

ACCEPTED AUTHOR VERSION OF THE MANUSCRIPT:

Genome-wide assessment of genetic diversity, population structure and selection signatures in Polish mountain sheep breeds using SNP microarrays

DOI: 10.2478/aoas-2026-0009

Anna Miksza-Cybulska^{1♦}, Aldona Kawęcka¹, Artur Gurgul²

¹Department of Sheep and Goat Breeding, National Research Institute of Animal Production, 32-083 Balice n. Kraków, Poland

²Center for Experimental and Innovative Medicine, University of Agriculture in Krakow, Rędzina 1c, 30-248 Kraków, Poland

♦Corresponding author: anna.miksza@iz.edu.pl

Received date: 27 October 2025

Accepted date: 5 January 2026

To cite this article: (2026). Miksza-Cybulska A., Kawęcka A., Gurgul A. (2026). Genome-wide assessment of genetic diversity, population structure and selection signatures in Polish mountain sheep breeds using SNP microarrays, Annals of Animal Science, DOI: 10.2478/aoas-2026-0009

This is unedited PDF of peer-reviewed and accepted manuscript. Copyediting, typesetting, and review of the manuscript may affect the content, so this provisional version can differ from the final version.

Genome-wide assessment of genetic diversity, population structure and selection signatures in Polish mountain sheep breeds using SNP microarrays*

Anna Miksza-Cybulska¹*, Aldona Kawęcka¹, Artur Gurgul²

¹Department of Sheep and Goat Breeding, National Research Institute of Animal Production, 32-083 Balice n. Kraków, Poland

²Center for Experimental and Innovative Medicine, University of Agriculture in Krakow, Rędzina 1c, 30-248 Kraków, Poland

*Corresponding author: anna.miksza@iz.edu.pl

Abstract

The study aimed to characterize the genetic structure with the use of SNP microarrays of three mountain sheep breeds: the Podhale Zackel (PZ), the Polish Mountain Sheep (PMS) and the Polish Colored Mountain Sheep (CMS), including the assessment of the size, nature and biological significance of genetic differences between them and to assess the usefulness of the obtained data in the implementation of conservation programs for populations at risk. The analyses were based on the data from OvineSNP50 BeadChip microarray (Illumina). The level of heterozygosity was similar for all analyzed breeds, and the value of the inbreeding coefficient of the studied populations was low. Comparing the genetic distances based on F_{ST} between the studied breeds, the highest values were noted between the PMS and CMS, and the lowest between the PZ and the PMS. Detailed studies of the genetic differentiation across the genome revealed the presence of differential selection signals. In addition, a common genome region located on chromosome 4 under selective pressure was detected, in which two genes were found (*ISPD* – a gene associated with the development of the nervous system, and *SOSTDC1* – a gene associated with the morphogenesis of hair follicles). The possibility of obtaining comprehensive information on the genetic diversity of the studied populations indicates the usefulness of SNP microarrays in determining the genetic variability of native sheep breeds.

Key words: mountain sheep, genetic diversity, microarrays, selection signatures

*This study was funded by the project: “Directions for the use and conservation of farm animal genetic resources in sustainable development”, co-financed by the National Centre for Research and Development under the Strategic Research and Development Programme “Environment, Agriculture and Forestry” – BIOSTRATEG, Decision No. BIOSTRATEG2/297267/14/NCBR/2016.

The tradition of mountain sheepherding in Poland has a centuries-old history. Mountain sheep are an integral element of the pastoral economy, not only being the basis of subsistence for the local community of the Polish Carpathians, but also because of their role in highlander culture and shaping the mountain landscape. They provide milk for the production of traditional products, wool, leather, and delicious meat. The breeds are perfectly adapted to the harsh climatic conditions and terrain of the mountains. Mountain sheep currently include three breeds: the Podhale Zackel, the Polish Mountain Sheep, and the Colored Mountain Sheep.

The sheep of the Polish Carpathians descend from the Vlach Zackel, which the Vlachs brought with them in the 14th century (Kłapyta, 2014), leaving a lasting mark on the economic

and cultural life of the inhabitants of the Polish Carpathians. From the Vlachs, the Podhale people adopted almost everything related to mountain pastoralism: the system of organizing grazing, along with all the associated rituals, the technology of processing sheep's milk and the accompanying nomenclature (Figure 1). The Podhale sheep had a white, mixed woolly coat, with dark spots on the muzzle, ears, and around the eyes. Black and dark brown individuals also occurred in the population. These sheep were used for a wide range of purposes, with the wool being used for the production of cloth, felt, and outerwear, while the milk was processed into traditional cheeses during their grazing in mountain pastures (Kawęcka and Krupiński, 2014). Several attempts were made to improve the sheep's utility by introducing various sheep breeds. After World War II, the Podhale Zackel was refined using Transylvanian Zackel imported from Romania and Friesian rams. As a result, sheep with increased body mass, higher milk yield, and a modified coat character were obtained. The wool became finer, its yield doubled, and the outer coat hair lengthened at the expense of the medullary hair. The refinement process resulted in a new breed, and the entire group of mountain sheep, including the Zackel, was named the Polish Mountain Sheep.

