

Article

Genomic Diversity Profiling and Breed-Specific Evolutionary Signatures of Selection in Arunachali Yak

Aneet Kour ^{1,*}, Saket Kumar Niranjana ², Mohan Malayaperumal ³, Utsav Surati ³, Martina Pukhrambam ¹, Jayakumar Sivalingam ⁴, Amod Kumar ² and Mihir Sarkar ¹ 

¹ ICAR-National Research Centre on Yak, Dirang 790101, Arunachal Pradesh, India; martina.pukhrambam@icar.gov.in (M.P.); mihir.sarkar@icar.gov.in (M.S.)

² ICAR-National Bureau of Animal Genetic Resources, Karnal 132001, Haryana, India; Saket.Niranjana@icar.gov.in (S.K.N.); amod.kumar@icar.gov.in (A.K.)

³ ICAR-National Dairy Research Institute, Karnal 132001, Haryana, India; mohanmalayaperumal@gmail.com (M.M.); utsav.surati@yahoo.com (U.S.)

⁴ ICAR-Directorate of Poultry Research, Hyderabad 500030, Telangana, India; jeyvet@gmail.com

* Correspondence: aneet.kour@icar.gov.in

Abstract: Arunachali yak, the only registered yak breed of India, is crucial for the economic sustainability of pastoralist Monpa community. This study intended to determine the genomic diversity and to identify signatures of selection in the breed. Previously available double digest restriction-site associated DNA (ddRAD) sequencing data of Arunachali yak animals was processed and 99,919 SNPs were considered for further analysis. The genomic diversity profiled based on nucleotide diversity, π ($\pi = 0.041$ in 200 bp windows), effective population size, N_e ($N_e = 83$) and Runs of homozygosity (ROH) (predominance of shorter length ROHs) was found to be optimum. Subsequently, 207 regions were identified to be under selective sweeps through de-correlated composite of multiple signals (DCMS) statistic which combined three individual test statistics viz. π , Tajima's D and $|iHS|$ in non-overlapping 100 kb windows. Mapping of these regions revealed 611 protein-coding genes including *KIT*, *KITLG*, *CDH12*, *FGG*, *FGA*, *FGB*, *PDGFRA*, *PEAR1*, *STXBP3*, olfactory receptor genes (*OR5K3*, *OR5H6* and *OR1E1*) and taste receptor genes (*TAS2R1*, *TAS2R3* and *TAS2R4*). Functional annotation highlighted that biological processes like platelet aggregation and sensory perception were the most overrepresented and the associated regions could be considered as breed-specific signatures of selection in Arunachali yak. These findings point towards evolutionary role of natural selection in environmental adaptation of Arunachali yak population and provide useful insights for pursuing genome-wide association studies in future.

Keywords: genomic diversity; selection signatures; double digest restriction-site associated DNA; Arunachali yak



Citation: Kour, A.; Niranjana, S.K.; Malayaperumal, M.; Surati, U.; Pukhrambam, M.; Sivalingam, J.; Kumar, A.; Sarkar, M. Genomic Diversity Profiling and Breed-Specific Evolutionary Signatures of Selection in Arunachali Yak. *Genes* **2022**, *13*, 254. <https://doi.org/10.3390/genes13020254>

Academic Editor: Anna V. Kukekova

Received: 7 December 2021

Accepted: 13 January 2022

Published: 28 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Mammalian body has evolved certain anatomical (enlarged heart and lungs, shorter tongue), physiological (increased oxygen availability in blood, lower metabolic rate), haematological (modifications in haemoglobin structure to improve oxygen affinity, increased average platelet volume and plasma fibrinogen concentrations) and morphological (small body size, distinct coat features like colour, texture and length) adaptations in order to cope up with the stressors of long-term exposure to cold and hypoxic conditions of high altitudes [1–8]. Most of these adaptations are genetically fixed in highlanders, thus conferring them protection against adverse climatic conditions [9]. Genomic studies have shown that the most common genes responsible for adaptation in highlander species include those involved in hypoxia-inducible factor (HIF) pathway like *EPAS1*, *EGLN1*, *EGLN2*, *VEGF*, *EPO*, *NOS* and *EDN* [10,11]. Further, different populations of the same species may

possess varying adaptations in their genomes in response to difference in altitude and vegetation [12,13].

Yak (*Bos grunniens*) is a unique species which has been exposed to intensive natural selection during the process of environmental adaptation in order to sustain the harsh environmental extremes of high altitude regions [14,15]. Artificial selection by man and breeding within small geographical territories may have further consolidated certain evolutionary signatures within its genome [16]. Elucidation of these regions subjected to positive selection in yak genome can help in maintaining genetic diversity and unravelling causal variants and genes subjected to selection.

There are around 58,000 yaks in India, around half of which are concentrated in the easternmost state, Arunachal Pradesh. [17]. Arunachali yak (Figure 1), the only registered native yak breed of India (till date) has been traditionally reared for food, fibre, transport and manure by the Monpa pastoralist community of Tawang and West Kameng districts of Arunachal Pradesh. These are medium-sized animals with compact body producing an average milk yield of 1.1–1.6 kg/day and the mean adult body weight in males and females is around 365 kg and 230 kg, respectively [18]. The breed resides in cold and humid climate at an altitude of 2700–4300 m above mean sea level and can sustain a temperature of $-40\text{ }^{\circ}\text{C}$ to $10\text{ }^{\circ}\text{C}$ [19]. This is in stark contrast to the yaks found in cold and arid regions of Ladakh (India) and Tibetan Autonomous Region (TAR) of China which dwell at an altitude of 4000–5000 m and 4500–5500 m respectively [20]. Hence, the possibility of distinct selection signatures in the yaks of Arunachal Pradesh cannot be ruled out.



Figure 1. Arunachali yak. (a) Arunachali yak male (b) Arunachali yak female.

A breed-specific, genome-wide scan for identifying signatures of selection can provide some interesting insights into the evolutionary history of the population and can help to

highlight new targets for selection and genetic improvement of the breed. Although the genome-wide SNPs (Single Nucleotide Polymorphisms) present in Arunachali yak breed have been primarily evaluated, a holistic study to explore the positively selected regions in the genome of lone registered yak breed of India is still pending. The present investigation was directed at profiling the genomic diversity and identifying the putative regions under selection in Arunachali yak population based on a genome-wide scan.

2. Methods

Double digest restriction-site associated DNA (ddRAD) sequencing technique employs a combination of both rare and frequent cutters or restriction enzymes to generate raw reads after precise size selection [21]. The ddRAD sequencing data for twenty animals of Arunachali yak available in NCBI (BioProject PRJNA577203) was used in the study. The complete analysis was performed on a computer with the following configuration: 12 GB RAM, 256 GB SSD and Microsoft i4 processor.

2.1. Quality Control of Sequencing Reads

After demultiplexing to obtain reads for each sample, quality control of the paired-end reads was performed and low quality sequences, sequences with contamination or artifacts were removed to avoid erroneous conclusions. Paired-end FASTQ files with raw reads were analysed in PRINSEQ lite v0.20.4 [22] to trim the adaptor sequences at the ends. The sequences with less than 30 phred quality score at any of the windows were discarded using Stacks [23].

2.2. Sequence Alignment and Variant Calling

The processed reads were further aligned to *Bos taurus* reference genome assembly ARS-UCD1.2 [24,25] using Bowtie 2 [26] due to lack of chromosome-wise assembly for yak genome (scaffold-based assembly is available at present). The resulting aligned reads in sequence alignment mode (SAM) format were converted to binary alignment mode (BAM) format (using 'samtools view' flag) in SAMtools [27]. Finally, the reads from all the samples were merged into a single binary call format (BCF) file using the flag 'samtoolsmpileup'.

VCftools [28] was employed to convert single BCF file into a VCF format which is the ideal format to be used for further analysis. Furthermore, it was used to filter out SNPs at different read depths (RD) viz. RD 2, RD 5 and RD 10 and those at RD 10 were retained for further analysis. SNPs at RD 10 were subjected to quality control (QC) and the ones with minor allele frequency (MAF) lesser than 0.05, call rate (CR) lesser than 90% and p -value for Hardy-Weinberg Equilibrium (H-WE) lesser than 0.0001 were removed from the data so as to avoid false positive results.

