



Genome-wide genetic characterization and selection signatures in Anatolian Merino sheep

Taki Karsli

Department of Animal Science, Faculty of Agriculture, Eskişehir Osmangazi University,
Eskişehir, 26160, Türkiye

Correspondence: Taki Karsli (takikarsli@ogu.edu.tr)

Received: 14 October 2024 – Revised: 13 December 2024 – Accepted: 8 January 2025 – Published: 26 February 2025

Abstract. Molecular characterization and identification of selection signals at the genome-wide level facilitate the enhancement of ongoing conservation and selection studies in farm animals. This study aimed to reveal genomic diversity and selection signatures in Anatolian Merino sheep via 351 539 bi-allelic single-nucleotide polymorphisms (SNPs) obtained from the double-digest restriction-site-associated DNA sequencing (ddRAD-seq) technique. Genetic variability parameters such as minor-allele frequency (MAF), nucleotide diversity (π), observed heterozygosity (H_O), and expected heterozygosity (H_E) were estimated to be 0.340, 0.235, 0.258, and 0.235, respectively, while the inbreeding coefficient was 0.027 based on runs of homozygosity (ROH). A decreasing trend was detected in the effective population size of Anatolian Merino sheep, the current population of which turned out to be descendants of an ancestral population covering 2500 individuals 400 generations ago. Significant selection signals were detected in 464 SNPs within 14 genomic intervals via the ROH approach, whereas 259 SNPs were categorized into 79 genomic intervals by integrated haplotype score (iHS) statistics. A total of 37 and 72 protein-coding genes overlapped with detected genomic intervals in ROH and iHS approaches, respectively. A survey of a sheep QTL database confirmed that selection signals covered 66 QTL-associated SNPs. A large part of the protein-coding genes under selection pressure were mainly associated with milk production (*ROBO1*, *FSIP2*, *COBLL1*, *PTPN12*, *GSAP*, *CCDC146*, *FGL2*, *FAM185A*, *FBXL13*, *LOXL2*, *R3HCC1*, *CHMP7*, *RHOBTB2*, *PEBP4*, *SRGAP3*, and *RAD18*) and udder morphology (*SYT1*, *GFPT2*, *MAPK9*, and *RASGEF1C*), while numerous genes turned out to have effects on total muscle area (*PCDH7*), bone density (*SDK2*), carcass traits (*MBL2*), fecal egg count (*LGSN*), and immunoglobulin A level in blood circulation (*RFXAP*, *SERTM1*, *BEND7*, *PRPF18*, *FRMD4A*, and *GRAMD1B*). The results of this study confirm that high-density next-generation sequencing (NGS) data could be utilized to characterize local sheep breeds to shape conservation programs and shed light on the past breeding practices of the populations whose phenotypic records are absent.

1 Introduction

Thanks to their higher adaptability, small ruminants such as goats and sheep are reared to produce milk and meat across diverse climatic zones under different breeding systems (Demir et al., 2022). In Türkiye, grassland-based rearing not only allows for efficiently utilizing lands which are not proper for crop production but also supports the incomes of smallholder farmers in rural areas in order to sustain the production of animal-derived protein resources in a traditional way (Demir, 2024a). Sheep rearing is mainly centered

on native Turkish breeds in the country, covering 91 % of 45 million heads in total (Karsli, 2024), while the remaining breeds are believed to be Merino and other crossbreeds. After being introduced to Anatolia in 1839, Merino sheep were subjected to crossbreeding studies with several native Turkish sheep breeds such as Kıvrıcık and Akkaraman to improve economically important traits, including wool characteristics (Koyuncu, 2024). Crossbreeding practices between Kıvrıcık and Merino have led to the obtainment of a new breed called Karacabey Merino (Koyuncu, 2024), whereas the Anatolian Merino breed was developed via crossbreeding practices be-

tween Akkaraman and German Mutton Merino (Kizilaslan et al., 2024). Although Merino-derived breeds were initially reared to produce high-quality fleece for the textile sector during the 1950s, the loss in the economic value of wool production has forced farmers to focus on meat and milk production since then (Behrem and Gül, 2022). It is believed that this shift in breeder preferences has led to the fixation of certain gene regions associated with milk and meat yield in the genome of Anatolian Merino sheep, increasing their frequency. Identifying these gene regions (potential candidate genes) across the whole genome will provide a foundation for future genomic selection studies in this breed

Advances in molecular genetics have facilitated the evaluation of local sheep populations at the genome-wide level in terms of different aspects, including genetic characterization and selection signatures. Genetic characterization is an efficient approach for designing management and conservation programs in local sheep populations (Alnajm et al., 2021). Numerous molecular genotyping methods are available to estimate genetic diversity parameters, such as observed (H_O) and expected heterozygosity (H_E), as well as the inbreeding level, which allows one to maintain production in the local sheep population by periodically monitoring variability across the genome. Besides, genome-wide genetic variability gives significant clues with regard to prioritizing populations with regard to which low-variability populations to prioritize in conservation programs (Ben Sassi-Zaidy et al., 2022). Selection signatures, defined as regions with low genetic variability, are valuable indicators for the detection of genomic regions exposed to artificial- and natural-selection pressure. Defined as continuous homozygous segments occurring among a significant number of animals in a certain breed, runs of homozygosity (ROH) is a popular approach to detect genomic regions with lower variability resulting from factors such as genetic drift, bottlenecks, inbreeding, and selection (Peripolli et al., 2017). As articulated by Tiwari et al. (2024), the high proportion of long and short ROH segments indicates long-term and recent selection practices, respectively. The integrated haplotype score (iHS) approach, on the other hand, allows for the detection of recent selection signatures by scanning unusual haplotypes around a single-nucleotide polymorphism (SNP) compared to the whole genome (Voight et al., 2006). These signatures, detected by different approaches, show potential for providing insights into the past breeding experiences (genetic drift, bottleneck effect, geographic expansion, etc.) of local populations and the evolution of sheep genomes throughout history (Kim et al., 2016).

