Detection of whole genome selection signatures of Pakistani Teddy goat

Rashid Saif1 · Jan Henkel2 · Tania Mahmood1 · Aniqa Ejaz1 · Fraz Ahmad1 · Saeeda Zia3

Received: 26 June 2021 / Accepted: 27 September 2021 / Published online: 5 October 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract
Background Natural and artificial selection tend to cause variability that contributes to shape the genome of livestock in a way that differentiates them among the animal kingdom. The particular aim here is to identify positive selection signatures with whole genome pooled-sequence data of Pakistani Teddy goat.

Methods and results Paired-end alignment of 635,357,043 reads of Teddy goat with (ARS1) reference genome assembly was carried out. Pooled-Heterozygosity (Hp) and Tajima’s D (TD) are applied for validation and getting better hits of selection signals, while pairwise FST statistics is conducted on Teddy vs. Bezoar (wild goat ancestor) for genomic differentiation, moreover annotation of regions under positive selection was also performed. Hp score with −ZHp > 5 detected six windows having highest hits on Chr. 29, 9, 25, 15 and 14 that harbor HRASLS5, LACE1 and AXIN1 genes which are candidate for embryonic development, lactation and body height. Secondly, −ZTD value of > 3.3 showed 4 windows with very strong hits on Chr.5 & 9 which harbor STIM1 and ADM genes related to body mass and weight. Lastly, −ZFST < − 5 generated four strong signals on Chr.5 & 12 harbor LOC102183233 gene. Other significant selection signatures encompass genes associated with wool production, prolificacy and coat colors traits in this breed.

Conclusions In brief, this study identified the genes under selection in Pakistani Teddy goat that will be helpful to refining the marker-assisted breeding policies and converging required production traits within and across other goat breeds and to explore full genetic potential of this valued species of livestock.

Keywords Pooled-seq · Teddy goat · Selection signatures · FST · TD · Hp statistics

Introduction

Capra hircus are small ruminants whose domestication dates back to ~ 10,000 years ago [1]. Goats are mainly reared for meat, milk and wool production [2]. Classical domestication and breeding practices allowed geneticists and animal breeders to explore the inheritance of the economical traits in this species. Meanwhile, industrial revolution and commercial needs triggered the masses to develop genomic technologies along with refining husbandry techniques to get maximum outputs by adaptation of this species to diverse environments and make them specialized for valued products. Estimated goat population in Pakistan is 76.1 million which produces ~ 940 thousand tonnes of milk and ~ 344,000 tonnes of meat annually [3]. Spontaneous phenotypic mutants have been studied and selected by artificial selection methods which are hallmarks of today’s goat breeds like coat colors, meat and milk production, fecundity and adaptation traits. Teddy goat in Punjab, Pakistan is characterized by its short stature, weighing approximately 23–34 kg, quality meat production and weather tolerant characters [4]. In our previous attempts for searching selective sweeps, we identified coat colour structural variants, SNPs and CNVs in 20 domesticated Pakistani/Swiss goat breeds [5], and particularly analyzed 8 Pakistani breeds for putative variants responsible for large body size traits [6].

Aim of the current study is to report complete breed defining hotspots with reduced heterozygosity called “selective sweeps” in this Teddy goat breed from Pakistan, which
is primarily raised for meat purposes in this region. Ultimate goal of this endeavor is to aid in selecting particular characteristics of this breed through artificial selection which could also be helpful to explore the full genetic potential of this breed in its successive generations and to conserve the genetic resources of this species.

**Materials and methods**

**Sample collection and whole genome pooled sequencing**

Whole blood samples of Teddy goat (n = 12) from home tract of this breed and Bezoar wild ancestor (n = 8) were collected from Punjab/Pakistan and Switzerland respectively. Genomic DNA of Pakistani breeds were extracted through standard protocol using TIANGEN biotech (Beijing) CO., LTD [7], while Bezoar DNA extraction was performed at Institute of Genetics, University of Bern, Switzerland. Extracted DNA were mixed into a single pool/breed in equimolar ratios. High-throughput sequencing was conducted using Illumina HiSeq3000 platform which generated 150 bp paired-end ~ 300mio reads. These sequences were further submitted to European Nucleotide Archive (ENA) under Project ID: PRJEB23815 with sample accession numbers ERS2037817 for Teddy and ERS2037806 for Bezoar. Characteristics & representative animals of Teddy (breed), Bezoar (wild ancestor) and San Clemente (reference genome) are shown in Table 1; Fig. 1.

