Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Historical gene flow between wild and domestic sheep: genome-wide insights into the genetic history of domestication

Sadegh Taheri
Ferdowsi University of Mashhad
Kimia Kazemi
Ferdowsi University of Mashhad
Hamidreza Khajavi
Ferdowsi University of Mashhad
Hossein Bazzi
h.bazzi@uoz.ac.ir

University of Zabol

Research Article

Keywords: Sheep domestication, wild sheep, population genomics, gene flow, introgression

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11257827/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Historical gene flow between wild and domestic sheep: genome-wide insights into the genetic history of domestication

Sadegh Taheri[1], Kimia Kazemi[1], Hamidreza Khajavi[1], and Hossein Bazzi[2]

[1] Department of Animal Science, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran

[2] Department of Animal Science, Faculty of Agriculture, University of Zabol, Zabol, Iran

*__Corresponding authors:__ Hossein Bazzi; h.bazzi@uoz.ac.ir

Abstract

Background: The domestication of sheep involved complex processes such as genetic divergence, gene flow, and selection. However, it remains unclear whether modern domestic breeds retain ancestral genomic components shared with wild sheep. In this study, we analyzed 22 sheep populations, including two wild populations and 20 commercially important dairy, meat, and wool breeds representing contrasting selection histories, to investigate genetic relationships, shared genomic components, historical gene flow, and genomic signatures of selection.

Results: Analyses of genetic distance, PCA, ADMIXTURE, and hierarchical clustering consistently revealed significant population differentiation. These analyses identified APM (wool), LME (meat), and LMI (dairy) as the domestic breeds genetically closest to the wild populations, while CHI (dairy), QEZ (meat), and ROM (wool) were found to be the most genetically distant. Demographic analyses further showed more recent inferred divergence and larger recent effective population sizes in near-wild than far-wild breeds .Cross-population IBD analysis demonstrated extensive but heterogeneous genomic sharing with wild sheep. APM showed the largest absolute amount of shared IBD, while QEZ exhibited the highest amount of wild-shared IBD when normalized for sample size and contained the greatest number of very long segments. Several extended wild-shared regions harbored biologically relevant genes, including *MSTN*, *APOE/APOC*, *FADS*, and *PRNP*. TreeMix identified three historical

migration events, indicating that divergence alone could not explain the observed population relationships. Additional analyses, including D-statistic, f4-ratio, and f-branch tests, revealed significant excess allele sharing in 19 of 35 tested configurations. These results demonstrated historical gene flow between wild sheep and both genetically proximate and distant domestic breeds, with particularly strong signals involving SMU and QEZ. Genome-wide selection scans identified several putative candidate regions in comparisons of far-wild vs. wild, near-wild vs. wild, and near-wild vs. far-wild sheep. Integrative analyses combining QTL overlap, PPI, MCODE, and STRING evidence prioritized candidate genes associated with production, reproduction, body condition, fleece quality, milk production, and metabolic traits. Noteworthy candidates included *GSK3B*, *UBXN7*, *CACNA1D*, *TCTEX1D2*, *CHDH*, *ACTR3*, and *TTC33*. Functional enrichment analyses indicated involvement in cell-cycle regulation, phosphorylation, cellular signaling, and cytoskeletal processes, while comparisons of near- versus far-wild sheep suggested enrichment for fat digestion and absorption, as well as insulin secretion.

Conclusion: These findings suggest that both divergence and historical gene flow with wild sheep have shaped the genomic history of domestic sheep. The heterogeneous wild-shared genomic ancestry and distinct selection signals among domestic breeds indicate that ancestral introgression, haplotype retention, and breed-specific selection have jointly contributed to the current genomic architecture of sheep. Overall, this study provides a genome-wide framework linking genetic affinity, ancestral gene flow, evidence of shared ancestral genomic components, and selection to the evolutionary history of domestic sheep.

Keywords Sheep domestication, wild sheep, population genomics, gene flow, introgression

Background

Recent advances in sequencing, such as genotyping by SNPs, have greatly enhanced the characterization of genetic relationships among livestock populations. Single-nucleotide polymorphism (SNP) markers provide genome-wide information that can be used to assess population structure, genetic differentiation, and patterns of shared ancestry among breeds and populations. Genome-wide SNP analyses have been highly valuable in revealing the genetic structure, relationships, and historical patterns of diversity among sheep breeds [1]. Genome-wide markers capture actual patterns of genetic variation across the genome and therefore provide a robust approach for investigating the evolutionary and demographic processes that have influenced domestic animal populations. In sheep, genome-wide analyses have revealed substantial genetic differentiation among breeds while also providing evidence of historical admixture and shared ancestry among geographically and phenotypically distinct populations. [1, 2]

Sheep (*Ovis aries*) are among the earliest domesticated livestock species, with archaeological evidence indicating that sheep domestication occurred in southwestern Asia approximately 10,000 years ago. [3] The domestication process was followed by population expansion, geographic dispersal, and subsequent breed formation, during which domestic sheep were exposed to diverse environmental conditions and human-driven selection. Whole-genome studies indicate that the genetic history of domestic sheep reflects a combination of domestication, demographic expansion, migration, genetic drift, admixture, and subsequent artificial and natural selection. [1, 2, 4]

The genetic relationship between domestic sheep and their wild relatives is therefore essential for understanding the evolutionary history of the species. The Asiatic mouflon (*Ovis*

orientalis) has long been regarded as the principal wild ancestor of domestic sheep, although genomic evidence indicates a more complex evolutionary history involving multiple wild *Ovis* populations. [4,5] Whole-genome analyses have identified substantial genetic differentiation among wild *Ovis* populations as well as historical gene flow among some wild lineages, suggesting that the evolutionary history of sheep involved repeated genetic exchange rather than complete reproductive isolation among ancestral populations. [5]

Historical gene flow between wild and domestic populations is highly relevant to understanding sheep domestication because introgression can introduce genetic variation into domestic populations that may subsequently be retained, modified, or lost through demographic and selective processes. Genome-wide resequencing of worldwide wild and domestic sheep has identified evidence of introgression and retained wild-derived variation in domestic populations, suggesting that gene flow from wild relatives contributed to the genomic diversity of modern sheep. [4] Similarly, genomic analyses of European mouflon and domestic sheep have detected patterns consistent with adaptive introgression from mouflon into domestic populations, including genomic regions potentially associated with immune and adaptive functions. [6]

These findings suggest that the genomic differentiation observed among contemporary domestic breeds may reflect not only domestication and subsequent artificial selection but also historical interactions with wild populations. Wild–domestic gene flow may have contributed to the genetic diversity and local adaptation of domestic sheep, while subsequent demographic processes and breed formation may have led to variable retention of wild-derived genomic variation among breeds. [4-6] Consequently, investigating the

extent to which domestic breeds share genomic segments with wild populations may provide additional insights into the historical processes underlying sheep domestication and breed diversification.

Identity-by-descent (IBD) analysis provides a complementary approach for investigating such historical relationships. IBD segments are genomic regions shared between individuals because they are inherited from a common ancestral source, and their distribution can provide information about shared ancestry and demographic history. [7] Compared with analyses based solely on allele-frequency differences, IBD-based approaches retain information about the sharing of contiguous genomic regions and can therefore provide greater resolution for investigating historical genetic relationships among populations. [7, 8] Patterns of IBD sharing between wild and domestic sheep may therefore identify genomic regions that have persisted across population boundaries and may represent a signal of historical genetic connectivity or introgression. [4, 8]

Despite increasing evidence for genetic exchange between wild and domestic sheep, the extent to which different domestic breeds retain genomic signatures associated with their historical relationships with wild populations remains incompletely understood. Domestic breeds have experienced heterogeneous demographic histories, geographic distributions, and selection pressures, which may have resulted in substantial differences in the retention of wild-associated genomic segments. Comparative evaluation of breeds showing different levels of genomic affinity to wild populations may therefore provide a valuable framework for investigating whether the extent and distribution of shared genomic segments vary according to the genetic relationship between domestic and wild populations.

Modern domestic sheep breeds have also undergone prolonged artificial selection for production-related traits, with dairy, meat, and wool breeds developed under distinct production systems, geographic settings, and breeding histories. These contrasting selection histories may have influenced the retention or differentiation of genomic regions shared between wild and domestic sheep. Comparing breeds with contrasting genomic affinities to wild populations across major production groups therefore provides an opportunity to examine how historical wild–domestic genetic relationships may intersect with subsequent breed-specific selection and specialization.

Therefore, this study aimed to investigate the genomic relationships between wild and domestic sheep across a broad panel of breeds and to characterize IBD haplotype segments shared between wild populations and domestic breeds. Domestic breeds were selected from three major production groups, dairy, meat, and wool, to represent contrasting levels of genomic affinity to the wild populations. Within each production group, one breed with relatively high affinity and one with relatively low affinity to the wild populations were selected, allowing comparisons to be made between genetically proximate and distant breeds while maintaining comparable production contexts. Based on genome-wide genetic relationships, domestic breeds showing relatively close and distant genomic affinities to wild populations were comparatively evaluated to determine whether breeds with greater genomic proximity to wild sheep exhibit distinct patterns of shared IBD segments. The genomic regions identified through these analyses were subsequently annotated to characterize the genes and biological functions potentially associated with the retained segments. Collectively, this approach was designed to offer genomic evidence consistent with historical connectivity and potential gene flow between wild and domestic sheep, while

providing insights into the retention of ancestral genomic variation during domestication, dispersal, and subsequent diversification of modern sheep breeds.