In theory, more reliable results are expected to be obtained with the denser genetic data used in bioinformatic analyses. In practice, on the other hand, using some high-resolution genotyping methods such as arrays of single-nucleotide polymorphisms (SNPs) and whole-genome sequencing (WGS) may not be practical or affordable for smallholders rearing local sheep breeds. Indeed, as articulated by Bilginer et

al. (2022), SNP arrays are biased due to being developed based on some reference breeds, which hinders the identification of novel or breed-specific genetic variations in local populations. On the contrary, WGS offers a way to detect almost all genomic variations across the local populations, but it remains expensive, and the sequencing cost per individual is not affordable in many cases (Pedrosa et al., 2021).

Over the last 2 decades, the discovery of next-generation sequencing (NGS) based on genomic library preparation methods has had a ground-breaking effect on the characterization of the local sheep population at a genome-wide level. Of these library preparation methods, double-digest restriction-site-associated DNA sequencing (ddRADseq) allows for the obtainment of millions of short DNA fragments belonging to different individuals which could be sequenced simultaneously (Peterson et al., 2012). The multiplexing process of the ddRADseq method not only decreases the cost of sequencing but also enables scientists to recover adequate genetic data with reduced representations of the genome, which eliminates the complexity of the whole genome. In contrast, ddRADseq includes several laboratory practices (barcode tagging, pooling, enrichment, etc.) which mainly limit the repetition and reproduction of the same genetic data. Besides, numerous sophisticated bioinformatic tools are required in each step (de-multiplexing, quality control, mapping, and filtering) to process raw sequence data into filtered SNP data. ddRADseq-based genetic data have been used for genetic characterization and/or for investigating the signatures of selection in several local species such as cattle (Demir et al., 2023a), sheep (Karsli, 2024), pigs (Vani et al., 2024), and fish (Magris et al., 2022).

By checking the literature, it was detected that the Anatolian Merino breed has been genotyped via microsatellite markers (Öner et al., 2014; Yilmaz et al., 2015; Ameer et al., 2020) for genetic characterization; however, no studies focusing on unraveling selection signatures are available. It is noteworthy that microsatellite markers are insufficient to reveal genome-wide genetic variability and selection signals due to the fact that they represent a small part of the genome, while NGS-based studies have recommended better genetic characterization of native Turkish livestock species (Karsli et al., 2020; Demir, 2024b). In this regard, this study aims, for the first time, to assess the genetic diversity and signatures of selection in Anatolian Merino sheep via genome-wide genetic data obtained from ddRAD sequencing.

2 Material and methods

2.1 Sampling and DNA isolation

By using pedigree information and oral interviews with animal keepers, a total of 20 unrelated animals belonging to the Anatolian Merino breed were sampled from three herds reared in the province of Eskişehir. In addition, phenotypic identification was conducted to validate the fact that these

animals represent the population. For this, 3 mL blood samples were collected from the jugular vein in Vacutainer tubes, with EDTA solution as an anticoagulant. Blood samples were stored at -20°C until DNA extraction was performed. By following the manufacturer's guidance, the GeneJET Genomic DNA Purification Kit (Thermo K0721) was deployed to isolate genomic DNA from the blood.

2.2 Preparation of genomic libraries and sequencing

As explained in detail by Peterson et al. (2012), the standard ddRADseq protocol was performed to create genomic libraries for the sequencing process. In brief, isolated DNA with the desired quantity and quality was digested with *EcoRI* and *MspI* restriction enzymes (New England Biolabs). AMPure XP beads (Beckman Coulter) were used to clean digested DNA fragments, which were further supplied with specific molecular barcodes and index primers via T4 DNA ligase. After the enrichment process by polymerase chain reaction (PCR), prepared DNA libraries were subjected to the Illumina NovaSeq 6000 instrument to obtain paired-end reads (2×150 base pair).

2.3 Raw data processing and variant calling

A methodology used by Demir et al. (2023b) was adopted to process raw short reads into filtered data. To do so, Stacks 2 software (Rochette et al., 2019) was run with supplied barcode and index information to assign each read to individuals. Fastp software (Chen et al., 2018) was run with default parameters for adapter removal and quality trimming. Clean reads were aligned to the *Ovis aries* (ARS-UI_v3.0) genome assembly via default parameters of the Burrows–Wheeler Aligner program (Li and Durbin, 2009). Aligned reads were converted into a Binary Alignment Map (BAM) file format using the SAMtools program (Danecek et al., 2021) with default parameters. The BCFtools (Danecek et al., 2021) pipeline was used to call variants in which all insertion–deletions (InDels) and SNPs not located on autosomal chromosomes were excluded. Bi-allelic SNPs passing specific criteria such as read depth ($20 \leq D \leq 500$) and base quality score ($Q \geq 20$) were kept in the data set. Furthermore, PLINK1.9 software (Chang et al., 2015) was utilized for filtering the data set in terms of SNPs ($-\text{maf } 0.05$ and $-\text{geno } 0.1$) and individuals ($-\text{mind } 0.1$) to call only informative variants with a high genotyping rate. In this study, approximately 12 million short reads were obtained from the sequencing analysis, in which a large part of the reads (98.5%) were mapped to the reference genome, with a coverage depth of nearly $30\times$. After raw data processing and filtering criteria, 351 539 bi-allelic SNPs were recovered across 20 animals.

