RESEARCH   **Open Access**

Genomic analysis uncovers novel candidate genes related to adaptation to tropical climates and milk production traits in native goats

Chenxi Zhang[1], Hojjat Asadollahpour Nanaei[1,2,3*], Niloufar Jafarpour Negari[4], Mahmoud Amiri Roudbar[5], Zeinab Amiri Ghanatsaman[3], Zhannur Niyazbekova[6] and Xiaojun Yang[1*]

Abstract

Background Since domestication, both evolutionary forces and human selection have played crucial roles in producing adaptive and economic traits, resulting in animal breeds that have been selected for specific climates and different breeding goals. Pakistani goat breeds have acquired genomic adaptations to their native climate conditions, such as tropical and hot climates. In this study, using next-generation sequencing data, we aimed to assess the signatures of positive selection in three native Pakistani goats, known as milk production breeds, that have been well adapted to their local climate.

Results To explore the genomic relationship between studied goat populations and their population structure, whole genome sequence data from native goat populations in Pakistan ($n = 26$) was merged with available worldwide goat genomic data ($n = 184$), resulting in a total dataset of 210 individuals. The results showed a high genetic correlation between Pakistani goats and samples from North-East Asia. Across all populations analyzed, a higher linkage disequilibrium (LD) level (-0.59) was found in the Pakistani goat group at a genomic distance of 1 Kb. Our findings from admixture analysis ($K = 5$ and $K = 6$) showed no evidence of shared genomic ancestry between Pakistani goats and other goat populations from Asia. The results from genomic selection analysis revealed several candidate genes related to adaptation to tropical/hot climates (such as; *KITLG*, *HSPB9*, *HSP70*, *HSPA12B*, and *HSPA12B*) and milk production related-traits (such as *IGFBP3*, *LPL*, *LEPR*, *TSHR*, and *ACACA*) in Pakistani native goat breeds.

Conclusions The results from this study shed light on the structural variation in the DNA of the three native Pakistani goat breeds. Several candidate genes were discovered for adaptation to tropical/hot climates, immune responses, and milk production traits. The identified genes could be exploited in goat breeding programs to select efficient breeds for tropical/hot climate regions.

Keywords Population genomics, Pakistan indigenous goats, Whole-genome sequence, Adaptation, Candidate genes

*Correspondence:
Hojjat Asadollahpour Nanaei
h.asadollahpour@agr.uk.ac.ir
Xiaojun Yang
yangxj@nwafu.edu.cn
Full list of author information is available at the end of the article

Zhang *et al. BMC Genomics*   (2024) 25:477

Page 2 of 11

Background

For thousands of years, domesticated animals have played important roles in human society and have been genetically adapted to a variety of environmental circumstances [1–3]. They have occupied a wide range of roles including; source of energy, food, material for construction, herding, and manure for primarily rural farmers. Based on recent paleogenomic evidence, goat (*Capra hircus*) were one of the first animals domesticated for use as livestock some 10,000 years ago from multiple wild bezoars in Southwest Asia [2]. Since domestication, they have spread throughout the globe, and currently there are more than 1 billion farmed goats with over 210 recognized breeds [4–6].

Generally, indigenous goat breeds have excellent adaptability to local climates, such as tropical and arid environments [7]. They have a high level of vitality, resistance to endemic diseases and low input regimes [8]. Today, local goats have become an important supply of high-quality products in the developing world through the conversion of feed from noncompetitive sources. Due to their ability to survive in harsh climates, they have increasingly become a profitable business for farmers in local regions [4, 9]. According to the FAO, more than 90% of goat populations are inhabited in developing countries (https://www. fao.org/faostat/en/#data/QCL), and Pakistan with more than 53 million heads and around 25 breeds is ranked third after China and India [10, 11]. Thus, goat farming is an important part of the livestock sector in this country. Among the local Pakistani goats, some of them are popular for their higher levels of milk production traits. For example, Beetal goat is a well-known breed used for dairy products in Pakistan due to its high production capacity (average 1.2-liters milk per day) and also its beauty [12]. Damani is another local dairy goat breed that originated in the Bannu and Dera Ismail Khan districts in NWF Province, Pakistan. Kamori goat, which is mainly raised for milk production (average 1.5-2-liters of milk per day), is popularly known as the cow of the poor farmer in the country [13, 14].

In the last ten years, numerous genetic studies have been carried out in order to explore genetic diversity and identify loci related to climate change and production traits in different goat breeds [15–18]. However, there are only a few genomic studies focusing on Pakistani native goats, and the majority of these studies have been carried out using limited microsatellite marker resources, which prevents them from detecting novel variation at the complete sequence level [19, 20].

Here we explored a thorough genetic analysis of worldwide goat genomes ($n = 210$), including seven populations from different geographical regions, to investigate population genetic diversity, genetic structure, and also the signatures of selection analysis. Our findings from selective genomic analysis revealed new genomic footprints related to adaptation to local climates and milk production traits that may be under natural/artificial selection in local goat populations.

Methods

Quality control checking, read mapping and SNP calling

The experimental samples were collected from the Sindh Agriculture University, Tandojam, Pakistan. The studied animals were not anesthetized or euthanized in order to perform this study. No goat individuals died in this study, and all animals stayed healthy after collecting blood samples. About 15 milliliters of whole blood ($n = 26$) were collected from each native goat. By using the standard phenol-chloroform extraction method, DNA was extracted from blood samples. Sequence data were generated using the Illumina Hiseq 2500. In addition, previously published genome data from worldwide domestic goat samples ($n = 184$) was collected from the Sequence Read Archive (https://trace.ncbi.nlm.nih.gov/Traces/sra). Supplementary Table S1 provides details on the samples utilized in this study. By applying the FASTQC program (Version 0.4.2), quality control checks on raw high throughput sequence data were carried out (https:// www.bioinformatics.babraham.ac.uk/projects/fastqc/), and low quality base pairs and adapters were then filtered from the original paired-end sequences by Trimmomatic software (version 0.38.1) [21]. BWA (mem algorithm) (Version 0.7.15) [22] was performed to map all pre-processed reads to the reference genome (GCF_001704415.1, ARS1). The SAMtools program was applied for manipulating SAM (.sam) files and converting files between different formats [23]. In order to minimize false positives, we removed PCR duplicates from the alignment files using the Picard toolkit (https://broadinstitute.github.io/picard/). Next, utilizing tools from the Genome Analysis Toolkit (GATK), base quality score recalibration (BQSR) and local realignment around known INDELs were carried out to refine the accuracy of the alignments [24]. The GATK best practices workflow was applied to identify and filter out the final genomic variants (SNPs). Haplotypes were inferred and imputed from variant calling files by using BEAGLE software (V.4.1) [25]. All discovered variations (9,764,316) were then filtered to ensure that they had at least 25 and 40 mapping and genotyping qualities, respectively. To eliminate potential genotyping errors, all discovered loci containing multiple alleles (more than 2 alleles) and inside clusters (lower than 3 SNPs in a 10 bp window) were eliminated from additional analysis [26].