**Mapping and variant calling**

Quality checks were performed on both pools FastQ files using FastQC (v0.11.8) software. While Trimomatic (v0.36) was applied for base quality filtration using SLIDINGWINDOW:4:20 MINLEN:2 parameters [8]. Filtered reads were aligned with ARS1 reference goat genome assembly using BWA-MEM algorithm v0.7.17 [9]. SAM files were converted to BAM using samtools view. BAM files were sorted on coordinate basis and adapter sequences were marked duplicate using Picard-SortSam and MarkDuplicates features respectively [9]. Single Nucleotide Variants (SNVs) were detected using samtools mpileup which yielded combined mpileup file of Teddy and Bezoar as well as separate pileup file of Teddy for Hp and Tajima’s D analysis [10]. Popoolation2 v1.201 tool scripts, mpileup2sync.jar with parameters –fastq-type sanger, –min-qual 20 and snp-frequency-diff.pl was applied on mpileup and pileup files, which generated synchronized (sync) combined mpileup and separate sync pileup files [11].

**Genome wide selection scanning**

Three statistical tests were applied for the detection of genomics selection imprints left in this Teddy breed genome and generated the SNVs. Resulting values were then Z-transformed by applying Z values = -(value-μ/σ) such that if any −ZHp value > 5, −ZTD > 3.3 and −ZFST < −5 as a best hit for considering that window under positive selection.

**Detecting selective sweeps using Pooled-Heterozygosity (Hp)**

First of all we calculated Hp score for both pools by using an in-house Ruby script which applies \(Hp = \frac{2Σn_{MAJ}Σn_{MIN}}{(Σn_{MAJ} + Σn_{MIN})^2}\), where \((n_{MAJ})\) and \((n_{MIN})\) are major and minor allele counts with window-size of 150 kb.

---

**Table 1** Phenotypic traits of the Teddy (breed), Bezoar (wild ancestor) and San Clemente (reference genome)

| Breed                  | Specific trait                                      | Origin   | Animals per pool |
|------------------------|-----------------------------------------------------|----------|------------------|
| Teddy (Capra hircus)   | Meat producing, tender meat quality, small to medium size | Pakistan | 12               |
| Bezoar (Capra aegagrus)| Wild ancestor                                       | Switzerland | 08               |
| San Clemente (ARS1 reference goat genome) | Meat producing, relatively small, mostly red and tan with black markings | USA | 01               |

---

Fig. 1 True representatives of a Teddy, b San Clemente (ARS1-reference) and c wild ancestor Bezoar. (Color figure online)
Detecting selective sweeps using Tajima’s D (TD) statistics

Neutrality statistics as classical Tajima’s D was computed which implies $D_b = \frac{d_b}{\sqrt{\text{Var}(d_b)}}$ in popoolation v1.2.2 tool script variance-sliding.pl [12] that was run on separate pileup files with –min-count 1 –min-coverage 3 –max-coverage 50 and –fastq-type sanger.

Detecting genomic differentiation using fixation index ($F_{ST}$) analysis

We also ran fst-sliding.pl script of popoolation2 v1.201 [13] based on $F_{ST} = \frac{s^2}{\bar{p}(1-\bar{p})} + \frac{s^2}{r}$ [14] with 50% overlapping window for each SNV value in earlier generated combined sync file with settings –min-count 2 –min-coverage 4 –max-coverage 50 –suppress-non informative and pool-size 12:8.

Goat reference genome

ARS1 goat reference genome accession number GCF_001704415.1 was obtained from NCBI and used for annotations.

SNP data visualization

R software was used for the construction of SNP density graph of $-ZH_p$, $-ZTD$ and $-ZF_{ST}$ scores using CMplot package [15]. For $-ZH_p$, $-ZTD$ and $-ZF_{ST}$ scores, Manhattan plots were constructed by qman package on R. Horizontal threshold lines were drawn on manhattan plots which signifies the chosen cutoff values ($-ZH_p > 5$, $-ZTD > 3.3$ and $-ZF_{ST} < -5$).