2.4 Bioinformatic analyses

To assess genome-wide genetic variability in Anatolian Merino sheep, several statistical parameters such as minor-allele frequency (MAF), π , H_O , and H_E were calculated. Of these parameters, MAF, H_O , and H_E were estimated using PLINK 1.9 software (Chang et al., 2015), whereas VCFtools (Danecek et al., 2011) was deployed to calculate the π value. SNeP v.1.1 software (Barbato et al., 2015) was run with default parameters to assess historical changes in the effective population size (N_e) of Anatolian Merino sheep up to 400 generations ago. The distribution of N_e values per generation was visualized by the ggplot2 package (Villanueva and Chen, 2019), implemented in the R environment (<https://www.r-project.org>, last access: 9 January 2025). ROH was defined to estimate the genome-wide inbreeding (F_{ROH}) value. Two statistical models, namely ROH and integrated haplotype score (iHS), were chosen to detect genomic regions under selection pressure. By running the detectRUNs (Biscarini et al., 2018) package with a consecutive-run algorithm in the R environment (<https://www.r-project.org>). ROH was defined according to the criteria adopted by Demir (2024b) to reveal selection signatures in Kangal Akkaraman sheep. The iHS values per SNP were calculated via the rehh package (Gautier and Vitalis, 2012) implemented in the R environment (<https://www.r-project.org>) after the SNPs were phased in chromosome-wise by the Beagle 5.4 software (Browning et al., 2021).

A methodology reported by Argun Karsli et al. (2024) was followed to detect outlier SNPs and gene annotation. For this purpose, the Manhattan plots for ROH and iHS values were constructed via the qqman package (Turner, 2014) implemented in the R environment (<https://www.r-project.org>), in which only the top 0.1% of SNPs were considered to be under selection according to an empirical distribution. Furthermore, non-overlapping genomic intervals of 200 kb (100 kb upstream and 100 kb downstream) were screened to detect protein-coding genes via the Biomart platform (<https://www.ensembl.org/info/data/biomart/index.html>, last access: 9 January 2025). The effects of detected protein-coding genes on phenotype were validated via the sheep QTL database (<https://www.animalgenome.org/cgi-bin/QTLdb/OA/index>, last access: 9 January 2025).

3 Results

The average MAF and π values were estimated to be 0.340 and 0.235, respectively, in Anatolian Merino sheep. A higher observed heterozygosity (0.258) was detected compared to the average expected value (0.235), most probably due to the sampling strategy used, in which unrelated animals were chosen from different herds. The lowest and highest F_{ROH} values ranged from 0.018 in chromosome 14 to 0.047 in chromosome 6, with a mean of 0.027. The linkage-disequilibrium-based algorithm showed that effective popu-

lation size decreased from one generation to another. Indeed, the current population turned out to be represented by 537, 1075, and 2500 individuals 86, 174, and 400 generations ago, respectively (Fig. 1).

A total of 1345 ROH islands, of which 1342 were clustered into 0–2 Mb based on their physical length, were observed. Only three islands turned out to be clustered into 2–4 Mb according to their physical length. Using the ROH and iHS approaches, a total of, respectively, 464 and 259 SNPs turned out to be under selection pressure (Fig. 1). Detailed information about genes under selection and their effects on phenotype can be found in a summary in the Supplement.

In the ROH approach, detected SNPs covered 14 genomic intervals distributed over 8 chromosomes. These genomic intervals overlapped with 37 protein-coding genes, in which the lowest number of genes was observed in chromosome 7 (*TEX9*), while seven different genes were detected in chromosomes 1 (*BTBD8*, *C1orf146*, *GLMN*, *RPAP2*, *GF11*, *EVI5*, and *ROBO1*) and 15 (*IZUMO1R*, *GPR83*, *MRE11*, *ANKRD49*, *AASDHPPT*, *KBTBD3*, and *MSANTD4*). A survey based on the sheep QTL database revealed that selection signals detected by the ROH approach overlapped with 24 QTL-related SNPs which were associated with milk content (*ROBO1*, *FSIP2*, *COBLL1*, *PTPN12*, *GSAP*, *CCDC146*, *FGL2*, *FAM185A*, *FBXL13*, *EMCN*, *DDIT4L*, *H2AZ1*, and *DNAJB14*), total muscle area (*PCDH7*), udder morphology, and immunoglobulin A level in blood (*SYBU*, *EBAG9*, *PKHD1L1*, *ENY2*, *NUDCD1*, and *TRHR*).

Regarding the iHS approach, selection signatures were detected in all chromosomes except for chromosome 25. Of these chromosomes, 3, 4, 6, 8, 9, 11, 15, 17, 20, 22, and 26 contained single genes, while selection signals were detected in 11 genes (*GRIN3A*, *RNF20*, *PGAP4*, *ALDOB*, *LOXL2*, *R3HCC1*, *CHMP7*, *RHOBTB2*, *PEBP4*, *LGSN*, and *SCYGR4*) in chromosome 2. The highest number of SNPs (35) was detected between 0–200 000 intervals of chromosome 20, which turned out to be a non-coding region according to up-to-date reference genome assembly. Selection pressure was detected at 79 genomic intervals containing 72 protein-coding genes. A total of 42 QTL-related SNPs were identified across these genomic intervals. These SNPs were previously reported to be associated with milk content (*LOXL2*, *R3HCC1*, *CHMP7*, *RHOBTB2*, *PEBP4*, *SRGAP3*, and *RAD18*), fecal egg and milk bacterial count (*LGSN* and *RUNX2*), udder morphology (*SYT1*, *GFPT2*, *MAPK9*, and *RASGEF1C*), immunoglobulin A level in blood (*RFXAP*, *SERTM1*, *BEND7*, *PRPF18*, *FRMD4A*, and *GRAMD1B*), bone density (*SDK2*), carcass traits (*MBL2*), and resistance or susceptibility to maedi-visna virus (*DLGAP1*).

