

RESEARCH

Open Access



# Dissecting genomes of multiple yak populations: unveiling ancestry and high-altitude adaptation through whole-genome resequencing analysis

Sheikh Firdous Ahmad<sup>1\*†</sup>, Munish Gangwar<sup>1†</sup>, Amit Kumar<sup>1\*</sup>, Amod Kumar<sup>2</sup>, Mahesh Shivanand Dige<sup>2</sup>, Girish Kumar Jha<sup>3</sup>, Gyanendra Kumar Gaur<sup>1,4</sup> and Triveni Dutt<sup>1</sup>

## Abstract

The present study was undertaken to elucidate the population structure and differentiation of Indian yak from Chinese and wild cohorts on genome-wide scale by identifying the selection sweeps and genomic basis of their adaptation across different comparisons while analyzing whole genome sequencing (WGS) data using latest bioinformatics tools. The study included 105 individuals from three distinct yak populations i.e., Indian yak ( $n = 29$ ); Chinese yak ( $n = 61$ ) and wild yak ( $n = 15$ ), hypothesized to be related along the evolutionary timescale. Efficient variant calling and quality control in GATK and PLINK programs resulted in around 1 million (1,002,970) high-quality (LD-independent) SNPs with an average genotyping rate of 96.55%. The PCA, ADMIXTURE and TREEMIX analysis revealed stratification of the yak groups into three distinct clusters. The empirical distribution pattern of minor allele frequency (MAF) of SNPs on genome-wide scale was also elucidated for three yak cohorts revealing unique distribution across five different bins. The selection signature analysis revealed candidate genes that are important for the adaptation of Indian yak against harsh environmental conditions in their habitats. Under *iHS* analysis, several genes were identified to be under selection pressure in Indian yak including *ABCA12*, *EXOC1*, *JUNB*, *KLF1*, *PRDX2*, *NANOS3*, *RFX1*, *RFX2*, and *CACNG7*. On the other hand, across population analysis revealed the genes like *NR2F2*, *OSBPL10*, *CIDEA*, *WFIKKN2*, *ADCY*, *THSD7A*, *ADGRB3*, *TRPC1*, *VASH2*, and *ABHD5* to be part of selective sweeps under these comparisons. A total of 53 genes were found common between intra- and inter-population selection signature analysis of Indian yak. Notably, the genes harbouring the SNPs under selection pressure were significant for adaptation traits including lipidogenesis, energy metabolism, thermogenesis, hair follicle formation, oxidation–reduction reactions, hypoxia and reproduction. These genes may be evaluated as candidate genes for livestock adaptation to harsh environmental conditions and to further the research and application in the present era of climate change.

**Keywords** Adaptation, MAF, Population structure, Selection signature, WGS, Yak

<sup>†</sup>Sheikh Firdous Ahmad and Munish Gangwar contributed equally to the work and share the first authorship.

\*Correspondence:

Sheikh Firdous Ahmad  
firdousa61@gmail.com

Amit Kumar  
vetamitchandan07@gmail.com

Full list of author information is available at the end of the article



## Introduction

Domestication of farm animals has driven the socio-economic transition of human civilizations from a nomadic pastoralist style to the agricultural settler. The domestication process of livestock began in the Middle-East around 11,000 years ago which later spread to other geographical areas [1, 2]. The domestication process has led to evolution of these animals with respect to their anatomical, physiological, haematological, and morphological features as adaptation to cope with environmental stressors, mainly related to climatic conditions, management, nutritional support and thermo-precipitation factors.

Yak (*Bos grunniens*) is a domesticated bovine species with wild yak (*Bos mutus*) as its progenitor. These are medium-sized docile animals, classed as bovines, with characteristic thick skirting of hair all over the body. Thick hair along with the black coat colour is considered an adaptation in yak in order to conserve body heat in cold high-altitude environment and adapt to rapid fluctuations in temperatures. Yak are inhabitants of the vast high-altitude terrains including the Himalayan regions of multiple nations i.e., Afghanistan (Gilgit-Baltistan), Nepal, India, Tibet (Tibetan plateau), Tajikistan, and Mongolia [3]. Though the wild yak population is currently dwindling, the gene flow between wild and domesticated cohorts is still ongoing; it thus promises an apt model to study various aspects of evolutionary processes [4]. It is one of the important livestock species that has evolved through millions of years of evolution, domestication and adaptation, especially with respect to their long-term exposure to cold and hypoxic conditions at high altitudes. These adaptations vary across morphological to biochemical characteristics and include enlarged heart and lungs, shorter tongue, altered haemoglobin structure for increased blood oxygen availability, lack of hypoxic pulmonary vasoconstriction, and optimal foraging behaviour [5–7].

In India, yak are distributed across few regions including Union Territory of Ladakh (dry arid climate), Himachal Pradesh (cold arid climate), and Arunachal Pradesh & Sikkim (wet and cold climate). Yak rearing provides a multitude of benefits to farmers and inhabitants, including milk, meat, wool, transportation, dung for fuel, and hides for various purposes, etc. Additionally, they are excellent pack animals for transporting goods in high-altitude areas. Based on near-optimal adaptation to harsh environment and subsequent usefulness, these animals have closely knit into the socio-culturo-economic fabric and livelihood of high-altitude societies. The global yak population is estimated to be around 14 million [8]. On the other hand, the wild yak population has now reached the vulnerability class as per the latest IUCN Red List, with less than 10,000 existing mature animals [9].

According to the latest Indian livestock census report of 2019, the yak population in India is 0.58 lakh (58,000 animals), recording a concerning decrease of 24.08 percent over previous levels of 2012. Courtesy the recent efforts from various stakeholders, the Indian yak populations have been documented and categorized into four groups based on different geographical locations wherein they are reared in India i.e., Arunachali, Himachali, Ladakhi, and Sikkimi. Besides their adaptive characteristics to varied agro-climatic conditions, the distinguishing features in these four populations include the coat colour, horn pattern, hair characteristics and unique markings.

Anthropogenic efforts of selection along with intricate events of natural selection, domestication and cross-species introgression have left its unique footprints on the yak genome [4, 10]. Molecular markers have proven helpful in planning breed improvement programs across various farm animal species. The genome-wide data on molecular markers may be helpful in elucidating the adaptation of highlander population and improve the understanding of complex traits. Single nucleotide polymorphisms (SNPs) are widely used genetic markers in genomic studies due to their nearly-even distribution and amenability as compared to other similar markers [11]. Technological advancements and cost reductions in next-generation sequencing technologies have enabled swift and precise identification of SNP markers for larger animal cohorts. Earlier, double-digestion restriction associated DNA (ddRAD) sequencing has been used to identify SNPs in Indian yak and study various aspects of population genomics [12, 13]. Although it is cost-effective than WGS, it has limitations in analyzing vast genomes and may not be accurate for model species where fairly accurate reference genomes are available. Using next-generation sequencing technology, WGS promises deeper insights into the genome of farm animal species, including genetic variants, insertion/deletion polymorphisms (indels), structural variations, copy number variations (CNVs), and quantitative trait loci analysis [14, 15].

The adaptations in these highlander livestock species have led to modification in gene frequencies that help them adapt to adverse climatic conditions [16, 17]. Genomic studies have helped elucidate the candidate genes that are linked to similar adaptations. These candidate genes gain additional significance in the era of climate change and global warming. Yak are important genetic resources to elucidate the genetic variants related to adaptation under challenging high-altitude and hypobaric conditions. The WGS approach has been utilized by a number of researchers in yak populations all over the world in order to investigate signatures of evolution and domestication [18], understanding of adaptation to high altitude [8], evolution and divergence [19]. However, our

understanding of evolutionary origin of Indian yak is not complete, especially in comparison of Chinese and wild yak populations. To the best of our knowledge, no study has ever reported the elucidation of population structure and selection sweeps in Indian yak as compared to wild ancestors and Chinese counterparts using WGS data. In view of the above points, the present study was aimed to elucidate the differentiation of Indian yak from wild and Chinese counterparts on genome-wide scale by identifying the selection sweeps across different comparisons while using WGS data and latest bioinformatics tools.

## Materials and methods

### Populations and resequencing data

The present study leverages the concept of data reuse in livestock genomics research, offering solutions to multiple research questions. An attempt was made to build upon the already generated WGS data from Indian, Chinese, and wild yak populations (for further details, refer [18, 20]). The WGS data was earlier generated on 30 yak across three populations i.e., Himachali, Ladakhi, and Arunachali ( $n=10$  per population) using the genomic DNA available in the repository at ICAR-National Bureau of Animal Genetic Resources, Karnal, India (accession code: PRJNA803425). The 30 Indian yak individuals were randomly selected in the analysis within the yak cohorts (Arunachali, Himachali and Ladakhi). The Illumina<sup>R</sup> NovaSeq 6000 platform, offering efficient resequencing of genomes with high efficiency and cost effectiveness [21], was employed using the S4 flowcell to generate 150 bp long paired-end reads. Qiu et al. [18] reported generation and analysis on 84 unrelated yak with Chinese ( $n=69$ ) and wild ( $n=15$ ) inheritance using paired end libraries of 500 bp reads *via* Hiseq 2000 platform. Standardized Illumina<sup>R</sup> protocols ensured consistent sequencing and base calling for all samples in both the studies. Hence, the present study was undertaken on 114 individuals from three distinct yak populations i.e., Indian yak ( $n=30$ ); Chinese yak ( $n=69$ ) and wild yak ( $n=15$ ), hypothesized to be related along the evolutionary timescale.