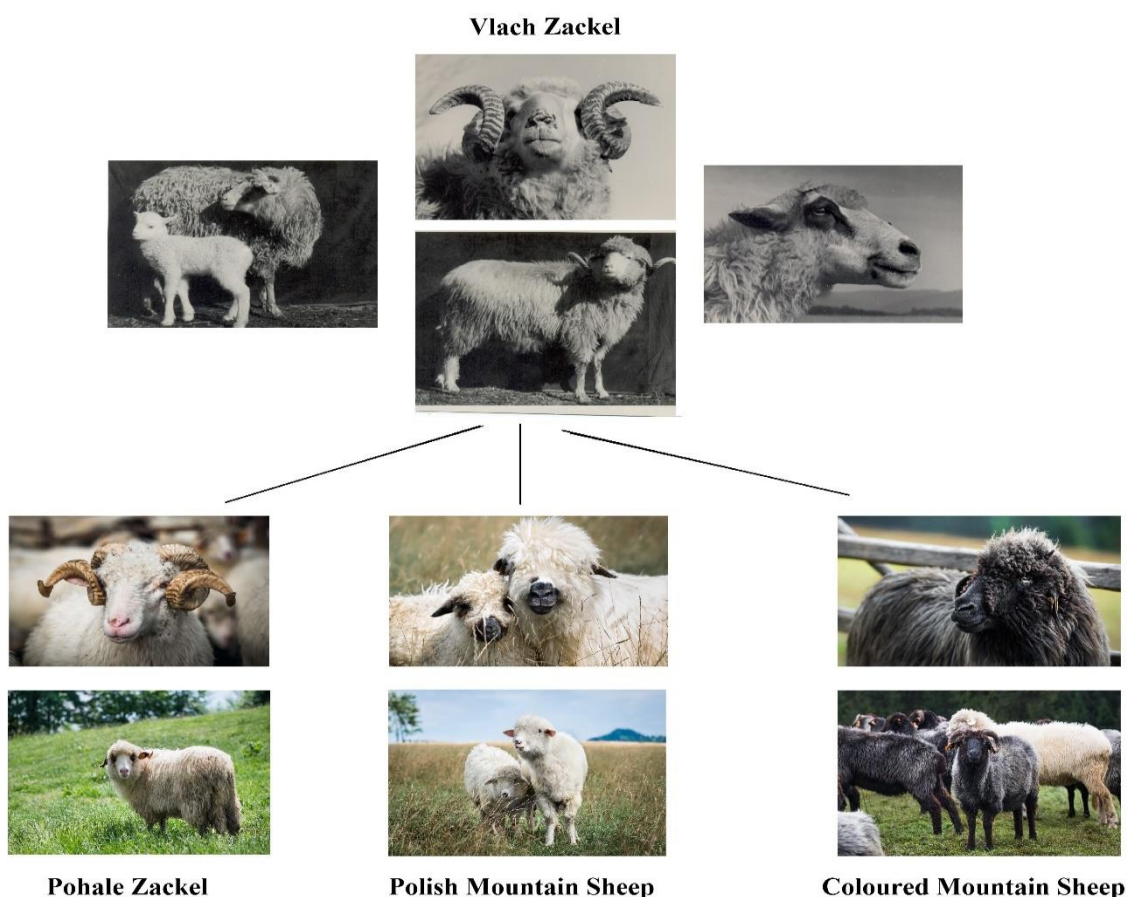


Figure 1. The origin of three mountain sheep breeds in Poland from the original Vlach Zackel (photos by W. Puchalski, M. Pasternak-Choros, A. Kawęcka)

The idea of preserving biodiversity and valuable resources of native breeds, each with its unique characteristics, has changed the direction of breeding efforts in mountain sheep farming. This was driven by the characteristics of this native population, such as excellent adaptation to local conditions, disease resistance, and a strong herding instinct, so important in traditional grazing. Work has been undertaken to restore the original Podhale Zackel breed by selecting individuals phenotypically consistent with the breed standard from the existing Polish

mountain sheep population. Work on separating the colored sheep from the Polish mountain sheep population began 25 years ago, by selecting individuals from Polish mountain sheep flocks based on their exterior, color, and coat characteristics (Polak et al., 2020). Given the high utility value of mountain sheep, their excellent adaptation to the environmental conditions of the Polish Carpathians, and their special importance for the landscape and heritage of these regions, this breed has been included in a genetic resources conservation program, as an expression of special concern for their preservation for future generations. Currently, 17 native breeds in Poland are included in the conservation program (Kawęcka et al., 2022).

An important element of the conservation and management of farm animal genetic resources is the estimation of diversity and analysis of the genetic structure of populations. Among the most important molecular tools used in farm animal research in recent decades are mitochondrial haplotype analysis, microsatellite sequence analysis, and single-nucleotide polymorphism (SNP) analysis. Currently, BeadChip microarrays (Illumina) are in widespread use for various animal species (Kawęcka et al., 2016). Recent genomic studies in sheep using similar approaches have provided valuable insights into genetic diversity, population structure, and selection patterns (Talebi et al., 2025). The aim of this study was to characterize genome-wide genetic diversity, population structure and signatures of selection in three Polish mountain sheep breeds (Podhale Zackel, Polish Mountain Sheep and Colored Mountain Sheep) using SNP microarray data, and to evaluate their implications for genetic resource conservation.

Material and methods

The study included 300 ewes belonging to three native breeds of mountain sheep: the Podhale Zackel (PZ), the Polish Mountain Sheep (PMS), and the Colored Mountain Sheep (CMS). The experimental material consisted of blood collected from randomly selected ewes from two different flocks in each breed group (Figure 2):

- PZ – 50 samples in Pieniny Spiskie and 50 samples in Bieszczady Mountains,
- PMS – 38 samples in Pieniny Spiskie and 65 samples in Pieniny Czorszyńskie,
- CMS – 50 samples in Pieniny Spiskie and 47 samples in Podhale.

Related individuals were not excluded before further analyses.

A 5 ml portion of blood was collected from the sheep in the morning before milking by a veterinarian, with the approval of the Ethical Committee (Local Ethical Committee No. II in Kraków – approval no. 1293/2016).

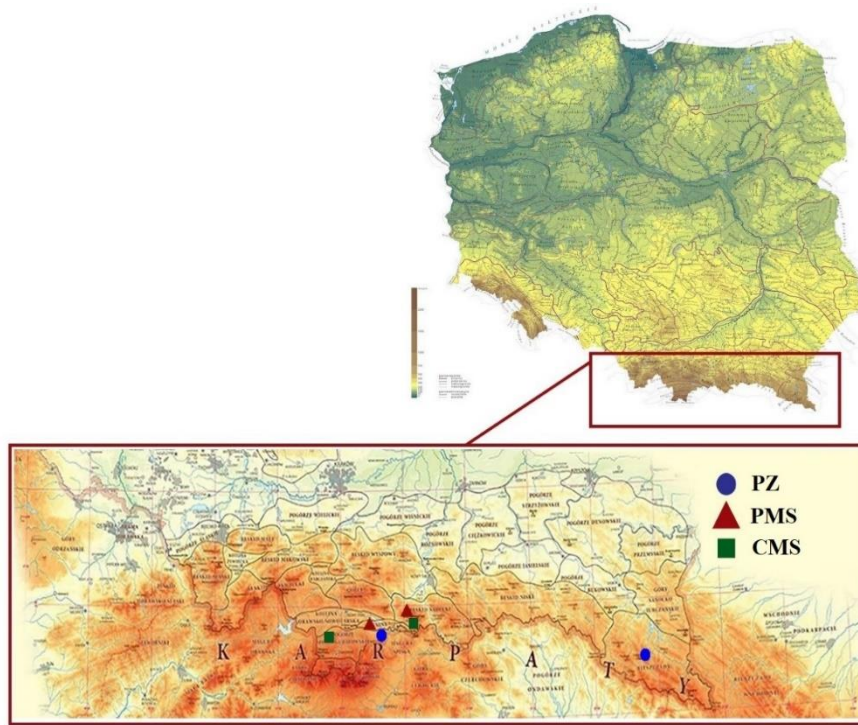


Figure 2. Map of the Polish Carpathians with the locations where the material was collected from the examined sheep

Genomic DNA was isolated from blood using the QuickGene DNA whole blood kit S (Kurabo, Tokyo, Japan) according to the manufacturer's protocol. After DNA concentration and quality control, samples were normalized and genotyped using the OvineSNP50 BeadChip microarray (Illumina, San Diego, CA, USA) and the HiSanSQ system (Illumina). Genotypes were determined using GenomeStudio software (Illumina). SNP filtering was performed before the actual calculations. The genotyped SNP positions were remapped to the ISCG Oar_v3.1 genome assembly. SNPs located on the Y chromosome and unmapped SNPs were removed. SNP markers with undetermined genomic positions, those with more than 10% missing genotypes, and monomorphic markers across all breeds (i.e., MAF <0.01) were excluded. Finally, SNP markers were tested for deviations from Hardy-Weinberg equilibrium in each breed, and as a consequence, markers that deviated very strongly from equilibrium ($P < 0.0001$) were rejected, which may indicate errors in the genotyping process. The presented results are after filtering.

The indicators of genetic variation considered in this study were the distribution of minor allele frequencies (MAF), mean expected heterozygosity (H_E) and mean observed heterozygosity (H_O), as well as the mean value of the inbreeding coefficient F_{IS} calculated according to Wright's method (Wright, 1949) according to the formula:

$$F_{IS} = 1 - \frac{H_O}{H_E}$$

To visualize the differentiation between mountain sheep breeds and the variation within populations, PLINK v1.90b5.3 64-bit software (Purcell, 2007) was used to generate the first three components of a principal components analysis (PCA).