Genome-wide patterns of genomic structure and admixture

Following the removal of low-quality sequencing reads, to assess the genetic relatedness among all studied goat individuals, we used 7,843,129 SNPs retrieved from data from full genome resequencing to construct a maximum-likelihood (ML) phylogenetic tree [27]. To manage and visualize the topological structure, the free available tool (iTOL, https://itol.embl.de/) was utilized. To generate pruned SNP data with approximate linkage equilibrium (LD), all genomic data was pruned for LD in the PLINK program ("indeppairwise 50 10 0.1" flag) [28]. In addition, PCA was carried out through the use of the genome wide complex trait (GCTA) statistical approach [29]. To estimate the genomic structure of the studied goat samples, we employed the admixture software clustering process (v. 1.3.0) with an ancestor population (K) size ranging from $K=2$ to $K=6$ (10,000 times over in each run) [30]. By applying CHROMOPAINTER and fineSTRUCTURE programs, the haplotype sharing patterns were explored for all populations [31]. Additionally, by using the PLINK program [28] with default parameters, we estimated different genomic features, including LD decay, inbreeding (F) ('-het' flag), and nucleotide diversity.

Genome-wide selective sweep analysis

To discover genomic selection signals related to physiological traits, two distinct approaches were applied. The genome-wide weighted FST was estimated [32] in order to calculate genetic distance among populations with different sample sizes [33]. In addition, by using VCFtools (V0.1.13), we investigated nucleotide diversity θπ (-Window-pi 50 kb -Window-pi-step 25 kb) for the studied populations [34]. The sliding window analyses were carried out with a 50 kb window size and a 25 kb step size for the entire genome. The average log2 (θπ Pakistan group/θπ Northeast Asia) and FST scores of SNPs in each window were computed.

Gene set enrichment and pathway analysis

To investigate the potential pathways associated with the discovered genomic regions, using the aforementioned methods (FST and log2 ratio), the Variant Effect Predictor (VEP) toolkit from Ensembl was used (https://ensembl.org/info/docs/tools/vep/index.html) to annotate all candidate genomic regions. Gene set enrichment analysis was then performed using a reliable and up-to-date tool for functional enrichment analysis (g: Profiler; https://biit.cs.ut.ee/gprofiler/). Finally, we used Benjamini-Hochberg's False Discovery Rate (FDR-BH) adjustment across all tests to correct thresholds for false discovery rates.

Results

Population structure and genomic diversity

All the studied goat individuals were assigned to seven different groups according to their geographical regions, including Europe ($n=20$), Africa ($n=56$), Pakistan ($n=26$), Bangladesh ($n=7$), China ($n=42$), Iran ($n=36$), and Iraq ($n=23$) (Fig. 1 and Supplementary Table S1).

To comprehend the genetic heritage of worldwide goat populations, we conducted PCA analysis on all the studied goat individuals (Fig. 1B). The PC1 and PC2 explained 8.56% and 5.18% of the total genotypic variation, respectively. PC1 divided European and African goat groups from each other and also from Asian goat individuals. In agreement with the previous studies, those samples from East African regions showed a relatively close relationship with those from West Asian groups [2, 3]. In contrast, PC2 indicated the variation between goat samples from northern Europe and those from African regions. With whole-genome sequence data, a ML phylogenetic tree divided all samples into three main geographic subgroups, including China-Bangladesh-Pakistan; Iran-Iraq; Africa-European (Fig. S1).

By utilizing a model-based Bayesian technique that ADMIXTURE software has put into place, we explored potential admixture between all studied populations (from $K=2$ to $K=6$) (Fig. 1C). The $K=2$ splits Asian goat samples from both African and European populations. At $K=4$, with the least amount of CV error (-0.52) (Fig. S2), we observed a division between the European and African goat populations (Fig. 1C). Ancestral genomic proportions at $K=5$ showed that Chinese and Bangladeshi samples were mainly assigned to the same genomic clade, while Pakistani goats were divided from the remaining Asian goat groups. We next used phylogenetic tree analysis using only the examined Pakistani goats in order to concentrate on the variety within the breeds of Pakistani goats. Our results revealed that Pakistani goat individuals can be clearly grouped into three different clades (Fig. S3).

To explore the patterns of haplotype sharing between the studied goat populations, we applied the algorithms available in CHROMOPAINTER and fineSTRUCTURE (Fig. S4). The results showed a relatively higher level of haplotype sharing in Pakistani samples with East Asian goat populations, compared to other goat groups. In order to assess the historical and evolutionary processes that shape the genetic structure of local goats from Pakistan, we explore the genetic divergence patterns as a function of θπ (Fig. 1D). The results revealed a higher correlation between the gene pool of Pakistani goats and East Asian individuals (samples from China) than other populations from South west Asia (Iran and Iraq) and European (Fig. 1D). Based on the complete mitochondrial genome sequences, five different haplotypes; A,
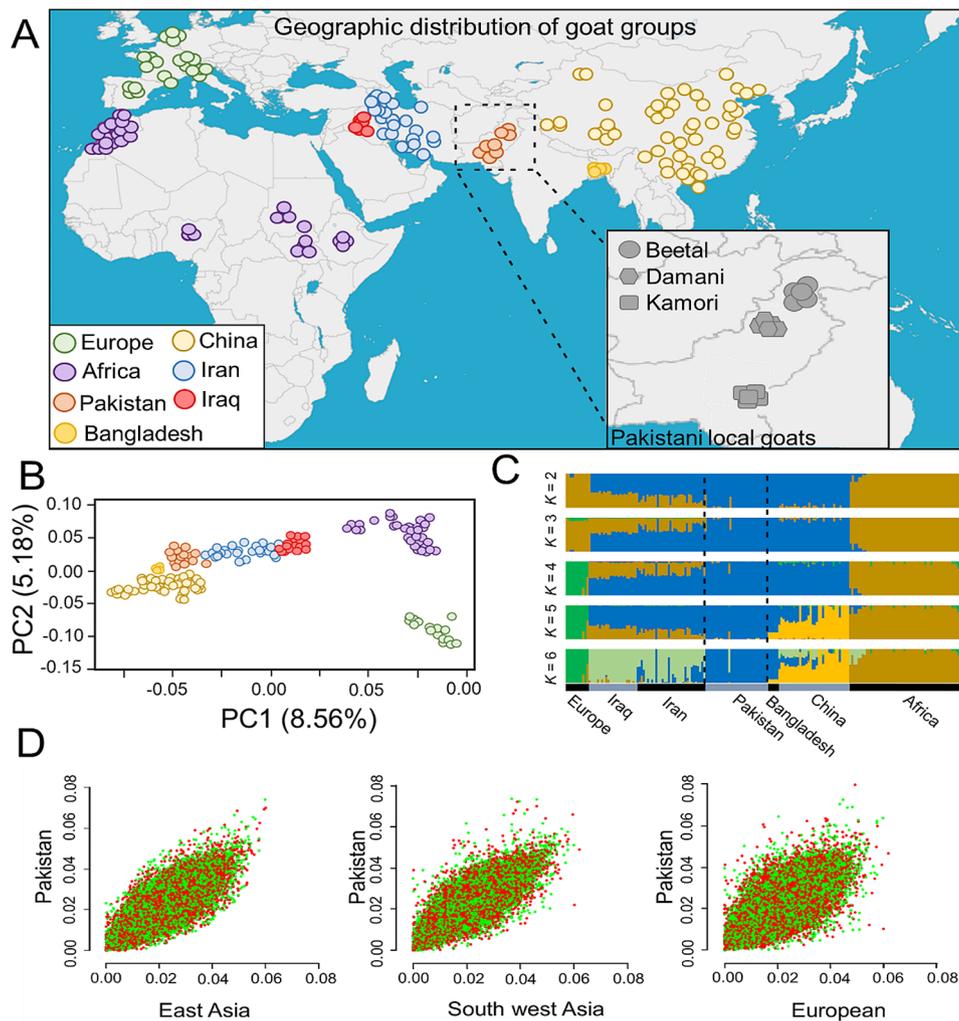
Fig. 1 Geographic distribution of domestic goats analyzed in the current study (**A**). The color of goat populations reflects their geographic regions. (**B**) PCA analysis with whole genome data, PC1 against PC2. (**C**) ADMIXTURE clustering analysis for studied goat populations (ancestral populations from $K=2$ to $K=6$). (**D**) Nucleotide diversity correlations (non-overlapping window with 50-kb) between Pakistani native goat groups and other goat populations (orange)