2.3. SNP Annotation and Identification of Deleterious Mutations

Structural and functional annotation of the high quality SNPs was carried out using SnpEff [29]. Further, we identified the deleterious mutations in the population based on SIFT (Sorting Intolerant from Tolerant) scores using VEP [30]. The mutations with SIFT score ≤ 0.05 was adjudged as deleterious to protein function and subsequently, were mapped to their location in the protein-coding genes. The obtained gene list was analysed in PANTHER [31] for statistical overrepresentation of gene ontology (GO) terms.

2.4. Genomic Diversity Estimation

Three indicators of genomic diversity-nucleotide diversity (π), effective population size (N_e) and runs of homozygosity (ROH) were considered to assess genetic variability in the population.

Π as a measure of genetic divergence within population was estimated in 100 kb windows of the genome using "window-pi" flag in VCftools [28] and an output file with the suffix "Windowed.pi" was generated. In order to calculate the average π for the whole population, 200 bp window was considered for more accurate estimate.

SNeP package in R [32] was used to determine N_e and trends in N_e trajectories on the basis of relationship between r^2 , N_e and recombination rate (c). A ped file was provided as an input in the package and svedf [33] was the mapping function used. The output file revealed the historical N_e estimates and genome-wide linkage disequilibrium (LD) patterns.

Additionally, ROH and genomic inbreeding coefficient based on ROH (F_{ROH}) in the population were investigated using the consecutive method in detectRUNS package in R [34]. ROH pattern was evaluated in 0–2 Mb, 2–4 Mb, 4–8 Mb and >8 Mb classes.

2.5. Selective Sweep Identification

Selective sweep in the population were identified based on de-correlated composite of multiple signals (DCMS) which is a composite measure of selection combining the power of different individual statistics. We used three statistical parameters viz., allele-frequency spectrum-based methods like nucleotide diversity (π) and Tajima's D and haplotype-based methods like integrated haplotype score (iHS) to estimate the combined DCMS scores.

Similar to the π estimates (calculated earlier), the Tajima's D was also estimated in windows of 100 kb using "TajimaD" flag in the VCFtools package. For calculating iHS scores, haplotype phasing was performed in Beagle [35] to determine the individual haplotypes. Thereafter, iHS scores for 100 kb windows were obtained using selscan package [36]. Un-standardised iHS scores were normalised using 'norm' flag. Normalised iHS scores within each 100 kb window were calculated.

Finally, all three parameters for each 100 kb window were fed into the MINOTAUR package in R [37] to calculate DCMS score for each window. First of all, genome-wide rank-based p -values were generated for each of three statistics using "stat_to_pvalue" function in the MINOTAUR package. Afterwards, covariance matrix was constructed based on 50,000 randomly sampled SNPs (with $\alpha = 0.75$) using "CovNAMcd" function in rrcovNA package in R [38]. This matrix was used to adjust for correlation among the statistics and was used in obtaining a DCMS statistic (using 'DCMS' function) in MINOTAUR package. The resulting statistic was fitted into a normal distribution using a robust linear model in MASS package in R [39]. Finally, the fitted DCMS scores were used as input along with the mean (μ) and standard deviation (SD) of the fitted model to calculate the p -values based on DCMS scores using 'pnorm' function in R. The SNP windows with p values < 0.01 were considered as 'significant' signatures of selection.

2.6. Mapping of Selective Sweeps and Gene Ontology

Based on the marker density and r^2 between adjacent SNPs, LD decay was calculated using GAPIT package in R [40,41]. SNP windows with significant p -values were extended by a certain distance (both upstream and downstream) based on LD decay. This was used to scan the genome for protein-coding genes based on *Bos taurus* genome assembly ARS-UCD 1.2 in UCSC Genome Browser [24,25].

2.7. Validation Based on QTL Mapping

In order to validate our findings, putative regions under selective sweep were examined for the presence of quantitative trait loci (QTL) based on Cattle QTLdb (<https://www.animalgenome.org/cgi-bin/QTLdb/BT/index>, accessed on 10 May 2021) [42] and overlaps between the selected regions and known QTL regions were assessed.

2.8. Gene Ontology

Gene ontology (GO) and statistical overrepresentation of GO terms was done in PANTHER [31] by providing the gene list from the putative regions under selective sweep as an input and *Bos taurus* annotation file as a background. A false discovery rate (FDR) value of ≤ 0.02 was set as the threshold for significance of GO terms.

3. Results

3.1. SNP Annotation

A total of 96.46 million raw reads of Arunachali yak were obtained for the analysis. *SphI* and *MluCI* restriction enzymes were used for creating reduced representation libraries of the samples. Thus, the sequences lacking the restriction cut sites for both the enzymes and those with phred score (base quality score) lesser than 30 were removed. Finally, 89.95 million good quality reads which constituted 93.25% of the total raw reads were generated after processing.

This was followed by alignment of the filtered good quality reads with *Bos taurus* reference genome assembly ARS-UCD 1.2. The average alignment rate across the sample datasets was 97.71% with >85% alignment rate for all the samples.

The variants (SNPs and Indels) were generated at different read depths viz., RD 2, RD 5 and RD 10. (Table 1) SNPs at RD 10 were considered for further analysis and were subjected to additional QC criteria like MAF > 0.05, CR > 90% and *p*-value for H-WE > 0.0001. Only 99,919 SNPs could pass these stringent criteria and were considered for further analysis.

Table 1. Total variants (including Single Nucleotide Polymorphism (SNP) and Insertion-Deletions (Indels)) at different read depths.

	RD 2	RD 5	RD 10
Total variants	755,786	681,556	634,582
SNPs	701,036	631,816	588,573
Indels	54,750	49,740	46,009

On structural and functional annotation of SNPs, it was revealed that the total number of effects were 157,173 and around 48% of these were present in the intronic region followed by 41% in the intergenic region. Only 0.63% of the total effects could be localised to the exonic region. The distribution of SNP effects in different regions of the genome is presented in Figure 2.

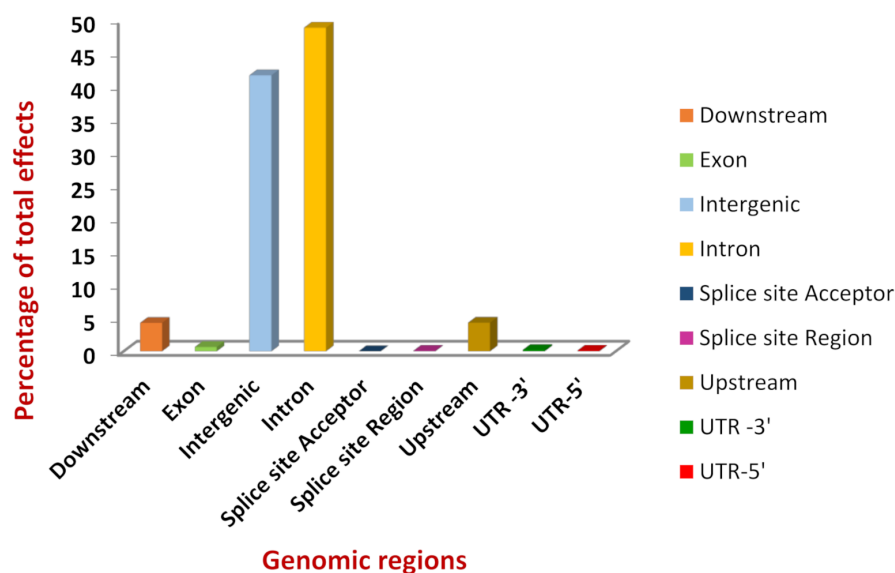


Figure 2. Diagrammatic representation of Single Nucleotide Polymorphism (SNP) effect distribution in the genome of Arunachali yak.

A total of 68.5% (623) of the variant effects were found to be silent causing no change in the amino acid synthesised, whereas 31.5% (286) were found to be mis-sense leading to the synthesis of a different amino acid. The missense: silent ratio was 0.459. In line

with our expectations, the number of transitions (Ts) were more than twice the number of transversions (Tv) i.e., Ts/Tv ratio was 2.879.

With respect to the impact of SNP effects, highly deleterious were about 0.001% (2), low and moderately deleterious constituted 0.5% (730) and 0.2% (286) respectively whereas modifiers were a humongous 99.4% (156,155) of the total effects. The two highly deleterious mutations were located in the non-coding region.