Effective population size was determined for all studied mountain sheep breeds based on linkage disequilibrium between markers using NeEstimatorv2 software (Do et al., 2014). To reduce computational time and to minimize the impact of SNP linkage, analyses were performed based on a panel of randomly selected 1,000 polymorphic SNPs for each studied breed. The narrow confidence intervals obtained with the Jackknife approach indicate that our

NeN_eNe estimates are stable and reliable. Furthermore, the estimates are highly consistent with those reported in our previous study, where NeN_eNe was calculated using LD-based methods and the full SNP dataset (Jasielczuk et al., 2023).

To identify genomic regions showing signs of differential selection between breeds, Wright's F-statistic (F_{ST}) analyses were performed, assessing the inbreeding parameters of a given subpopulation compared to the entire population (distance between populations), using PLINK v1.90b5.3. For each SNP, the F_{ST} value was calculated, obtained by pairwise comparisons between breeds, i.e., PZvsPMS, PZvsCMS, and PMSvsCMS (Wright, 1949; Akey et al., 2002). The F_{ST} value was then standardized according to methodology of Wright (1949) and Akey et al. (2002), for each breed using the formula:

$$d_i = \frac{F_{ST}^{ij} - E[F_{ST}^{ij}]}{sd[F_{ST}^{ij}]}$$

where:

$E[F_{ST}^{ij}]$ – average F_{ST} ,

$sd[F_{ST}^{ij}]$ – standard deviation for F_{ST} .

0.1% top d_i values were considered as significant signals of differential selection.

Quality-filtered genotypes were additionally subjected to phasing using default parameters in FASTPHASE software (Scheet and Stephens, 2006). Next, for each breed, positive selection signatures were identified using the REHH test using SWEEP v.1.1 software (Sabeti et al., 2002). This procedure yielded a dataset that was then analyzed by calculating the number of haplotype blocks for which haplotype frequencies were greater than 0.25 and for which the probability of the statistical test for REHH was lower than the significance threshold of $P = 0.05$. All identified genomic regions potentially being under selection pressure were analyzed for their distribution in the genome, the genes they encode, and their functional significance for the selected traits. To this end, we first used the Ensembl database (<https://genome.ucsc.edu/cgi-bin/hgTables>) to identify genes encoded by the indicated regions. Subsequently, the resulting genes were analyzed for their function using the online bioinformatics tool DAVID Bioinformatics Resources 6.8, NIAID/NIH. This tool also allowed us to associate the annotated genes with their corresponding metabolic pathways and to test the statistical significance of the enrichment of individual pathways by the detected genes (Huang et al., 2009).

Results

A total of 47,592 SNP markers were used for the analyses. Missing genotypes were observed for 3,281 SNPs, with an average missing genotype count of 0.0198. The mean minor allele frequencies (MAF) for the whole Mountain Sheep group was 0.300 (± 0.124), with the highest mean value observed on chromosome 21 (0.309 ± 0.121). The distribution of the minor allele frequencies for Mountain Sheep is presented in a histogram (Figure 3). The highest mean frequency of the minor allele was observed for the Podhale Zackel (0.302), while for the PMS and CMS, these values were lower and similar (0.297 and 0.298, respectively). The highest values of mean expected and observed heterozygosity were noted for the Podhale sheep, while the F_{IS} value for all mountain sheep breeds was below zero.

The mean expected heterozygosity of the marker panel was 0.390 (± 0.109), and the mean observed heterozygosity was 0.388 (± 0.111).

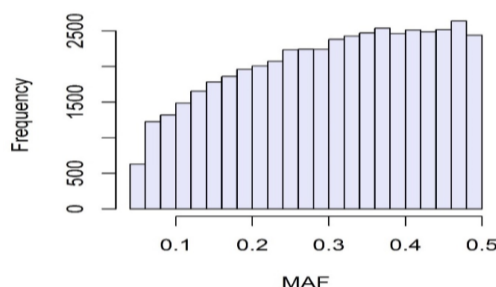


Figure 3. Frequency distribution of the rarer allele for tested sheep

Significant ($P < 0.05$) deviations from Hardy-Weinberg equilibrium were observed for 2,749 markers, and 718 markers showed highly significant ($P < 0.01$) deviations. The highest number of markers significantly and highly significantly deviating from the Hardy-Weinberg equilibrium was observed for PMS (2207 and 479 markers, respectively). The mean values of the genetic variation indices for the breeds are summarized in Table 1.

Table 1. Genetic variability indexes for the Mountain Sheep subpopulations

Breed	MAF	H _O	H _E	F _{IS}	HWE P<0.05	HWE P<0.01
PZ	0.302 (± 0.122)	0.397 (± 0.116)	0.392 (± 0.107)	-0.0125	1696	322
PMS	0.297 (± 0.124)	0.391 (± 0.119)	0.386 (± 0.110)	-0.0122	2207	479
CMS	0.298 (± 0.124)	0.393 (± 0.119)	0.388 (± 0.110)	-0.0137	1790	346
Mean to all breeds	0.299	0.394	0.389	-0.0128	1898	401

Effective population size was determined for the sheep breeds studied, and significant variation was observed among them. Individual breeds had effective population sizes ranging from 64 to 103 individuals. The results are summarized in Table 2.

Table 2. Effective population size (N_e) and 95% confidence interval of N_e for the studied sheep breeds

Breed	N _e	95% CI down	95% CI up
PZ	94.2	92.7	95.7
PMS	103	101.3	104.7
CMS	64.6	63.8	65.4

Principal component analysis (PCA) was used to graphically illustrate the interbreed and intrabreed variability of Mountain Sheep. For each individual, the values of the first three components of variability were determined (PC1, PC2, and PC3).

Figure 4 presents a visualization of the data for PC1 and PC2 in mountain sheep. It allows for the observation of three distinct clusters, representing the breed lines studied. The

figure for the PC1 and PC3 data for mountain sheep, similarly to the first two components (PC1 and PC2), shows a group of poorly differentiated individuals in the middle section. In this layout, the internal genetic variability of the Polish mountain sheep group is even more pronounced.

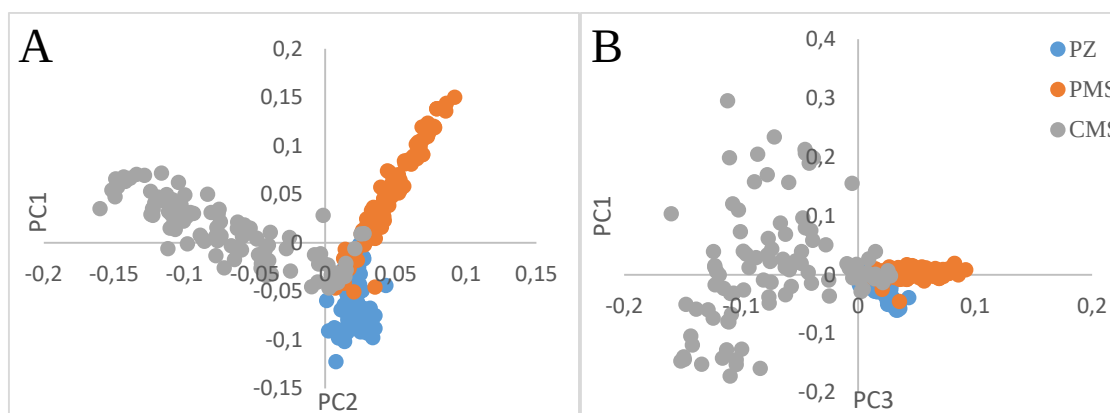


Figure 2. Visualization of genetic differentiation based on PCA - A - PC1 vs. PC2 components for mountain sheep; B - PC1 vs. PC3 components for mountain sheep

The highest values of mean and weighted mean genetic distance (F_{ST}) between breeds were recorded between PMS and CMS (0.0236 and 0.0254), and the lowest between PZ and PMS (0.0142 and 0.0150). The mean and weighted mean F_{ST} values are summarized in Table 3.