### Alignment, post-alignment processing and variant calling

The raw sequencing data was processed for quality trimming using FastQC v0.12.0 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>; accessed - May 2024), while adapters and poor quality reads were removed using TrimGalore v0.6.5 [22, 23]. The quality of trimmed reads was reassessed using FastQC. The latest BosGru3.1 domestic yak, chromosome-level, reference genome assembly (*Bos grunniens* reference genome BosGru3.1; GenBank: GCA\_005887515.3; genome size: 2.8 Gb with 312 unplaced scaffolds; latest update: April 2021) was

indexed using the Burrows-Wheeler Aligner (BWA, v0.7.17) with default settings [24]. Post read-filtering and genome indexing, the trimmed reads were aligned to the reference genome assembly using the BWA program with default settings. Subsequently, the BAM files were sorted and indexed using SAMtools v1.17 [25]. The SM read tags were then appended to the mapped reads. Subsequently, the PCR duplicates in BAM files were identified *via* the *MarkDuplicates* function of Picard tools v2.25.1 (<https://broadinstitute.github.io/picard/>, accessed - May 2024) and finally removed. Genome Analysis ToolKit (GATK, v4.4.0.0, available at <https://software.broadinstitute.org/gatk>, accessed - May 2024) was used for the variant calling using standard methodology [26].

The *HaplotypeCaller* function of GATK toolkit was used to identify the genetic variants (SNPs and indels) from WGS data on 114 yak samples individually. It employed probabilistic modelling to reconstruct haplotypes and identify genetic variants based on the differences between these reconstructed haplotypes. The *CombineVCF* tool of GATK was used to merge multiple variant calling format (VCF) files into a single dataset for further analysis. Subsequently, the *GenotypeGVCFs* tool of GATK was employed to convert genome VCF (GVCF) files back into a standard VCF format containing the genotyping information for specific samples. The *SelectVariants* tool was used to extract a subset of SNP variants from a VCF file. Ultimately, the filtration of genetic variants was undertaken using the *VariantFiltration* tool in GATK while employing a combination of basic expression and mapping quality criteria, removing variants with low confidence, mapping errors, and/or strand bias. The various thresholds used for variant filtration included that of quality depth (QD) score below 2; Phred-scaled genotype quality (QUAL) score below 30; mapping quality (MQ) score lower than 40; Fisher strand (FS) bias score exceeding 60, strand odds ratio (SOR) score greater than 3, mapping quality rank sum test (MQRankSum) score less than  $-12.5$  and Read Position Rank Sum Test (ReadPosRankSum) score less than  $-8$ . Overall, the filtration process was kept stringent enough to remove variants that were likely to be unreliable due to low confidence, mapping errors, strand bias or other similar issues.

### Quality control of genome-wide SNP data

Initially, the SNP variant files in variant calling format (VCF) were converted to PLINK-friendly binary files using PLINK v1.9 program [27]. Subsequently, the SNP markers with no definite coordinates, and those present on sex chromosomes or mitochondria were removed; filtering only the autosomal markers with definite coordinates for downstream processing. Furthermore, the

quality control thresholds pertaining to genotyping rate of individuals (10%), call rate of SNPs (10%), minor allele frequency (5%) and Mendelian errors (5% Mendel errors for families and 10% Mendel error rate for SNPs) were applied on the WGS data. Post-filtering, in order to remove the SNP markers contributing redundant information due to their close physical proximity on the chromosomes, the markers in linkage disequilibrium with  $r^2$  estimate greater than 0.2 were removed using *indep-pairwise* function in PLINK. The window size of 50 SNPs in a group with physical distance of 5 base pairs was considered for LD pruning. The filtered dataset with independent and non-redundant markers was used for downstream analysis including assessment of population structure, genetic diversity, and identification of selection sweeps.

### Population structure

In order to gain a comprehensive understanding of the population structure and potential historical events that shaped the genetic diversity of the yaks under study, the population structure of different yak populations was delineated using a suite of complementary bioinformatics tools. Initially, principal component analysis (PCA) was performed using WGS data on animals from three yak populations by employing Genome-wide Complex Trait Analysis (GCTA) v1.93 program [28]. Briefly, a genomic relationship matrix (GRM) was elucidated using binary PLINK-formatted files as input. GRM was subsequently used to delineate the eigenvalues and eigenvectors for the first 20 principal components (PCs). The eigenvectors, corresponding to the first three PCs, were plotted in an R-programming environment using *ggplot2* R-package v3.4.4 [29] for visualization. Moreover, TreeMix v1.13 program [30] was employed to infer the population splits, admixture events, potential historical relationship or gene flow among yak population using a maximum-likelihood approach while assuming two migration events. A customized python script was used to convert the PLINK binary data into TreeMix-friendly format while plotting and visualization was done in R-programming environment. Finally, a model-based approach was employed in ADMIXTURE v1.30 program [31] to elucidate the genomic breed clustering among different yak populations. Additionally, the phylogenetic relationship was assessed among three yak cohorts based on genetic distances in Molecular Evolutionary Genetics Analysis (MEGA) v11.0.13 program [32].

### Genetic diversity: Linkage disequilibrium and minor allele frequency

The genetic diversity of Indian yak cohort was elucidated in terms of empirical distribution pattern of minor allele frequency and proportional distribution across different bins. A total of 5 bins i.e., I:  $x < 0.10$ ; II:  $0.10 < x < 0.20$ , III:  $0.20 < x < 0.30$ , IV:  $0.30 < x < 0.40$  and V:  $x > 0.40$  were analyzed to categorize the MAF values. The MAF levels were elucidated using PLINK program while customized scripts were used to elucidate the proportion of markers in different bins. The plots were generated for graphical representations of the data. The linkage disequilibrium among the markers was assessed based on  $r^2$  parameters in PLINK program.

### Selection signature analysis

In order to elucidate the selection signatures keeping Indian yak as a central point, two approaches were employed i.e., intra-population (within the Indian yak population) and inter-population (comparing Indian yak with Chinese yak and wild yak populations).

#### *Within-population selection signatures in Indian yak (iHS)*

The intra-population signatures of selection within the Indian yak population were investigated using the haplotype-based *iHS* statistic [33]. The VCF files, phased in BEAGLE program [34], were used in *rehh* package within R-programming environment [35, 36] to calculate the *iHS* statistic for each marker during the selection scan. The *iHS* statistic was indicative of the decay of extended haplotype homozygosity (EHH) surrounding a specific SNP at genome-wide scale within a population. The haplotype files were initially converted into a format compatible with *rehh*. Subsequently, a chromosome-wise scan was performed across all markers to compute the integrated EHH (iHH) values across the yak autosomes. The results for each chromosome were then combined to generate genome-wide estimates (for autosomes only). Following this, the unstandardized genome-wide *iHS* scores were calculated. Standardization of *iHS* was then performed within defined bins, excluding markers with a minor allele frequency (MAF) less than or equal to 0.01 [33]. The absolute *iHS* values were averaged across non-overlapping 50 kb windows using default settings [37]. Manhattan plots were generated using *Manhattan* package to visualize the *iHS* estimates and corresponding *p*-values on a logarithmic scale for all SNP markers across the bovine genome. Genomic regions with an average *iHS* score exceeding five standard deviations above the mean were considered putative candidate regions under selection. These regions were then subjected to further



downstream analysis. The functional significance of different genes was identified using mining approach from various databases.

#### **Across population selection signatures (XP-EHH)**

Here, the cross-population extended haplotype homozygosity (XP-EHH) statistic [37] was used to detect selection signatures across different comparisons involving Indian yak. The regions under selection in these comparisons were identified by comparing the EHH profiles of two populations (target and reference). In essence, the EHH estimates surrounding each SNP on a genome-wide scale were compared between contrasting populations using haplotype-phased *VCF* files in *rehh* package. For each SNP, the unstandardized XP-EHH score represented the log ratio of the integrated EHH (iES) values from the two compared populations. Following this, the XP-EHH scores were standardized using the mean and standard deviation of all unstandardized scores (unXP-EHH). The *p*-values were then transformed on a two-sided logarithmic scale. Positive and negative XP-EHH values indicated the selective sweeps in the observed and reference populations, respectively. The Indian yak was considered the observed population for all comparisons while that of Chinese and wild cohorts were taken as reference. XP-EHH scores and corresponding *p*-values (logarithmic scale) were calculated for each SNP. These results were visualized as Manhattan plots for each comparison (Indian yak vs. each reference population). The functional significance of different genes was identified using mining approach from various databases.