Quantile–quantile plots and histogram

The $-ZH_p$, $-ZTD$ and $-ZF_{ST}$ values were plotted against expected values as normal Q–Q plots using the function “qqnorm” on R software [16]. Standard normal distribution diagonal line which represents the expected values was drawn with “qline” function on R. To check the distribution of $-ZH_p$, $-ZTD$ and $-ZF_{ST}$ values across all autosomes, histograms were constructed using the function “hist”.

Results

Quality checks and SNPs calling

Total 248,890,548 variants were called from the Teddy genome using aforementioned three applied statistics after quality checks, trimming and mapping steps using ARS1 reference genome assembly. A total of 635,357,043 (95.83%) reads passed the quality threshold for onward genome wide positive selection signature scanning (Supplementary Fig. S1). $H_p$ analysis called 26,115,502 SNPs and after applying $-ZH_p > 5$ threshold, 33,324 SNPs were obtained. TD analysis called 66,775,895 SNPs and after applying $-ZTD > 3.3$ threshold, 45,512 SNPs were obtained. Similarly, $F_{ST}$ analysis called 155,999,151 SNPs and after setting threshold of $F_{ST} < -5$, 201,361 SNPs were obtained for downstream analysis. Distribution of these SNPs within 10,000 MB window are shown in (Supplementary Fig. S2).

Selective sweeps and harbor genes

The windows under positive selection were obtained after setting thresholds on the basis of previously published data and the rationale observation of current data. Further fine-mapping and annotations were performed which revealed that body weight/mass, reproduction, milk production, litter size, wool production and coat color related genes harbor in these selective sweeps (Table 2).

Selection footprints by Pooled-Heterozygosity analysis

$H_p$ is applied using sliding window approach. Based on the generated scores (Fig. 2), six windows have highest $-ZH_p$ values of 11.91, 11.45, 11.20, 11.34, 10.24 and 9.18 harbor Chr.29:42,075–42,225 kb, next two regions on Chr.9:28,875–29,100 kb, Chr.25 in a window from 150 to 300 kb, Chr.15:81,750–81,900 kb and Chr.14:16,050–16,200 kb region having 284, 337, 423, 94, 345 and 409 SNPs respectively. Total 82 windows exhibits strong signals ($-ZH_p > 5$) which harbor forty genes related to body mass/weight, body height, milk/lactation, coat color, hair fleece development involved in wool production, number of teats and litter size, for embryonic development and reproduction. Twenty windows comprise LOCs (genes not having orthologues), while 6 windows harbor no genes (Supplementary Table S1).

In addition, theoretical distribution of millions of SNPs on x-axis are mapped with observed $-ZH_p$ scores on y-axis (Supplementary Fig. S3a), where black line shows deviated SNPs from the tail at both ends as compared to red line of theoretical distribution. Likewise, histogram of $-ZH_p$ scores across all autosomes on x-axis vs. its frequencies on y-axis reveal that only handful of SNPs have either very low or very high $-ZH_p$ values (Supplementary Fig. S3b).
Genetic hitchhiking by Tajima’s D statistics

By applying TD statistics, 41 top hit windows of 150 kb with the cutoff threshold of > 3.3 harbor strong selective sweeps (Fig. 3).

The 4 putative windows containing outliers are positioned on Chr.15:32,100–32,250 kb, 32,175–32,325 kb, 33,975–34,125 kb and 39,900–40,050 kb region with −ZTD values 4.64, 4.62, 4.50 and 4.10 including 857, 836, 810 and 1024 number of SNPs respectively. Other significant signatures in the genome are located on Chr.29:39,525,000–39,675,000 kb region, Chr.15:32,025,000–32,175,000 kb, Chr.11:15,075,000–15,225,000 kb, and on Chr.9:29,250,000–29,400,000 kb containing 1345, 969, 1089 and 1084 SNPs in total respectively. Further, the identified twenty genes under selection regions are related to body mass/weight, body height, lactation and reproduction while rest of 3 regions do not harbor any genes and 3 windows appeared with LOC (Supplementary Table S2).

The distribution of −ZTD values across all autosomes are observed as shown in black line against the expected standard normal distribution in red line (Supplementary Fig. S4a) along with frequency distribution graph of −ZTD values (Supplementary Fig. S4b). A plethora of observed polymorphisms at the tail of the Q–Q plot deviates from the bulk of empirical distribution which deemed statistically significant and illustrates SNPs that are predominantly linked with particular traits of Teddy.