Materials and Methods

Resource population

The genomic data analyzed in this study were sourced from the WIDDE database (http://widde.toulouse.inra.fr/widde/), a publicly accessible resource that provides genotype data for sheep. The dataset includes 22 sheep populations representing three major production types: 10 meat breeds, 6 wool breeds, 4 dairy breeds, and 2 wild populations (Additional File 1, Table S1). According to the information provided by the WIDDE database, these populations were genotyped using the Illumina OvineSNP50 BeadChip. The two wild populations included the European mouflon (EMU) and the Sardinian mouflon (SMU). In contrast, the domestic populations comprise dairy, meat, and wool breeds, each with distinct production systems and breeding histories. This diverse dataset allows for comparative analysis of genomic variation across different production types and wild populations. A complete list of populations, breed abbreviations, production types, and sample sizes can be found in Table S1.

Quality Control

Genotype data for 22 sheep populations were obtained from the WIDDE database. Quality control procedures were performed using PLINK v1.9 software [9]. Individuals with a genotyping call rate below 85% were excluded from the dataset. Additionally, SNP markers

were removed if they met any of the following criteria: (1) genotyping call rate below 90%, (2) minor allele frequency (MAF) less than 5% [10].

Genetic distances IBS-based

To characterize the genomic affinity of domestic breeds to the wild populations, genetic distances between domestic and wild breeds were calculated using the --distance 1-ibs flat-missing square in PLINK v1.9 [9]. Pairwise values were calculated across individuals and subsequently summarized at the population/breed level to generate a breed-level distance matrix. A heatmap was generated using the pheatmap v1.0.13 package in the R v4.5.0 software [12]. To minimize potential confounding by production type, domestic breeds were considered separately within the three major production groups (dairy, meat, and wool). For each domestic breed, the mean genetic distance to the two wild populations (EMU and SMU) was calculated. Within each production group, the breed with the lowest mean distance was classified as the near-wild breed, whereas the breed with the highest mean distance was classified as the far-wild breed. This matched-pair design allowed subsequent comparisons to contrast breeds with different genomic affinities to wild sheep while retaining representation of the major production systems.

Population structure and clustering analysis

To evaluate genetic relationships among the populations studied and assess the consistency of breed classification based on mean genetic distance to wild populations, we performed principal component analysis (PCA), model-based admixture analysis, and hierarchical clustering using a QC-filtered SNP dataset. The dataset was initially subjected to LD pruning using PLINK v1.9 [9]. SNPs were pruned with a 50-SNP sliding window, a 5-SNP

step size, and an LD threshold of r2 = 0.20 [11]. This process resulted in the retention of 30,134 SNP markers for subsequent analyses. PCA was conducted in PLINK v1.9 to visualize the overall genomic structure and genetic differentiation among individuals [9]. The resulting principal components were plotted using the ggplot2 v3.4.4 package in R v4.5.2 [13]. Population structure was further investigated using ADMIXTURE v1.3.0, which estimates individual ancestry proportions under different numbers of assumed ancestral populations (K = 1–22) using a maximum-likelihood framework [14]. The resulting ancestry proportions were visualized as stacked bar plots using R v4.5.2 to compare the genetic composition and inferred population structure of the wild and domestic populations. Additionally, we performed hierarchical clustering on the breed-level genetic distance matrix using the Ward.D2 method, as implemented in the hclust function in R v4.5.2 [11]. We used the resulting dendrogram to visualize clustering relationships among the populations. These complementary analyses allowed us to evaluate whether the population groupings and the identification of genetically close and distant domestic breeds, based on genetic distance, were consistent with the broader genomic structure of the populations studied.

Segment Sharing

To identify genomic segments shared between near-wild and wild, and far-wild and wild sheep populations, genotype data were processed using bcftools v1.19 to ensure consistent variant ordering and formatting before haplotype phasing [15]. Haplotype phasing was performed using Beagle v5.4 [16], after which the phased genotypes were analyzed with hap-IBD v1.0 [17] to identify IBD segments. Due to the unavailability of duplicate genotypes and pedigree-based trios, we were unable to directly estimate the

switch-error rate. To evaluate phasing reproducibility, we independently phased the six populations twice using Beagle v5.4 with different random seeds [16, 17]. We assessed phase-transition concordance between the two independent phasing runs and examined the robustness of downstream IBD detection by comparing both the number and cumulative length of IBD segments identified in the two analyses. The hap-IBD analysis utilized a minimum seed length of 0.5 cM, a minimum output segment length of 0.5 cM, and required a minimum of 30 markers per IBD segment. We identified IBD segments shared between each domestic breed (near and far) and the two wild populations. The detected wild-shared IBD segments were characterized for each domestic breed by calculating the number of shared segments, the cumulative segment length, and the genomic coverage.

Gene Flow and Introgression

The relationships between wild and domestic sheep populations (near and far), as well as potential gene flow among them, were examined using TreeMix v1.13 [18]. The analysis involved 409 individuals from eight different populations: two wild populations (EMU and SMU) and six domestic breeds (APM, CHI, LME, LMI, QEZ, and ROM). Genotype data were prepared for TreeMix by converting individual genotypes into allele counts for each population. In the primary analysis, EMU and SMU served as reference populations to root the population graph, consistent with previous studies on the relationships between mouflon and domestic sheep [5]. Given the complex and possibly feral history of domestic-derived Sardinian mouflon [5], we also conducted a sensitivity analysis using EMU alone as the rooting population while keeping SMU as a separate population. Since the OvineSNP50 array is subject to ascertainment bias, we conducted an additional sensitivity analysis to

assess its potential impact on migration inference in TreeMix. For this analysis, we restricted the TreeMix dataset to 15,379 SNPs that were polymorphic in both wild populations (EMU and SMU). This adjustment helped minimize the influence of SNPs that may not provide reliable information for wild populations due to ascertainment bias associated with the array. We evaluated TreeMix models with 0 to 5 migration edges (m = 0–5) under three rooting schemes, and we selected the optimal model based on improvements in likelihood and model complexity [19]. We then compared the number, population-level patterns, and estimated weights of migration edges between the original dataset and the ascertainment-filtered dataset. To ensure the reliability of the inferred population relationships, we performed 100 bootstrap replicates using blocks of 500 SNPs to account for linkage disequilibrium [20]. The final population graph and migration edges were visualized and interpreted based on their direction and estimated weight.

D-statistics, f4-ratio and f-branch analyses

D-statistics and f4-ratio statistics were calculated using the Dtrios module of Dsuite v0.5 [21]. The analysis included six domestic sheep breeds (APM, CHI, LME, LMI, QEZ, and ROM), while the two wild populations were EMU and SMU. Based on the population relationships inferred from genetic distance, clustering, and phylogenetic analyses, ROM represented the most divergent domestic lineage in the inferred population topology and was selected as the rooting population for the D-statistics and f-branch analyses. Since ROM is a domestic breed rather than a wild or ancestral lineage, using it as the rooting population may lead to uncertainties in determining the ancestral and derived allele polarization for certain quartet configurations. This rooting creates a consistent framework for analyzing the

patterns of excess allele sharing between the two wild populations (EMU and SMU) and the other domestic breeds. However, the potential influence of the domestic outgroup on allele polarization was taken into account when interpreting the results. Statistical significance was assessed using the corresponding Z-scores and P-values, with D-statistics considered significant at $|Z| > 2$ and $P < 0.05$. The population tree used for the analysis was the eight-population TreeMix tree generated without migration edges ($m = 0$), representing the inferred population relationships in the absence of gene-flow edges. To further characterize lineage-specific patterns of allele sharing, the Fbranch module of Dsuite was applied using the TreeMix population topology [21, 22]. The resulting f-branch statistics were used to quantify excess allele sharing between individual populations and ancestral branches of the population tree. The f-branch results were visualized using the dtools.py script provided with Dsuite, together with the corresponding population topology. Higher f-branch values were interpreted as stronger signals of excess allele sharing between a population and the descendants of a given ancestral branch, providing evidence consistent with historical gene flow or admixture [21, 22].

Estimates of effective population size

Recent trajectories of effective population size (Ne) were estimated using SNeP v1.1, a tool that infers historical Ne from linkage disequilibrium patterns in genome-wide SNP data [23, 24]. Also, to estimate relative divergence times, folded two-dimensional joint site-frequency spectra (SFS) were generated for each domestic breed compared to a pooled wild group (EMU + SMU), with each population represented by 40 gene copies. This common projection level was used to standardize the number of gene copies each population

contributes to the joint site-frequency spectrum and to minimize differences due to varying sample sizes. This projection was not regarded as an ascertainment correction because the OvineSNP50 array was developed through a predefined SNP discovery process, which can influence the observed allele frequency spectrum [25]. Demographic parameters were inferred using moments [26, 27] under a two-population split-without-migration model (M1). This model estimates the relative effective population sizes of the two descendant populations and the scaled divergence time (Tdiv) through maximum-likelihood optimization of the joint SFS. As part of a sensitivity analysis, we evaluated an alternative asymmetric migration model (M2) for all comparisons between domestic breeds and wild populations. This assessment aimed to determine the potential impact of post-divergence gene flow on the estimated divergence times. Since model M1 assumes strict isolation following the population split, its estimates of Tdiv were regarded as relative divergence-time estimates that are dependent on the model, rather than as calibrated historical dates. This approach aligns with methods previously used to reconstruct the demographic history and divergence of domestic sheep [28, 29, 4].