4 Discussion

Genetic diversity parameters in native Turkish sheep breeds have mainly been assessed via microsatellite markers for the

last 2 decades (Yilmaz et al., 2014; Ameer et al., 2020; Karsli et al., 2020; Kirikci et al., 2020), while SNP data obtained from NGS and array technologies are preferred to monitor genome-wide genetic variability nowadays (Ceccobelli et al., 2023; Bayraktar, 2024; Demir, 2024b; Karsli, 2024). Bayraktar (2024) reported higher observed heterozygosity (ranging from 0.348 to 0.356) compared to expected values (between 0.310 and 0.340) in three Anatolian sheep breeds known as Sakız, Norduz, and Karakaş, which were genotyped with the Illumina OvineSNP50 BeadChip. Inbreeding coefficient values, on the other hand, were declared to range from -0.063 (Sakız) to -0.019 (Karakaş) (Bayraktar, 2024). In another study, four native Turkish sheep breeds (Akkaraman, Güney Karaman, Karakaş, and Morkaraman) were genotyped via ddRADseq to identify genome-wide variability (Karsli, 2024). Similarly, observed heterozygosity (0.296–0.315) was reported to be higher than expected (0.287–0.299), together with negative values of the inbreeding coefficient (ranging from -0.060 to -0.034). The author declared that MAF values were between 0.311 in the Morkaraman breed and 0.316 in the Akkaraman and Güney Karaman breeds (Karsli, 2024). Similarly, a 50 K Ovine SNP chip-based study showed a higher observed heterozygosity (0.399) than expected (0.352) in Anatolian Merino sheep (Ceccobelli et al., 2023). On the contrary, another up-to-date study conducted by Demir (2024b) highlighted that the observed heterozygosity (0.290) was slightly lower than the expected heterozygosity (0.300), together with an MAF value of 0.320 in Kangal Akkaraman sheep, genotyped via a genotyping-by-sequencing (GBS) technique. The author indicated that the inbreeding coefficient was close to zero (0.01) in Kangal Akkaraman sheep according to the ROH-based inbreeding coefficient (F_{ROH}) (Demir, 2024b). Compared to Kangal Akkaraman sheep, a higher genomic inbreeding was detected in Anatolian Merino sheep ($F_{ROH} = 0.027$), which may require conservation programs to prevent potential inbreeding depression and loss of genetic diversity in the future. On the other hand, higher heterozygosity values were reported by SNP array technologies (Ceccobelli et al., 2023; Bayraktar, 2024) compared to in NGS-based studies (Demir, 2024b; Karsli, 2024). These inconsistent results may be explained by differences between molecular genotyping methods in which SNP arrays are designed to detect stable variations with higher variability. On the other hand, library preparation methods such as ddRADseq and GBS allow for sequencing random reads in which variations may be lower across the population. Still, this study shows consistency with previous studies in terms of showing higher heterozygosity than the expected value (Bayraktar, 2024; Karsli, 2024). However, it seems that the Anatolian Merino breed conserves lower heterozygosity and higher inbreeding compared to other local Turkish sheep breeds. This finding is not surprising due to the fact that the Turkish Merino breed has been derived from a small population and subjected to selection practices. These kinds of breeding practices have the

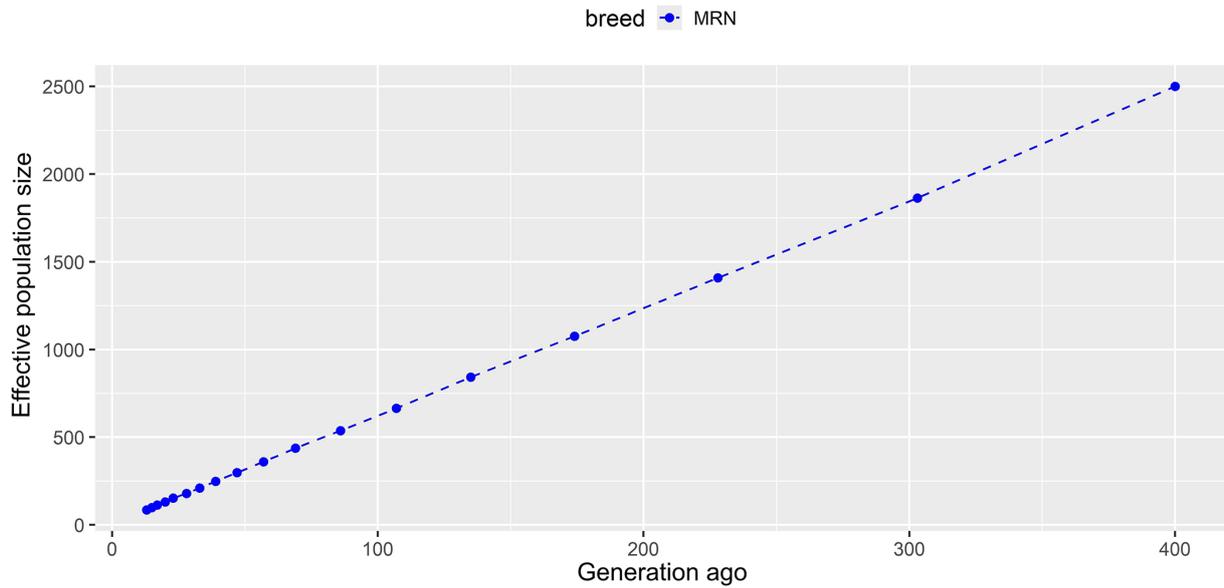


Figure 1. Historical changes in effective population size of Anatolian Merino sheep.