## **Results and discussion**

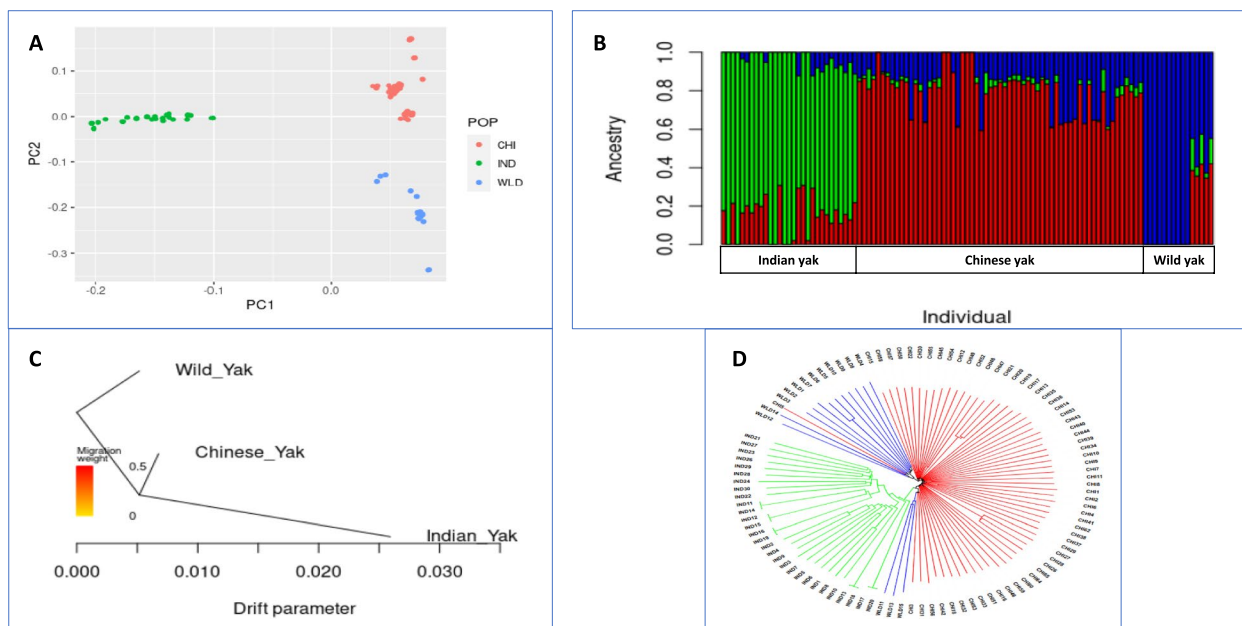
Yak is an important animal genetic resource that contributes significantly to the socio-economic fabric of masses in hilly and mountainous areas across various regions of the globe including India. It possesses unique characteristics with remarkable adaptation to challenges in hilly and mountainous terrains, mainly related to cold and hypoxic environment with limited feed availability. Molecular genetic analysis using genome-wide SNP markers, deduced from WGS analysis on multiple yak lineages, provides an optimal opportunity to comprehensively elucidate the deep insights into the genetic variants and biological pathways involved in the evolution and adaptation of these unique populations. The present study attempted to reveal deep insights into the evolution and adaptation of Indian Yak population in contrast to Chinese and wild counterparts while employing latest tools of bioinformatics and population genomics on WGS data.

#### **Variant calling and population structure analysis**

Efficient variant calling and quality control ensure unbiased and reliable predictions on the studied population. Using strict quality control during variant calling in GATK and subsequently post-SNP identification in PLINK resulted in around 1 million (1,002,970) high-quality SNPs on 105 individuals with an average genotyping rate of 96.55%. A total of nine problematic individuals (one Indian and eight Chinese yak) were removed based on the results from population structure analysis i.e., principal component analysis and model-based analysis in ADMIXTURE program. On population structure analysis, three population groups were stratified using PCA. The first two PCs comprehensively stratified various populations (Indian, Chinese and wild yaks) into separate clusters (Fig. 1A). It is noteworthy that PC#1 stratified Indian yak away from other two populations; while PC#2 stratified the Chinese and wild yak populations from each other. The Chinese and wild yak population were grouped closer to each other. The first three PCs were able to describe 10.28% of the total variation in the dataset for 105 animals with 1,002,970 SNP markers (matrix size: 105 × 1,002,970). On the other hand, the model-based clustering analysis of population structure in the ADMIXTURE program revealed early stratification of Indian yak from the other two populations at  $K=2$ . Furthermore, at  $K=3$ , the three clusters were identified as depicted in Fig. 1B. At  $K=3$ , the individuals of Indian, Chinese and wild yak population showed average genomic breed clustering estimates of approximately 82, 80 and 84% into different clusters, respectively. Similarly, TreeMix analysis revealed a similar stratification pattern using likelihood-based phylogenetic analysis with an assumption of twin migration events across three yak populations. Figure 1C depicts the likelihood-based phylogenetic tree from TreeMix analysis. Finally, the neighbour-joining tree depicting the level of genetic relationship among the three yak cohorts has been depicted in Fig. 1D.

#### **Extent of MAF and linkage disequilibrium in yak cohorts**

The empirical distribution of MAF estimates for genome-wide SNPs was revealed for Indian yak which revealed highest proportion of markers (25.92%) to fall in bin I with MAF lesser than 0.10. On the other hand, the proportion of markers in other four bins was 22.66% (bin II:  $0.10 < x < 0.20$ ), 14.57% (bin III:  $0.20 < x < 0.30$ ), 16.07% (bin IV:  $0.30 < x < 0.40$ ) and 20.78% (bin V:  $x > 0.40$ ), respectively. On the other hand, the summary of the SNP markers at whole-genome scale along with average minor allele frequency estimates for each autosomal chromosome (BosGru) has been summarized in Table 1. The



**Fig. 1** Population stratification of three yak cohorts based on WGS data; **A** The principal component analysis plot showing the stratification (and clustering) of different Yak populations under study. **B** The admixture plot (K=3) showing the population structure of individuals from three Yak populations. **C** TreeMix based plot for three Yak populations. **D** Neighbour-Joining tree depicting the relation between three yak cohorts. (PC: Principal Component; POP: Population; CHI: Chinese Yak; IND: Indian Yak, WLD: Wild Yak)

ROH distribution and inbreeding analysis cohorts have been reported earlier for Indian yak cohorts [38]. Table 2 presents the distribution pattern of MAF across different bins in three yak cohorts (Indian, Chinese and wild) based on WGS data. A higher proportion of the markers in three cohort represented MAF bin I, II or V. Whereas the proportion of markers with intermediate frequencies (bins III and IV) were lesser as compared to other bins. The trend was similar across the three yak cohorts. The average  $r^2$  estimate for all combination of SNPs on genome-wide scale in Indian, Chinese and wild yak cohorts was 0.34, 0.36 and 0.40, respectively.

**Selection signature analysis**

The present study employed twin approaches to elucidate the genomic basis of adaptation in yak through intra- and inter-population selection sweeps.

**Intra-population *iHS* selection signals within Indian yak**

Adaptation of animals (yak) to high-altitude conditions poses multiple challenges in terms of hypobaric atmosphere and increased exposure to harsh sun rays [39]. The climatic conditions in high-altitude regions are mostly erratic and animals need to develop adequate rapid signalling mechanisms to counter these adversities. Similarly, special regulation of energy and lipid metabolism is required for adaptation to extremes of temperature, especially in terms of thermoregulation [40]. Multiple

SNPs were found to be under selection pressure and part of selective sweeps through intra-population analysis in Indian yak using *iHS* methodology. These SNPs were present on multiple autosomes of yak genome including BosGru\_5, BosGru\_6, BosGru\_7, BosGru\_8, BosGru\_14, BosGru\_17, BosGru\_18 and BosGru\_20 (Fig. 2). Several genes were found overlapping these SNPs within the 20 kb window including *ABCA12*, *EXOC1*, *JUNB*, *KLF1*, *PRDX2*, *NANOS3*, *RFX1*, *RFX2*, and *CACNG7*. The functioning of *ABCA12* gene i.e., ATP-binding cassette sub-family A member 12 (mapped on BosGru\_2) has earlier been related to the skin protection of animals against harsh solar radiations in high-altitude habitats [41]. On the other hand, the comparative genomic analysis of the wild African goat (Nubian ibex) revealed the *ABCA12* and other genes, related to skin barrier development and function, to be under positive selection pressure [42]. Another gene i.e., *EXOC1* or exocyst complex component 1 (mapped on BosGru\_6), known for its critical role in protein and lipid transportation, has been reported to play an important role in cellular functions across various organisms [43]. The *JUNB* gene i.e., JunB proto-oncogene, AP-1 transcription factor subunit (mapped on BosGru\_8) plays crucial role in regulating adipogenesis in yak, influencing the process through which the pre-adipocytes mature into fully functional adipocytes [44]. Adipogenesis is essential for energy storage, insulation, and metabolic regulation, making it a vital biological process

**Table 1** Summary of the SNP markers at whole-genome scale along with average minor allele frequency (MAF) estimates for each autosomal chromosome (BosGru) in Indian yak

Chromosome	Chromosome size (length in mb)	No. of SNP Markers	Average SNP interval (mb)	Average MAF
BosGru_1	172.39	51,315	0.0034	0.2196
BosGru_2	150.29	41,574	0.0036	0.2163
BosGru_3	132.31	50,933	0.0026	0.2455
BosGru_4	130.91	44,709	0.0029	0.2335
BosGru_5	129.74	54,937	0.0024	0.2447
BosGru_6	128.86	46,226	0.0028	0.2278
BosGru_7	124.23	44,513	0.0028	0.2341
BosGru_8	120.92	41,987	0.0029	0.2371
BosGru_9	118.90	34,589	0.0034	0.2185
BosGru_10	114.39	37,814	0.0030	0.2236
BosGru_11	96.16	43,135	0.0022	0.2344
BosGru_12	92.12	34,109	0.0027	0.2414
BosGru_13	90.18	41,557	0.0022	0.2344
BosGru_14	89.19	33,880	0.0026	0.2438
BosGru_15	85.77	44,937	0.0019	0.2326
BosGru_16	81.27	29,025	0.0028	0.2242
BosGru_17	77.38	26,254	0.0029	0.2311
BosGru_18	75.87	29,509	0.0026	0.2322
BosGru_19	72.32	28,474	0.0025	0.2400
BosGru_20	71.41	35,695	0.0020	0.2552
BosGru_21	68.96	22,366	0.0031	0.2263
BosGru_22	66.29	20,035	0.0033	0.2187
BosGru_23	60.85	18,160	0.0034	0.2161
BosGru_24	57.70	37,710	0.0015	0.2372
BosGru_25	57.40	18,795	0.0031	0.2270
BosGru_26	50.81	20,233	0.0025	0.2355
BosGru_27	49.70	24,849	0.0020	0.2398
BosGru_28	48.02	15,268	0.0031	0.2243
BosGru_29	46.15	30,382	0.0015	0.2423