Detection of selection signatures

Selection signatures were identified through a combined analysis of genetic differentiation (F_{ST}) and nucleotide diversity (π). Using PCA, admixture analysis, hierarchical clustering, and genetic distance analysis, the three domestic breeds genetically closest to the wild populations, LMI, LME, and APM, were classified as the near-wild group. In contrast, CHI, QEZ, and ROM were categorized as the far-wild group. Three pairwise comparisons were conducted: near-wild versus wild, far-wild versus wild, and near-wild versus far-wild

populations. F_{ST} and π were calculated across the autosomes using 100-kb sliding windows with a 10-kb step size in VCFtools v1.16 [30]. To ensure reliable estimation of window-based statistics under the available SNP density, only genomic windows containing at least 5 SNPs were retained for downstream analyses. The Weir and Cockerham estimator was used to calculate F_{ST} [31], and the π ratio was computed and transformed into $\log_2(\pi_1/\pi_2)$. We standardized F_{ST} values as $Z(F_{ST})$. Windows that fell within the top 5% of both $Z(F_{ST})$ and the $\log_2$ (π ratio) distributions were defined as putative candidate regions under divergent selection [32, 33]. To assess the robustness of the pooled-group analysis and account for potential breed heterogeneity within each group, additional pairwise comparisons were conducted among matched production classes: APM versus ROM for wool, LME versus QEZ for meat, and LMI versus CHI for dairy. For these breed-level comparisons, F_{ST} and π were computed using larger 500-kb windows with a 250-kb step to increase the number of informative SNPs per window and improve the stability of window-based estimates given the smaller sample sizes of individual breeds. The same putative candidate regions framework was then applied to these comparisons. Putative candidate regions identified in the three pairwise comparisons were compared to identify shared signals. Overlapping or adjacent windows supported by at least two pairwise comparisons were merged into candidate genomic regions.

Gene report, QTL report, and PPI network

Genomic regions identified through the IBD and selection-signal analyses were annotated against the sheep reference genome using BEDTools v2.31.1 [34], with the Oar_v3.1 reference genome annotation (GTF). Genomic intersections between the identified

regions and annotated genes were determined using BEDTools v2.31.1 intersect. Previously reported quantitative trait loci (QTLs) were retrieved from the Animal QTL database (www.animalgenome.org/cgi-bin/QTLdb/OA/index). For each candidate gene, genomic overlap with reported QTL intervals was assessed, and the corresponding QTL identifier and associated trait information were recorded [35]. To further characterize functional relationships among candidate genes, protein–protein interaction (PPI) networks were constructed [36] using the STRING database with a high-confidence interaction threshold of 0.7 [37], and networks were visualized and analyzed in Cytoscape [38]. Network topology was evaluated using Network Analyzer, with node degree used to characterize individual gene connectivity [39]. MCODE was subsequently applied to identify densely connected network modules using the specified parameters (degree cutoff = 2, node score cutoff = 0.2, k-core = 2, and maximum depth = 100) [40]. A minimum MCODE score of 0.4 and at least 3 nodes per cluster were required [41]. Modules were retained when they met the predefined minimum score and node-number criteria. Importantly, the integrative analysis was performed separately for the three genomic contrasts (Far_Wild, Near_Far, and Near_Wild) and subsequently compared across contrasts. For each contrast, candidate genes were characterized using complementary evidence from QTL overlap, PPI connectivity, MCODE module membership, and STRING interactions.

Functional enrichment analysis

To investigate the biological functions and molecular pathways associated with genes linked to discriminant SNPs, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed using ShinyGO [42]. GO enrichment

was assessed across the three main ontology domains, including biological process (BP), molecular function (MF), and cellular component (CC). In contrast, KEGG enrichment analysis was used to identify potentially relevant metabolic and signalling pathways. ShinyGO integrates gene annotations primarily derived from Ensembl and STRING databases and provides functional enrichment analysis based on the hypergeometric test, with multiple-testing correction performed using the Benjamini–Hochberg false discovery rate (FDR) procedure [43]. Gene lists corresponding to the putative candidate genes identified from the discriminant SNP analysis were submitted to ShinyGO using official gene identifiers, and *Ovis aries* was selected as the reference species. Enriched GO terms and KEGG pathways were considered statistically significant at an FDR < 0.05 [35]. We then summarized and visualized the enriched terms and pathways to support biological interpretation of the genomic regions associated with breed differentiation.

Results

Quality control of genotype data

A total of 50,384 SNP markers from 1,225 sheep were obtained from the WIDDE database for population-level genomic analysis. Quality control was performed using PLINK v1.9. The initial overall genotyping rate was 85.06%. SNP markers with a call rate below 90% were excluded, leading to the removal of 13,616 markers due to missing genotype data, particularly in wild populations. Additionally, 2,444 markers were removed based on a minor allele frequency (MAF) threshold of 5%. No individuals were excluded based on the individual-level call-rate threshold of 85%. After filtering, 34,324 SNP markers across all

1,225 sheep were retained for further population genomic analyses, resulting in an overall genotyping rate of 99.00% (Table 1).

Genetic distance and population structure

Following genotype quality control, genetic distances among the 22 sheep populations were estimated to characterize the genomic relationships between the two wild populations (EMU and SMU) and the 20 domestic breeds. The domestic breeds represented three major production groups, including dairy, meat, and wool breeds**.** To account for the genetic background of each production group, the genetic distance of each domestic breed to both wild populations was calculated, and the mean genetic distance to EMU and SMU was used to rank breeds within each production group. Within each group, the breed with the lowest mean distance to the two wild populations was considered the genetically closest**,** whereas the breed with the highest mean distance was considered the genetically most distant. Based on this criterion, six domestic breeds were selected for subsequent analyses, including APM (wool), LME (meat), and LMI (dairy) as the breeds closest to the wild populations, and CHI (dairy), QEZ (meat), and ROM (wool) as the most distant breeds. Thus, one genetically close and one genetically distant breed were selected from each of the three production groups (Fig. 1).

To determine whether the distance-based classification aligned with broader patterns of genomic structure, we performed PCA, ADMIXTURE, and hierarchical clustering using the QC-filtered SNP dataset. These four complementary approaches revealed generally consistent patterns of population structure and genetic differentiation among the 22 sheep populations (Fig. 2). The ADMIXTURE analysis at K = 4 identified the major components of

genomic ancestry across the populations (Fig. 2A), while K = 20 offered a more detailed view of population-specific ancestry patterns and showed varying degrees of admixture among several domestic breeds (Fig. 2B). The PCA analysis grouped the populations into distinct genomic categories, with the two wild populations (EMU and SMU) positioned differently from most domestic breeds. Among the domestic breeds, there was a range of genetic proximity to the wild populations. LMI, LME, and APM were relatively close to the wild populations within the dairy, meat, and wool categories, respectively. In contrast, CHI, STX, and ROM were more distantly positioned from the wild populations within their production types (Fig. 2C). Hierarchical clustering supported the population structure observed in the PCA and ADMIXTURE analyses, revealing that LME, LMI, and APM formed a closely related cluster. Meanwhile, QEZ and CHI clustered together on a separate branch. ROM, while occupying a distinct branch, was relatively close to CHL and ISF, indicating additional breed-specific relationships beyond the Near-Wild and Far-Wild classifications (Fig. 2D). Overall, these complementary analyses aligned well with genetic-distance results, providing independent visual support for classifying APM, LME, and LMI as Near-Wild representatives, and CHI, QEZ, and ROM as Far-Wild representatives, while also uncovering additional relationships among individual domestic breeds.

Segment sharing

Cross-population IBD analysis revealed marked differences in the extent of genomic sharing between domestic sheep breeds and wild populations. Among the six investigated breeds, APM showed the greatest overall sharing with wild sheep, with 1,769.24 Mb

distributed across 423 IBD segments, followed by QEZ (1,489.32 Mb; 397 segments), whereas CHI exhibited the lowest level of sharing (326.58 Mb; 91 segments), supporting the inferred gradient of genetic affinity (APM > QEZ > LME > ROM > LMI > CHI) (Table 2). The contribution of the two wild populations was generally asymmetric, with EMU accounting for the majority of shared IBD in APM, LME, and CHI (approximately 70–76%), while QEZ and ROM showed a more balanced contribution from EMU and SMU, suggesting shared genomic components with both wild populations (Table 2; Additional File 1, Fig. S1). Long IBD segments (>4 Mb), indicative of relatively recent shared ancestry or the persistence of extended haplotypes, were most abundant in APM (210 segments), followed by QEZ (137) and LME (67). Notably, QEZ harbored the highest number of very long segments (>8 Mb; 10 segments), whereas only one such segment was detected in each of LMI and ROM, potentially reflecting recent introgression, ancestral haplotype retention, or selection on genomic regions shared with wild sheep. After normalization for breed-specific sample size, QEZ displayed the highest level of wild-related IBD sharing (42.6 Mb per individual), followed by ROM (23.1 Mb) and APM (18.1 Mb), indicating that despite its lower absolute sharing than APM, QEZ carries a disproportionately high burden of wild-shared genomic segments at the individual level (Table 2). These results indicate heterogeneous patterns of wild genomic ancestry among domestic breeds, likely reflecting differences in domestication history, breeding practices, and selection intensity. To evaluate the robustness of the IBD results

against phasing stochasticity, each dataset was independently phased twice using different random seeds. The phase-transition concordance between the replicate phasing runs was 96.02% for the near-wild dataset and 81.64% for the far-wild dataset. Despite this variation in phasing reproducibility, the downstream IBD detection remained highly stable. The differences in the number of detected IBD segments were only 0.042% for the near-wild dataset and 0.045% for the far-wild dataset. In terms of cumulative IBD length, the differences were 0.071% for the near-wild dataset and 0.499% for the far-wild dataset (Additional File 2, Table S2).