3.2. Identification of Deleterious Mutations

A total of 166 SNPs with mis-sense deleterious mutations and SIFT score ≤ 0.05 were mapped to their genomic locations and a gene list with 106 genes was generated (Supplementary Table S1). These included those involved in reproduction (*FANCD2*, *ADAM18*, *NUF2*), immunity (*TET2*, *GPR33*), development (*TET2*, *TBX19*) and signalling (*STIL*, *VRK2*, *HTR4*, *ADGRA3*) processes. On functional annotation of this gene list, no statistically significant results were found for GO terms related to biological processes or pathways which reflected that the mutations were not deleterious enough to effect any change in protein function.

3.3. Genomic Diversity Estimation

Genomic diversity measured using π with 200 bp windows across the genome was found to be 0.041 indicating sufficient variability in Arunachali yak population.

N_e for the most recent generation (i.e., 13 generations ago) was estimated to be 83. However, a serious declining trend in N_e was emerging in the past more than 100 generations which is depicted in Figure 3 (and Supplementary Table S2). In order to investigate the cause for the drastic reduction in N_e over the generations, ROH were calculated. On an average, there were a total of 371 runs while the average length of ROH in the whole population was 875 kb (minimum length was 11 kb while the maximum length was 9 Mb) (Supplementary Figure S1). The average genomic inbreeding coefficient estimated based on ROH (F_{ROH}) was 0.134. The estimates of the number of ROH and F_{ROH} in the population are presented in Supplementary Table S3, whereas F_{ROH} estimates across different chromosomes in the population are depicted in Supplementary Figure S2. To draw meaningful conclusions from the length of ROH patterns, the ROH were categorised based on: 0–2 Mb, 2–4 Mb, 4–8 Mb and >8 Mb lengths. A high proportion of ROH (94.33%) were predominantly falling in the short category i.e., 0–2 Mb followed by 4.66% in 2–4 Mb and 0.96% in 4–8 Mb category. Only 0.03% of ROH belonged to the longer length (>8 Mb) category in the population. The F_{ROH} estimates for the different categories of ROH are elucidated in Supplementary Table S4.

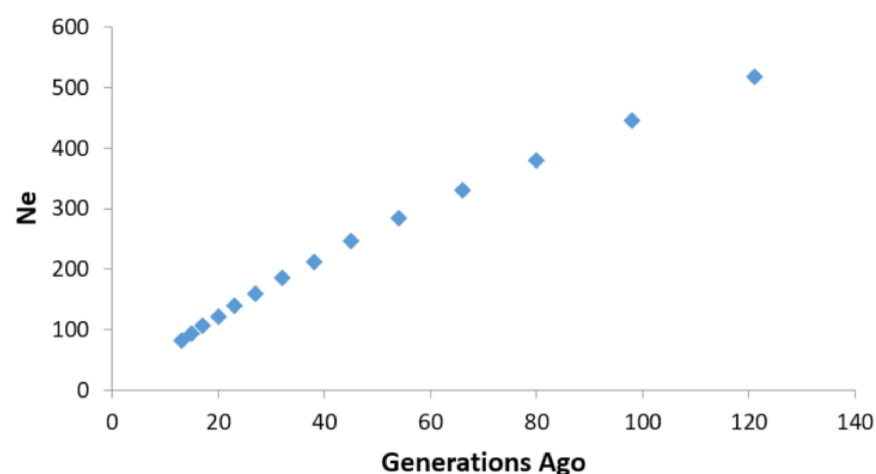


Figure 3. Scatter plot showing N_e over the generations (starting from >100 generations ago) based on LD.

3.4. Selective Sweep Identification

The individual statistics related to allele frequency spectrum-based methods (like Tajima's D and π) for identification of selective sweep were calculated in non-overlapping sliding window of 100 kb each. Similarly for haplotype-based methods like iHS, the phased haplotype file was used and the unstandardised iHS values were computed individually for each SNP location across each *Bos taurus* autosome (BTA). Normalised iHS ($|iHS|$) scores for each 100 kb window were estimated by averaging iHS scores for the SNPs present within that window. The plot for normalised iHS scores against SNP positions is shown in Figure 4. The resulting values for three individual test statistics viz., Tajima's D , π and iHS at each 100 kb window across the genome were combined by decomposition of p -values for each of the test statistics.

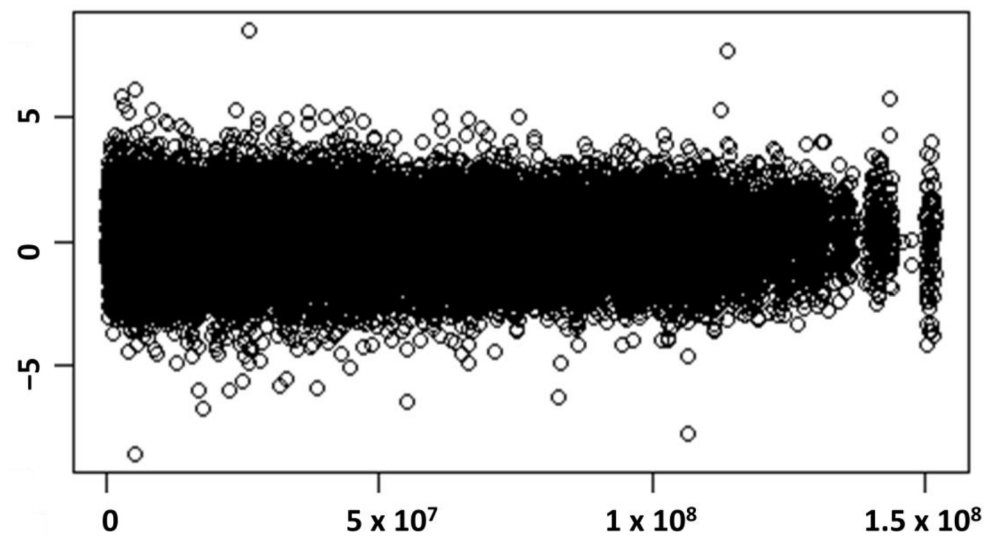


Figure 4. Plot for normalised iHS scores (Y-axis) against SNP position (X-axis) across the genome.

DCMS statistics were calculated individually for each BTA in non-overlapping 100 kb SNP windows. On conversion of scores to p -values, 207 regions were found to be putatively under selection ($p < 0.01$).

Linkage disequilibrium (LD) decay report was generated to reveal the extent of LD in the genome and to identify the candidate selection regions more comprehensively. It was found that r^2 value decayed below 0.2 after 200 kb distance between the adjacent SNPs (Figure 5).

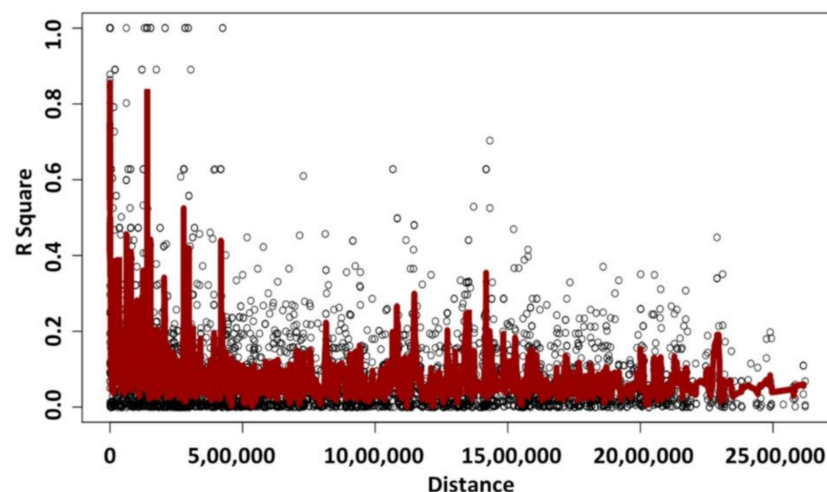


Figure 5. LD decay based on marker density and r^2 in the population.

Hence, the identified region was extended 200 kb upstream and downstream to unveil the genes in the extended region. These putatively selected regions encompassed 611 protein-coding genes in the identified sweep positions. BTA7 had the maximum number of regions subjected to positive selection (22) whereas BTA24, 25, 28 and 29 had minimum number of selective sweeps (2 each). The detailed list of the selective sweep regions along with the number of variants and associated genes is presented in Supplementary Table S5.