Table 3. Average (above the diagonal) and weighted average (below the diagonal) F_{ST} distances among the studied breeds

Breed	PZ	PMS	CMS
PZ	-----	0.0142364	0.0197141
PMS	0.014995	-----	0.0236004
CMS	0.0208261	0.0254135	-----

Based on the F_{ST} values for individual SNPs and the mean F_{ST} values for individual breed pairs, standardized F_{ST} values (d_i) were determined. In individual mountain sheep breeds, 11 to 17 genomic regions with strong F_{ST} -based selection signals were detected, ranging in size from 305.9 kbp to 1.5 Mbp. The most selection signals were detected in the Podhale Zackel, and the least in the Polish Mountain Sheep. The most signatures were noted on chromosome 2 in the Podhale Zackel and the Polish Mountain Sheep, while in the Colored Mountain Sheep on chromosome 14. No signals of differential selection were detected on chromosomes 10–12, 15, 17–22, and 25–28. Three signatures occupying common genomic regions were identified for the Podhale Zackel and the Polish Mountain Sheep, located on chromosomes 2, 3, and 16, and one signal located on chromosome 4, was common to all breeds. The visualization of selection signals in the form of Manhattan plots for the studied breeds is shown in Figure 5. The dashed line indicates the top 0.1% of d_i values in each breed. This corresponds to the top 0.1% of d_i values from the empirical distribution. This threshold was selected based on previous literature using the same cutoff (Pérez O'Brien et al., 2014; Purfield

et al., 2017; Lukic et al., 2023) and was additionally motivated by the need to obtain a number of significant results that can be reasonably interpreted and analyzed.

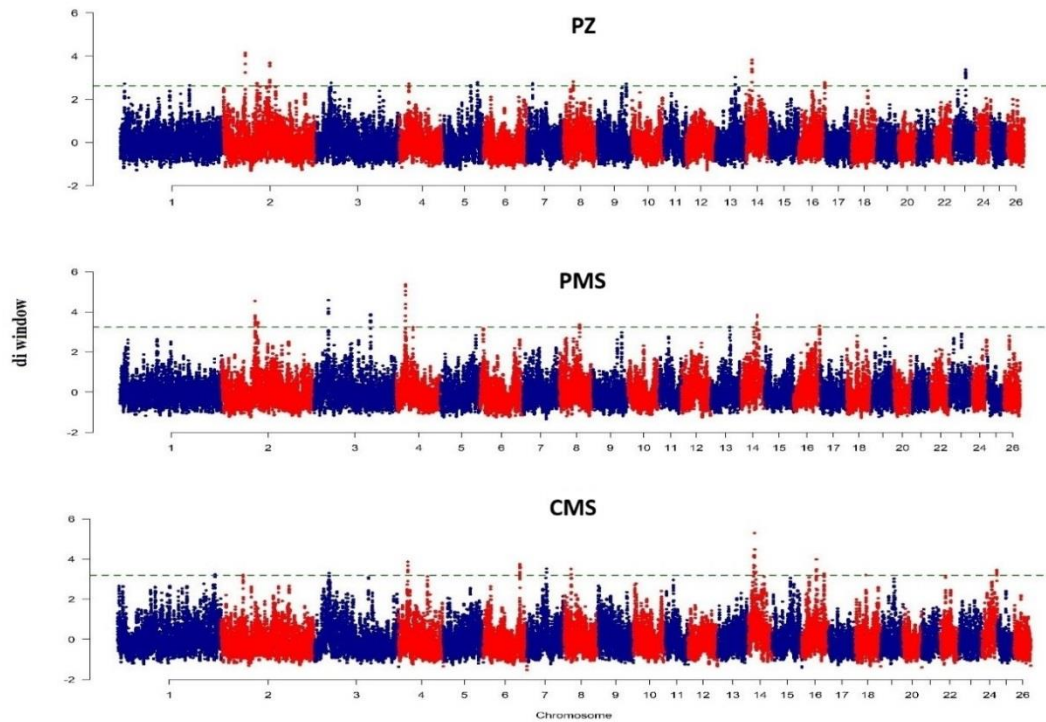


Figure 5. Di values in all analyzed autosomes of mountain sheep

Annotation of the obtained regions under selection using the Ensembl database identified 92 genes for the Podhale Zackel, 90 for the Polish Mountain Sheep, and 100 for the Colored Mountain Sheep. From this pool, 56 genes with known function for ZN, 55 for PMS, and 59 for CMS were used to analyze metabolic pathways using the DAVID 6.8 online tool. Of these, two genes were common to PZ and PMS (*C1D*, *DMRTA1*), four were common to PZ and CMS (*GNAQ*, *LONP2*, *N4BP1*, *SIAH1*), and two were common to all three breeds (*ISPD*, *SOSTDC1*). Functional classification of well-annotated genes showed that genes detected in regions under selective pressure were enriched in processes related to cellular responses, regulation of kinase and enzyme activity, and ATP binding. The most statistically significantly enriched processes were observed for the Podhale Zackel. For the PZ and CMS, one process (ATP binding) and a group of three genes common to both breeds (*ABCC11*, *ABCC12*, and *LONP2*) were observed. No statistically significantly enriched processes were observed for the PMS.

It was also found that all breeds share a common selection signal located in a region of the chromosome 4. Annotation of this region in the Ensembl database identified two genes, namely *ISPD* (Isoprenoid synthase domain containing; Chr 4 Exon 10, ENSOARG00000008410, Sequence: NC_056057.1) and *SOSTDC1* (Sclerostin domain containing 1; Chr 4 Exon 2, ENSOARG00000008418, Sequence: NC_056057.1).

Analyses conducted using Sweep v.1.1 software identified 1,708 (PZ), 1,901 (PMS), and 1,963 (CMS) core haplotype blocks (CRs), with an average length of 89.4 (58.5) kbp for PZ, 95.9 (127.2) for CMS and 96.3 (125.9) kbp for PMS. For all haplotype blocks, from 13,338 (PZ) to 15,752 (CMS) REHH tests were performed. Tests with haplotype frequencies lower than 0.25 were then removed from the results, resulting in 5,606 (PZ) to 6,501 (CMS) core haplotypes. From the remaining tests, those that were statistically significant, i.e., $P < 0.05$ (representing the strongest selection signals), were selected. The resulting REHH p values were

logarithmized and plotted against chromosomal position to visualize the core haplotypes and patterns of selection in the genome for each breed (summarized in Figure 6, where the dashed line represents the significance threshold of 0.05). There were between 279 (PZ) and 326 (CMS) significant haplotype cores (Table 4). After combining haplotypes spanning the same cores, 210 haplotype blocks remained for PZ, 215 for PMS, and 216 for CMS. The lengths of the haplotype cores, their numbers, and the other values described above are collected in the supplementary materials.