Further characterization of the wild-shared IBD segments identified several extended genomic regions that were unevenly distributed across the six domestic breeds. Among the 20 regions ranked by genomic length, the largest was in ROM on chromosome 2 (14.86 Mb), followed by APM on chromosome 14 (10.17 Mb), LMI on chromosome 21 (9.91 Mb), and QEZ on chromosome 13 (9.08 Mb). APM was particularly prominent, accounting for 11 of the 20 largest wild-shared regions (Additional File 1, Fig. S2). Several extended regions were also identified in the Far-wild breeds, including QEZ and ROM. Annotation of these regions revealed substantial variation in gene content, with some segments containing hundreds of genes. Notably, the APM region on chromosome 14 included 428 genes, while the LMI region on chromosome 21 had 298 genes. Biologically relevant putative candidate genes identified in these regions include *MSTN* in the ROM chromosome 2 region, *APOE/APOC* genes in the APM chromosome 14 region, *FADS* genes in the LMI chromosome 21 region, and *PRNP* in the QEZ chromosome 13 region. These findings underscore several extended wild-shared

genomic regions that contain genes potentially important for muscle development, lipid metabolism, and other processes related to production and adaptation (Fig. 3). For further information, please refer to Additional File 2, Table S3.

Gene Flow and Introgression

TreeMix analysis was conducted using two different rooting schemes. In the primary analysis, both EMU and SMU were designated as the root populations, while in a sensitivity analysis, only EMU was used as the root. The results from both rooting schemes showed high consistency, producing the same optimal model (m = 3), identical log-likelihood (261.524), and nearly identical migration weights: 0.307055 versus 0.306904, 0.390523 versus 0.390752, and 0.118193 versus 0.118193. Therefore, the inferred population relationships and major migration signals remained robust regardless of the rooting scheme used. TreeMix models including 0 to 5 migration edges were evaluated based on log-likelihood improvement and Akaike information criterion (AIC). The likelihood increased substantially when migration edges were added from m=0 to m=3, indicating that historical gene flow events explain a significant proportion of the covariance among the eight sheep populations. However, additional migration edges beyond m=3 resulted in only minor improvements in likelihood (1.51% and 0.60% for m=4 and m=5, respectively), suggesting that these models mainly capture residual variation rather than biologically meaningful migration events. Therefore, the three-migration model (m=3) was selected as the optimal TreeMix model because it provided the best compromise between likelihood improvement, model complexity, and biological interpretability (Additional File 2, Table S4).

To evaluate the potential impact of ascertainment bias from the OvineSNP50 dataset on migration inference using TreeMix, we conducted an additional sensitivity analysis with 15,379 SNPs that were polymorphic in both wild populations (EMU and SMU). In the ascertainment-filtered dataset, we retained a three-migration model, with the log-likelihood increasing from 256.286 for two migrations (m = 2) to 265.157 for three migrations (m = 3). Adding more migration edges resulted in smaller improvements in likelihood: a change of 2.625 for four migrations (m = 4) and 0.754 for five migrations (m = 5), suggesting a plateau in model enhancement beyond three migration events. Although the major population-level migration patterns remained identifiable after filtering, the estimated migration weights changed significantly. In the original analysis, the weights were 0.307, 0.391, and 0.118, while after filtering, they adjusted to 0.085, 0.207, and 0.474, respectively. These findings indicate that while the number and general population-level patterns of inferred migration events were relatively robust against ascertainment filtering, the magnitude of individual migration signals was sensitive to SNP ascertainment (Additional File 2, Table S5).

To evaluate the robustness of the selected TreeMix model, 100 bootstrap replicates were performed using the three-migration-edge model (m=3). Among the 100 replicates, 84 successfully completed the analysis, whereas 16 replicates failed due to optimization convergence issues. The successful bootstrap replicates showed a mean log-likelihood of 259.13 (SD = 5.46), indicating consistent model performance across resampled datasets. These results support the stability of the inferred population relationships and migration patterns (Additional File 2, Table S6). The inferred TreeMix topology separated the two wild populations (EMU and SMU) as the basal lineages, supporting their distinct evolutionary position relative to domesticated breeds. Among domesticated populations, LME and LMI

formed a closely related cluster, indicating a shared demographic history and/or relatively recent divergence. Similarly, CHI and QEZ showed a close genetic relationship, forming a separate lineage within the domesticated populations. APM was positioned intermediate between the CHI/QEZ lineage and other domesticated breeds, suggesting partial affinity with this group. In contrast, ROM displayed a more divergent position, consistent with its larger genetic distance from several other domesticated populations observed in previous analyses. The migration model identified three historical gene flow events among populations, with estimated migration weights of 0.307, 0.391, and 0.118 (Fig. 4). These inferred migration edges suggest that the current genetic structure of domesticated sheep populations cannot be explained solely by population divergence and that historical admixture events have contributed to the observed genomic relationships. The strongest migration signal (weight = 0.391) involved the lineage associated with the QEZ/CHI cluster, suggesting substantial historical genetic exchange affecting this group. Additional migration signals (weights = 0.307 and 0.118) indicate gene flow involving branches connected to wild-derived and domesticated lineages, suggesting possible introgression or ancestral genetic contributions retained during domestication and breed formation (Fig. 4). In the primary analysis, the most significant migration signal (weight = 0.391) was linked to a lineage associated with the QEZ/CHI cluster. The other two migration edges involved branches connected to wild-derived and domesticated lineages. These inferred migration paths suggest that the current genetic structure of domesticated sheep populations cannot be solely explained by population divergence; rather, historical admixture events have also contributed to the observed genomic relationships.

However, when examining the ascertainment-filtered analysis, we observed substantial changes in the estimated migration weights. The three signals changed to 0.085, 0.207, and 0.474, respectively. While the three-migration model and the major population-level migration patterns remained detectable after restricting the analysis to SNPs polymorphic in both wild populations, the magnitude of individual migration signals was sensitive to SNP ascertainment. In particular, the increase in the migration signal related to SMU and the reduction in the other migration weights suggest that the individual TreeMix weights should not be viewed as precise estimates of the ancestry contributed by specific migration events. Overall, the close association of certain breeds with wild populations, along with the persistence of migration signals after ascertainment-aware filtering, provides evidence that both population divergence and historical gene flow have shaped the current genomic architecture of sheep populations. Nevertheless, due to the sensitivity of migration weights to SNP ascertainment, the TreeMix results should primarily be interpreted as indications of the existence and general direction of historical gene flow rather than as precise estimates of admixture magnitude.

D-statistics, f4-ratio and f-branch analyses

To investigate historical gene flow between wild sheep populations and domestic populations with different levels of genetic affinity to the wild lineages, D-statistic, f4-ratio, and f-branch analyses were performed using ROM as the outgroup. The inferred topology placed EMU and SMU in a common wild clade, LMI and LME in a separate sister clade, and QEZ and CHI in another clade associated with APM. Significant allele-sharing signals were detected between the wild populations and both genetically proximate and distant domestic

524 populations, indicating that excess allele sharing was not restricted to domestic populations
525 closely related to the wild lineage. Among the 35 tested population quartets, 19 showed
526 significant D-statistic signals ($P < 0.05$), indicating excess allele sharing consistent with
527 historical gene flow or introgression (Additional File 1, Fig. S3; Table 3). The D-statistic
528 results showed significant excess allele sharing involving EMU and SMU with APM, LME, and
529 LMI, consistent with historical gene flow between the wild and genetically proximate
530 domestic populations. However, significant signals were also detected between the wild
531 populations and the distant domestic populations QEZ and CHI. The strongest allele-sharing
532 signal involving the wild populations was observed for SMU and QEZ, suggesting a relatively
533 strong excess of shared alleles consistent with historical gene flow between these
534 populations (Table 3). Further details of all D-statistic, f4-ratio, and allele-sharing results are
535 provided in Additional File 2, Table S7. F-branch analysis further supported this pattern and
536 identified stronger excess allele sharing between SMU and the QEZ–CHI lineage than
537 between the wild lineage and the LMI–LME clade. APM also showed allele-sharing signals
538 with both wild populations (Fig. 5). These results indicate that excess allele sharing between
539 wild and domestic sheep was not limited to genetically proximate populations and was also
540 detected in more distantly related domestic populations, particularly QEZ, consistent with
541 historical gene flow.

542 Local F-statistic analyses provided further insight into the genome-wide patterns of
543 introgression, revealing that these signals were not uniformly distributed across the genome
544 but were concentrated in specific regions. The strongest local signal was detected for the
545 LMI-QEZ-APM topology on chromosome 23, with values of $D = 0.0696$, $f_d = 0.246$, and $f_{dM} =$
546 0.138. Additional strong signals for the same topology were observed on chromosome 1, with

547 f_d values of 0.243 and 0.222. Similarly, the APM-QEZ-LME topology displayed elevated local
548 introgression statistics on chromosomes 1 (f_d = 0.197) and 23 (f_d = 0.181). For the APM-QEZ-
549 ROM topology, prominent signals were detected on chromosomes 23 and 22, with maximum
550 f_d values of 0.181 and 0.144, respectively. The APM-CHI-LME topology also exhibited a
551 significant local signal on chromosome 7, with an f_d value of 0.174. The recurrence of elevated
552 f_d values in overlapping or adjacent genomic regions across multiple population topologies
553 suggests that historical allele sharing was concentrated in specific genomic segments rather
554 than being evenly distributed throughout the genome (Additional File 1, Fig. S4).