Table 2  List of genes under positive selection in Teddy breed and their associated traits

| Genes under selection | Selection traits | References |
|-----------------------|------------------|------------|
| ITGB1, LRRIQ3, CCDC152, NBEA, ADM, NUP98, GPR21, DPH6, GNBI, IPO9, RELN, ASH11, STRBP, STIM1, TGFBR3, MYCBP2, MAB21L1, KREME1, SH2B3, FAM149B1, XRCC4, FGny, TRPC7, HMBOX1, SEC63, PRRC2C, SSU72, ZNF3, VWDE, SP8, CEP57L1, NR2E1 | Meat production, quality and tenderness, body mass, body weight at birth, skeleton development | [17–19] |
| BIRC6, RTREL1, IFT88, DCHS2, TADA2A, SYNRG, HRASLS5, PRR12, SPEF2 | Reproductive performance, embryonic development, fecundity rate, number of teats, litter size | [20, 21] |
| FOXO3, LACE1, TTC27, RTREL1, TDRD3, VPS13B | Milk production and lactation persistency | [22, 23] |
| SCAPER, AXIN1, STRBP, TCDC152, DLG4, DCBLD1, PDE5A, GATAD2B | Body height, Short stature | [24] |
| PKIA | Hair fleece development, wool production | [25] |
| KIT | Coat color | [5] |

*Denotes genes involved in more than one functions

Fig. 2  Manhattan plot demonstrating −ZHp values from Teddy goat breed. The blue horizontal line shows the suggested significant cutoff threshold of −ZHp > 5 for better hits. (Color figure online)
Genomic differentiation by fixation index

Highly differentiated regions of Teddy vs. Bezoar (Fig. 4) are selected by setting the permissive threshold of $-Z_{FST} < -5$ that comprise 48 windows.

On Chr.5, two of the regions are under positive selection as they appeared very far from our set threshold having $F_{ST}$ value of $-12.30$ and $-8.47$ containing 297 and 1236 SNPs at 98,925–99,150 kb region, while the other two very strong signal on Chr.12 were found in the genomic region of 98.925–99,075 kb comprising of 1486 and 3145 SNPs with $F_{ST}$ value of $-9.58$ and $-7.94$ respectively. Fine mapping of under selection windows harbor five genes related to milk, meat and reproduction while twelve windows have LOCs and 32 regions are devoid of genes (Supplementary Table S3).

To evaluate the significance of $-Z_{FST}$ scores, the Q–Q plot is also constructed which shows that the observed (sample) quantiles of the $-Z_{FST}$ values detected against the expected (theoretical) quantiles has some outliers at tail (Supplementary Fig. S5a). These deviated SNPs whose values are far from zero are expected to be responsible for the genetic differentiation between the two populations. Moreover, the frequencies of fixation index profiles are also

Fig. 3 Manhattan plot of Tajima’s D values. Horizontal red bar is drawn to indicate the significant threshold of $-Z_{TD} > 3.3$. (Color figure online)

Fig. 4 Manhattan plot of $-Z_{FST}$ scores. Horizontal red line points the preferred significant threshold for absolute values of $-Z_{FST} < -5$. (Color figure online)
observed by computing histogram (Supplementary Fig. S5b).

**Common selective sweeps observed by Hp, TD and FST statistics**

Several significant windows were found common among at least two of the three applied statistics and proposed as the best selection hits. Total 18 common positive signatures harbor 12 genes associated with meat, milk, body height and reproduction traits (Table 3).

**Discussion**

Selective sweep regions in the Teddy breed which either under natural or artificial selection are detected by three statistics e.g. *Hp, TD* and *FST*. The main trait of Teddy goat breed is its short stature and tender meat production whose consumption in Pakistan is preferred among other mutton breeds/species due to its soft and leaner properties on various religious occasions. Also improved reproductive performance and better fertility rate are other valued attributes of this goat breed. Recently, this meat producing Teddy breed is artificially selected for milk yield, as we find six genes affecting lactation. Similarly, strong selection for wool production and coat color in Teddy goats due to economic interest might have led to the detected selective sweeps.