**Table 2** The minor allele frequency (MAF) distribution patterns across different bins in three yak cohorts (Indian, Chinese and wild) based on whole-genome resequencing data

MAF Bins	MAF distribution (Number of markers and percentage in parenthesis)		
	Indian yak	Chinese yak	Wild yak
I: $x < 0.10$	260,050 (25.92%)	267,509 (26.67%)	223,381 (22.27%)
II: $0.10 < x < 0.20$	227,283 (22.66%)	242,885 (24.22%)	252,725 (25.20%)
III: $0.20 < x < 0.30$	146,093 (14.57%)	138,105 (13.77%)	171,262 (17.07%)
IV: $0.30 < x < 0.40$	161,158 (16.07%)	131,300 (13.09%)	152,606 (15.22%)
V: $x > 0.40$	208,386 (20.78%)	223,171 (22.25%)	202,996 (20.24%)

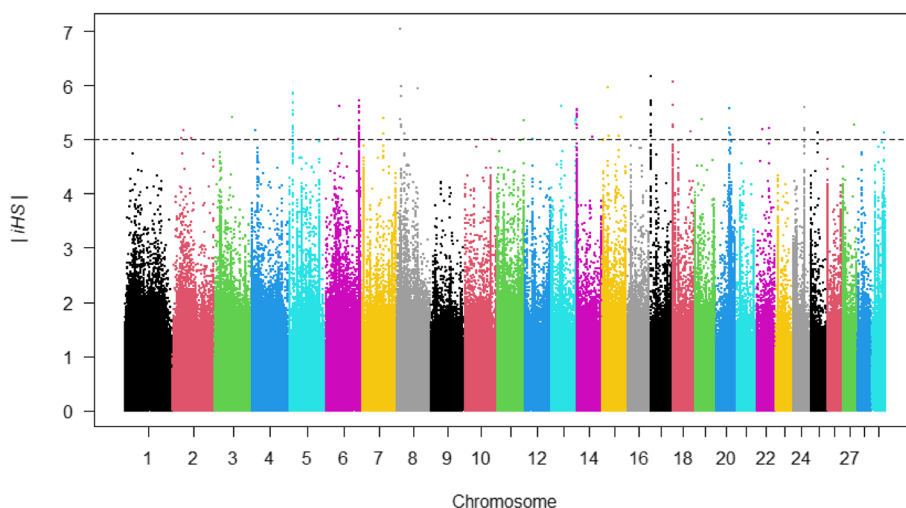
for yak, especially given their adaptation to harsh environments [45]. The *KLF1* gene i.e., Krüppel-like factor 1 (mapped on BosGru\_8) encodes erythroid-specific transcription factor that mediates erythropoiesis, the process of red blood cell formation. As a key regulator within the erythroid lineage, *KLF1* controls the expression of numerous genes essential for the development, differentiation, and maturation of erythrocytes [46]. Erythroid specializations are important in oxygen-deficient and hypobaric environments at high-altitude areas where yak are reared. The expression of *PRDX2* gene i.e., peroxiredoxin 2 (mapped on BosGru\_8) has been linked to oxidation–reduction process in animal cells [47]. These genes signify their role in various adaptation traits including protection against solar radiations, adipogenesis, erythropoiesis and oxidation–reduction processes.

On the other hand, other genes (*CACNG7*, *KHDRBS2*, *NANOS3*, *RFX1* and *RFX2*) found to be part of selection sweeps were linked with reproduction traits in yak. The *CACNG7* gene i.e., voltage-dependent calcium channel gamma-7 subunit (mapped on BosGru\_20) plays a role in the oxytocin signalling pathway specifically within yak physiology [48]. Among the other genes, *KHDRBS2* i.e., KH RNA binding domain containing, signal transduction associated 2 (mapped on BosGru\_24) has been linked with reproduction traits in goats and cattle, (Brahman and Colombian cattle) as reported by Strillacci et al. [49]. On the other hand, the genes like *NANOS3* i.e., nanos C2HC-type zinc finger 3 (mapped on BosGru\_8), *RFX1* (mapped on BosGru\_8), and *RFX2* (mapped on BosGru\_8) are regulatory factors that have been related to various aspects of spermatogenesis in different livestock species [50, 51]. Overall, the results help pinpoint the genes with known reproductive functions in yak or other related species that underwent selection sweeps and play a crucial role in yak adaptation. The genetic variations in these genes may have been beneficial and spread rapidly in the Indian yak population.

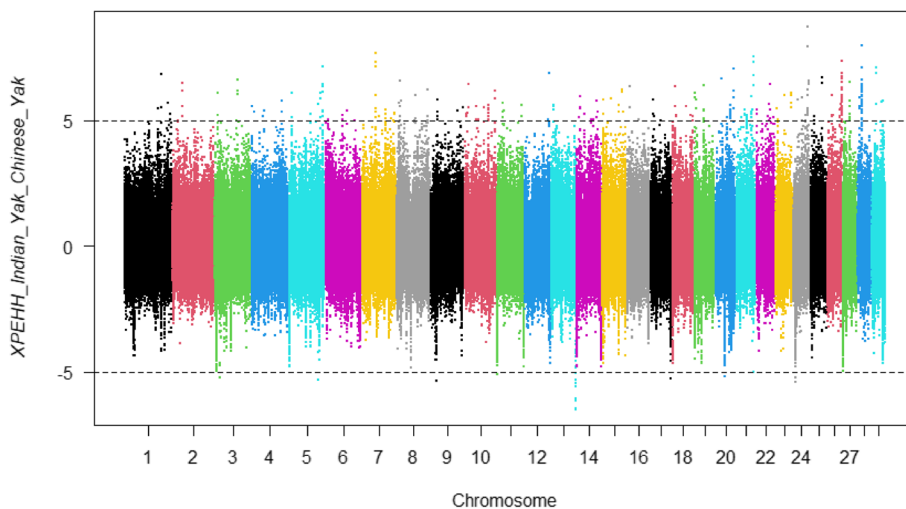
**Cross population selection signals**

In the present study, the Indian yak were compared with Chinese and wild groups to deduce the selection signals in respective comparisons. In all comparisons, the Indian yak genome was used as a query while that of Chinese and wild yak were used as a reference. A 2 mb window was used for predicting the functionality of SNPs found to be part of selective sweeps. The selective sweeps were therefore evident in gene-rich areas.

*XPEHH selection signals: Indian versus Chinese yak* In Indian versus Chinese yak comparison, multiple SNPs were found to be a part of selection signals in both the cohorts (Fig. 3) that were present in the vicinity of vari-



**Fig. 2** Manhattan plot showing evolutionary footprints within Indian yak in terms of distribution of *iHS* values



**Fig. 3** Manhattan plot showing evolutionary footprints on comparison of Indian yak with Chinese yak cohort

ous protein-coding genes with significant effects towards improving their adaptability. A total of 1447 genes were found to be part of selective sweeps in Indian yak cohort, while 241 genes were found to be in the vicinity of selective sweeps in Chinese yak. These genes included *NR2F2*, *OSBPL10*, *CIDEA*, *WFIKKN2*, *ADCY*, *THSD7A*, *ADGRB3*, *TRPC1*, *VASH2*, *ABHD5*, and others. The complete list delineating the genes in the vicinity of selective sweeps in Indian and Chinese yak cohorts have been presented in Supplementary Table S1. The *ABHD5* gene i.e., abhydrolase domain-containing protein 5, lysophosphatidic acid acyltransferase, mapped on BosGru\_22, has been reported to be associated with regulation of lipid metabolism in animals which is, in turn, essential for diverse cellular functions [52]. Similarly, the *NR2F2* gene i.e., nuclear

receptor subfamily 2, group F, member 2 (mapped on BosGru\_17) is a member of the steroid/thyroid hormone receptor family that has been related to higher fat content modulating adipogenesis in Angus beef cattle based on a transcriptomic analysis as compared to hybrid cattle [53]. Another gene i.e., *OSBPL10* or oxysterol binding protein like 10 (mapped on BosGru\_22) is involved in signalling and transport of lipids [54]. On the similar lines, Zhang et al. [55] reported the role of *WFIKKN2* gene i.e., WAP, follistatin/kazal, immunoglobulin, kunitz and netrin domain containing 2 (mapped on BosGru\_19) in adipocyte differentiation and muscle fat metabolism (via miR-22-3p signalling) in Yanbian cattle. Another gene, i.e., *ADCY* or adenylate cyclase (mapped on BosGru\_26) is a member of the cyclase gene family that has



been identified as a key factor in the adaptation of birds to elevated areas with vast elevated mountain ranges [56]. Its biological roles have been implicated in modulation of lipolysis, thermogenesis, and glucagon signalling in vertebrates [57] with expression across 27 different body tissues, including the lung, thyroid, and large-B cells [58]. The *CIDEC* gene i.e., cell death-inducing DFFA-like effector C, located on BosGru\_22, has been reported to play a crucial role in adipose tissue and liver development [59]. These findings indicate the relevance of these genes in regulation of energy (especially lipid) metabolism and thermoregulation in adaptation of animals to challenging hilly and mountainous environment and may be further evaluated using candidate gene approach for gaining deeper understanding on yak adaptation.