555 **Estimates of effective population size**

556 For joint SFS construction, all populations were down-projected to 40 gene copies (20
557 diploid individuals) to standardize the number of sampled chromosomes across
558 populations. The comparison of the two demographic models showed that the split-without-
559 migration model (M1) had a better fit than the alternative migration model (M2) for all six
560 comparisons between domestic breeds and wild populations, as indicated by lower AIC
561 values (Additional File 2, Table S8). Therefore, M1 estimates were used for the primary
562 interpretation of relative divergence times. Under the M1 model, CHI exhibited the oldest
563 inferred divergence from wild sheep (Tdiv = 0.002083), followed by ROM (0.001751) and
564 QEZ (0.001710). In contrast, LMI (0.001575), APM (0.001551), and LME (0.001488)
565 displayed progressively more recent inferred divergences (Fig. 6). As a result, all three Far-
566 Wild breeds showed older inferred divergence times compared to the three Near-Wild
567 breeds, with the mean Tdiv being approximately 20% higher in Far-Wild breeds than in Near-
568 Wild breeds. Recent SNeP estimates, which date back about 13 generations, revealed

significant differences in Ne among domestic populations. These ranged from 95 in CHI to 310 in APM, with ROM, QEZ, LME, and LMI having Ne values of 107, 195, 230, and 249, respectively. The corresponding wild populations had recent Ne estimates of 42 for EMU and 55 for SMU (Additional File 2, Table S9; Fig. 6). Overall, Near-Wild breeds had a larger recent Ne than Far-Wild breeds on average. Notably, CHI had both the oldest inferred divergence and the smallest recent Ne among the six domestic breeds.

Genome-wide Selection scan, Gene report, QTL report, and PPI network

The Far-wild vs wild comparison identified 471 candidate genomic regions showing elevated genomic differentiation or altered nucleotide diversity (Fig. 7A). Functional characterization of genes located within these regions identified several candidates supported by complementary evidence from QTL overlap, PPI connectivity, MCODE clustering, and STRING interactions. Among the prioritized genes, *GSK3B* showed the strongest integrative support, with evidence from QTL overlap, PPI connectivity, MCODE membership, and STRING interaction, and was located within a QTL associated with body condition score. Other highly connected candidate genes included *UBXN7*, *FIGN*, and *ZDHHC19*, with QTL associations involving body condition score and lambing potential. Additional prioritized genes were associated with production-related QTLs, including *PRKCB*, *FOXP1*, *PTPRG*, *NCAPH*, *LMBR1*, *EYA3*, *KANSL3*, *ARID5A*, *PPP1R3A*, and *RNF32*, with reported associations with milk yield, clean fleece weight, and pregnancy rate (Fig. 7B). GO enrichment analysis further revealed significant enrichment of several biological processes and pathways among the prioritized candidate genes (Fig. 7C). The most strongly enriched terms included the mitotic cell cycle, phosphate-containing compound metabolic process,

and focal adhesion. Additional enriched categories included cGMP–PKG signaling, protein phosphorylation, chemokine signaling, oxytocin signaling, and cytoskeleton organization, although these showed weaker statistical support (FDR > 0.05). Together, these results indicate that genomic differentiation between Far-wild domestic breeds and wild sheep encompasses candidate genes involved in cell-cycle regulation, phosphorylation and cellular signaling, cytoskeletal organization, and cell–matrix interactions. The top 30 prioritized candidate genes, together with their corresponding QTLs, are provided in Additional File 2, Table S10.

The comparison between near-wild and wild sheep identified 472 candidate genomic regions showing elevated genomic differentiation or altered nucleotide diversity (Fig. 8A). Integrative characterization of genes located within these regions identified several candidate genes supported by complementary evidence from QTL overlap, PPI connectivity, MCODE clustering, and STRING interactions. Among the prioritized candidates, *CACNA1D*, *TCTEX1D2*, and *RXRG* showed the strongest integrative support, each supported by QTL overlap, PPI connectivity, MCODE membership, and STRING interactions. *UBXN7* exhibited the highest network connectivity among the prioritized genes, while *PRKCB*, *PCYT1A*, *OXSR1*, *ARHGAP31*, *ZDHHC19*, *IPCEF1*, *TM4SF19*, and *NCAPH* also showed relatively high network connectivity. The identified candidates overlapped QTLs associated with economically relevant traits, including clean fleece weight, body condition score, milk fat yield, milk yield, mean fiber diameter, and lambing potential (Fig. 8B). GO enrichment analysis revealed several suggestive enrichment signals, although none reached the conventional FDR < 0.05 threshold (Fig. 8C). The most enriched categories included gamete generation, spermatogenesis, CENP-A-containing chromatin assembly, and cell adhesion molecule

614 binding. Several signaling-related pathways also showed enrichment, including MAPK
615 signaling, Rap1 signaling, calcium signaling, and GnRH signaling, with fold enrichments
616 ranging from 3.06 to 5.56. In addition, protein kinase activity showed a 2.18-fold enrichment
617 (FDR = 0.199). Overall, the enrichment profile suggests that candidate genes differentiating
618 near-wild domestic breeds from wild sheep are functionally related to cell adhesion, kinase-
619 mediated signaling, calcium-dependent signaling, and reproductive or gametogenic
620 processes, although these functional associations should be considered suggestive given the
621 lack of FDR-significant terms. The top 30 prioritized candidate genes together with their
622 corresponding QTLs are provided in Additional File 2, Table S11.

623 The comparison between near-wild and far-wild domestic breeds identified the
624 highest number of candidate genomic regions among the three population contrasts, with
625 649 candidate regions showing elevated genomic differentiation or altered nucleotide
626 diversity (Fig. 9A). Integrative characterization of genes within these regions identified
627 several candidates supported by complementary evidence from QTL overlap, PPI
628 connectivity, MCODE clustering, and STRING interactions. Among the prioritized genes,
629 *CHDH*, *ACTR3*, and *TTC33* showed the strongest integrative support, with each receiving
630 evidence from QTL overlap, PPI connectivity, MCODE membership, and STRING interactions.
631 *GSK3B* exhibited the highest network connectivity among the prioritized candidates,
632 followed by *CHDH*, *TTC21A*, *CACNA1D*, *ACTR3,* and *CPEB1*. The prioritized genes overlapped
633 QTLs associated with several production and reproductive traits, including clean fleece
634 weight, milk fat yield, body condition score, milk yield, pregnancy rate, mean fiber diameter,
635 and muscle density (Fig. 9B). GO enrichment analysis identified two notable enrichment
636 signals among the prioritized genes (Fig. 9C). Fat digestion and absorption showed the

strongest enrichment, followed by insulin secretion. Although neither pathway reached the conventional FDR < 0.05 threshold, the strong fold enrichment, particularly for fat digestion and absorption, indicates a potentially relevant functional signal associated with differentiation between near-wild and far-wild domestic sheep. The enrichment of genes involved in lipid digestion and absorption and insulin-related regulation may reflect functional differences in metabolic processes between the two domestic groups. The top 30 prioritized candidate genes together with their corresponding QTLs are provided in Additional File 2, Table S12.

Robustness of Pooled Near–Far Selection Signals

To assess the robustness of the pooled Near–Far selection signals, three pairwise comparisons involving matched production classes were analyzed: APM versus ROM (wool), LME versus QEZ (meat), and LMI versus CHI (dairy). These comparisons identified 158, 208, and 153 putative candidate windows, respectively, yielding a total of 519 outlier windows and 480 unique genomic windows across all analyses. Notably, 39 of these windows were supported by at least two pairwise comparisons, but no single window was common to all three comparisons. After consolidating overlapping or adjacent supported windows, this analysis resulted in 30 candidate genomic regions, including two regions that contained outlier windows from all three pairwise comparisons. When the pooled Near–Far results were specifically examined alongside the matched pairwise analyses, it was found that 38 out of the 649 pooled Near–Far outlier windows were supported by at least two pairwise comparisons. These were further combined into 21 independent genomic regions encompassing a total of 48 annotated genes. The support for these 21 regions came from

various combinations of two matched comparisons: 13 regions were supported by both LME versus QEZ and LMI versus CHI, five regions by APM versus ROM and LME versus QEZ, and three regions by APM versus ROM and LMI versus CHI. No pooled Near–Far region received support from all three pairwise comparisons. Detailed information on genomic coordinates, pairwise support, and annotated genes for these regions is available in Additional File 2, Table S13.