Table 4. Characteristics of haplotype cores for PZ, PMS and CMS

Specification	PZ	PMS	CMS
Number of haplotype cores	1708	1901	1963
Average haplotype core length (SD)	89.4 (58.5) kbp	95.9 (127.2) kbp	96.3 (125.9) kbp
Number of REHH tests	13338	14818	15752
Number of haplotypes more frequent than 0.25	5606	6329	6501
Number of cores for which haplotype frequency >0.25 and $P < 0.05$	279	326	321
Number of unique haplotypes	210	215	216

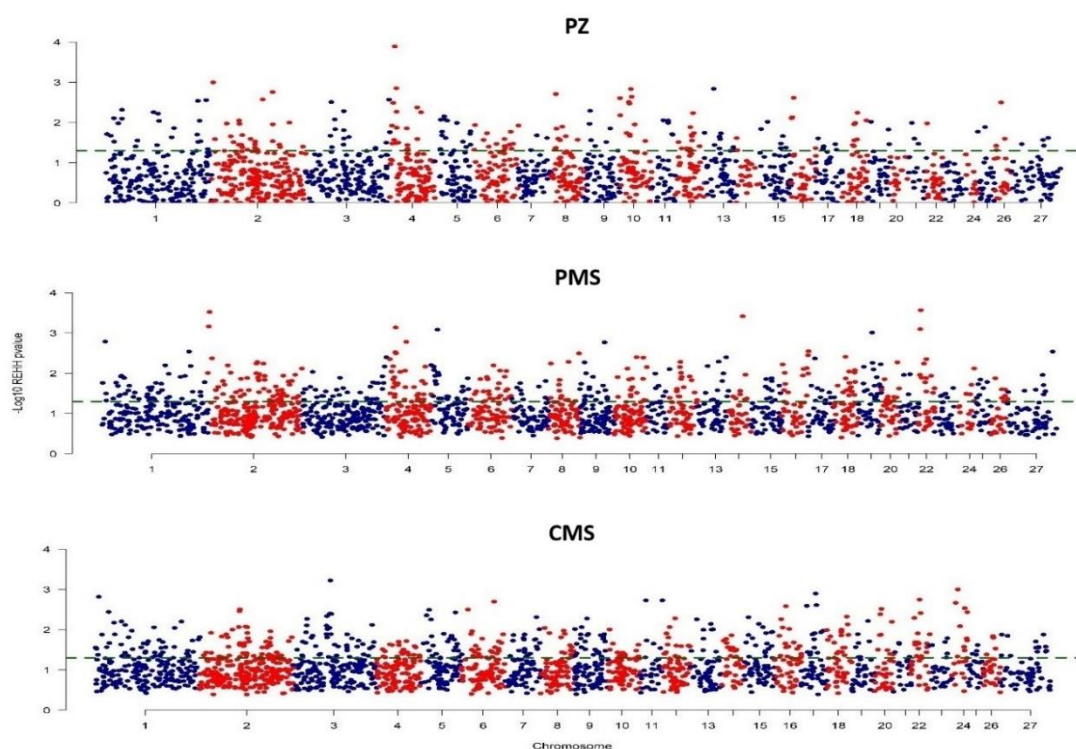


Figure 6. Genomic map of REHH probability values for all core haplotypes showing frequencies >0.25 for mountain sheep.

The green dashed line indicates a REHH probability level of 0.05

The genomic positions of statistically significant core haplotypes were used to identify genes potentially associated with the identified selection signals. The procedure was performed similarly to that for differential selection, using the ENSEMBL database for gene identification. This resulted in 197 genes for PZ, 329 for PMS, and 264 for CMS. The DAVID was then used

to determine the functions of the identified genes. For the Podhale Zackel, the program identified 197 genes with known function, 329 genes for the Polish Mountain Sheep, and 156 genes for the Colored Mountain Sheep. In the Podhale Zackel, only one process controlled by a group of 17 genes was statistically significantly enriched, and it was related to ATP binding, similar to the results obtained from the FST analysis. For the Polish Mountain Sheep, five processes involving 22 genes were detected, while for the Colored Mountain Sheep, seven processes involving 47 different genes were identified. For the Podhale Zackel and the Colored Mountain Sheep, the same process was identified: ATP binding, for which two genes (*PAN3* and *NUBPL*) were common to both breeds.

Discussion

Effective management of livestock genetic resources requires comprehensive knowledge of breeds, including data on population size and structure, geographic distribution, production environment, and genetic diversity within and between breeds. Combining different types of data allows for a more comprehensive assessment of biodiversity within and between breeds, thus facilitating effective resource management. The genetic structure of a population is described by a number of parameters, such as the number of alleles, heterozygosity, inbreeding, variation between populations, and so on. Populations may differ in one or more characteristics that describe their genetic structure. Genetic structure parameters can indicate the population's fitness and trends in change (Groeneveld et al., 2010).

Heterozygosity is an important indicator for assessing the genetic variability of domestic animals. Heterozygosity reflects the genetic potential and adaptability of a population to its natural environment (Zhang et al., 2018). Deniskova et al. (2018), analyzing the genotype of 25 Russian local breeds using the OvineSNP50 BeadChip microarray, found heterozygosity values ranging from 0.354 to 0.359, comparable to those observed in other sheep breeds worldwide. Ciani et al. (2014), using the OvineSNP50 BeadChip, analyzed the genetic structure of 19 Italian breeds. The authors found distinct breeds inhabiting Sardinia, which are a consequence of both sporadic introgressions of wild mouflon in ancient times and the island's long-term genetic isolation from mainland sheep. The authors presented arguments that challenge, from a genomic perspective, the current classification of Bergamasca and Biellese sheep into two distinct breeds. Expected heterozygosity in the Italian breeds ranged from 0.33 to 0.37. These values were consistent with those observed by Kijas et al. (2012) for southern and western European sheep, analyzed using the IlluminaOvineSNP50BeadChip. Their study reported observed heterozygosity of 0.33 to 0.38, with the exception of the Macarthur Merino, for which the value was 0.22. The authors suggest that, in sheep, breed heterozygosity revealed a weak association with increasing distance from the "center of domestication". One likely explanation is the widespread use of Merino rams in Europe, which began after the Middle Ages, resulting in the widespread dissemination of haplotypes between Merino and other breeds. In our own studies, heterozygosity values for mountain sheep ranged from 0.393 to 0.397. These values were comparable to those obtained in European sheep and higher than those in Asian sheep (Zhang et al., 2018; Deniskova et al., 2018; Ciani et al., 2014; Kijas et al., 2012). As suggested by the authors of the studies cited above, this may indicate a higher level of domestication of Polish native mountain sheep breeds.

Abied et al. (2020) presented analyses conducted on five Chinese native sheep breeds. They used high-resolution 600K SNP microarrays. The calculations showed that the FIS inbreeding coefficient ranged from -0.03 for Karakul sheep to -0.004 for Wadi sheep. Our own research did not reveal a risk of increased inbreeding indices in any of the studied populations (FIS ranging from -0.0122 to -0.0137). This may indicate proper breeding practices and appropriate ram selection.