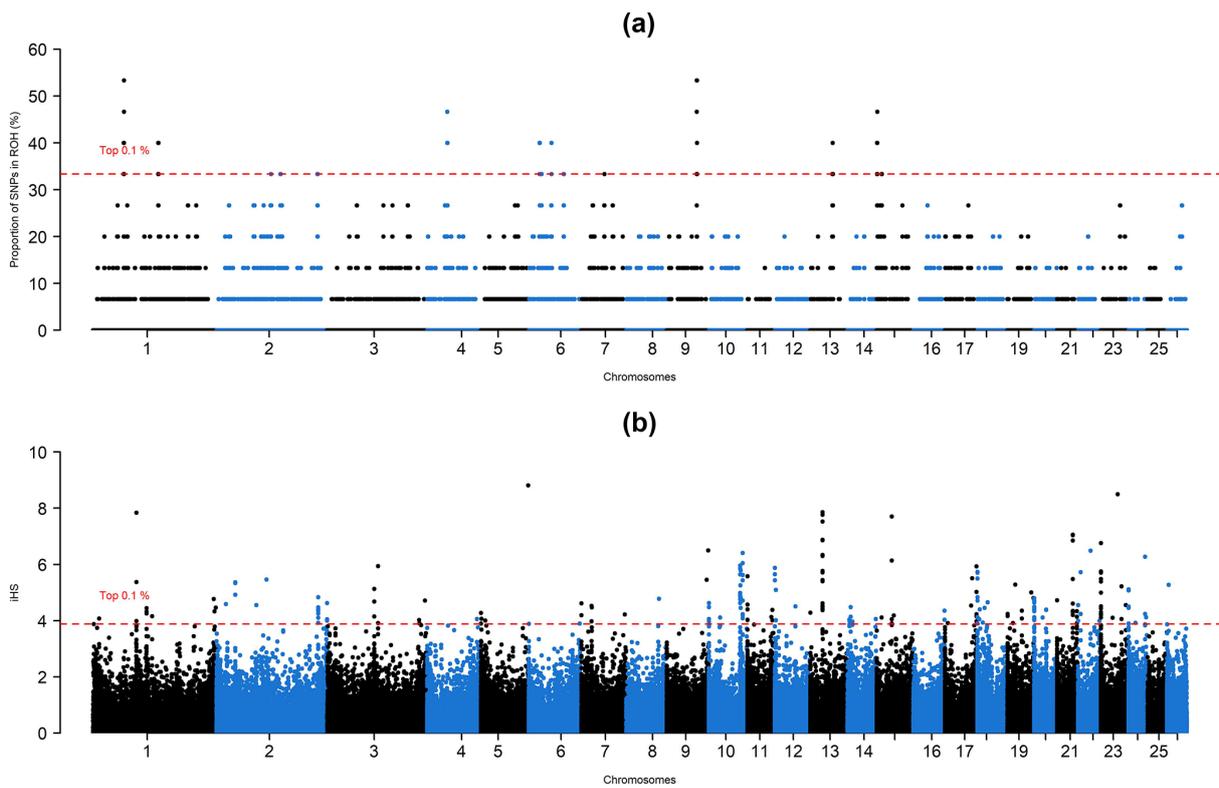


Figure 2. Manhattan plot for (a) percentage of ROH distribution and (b) iHS values per SNP.

potential to decrease heterozygosity and increase inbreeding at the genome-wide level (Karsli and Balcioğlu, 2019). A dramatic and continuous reduction was observed in effective population size, which was estimated to be 2500 individuals 400 generations ago, in Anatolian Merino sheep. As afore-

mentioned, the loss in the economic value of wool production over the last 6 decades, as well as lower adaptation to environmental challenges, may lead to a decrease in the effective population size of Anatolian Merino sheep. This decreasing trend was also reported by Ceccobelli et al. (2023) via 22 682

SNPs, in which the effective size of Anatolian Merino sheep was estimated to be 105 and 332 animals 13 and 50 generations ago, respectively. This trend has been declared for numerous local sheep breeds across the world (Ablondi et al., 2020; Zhang et al., 2020; Zinovieva et al., 2020; Liu et al., 2021). For example, a GBS-based study on 22 Kangal Akkaraman sheep confirmed that the actual population was descended from 978 animals 150 generations ago (Demir, 2024b).

In this study, a total of 37 and 72 protein-coding genes were identified to be under selection pressure in Anatolian Merino sheep via ROH and iHS approaches, respectively. Similarly, a lower number of genes were detected via ROH (32 genes) compared to via iHS (136 genes) in four Anatolian sheep breeds (Argun Karsli et al., 2024). Compared to other statistical approaches, a lower number of genes under selection pressure detected via ROH were also reported in several livestock species, including some native Indian sheep (Saravanan et al., 2021) and cosmopolitan beef cattle (Moravčíková et al., 2019). As highlighted by Argun Karsli et al. (2024), these differences may occur due to the different assumptions of various statistical models in which strict criteria allow for monitoring a lower number of genomic intervals in the ROH approach. In this study, no common selection signals across protein-coding genes were detected via either the ROH or iHS approach. As mentioned by Demir et al. (2023b), this could be observed since each method scans different signals under different statistical assumptions. A total of 28 protein-coding genes detected by the iHS approach in Turkish Merino sheep were also reported to be under selection pressure in some native Turkish sheep breeds (Argun Karsli et al., 2024). Among these common genes, *R3HCC1*, *CHMP7*, *RHOBTB2*, and *GRAMD1B* were reported to be under selection pressure in all studied sheep breeds (Argun Karsli et al., 2024). According to the sheep QTL database, these genes are associated with milk content (*R3HCC1*, *CHMP7*, *RHOBTB2*) and immunoglobulin A level in blood (*GRAMD1B*). Another gene observed in this study (*LGSN*) was also reported to be under selection pressure in six native Anatolian (Anatolian Black, East Anatolian Red, South Anatolian Yellow, South Anatolian Red, Turkish Grey Steppe, and Zavot) and two cosmopolitan (Holstein Friesian and Brown Swiss) cattle breeds reared in Türkiye via the iHS approach (Demir et al., 2023b). The authors highlighted that *LGSN* plays a key role in the visual modality in humans, while it is associated with fecal egg and milk bacterial count in sheep according to the QTL database. It is known that the downregulation of the *DLG1* protein may increase the infection rate of human immunodeficiency virus (HIV) in mammals (Perugi et al., 2009), while the *DLGAP1* gene plays a key role in maintaining the regulation of the *DLG1* protein. Indeed, White et al. (2012) aimed to investigate the genetic basis of ovine lentivirus in 964 animals belonging to Rambouillet, Polypay, and Columbia sheep breeds via a genome-wide association study in which the *DLGAP1*

gene turned out to be directly associated with resistance to ovine lentivirus. Significant selection signals overlapping with this gene were detected in Anatolian Merino sheep; this could be subjected to selection studies together with the *LGSN* gene to improve adaptation to environmental challenges.