Discussion

The genetic relationships between wild and domestic sheep populations offer valuable insights into the demographic and evolutionary processes that have shaped modern sheep breeds. Previous genome-wide studies indicate that domestic sheep breeds exhibit significant genetic diversity, reflecting differences in geographic origin, demographic history, migration, and breed formation [1]. Comparisons between mouflon and domestic sheep have shown that domestication did not completely separate the wild and domestic lineages. Instead, present-day breeds retain varying levels of genetic affinity [35, 44]. In this study, we conducted an IBS-based genetic distance analysis, which revealed a clear gradient in the genetic affinity of domestic breeds to two wild populations. This pattern suggests that different breeds retain varying degrees of genomic similarity with wild sheep. The domestic breeds selected for the study represent three major production categories: dairy, meat, and wool. This includes commercially important breeds that have undergone significant breed formation and directional selection. Importantly, the differences in genetic affinity were not limited to any single production category. Breeds exhibiting either high or low affinity to wild sheep were found across all three groups, indicating that production purpose alone does not

account for the observed genetic relationships. These differences may instead reflect a combination of factors, including ancestral variation, demographic history, migration, genetic drift, and the development of specific breeds. Overall, the patterns observed in this study provide a valuable genomic framework for understanding how historical demographic processes and potential gene flow have influenced the diverse relationships between modern domestic breeds and wild sheep.

Among the domestic sheep breeds, APM, LME, and LMI exhibited the closest genetic affinity to wild populations, while CHI, QEZ, and ROM were positioned at the opposite end of the genetic-distance spectrum. This differentiation was consistently supported by PCA, ADMIXTURE, and hierarchical clustering analyses. These results indicate that the observed groupings were not solely influenced by the IBS-based distance measure but also reflected the broader genomic structure of the populations. Interestingly, the genomic similarity to wild sheep did not follow the same ranking as the overall genetic distance. APM, which showed the highest overall affinity to wild populations, also had the greatest amount of shared IBD sequences. In contrast, while QEZ ranked among the more distant breeds based on the IBS analysis, it demonstrated a relatively high level of shared IBD. This apparent discrepancy is biologically significant: overall genetic distance reflects genome-wide differences that have accumulated across the genome, whereas IBD segments capture specific stretches of genomic sequence inherited from a common ancestor. As a result, a breed can appear genetically distant from wild sheep on a genome-wide scale while still retaining substantial shared haplotypes. Similar patterns of varied relationships between domestic breeds and wild sheep have been documented in previous genome-wide studies. For instance, whole-genome analyses have shown that domestic sheep form geographically

704 structured genetic groups, while individual breeds can retain different levels of affinity to
705 mouflon and other wild relatives [1, 32, 45, 46]. In a large-scale analysis of domestic and wild
706 sheep worldwide, Lv et al. (2022) [4] found that European mouflon was closely related to
707 domestic sheep and identified significant variation in the contribution of wild-derived
708 ancestry among domestic breeds. More specifically, Li et al. (2020) [47] detected long
709 haplotypes derived from Iranian mouflon and other wild relatives within domestic sheep,
710 revealing considerable variation in the extent of this wild-derived ancestry across different
711 populations. These findings align with the present results and support the notion that
712 genome-wide genetic distance and the retention of specific shared haplotypes represent
713 related yet distinct aspects of the historical relationship between domestic and wild sheep.

714 Previous studies in sheep have supported the biological significance of several genes
715 located within the extended wild-shared IBD regions. One prominent example is the *MSTN*
716 gene, found within the large ROM region on chromosome 2. *MSTN*, also known as myostatin
717 or *GDF8*, is a well-known negative regulator of skeletal muscle growth. Variations in this gene
718 have been consistently linked to traits such as muscularity, growth, carcass composition, and
719 meat production in sheep [48]. Notably, *MSTN* has been shown to contribute to the increased
720 muscularity characteristic of Texel sheep [49]. Recent research has further validated the
721 importance of *MSTN* variants in relation to body weight and growth traits in sheep [50]. The
722 presence of *MSTN* within the exceptionally long wild-shared region observed in ROM is
723 significant, as it places a gene with a documented role in muscle development within a
724 genomic segment that has been preserved between wild and domestic sheep. The large wild-
725 shared region of the APM on chromosome 14 contains several significant genes, including
726 *APOE*, which is known for its role in lipid transport and homeostasis [51]. Recent studies

727 focusing specifically on sheep have provided evidence linking *APOE* to adipose biology [52].
728 In various ovine adipose tissues, *APOE* has been associated with lipid storage and transport
729 processes. Additionally, a recent integrated genome-wide association study (GWAS) and
730 transcriptomic analysis in Hu sheep identified *APOE* as one of the genes associated with fatty
731 acid homeostasis and the deposition of perirenal fat [53]. Therefore, the presence of *APOE*,
732 along with other metabolically relevant genes such as *LIPE*, *FFAR1/2/3*, *FGF21*, and *AKT2*,
733 within this exceptionally long wild-shared region of IBD suggests that the APM region is a
734 genomic interval enriched in lipid and energy metabolism. The LMI wild-shared region on
735 chromosome 21 contains an extensive cluster of *FADS* genes, including *FADS1*, *FADS2*, and
736 *FADS3*. In sheep, *FADS1* and *FADS2* have been frequently linked to fatty acid metabolism and
737 the synthesis of polyunsaturated fatty acids. Their expression and activity are associated
738 with variations in fatty acid composition across muscle and adipose tissues, as well as with
739 dietary responses [54-57]. Furthermore, these genes have been associated in a genome-wide
740 study of body size in Duolang sheep with fatty acid metabolism, adipose deposition, and
741 production-related traits in sheep [58]. Therefore, the presence of the *FADS* gene cluster
742 within this extended wild-shared IBD segment suggests that the LMI region is a candidate
743 genomic interval enriched in fatty acid metabolic functions. Finally, the *PRNP* gene identified
744 within the large QEZ wild-shared region on chromosome 13 serves as a significant example
745 of a locus linked to disease resistance. In sheep, variations in the *PRNP* gene are strongly
746 associated with susceptibility to classical scrapie, with certain genotypes offering greater
747 resistance [59]. Recent studies conducted on European and Middle Eastern sheep
748 populations have further confirmed considerable variation in the *PRNP* gene and its impact
749 on scrapie susceptibility and resistance [60, 61]. The presence of *PRNP* within this wild-

shared region suggests that genomic segments retained between wild and domestic sheep may contain loci pertinent to pathogen resistance and survival. This continuity likely reflects the influences of both natural and artificial selection during the domestication process.

The TreeMix analysis indicated that the genetic relationships among the eight populations cannot be fully explained by population divergence alone. The Sardinian mouflon has a complex history due to human-mediated sheep dispersal [5]. To evaluate the potential impact of the choice of rooting population on the inferred migration graph, we reassessed the analysis using TreeMix. Specifically, we reran the analysis using only the EMU population as the root, instead of using both EMU and SMU together. Both rooting approaches resulted in the same optimal model (m = 3) and produced identical log-likelihood values, with nearly identical migration weights. This indicates that the primary inferred migration patterns and population relationships were consistent regardless of the rooting choice, suggesting that including SMU as a reference population did not significantly influence the TreeMix results. The separation of the EMU and SMU populations from domestic populations, along with the close positioning of LME–LMI and CHI–QEZ, aligns with the genetic structure identified through distance, PCA, ADMIXTURE, and clustering analyses. Importantly, the inclusion of migration edges significantly improved the fit of the population graph, suggesting that historical admixture played a role in the covariance structure observed among these populations. The inferred migration edges involving the QEZ/CHI lineage and branches connected to both wild and domestic populations are particularly significant. They imply that some of the genetic relationships we observe today may have resulted from historical exchanges rather than solely from shared divergence. Similar patterns have been observed in genome-wide studies of sheep, where TreeMix has identified migration edges

between wild relatives and domestic populations. For example, Barbato et al. (2017) [5] detected bidirectional introgression between European mouflon and domestic sheep, with particularly strong signals involving a Sardinian mouflon population. Similarly, whole-genome analyses have identified introgression between wild sheep relatives and domestic populations, including signals involving European mouflon, Asiatic mouflon, and urial [62]. Recent research on Iranian sheep has indicated that TreeMix signals suggest gene flow between domestic sheep and Asiatic mouflon [63]. This provides further support for the idea of historical genetic exchange between geographically associated wild and domestic sheep populations. However, caution is needed when interpreting the TreeMix migration edges in this study, as the OvineSNP50 array does not provide an ascertainment-neutral representation of genetic variation. SNP ascertainment can distort allele frequency distributions and affect population genetic inferences, especially when certain variants are more frequently discovered in specific populations or breeds [64, 65]. This is particularly relevant to our analysis because EMU and SMU represent wild populations and are used as root populations.

To assess the potential influence of this bias, we repeated the TreeMix analysis using 15,379 SNPs that were polymorphic in both wild populations. The three-migration model remained consistent after this filtering, and the key migration patterns at the population level were still detectable. However, the estimated migration weights changed significantly, indicating that the magnitude of individual migration signals is influenced by SNP ascertainment. This sensitivity is crucial when interpreting migration edges involving wild-lineage branches. In the initial analysis, the estimated migration weights were 0.307, 0.391, and 0.118. After the ascertainment-filtered analysis, these weights changed to 0.085, 0.207, and 0.474. These results suggest that while the overall number and general pattern of inferred migration

events are relatively robust, the individual migration weights should not be viewed as precise estimates of the proportion of ancestry contributed by specific migration events. Importantly, the persistence of migration signals after applying ascertainment-aware filtering suggests that the TreeMix inferences are not solely due to the representation of wild variation on the OvineSNP50 array. Nevertheless, the sensitivity of migration weights underscores the necessity for caution when assigning biological significance to the magnitude or precise direction of individual TreeMix edges. TreeMix migration edges represent the covariance in allele frequency patterns consistent with historical gene flow, but they do not definitively establish the exact timing, direction, or demographic proportion of an admixture event. This interpretation is consistent with previous whole-genome studies in sheep, in which TreeMix signals have been complemented by D-statistics, f-statistics, and other approaches to provide stronger evidence for introgression.