3.5. Functional Annotation of the Selected Regions

The retrieved protein-coding genes included those involved in reproduction (*SPATA22*, *OXTR*, *MIAP*), growth (*SEMA3D*, *SEMA5A*, *SEMA4F*, *SEMA3A*, *SLIT2*), immunity (*SKINT1*, *KIT*, *TCAM1*, *OASL*, *TXK*, *TEC*) and behaviour (*PRLH*) etc. Interestingly, several genes regulating climatic adaptation were also found in the sweep regions including *IGF1R* for body size, *KIT* and *KITLG* genes for pigmentation, *CDH12* for subcutaneous fat, *OR5K3*, *OR5H6* and *OR1E1* genes for olfaction, *TAS2R1*, *TAS2R3* and *TAS2R4* genes for taste and *FGG*, *FGA*, *FGB*, *PDGFRA*, *PEAR1*, *STXBP3* genes for platelet aggregation and activation.

Subsequently, in order to perform further phenotypic annotation and validation of our findings, we identified QTL from the CattleQTL database that overlapped with the selected regions (Supplementary Table S6). The putative regions of selection found in our study overlapped with QTLs for growth (body weight, dry matter intake, average daily gain etc.), reproduction (conception rate, heifer pregnancy rate, fertility index, calving ease, sperm motility, twinning etc.), morphological adaptation (eye pigmentation, facial area pigmentation, coat colour, coat texture, degree of spotting, white spotting etc.) and physiological adaptation (lung percentage, lung weight, cold tolerance, haematocrit, subcutaneous fat, red blood cell distribution width, packed cell volume (PCV) variance, final packed RBC volume, methane production, body temperature, respiratory rate etc.). QTLs mapped in relation to climatic adaptation traits are described in Table 2.

Table 2. QTLs related to important traits mapped within the selective sweep regions.

Traits	BTA	SNP Window (bp)	QTL Region (bp)	QTL ID
Lung percentage	2	44,800,001–45,300,000	3,033,483–67,663,650	12152
Lung weight	22	12,300,001–12,800,000	10,166,113–34,111,868	12164
Kidney, pelvic and heart fat percentage	2	7,600,001–8,100,000	5,704,346–14,389,811	4857
	3	43,900,001–44,400,000	37,279,925–50,679,768	1352
	6	24,500,001–25,000,000	3,390,665–51,352,653	12153
	7	1,900,0001–19,500,000	7,081,101–22,650,129	4866
	11	54,300,001–54,800,000	20,106,495–66,060,688	15732
	16	72,800,001–73,300,000	72,931,855–72,931,895	152037
Subcutaneous fat	14	35,300,001–35,800,000	31,219,729–43,140,076	20704
	16	31,900,001–32,400,000	31,970,184–31,970,224	157073
	19	47,600,001–48,100,000	47,922,295–48,033,350	18940
Cold tolerance	7	19,000,001–19,500,000	17,086,793–88,971,675	31181
	25	22,800,001–23,300,000	22,013,058–22,964,883	31197
Heat tolerance	12	24,800,001–25,300,000	25,066,554–25,986,217	31189
Haematocrit	11	9,400,001–10,200,000	9,016,028–9,945,427	213425
PCV Variance	17	12,100,001–12,600,000	4,092,903–14,340,160	10533
	20	16,700,001–17,200,000	12,158,768–22,679,451	10538
Final packed RBC volume	17	12,100,001–12,600,000	4,092,903–14,340,160	10534
Percent decrease in PCV up to day 100 after challenge	13	45,900,001–46,400,000	17,709,118–53,561,417	10525
Red blood cell distribution width	15	15,000,001–15,500,000	14,272,339–15,253,334	213481

Table 2. Cont.

Traits	BTA	SNP Window (bp)	QTL Region (bp)	QTL ID
Haemoglobin	7	1,000,001–1,500,000	971,984–1,887,948	213432
Mean corpuscular haemoglobin concentration	20	41,000,001–41,500,000	41,051,481–42,052,137	213445
Methane production	20	63,300,001–64,300,000	63,407,165–63,407,205	165056
Respiratory rate	24	28,500,001–29,000,000	28,907,134–28,907,174	57040
Eye area pigmentation	6	69,600,001–70,300,000	69,807,007–69,807,047	37348
	4	35,600,001–36,100,000	35,939,851–35,939,891	37346
	5	17,500,001–18,300,000	18,206,797–18,206,837	21151
Facial pigmentation	6	69,600,001–70,300,000	69,807,007–69,807,047	37364
Degree of spotting	18	8,300,001–8,800,000	8,591,333–10,117,552	125378
Coat colour	18	20,100,001–20,600,000	20,060,029–20,212,668	6270
Coat texture	20	37,000,001–37,500,000	37,179,938–37,179,978	32197
White spotting	22	41,700,001–42,200,000	42,111,442–42,111,482	166867

Functional annotation of all the genes in the list was performed to highlight the statistical overrepresentation of the GO terms ($FDR \leq 0.02$). GO terms related to platelet aggregation (GO:0070527) and detection of chemical stimulus involved in sensory perception of smell (GO:0050911) were found to be highly significant ($FDR \leq 0.02$). Other related parent GO terms like platelet activation (GO:0030168), detection of chemical stimulus involved in sensory perception (GO:0050907) and sensory perception of smell (GO:0007608) were also significantly overrepresented in the analysis (Table 3). The complete description of the selective sweep region along with the significant biological processes is presented in Figure 6.

Important genes identified in relation to these processes included: olfactory receptor (OR) genes like *OR5K3*, *OR5H6* and *OR1E1* initiating a neuronal response for triggering sensation of smell [43] and taste receptor genes like *TAS2R1* (Taste 2 Receptor Member 1), *TAS2R3* (Taste 2 Receptor Member 3) and *TAS2R4* (Taste 2 Receptor Member 4) which encode for bitter taste perception [44]. Additionally, as a part of physiological adaptations, fibrinogen genes like *FGG* (Gamma), *FGA* (Alpha) and *FGB* (Beta) along with *PDGFRA* (Platelet Derived Growth Factor Receptor Alpha) gene facilitate platelet aggregation and connective tissue remodelling, thus facilitating wound healing [45,46]. Further, *STXBP3* (Syntaxin Binding Protein 3) and *PEAR1* (Platelet Endothelial Aggregation Receptor 1) genes act to induce platelet activation and secretion [47,48].

Table 3. Significant GO terms related to various biological processes ($FDR \leq 0.02$).

GO Biological Process	No. of Genes	Fold Enrichment	Raw <i>p</i> -Value	FDR
Platelet aggregation	7	11.19	1.02×10^5	2.09×10^2
■ Platelet activation	9	7.35	1.08×10^5	1.94×10^2
■ Homotypic cell-cell adhesion	8	9.30	7.52×10^6	2.16×10^2
■ Cell-cell adhesion	28	2.99	7.97×10^7	3.82×10^3
■ Cell adhesion	38	2.20	1.34×10^5	1.93×10^2
■ Biological adhesion	38	2.19	1.44×10^5	1.88×10^2
Detection of chemical stimulus involved in sensory perception of smell	3	11	8.38×10^9	6.02×10^5
■ Detection of chemical stimulus involved in sensory perception	8	28	1.16×10^5	1.85×10^2
■ Detection of chemical stimulus	8	28	8.51×10^6	2.04×10^2
■ Sensory perception of chemical stimulus	8	27	4.41×10^6	1.58×10^2
■ Sensory perception of smell	3	11	5.85×10^9	8.41×10^5