Another set of genes, found to be part of selection sweep under this comparison, was related to adaptation to hypoxic and climatic stressors with regulation of various metabolic pathways. These genes mainly included *ADGRB3*, *THSD7A*, *TRPC*, and *VASH2*. One of these putative candidate genes namely *ADGRB3* i.e., adhesion G protein-coupled receptor B3 (mapped on BosGru\_10) supports the new high-altitude adaptation theories, providing insights into the molecular mechanisms of altitude adaptation in different yak population [60]. The expression of the *THSD7A* gene i.e., thrombospondin type 1 domain-containing 7A (mapped on BosGru\_4) has been reported to aid pregnant Tibetan pigs overcome reproductive challenges caused by high-altitude hypoxia due to its role in embryonic angiogenesis [61]. On the other hand, The role of *TRPC4* gene i.e., transient receptor potential cation channel subfamily C member 4 (mapped on BosGru\_15) has been reported in body temperature regulation [62]. Furthermore, Yang et al. [63] reported the *VASH2* gene or vasohibin 2 to interact and regulate various aspects of heat stress and oxidative stress pathways in cattle-yak hybrids.

Similar to findings from *iHS* methodology, many genes, found to be part of selection footprints in this comparison were related to reproduction performance of yaks. These affected various aspects of spermatogenesis and mainly included *ALKBH5*, *GLI3*, *KIF*, *BCL2*, *LGR4*, *RAD52* and *YTHDC2*. The expression of *ALKBH5* gene i.e., alkB homolog 5, RNA demethylase (mapped on BosGru\_19) in spermatogonia, primary spermatocytes, and round spermatids of cattle and yak testis highlights its crucial role in reproductive performance of animals [64]. On the other hand, the functioning of *GLI3* i.e., GLI family zinc finger 3 (mapped on BosGru\_4) and *KIF22* i.e., kinesin family member 22 (BosGru\_26) genes has been implicated in various stages of spermatogenesis in yak [65, 66]. The *BCL2* gene i.e., B-cell lymphoma 2 (mapped on BosGru\_21) has role in yak placenta physiology during

gestation and calving periods, involved mainly in placental homeostasis, growth, and remodelling [67]. The *LGR4* gene or leucine-rich repeat-containing G protein-coupled receptor 4 has been reported to regulate the maturation of corpus luteum via modulation of the WNT-mediated EGFR-ERK signalling pathway [68]. Similarly, the expression of *YTHDC2* gene i.e., YTH N6-methyladenosine RNA binding protein C2 (mapped on BosGru\_11) has been implicated in different aspects of yak reproduction [69, 70]. On the other hand, the expression of *ANKK1* gene i.e., ankyrin-1 (mapped on BosGru\_28) has been linked to meat quality and tenderness, as evidenced by multiple reports [71, 72]. Ma et al. [73] reported the activity of the *ATPAF2* gene (mapped on BosGru\_19) to be associated with differential protein expression in brain mitochondria through proteomics study, providing insights into the plateau adaptability of yak.

Production of highly reactive oxygenated free radicals in cells is a tremendous challenge for animals reared under stressful climatic conditions [74]. The oxidation-redox potential needs to be maintained inside cells in yak; the oxygen-dependent expression of the cytochrome c oxidase subunit 4–2 gene, regulated by the transcription factor *CHCHD2*, plays a crucial role in their adaptation to high altitudes [75]. The expression of *COX6A2* gene i.e., cytochrome c oxidase subunit 6A2 (mapped on BosGru\_26) has earlier been related to oxidative phosphorylation in gayal and yak species, contributing to hypoxia adaptation in these animals [76]. Another gene i.e., *DKK2* or dickkopf WNT signaling pathway inhibitor 2 (mapped on BosGru\_6) has earlier been reported to be under positive selection across species indicating convergent molecular evolution, and related to adaptability in cattle, sheep and goat inhabited in various areas of the Tibetan plateau [77]. These genes gain significance in understanding the molecular basis of adaptation in yak and other similar species that are reared in hypoxic high-altitude conditions and should be evaluated further under diverse populations and environments.

The thick coat and special hair follicles contribute towards the adaptation of yak to high-altitude environments by providing insulation besides protection against ultraviolet (UV) radiations, wind and moisture. The development of hair follicles in yak is different from other low-lying livestock species including cattle. It is intricately associated with the adaptation of yak to high-altitude climatic adversities. The gene *FGF* i.e., Fibroblast growth factor (mapped on BosGru\_23), whose expression in hair follicles has been associated with the catagen and telogen phases of the hair growth cycle. Its expression has been related to keratin regulation and inhibition of the proliferation of stromal cells and strongly suppresses the growth of hair follicles, contributing towards

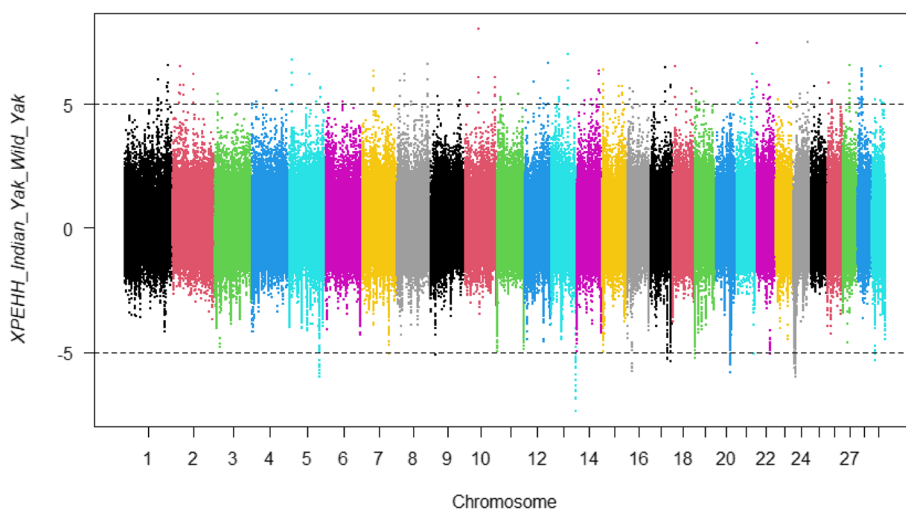
the improved adaptability of yaks to high altitudes [78]. The expression of another gene i.e., *MAP2K* or mitogen-activated protein kinase kinase (mapped on BosGru\_19) has also been related to hair growth in livestock species [79]. Guo et al. [70] reported the *ITGAD* gene i.e., integrin subunit alpha D (mapped on BosGru\_26) to be up-regulated during anestrus in yak.

On the other hand, the genes found to harbour selection signals in Chinese yak included 241 protein-coding genes. The gene *ACLY* is a potential candidate gene that involved in fatty acid synthesis in pigs [80]. Similar, the *FANCA* gene or Fanconi anemia complementation group A have been reported to play significant roles in cattle-yak male-sterility [81]. Another gene i.e., *HSPAIL* or Heat shock 70 kDa protein 1L is involved in protect cells from elevated body temperatures under high altitude conditions with direct exposure to UV radiations [82]. On the other hand, the expression of *KRTAP16-1*, *KRTAP17-1*, and *KRTAP3-1* genes has been linked to evolution of hair follicles in mammals [83]. Similarly, the *MAGEL2* gene i.e., MAGE family member L2 have been play significant roles in associated with hair follicle development in Tianzhu white yak [84]. This gene was part of selective sweep in both Chinese and wild yak cohorts when compared to Indian yak on genome-wide scale. Furthermore, the expression of *MC1R* gene has been associated with pigmentation of coat colour in North American yak population [85].

**XPEHH selection signals: Indian versus wild yak** Various genes were found to harbour or be present in vicinity of SNPs identified during selection sweep analysis on comparing the genomes of Indian yak with wild counterparts. These SNPs were present on multiple autosomes including BosGru\_1, BosGru\_2, BosGru\_5, BosGru\_8, BosGru\_10,

BosGru\_21, BosGru\_22 and BosGru\_28 among others (Fig. 4). The overlapping genes within assessed windows included *ADAMTS14*, *CELFI*, *LGR5*, *LIMCH1*, *MTCH2*, *NDUFS3* and *PIK3C2G*. The complete list delineating the genes in the vicinity of selective sweeps in Indian and wild yak cohorts have been presented in Supplementary Table S1. The expression of *ADAMTS14* gene i.e., ADAM metalloproteinase with thrombospondin type 1 motif 14 (mapped on BosGru\_22) has been reported to be involved in various aspects of follicular dynamics in granulosa and theca cells of cattle [86]. The expression of *CELFI* gene i.e., CUGBP Elav-like family member 1 (mapped on BosGru\_13) has been reported in muscle tissues of animals, playing a crucial role in regulating the muscle fiber length and diameter [87]. Another gene i.e., *LGR5* or leucine-rich repeat containing G protein-coupled receptor 5 (mapped on BosGru\_5) plays a role in macrophage-induced AKT/ $\beta$ -catenin-dependent stem cell activation and hair follicle regeneration through TNF pathways [88]. The *MTCH2* gene i.e., mitochondrial carrier 2 (mapped on BosGru\_13) plays a role in lipid homeostasis via mitochondrial carrier protein action and promotes adipogenesis in intramuscular preadipocytes through an m6A-YTHDF1-dependent mechanism [89], which is an adaptation to cold climatic conditions of high altitudes where yak are reared. The *NDUFS3* gene or NADH dehydrogenase [ubiquinone] iron-sulfur protein 3, mitochondrial (BosGru\_13) that has earlier been related to oxidative phosphorylation and hypoxia adaptation [90] was also found to harbour SNPs under selection pressure in Indian yak.