B, C, D and G were detected in the studied goat groups (Fig. 2A).

The majority of samples belong to haplogroup A (−89%), while compared to other haplogroups, haplogroup C showed the lowest frequency (−0.05) and was only presented in European goat individuals (Fig. 2A). The Pakistani goat individuals with four haplogroups (A, B, G and D) were the most diverse group, while the African (A and G), European (A and C) and Iraqi goat (A and B) groups harbored only two haplogroups. Genomic inbreeding coefficients were further calculated per individual (ranged from −0.02 to −0.54) and each goat population (ranged from 0.22 to 0.09 for European and Iranian goat populations, respectively) (Fig. 2B). We then estimated genome-wide LD (R^2) decay rates between adjacent nucleotides across the whole sequence, which could

be informative for the effective population size, non-random mating and migration events.

Our findings from the LD decay up to a distance of 120 kb are shown in Fig. 2C. At a marker pair distance of about 1 Kb, we discovered the r^2 scores were higher for all studied goat groups (ranging from −0.575 to −0.592 for Iranian and Pakistani goats, respectively); however, a progressive decline was noted when the physical distance between SNPs increased up to 20 Kb. Average nucleotide diversity within each goat group was calculated, and compared to other groups, African and Iranian populations showed slightly lower and higher diversity, respectively (Fig. 2D). Our demographic analyses based on the Sequential Markov Coalescent (SMC) method reveal that the divergence times between different groups predated the estimated
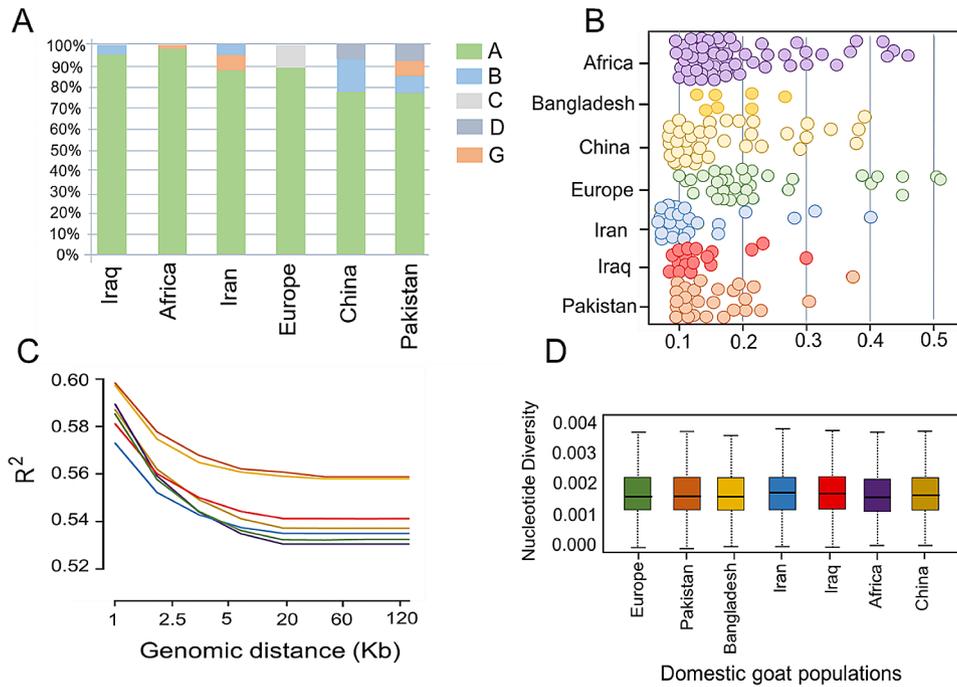
Fig. 2 Mitochondrial genome haplotype analysis (**A**). The majority of samples fell into haplogroup A. (**B**) Inbreeding coefficient. (**C**) Linkage disequilibrium (LD) decay is calculated as the squared genomic correlation coefficient by pairwise physical distance in all domestic groups. (**D**) Nucleotide diversity across the complete genome, estimated in a sliding window of 50 KB, with increments of 20 KB
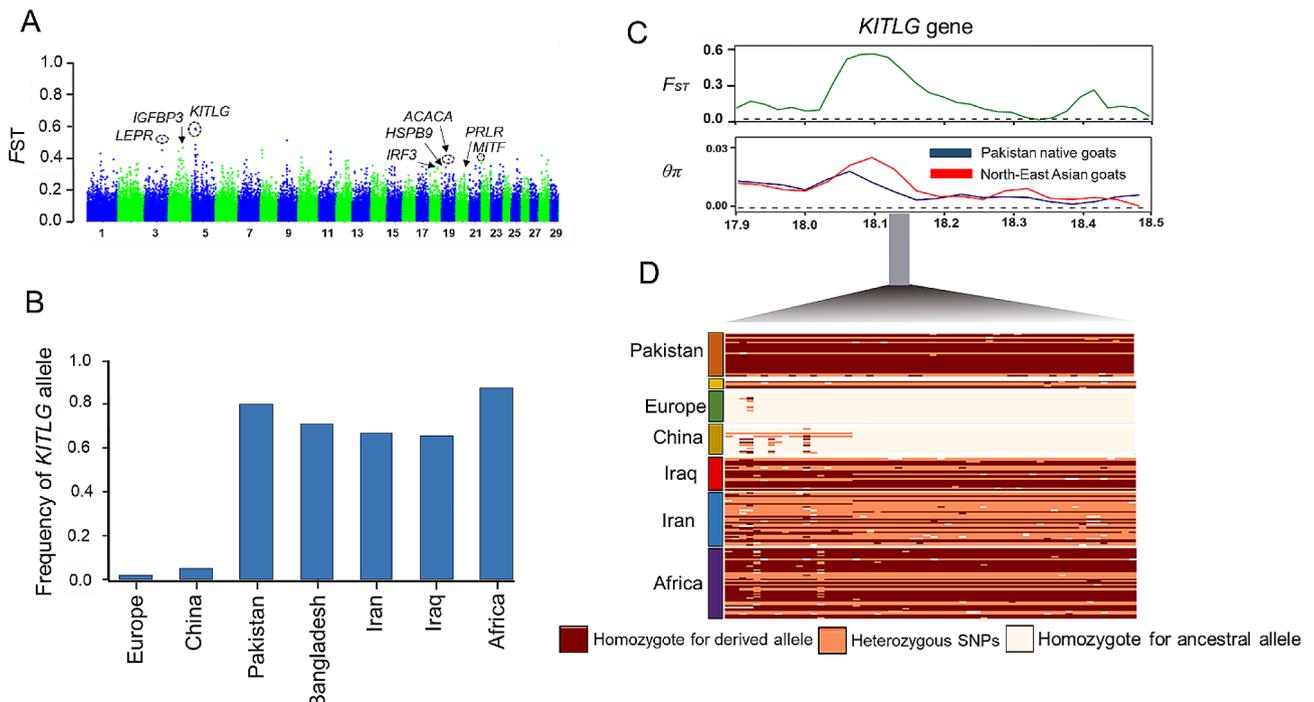
Fig. 3 Genomic landscape of population differentiation by *F*ST between Pakistani native goats and Northeast Asian breeds (**A**). (**B**) The *KITLG* allele frequency (chromosome 5: 18, 100–18, 150 kb) in each goat population. (**C**) The putative sweep region was further validated by *F*ST and π tests. (**D**) The SNP frequency pattern in the *KITLG* gene (18, 100–18, 150 kb) is shown