5 Conclusions

In this study, genomic diversity and selection signals in Anatolian Merino sheep were assessed for the first time by 351 539 bi-allelic SNPs recovered from the ddRADseq technique combined with the Illumina Nova-Seq 6000 instrument. Compared to previous studies on native Turkish sheep breeds, a lower genetic variability and a higher inbreeding coefficient were detected in Anatolian Merino sheep. According to current findings, on the other hand, it can be stated that adequate genetic variability has been conserved at the genome-wide level, which is promising for sustaining current and future demands for economically important traits. However, the effective population size shows a declining trend per generation. This trend may negatively affect the agricultural sector in the future. Therefore, comprehensive and effective breeding programs should be adopted by farmers to prevent bottleneck effects in the future while periodically monitoring genetic variability to foresee future challenges. ROH and iHS approaches revealed significant selection signals in 109 protein-coding genes, most of which were directly linked to milk production. Since Merino sheep are commonly raised for wool production, selection signals are expected to occur in wool-related genes. However, as mentioned above, the declining demand for high-quality fleece in the textile sector since the 1950s has forced farmers to focus on meat and milk traits. Therefore, it is not surprising to identify selection signals in milk-related genes in Anatolian Merino sheep. This fact also confirms that selection signature analyses based on high-density SNP data are significantly efficient in revealing deeper information about the past breeding history of livestock species without requiring phenotypic records.

Data availability. The data set used in this study is available for scientific purposes only upon request via the Material Transfer Agreement (MTA) signed by the corresponding author.

Supplement. The supplement related to this article is available online at <https://doi.org/10.5194/aab-68-161-2025-supplement>.

Author contributions. TK designed the study, performed the bioinformatic analyses, and prepared and edited the paper.

Competing interests. The author has declared that there are no competing interests.

Ethical statement. This study has been reviewed and approved by the Eskişehir Osmangazi University Animal Experiments Local Ethic Committee (protocol no. HADYEK-998/2024).

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. The numerical calculations reported in this paper were fully/partially performed at TÜBİTAK ULAKBİM High Performance and Grid Computing Center (TRUBA resources).

Financial support. This research has been supported by the Scientific Research Projects Coordination Unit of Eskişehir Osmangazi University (grant no. FHD-2024-3165).

Review statement. This paper was edited by Henry Reyer and reviewed by Yasemin Öner and one anonymous referee.

References

Ablondi, M., Dadousis, C., Vasini, M., Eriksson, S., Mikko, S., and Sabbioni, A.: Genetic diversity and signatures of selection in a native Italian horse breed based on SNP data, *Animals*, 10, 1005, <https://doi.org/10.3390/ani10061005>, 2020.

Alnajm, H., Alijani, S., Javanmard, A., Rafat, S. A., and Hasanpur, K.: Genetic diversity analysis of four sheep breeds of Iran: Towards genetic maintenance and conservation decision, Iran, *J. Appl. Anim. Sci.*, 11, 527–538, 2021.

Ameur, A. A., Yilmaz, O., Nezih, A., Cemal, I., and Gaouar, S. B. S.: Assessment of genetic diversity of Turkish and Algerian native sheep breeds, *Acta Agric. Slov.*, 115, 5–14, <https://doi.org/10.14720/aas.2020.115.1.1229>, 2020.

Argun Karsli, B., Demir, E., Bilginer, U., Dogru, H., Karsli, T., and Kaya, S.: Genome-wide discovery of selection signatures in four Anatolian sheep breeds revealed by ddRADseq, *Sci. Rep.*, 14, 20518, <https://doi.org/10.1038/s41598-024-71617-7>, 2024.

Barbato, M., Orozco-terWengel, P., Tapio, M., and Bruford, M. W.: SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data, *Front. Genet.*, 6, 109, <https://doi.org/10.3389/fgene.2015.00109>, 2015.

Bayraktar, M.: Analysing the genetic diversity of three sheep breeds in Turkey and nearby countries using 50 K SNPs data, *Anim. Biotechnol.*, 35, 2329106, <https://doi.org/10.1080/10495398.2024.2329106>, 2024.

Behrem, S. and Gül, S.: Effects of age and body region on wool characteristics of Merino sheep crossbreeds in Turkey, *Turk. J. Vet. Anim. Sci.*, 46, 235–247, <https://doi.org/10.55730/1300-0128.4171>, 2022.

Ben Sassi-Zaidy, Y., Mohamed-Brahmi, A., Aloulou, R., Charfi-Cheikhrouha, F., Cendron, F., Tormen, N., and Cassandro, M.: Genetic characterization and alternative preservation ways of locally adapted sheep breeds: Cases of private and public sheep sectors in Tunisia and Italy, *Biology*, 11, 1623, <https://doi.org/10.3390/biology11111623>, 2022.

Bilginer, U., Ergin, M., Demir, E., Yolcu, H. I., and Karsli, B. A.: Detection of genetic diversity in cattle by microsatellite and SNP markers – a review, *Anim. Sci. Pap. Rep.*, 40, 375–392, 2022.

Biscarini, F., Cozzi, P., Gaspa, G., and Marras, G.: DetectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes, in: *Te Compr R Arch Network*, <https://rdrr.io/cran/detectRUNS/f/vignettes/detectRUNS.vignette.Rmd>, last access: 14 October 2018.

Browning, B. L., Tian, X., Zhou, Y., and Browning, S. R.: Fast two-stage phasing of large-scale sequence data, *Am. J. Hum. Genet.*, 108, 1880–1890, <https://doi.org/10.1016/j.ajhg.2021.08.005>, 2021.