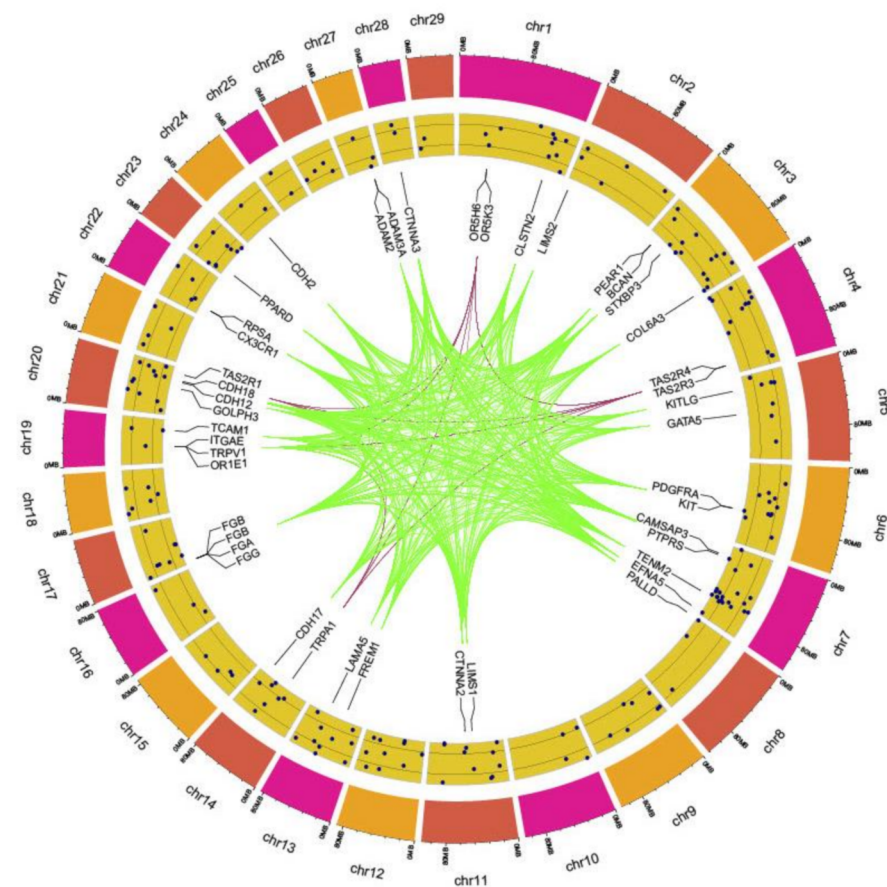


Figure 6. A circos plot depicting putative regions of selective sweep. Outer-Inner: *Bos taurus* auto-somes (BTAs); SNP windows; genes associated with a significant biological process with pink colour showing genes involved in olfaction and green colour depicting genes involved in platelet aggregation.

4. Discussion

The genome of Arunachali yak, the lone registered yak breed of India (till date) has evolved majorly as a result of adaptations in cold and humid environments of Arunachal Pradesh and also due to selection efforts of *Monpa* pastoralists of the region. Small and scattered herd size threatens to introduce inbreeding in the population [49]. Knowledge of genetic diversity of a population is critical for genetic improvement of economic traits and serves as an important guide to update the breeding goals and plans, as per the need and situation [50]. Complementarily, identification of selection signatures or adapted genotypes may reveal new targets for selection and may lead to better informed breeding decisions. This will lead to overall improvement in animal productivity and fitness and will contribute greatly to the economic and food security of the pastoralists [51,52]. Hence, genomic diversity estimation and identification of selective sweeps are the twin preliminary objectives of any genetic improvement programme.

4.1. Genomic Diversity in Arunachali Yak

In Arunachali yak, the genomic diversity profile was generated using three parameters viz., π , N_e and ROH. The average value of π in the population was 0.041 in 200 bp windows across the genome. This indicated sufficient genomic diversity in the population considering the narrow window size. Comparatively lower (0.0011 in 100 kb windows sliding in 10 kb steps) estimates of π have been reported in native yaks of China [53], possibly due to long-term artificial selection and breeding interventions going on in the yak breeds of the country [54]. In a previous study, higher estimates ($\pi = 0.3058$) were reported from the same samples of Arunachali yak [55]. This variation may be attributed to alignment of reads to *Bos mutus* reference genome and a different (probably larger)

window size (not mentioned in their study) used to estimate genomic divergence. Further, Sharma et al. (2018) [49] found some evidence of inbreeding and lower genetic diversity in Arunachali yak population based on microsatellite markers. The contrasting findings may have arisen due to the lesser dense cattle-specific microsatellite markers used in the study.

The effective population size was found to be optimum (83) for the most recent generation (i.e., 13 generations ago). The general consensus based on long-term selection experiments suggests that N_e of 50–100 is viable for long-term survival of the population [56,57]. Bull rotation practices of pastoralists and chance mating with diverse populations like wild yaks during their course of migration may actually be factors influencing the effective population size in Arunachali yaks. Yet the historical trends indicate that a serious decline in N_e has been witnessed over the past more than 100 generations. This calls for persistent efforts to increase the effective population size to greater than 100 in the population in order to maintain the breed for eternity [56]. Nonetheless, errors in N_e estimation owing to small sample size used in the study cannot be ruled out [58].

Higher prevalence of shorter and medium length ROH (0–2 Mb and 2–4 Mb) than longer ROHs (4–8 Mb and >8 Mb) in the population was indicative of past inbreeding and shared ancestry between the parents long ago or some population bottleneck in the past. Absence of longer ROHs also indicates optimum genetic diversity and that there has been almost no recent inbreeding in the population.

Hence, the genomic diversity profiling revealed that inbreeding is not a big threat in the breed, as of now and there is a potential for genetic improvement of the population by exploiting the genetic variability. This can be further corroborated with a rising population trend of the breed over the years, sound breeding practices of yak pastoralists (like bull exchange and bull rotation) and possible introgression of wild yak during migration. However, planned mating programme in the breed is the need of the hour to ensure its continual viability and sustainability.

4.2. Genomic Regions of Adaptive Change in Arunachali Yak

Predominance of shorter ROHs hinted at the distant demographic and selective events which resulted in repeated fragmentation of chromosomal segments due to recombination. So, the identification of putative regions under selective sweeps or signatures of selection was important to bring further clarity. We used composite measure of selection like DCMS to highlight the selective sweeps in the population. As compared to individual test statistics, DCMS can provide an unbiased and more precise criterion to identify genetic variants under selection by improving signal to noise ratio [59]. Consequently, candidate genes related to the putative selective sweeps can be identified with greater power and accuracy.

GO analysis showed that biological processes associated with adaptation like detection of chemical stimulus involved in sensory perception of smell (GO:0050911), platelet aggregation (GO:0070527) and the related parent GO terms were significantly overrepresented ($FDR \leq 0.02$). Olfactory sensation is one of the most genetically evolved physiological adaptation in ruminants at high altitude, and olfactory receptor (OR) genes were highly enriched for hypoxia response in yaks showing evidence of positive selection [10,13,60]. In order to adapt to highly difficult environmental conditions and in response to the diversity in the distribution of vegetation at different altitudes where yaks thrive, OR genes have been highly selected and have undergone rapid evolution during domestication [60]. Moreover, these genes have been implicated in animal adaptation by promoting growth and development of hair follicles [61].

Exposure to high altitude and hypoxic conditions also induces platelet hyperreactivity, leading to enhanced platelet adhesion, activation and aggregation [62,63]. Increased platelet function and fibrinogen levels have been documented as an important component of body's response to chronic hypobaric hypoxia conditions in high altitudes. In a study [64] related to divergent climatic adaptation in yaks and cattle, platelet activation was found to be an enriched biological process encompassing *FGG* and *FGA* (genes also identified in our study). Furthermore, *PDGFRA* (platelet-derived growth factor receptor, alpha polypeptide)

gene has been identified to be putatively under selective sweep in Kholmogor and Yaroslavl cattle breeds residing in high altitude cold and harsh climatic region of Russia [65].