Table 1 Two statistical methods (*F*ST and log2 θπ ratio) were used to identify genes involving hot climate and immune response traits in Pakistani indigenous goats

Statistical-approaches	Gene	Chr.*	Position (Kb)	Summary of gene function
*F*ST (top 1%)	*KITLG*	5	18,075–18,125	Adaptation to hot climate [3]
	HSPB9	19	41,875–41,925	Heat stress [51]
	DNAJB14	6	24,975–25,025	Heat stress [63]
	HSPA12B	13	50,800–50,850	Heat stress [52]
	DNAJC1	13	22,425–22,475	Heat stress [64]
	IRF3	18	57,025–57,075	Immune response [54]
	MITF	22	31,525–31,575	Innate immune signaling [55]
	RFC2	25	32,925–32,975	Immune response [56]
Log2 (θπ·Group-1 /θπ Group-2) ** (top 1%)	*KITLG*	5	18,100–18,150	Adaptation to hot climate [3]
	NBEA	12	60,575–60,625	Adaptation to hot climate [65]
	HSP70	23	22,400–22,450	Heat stress [53, 66]
	HSPA12B	13	22,425–22,475	Heat stress [52]
	HSPB9	19	22,425–22,475	Heat stress [51]
	DNAJB14	6	24,975–25,025	Heat stress [63]
	DNAJC1	13	22,425–22,475	Heat stress [64]
	IL23A	5	56,350–56,400	Resistance to gastrointestinal parasite [57]
	MITF	22	31,525–31,575	Immune response [55]
	STAT2	5	56,350–56,400	Immune response [67]
	IRF3	18	57,025–57,075	Immune response [54]
	CDH9	20	44,550–44,600	Environmental adaptation [68]
	LRFN5	21	52,050–52,100	Immune response [69, 70]

* Chromosome

** Group-1 = Pakistani native goats, Group-2 = North-East Asian native goat group

domestication time around 10,500 years ago (Fig. S5) [2].

Signals of genomic positive selection associated with adaptive and milk production traits

The adaptation of domestic animals located in tropical regions to hot environmental conditions may have influenced their physical characteristics, such as hair color and skin pigmentation, which are directly associated with heat loss capacity [35]. For example, it has been reported that indigenous sheep breeds in hot and desert climates have coarse wool coats [36], while the majority of goat populations in cold geographical regions produce cashmere wool [37]. Generally, Pakistan has a tropical climate with an extreme annual range of temperature, which ranks it globally among the top ten most prone nations to climate change [38]. In order to discover candidate genes related to adaptations to local climate and milk production traits, we compared local goat breeds from Pakistan with those from East Asian samples using the *F*ST statistic and also differences in nucleotide diversity (π ln-ratio East Asian goats/local dairy goats from Pakistan).

Our findings from the signature of selection statistics provided a total of 145 and 147 genomic windows in the top *F*ST and log2 ratio (1% cutoff) values, respectively (Fig. 3, Tables S2 and S3). To explore the

potential genetic mechanisms linked with biological traits in goats, all genomic regions were then annotated with their biological pathways (Tables S4 and S5). The strongest *F*ST-outlier signal contained one candidate gene, the *KITLG* gene, located on chromosome 5 (18,100–18,150 kb), that is reportedly associated with hair-related traits in humans [39] and also adaptation to local climate in livestock species [40]. We further observed that the selected genomic region had a high frequency in goat populations from Africa, South and Southwest Asia, compared with those samples from Northeast Asia (China) and Europe (Fig. 3B). The low π scores and high *F*ST values revealed a strong signature at this locus in Pakistani goat individuals (Fig. 3C). Furthermore, we identified several other genes with significant selection signals associated with milk production traits (such as; *PRLR*, *ACACA*, *LEPR*, *LPL* and *IGFBP3*) and environmental adaptation (such as; *NBEA*, *HSP70*, *HSPB9* and *CDH9*).

KEGG pathways mapping and over-representation analysis (ORA) revealed some functional categories that are significantly associated with adaptation to local climate and immune response traits including; "regulation of response to stress" (GO:0080134), "immune effector process" (GO:0002252), "regulation of immune system process" (GO:0002682). Additionally, we identified "regulation of phosphorylation"

(GO:0042325), "phosphorus metabolic process" (GO:0006793) and "phosphorylation" (GO:0016310) categories that are related with milk production traits (Tables S4 and S5).