Ceccobelli, S., Landi, V., Senczuk, G., Mastrangelo, S., Sardina, M. T., Ben-Jemaa, S., Persichilli, C., Karsli, T., Bâlteanu, V. A., Raschia, M. A., Poli, M. A., Ciappesoni, G., Muchadeyi, F.C., Dzomba, E. F., Kunene, N. W., Lühken, G., Deniskova, T. E., Dotsev, A. V., Zinovieva, N.A., Zsolnai, A., Anton, I., Kusza, S., Carolino, N., Santos-Silva, F., Kawecka, A., Świątek, M., Niżnikowski, R., Špehar, M., Anaya, G., Granero, A., Perloiro, T., Cardoso, P., Grande, S., de los Santos, B. L., Danchin-Burge, C., Pasquini, M., Martínez, A. M., Bermejo, J. V. D., Lasagna, E., Ciani, E., Sarti, F. M., and Pilla, F.: A comprehensive analysis of the genetic diversity and environmental adaptability in worldwide Merino and Merino-derived sheep breeds, *Genet. Sel. Evol.*, 55, 1–18, <https://doi.org/10.1186/s12711-023-00797-z>, 2023.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J.: Second-generation PLINK: rising to the challenge of larger and richer datasets, *GigaScience*, 4, 1–16, <https://doi.org/10.1186/s13742-015-0047-8>, 2015.

Chen, S., Zhou, Y., Chen, Y., and Gu, J.: Fastp: an ultra-fast all-in-one FASTQ preprocessor, *Bioinformatics*, 34, 884–890, <https://doi.org/10.1093/bioinformatics/bty560>, 2018.

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., McVean, G., and Durbin, R.: The variant call format and VCFtools, *Bioinformatics*, 27, 2156–2158, <https://doi.org/10.1093/bioinformatics/btr330>, 2011.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitham, A., Keane, T., McCarthy, S. A., Davies, R. M., and Li, H.: Twelve years of SAMtools and BCFtools, *GigaScience*, 10, giab008, <https://doi.org/10.1093/gigascience/giab008>, 2021.

Demir, E.: Genomic Diversity and Autozygosity-Based Signatures of Selection in Kagal Akkaraman Sheep via Genotyping-by-Sequencing, *Kafkas Univ. Vet. Fak.*, 30, 669–675, <https://doi.org/10.9775/kvfd.2024.32339>, 2024b.

Demir, E.: Microsatellite-based bottleneck analysis and migration events among four native Turkish goat breeds, *Arch. Anim. Breed.*, 67, 353–360, <https://doi.org/10.5194/aab-67-353-2024>, 2024a.

