy species from modern woolly sheep. *Mol Biol Evol.* (2017) 34:1722–9. doi: 10.1093/molbev/msx114
49. Yilmaz O, Kizilaslan M, Arzik Y, Behrem S, Ata N, Karaca O, et al. Genome-Wide association studies of preweaning growth and in vivo carcass composition traits in esme sheep. *J Animal Breeding Genet.* (2022) 139:26–39. doi: 10.1111/jbg.12640
50. Song N, Wang X, Gui L, Raza SHA, Luoreng Z, Zan L. MicroRNA-214 regulates immunity-related genes in bovine mammary epithelial cells by targeting Nfatc3 and Traf3. *Mol Cell Probes.* (2017) 35:27–33. doi: 10.1016/j.mcp.2017.06.002
51. Pereira G, Guo Y, Silva E, Silva MF, Bevilacqua C, Charpigny G, et al. Subclinical endometritis differentially affects the transcriptomic profiles of endometrial glandular, luminal, and stromal cells of postpartum dairy cows. *J Dairy Sci.* (2022). doi: 10.3168/jds.2022-21811
52. Marina H, Pelayo R, Suárez-Vega A, Gutiérrez-Gil B, Esteban-Blanco C, Arranz J. Genome-wide association studies (GWAS) and post-GWAS analyses for technological traits in assaf and churra dairy breeds. *J Dairy Sci.* (2021) 104:11850–66. doi: 10.3168/jds.2021-20510