There is a close relationship between the inbreeding growth rate (ΔF) and the effective population size (N_e), as $\Delta F = 1/2N_e$, where ΔF represents the inbreeding growth rate and N_e represents the effective population size (Filistowicz, 2015). The effective population size (N_e) is one of the most important parameters considered in analyzing the condition of a conserved population and assessing its threat status. In Germany, using only this indicator, the following thresholds were adopted for assessing the threat status of breeds: $N_e \leq 50$ – only phenotypic protection is possible; $50 < N_e \leq 200$ – the population requires protective measures; $200 < N_e \leq 1000$ – the population requires monitoring, $N_e > 1000$ – the population is not threatened (Krupiński and Polak, 2018). The effective population size in the study by Deniskova et al. (2019) varied for individual sheep groups, ranging from $N_e = 128$ for the Kyrgyz wire-haired sheep to $N_e = 660$ for the Gissar sheep. The authors suggest that even the lowest values obtained were higher than in previous studies by our own and other authors, or consistent with observations reported by the SheepHapMap project (Kijas et al., 2012). They suspect that the larger effective population sizes likely correspond to the widespread and high popularity of these studied breeds in Central Asian countries. According to Deniskova et al. (2019), the low effective population size values of the Kyrgyz Alai and Tien-Shan breeds were likely due to a limited founder population, as both breeds were created and bred within scientifically developed breeding programs.

Preliminary results of molecular studies using SNP microarrays comparing effective population size between Polish breeds covered by genetic resources conservation programs were presented by Gurgul et al. (2019). They showed variation in this parameter between species. The highest mean N_e values were found in horses and cattle (97 and 94, respectively), relatively low in pigs (46), and intermediate in sheep (60). The results obtained in our study showed significant variation within the mountain sheep population itself, from the lowest $N_e = 64.6$ for the Colored Mountain Sheep to $N_e = 103$ for the Polish Mountain Sheep.

Visualization of genetic diversity using the PCA method, based on the genetic profiles of individuals, revealed a clear division of the studied population into three clusters, corresponding to separate breeds. The analysis revealed that all breeds had distinct genetic structures, and individuals within each breed clustered closely together. However, some individuals were found in the area overlapping for all three populations. For some individuals from each breed, significant similarity to other breeds was observed. This may indicate a close relationship between the studied individuals, currently belonging to separate breeds but historically originating from the same mountain sheep population. These data likely reflect the common phylogenetic origin of all three mountain sheep breeds. Based on the PCA results, it can also be concluded that the Podhale Zackel represents the most homogeneous genetic group, while the Colored Mountain Sheep represents the most internally diverse group.

Talebi et al. (2025) investigated the genomic consequences of recent, targeted crossbreeding in Moghani sheep by comparing purebred animals with Texel- and Booroola-derived paternal backcrosses, using genotyping-by-sequencing to assess population structure, genetic differentiation, and inbreeding. PCA clearly separated purebred Moghani sheep from all crossbred groups, indicating strong genomic restructuring due to introgression, while pairwise F_{ST} values (approximately 0.007–0.018) confirmed increased differentiation between pure and crossbred populations, particularly in Texel-derived lines. Although crossbred lambs showed higher numbers of polymorphic variants, they also exhibited longer runs of homozygosity and higher ROH-based inbreeding coefficients, indirectly suggesting reduced effective population sizes and potential loss of adaptive resilience. F_{ST} was interpreted together with ROH and heterozygosity to show that introgression did not uniformly increase genetic health, as higher differentiation coincided with increased genomic inbreeding in some crosses. In contrast, our study on native Polish mountain sheep breeds within a conservation framework using SNP microarrays, where PCA revealed three distinct but partially overlapping breed

clusters reflecting shared historical ancestry rather than recent admixture, and F_{ST} values (approximately 0.014–0.025) indicated moderate differentiation typical of closely related breeds. Importantly, we directly estimated effective population size, obtaining N_e values between about 64 and 103, which, together with low inbreeding coefficients, supported the demographic stability of these conserved populations. Overall, while both studies applied similar population-genetic tools, Talebi et al. (2025) emphasized short-term genomic changes and trade-offs arising from crossbreeding for productivity, whereas we focused on long-term genetic structure, differentiation, and demographic viability in the context of breed conservation.

It is well known that selection increases the frequency of particular alleles at different genomic loci in a selected population and creates unique genetic patterns in their DNA sequence. Several methods are available to scan the genome for selective signatures by assessing patterns of variation within and between breeds. Among them, F_{ST} -based signatures and extended haplotype homozygosity (EHH) are a reliable approach for detecting genomic regions subject to relatively recent selective pressure (Bomba et al., 2015).

A comparative study of selection signatures in two species was conducted by Kim et al. (2016), who analyzed 50K BeadChip microarray data from indigenous sheep and goat populations from the hot and arid environment of Egypt, where natural selection played a major role. Genomic regions encompassing 119 genes were identified, most of which were responsible for metabolism. Notable signatures included genes influencing traits related to adaptation to unfavorable environmental conditions. Among the identified genes were genes responsible for thermoregulation (*FGF2*, *PLCB1*), digestion (*MYH*, *TRHDE*), body size (*BMP2*, *GJB3*), and immune response (*IL2*, *GRIA1*). The authors also observed shared signatures between sheep and goats, suggesting shared selection in these two species, which are ultimately closely related. Furthermore, sheep and goats are useful as model organisms for mapping selection signatures and genomic linkages.

Abied et al. (2020) analyzed population structure and selection signatures among five indigenous Chinese sheep breeds (using high-density 600K SNP genotypes) selected from different geographic locations with extremely dry or humid conditions. Genomic regions under selection identified by F_{ST} and XP EHH methods frequently overlapped across breeds and included genes associated with adaptation to extremely dry or humid environments, innate and adaptive immune responses, and traits related to growth, wool, milk, and reproduction.

In another previous work, Eydivandi et al. (2021) analyzed 34,206 SNP markers from 450 individuals of 15 sheep breeds, including six indigenous Middle Eastern sheep breeds. Using both F_{ST} and REHH statistical analysis, 629 regions and 256 genes were detected as being under selection, respectively. These regions encoded genes identified as influencing economic traits, namely *CIDEA*, *HHATL*, *MGST1*, *FADS1*, *RTL1*, and *DGKG*. Both F_{ST} and REHH identified 60 common genes as selection signatures. Among them, four candidate genes were identified: *NT5E*, *ADA2*, *C8A*, and *C8B*, which were enriched in two significant functional categories from the Gene Ontology (GO) database, related to adenosine metabolism. The authors emphasize the high value of the acquired knowledge about genomic regions subject to selection pressure, which will enable future insights into the role of the selected genes in local adaptation in the studied populations.

Using F_{ST} statistics, Gurgul et al. (2021) identified genomic loci with the greatest genetic divergence between seven analyzed breeds of Polish native sheep: Blackface, Colored Merino, Old Type Merino, Polish Mountain Sheep, Świniarka, Uhruska, and Wrzosówka. The identified genomic regions with the strongest selection signals encoded 781 distinct genes. The *RXFP2* gene locus, responsible for horn size/type in sheep, was identified as the most prominent selective signal.

The results of Yurchenko et al. (2019) showed that the genomes of 15 Russian sheep breeds contain numerous regions likely subject to selection pressure. More than half of these regions contained well-known candidate genes related to morphology, adaptation, and domestication (*KITLG*, *KIT*, *MITF*, and *MC1R*), wool quality (*DSG*, *DSC*, and *KRT*), reproduction (*CMTM6*, *HTRA1*, *GNAQ*, *UBQLN1*, and *IFT88*), and milk production (*ABCG2*, *SPP1*, *ACSS1*, and *ACSS2*). Furthermore, the authors observed selection signatures in gene regions whose function in the human genome is associated with sensory neuropathies and genetic disorders affecting the inability to sense pain and temperature. This may indicate a possible mechanism for adaptation to harsh climatic conditions.