- Demir, E., Ceccobelli, S., Bilginer, U., Pasquini, M., Attard, G., and Karsli, T.: Conservation and selection of genes related to environmental adaptation in native small ruminant breeds: a review, *Ruminants*, 2, 255–270, <https://doi.org/10.3390/ruminants2020017>, 2022.
- Demir, E., Moravčíková, N., Kaya, S., Kasarda, R., Bilginer, Ü., Doğru, H., Balcioglu, M. S., and Karsli, T.: Genome-wide screening for selection signatures in native and cosmopolitan cattle breeds reared in Türkiye, *Anim. Genet.*, 54, 721–730, <https://doi.org/10.1111/age.13361>, 2023a.
- Demir, E., Moravčíková, N., Kaya, S., Kasarda, R., Doğru, H., Bilginer, Ü., Balcioglu, M. S., and Karsli, T.: Genome-wide genetic variation and population structure of native and cosmopolitan cattle breeds reared in Türkiye, *Anim. Biotechnol.*, 34, 3877–3886, <https://doi.org/10.1080/10495398.2023.2235600>, 2023b.
- Gautier, M. and Vitalis, R.: rehh: An R package to detect footprints of selection in genome-wide SNP data from haplotype structure, *Bioinformatics*, 28, 1176–1177, <https://doi.org/10.1093/bioinformatics/bts115>, 2012.
- Karsli, B. A., Demir, E., Fidan, H. G., and Karsli, T.: Assessment of genetic diversity and differentiation among four indigenous Turkish sheep breeds using microsatellites, *Arch. Anim. Breed.*, 63, 165–172, <https://doi.org/10.5194/aab-63-165-2020>, 2020.
- Karsli, B. A.: Genetic diversity and population structure of four Anatolian sheep revealed by genome-wide ddRADseq data, *Small Rum. Res.*, 235, 107285, <https://doi.org/10.1016/j.smallrumres.2024.107285>, 2024.
- Karsli, T. and Balcioglu, M. S.: Genetic characterization and population structure of six brown layer pure lines using microsatellite markers, *Asian-Australas. J. Anim. Sci.*, 32, 49–57, <https://doi.org/10.5713/ajas.17.0870>, 2019.
- Kim, E. S., Elbeltagy, A. R., Aboul-Naga, A. M., Rischkowsky, B., Sayre, B., Mwacharo, J. M., and Rothschild, M. F.: Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment, *Heredity*, 116, 255–264, <https://doi.org/10.1038/hdy.2015.94>, 2016.
- Kirikci, K., Cam, M. A., and Mercan, L.: Genetic diversity and relationship among indigenous Turkish Karayaka sheep subpopulations, *Arch. Anim. Breed.*, 63, 269–275, <https://doi.org/10.5194/aab-63-269-2020>, 2020.
- Kizilaslan, M., Arzik, Y., and Behrem, S.: Genetic parameters for ewe lifetime productivity traits in Central Anatolian Merino sheep, *Small Rum. Res.*, 233, 107235, <https://doi.org/10.1016/j.smallrumres.2024.107235>, 2024.
- Koyuncu, M.: Merino sheep and wool, *J. Anim. Prod.*, 65, 88–99, <https://doi.org/10.29185/hayuretim.1498665>, 2024.
- Li, H. and Durbin, R.: Fast and accurate short read alignment with Burrows–Wheeler transform, *Bioinformatics*, 25, 1754–1760, <https://doi.org/10.1093/bioinformatics/btp324>, 2009.
- Liu, J., Shi, L., Li, Y., Chen, L., Garrick, D., Wang, L., and Zhao, F.: Estimates of genomic inbreeding and identification of candidate regions that differ between Chinese indigenous sheep breeds, *J. Anim. Sci. Biotechnol.*, 12, 1–14, <https://doi.org/10.1186/s40104-021-00608-9>, 2021.
- Magris, G., Marroni, F., D’agaro, E., Vischi, M., Chiabà, C., Scaglione, D., Kijas, J., Messina, M., Tibaldi, E., and Morgante, M.: ddRAD-seq reveals the genetic structure and detects signals of selection in Italian brown trout, *Genet. Sel. Evol.*, 54, 8, <https://doi.org/10.1186/s12711-022-00698-7>, 2022.
- Moravčíková, N., Kasarda, R., Vostrý, L., Krupová, Z., Krupa, E., Lehocká, K., Olšanská, B., Trakovická, A., Nádaský, R., Židek, R., Belej, L., and Golian, J.: Analysis of selection signatures in the beef cattle genome, *Czech J. Anim. Sci.*, 64, 491–503, <https://doi.org/10.17221/226/2019-CJAS>, 2019.
- Öner, Y., Üstüner, H., Orman, A., Yılmaz, O., and Yılmaz, A.: Genetic diversity of Kıvrıkcık sheep breed reared in different regions and its relationship with other sheep breeds in Turkey, *Ital. J. Anim. Sci.*, 13, 3382, <https://doi.org/10.4081/ijas.2014.3382>, 2014.
- Pedrosa, V. B., Schenkel, F. S., Chen, S. Y., Oliveira, H. R., Casey, T. M., Melka, M. G., and Brito, L. F.: Genomewide association analyses of lactation persistency and milk production traits in Holstein cattle based on imputed whole-genome sequence data, *Genes*, 12, 1830, <https://doi.org/10.3390/genes12111830>, 2021.
- Peripolli, E., Munari, D. P., Silva, M. V. G. B., Lima, A. L. F., Irngang, R., and Baldi, F.: Runs of homozygosity: current knowledge and applications in livestock, *Anim. Genet.*, 48, 255–271, <https://doi.org/10.1111/age.12526>, 2017.
- Perugi, F., Muriaux, D., Ramirez, B. C., Chabani, S., Decroly, E., Darlix, J. L., Blot, V., and Pique, C.: Human Discs Large is a new negative regulator of human immunodeficiency virus-1 infectivity, *Mol. Biol. Cell*, 20, 498–508, <https://doi.org/10.1091/mbc.e08-02-0189>, 2009.
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E.: Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species, *PLoS one*, 7, e37135, <https://doi.org/10.1371/journal.pone.0037135>, 2012.
- Rochette, N. C., Rivera-Colón, A. G., and Catchen, J. M.: Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics, *Mol. Ecol.*, 28, 4737–4754, <https://doi.org/10.1111/mec.15253>, 2019.
- Saravanan, K. A., Panigrahi, M., Kumar, H., Bhushan, B., Dutt, T., and Mishra, B. P.: Genome-wide analysis of genetic diversity and selection signatures in three Indian sheep breeds, *Livest. Sci.*, 243, 104367, <https://doi.org/10.1016/j.livsci.2020.104367>, 2021.
- Turner, S. D.: qqman: An R package for visualizing GWAS results using QQ and manhattan plots, *Biorxiv* 005165, <https://doi.org/10.1101/005165>, 2014.
- Vani, A., Kumar, A., Mahala, S., Janga, S. C., Chauhan, A., Mehrotra, A., De, A. K., Sahu, A. R., Ahmad, S. F., Vempadapu, V., and Dutt, T.: Revelation of genetic diversity and genomic footprints of adaptation in Indian pig breeds, *Gene*, 893, 147950, <https://doi.org/10.1016/j.gene.2023.147950>, 2024.
- Tiwari, M., Gujar, G., Shashank, C. G., and Ponsuksili, S.: Selection signatures for high altitude adaptation in livestock: A review, *Gene*, 927, 148757, <https://doi.org/10.1016/j.gene.2024.148757>, 2024.
- Villanueva, R. A. M. and Chen, Z. J.: ggplot2: elegant graphics for data analysis, *Meas-Interdiscip. Res.*, 17, 160–167, <https://doi.org/10.1080/15366367.2019.1565254>, 2019.
- Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K.: A map of recent positive selection in the human genome, *PLoS Biol.*, 4, e154, <https://doi.org/10.1371/journal.pbio.0040072>, 2006.

- White, S. N., Mousel, M. R., Herrmann-Hoesing, L. M., Reynolds, J. O., Leymaster, K. A., Neibergs, H. L., Lewis, G. S., and Knowles, D. P.: Genome-wide association identifies multiple genomic regions associated with susceptibility to and control of ovine lentivirus, *PLoS One*, 7, e47829, <https://doi.org/10.1371/journal.pone.0047829>, 2012.
- Yilmaz, O., Cemal, I., and Karaca, O.: Genetic diversity in nine native Turkish sheep breeds based on microsatellite analysis, *Anim. Genet.*, 45, 604–608, <https://doi.org/10.1111/age.12173>, 2014.
- Yilmaz, O., Sezenler, T., Sevim, S., Cemal, I., Karaca, O., Yaman, Y., and Karadağ, O.: Genetic relationships among four Turkish sheep breeds using microsatellites, *Turk. J. Vet. Anim. Sci.*, 39, 576–582, <https://doi.org/10.3906/vet-1411-46>, 2015.
- Zhang, J., Nie, C., Li, X., Ning, Z., Chen, Y., Jia, Y., Han, J., Wang, L., Lv, X., Yang, W., and Qu, L.: Genome-wide population genetic analysis of commercial, indigenous, game, and wild chickens using 600 K SNP microarray data, *Front. Genet.*, 11, 543294, <https://doi.org/10.3389/fgene.2020.543294>, 2020.
- Zinovieva, N. A., Dotsev, A. V., Sermyagin, A. A., Deniskova, T. E., Abdelmanova, A. S., Kharzinova, V. R., Sölkner, J., Reyer, H., Wimmers, K., and Brem, G.: Selection signatures in two oldest Russian native cattle breeds revealed using high-density single nucleotide polymorphism analysis, *PLoS One*, 15, e0242200, <https://doi.org/10.1371/journal.pone.0242200>, 2020.