Overall, the results from the TreeMix analysis in this study indicate a demographic history marked by both divergence and subsequent admixture, rather than a simplistic model in which domestic and wild populations evolved independently after domestication. The alignment between the TreeMix findings and the independent genetic analyses reinforces the idea that historical gene flow played a role in shaping the observed genomic relationships. Additionally, the ascertainment sensitivity analysis suggests that the migration weights from TreeMix should be understood as qualitative or comparative indicators of historical gene flow, rather than as precise measurements of admixture proportions.

The D-statistic, f4-ratio, and f-branch analyses provide further evidence that historical genetic exchange between wild and domestic sheep was not confined to the breeds with the

highest overall genomic affinity to wild populations. Although APM, LME, and LMI were genetically closer to wild populations, significant allele-sharing signals were also detected in the more divergent QEZ and CHI populations. This pattern is particularly noteworthy for QEZ, which demonstrated a strong association with the SMU lineage, despite its relatively distant position in the genome-wide genetic distance analysis. The agreement between the D-statistic and f-branch results suggests that genome-wide genetic distance and patterns of excess allele sharing capture different aspects of population history. A breed may appear relatively distinct from wild sheep while still retaining evidence of allele sharing consistent with historical gene flow from a wild lineage. Similar patterns of gene flow between wild and domestic sheep have been reported. For instance, Lv et al. (2022) [4] used D-statistics and f_d-based analyses across a global collection of domestic and wild sheep to identify substantial but uneven contributions from various wild Ovis lineages, demonstrating that introgression was not uniformly distributed among breeds. Additionally, Chen et al. (2021) [44] detected significant gene-flow signals between wild sheep and specific domestic breeds using D-statistics and TreeMix, including contributions from European mouflon, Asiatic mouflon, urial, and argali. These findings further support the idea of a complex history of sheep diversification rather than a straightforward separation between wild and domestic lineages. Furthermore, genomic analyses involving more than 1,000 domestic sheep have shown that long divergent haplotypes derived from wild Ovis species occur at varying frequencies among breeds, indicating that the contribution of wild ancestry can differ significantly across the domestic sheep gene pool [46]. The D-statistic results in this study should be interpreted cautiously, as ROM was used as a domestic outgroup rather than an independent wild or ancestral reference. While ROM's divergent position was supported by population-level

841 analyses, its domestic background may create uncertainties in allele polarization for some
842 quartet configurations. Thus, the 19 significant D-statistic configurations are better viewed
843 as evidence of historical gene flow rather than definitive proof of introgression. In this
844 context, the strong signal between SMU and QEZ indicates historical allele sharing, but
845 interpretations should consider the outgroup choice. Overall, the genetic relationships
846 observed suggest a mix of genome-wide divergence and lineage-specific allele sharing, with
847 QEZ exemplifying this distinction.

848 The demographic analyses offer additional support for the genetic differences
849 observed among the six domestic sheep breeds. The consistently older inferred divergence
850 times of the Far-Wild breeds compared to the Near-Wild group indicate that a genetic affinity
851 to wild sheep is linked not only to greater genome-wide similarity and increased sharing of
852 IBD segments but also to a more recent separation from wild populations. This pattern is
853 particularly evident in the CHI breed, which showed the oldest inferred divergence estimate
854 and the smallest recent Ne. In contrast, the APM and LME breeds, which demonstrated some
855 of the strongest genome-wide affinities and IBD sharing with wild sheep, had the most recent
856 divergence estimates. This alignment among genetic distance, haplotype sharing, and
857 demographic inference supports the idea that the current structure of domestic sheep is a
858 result of a combination of domestication, subsequent population differentiation, and
859 recurrent genetic exchange with wild populations, rather than a straightforward divergence
860 from a common wild ancestor. Previous whole-genome studies have also revealed significant
861 breed-specific demographic trajectories, including reductions in Ne linked to domestication
862 and later breed formation. Furthermore, genomic and paleogenomic evidence suggests that

863 domestic sheep populations experienced repeated episodes of admixture, population
864 replacement, and restructuring following their initial domestication [4, 28, 66, 67].

865 Caution is needed when interpreting demographic estimates from SNP data on the
866 OvineSNP50 array, as its design can alter the allele-frequency spectrum relative to whole-
867 genome sequencing. Commercial SNP panels, such as the Ovine 50K BeadChip, often exhibit
868 ascertainment bias, which can influence estimates of genomic diversity and population
869 differentiation [25]. The roughly 20% higher mean Tdiv for Far-Wild breeds might reflect
870 differences in the site frequency spectrum (SFS) rather than divergence history alone, since
871 ascertainment effects vary across populations with different demographic histories.
872 Although a standardized projection of 40 gene copies per population aided sampling, it did
873 not eliminate these biases. Therefore, Tdiv estimates should be viewed as relative measures
874 that indicate broader genetic differentiation patterns rather than absolute historical
875 divergence times.

876 The three comparative analyses revealed distinct sets of candidate genes that reflect
877 varying degrees of genomic divergence between domestic and wild sheep. In the comparison
878 between Far-wild and wild sheep, the prioritized genes included *GSK3B*, *UBXN7*, *FIGN*,
879 *ZDHHC19*, *PRKCB*, *FOXP1*, *PTPRG*, *PPP1R3A*, and *RNF32*. These genes are associated with
880 QTLs related to body condition score, lambing potential, milk yield, clean fleece weight, and
881 pregnancy rate. Functionally, these candidate genes are involved in metabolic regulation,
882 reproductive processes, and cellular signaling. Notably, *GSK3B* has been linked to resilience-
883 related traits, such as body condition score, somatic cell count [68], and regulating the
884 proliferation of mammary cells [69], in dairy sheep, while *PRKCB* is associated with

reproductive regulation and high-fecundity phenotypes in sheep through calcium-, cAMP-, and MAPK-related signaling pathways [70]. Additionally, *PPP1R3A* has been identified as a key regulator in energy storage, with recent multi-omics analyses highlighting its role in differences in tail-fat metabolism and glycogen utilization among sheep with contrasting tail-fat phenotypes [71, 72]. These gene-level associations are supported by the enrichment of pathways related to the mitotic cell cycle, phosphate-containing compound metabolism, and focal adhesion, as well as functions involving cGMP–PKG, protein phosphorylation, chemokines, and oxytocin. Overall, the combination of production-related QTLs and metabolic and cellular pathways indicates that the greater differentiation of Far-wild domestic breeds from wild sheep is linked to genomic components associated with body condition, energy storage, reproduction, and economically important production traits.

The comparison between Near-wild and wild sheep identified several genetic candidates: *CACNA1D*, *TCTEX1D2*, *RXRG*, *UBXN7*, *PRKCB*, *PCYT1A*, *OXSR1*, *ARHGAP31*, *ZDHHC19*, and *NCAPH*. These candidates are associated with QTLs related to various traits, including clean fleece weight, body condition score, milk yield, milk-fat yield, mean fiber diameter, and lambing potential. These genetic candidates can be grouped into four broad categories: calcium-dependent signaling, lipid and metabolic regulation, reproduction, and cellular organization. For instance, *CACNA1D* has been detected in transcriptomic studies of the ovine mammary gland within the oxytocin-signaling pathway [73]. *RXRG* has been linked to lipid metabolism in sheep under nutritional stress and in pigs [74, 75]. Additionally, *PRKCB* connects this group to reproductive signaling, as its expression is associated with differences in fecundity and gonadotropin-releasing hormone (GnRH) regulatory pathways in sheep [68, 70, 76]. The enrichment of processes related to gametogenesis,

908 spermatogenesis, cell adhesion, MAPK, Rap1, calcium, and GnRH signaling provides a
909 functional context for the observed QTLs related to milk production, fleece quality, body
910 condition, and reproduction. Although these GO signals were not significant after FDR
911 correction, their alignment with production- and reproduction-related QTLs suggests that
912 Near-wild breeds maintain a close genomic relationship with wild sheep while already
913 exhibiting differentiation in loci relevant to traits selected through domestication.

914 The comparison of Far-wild and Near-wild regions identified numerous candidate
915 regions and highlighted several key genes: *CHDH*, *ACTR3*, *TTC33*, *GSK3B*, *TTC21A*, *CACNA1D*,
916 and *CPEB1*. These genes are associated with QTLs related to clean fleece weight, milk and
917 milk fat yield, body condition score, pregnancy rate, mean fiber diameter, and muscle
918 density. They encompass a range of functions, including metabolic, reproductive, signaling,
919 and cytoskeletal roles. Notably, the repeated identification of *GSK3B* and *CACNA1D* across
920 different comparisons suggests a connection between metabolic regulation, body condition,
921 and calcium-dependent signaling in distinguishing features between wild and domestic
922 breeds [68, 73]. *CPEB1* adds another layer by affecting reproductive capacity; recent
923 research highlights its role in follicular development through a supported regulatory
924 network in ovarian tissue [77]. At the pathway level, the observed enrichment signals for fat
925 digestion and absorption and insulin secretion may point to differences in lipid utilization,
926 nutrient sensing, and energy allocation between Near-wild and Far-wild domestic breeds.
927 Although these pathway enrichments did not reach the conventional FDR < 0.05 threshold,
928 their biological relevance is consistent with recent sheep studies that observed coordinated
929 changes in lipid metabolism, glucose utilization, and adipose tissue biology among breeds
930 with differing fat deposition traits [71, 78]. Furthermore, genomic evidence links population

history and specific mutations to fat-tail adipogenesis in fat-tailed sheep [29]. Thus, this comparison reveals an additional layer of differentiation within domestic sheep, particularly concerning metabolic and production-related functions. The matched pairwise analyses provided partial support for the pooled Near–Far selection signals. Out of the 649 pooled outlier windows identified, 38 were supported by at least two pairwise comparisons. These were consolidated into 21 genomic regions, which encompass a total of 48 annotated genes. This partial concordance suggests that some of the genomic differentiation between the Near-wild and Far-wild groups is detectable in production-class-matched comparisons. However, the incomplete overlap indicates that the two approaches capture partially distinct components of genomic variation.