In the mountain sheep analyzed in this study, a region under selective pressure common to all breeds was detected. Among the genes located in this region are *ISPD* and *SOSTDC1*. Although the function of *ISPD* in mammals is unknown, mutations in this gene lead to reduced functional glycosylation of dystroglycan, which is a common cause of various forms of dystroglycanopathy (Cirak et al. 2013). Gouveia et al. (2017), in studies conducted on local Spanish sheep breeds, also detected the selection signal spanning the *ISPD* gene. They suggest a role for the *ISPD* gene in nervous system development. The *SOSTDC1* gene, in turn, encodes a sclerostin-containing protein, an alternative protein for the differentiation of follicular helper cells and regulatory follicular T cells during acute viral infection (Hu et al. 2019). This allows us to assume that the presence of this gene in the selection signature region has a positive effect on resistance to viral infections.

Närhi et al. (2012) investigated the effect of the *SOSTDC1* gene on modulating the hair follicle and mammary gland morphogenesis pathway in mice. Mice with an inactivated *SOSTDC1* gene were used to investigate the role of this gene in modulating the bone morphogenetic protein (BMP) signaling pathway and the WNT family of proteins, which control the initiation and formation of skin appendages. They observed that the *SOSTDC1* gene did not affect the hair growth cycle in the skin but limited the number of developing hair follicles and whiskers on the face of mice. Furthermore, this gene influences the size and shape of the mammary glands, and its knockout promotes the development of supernumerary mammary nipples. According to Närhi, *SOSTDC1* can be hypothesized to attenuate signaling of the WNT/ β -catenin pathway. Nie et al. (2018) conducted research on tissues collected from fetal Tibetan sheep to investigate the influence of long non-coding RNAs (lncRNAs) and mRNAs on the primary induction of wool follicles in sheep. RNA sequencing revealed signals in the WNT, BMP, and FGF (fibroblast growth factor) protein pathways. GO and KEGG analyses revealed enriched biological processes of morphogenesis and transforming growth factor- β , as well as intercellular matrix-receptor interactions. mRNA and lncRNA interaction networks revealed transcripts likely involved in the induction of primary wool follicles. Among the genes identified as hair follicle marker genes, *EDAR*, *FOXI3*, *FGF20*, *SHH*, and *SOSTDC1* showed increased expression when primary wool follicles were initiated in sheep.

Findings and conclusions

This study characterized the genetic structure of Polish native sheep breeds, classified as mountain sheep, using SNP microarrays. It was found that the studied populations have distinct genetic structures, with a certain pool of individuals sharing a similar genetic profile. Native mountain sheep breeds were characterized by similar levels of heterozygosity. The lowest effective population size was observed in CMS, which is the least favorable from the point of view of genetic diversity. The highest genetic distance was observed in PMS vs. CMS, and the lowest in PZ vs. PMS. The low inbreeding coefficient indicates a lack of inbreeding risk. In genomic regions with strong signals of differential selection, 92 genes were identified for the Podhale Zackel, 90 for the Polish Mountain Sheep, and 100 for the Colored Mountain Sheep. Two genes (*ISPD* – nervous system development and *SOSTDC1* – hair follicle

morphogenesis) were found in the common selection signal on OAR 4 across the breeds. Overall, SNP microarrays are an effective tool for assessing genetic variation and structure, which is definitely useful in genetic resource conservation programs, and may help maintain biodiversity and manage native populations, also by conscious selection of individuals for breeding in order to preserve the desired characteristics, such as adaptation to unfavorable environmental conditions.

Nevertheless, it must be acknowledged that the conducted studies were subject to certain limitations, such as sample size and SNP panel density.

References

- Abied A., Bagadi A., Bordbar F., Pu Y., Augustino S.M., Xue X., Xing F., Gebreselassie G., Han J.-L., Mwacharo J.M., Ma Y., Zhao Q. (2020). Genomic diversity, population structure, and signature of selection in five Chinese native sheep breeds adapted to extreme environments. *Genes*, 11: 494.
- Akey J.M., Zhang G., Zhang K., Jin L., Shriver M.D. (2002). Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.*, 12: 1805–1814.
- Bomba L., Nicolazzi E.L., Milanesi M., Negrini R., Mancini G., Biscarini F., Stella A., Valentini A., Ajmone-Marsan P. (2015). Relative extended haplotype homozygosity signals across breeds reveal dairy and beef specific signatures of selection. *Genet. Sel. Evol.*, 47: 25.
- Ciani E., Crepaldi P., Nicoloso L., Lasagna E., Sarti F.M., Moioli B., Napolitano F., Carta A., Usai G., D'Andrea M., Marletta D., Ciampolini R., Riggio V., Occidente M., Matassino D., Kompan D., Modesto P., Macciotta N., Ajmone-Marsan P., Pilla F. (2014). Genome-wide analysis of Italian sheep diversity reveals a strong geographic pattern and cryptic relationships between breeds. *Anim. Genet.*, 45: 256–266.
- Cirak S., Foley A.R., Herrmann R., Willer T., Yau S., Stevens E., Torelli S., Brodd L., Kamynina A., Vondracek P., Roper H., Longman C., Korinthenberg R., Marrosu G., Nürnberg P., UK10K Consortium, Michele D.E., Plagnol V., Hurles M., Moore S.A., Sewry C.A., Campbell K.P., Voit T., Muntoni F. (2013). *ISPD* gene mutations are a common cause of congenital and limb-girdle muscular dystrophies. *Brain*, 136: 269–281.
- Deniskova T., Dotsev A., Selionova M.I., Kunz E., Medugorac I., Reyer H., Wimmers K., Barbato M., Traspov A.A., Brem G., Zinovieva N.A. (2018). Population structure and genetic diversity of 25 Russian sheep breeds based on whole-genome genotyping. *Genet. Sel. Evol.*, 50: 29.
- Deniskova T., Dotsev A., Lushihina E., Shakhin A., Kunz E., Medugorac I., Reyer H., Wimmers K., Khayatzadeh N., Sölkner J., Sermyagin A., Zhunushev A., Brem G., Zinovieva N. (2019). Population structure and genetic diversity of sheep breeds in the Kyrgyzstan. *Front. Genet.*, 12: 1311.
- Do C., Waples R.S., Peel D., Macbeth G.M., Tillett B.J., Ovenden J.R. (2014). Ne Estimator v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Mol. Ecol. Res.*, 14: 209–214.
- Eydivandi S., Roudbar M.A., Ardestani S.S., Momen M., Sahana G. (2021). A selection signatures study among Middle Eastern and European sheep breeds. *J. Anim. Breed. Genet.*, 138: 574–588.
- Filistowicz A. (2015). Inbred w doskonałych i chronionych populacjach zwierząt gospodarskich (in Polish). *Prz. Hod.*, 2: 3–6.
- Gouveia J.J.S., Paiva S.R., McManus C.M., Caetano A.R., Kijas J.W., Facó O., Azevedo H.C., de Araujo A.M., de Souza C.J.H., Yamagishi M.E.B., Carneiro P.L.S., Lôbo R.N.B., de