On the other hand, many genes were found to be part of selective sweeps in wild yak population when compared to Indian cohort on genome-wide scale. The *UCPI* gene or uncoupling protein 1, located on



**Fig. 4** Manhattan plot showing evolutionary footprints on comparison of Indian yak with wild yak cohort

BosGru\_16, is a potential candidate gene that involved in thermogenesis processes for adaptation in cold region in warm-blooded organisms [91]. On the other hand, the expression of *MGAT4D* gene or Mannosyl-glycoprotein N-acetylglucosaminyltransferase family member D (mapped on BosGru\_16) has been linked an intrinsic protector of testicular germ cells from mild heat stress in mammals [92]. The *SETD7* gene or SET domain containing 7, histone lysine methyltransferase, located on BosGru\_16, is involved in the related to muscle development in yak [93].

A total of 357 genes were common as part of selective sweeps in Indian yak across both the comparisons i.e., Indian yak versus Chinese as well as wild yak using XPEHH methodology. The important genes included *ANGEL2*, *CALNI*, *CTNNA3*, and *DNAJC15*, playing crucial roles in the adaptation of Indian yak to harsh local conditions. The *ANGEL2* gene has been reported to play a regulatory role in muscle development in bovines and mitochondrial functioning [88, 94]. Similarly, Hardie et al. [95] reported the *CALNI* gene, located on BosGru\_26, to play a crucial role in regulating dry matter intake in cattle. Similar to the results from the present study, Sun et al. [96] identified *CTNNA3* gene (mapped on BosGru\_27) to be under selection pressure in Youzhou dark goats, as determined through *iHS* and *ROH* (runs of homozygosity) tools using genome-wide SNP markers. It has been identified as a promising candidate for influencing growth traits in beef cattle and ovines [97, 98]. On the other hand, the *DNAJC15* gene encodes a heat shock protein that is linked to various economic and functional traits in cattle, playing a significant role in their reproductive health and efficiency [99].

## Conclusion

The present study is first to reveal the population structure and differentiation of global yak populations viz. Indian, Chinese and wild using whole genome resequencing data and latest bioinformatics tools. The computational analysis identified around 1 million high-quality genome-wide SNPs that were used for downstream analysis on population genomics. The present study has confirmed the presence of different lineages of world-wide yak populations. Further, the selection sweeps analysis revealed important genes spotting their role in adaptations and colonization of Indian yak under harsh environmental conditions. Notably, the genes harbouring the SNPs under selection pressure were significant for adaptation traits including lipidogenesis, energy metabolism, oxidation–reduction reactions, hypoxia and reproduction. A set of important genes that included *ANGEL2*, *CALNI*, *CTNNA3*, and *DNAJC15* were found common along different selection signature methodologies and same shall be targeted for better understanding

of yak adaptation to harsh climatic conditions in high-altitude habitats. These genes may be exploited as candidate genes in yak cohorts providing adaptability to harsh environment and could be utilised in molecular breeding and conservation genomics in the era of climate change.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11387-2>.

Supplementary Material 1.

## Acknowledgements

The authors acknowledge the facilities and support from ICAR-Indian Veterinary Research Institute, Izatnagar.

## Authors' contributions

SFA: Data curation, Formal analysis, Investigation, Methodology, Resources, Writing- Original draft, Writing- Review & Editing. MG: Data curation, Formal analysis, Investigation, Resources, Writing- Original draft. AtK: Conceptualization, Data curation, Writing- Review & Editing, Supervision. AdK and MSD: Data curation, Formal analysis, Writing- Review & Editing. GKJ, GKG and TD: Funding acquisition, Supervision, Project Administration.

## Funding

This work was supported by grants from the Centre for Bioinformatics (CABin) project of Indian Council of Agricultural Research, New Delhi, India.

## Data availability

The data used in the present study is available in online repositories with accession No. PRJNA803425 and PRJNA285834. The researchers who submitted the data sets (PRJNA803425 and PRJNA285834) to these online platforms retain all rights (intellectual property and other rights) and liabilities for the data.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

### Author details

<sup>1</sup>ICAR-Indian Veterinary Research Institute, Uttar Pradesh, Izatnagar, Bareilly 243 122, India. <sup>2</sup>ICAR-National Bureau of Animal Genetic Resources, Karnal, Haryana 132 001, India. <sup>3</sup>ICAR-Indian Agricultural Statistics Research Institute, Pusa, New Delhi 110 012, India. <sup>4</sup>Assistant Director General (Animal Production and Breeding), Indian Council of Agricultural Research, Krishi Bhawan, New Delhi 110 001, India.

Received: 9 August 2024 Accepted: 19 February 2025

Published online: 03 March 2025

## References

- Larson G, Piperno DR, Allaby RG, Purugganan MD, Andersson L, Arroyo-Kalin M, et al. Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences of the United States of America*. 2014;111(17): 6139–46.
- Diamond J. Evolution, consequences and future of plant and animal domestication. *Nature*. 2002;418:700–7.

3. Don E, Wilson & DeeAnn M. Reeder (editors). 2005. *Mammal Species of the World. A Taxonomic and Geographic Reference* (3rd ed), Johns Hopkins University Press, 2,142.
4. Liu X, Liu W, Lenstra JA, Zheng Z, Wu X, Yang J, et al. Evolutionary origin of genomic structural variations in domestic yaks. *Nat Commun*. 2023;14(1):5617. <https://doi.org/10.1038/s41467-023-41220-x>.
5. Guan J, Long K, Ma J, Zhang J, He D, Jin L, et al. Comparative analysis of the microRNA transcriptome between yak and cattle provides insight into high-altitude adaptation. *PeerJ*. 2017;5:e3959. <https://doi.org/10.7717/peerj.3959>.
6. Yang J, Jin ZB, Chen J, Huang XF, Li XM, Liang YB, et al. Genetic signatures of high-altitude adaptation in Tibetans. *Proc Natl Acad Sci U S A*. 2017;114(16):4189–94. <https://doi.org/10.1073/pnas.1617042114>.
7. Edea Z, Dadi H, Dessie T, Kim KS. Genomic signatures of high-altitude adaptation in Ethiopian sheep populations. *Genes Genomics*. 2019;41(8):973–81. <https://doi.org/10.1007/s13258-019-00820-y>.
8. Qiu Q, Zhang G, Ma T, Qian W, Ye Z, Cao C, et al. The yak genome and adaptation to life at high altitude. *Nat Genet*. 2012;44(8):946–9. <https://doi.org/10.1038/ng.2343>.
9. Buzzard P & BJ. *Bos mutus*. IUCN Red List of Threatened Species. 2016.
10. Das PJ, Kour A, Deori S, Begum SS, Pukhrabam M, Maiti S, et al. Characterization of Arunachali Yak: A Roadmap for Pastoral Sustainability of Yaks in India. *Sustainability* (Switzerland). 2022;14(19):12655. <https://doi.org/10.3390/su141912655>.
11. Xia W, Luo T, Zhang W, Mason AS, Huang D, Huang X, et al. Development of high-density snp markers and their application in evaluating genetic diversity and population structure in elaeis guineensis. *Front Plant Sci*. 2019;10:130. <https://doi.org/10.3389/fpls.2019.00130>.
12. Sivalingam J, Vineeth MR, Surya T, Singh K, Dixit SP, Niranjan SK, et al. Genomic divergence reveals unique populations among Indian Yaks. *Sci Rep*. 2020;10(1):3636. <https://doi.org/10.1038/s41598-020-59887-3>.
13. Kour A, Niranjan SK, Malayaperumal M, Surati U, Pukhrabam M, Sivalingam J, et al. Genomic Diversity Profiling and Breed-Specific Evolutionary Signatures of Selection in Arunachali Yak. *Genes (Basel)*. 2022;13(2):254. <https://doi.org/10.3390/genes13020254>.
14. Ahmad SF, Chandrababu Shailaja C, Vaishnav S, Kumar A, Gaur GK, Janga SC, et al. Read-depth based approach on whole genome resequencing data reveals important insights into the copy number variation (CNV) map of major global buffalo breeds. *BMC Genomics*. 2023;616:24. <https://doi.org/10.1186/s12864-023-09720-8>.
15. Ng PC, Kirkness EF. Whole Genome Sequencing. 2010. p. 215–26.
16. Brutsaert TD. Population genetic aspects and phenotypic plasticity of ventilatory responses in high altitude natives. *Respir Physiol Neurobiol*. 2007;158(2-3):151–60. <https://doi.org/10.1016/j.resp.2007.03.004>.
17. Gangwar M, Ahmad SF, Ali AB, Kumar A, Gaur GK, et al. Identifying low-density, ancestry-informative SNP markers through whole genome resequencing in Indian, Chinese, and wild yak. *BMC Genomics*. 2024;25:1043.
18. Qiu Q, Wang L, Wang K, Yang Y, Ma T, Wang Z, et al. Yak whole-genome resequencing reveals domestication signatures and prehistoric population expansions. *Nat Commun*. 2015;6:10283. <https://doi.org/10.1038/ncomms10283>.
19. Chai Z xin, Xin J wei, Zhang C fu, Dawayangla, Luosang, Zhang Q, et al. Whole-genome resequencing provides insights into the evolution and divergence of the native domestic yaks of the Qinghai–Tibet Plateau. *BMC Evol Biol*. 2020;20:137. <https://doi.org/10.1186/s12862-020-01702-8>.
20. Kumar A, Dige M, Niranjan SK, Ahlawat S, Arora R, Kour A, et al. Whole genome resequencing revealed genomic variants and functional pathways related to adaptation in Indian yak populations. *Anim Biotechnol*. 2024;35(1):2282723. <https://doi.org/10.1080/10495398.2023.2282723>.
21. Modi A, Vai S, Caramelli D, Lari M. The Illumina Sequencing Protocol and the NovaSeq 6000 System. In: *Methods in Molecular Biology*. 2021;224:15–42. [https://doi.org/10.1007/978-1-0716-1099-2\\_2](https://doi.org/10.1007/978-1-0716-1099-2_2).
22. Andrews S. FastQC. A Quality Control tool for High Throughput Sequence Data. *soil*. 2010.
23. Krueger F. *Babraham Bioinformatics - Trim Galore!* Version 0.4.4. 2017.
24. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
25. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10(2):giab008. <https://doi.org/10.1093/gigascience/giab008>.
26. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–303. <https://doi.org/10.1101/gr.107524.110>.
27. 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. *The American Journal of Human Genetics*. 2007;81:559–75. <https://doi.org/10.1086/519795>.
28. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A Tool for Genome-wide Complex Trait Analysis. *The American Journal of Human Genetics*. 2011;88:76–82.
29. Wilkinson L. ggplot2: Elegant Graphics for Data Analysis by WICKHAM, H. *Biometrics*. 2011;67.
30. Pickrell JK, Pritchard JK. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genet*. 2012;8: e1002967.
31. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19:1655–64.
32. Kumar S, Nei M, Dudley J, Tamura K. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform*. 2008;9(4):299–306. <https://doi.org/10.1093/bib/bbn017>.
33. Voight BF, Kudravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol*. 2006;4(4):e154.
34. Browning BL, Tian X, Zhou Y, Browning SR. Fast two-stage phasing of large-scale sequence data. *Am J Hum Genet*. 2021;108(10):1880–90.
35. Gautier M, Vitalis R, Reh An R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics*. 2012; 2012;28(8):1176–7. <https://doi.org/10.1093/bioinformatics/bts115>.
36. Gautier M, Klassmann A, Vitalis R. rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Mol Ecol Resour*. 2017;17:78–90.
37. Ma Y, Ding X, Qanbari S, Weigend S, Zhang Q, Simianer H. Properties of different selection signature statistics and a new strategy for combining them. *Heredity (Edinb)*. 2015;115(5):426–436. <https://doi.org/10.1038/hdy.2015.42>.
38. Mahar K, Gurao A, Kumar A, Pratap Singh L, Chitkara M, Gowane GR, et al. Genomic inbreeding analysis reveals resilience and genetic diversity in Indian yak populations. *Gene*. 2024;928: 148787.
39. Ayalew W, Chu M, Liang C, Wu X, Ping Y. Adaptation mechanisms of yak (*Bos grunniens*) to high-altitude environmental stress. *Animals*. 2021;11(8):2344. <https://doi.org/10.3390/ani11082344>.
40. Ahmad HI, Mahmood S, Hassan M, Sajid M, Ahmed I, Shokrollahi B, et al. Genomic insights into Yak (*Bos grunniens*) adaptations for nutrient assimilation in high-altitudes. *Sci Rep*. 2024;14(1):5650.
41. Akiyama M. The roles of ABCA12 in keratinocyte differentiation and lipid barrier formation in the epidermis. *Dermato-Endocrinology*. 2011;3(2):107–112. <https://doi.org/10.4161/derm.3.2.15136>.
42. Chebii VJ, Oyola SO, Kotze A, Entfellner JBD, Musembi Mutuku J, Agaba M. Genome-wide analysis of nubian ibex reveals candidate positively selected genes that contribute to its adaptation to the desert environment. *Animals*. 2020;10(11):2181. <https://doi.org/10.3390/ani10112181>.
43. Walsh TG, Li Y, Williams CM, Aitken EW, Andrews RK, Poole AW. Loss of the exocyst complex component EXOC3 promotes hemostasis and accelerates arterial thrombosis. *Blood advances*. 2021;5(3):674–86.
44. Zhang X, Ding X, Wang C, Le Q, Wu D, Song A, et al. Depletion of JunB increases adipocyte thermogenic capacity and ameliorates diet-induced insulin resistance. *Nat Metab*. 2024;6(1):78–93. <https://doi.org/10.1038/s42255-023-00945-1>.
45. Zhang Z, Zhang Y, Bao Q, Gu Y, Liang C, Chu M, et al. The Landscape of Accessible Chromatin during Yak Adipocyte Differentiation. *Int J Mol Sci*. 2022;23:9960.
46. Caria CA, Faà V, Ristaldi MS. Krüppel-Like Factor 1: A Pivotal Gene Regulator in Erythropoiesis. *Cells*. 2022;11:3069.
47. Pei J, Zhao S, Yin M, Wu F, Li J, Zhang G, et al. Differential proteomics of placentas reveals metabolic disturbance and oxidative damage participate yak spontaneous miscarriage during late pregnancy. *BMC Vet Res*. 2022;18(1):248. <https://doi.org/10.1186/s12917-022-03354-w>.