Other identified (but not statistically overrepresented) genes in the sweep regions included: *KIT* (KIT proto-oncogene, receptor tyrosine kinase) and *KITLG* (KIT ligand) which are known as key genes involved in the pathway for coat pigmentation [66,67]; *PRLH* (prolactin releasing hormone) gene is known to regulate food intake by relaying satiety signals consequently, influencing feeding behaviour [68]; *IGF1R* (insulin like growth factor 1 receptor) gene mediating the action of *IGF1* gene, thus stimulating body growth [69]. Immune-response related genes like *SKINT1* (selection and upkeep of intraepithelial T cells protein 1), *TCAM1* (testicular cell adhesion molecule 1), *OASL* (2'-5'oligoadenylate synthetase like), *TXK* (TXK tyrosine kinase) and *TEC* (tyrosine protein kinase) were identified. *SKINT1* mediates T-cell differentiation in thymus [70] whereas *TCAM1* gene facilitates immune response during meiosis [71]. *OASL* gene induces innate immunity in response to viral attack [72] while *TXK* and *TEC* genes further contribute to adaptive immune response of the body [7,73]. QTL analysis further validated the existence of QTLs for growth, immunity and adaptation traits overlapping with the putative selective sweeps identified in our study. Most of the regions identified to be positively selected in yaks have their significance in environmental adaptation including physiological modifications, coat colour and skin pigmentation, olfactory sensation, immunity and immune response and hypoxia-related adaptations [10,13,74]. Majority of these pathways like the HIF (hypoxia-inducible factor) pathway have been related to environmental information processing, environmental adaptations, organismal systems and metabolism [74,75]. Most of the species residing at higher altitudes seem to have undergone convergent evolution with respect to genes present in HIF (hypoxia-inducible factor) pathway like *EGLN1*, *EGLN2*, *EGLN3* and *EPAS1* [10]. However, we did not observe any evidence of positive selection for hypoxia-associated genes in our study. It may be due to the fact that yak habitats and vegetation differ greatly within yak-rearing states of India. Trans-Himalayan states like Jammu and Kashmir and Himachal Pradesh (and also, Tibetan region) encounter an extremely cold and arid climate, whereas Arunachal Pradesh, Sikkim and Uttarakhand experience cold as well as humid climate throughout the year [76]. It is further substantiated by the statement that hypoxia-related genes formed an important component for adaptation in the trans-Himalayan region of Ladakh [77]. Thus, it can be safely assumed that adaptations resulting from natural selection may also vary between the yaks found in different climatic and vegetative conditions. In fact, studies such as the present one are indispensable to reveal the breed-specific signals or divergence signals in order to shed light on the 'breed signatures' [78].

Our findings of selective sweeps in Arunachali yak are mostly a reiteration of the earlier studies for identifying selection signatures in yaks. However, absence of selection signals for important physiological adaptations related to hypoxia in the breed is reflective of divergent selection within the Himalayan landscape due to differing habitats and ecological conditions prevailing in the specific regions. Future studies with larger sample sizes can unveil more interesting insights for further carrying out genome-wide association studies.

5. Conclusions

This study concluded that there was optimum genomic diversity in Arunachali yak breed and genes like *FGG*, *FGA*, *FGB*, *PDGFRA*, *PEAR1*, *STXBP3* and *OR5K3*, *OR5H6*, *OR1E1*, *TAS2R1*, *TAS2R3* and *TAS2R4* have been subjected to strong positive selection for adaptation in the breed. Presence of divergent selection signals further paves the way for identifying breed signatures and designing a genome-wide association study (GWAS) for improving animal productivity and fitness in future.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13020254/s1>, Figure S1: Distribution of mean ROH lengths across different categories in different individuals. (Description: This figure represents the distribution of ROH lengths in four different categories for each individual). Figure S2: F_{ROH} estimates across different chromosomes in different individuals of the population. (Description: This figure shows the

chromosome-wise estimates of F_{ROH} for all the individuals in the population); Table S1: Mapping of deleterious SNPs to their genomic location. (Description: This excel sheet contains a list of genomic variants or SNPs which were found to be deleterious in nature. These SNPs have been mapped to their genomic locations and the gene names have also been mentioned). Table S2: Trends of Effective population size (N_e) in the population across past >100 generations. (Description: The table shows the N_e values over past more than 100 generations based on r^2). Table S3: Estimates of Runs of Homozygosity (ROH) and Inbreeding coefficient based on ROH (F_{ROH}) in the population. (Description: This table presents Runs of Homozygosity (ROH) and inbreeding coefficient based on ROH for each sample considered in the study as well as the average for the whole population). Table S4: Genomic inbreeding coefficients based on ROH (F_{ROH}) for segments of different lengths in the individuals of the population. (Description: ROHs have been categorized into those with different lengths viz. 0–2 Mb, 2–4 Mb, 4–8 Mb and >8 Mb. This table shows the inbreeding coefficient based on ROH for different ROH length categories). Table S5: Regions of selective sweep in Arunachali yak genome. (Description: This excel sheet presents chromosome-wise distribution of selective sweep regions with details and the genes present in those regions). Table S6: Identified QTL regions overlapping with already reported QTLs for various traits. (Description: This file contains separate excel sheets for each chromosome mentioning the reported QTL regions overlapping with the identified QTL regions with special focus on growth and adaptation traits).

Author Contributions: A.K. (Aneet Kour), S.K.N. and M.S. conceived and designed the study; M.M., U.S. and M.P. helped in statistical analysis; J.S. provided the data; A.K. (Amod Kumar) helped in data curation; A.K. (Aneet Kour) and S.K.N. wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets related to this publication were generated by ICAR-NBAGR, Karnal and can be accessed through NCBI (BioProject PRJNA577203).

Acknowledgments: The authors are grateful to ICAR-NBAGR and ICAR-NRC on Yak for providing all the necessary support required for this study. We all are indebted to Indian Council of Agricultural Research (ICAR) for giving us an opportunity to serve the country.

Conflicts of Interest: All the authors declare that they have no conflict of interests.

References

- Wiener, G.; Han, J.; Long, R. *The Yak*, 2nd ed.; Regional Office for Asia and the Pacific, Food and Agriculture Organization of the United Nations: Bangkok, Thailand, 2003.
- Shao, B.; Long, R.; Ding, Y.; Wang, J.; Ding, L.; Wang, H. Morphological adaptations of yak (*Bos grunniens*) tongue to the foraging environment of the Qinghai-Tibetan Plateau. *J. Anim. Sci.* **2010**, *88*, 2594–2603. [[CrossRef](#)] [[PubMed](#)]
- Hochachka, P.W.; Buck, L.T.; Doll, C.J.; Land, S.C. Unifying theory of hypoxia tolerance: Molecular/metabolic defense and rescue mechanisms for surviving oxygen lack. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 9493–9498. [[CrossRef](#)] [[PubMed](#)]
- Storz, J.F.; Scott, G.R.; Cheviron, Z.A. Phenotypic plasticity and genetic adaptation to high-altitude hypoxia in vertebrates. *J. Exp. Biol.* **2010**, *213*, 4125–4136. [[CrossRef](#)] [[PubMed](#)]
- Weber, R.E. High-altitude adaptations in vertebrate hemoglobins. *Respir. Physiol. Neurobiol.* **2007**, *158*, 132–142. [[CrossRef](#)]
- Vij, A.G. Effect of prolonged stay at high altitude on platelet aggregation and fibrinogen levels. *Platelets* **2009**, *20*, 421–427. [[CrossRef](#)]
- Yang, J.; Jin, Z.B.; Chen, J.; Huang, X.F.; Li, X.M.; Liang, Y.B.; Mao, J.Y.; Chen, X.; Zheng, Z.; Bakshi, A.; et al. Genetic signatures of high-altitude adaptation in Tibetans. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 4189–4194. [[CrossRef](#)]
- Edea, Z.; Dadi, H.; Dessie, T.; Kim, K.S. Genomic signatures of high-altitude adaptation in Ethiopian sheep populations. *Genes Genom.* **2019**, *41*, 973–981. [[CrossRef](#)]
- Brutsaert, T.D. Population genetic aspects and phenotypic plasticity of ventilatory responses in high altitude natives. *Respir. Physiol. Neurobiol.* **2007**, *158*, 151–160. [[CrossRef](#)]
- Friedrich, J.; Wiener, P. Selection signatures for high-altitude adaptation in ruminants. *Anim. Genet.* **2020**, *51*, 157–165. [[CrossRef](#)]
- Moore, L.G. HYPOXIA AND REPRODUCTIVE HEALTH: Reproductive challenges at high altitude: Fertility, pregnancy and neonatal well-being. *Reproduction* **2021**, *161*, F81–F90. [[CrossRef](#)]