Our study found 100 promising genes in this breed from which 32 genes influence meat production. Examples include *LRRIQ3, TGFBR3* and *FGGY* that also controls the number, diameter of muscle fiber which affects meat quality, tenderness, fat deposition and muscle growth in swine and average daily gain in Nellore cattle [26–29]. Nine genes putatively associated with embryonic development, high fecundity rate that additionally control litter size are also observed e.g. *BIRC6, RTEL1, SYNRG* etc. [20, 25]. Similarly, we identified *FOXO3, LACE1, TTC27, RTEL1, TDRD3* and *VPS13B* related to lactation [22, 23, 30]. The genetic architecture underlying body height trait in Teddy presents 8 candidate genes e.g. *SCAPER* and *AXIN1* associated with body height trait in humans [24]. Other selection hits include genes *PKIA* and *KIT* which functions for hair fleece development enhancing wool production and in defining coat color phenotypes respectively [5, 25, 31].

In conclusion, we found several strong signals around distinguished genomic regions especially harboring Chr. 5, 9, 12, 15, 14, 25 and 29. A set of common significant windows were also revealed by more than one aforementioned statistical methods which are located on Chr. 3, 4, 5, 9, 11, 12, 13, 15, 17, 21. Further fine-mapping is still needed for more comprehensive understanding of selective sweeps in this breed. This research provides genome wide maps of selection footprints in this Pakistani goat, that will help in better understanding of genomic architecture of this breed and for further improvements of its existing production traits.

| Chr | Common selective sweeps | Statistical tests | Gene | Selected traits | Significant values |
|-----|-------------------------|------------------|------|-----------------|--------------------|
| 3   | 68,850–69,000 kb        | *Hp, TD*          | TGFBR3 | Skeleton development | $-Z_{Hp} = 6.085, -Z_{TD} = 3.680$ |
| 4   | 90,900–91,050 kb        | *Hp, TD*          | SP8   | Body weight     | $-Z_{Hp} = 5.037, -Z_{TD} = 3.341$ |
| 5   | 98,925–99,075 kb        | *Hp, FST*         | LOC102183233 | – | $-Z_{Hp} = 7.710, -Z_{FST} = -8.474$ |
| 5   | 99,000–99,150 kb        | *Hp, FST*         | No gene | – | $-Z_{Hp} = 8.111, -Z_{FST} = -12.305$ |
| 9   | 28,800–28,950 kb        | *Hp, TD*          | FOXO3 | Milk production | $-Z_{Hp} = 7.788, -Z_{TD} = 3.305$ |
| 9   | 28,875–29,025 kb        | *Hp, TD*          | LACE1 | Milk production | $-Z_{Hp} = 11.450, 11.203, 8.651$ |
|     | 28,950–29,100 kb        |                  |      | –               | $Z_{TD} = 3.988, 3.984, 3.783$ |
|     | 29,025–29,175 kb        |                  |      |                 |                    |
| 11  | 15,075–15,225 kb        | *Hp, TD*          | TTC27 | Milk production | $-Z_{Hp} = 6.111, -Z_{TD} = 3.900$ |
| 12  | 13,725–13,875 kb        | *Hp, FST*         | BIRC6 | Embryonic development | $-Z_{Hp} = 6.05, -Z_{FST} = 3.630$ |
| 12  | 13,800–13,950 kb        | *Hp, FST*         | No gene | – | $-Z_{Hp} = 8.839, -Z_{FST} = -9.588$ |
| 13  | 53,325–53,475 kb        | *Hp, TD*          | RTEL1 | Milk production, reproduction | $-Z_{Hp} = 7.638, -Z_{TD} = 3.574$ |
| 13  | 19,275–19,425 kb        | *Hp, FST*         | ITGB1 | Milk production | $-Z_{Hp} = 8.234, -Z_{FST} = -7.585$ |
| 15  | 32,100–32,250 kb        | *Hp, TD*          | STIM1 | Body weight/mass | $-Z_{Hp} = 5.808, 5.753, -Z_{TD} = 4.628, 4.642$ |
|     | 32,125–32,325 kb        |                  |      |                 |                    |
| 17  | 56,700–56,850 kb        | *Hp, FST*         | No gene | – | $-Z_{Hp} = 7.221, -Z_{FST} = -5.519$ |
| 21  | 31,350–31,500 kb        | *Hp, TD*          | SCAPER | Body height | $-Z_{Hp} = 5.864, -Z_{TD} = 3.491$ |
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11033-021-06726-x.