48. Wang T, Ma X, Zheng Q, Ma C, Zhang Z, Pan H, et al. A comprehensive study on the longissimus dorsi muscle of Ashdan yaks under different feeding regimes based on transcriptomic and metabolomic analyses. *Anim Biotechnol.* 2024;35(1):2294785. <https://doi.org/10.1080/10495398.2023.2294785>.
49. Strillacci MG, Vevey M, Blanchet V, Mantovani R, Sartori C, Bagnato A. The genomic variation in the aosta cattle breeds raised in an extensive alpine farming system. *Animals.* 2020;10(12):2385. <https://doi.org/10.3390/ani10122385>.
50. Horvath GC, Kistler WS, Kistler MK. RFX2 is a potential transcriptional regulatory factor for histone H1t and other genes expressed during the meiotic phase of spermatogenesis. *Biol Reprod.* 2004;71(5):1551–9. <https://doi.org/10.1095/biolreprod.104.032268>.
51. Cai X, Yu S, Mipam TD, Yang F, Zhao W, Liu W, et al. Comparative analysis of testis transcriptomes associated with male infertility in cattleyak. *Theriogenology.* 2017; 88:28–42. <https://doi.org/10.1016/j.theriogenology.2016.09.047>.
52. Schratte M, Lass A, Radner FPW. ABHD5—A Regulator of Lipid Metabolism Essential for Diverse Cellular Functions. *Metabolites.* 2022;12(11):1015. <https://doi.org/10.3390/metabo12111015>.
53. Song C, Huang Y, Yang Z, Ma Y, Chaogetu B, Zhuoma Z, et al. RNA-Seq Analysis Identifies Differentially Expressed Genes in Subcutaneous Adipose Tissue in Qaidaford Cattle, Cattle-Yak, and Angus Cattle. *Animals.* 2019;9(12):1077. <https://doi.org/10.3390/ani9121077>.
54. Sierra B, Triska P, Soares P, Garcia G, Perez AB, Aguirre E, et al. OSBPL10, RXRA and lipid metabolism confer African-ancestry protection against dengue haemorrhagic fever in admixed Cubans. *PLoS Pathog.* 2017;13:e1006220.
55. Wang X, Yang C, Guo F, Zhang Y, Ju Z, Jiang Q, et al. Integrated analysis of mRNAs and long noncoding RNAs in the semen from Holstein bulls with high and low sperm motility. *Sci Rep.* 2019;9(1):2092. <https://doi.org/10.1038/s41598-018-38462-x>.
56. Wang L, Liu F, Zhao L, Xu Y, Zhang T, Wen L. A test of genetic divergence of a bird existing in the Sichuan Basin and its surrounding mountain ranges. *Avian Res.* 2023;14:100144.
57. Sazzini M, Abondio P, Sarno S, Gneccchi-Ruscione GA, Ragno M, Giuliani C, et al. Genomic history of the Italian population recapitulates key evolutionary dynamics of both Continental and Southern Europeans. *BMC Biol.* 2020;18(1):51. <https://doi.org/10.1186/s12915-020-00778-4>.
58. Tang Y, Wang T, Zhang A, Zhu J, Zhou T, Zhou YL, et al. ADCY9 functions as a novel cancer suppressor gene in lung adenocarcinoma. *J Thorac Dis.* 2023;15(3):1018–35. <https://doi.org/10.21037/jtd-22-1027>.
59. Matsusue K. A physiological role for fat specific protein 27/cell death-inducing DFF45-like effector C in adipose and liver. *Biol Pharm Bull.* 2010;33(3):346–50. <https://doi.org/10.1248/bpb.33.346>.
60. Guang-Xin E, Yang BG, Zhu Y Bin, Duang XH, Basang WD, Luo XL, et al. Genome-wide selective sweep analysis of the high-altitude adaptability of yaks by using the copy number variant. *3 Biotech.* 2020;10(6):259. <https://doi.org/10.1007/s13205-020-02254-w>.
61. Ma YF, Han XM, Huang CP, Zhong L, Adeola AC, Irwin DM, et al. Population Genomics Analysis Revealed Origin and High-altitude Adaptation of Tibetan Pigs. *Sci Rep.* 2019;9(1):11463. <https://doi.org/10.1038/s41598-019-47711-6>.
62. Colombi D, Perini F, Bettini S, Mastrangelo S, Abeni F, Conte G, et al. Genomic responses to climatic challenges in beef cattle: A review. *Anim Genet.* 2024;55:854–70.
63. Yang Q, Xie Y, Pan B, Cheng Y, Zhu Y, Fei X, et al. The Expression and Epigenetic Characteristics of the HSF2 Gene in Cattle-Yak and the Correlation with Its Male Sterility. *Animals.* 2024;14:1410.
64. Chen H, Zhang J, Yan Y, Zhu C, Wang L, Fu S, et al. N6-methyladenosine RNA demethylase ALKBH5 is testis-specifically downregulated in hybrid male sterile dzo and is a target gene of bta-miR-200a. *Theriogenology.* 2022;187:51–7.
65. Wu S, Mipam T, Xu C, Zhao W, Shah MA, Yi C, et al. Testis transcriptome profiling identified genes involved in spermatogenic arrest of cattleyak. *PLoS ONE.* 2020;15:e0229503.
66. Cao M, Pei J, Xiong L, Guo S, Wang X, Kang Y, et al. Analysis of Chromatin Openness in Testicle Tissue of Yak and Cattle-Yak. *Int J Mol Sci.* 2022;23:15810.
67. Fan J, Yu S, Cui Y, Xu G, Wang L, Pan Y, et al. Bcl-2/Bax protein and mRNA expression in yak (*Bos grunniens*) placentomes. *Theriogenology.* 2017;104:23–9.
68. Pan H, Cui H, Liu S, Qian Y, Wu H, Li L, et al. Lgr4 Gene Regulates Corpus Luteum Maturation Through Modulation of the WNT-Mediated EGFR-ERK Signaling Pathway. *Endocrinology.* 2014;155:3624–37.
69. Terefe E, Belay G, Han J, Hanotte O, Tijjani A. Genomic adaptation of Ethiopian indigenous cattle to high altitude. *Front Genet.* 2022;13:960234.
70. Guo S, Cao M, Wang X, Xiong L, Wu X, Bao P, et al. Changes in transcriptomic profiles in different reproductive periods in yaks. *Biology (Basel).* 2021;10(12):1229. <https://doi.org/10.3390/biology10121229>.
71. Aslan O, Sweeney T, Mullen AM, Hamill RM. Regulatory polymorphisms in the bovine Ankyrin 1 gene promoter are associated with tenderness and intramuscular fat content. *BMC Genet.* 2010;11:111. <https://doi.org/10.1186/1471-2156-11-111>.
72. Horodyska J, Sweeney T, Ryan M, Hamill RM. Novel SNPs in the Ankyrin 1 gene and their association with beef quality traits. *Meat Sci.* 2015;108:88–96.
73. Ma X, Zhang Q, La Y, Fu D, Jiang H, Bao P, et al. Differential Abundance of Brain Mitochondrial Proteins in Yak and Cattle: A Proteomics-Based Study. *Front Vet Sci.* 2021;8:663031. <https://doi.org/10.3389/fvets.2021.663031>.
74. Yang S, Liu J, Gu Z, Liu P, Lan Q. Physiological and Metabolic Adaptation to Heat Stress at Different Altitudes in Yaks. *Metabolites.* 2022;12(11):1082. <https://doi.org/10.3390/metabo12111082>.
75. Aras S, Pak O, Sommer N, Finley R, Hüttemann M, Weissmann N, et al. Oxygen-dependent expression of cytochrome c oxidase subunit 4–2 gene expression is mediated by transcription factors BBP1, CXXC5 and CHCHD2. *Nucleic Acids Res.* 2013;41:2255–66.
76. Ma J, Zhang T, Wang W, Chen Y, Cai W, Zhu B, et al. Comparative Transcriptome Analysis of Gayal (*Bos frontalis*), Yak (*Bos grunniens*), and Cattle (*Bos taurus*) Reveal the High-Altitude Adaptation. *Front Genet.* 2022;12:778788.
77. Chebii VJ, Mpolya EA, Muchadeyi FC, Entfellner JBD. Genomics of adaptations in ungulates. *Animals.* 2021;11(6):1617.
78. Kimura-Ueki M, Oda Y, Oki J, Komi-Kuramochi A, Honda E, Asada M, et al. Hair cycle resting phase is regulated by cyclic epithelial FGF18 signaling. *J Invest Dermatol.* 2012;132(5):1338–45. <https://doi.org/10.1038/jid.2011.490>.
79. 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(1):2815. <https://doi.org/10.1038/s41467-020-16485-1>.
80. Davoli R, Braglia S, Zappaterra M, Redeghieri C, Comella M, Zambonelli P. Association and expression analysis of porcine ACLY gene related to growth and carcass quality traits in Italian Large White and Italian Duroc breeds. *Livest Sci.* 2014;165:1–7.
81. Zhao S, Sun W, Chen SY, Li Y, Wang J, Lai S, et al. The exploration of miRNAs and mRNA profiles revealed the molecular mechanisms of cattle-yak male infertility. *Front Vet Sci.* 2022;9:974703. <https://doi.org/10.3389/fvets.2022.974703>.
82. Zhou X, Bao P, Zhang X, Guo X, Liang C, Chu M, et al. Genome-wide detection of RNA editing events during the hair follicles cycle of Tianzhu white yak. *BMC Genomics.* 2022;23:737.
83. Litman T, Stein WD. Ancient lineages of the keratin-associated protein (KRTAP) genes and their co-option in the evolution of the hair follicle. *BMC Ecol Evol.* 2023;23:7.
84. Bao Q, Ma X, Jia C, Wu X, Wu Y, Meng G, et al. Resequencing and Signatures of Selective Scans Point to Candidate Genetic Variants for Hair Length Traits in Long-Haired and Normal-Haired Tianzhu White Yak. *Front Genet.* 2022;13:798076. <https://doi.org/10.3389/fgene.2022.798076>.
85. Petersen JL, Kalbfleisch TS, Parris M, Tietze SM, Cruickshank J. MC1R and KIT Haplotypes Associate With Pigmentation Phenotypes of North American Yak (*Bos grunniens*). *J Hered.* 2020;111:182–93.
86. Hernández-Delgado P, Felix-Portillo M, Martínez-Quintana JA. ADAMTS Proteases: Importance in Animal Reproduction. *Genes (Basel).* 2023;14:1181.
87. Berger DS, Ladd AN. Repression of nuclear CELF activity can rescue CELF-regulated alternative splicing defects in skeletal muscle models of myotonic dystrophy. *PLoS Curr.* 2012. RNN105. <https://doi.org/10.1371/currents.RRN1305>.