Discussion

Genomic diversity and population structure

Knowledge about the genetic architecture and variability can advance our understanding of the evolution of a population, which could be useful for breeding and genetic conservation programs [41]. In this research, by collecting sequencing data from worldwide goat populations, we investigated the genomic structure of Pakistani native goats and their phylogenic relationship with other goat populations. Our phylogenetic analysis showed that the native goats from Pakistan are the sister taxon to samples from Bangladesh, which could be considered a separate clade from other goat breeds from Southwest Asia; including samples from Iran and Iraq countries. Through Admixture and PCA analysis, the genetic distance patterns between the analyzed goat groups were also observed. We further observed a relatively higher correlation of nucleotide diversity between Pakistani goat breeds and samples from North-East Asia, compared with samples from Southwest Asia. These results agree with those of the earlier research, which suggested that the genomic architecture of the goat populations from Southwest Asia has been affected by broad historical processes such as gene flow from ancient periods [2]. In line with previous studies on different species [3, 16, 26, 42], we discovered that all groups experienced the same pattern of reduction in LD decay as the genomic distance (Kb) increased, while a relatively rapid change in LD over increasing physical distance was detected in the African population (Fig. 2C). The r^2 values found in this investigation ranged from -0.53 to -0.57 at a marker pair distance of 120 Kb, which were consistent with those previously reported for Iranian and Chinese native goat groups [16, 42]. Furthermore, the results demonstrated that Pakistani goats had much higher levels of LD across all genomic lengths than other goat groups, which could be due to recent artificial selection for economic traits in these breeds [43].

Selection signals related to adaptation to tropical environmental and milk production traits

Due to their long history of natural selection, domestic animals are an excellent model for studying the genetic variants enabling adaptation to local climates and also for biomedical research [44]. Generally, due to their fast metabolic rate and growth performance, farm animals are susceptible to heat stress [45]. Prior research has demonstrated that heat tolerance is one of the most demanding challenges for domestic animals that inhabit tropical regions, which can negatively impact their health and production [46, 47]. In several independent studies, it has been confirmed that there is variation in the performance of goats from different breeds under high temperature conditions [48]. Thus, genetic selection could be a cost-effective tool to help improve the thermotolerance of animals in hot regions. Therefore, identifying and selecting genes related to heat tolerance is expected to provide a potential and long-term solution to this issue. In this study, in order to understand how native goats from Pakistan have adapted to the tropical climates, we compared them with samples from Northeast Asia to uncover loci that have been exposed to long-term natural selection. Our findings from the detection of selection signals revealed that some candidate loci inside the regions of high confidence selection (highest 1% log2 θπ ratio and 1% *F*ST values) may be related to adaptation to local climate and immune response processes (Table 1). We observed the strongest genomic signal (both θπ ratio and *F*ST methods) related to hair and skin pigmentation (*KITLG*) located on chromosome 5 [3, 39]. The protein related to this gene, the tyrosine kinase protein KIT, is required for the growth and differentiation of various cell types, such as melanocyte proliferation, pigment production, blood cells, and also germ cells [49]. A previous study on the human genome has reported a causative base-pair change at the enhancer of this gene (encoding the *KIT* ligand), which controls the expression of hair follicles [39]. So far, various mutations of the *KITLG* gene related to coat color and environmental adaptation [39, 40, 50] have been reported. Due to the fact that the frequency of the *KITLG* allele was higher in populations residing in hot climate conditions (such as in South Asia and Africa), we speculate that this gene may have played a role in the evolution of goat lineages (Fig. 3D). Aside from this gene, several genes for heat shock protein (HSP) were also discovered, such as; *HSPB9*, *HSP70* and *HSPA12B*, may have an impact on attributes related to heat tolerance in goats and different livestock species [51–53] (Table 1). We further discovered a few potential loci on different chromosomes that could be important in the goat immune system. For instance, we found *IL23A*, *IRF3, RFC2* and *MITF* genes that are related to immune response traits [54–57].

Due to the fact that studied Pakistani native goats are considered to be among the highest milk yielding breeds in Pakistan [12–14], comparative genome analysis of this group with other native breeds that usually produce low levels of milk production is a promising method for determining genetic diversity in milk yield characteristics. To accomplish this goal, we employed two statistical techniques: comparisons between two groups and comparisons within a group. Within the genomic regions with greater *F*ST scores (top 0.01)

Table 2 Identified candidate genes related with milk production-related traits

Statistical-approaches	Gene	Chr.*	Position (Kb)	Summary of gene function
FST (top 1%)	*LEPR*	3	41,100 − 41,150	Milk fat and protein [71, 72]
	LPL	8	66,725 − 66,775	Milk fat yield [16, 62]
	IGFBP3	4	44,100 − 44,150	Milk performance [73]
	TRNAC-ACA	6	66,425 − 66,475	Milk composition traits [74]
	TRNAS-GGA	4	61,925 − 61,975	Milk composition traits [74]
	PRLR	20	38,950 − 39,000	Milk production traits [75, 76]
	ABCB1	4	87,525 − 87,575	Milk production traits [77]
	ACACA	19	13,325 − 13,375	Milk fat content [78, 79]
	TSHR	10	10,650 − 10,700	Milk production [80]
	SUPT3H	23	30,575 − 30,625	Milk protein percentage [81]
Log2 (θπ·Group-1 /θπ Group-2) ** (top 1%)	*ARFGEF1*	14	50,700 − 50,750	Milk production [82, 83]
	TRNAC-ACA	5	16,725 − 16,775	Milk composition traits [74]
	TRNAS-GGA	4	12,125 − 12,175	Milk composition traits [74]
	LEPR	3	41,100 − 41,150	Milk fat and protein [71, 72]
	LPL	8	66,750 − 66,800 and 66,725 − 66,775	Milk fat yield [16, 62]
	IGFBP3	4	44,125 − 44,175 and 44,100 − 44,150	Milk performance [73]
	TSHR	10	10,625 − 10,675	Milk production [80]
	AGPAT1	23	22,700 − 22,750	Milk fatty acid [84, 85]
	SUPT3H	23	30,500 − 30,550	Milk protein percentage [82]

* Chromosome

** Group-1 = Pakistani native goats, Group-2 = North-East Asian native goat group

and reduced nucleotide diversity (cutoff 1% log2 θπ ratio), we detected some candidate genes associated with milk production traits. Protein-encoding genes were listed in Table 2.

For example; in the cluster of signals discovered by both statistical methods, we found *IGFBP3* as a putative candidate gene on chromosome 4 (44.10-44.15 MB) (Table 2). This gene encodes a protein with an IGFBP domain, which is a subfamily of the insulin-like growth factor binding protein [58]. It has been confirmed that the mutations in this gene are related to the milk production traits in both goat [59] and sheep [60] species. Another gene associated with milk, *LPL*, was found in one of the selection regions of chromosome 8, which is a crucial potential gene that plays a central role in human plasma triglyceride metabolism [61]. Previous studies have reported that genetic mutations in the *LPL* gene are related to milk production traits (e.g., milk yield, fat content, and protein yield) and composition traits (protein and fat percentages) in dairy goats and also in different livestock species [16, 62].

Conclusions

By exploring genome-wide data from worldwide goat populations, we provided valuable insight into the genomic architecture of three local goat populations in Pakistan. The results indicate that there is a relatively higher genetic affinity between studied goat samples from Pakistan and those from North-East Asia, compared with samples from other geographical regions. We further observed high levels of LD decay in Pakistani goat samples, which may be a consequence of recent human selection for interested traits (such as milk production traits). Furthermore, we discovered multiple candidate genes related to local climate adaptation, immune response, and milk production traits. The genomic regions discovered in this study will help us better understand the mechanisms of selection and identify the targets of selection in goat breeds located in tropical regions.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10387-y.