Acknowledgements Authors are obliged to Prof. Dr. Tosso Leeb & Dr. Vidhya Jagannathan for their continuous support and valued time for accomplishing this work. Thanks to Dr. Mushtaq Ahmad Gondal and Mr. Muhammad Sidique for helping in sample collection.

Author contributions RS envisaged the idea, involved in critical thinking, analysis of data, editing, proofread and correspondence with journal. JH was involved in understandings the software and Linux features, analysis of data and provision of in-house Hp script. TM, AE and FA helped in data analysis and initial write-up. SZ helped in understanding the statistical methods and proof reading of the manuscript.

Funding Sequencing of Pakistani goats were funded by Swiss National Science Foundation (31003A_172964), when RS was postdoc fellow at University of Bern under Swiss Government Excellence Scholarship/ Hans Sigrist Foundation and now further analyzing this already generated publicly available data.

Data availability Relevant data is available in the manuscript including supplementary files and supporting information. This article has already been submitted to a preprint platform distributed under a Creative Commons CC BY license having DOI number https://doi.org/10.20944/preprints202012.0421.v1.

Code availability FST, TD scripts are available publicly, while Hp statistics was in-house written ruby script. Manhattan and other supplementary plots were plotted using Bioconductor qMan package using R.

Declarations

Conflict of interest There are no competing interests among the authors.

Ethical approval All Pakistani animals were sampled with the consent of their owners by obeying the local regulations laid down by “Ethical Review Committee” UVAS-Pakistan, while sampling of Swiss animals was permitted by the “Cantonal Committee for Animal Experiments” (Canton of Bern; permit 75/16).

Consent to participation Authors informed to the farmers and took their consent before collecting samples of their animals.

References

1. Naderi S et al (2008) The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. Proc Natl Acad Sci USA 105:17659–17664. https://doi.org/10.1073/pnas.0804782105

2. Frantz LA, Bradley DG, Larson G, Orlando L (2020) Animal domestication in the era of ancient genetics. Nat Rev Genet 21:449–460. https://doi.org/10.1038/s41576-020-0225-0

3. Ejaz Wasti MSH, Asif M, Shoaib Malik M, Husnain Z, Khatoon S, Shah A, Qureshi A, Farooq O, Saddiq S, Mazhar N, Shujaat Malik Awan MAH, Arslan M, Syed Muhammad Azeem H, Anwar F, Faisal Shamim M (2018–2019) Pakistan economic survey (ed. by Economic Adviser’s Wing FD)

4. Kuthu Z, Javed K, Awan K, Ahmad N, Ahad A (2017) A study on phenotypic and genetic correlations between birth weight and other growth traits in Teddy goat. J Dairy Vet Anim Res. https://doi.org/10.15406/jdvar.2017.05.00158

5. Henkel J et al (2019) Selection signatures in goats reveal copy number variants underlying breed-defining coat color phenotypes. PLoS Genet 15:e1008536. https://doi.org/10.1371/journal.pgen.1008536

6. Saif R, Henkel J, Jagannathan V, Drögemüller C, Flury C, Leeb T (2020) The LCORL locus is under selection in large-sized Pakistani goat breeds. Genes 11:168. https://doi.org/10.3390/genes11020168

7. TIANGEN BIOTECH (BEIJING) CO., LTD. https://www.tiangen.com/en/?product1/t1/4.html

8. Bolger AM, Mosh C, Usadel B (2014) Trimomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

9. Li H et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

10. Rubin C-J et al (2012) Strong signatures of selection in the domestic pig genome. Proc Natl Acad Sci USA 109:19529–19536. https://doi.org/10.1073/pnas.1217149109

11. Koller R, Pandey RV, Schlötterer C (2011) PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics 27:3435–3436. https://doi.org/10.1093/bioinformatics/btr289

12. Korneliussen TS, Molle T, Albrechtsen A, Nielsen R (2013) Calculation of Tajima’s D and other neutrality test statistics from low depth next-generation sequencing data. BMC Bioinform 14:1–14. https://doi.org/10.1186/1471-2105-14-289