88. Wang X, Chen H, Tian R, Zhang Y, Drutskaya MS, Wang C, et al. Macrophages induce AKT/ $\beta$ -catenin-dependent Lgr5+ stem cell activation and hair follicle regeneration through TNF. *Nat Commun.* 2017;8:14091.
89. Jiang Q, Sun B, Liu Q, Cai M, Wu R, Wang F, Yao Y, Wang Y, Wang X. MTH2 promotes adipogenesis in intramuscular preadipocytes via an m6A-YTHDF1-dependent mechanism. *FASEB J.* 2018;33(2):2971.
90. Ma J, Zhang T, Wang W, Chen Y, Cai W, Zhu B, et al. Comparative Transcriptome Analyses of Gayal (*Bos frontalis*), Yak (*Bos grunniens*), and Cattle (*Bos taurus*) Reveal the High-Altitude Adaptation. *Front Genet.* 2022;12:778788. <https://doi.org/10.3389/fgene.2021.778788>.
91. Nikanorova AA, Barashkov NA, Pshennikova VG, Gotovtsev NN, Romanov GP, Solovyev AV, et al. Relationships between Uncoupling Protein Genes UCP1, UCP2 and UCP3 and Irisin Levels in Residents of the Coldest Region of Siberia. *Genes (Basel).* 2022;13:1612.
92. Akintayo A, Liang M, Bartholdy B, Batista F, Aguilar J, Prendergast J, et al. The Golgi Glycoprotein MGAT4D is an Intrinsic Protector of Testicular Germ Cells From Mild Heat Stress. *Sci Rep.* 2020;10:2135. <https://doi.org/10.1038/s41598-020-58923-6>.
93. Ma X, Jia C, Chu M, Fu D, Lei Q, Ding X, et al. Transcriptome and DNA Methylation Analyses of the Molecular Mechanisms Underlying with Longissimus dorsi Muscles at Different Stages of Development in the Polled Yak. *Genes (Basel).* 2019;10(12):970. <https://doi.org/10.3390/genes10120970>.
94. Clemente P, Calvo-Garrido J, Pearce SF, Schober FA, Shigematsu M, Siira SJ, et al. ANGEL2 phosphatase activity is required for non-canonical mitochondrial RNA processing. *Nat Commun.* 2022;13(1):5750.
95. Hardie LC, VandeHaar MJ, Tempelman RJ, Weigel KA, Armentano LE, Wiggans GR, et al. The genetic and biological basis of feed efficiency in mid-lactation Holstein dairy cows. *J Dairy Sci.* 2017;100(11):9061–75. <https://doi.org/10.3168/jds.2017-12604>.
96. Sun X, Niu Q, Jiang J, Wang G, Zhou P, Li J, Chen C, Liu L, Xu L, Ren H. Identifying candidate genes for litter size and three morphological traits in Youzhou dark goats based on genome-wide SNP markers. *Genes.* 2023;14(6):1183.
97. Yu H, Yu S, Guo J, Cheng G, Mei C, Zan L. Genome-Wide Association Study Reveals Novel Loci Associated with Body Conformation Traits in Qinchuan Cattle. *Animals.* 2023;13(23):3628. <https://doi.org/10.3390/ani13233628>.
98. Zhao L, Li F, Yuan L, Zhang X, Zhang D, Li X, et al. Expression of ovine CTNNA3 and CAP2 genes and their association with growth traits. *Gene.* 2022;807:145949.
99. Zhang B, Peñagaricano F, Driver A, Chen H, Khatib H. Differential expression of heat shock protein genes and their splice variants in bovine preimplantation embryos. *J Dairy Sci.* 2011;94(8):4174–82. <https://doi.org/10.3168/jds.2010-4137>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.