Supplementary Material 1

Supplementary Material 2

Data availability
The data used to support the findings of this study have been archived at the NCBI SRA Database under the BioProject accession number PRJNA1087734. The fastq files used in this study were downloaded from the public sequence database (https://trace.ncbi.nlm.nih.gov/Traces/sra).

Declarations

Ethics approval and consent to participate
All experimental protocols were carried out following relevant guidelines and regulations. The protocols for all animal experiments were approved by the Northwest Agriculture and Forestry University Animal Research Ethics Committee (Yangling, Shaanxi, China) (Permit code: NWAFU 2023011), and all methods strictly obeyed the Guide for the ARRIVE guidelines.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
¹College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi 712100, China
²College of Life Sciences, Northwest A&F University, Yangling 712100, China
³Animal Science Research Department, Fars Agricultural and Natural Resources Research and Education Center, Agricultural Research, Education and Extension Organization (AREEO), Shiraz, Iran
⁴Department of Animal Science, Shahid Bahonar University of Kerman, Kerman 7616914111, Iran
⁵Department of Animal Science, Safiabad-Dezful Agricultural and Natural Resources Research and Education Center, Agricultural Research, Education and Extension Organization (AREEO), Dezful 333, Iran
⁶Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi 712100, China

References
1. Alberto FJ, Boyer F, Orozco-terWengel P, Streeter I, Servin B, Villemereuil PD, et al. Convergent genomic signatures of domestication in sheep and goats. Nat Commun. 2018;9(1):813.
2. Daly KG, Delser PM, Mullin VE, Scheu A, Mattiangeli V, Teasdale MD, et al. Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. Science. 2018;361(6397):85–8.
3. Asadollahpour Nanaei H, Cai Y, Alshawi A, Wen J, Hussain T, Fu WW, Xu NY, Essa A, Lenstra JA, Wang X, Jiang Y. Genomic analysis of indigenous goats in Southwest Asia reveals evidence of ancient adaptive introgression related to desert climate. Zool Res. 2023;44(1):18–27.
4. Miller BA, Lu CD. Current status of global dairy goat production: an overview. Asian-Australas J Anim Sci. 2019;32(8):1219–32. https://doi.org/10.5713/ajas.19.0253.
5. Ahlawat SPS. Goat genetic resource of India. National bureau of animal genetic resources P.O. box. 129, Karnal, 132001 (Haryana) India. 2006.
6. Saleh AA, Rashad AMA, Hassanine NAM, Sharaby MA, Sallam SMA. History of the Goat and Modern Versus Old Strategies to enhance the genetic performance. IntechOpen. 2023. https://doi.org/10.5772/intechopen.1001106.
7. Silanikove N, Koluman N. Impact of climate change on the dairy industry in temperate zones: predications on the overall negative impact and on the positive role of dairy goats in adaptation to earth warming. Small Rumin Res. 2015;123:27–34. https://doi.org/10.1016/j.smallrumres.2014.11.005.
8. Mandal M, Mishra C, Dash SK, Priyadarshini P, Sabat SS, Swain L, Sahoo M. Genomic insight to the disease resistance in goat. J Pharm Innov. 2018;7(2):98–103.
9. Utaaker KS, Chaudhary S, Kifleyohannes T, Robertson LJ. Global Goat! Is the Expanding Goat Population an important Reservoir of Cryptosporidium? Front Vet Sci. 2021;8:648500. https://doi.org/10.3389/fvets.2021.648500.
10. Khan MFU, Ashfaq F. Meat production potential of small ruminants under the arid and semi-arid conditions of Pakistan. J Agric Mar Sci. 2010;15:33–9.
11. Babar MEl, Hussain T, Ahmad MS, Ali A, Abbas K, Ali MM. Evaluation of Pakistani goat breeds for genetic resistance to Haemonchus Contortus. Acta Vet Brno. 2015;84:231–5. https://doi.org/10.2754/avb201584030231.
12. Prasad H, Tewari HA, Sengar OPS. Milk yield and composition of the beetal breed and their crosses with Jamunapari, Barbari and Black Bengal breeds of goat. Small Rumin Res. 2005;58(2):195–9.
13. Talpur FN, Bhanger MI, Memon NN. Milk fatty acid composition of indigenous goat and ewe breeds from Sindh, Pakistan. J Food Compos Anal. 2009;29(1):59–64.
14. Bilal G, Moaeen-ud-Din M, Khan M, Reecy J. Productive and reproductive performance of Goat breeds of Sindh. J Anim Sci. 2018;96(Suppl 3):258–9. https://doi.org/10.1093/jas/sky404.566.
15. Massender E, Brito LF, Maignel L, Oliveira HR, Jafarikia M, Baes CF, Brian Sullivan B, Schenkel FS. Single-step genomic evaluation of milk production traits in Canadian Alpine and Saanen dairy goats. J Dairy Sci. 2022;105(3):2393–407.
16. Amiri Ghanatsaman Z, Ayatolahi Mehrgardi A, Asadollahpour Nanaei H, Esmailizadeh A. Comparative genomic analysis uncovers candidate genes related with milk production and adaptive traits in goat breeds. Sci Rep. 2023;13(1):8722. https://doi.org/10.1038/s41598-023-35973-0.
17. Martin P, Palhière I, Maroteau C, et al. A genome scan for milk production traits in dairy goats reveals two new mutations in Dgat1 reducing milk fat content. Sci Rep. 2017;7:1872. https://doi.org/10.1038/s41598-017-02052-0.
18. Xiong J, Bao J, Hu W, Shang M, Zhang L. Whole-genome resequencing reveals genetic diversity and selection characteristics of dairy goat. Front Genet. 2023;6(13):1044017. https://doi.org/10.3389/fgene.2022.1044017.
19. Hussain T, Babar ME, Sadia H, Shaheen M, Nadeem A, Ali A, Wajid A, Shah SA. Microsatellite markers based genetic diversity analysis in Damani and Nachi goat breeds of Pakistan. Pak Vet J. 2013;33(4):520–2.
20. Hussain T, Shaheen M, Babar ME, Musthafa MM, Nadeem A, Nawaz A, Javed MM, Marikar F. Molecular Diversity Analysis of Jattal and Dera Din Panah Goat breeds of Pakistan using microsatellite markers. J Hellenic Vet Med Soc. 2018;69(1):791–6.
21. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinform. 2014;30(15):2114–20.
22. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinform. 2009;15:1754–60.
23. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Mart G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. Bioinform. 2009;25:2078–9. 1000 Genome Project Data Processing Subgroup.
24. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S. The genome analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.
25. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet. 2007;81(5):1084–97. https://doi.org/10.1086/521987.
26. Asadollahpour Nanaei H, Esmailizadeh A, Ayatollahi Mehrgardi A, et al. Comparative population genomic analysis uncovers novel genomic footprints and genes associated with small body size in Chinese pony. BMC Genom. 2020;21:496. https://doi.org/10.1186/s12864-020-06887-2.

27. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE. 2010;5(3):e9490. https://doi.org/10.1371/journal.pone.0009490.

28. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for wholegenome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.

29. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76–82.

30. Alexander DH, Novembrem J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655–64.

31. Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of population structure using dense haplotype data. PLoS Genet. 2012;8:e1002453.

32. Weir BS, Cockerham CC. Estimating F-Statistics for the analysis of Population structure. Evol. 1984;38(6):1358–70.

33. Spolaore E, Wacziarg R. The Diffusion of Development. Q J Econ. 2009;124(2):469–529. https://doi.org/10.1162/qjec.2009.124.2.

34. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. Bioinform. 2011;27(15):2156–8.

35. Mota-Rojas D, Titto CG, de Mira Geraldo A, Martínez-Burnes J, Gómez J, Hernández-Ávalos I, Casas A, Domínguez A, José N, Bertoni A, Reyes B, Pereira AMF. Efficacy and function of feathers, hair, and Glabrous skin in the Thermoregulation strategies of domestic animals. Anim (Basel). 2021;6(1112). https://doi.org/10.3390/ani11123472.

36. Saadatabadi L, Mohammadabadi M, Nanaei HA, Ghanatsaman ZA, Stavetska RV, Kalashnyk O, Kochuk-Yashchenko OA, Kucher DM. Unraveling candidate genes related to heat tolerance and immune response traits in some native sheep using whole genome sequencing data. Small Rumin Res. 2023;225:107018.

37. Cai Y, Fu WW, Cai D, et al. Ancient genomes reveal the Evolutionary History and Origin of Cashmere-Producing goats in China. Mol Biol Evol. 2020;37:2099–109. https://doi.org/10.1093/molbev/msaa103.

38. Yaqoob E, Javed S, Khan SA. Trauma care in the face of climate change in Pakistan. Correspondence. 2023;401:1769–70. https://doi.org/10.1016/S0140-6736(23)00927-3.

39. Guenther CA, Tasic B, Luo L, Bedell MA, Kingsley DM. A molecular basis for classic blond hair color in europeans. Nat Genet. 2014;46(7):748–52. https://doi.org/10.1038/ng.2991.

40. Yurchenko AA, Daetwyler HD, Yudin N, et al. Scans for signatures of selection in Russian cattle breed genomes reveal new candidate genes for environmental adaptation and acclimation. Sci Rep. 2018;8:12984. https://doi.org/10.1038/s41598-018-31304-w.

41. Psifidi A, Banos G, Matika O, et al. Genome-wide association studies of immune, disease and production traits in indigenous chicken ecotypes. Genet Sel Evol. 2016;48:74.

42. Zheng Z, Xihong W, Li M, Li Y, Yang Z, Xiaolong W, Pan X, Gong M, Zhang Y, Guo Y, et al. The origin of domestication genes in goats. Sci Adv. 2020;6:eaaz5216.

43. Saif R, Mahmood T, Ejaz A, Fazlani SA, Zia S. Whole-genome selective sweeps analysis in Pakistani Kamori goat. Gene Rep. 2022;26:101429. https://doi.org/10.1016/j.genrep.2021.101429.

44. Anderson L. Domestic animals as models for biomedical research. Ups J Med Sci. 2016;121(1):1–11. https://doi.org/10.3109/03009734.2015.1091522.

45. Gonzalez-Rivas PA, Chauhan SS, Ha M, Fegan N, Dunshea FR, Warner RD. Effects of heat stress on animal physiology, metabolism, and meat quality: a review. Meat Sci. 2020;162:108025.

46. Habeeb AAM. Deterioration effects of heat stress on farm animals performance in tropical and subtropical regions. World J Biol Pharm Health Sci. 2020;4(2):007–25.

47. Thornton P, Nelson G, Mayberry D, Herrero M. Increases in extreme heat stress in domesticated livestock species during the twenty-first century. Glob Chang Biol. 2021;27:5762–72. https://doi.org/10.1111/gcb.15825.

48. Kaushik R, Arya A, Kumar D, Goel A, Rout PK. Genetic studies of heat stress regulation in goat during hot climatic condition. J Therm Biol. 2023;113:103528. https://doi.org/10.1016/j.jtherbio.2023.103528.

49. Morrison-Graham K, Takahashi Y. Steel factor and c-Kit receptor: from mutants to a growth factor system. BioEssays. 1993;15(2):77–83.

50. Talenti A, Bertolini F, Williams J, Moaeen-Ud-Din M, Frattini S, Coizet B, Pagnacco G, Reecy J, Rothschild MF, Crepaldi P, Italian Goat Consortium. Genomic Analysis Suggests KITLG is responsible for a Roan Pattern in two Pakistani Goat breeds. J Hered. 2018;109(3):315–9. https://doi.org/10.1093/jhered/esx093.

51. Freitas PHF, Wang Y, Yan P, Oliveira HR, Schenkel FS, Zhang Y, Xu Q, Brito LF. Genetic Diversity and Signatures of Selection for Thermal Stress in Cattle and other two Bos Species adapted to Divergent climatic conditions. Front Genet. 2021;12. https://doi.org/10.3389/fgene.2021.604823.

52. Thirunavukkarasu M, Rishi MT, Pradeep SR, Swaminathan S, Accorsi D, Palesty JA, Maulik N. Heat shock protein A12B gene therapy improves perfusion, promotes neovascularization, and decreases fibrosis in a murine model of Hind limb ischemia. Surgery. 2021;170(3):969–77.

53. Habib HN, Saleh WMM, Gheni QJ. Genetic profiling of HSP70 gene in local Iraqi goats. Brazilian J Vet Med. 2022;5:44:e004121. https://doi.org/10.29374/2527-2179.bjvm004121.

54. Okpeku M, Esmailizadeh A, Adeola AC, Shu L, Zhang Y, Wang Y, Sanni TM, Imumorin IG, Peters SO, Zhang J, Dong Y, Wang W. Genetic Variation of Goat Interferon Regulatory Factor 3 Gene and its implication in Goat Evolution. PLoS ONE. 2016;11(9):e0161962. https://doi.org/10.1371/journal.pone.0161962.

55. Ballotti R, Cheli Y, Bertolotto C. The complex relationship between MITF and the immune system: a Melanoma ImmunoTherapy (response) factor? Mol Cancer. 2020;19:170. https://doi.org/10.1186/s12943-020-01290-7.

56. Zhao X, Wang Y, Li J, et al. RFC2: a prognosis biomarker correlated with the immune signature in diffuse lower-grade gliomas. Sci Rep. 2022;12:3122. https://doi.org/10.1038/s41598-022-06197-5.

57. Onzima RB, et al. Genome-wide characterization of selection signatures and runs of homozygosity in Ugandan goat breeds. Front Genet. 2018;9:318.

58. Kim H, Ali O, Shim M, et al. Insulin-like growth factor binding Protein-3 induces insulin resistance in Adipocytes in Vitro and in rats. Vivo Pediatr Res. 2007;61:159–64. https://doi.org/10.1203/pdr.0b013e31802d8a30.