12. Song, S.; Yao, N.; Yang, M.; Liu, X.; Dong, K.; Zhao, Q.; Pu, Y.; He, X.; Guan, W.; Yang, N.; et al. Exome sequencing reveals genetic differentiation due to high-altitude adaptation in the Tibetan cashmere goat (*Capra hircus*). *BMC Genom.* **2016**, *17*, 122. [[CrossRef](#)] [[PubMed](#)]
13. Guang-Xin, E.; Yang, B.G.; Zhu YBin Duang, X.H.; Basang, W.D.; Luo, X.L.; An, T.W. Genome-wide selective sweep analysis of the high-altitude adaptability of yaks by using the copy number variant. *3 Biotech* **2020**, *10*, 259. [[CrossRef](#)] [[PubMed](#)]
14. Qiu, Q.; Zhang, G.; Ma, T.; Qian, W.; Ye, Z.; Cao, C.; Hu, Q.; Kim, J.; Larkin, D.M.; Auvil, L.; et al. The yak genome and adaptation to life at high altitude. *Nat. Genet.* **2012**, *44*, 946–949. [[CrossRef](#)] [[PubMed](#)]
15. de Simoni Gouveia, J.J.; da Silva, M.V.G.B.; Paiva, S.R.; de Oliveira, S.M.P. (2014). Identification of selection signatures in livestock species. *Genet. Mol. Biol.* **2014**, *37*, 330–342. [[CrossRef](#)] [[PubMed](#)]
16. Wilkins, A.S.; Wrangham, R.W.; Fitch, W.T. The “domestication syndrome” in mammals: A unified explanation based on neural crest cell behavior and genetics. *Genetics* **2014**, *197*, 795–808. [[CrossRef](#)]
17. 20th Livestock Census-2019. 2019. Available online: <https://dahd.nic.in/sites/default/files/Key%20Results%2BAnnexure%2018.10.2019.pdf> (accessed on 7 May 2021).
18. Das, P.J. Arunachali Yak: An alpine Treasure on Northeast India. In *Arunachali Yak*; Das, P.J., Deori, S., Deb, S.M., Eds.; ICAR-National Research Centre on Yak: Dirang, India, 2016; pp. 158–199.
19. Ramesha, K.P. *Yak—Moving treasure of the Himalayas*. ICAR-NRC on Yak, Dirang, Arunachal Pradesh, India. 2008. Available online: <https://books.google.co.in/books?id=MozDxQEACAAJ> (accessed on 10 May 2021).
20. FAO. *The Yak*, 2nd ed.; 2003. Available online: <http://www.fao.org/3/ad347e/ad347e08.html> (accessed on 12 May 2021).
21. Peterson, B.K.; Weber, J.N.; Kay, E.H.; Fisher, H.S.; Hoekstra, H.E. Double Digest RADseq: An inexpensive method for *De Novo* SNP Discovery and Genotyping in Model and Non-Model species. *PLoS ONE* **2012**, *7*, e37135. [[CrossRef](#)]
22. Schmieder, R.; Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **2011**, *27*, 863–864. [[CrossRef](#)]
23. Catchen, J.M.; Amores, A.; Hohenlohe, P.; Cresko, W.; Postlethwait, J.H. Stacks: Building and genotyping loci de novo from short-read sequences. *G3-Genes Genom. Genet.* **2011**, *1*, 171–182. [[CrossRef](#)]
24. UCSC Genome Browser. Available online: <https://genome.ucsc.edu/> (accessed on 10 May 2021).
25. Zimin, A.V.; Delcher, A.L.; Florea, L.; Kelley, D.R.; Schatz, M.C.; Puiu, D.; Hanrahan, F.; Pertea, G.; Van Tassell, C.P.; Sonstegard, T.S.; et al. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* **2009**, *10*, R42. [[CrossRef](#)]
26. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)]
27. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)] [[PubMed](#)]
28. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [[CrossRef](#)] [[PubMed](#)]
29. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **2012**, *6*, 80–92. [[CrossRef](#)] [[PubMed](#)]
30. Ensembl. Available online: <https://www.ensembl.org/vep> (accessed on 11 May 2021).
31. Mi, H.; Dong, Q.; Muruganujan, A.; Gaudet, P.; Lewis, S.; Thomas, P.D. PANTHER version 7: Improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res.* **2009**, *38*, D204–D210. [[CrossRef](#)] [[PubMed](#)]
32. Barbato, M.; Orozco-terWengel, P.; Tapio, M.; Bruford, M.W. SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front. Genet.* **2015**, *6*, 109. [[CrossRef](#)] [[PubMed](#)]
33. Sved, J.A.; Feldman, M.W. Correlation and probability methods for one and two loci. *Theor. Popul. Biol.* **1973**, *4*, 129–132. [[CrossRef](#)]
34. Marras, G.; Gaspa, G.; Sorbolini, S.; Dimauro, C.; Ajmone-Marsan, P.; Valentini, A.; Williams, J.L.; MacCiotta, N.P.P. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Anim. Genet.* **2015**, *46*, 110–121. [[CrossRef](#)]
35. Browning, S.R.; Browning, B.L. 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*, 1084–1097. [[CrossRef](#)]
36. Szpiech, Z.A.; Hernandez, R.D. Selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Mol. Biol. Evol.* **2014**, *31*, 2824–2827. [[CrossRef](#)]
37. Lotterhos, K.E.; Card, D.C.; Schaal, S.M.; Wang, L.; Collins, C.; Verity, B. Composite measures of selection can improve the signal-to-noise ratio in genome scans. *Methods Ecol. Evol.* **2017**, *8*, 717–727. [[CrossRef](#)]
38. Todorov, V.; Templ, M.; Filzmoser, P. Detection of multivariate outliers in business survey data with incomplete information. *Adv. Data Anal. Classif.* **2011**, *5*, 37–56. [[CrossRef](#)]
39. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S*, 4th ed.; Springer: New York, NY, USA, 2002; ISBN 0-387-95457-0. Available online: <https://www.stats.ox.ac.uk/pub/MASS4/> (accessed on 27 April 2021).
40. Zhang, Z.; Ersoz, E.; Lai, C.Q.; Todhunter, R.J.; Tiwari, H.K.; Gore, M.A.; Bradbury, P.J.; Yu, J.; Arnett, D.K.; Ordovas, J.M.; et al. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **2010**, *42*, 355–360. [[CrossRef](#)] [[PubMed](#)]
41. Lipka, A.E.; Tian, F.; Wang, Q.; Peiffer, J.; Li, M.; Bradbury, P.J.; Gore, M.A.; Buckler, E.S.; Zhang, Z. GAPIT: Genome association and prediction integrated tool. *Bioinformatics* **2012**, *28*, 2397–2399. [[CrossRef](#)] [[PubMed](#)]