13. Wang X et al (2016) Whole-genome sequencing of eight goat populations for the detection of selection signatures underlying production and adaptive traits. Sci Rep 6:38932. https://doi.org/10.1038/srep38932

14. Guo J et al (2018) Whole-genome sequencing reveals selection signatures associated with important traits in six goat breeds. Sci Rep 8:1–11. https://doi.org/10.1038/s41598-018-28719-w

15. Zhou C et al (2019) Genome-wide association study for milk protein composition traits in a Chinese Holstein population using a single-step approach. Front Genet 10:72. https://doi.org/10.3389/fgen.2019.00072

16. Shaffer L, Young T, Guess F, Bensmail H, León R (2008) Using R software for reliability data analysis. Int J Reliabil Appl 91:53–70

17. Raza SHA et al (2020) Bioinformatics analysis and genetic polymorphisms in genome region of the bovine SH2B2 gene and their associations with molecular breeding for body size traits in qinchuan beef cattle. Biosci Rep. https://doi.org/10.1042/BSR20192113

18. Tao L, He X, Pan L, Wang J, Gan S, Chu M (2020) Genome-wide association study of body weight and conformation traits in neonatal sheep. Anim Genet 51:336–340. https://doi.org/10.1111/age.12904

19. Zonaed Siddiki A, Miah G, Islam M, Kumkum M, Rum MH, Baten A, Hossain MA (2020) Goat genomic resources: the search for genes associated with its economic traits. Int J Genomics. https://doi.org/10.1155/2020/5940205

20. El-Halawany N, Zhou X, Al-Tohamy AF, El-Sayd YA, Shawky A, Michal JJ, Jiang Z (2016) Genome-wide screening of candidate genes for improving fertility in Egyptian native Rahmani sheep. Anim Genet. https://doi.org/10.1111/age.12437

21. Yurchenko AA et al (2019) High-density genotyping reveals signatures of selection related to acclimation and economically important traits in 15 local sheep breeds from Russia. BMC Genomics 20:294. https://doi.org/10.1186/s12864-019-5537-0
22. Nayeri S et al (2019) Genome-wide association analysis for β-hydroxybutyrate concentration in milk in Holstein dairy cattle. BMC Genet 20:58. https://doi.org/10.1186/s12863-019-0761-9

23. Zhang S et al (2019) Comparative transcriptomic and proteomic analyses identify key genes associated with milk fat traits in Chinese Holstein cows. Front Genet 10:672. https://doi.org/10.3389/fgen.2019.00672

24. Carty CL, Johnson NA, Hutter CM, Reiner AP, Peters U, Tang H, Kooperberg C (2012) Genome-wide association study of body height in African Americans: The women’s health initiative SNP health association resource (share). Hum Mol Genet 21:711–720. https://doi.org/10.1093/hmg/ddr489

25. Xu S-S, Li M-H (2017) Recent advances in understanding genetic variants associated with economically important traits in sheep (Ovis aries) revealed by high-throughput screening technologies. Front Agric Sci Eng 4:279–288

26. Nonneman D, Shackelford S, King D, Wheeler T, Wiedmann R, Snelling W, Rohrer G (2013) Genome-wide association of meat quality traits and tenderness in swine. J Anim Sci 91:4043–4050. https://doi.org/10.2527/jas.2013-6255

27. Santana M et al (2014) Genome-wide association study for feedlot average daily gain in Nellore cattle (Bos indicus). J Anim Breed Genet 131:210–216. https://doi.org/10.1111/jbg.12084

28. Jeong H et al (2015) Exploring evidence of positive selection reveals genetic basis of meat quality traits in Berkshire pigs through whole genome sequencing. BMC Genet 16:1–9. https://doi.org/10.1186/s12863-015-0265-1

29. Zhang L et al (2020) Genomic variants associated with the number and diameter of muscle fibers in pigs as revealed by a genome-wide association study. Animal 14:475–481. https://doi.org/10.3389/fgene.2019.00672

30. Gao Y et al (2017) CNV discovery for milk composition traits in dairy cattle using whole genome resequencing. BMC Genomics 18:265. https://doi.org/10.1186/s12864-017-3636-3

31. Zheng Z et al (2020) The origin of domestication genes in goats. Sci Adv 6:eaaz5216. https://doi.org/10.1126/sciadv.aaz5216

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