59. Sun Y, Cai W, Zhang X, Hui T, Xu Y, et al. Association analysis for SNPs of MSTN and IGFBP-3 genes with body size and other production traits in Liaoning Cashmere Goats. Anim Biotechnol. 2023;34(5):1796–806. https://doi.org/10.1080/10495398.2022.2051043.

60. AL-Khuzai HMH, AL-Anbari NN. Relationship of igfbp-3 gene polymorphism with some of economical traits in Iraqi Awassi ewes. 2019. https://www.iasj.net/iasj/download/88126097c7cdbc8d.

61. Kovrov O, Landfors F, Saar-Kovrov V, Näslund U, Olivecrona G. Lipoprotein size is a main determinant for the rate of hydrolysis by exogenous LPL in human plasma. J Lipid Res. 2022;63(1):100144.

62. Moioli B, D'Andrea M, Pilla FJSRR. Candidate genes affecting sheep and goat milk quality. Small Rumin Res. 2007;68:179–92.

63. Li G, Zhang S, Wang H, Liang L, Liu Z, Wang Y, Xu B, Zhao H. Differential expression characterisation of the heat shock proteins DnaJB6, DnaJshv, DnaJB13, and DnaJB14 in Apis cerana cerana under various stress conditions. Front Ecol Evol. 2022;10. https://doi.org/10.3389/fevo.2022.873791.

64. Ritossa F. A new puffing pattern induced by temperature shock and DNP in Drosophila. Experientia. 1962;18:571–3.

65. Mohamadipoor Saadatabadi L, Mohammadabadi MR, Amiri Ghanatsaman Z, et al. Signature selection analysis reveals candidate genes associated with production traits in Iranian sheep breeds. BMC Vet Res. 2021;17:369. https://doi.org/10.1186/s12917-021-03077-4.

66. Liu H, Zhang M, Han H, et al. Comparison of the expression of cytokine genes in the bursal tissues of the chickens following challenge with infectious bursal disease viruses of varying virulence. Virol J. 2010;7:364. https://doi.org/10.1186/1743-422X-7-364.

67. Cheon H, et al. IFNβ-dependent increases in STAT1, STAT2, and IRF9 mediate resistance to viruses and DNA damage. EMBO J. 2013;32:2751–63.

68. Serranito B, et al. Local adaptations of Mediterranean sheep and goats through an integrative approach. Sci Rep. 2021;11:1–17.

69. Wang MD, Dzama K, Hefer CA, Muchadeyi FC. Genomic population structure and prevalence of copy number variations in South African Nguni cattle. BMC Genom. 2015;16:1–16.

70. Manzari Z, Mehrabani-Yeganeh H, Nejati-Javaremi A, Moradi MH, Gholizadeh M. Detecting selection signatures in three Iranian sheep breeds. Anim Genet. 2019;50:298–302.

71. Szyda J, Komisarek J. Statistical modeling of candidate gene effects on milk production traits in dairy cattle. J Dairy Sci. 2007;90(6):2971–9. https://doi.org/10.3168/jds.2006-724.

72. Ghanbari Baghenoey S, Ansari Mahyari S, Asadollahpour Nanaei H, Rostami M, Edriss MA. Association of LEPR gene polymorphism with milk yield and age at first calving in the Iranian holstein dairy cows. J Livest Sci Technol. 2014;2(2):39–42.

Zhang *et al. BMC Genomics*   (2024) 25:477

Page 11 of 11

73. Zhou F, Yang Q, Lei C, Chen H, Lan X. Relationship between genetic variants of POU1F1, PROP1, IGFBP3 genes and milk performance in Guanzhong dairy goats. Small Rumin Res. 2016;140:40–5.

74. 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.

75. El-Shorbagy HM, Abdel-Aal ES, Mohamed SA, El-Ghor AA. Association of PRLR, IGF1, and LEP genes polymorphism with milk production and litter size in Egyptian Zaraibi goat. Trop Anim Health Prod. 2022;54(5):321. https://doi.org/10.1007/s11250-022-03316-2.

76. Hou JX, An XP, Song YX, Wang JG, Ma T, Han P, Fang F, Cao BY. Combined effects of four SNPs within goat PRLR gene on milk production traits. Gene. 2013;529(2):276–81. https://doi.org/10.1016/j.gene.2013.07.057.

77. Freebern E, Santos DJA, Fang L, Jiang J, Gaddis KLP, Liu GE, VanRaden PM, Maltecca C, Cole JB, Ma L. GWAS and fine-mapping of livability and six disease traits in Holstein cattle. BMC Genom. 2020;21(1):41. https://doi.org/10.1186/s12864-020-6461-z.

78. Matsumoto H, Sasaki K, Bessho T, et al. The SNPs in the ACACA gene are effective on fatty acid composition in holstein milk. Mol Biol Rep. 2012;39:8637–44. https://doi.org/10.1007/s11033-012-1718-5.

79. Kęsek-Woźniak MM, Wojtas E, Zielak-Steciwko AE. Impact of SNPs in ACACA, SCD1, and DGAT1 genes on fatty acid profile in bovine milk with regard to lactation phases. Anim (Basel). 2020;810(6):997.

80. Liu D, Xu Z, Zhao W, Wang S, Li T, Zhu K, Liu G, Zhao X, Wang Q, Pan Y, Ma P. Genetic parameters and genome-wide association for milk production traits and somatic cell score in different lactation stages of Shanghai Holstein population. Front Genet. 2022;13:940650. https://doi.org/10.3389/fgene.2022.940650.

81. Jiang J, Ma L, Prakapenka D, VanRaden PM, Cole JB, Da Y. A large-scale genome-wide Association study in U.S. Holstein cattle. Front Genet. 2019;14:10412. https://doi.org/10.3389/fgene.2019.00412.

82. Saravanan KA, Panigrahi M, Kumar H, Parida S, Bhushan B, Gaur GK, et al. Genomic scans for selection signatures revealed candidate genes for adaptation and production traits in a variety of cattle breeds. Genomics. 2021;113:955–63. https://doi.org/10.1016/j.ygeno.2021.02.009.

83. Menon R, Patel AB, Joshi C. Comparative analysis of SNP candidates in disparate milk yielding river buffaloes using targeted sequencing. PeerJ. 2016;4:e2147. https://doi.org/10.7717/peerj.2147.

84. Nafikov RA, Schoonmaker JP, Korn KT, Noack K, Garrick DJ, Koehler KJ, Minick-Bormann J, Reecy JM, Spurlock DE, Beitz DC. Polymorphisms in lipogenic genes and milk fatty acid composition in Holstein dairy cattle. Genomics. 2014;104(6):572–81.

85. Gebreyesus G, Buitenhuis AJ, Poulsen NA, Visker M, Zhang Q, van Valenberg HJF, et al. Multi-population GWAS and enrichment analyses reveal novel genomic regions and promising candidate genes underlying bovine milk fatty acid composition. BMC Genom. 2019;20(1):178. https://doi.org/10.1186/s12864-019-5573-9.

Publisher's Note