42. Cattle QTL Database. Available online: <https://www.animalgenome.org/cgi-bin/QTLdb/BT/index> (accessed on 28 May 2021).
43. Malnic, B.; Godfrey, P.A.; Buck, L.B. The human olfactory receptor gene family. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 2584–2589. [[CrossRef](#)]
44. Jeruzal-Swiątecka, J.; Fendler, W.; Pietruszewska, W. Clinical role of extraoral bitter taste receptors. *Int. J. Mol. Sci.* **2020**, *21*, 5156. [[CrossRef](#)]
45. Farrell, D.H.; Thiagarajan, P.; Chung, D.W.; Davie, E.W. Role of fibrinogen α and γ chain sites in platelet aggregation. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 10729–10732. [[CrossRef](#)]
46. Horikawa, S.; Ishii, Y.; Hamashima, T.; Yamamoto, S.; Mori, H.; Fujimori, T.; Shen, J.; Inoue, R.; Nishizono, H.; Itoh, H.; et al. *PDGFR α* plays a crucial role in connective tissue remodeling. *Sci. Rep.* **2015**, *5*, 17948. [[CrossRef](#)]
47. Golebiewska, E.M.; Poole, A.W. Secrets of platelet exocytosis—What do we really know about platelet secretion mechanisms? *Br. J. Haematol.* **2014**, *165*, 204–216. [[CrossRef](#)]
48. Izzi, B.; Gianfagna, F.; Yang, W.Y.; Cludts, K.; De Curtis, A.; Verhamme, P.; Di Castelnuovo, A.; Cerletti, C.; Donati, M.B.; De Gaetano, G.; et al. Variation of *PEAR1* DNA methylation influences platelet and leukocyte function. *Clin. Epigenet.* **2019**, *11*, 151. [[CrossRef](#)]
49. Sharma, H.; Sharma, R.; Ahlawat, S.; Das, P.J.; Jayakumar, S.; Tandia, M.S. Cattle microsatellite markers successfully established diversity status of Arunachali yak (only registered yak breed of India). *Indian J. Anim. Sci.* **2018**, *88*, 1051–1057.
50. Blott, S.C.; Williams, J.L.; Haley, C.S. Genetic relationships among European cattle breeds. *Anim. Genet.* **1998**, *29*, 273–282. [[CrossRef](#)] [[PubMed](#)]
51. Mirkena, T.; Duguma, G.; Haile, A.; Tibbo, M.; Okeyo, A.M.; Wurzinger, M.; Sölkner, J. Genetics of adaptation in domestic farm animals: A review. *Livest. Sci.* **2010**, *132*, 1–12. [[CrossRef](#)]
52. Rege, J.E.O.; Gibson, J.P. Animal genetic resources and economic development: Issues in relation to economic valuation. *Ecol. Econ.* **2003**, *45*, 319–330. [[CrossRef](#)]
53. Chai, Z.X.; Xin, J.W.; Zhang, C.F.; Dawayangla, L.; Zhang, Q.; Pingcuozhandui, L.C.; Zhu, Y.; Cao, H.W.; Wang, H.; Han, J.L.; et al. Whole-genome resequencing provides insights into the evolution and divergence of the native domestic yaks of the Qinghai–Tibet Plateau. *BMC Evolut. Biol.* **2020**, *20*, 137. [[CrossRef](#)] [[PubMed](#)]
54. Ma, Z.J.; Guang-Zhen, L.; Chen, S.M.; Han, J.L.; Hanif, Q. Rich maternal and paternal genetic diversity and divergent lineage composition in wild yak (*Bos mutus*). *Anim. Biotechnol.* **2021**, 1–4. [[CrossRef](#)]
55. Jayakumar, S.; Vineeth, M.R.; Surya, T.; Singh, K.; Dixit, S.P.; Niranjana, S.K.; Tandia, M.S.; Gupta, I.D.; Ravikumar, D. Genomic divergence reveals unique populations among Indian Yaks. *Sci. Rep.* **2020**, *10*, 3636. [[CrossRef](#)]
56. Meuwissen, T.H.E.; Woolliams, J.A. Effective sizes of livestock populations to prevent a decline in fitness. *Theor. Appl. Genet.* **1994**, *89*, 1019–1026. [[CrossRef](#)]
57. Meuwissen, T. Genetic management of small populations: A review. *Acta Agric. Scand. A Anim. Sci.* **2009**, *59*, 71–79. [[CrossRef](#)]
58. Sved, J.A.; Cameron, E.C.; Gilchrist, A.S. Estimating effective population size from linkage disequilibrium between unlinked loci: Theory and application to fruit fly outbreak populations. *PLoS ONE* **2013**, *8*, e69078. [[CrossRef](#)]
59. 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* **2015**, *115*, 426–436. [[CrossRef](#)]
60. Zhang, S.; Zhang, K.; Peng, X.; Zhan, H.; Lu, J.; Xie, S.; Zhao, S.; Li, X.; Ma, Y. Selective sweep analysis reveals extensive parallel selection traits between large white and Duroc pigs. *Evolut. Appl.* **2020**, *13*, 2807–2820. [[CrossRef](#)]
61. Busse, D.; Kudella, P.; Grüning, N.M.; Gisselmann, G.; Ständer, S.; Luger, T.; Jacobsen, F.; Steinsträßer, L.; Paus, R.; Gkogkolou, P.; et al. A synthetic sandalwood odorant induces wound-healing processes in human keratinocytes via the olfactory receptor OR2AT4. *J. Investig. Dermatol.* **2014**, *134*, 2823–2832. [[CrossRef](#)] [[PubMed](#)]
62. Lehmann, T.; Mairbäurl, H.; Pleisch, B.; Maggiorini, M.; Bärtsch, P.; Reinhart, W.H. Platelet count and function at high altitude and in high-altitude pulmonary edema. *J. Appl. Physiol.* **2006**, *100*, 690–694. [[CrossRef](#)]
63. Tyagi, T.; Ahmad, S.; Gupta, N.; Sahu, A.; Ahmad, Y.; Nair, V.; Chatterjee, T.; Bajaj, N.; Sengupta, S.; Ganju, L.; et al. Altered expression of platelet proteins and calpain activity mediate hypoxia-induced prothrombotic phenotype. *Blood* **2014**, *123*, 1250–1260. [[CrossRef](#)] [[PubMed](#)]
64. Freitas, P.H.F.; Wang, Y.; Yan, P.; Oliveira, H.R.; Schenkel, F.S.; Zhang, Y.; Xu, Q.; Brito, L.F. 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*, 102. [[CrossRef](#)] [[PubMed](#)]
65. Zinovieva, N.A.; Dotsev, A.V.; Sermyagin, A.A.; Deniskova, T.E.; Abdelmanova, A.S.; Kharzinova, V.R.; Sölkner, J.; Reyer, H.; Wimmers, K.; Brem, G. Selection signatures in two oldest Russian native cattle breeds revealed using high-density single nucleotide polymorphism analysis. *PLoS ONE* **2020**, *15*, e0242200. [[CrossRef](#)]
66. Amyere, M.; Vogt, T.; Hoo, J.; Brandrup, F.; Bygum, A.; Boon, L.; Vikkula, M. *KITLG* mutations cause familial progressive hyper- and hypopigmentation. *J. Investig. Dermatol.* **2011**, *131*, 1234–1239. [[CrossRef](#)]
67. Picardo, M.; Cardinali, G. The genetic determination of skin pigmentation: *KITLG* and the *KITLG/c-kit* pathway as key players in the onset of human familial pigmentary diseases. *J. Investig. Dermatol.* **2011**, *131*, 1182–1185. [[CrossRef](#)]
68. Takayanagi, Y.; Matsumoto, H.; Nakata, M.; Mera, T.; Fukusumi, S.; Hinuma, S.; Ueta, Y.; Yada, T.; Leng, G.; Onaka, T. Endogenous prolactin-releasing peptide regulates food intake in rodents. *J. Clin. Investig.* **2008**, *118*, 4014–4024. [[CrossRef](#)]

69. Rodriguez, S.; Gaunt, T.R.; Day, I.N.M. Molecular genetics of human growth hormone, insulin-like growth factors and their pathways in common disease. *Hum. Genet.* **2007**, *122*, 1–21. [[CrossRef](#)]
70. Turchinovich, G.; Hayday, A.C. *SKINT-1* identifies a common molecular mechanism for the development of interferon- γ -secreting versus interleukin-17-secreting $\gamma\delta$ T cells. *Immunity* **2011**, *35*, 59–68. [[CrossRef](#)] [[PubMed](#)]
71. Kurihara, M.; Otsuka, K.; Matsubara, S.; Shiraishi, A.; Satake, H.; Kimura, A.P. A testis-specific long non-coding RNA, lncRNA-Tcam1, regulates immune-related genes in mouse male germ cells. *Front. Endocrinol.* **2017**, *8*, 299. [[CrossRef](#)] [[PubMed](#)]
72. Zhu, J.; Ghosh, A.; Sarkar, S.N. OASL—A new player in controlling antiviral innate immunity. *Curr. Opin. Virol.* **2015**, *12*, 15–19, Elsevier, B.V.. [[CrossRef](#)] [[PubMed](#)]
73. Mihara, S.; Suzuki, N. Role of *TXK*, a member of the Tec family of tyrosine kinases, in immune-inflammatory diseases. *Int. Rev. Immunol.* **2007**, *26*, 333–348. [[CrossRef](#)]
74. Guangxin, E.; Yang, B.G.; Basang, W.D.; Zhu, Y.B.; An, T.W.; Luo, X.L. Screening for signatures of selection of Tianzhu white yak using genome-wide re-sequencing. *Anim. Genet.* **2019**, *50*, 534–538. [[CrossRef](#)]
75. Guang-Xin, E.; Basang, W.D.; Zhu, Y.B. Whole-genome analysis identifying candidate genes of altitude adaptive ecological thresholds in yak populations. *J. Anim. Breed. Genet.* **2019**, *136*, 371–377. [[CrossRef](#)]
76. Pal, R.N. *Yak Husbandry in India*; FAO Regional Office for Asia and the Pacific: Rome, Italy, 2003.
77. Verma, P.; Sharma, A.; Sodhi, M.; Thakur, K.; Bharti, V.K.; Kumar, P.; Giri, A.; Kalia, S.; Swami, S.K.; Mukesh, M. Overexpression of genes associated with hypoxia in cattle adapted to Trans Himalayan region of Ladakh. *Cell Biol. Int.* **2018**, *42*, 1141–1148. [[CrossRef](#)]
78. Gutiérrez-Gil, B.; Arranz, J.J.; Wiener, P. An interpretive review of selective sweep studies in *Bos taurus* cattle populations: Identification of unique and shared selection signals across breeds. *Front. Genet.* **2015**, *6*, 167. [[CrossRef](#)]