- Oliveira S.M.P., da Silva M.V.G.B. (2017). Genome-wide search for signatures of selection in three major Brazilian locally adapted sheep breeds. *Livest. Sci.*, 197: 36–45.
- Groeneveld L.F., Lenstra J.A., Eding H., Toro M.A., Scherf B., Pilling D., Negrini R., Finlay E.K., Jianlin H., Groeneveld E., Weigend S., GLOBALDIV Consortium. (2010). Genetic diversity in farm animals – a review. *Anim Genet.*, 41, Suppl., 1: 6–31.
- Gurgul A., Jasielczuk I., Ropka-Molik K., Piestrzyńska-Kajtoch A., Szmatoła T., Bugno-Poniewierska M., Krupiński J. (2019). Ocena zmienności genetycznej oraz genetycznej podatności na wybrane jednostki chorobowe polskich ras zachowawczych oraz innych małych populacji zwierząt gospodarskich utrzymywanych w Polsce w oparciu o klasyczne i wielkoskalowe badania genetyczne. Monografia “Rasy rodzime zwierząt gospodarskich w badaniach projektu Biostrateg II w latach 2016–2019” (in Polish), Instytut Zootechniki Państwowy Instytut Badawczy, pp. 83–131.
- Gurgul A., Jasielczuk I., Miksza-Cybulska A., Kawęcka A., Szmatoła T., Krupiński J. (2021). Evaluation of genetic differentiation and genome-wide selection signatures in Polish local sheep breeds. *Lives. Sci.*, 251: 104635.
- Hu J., Wu J., Li Y., Wang Z., Tang J., Li Z., Hu L., Huang Q., Ye L., Xu L. (2019). Sclerostin domain-containing protein 1 is dispensable for the differentiation of follicular helper and follicular regulatory T cells during acute viral infection. *Am. J. Transl. Res.*, 11: 3722–3736.
- Huang D., Sherman B., Lempicki R. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, 4: 44–57.
- Jasielczuk I., Szmatoła T., Miksza-Cybulska A., Kawęcka A., Gurgul A. (2023). Linkage disequilibrium, historical effective population size and haplotype blocks in selected native sheep breeds. *Small Rumin. Res.*, 228: 107109.
- Kawęcka A., Krupiński J. (2014). Sheep in the Polish Carpathians: genetic resources conservation of the Podhale Zackel and Coloured Mountain Sheep. *Geomatics, Landmanagement and Landscape*, 1: 35–45.
- Kawęcka A., Gurgul A., Miksza-Cybulska A. (2016). The use of SNP microarrays for biodiversity studies of sheep – a review. *Ann. Anim. Sci.*, 16: 975–987.
- Kawęcka A., Pasternak M., Miksza-Cybulska A., Puchała M. (2022). Native sheep breeds in Poland – importance and outcomes of genetic resources protection programmes. *Animals*, 12: 1510.
- Kijas J.W., Lenstra J.A., Hayes B., Boitard S., Porto Neto L.R., San Cristobal M., Servin B., McCulloch R., Whan V., Gietzen K., Paiva S., Barendse W., Ciani E., Raadsma H., McEwan J., Dalrymple B., International Sheep Genomics Consortium Members (2012). Genome-wide analysis of the world’s sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol.*, DOI: 10(2): e1001258.
- Kim E.-S., Elbeltagy A.R., Aboul-Naga A.M., Rischkowsky B., Sayre B., Mwacharo J.M., Rothschild M.F. (2016). Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity*, 116: 255–264.
- Kłapyta P. (2014). Wołoskie osadnictwo w Karpatach w aspekcie historyczno-geograficznym (in Polish). *Proc. Conf.: Kultura pasterska łuku Karpat i jej oddziaływania na kulturę Babiogórców*. Zawoja, September, 2014.
- Krupiński J., Polak G. (2018). Conservation of livestock biodiversity in conditions of sustainable agriculture (in Polish). *Prz. Hod.*, 5.
- Lukic B., Curik I., Drzaic I., Galić V., Shihabi M., Vostry L., Cubric-Curik V. (2023). Genomic signatures of selection, local adaptation and production type characterisation of East Adriatic sheep breeds. *J. Anim. Sci. Biotechnol.*, 14: 142.

- Närhi K., Tummers M., Ahtiainen L., Itoh N., Thesleff I., Mikkola M. (2012). Sostdc1 defines the size and number of skin appendage placodes. *Develop. Biol.*, 364: 149–161.
- Nie Y., Li S., Zheng X., Chen W., Li X., Liu Z., Hu Y., Qiao H., Qi Q., Pei Q., Cai D., Yu M., Mou C. (2018). Transcriptome reveals long non-coding RNAs and mRNAs involved in primary wool follicle induction in carpet sheep fetal skin. *Front. Physiol.*, 9: 446.
- Pérez O'Brien A.M., Utsunomiya Y.T., Mészáros G., Bickhart D.M., Liu G.E., Van Tassell C.P., Sonstegard T.S., Da Silva M.V.B., Garcia J.F., Sölkner J. (2014). Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle. *Genet. Sel. Evol.*, 46: 19.
- Polak G., Krupiński J., Martyniuk E., Calik J., Kawęcka A., Krawczyk J., Majewska A., Sikora J., Sosin-Bzducha E., Szyndler-Nędza M., Tomczyk-Wrona I. (2020). The risk status of Polish local breeds under conservation programmes – new approach. *Ann. Anim. Sci.*, 21: 125–140.
- Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., de Bakker P.I.W., Daly M.J., Sham P.C. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.*, 81.
- Purfield D.C., McParland S., Wall E., Berry D.P. (2017). The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *PLoS ONE*, 12(5), e0176780.
- Sabeti P.C., Reich D.E., Higgins J.M., Levine H.Z., Richter D.J., Schaffner S.F., Gabriel S.B., Platko J.V., Patterson N.J., McDonald G.J., Ackerman H.C., Campbell S.J., Altshuler D., Cooper R., Kwiatkowski D., Ward R., Lander E.S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, 419: 832–837.
- Scheet P., Stephens M. (2006). A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.*, 78: 629–644.
- Talebi R., Mardi M., Zeinalabedini M., Szmatoła T., Alamouti M.K., Taklimi S.A.J., Ghaffari M.R. (2025). Genomic architecture of purebred and crossbred Moghani lambs with Texel and Booroola sheep. *Sci. Rep.*, 15: 22833.
- Wright S. (1949). The genetical structure of populations. *Ann. Eugen.*, 15: 323–354.
- Yurchenko A.A., Deniskova T.E., Yudin N.S., Dotsev A.V., Khamiruev T.N., Selionova M.I., Egorov S.V., Reyer H., Wimmers K., Brem G., Zinovieva N.A., Larkin D.M. (2019). High-density genotyping reveals signatures of selection related to acclimation and economically important traits in 15 local sheep breeds from Russia. *BMC Genom.*, 20: 294.
- Zhang M., Peng W., Hu X. (2018). Global genomic diversity and conservation priorities for domestic animals are associated with the economies of their regions of origin. *Sci. Rep.*, 8: 11677.

Websites:

- https://www.ncbi.nlm.nih.gov/projects/SNP/docs/rs_attributes.html
- <http://www.allelefrequencies.net/>
- <http://www.ensembl.org/index.html>
- <https://genome.ucsc.edu/cgi-bin/hgTables>
- <https://david.ncifcrf.gov/list.jsp>
- https://pqstat.pl/?mod_f=test_pca
- <https://www.illumina.com>
- http://www.bioroznorodnosc.izoo.krakow.pl/system/files?file=2008.108.691_0.pdf
- <https://www.gov.pl/web/rolnictwo/produkty-regionalne-i-tradycyjne>
- <https://www.animalgenome.org/cgi-bin/QTLdb/OA/index>

Received: 27 X 2025

Accepted: 5 I 2026