The three comparisons reveal a clear gradient of genomic differentiation. The comparison between Far-wild and wild primarily highlighted functions related to production, reproduction, body condition, and energy storage. The Near-wild and wild comparison retained these production-related signals, while also incorporating components associated with calcium and reproductive signaling. In contrast, the comparison between Far-wild and Near-wild focused on metabolic specialization, particularly concerning lipid digestion and insulin-related functions. The alignment of these candidate genes with QTLs for traits such as milk production, body condition related to meat, wool, fertility, and fat storage supports the idea of a gradual differentiation of the domestic genome throughout the domestication process and subsequent breed-specific selection.

Conclusion

This study demonstrates that the genomic history of domestic sheep has been shaped by the combined effects of historical gene flow, retention of wild-shared genomic segments, and subsequent breed-specific differentiation. Contrasting levels of genomic affinity to wild sheep were observed among commercially important dairy, meat, and wool breeds, while IBD and allele-sharing analyses revealed that signals of wild-related genomic variation were heterogeneous and extended beyond the breeds most closely related to wild populations. Evidence from TreeMix, D-statistics, f4-ratio, and f-branch analyses further supported recurrent historical gene flow between wild and domestic lineages, including genetically distant breeds. Demographic analyses showed distinct breed-specific patterns of divergence and recent effective population size. Near-wild breeds exhibited a more recent inferred divergence from wild sheep and, on average, a larger recent Ne than far-wild breeds. The distinct genomic differentiation patterns identified between near-wild and far-wild breeds, together with candidate genes and production-related QTLs, suggest that subsequent breed-specific selection has further shaped these inherited genomic signals. Overall, the findings provide genome-wide evidence that domestication in sheep was not a simple process of divergence from wild ancestors, but a dynamic history involving gene flow, genomic retention, and selection during breed formation and diversification.

List of abbreviations

AIC: Akaike information criterion

BP: biological process

CC: cellular component

EMU: European mouflon

FDR: false discovery rate

F_{ST}: genetic differentiation

GnRH: gonadotropin-releasing hormone

GO: Gene Ontology

GWAS: genome-wide association study

IBD: identity-by-descent

KEGG: Kyoto Encyclopedia of Genes and Genomes

LD: Linkage disequilibrium

MAF: minor allele frequencies

MF: molecular function

Ne: effective population size

PCA: principal component analysis

PPI: protein–protein interaction

QC: quality control

QTLs: quantitative trait loci

SFS: site-frequency spectrum

SMU: Sardinian mouflon

SNP: single-nucleotide polymorphism

Tdiv: scaled divergence time

π: nucleotide diversity

Declarations

Ethics approval and consent to participate

This study utilized publicly available genotype data sourced from the WIDDE database. No animals were sampled or handled in this study; therefore, ethical approval and consent to participate were not necessary.

Consent for publication

Not applicable.

Availability of data and materials

The genotype data used in this study are available publicly through the WIDDE database (Web-Interfaced Next Generation Database dedicated to Genetic Diversity Exploration): http://widde.toulouse.inra.fr/widde/.

Competing interests

The authors declare no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' contributions

Sadegh Taheri: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Validation, Visualization, Writing – review & editing.

Kimia Kazemi: Validation, Writing – original draft, Writing – review & editing.

Hamidreza Khajavi: Writing – original draft.

Hossein Bazzi: Methodology, Project administration, Supervision, Writing – review & editing. All authors read and approved the final manuscript.

Acknowledgements

The authors gratefully acknowledge the computational resources and technical support provided by the data analysis staff at Ferdowsi University of Mashhad. We also sincerely thank **Dr. Ali Javadmanesh**, Department of Animal Science, for his assistance in facilitating access to these resources.

Reference

1. Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, et al. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol. 2012;10:1001258.
2. Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, et al. A genome-wide survey of SNP variation reveals the genetic structure of sheep breeds. PLoS One. 2009;4:4668.
3. Zeder MA. Domestication and early agriculture in the Mediterranean Basin: origins, diffusion, and impact. Proc Natl Acad Sci U S A. 2008;105:11597-11604.

4. Lv FH, Cao YH, Liu GJ, Luo LY, Lu R, Liu MJ, et al. Whole-genome resequencing of worldwide wild and domestic sheep elucidates genetic diversity, introgression, and agronomically important loci. Mol Biol Evol. 2022;39:353.
5. Barbato M, Hailer F, Orozco-terWengel P, Kijas J, Mereu P, Cabras P, et al. Genomic signatures of adaptive introgression from European mouflon into domestic sheep. Sci Rep. 2017;7:7623.
6. Kaptan D, Atağ G, Vural KB, Morell Miranda P, Akbaba A, Yüncü E, et al. The population history of domestic sheep revealed by paleogenomes. Mol Biol Evol. 2024;41:158.
7. Browning SR, Browning BL. Identity by descent between distant relatives: detection and applications. Annu Rev Genet. 2012;46:617-633.
8. Blondeau Da Silva S, Mwacharo JM, Li M, Ahbara A, Muchadeyi FC, Dzomba EF, et al. IBD sharing patterns as intra-breed admixture indicators in small ruminants. Heredity. 2024;132:30-42.
9. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a toolset for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559-575.
10. Taheri S, Saedi N, Zerehdaran S, Javadmanesh A. Identification of selection signatures in Capra hircus and Capra aegagrus in Iran. Anim Sci J. 2023;94:13864.
11. Taheri S, Karimi MO, Saedi N, Zerehdaran S, Shariati MM, Gholizadeh M, et al. Genetic diversity, population structure, and combined detection of selection signatures in Iranian versus Afghan Baluchi sheep. PLoS One. 2026;21:0350262.
12. Kolde R, Kolde MR. Package 'pheatmap'. R package. 2015;1:790.

13. Gómez-Rubio V. ggplot2-elegant graphics for data analysis. J Stat Softw. 2017;77:1-3.
14. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655-1664.
15. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10:008.
16. Browning BL, Tian X, Zhou Y, Browning SR. Fast two-stage phasing of large-scale sequence data. Am J Hum Genet. 2021;108:1880-1890.
17. Zhou Y, Browning SR, Browning BL. A fast and simple method for detecting identity-by-descent segments in large-scale data. Am J Hum Genet. 2020;106:426-437.
18. Pickrell J, Pritchard J. Inference of population splits and mixtures from genome-wide allele frequency data. Nat Precedings. 2012:1-1.
19. Fitak RR. OptM: estimating the optimal number of migration edges on population trees using Treemix. Biol Methods Protoc. 2021;6:017.
20. Muthusamy M, Akinsola OM, Pal P, Ramasamy C, Ramasamy S, Musa AA, et al. Comparative genomic insights into adaptation, selection signatures, and population dynamics in indigenous Indian sheep and foreign breeds. Front Genet. 2025;16:1621960.
21. Malinsky M, Matschiner M, Svardal H. Dsuite - fast D-statistics and related admixture evidence from VCF files. Mol Ecol Resour. 2021;21:584-595.

22. Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, et al. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. Nat Ecol Evol. 2018;2:1940-1955.
23. Barbato M, Orozco-terWengel P, Tapio M, Bruford MW. SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Front Genet. 2015;6:109.
24. Taheri S, Zerehdaran S, Javadmanesh A. Genetic diversity in some domestic and wild sheep and goats in Iran. Small Rumin Res. 2022;209:106641.
25. Benjelloun B, Boyer F, Streeter I, Zamani W, Engelen S, Alberti A, et al. An evaluation of sequencing coverage and genotyping strategies to assess neutral and adaptive diversity. Mol Ecol Resour. 2019;19:1497-515.
26. Jouganous J, Long W, Ragsdale AP, Gravel S. Inferring the joint demographic history of multiple populations: beyond the diffusion approximation. Genetics. 2017;206:1549-1567.
27. Bourgeois YX, Warren BH. An overview of current population genomics methods for the analysis of whole-genome resequencing data in eukaryotes. Mol Ecol. 2021;30:6036-6071.
28. Morell Miranda P, Soares AE, Günther T. Demographic reconstruction of the Western sheep expansion from whole-genome sequences. G3: Genes, Genomes, Genetics. 2023;13:199.
29. Jin M, Liu G, Liu E, Wang L, Jiang Y, Zheng Z, et al. Genomic insights into the population history of fat-tailed sheep and identification of two mutations that contribute to fat tail adipogenesis. J Adv Res. 2025.

30. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27:2156-2158.
31. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984;38:1358-1370.
32. Yang JI, Li WR, Lv FH, He SG, Tian SL, Peng WF, et al. Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. Mol Biol Evol. 2016;33:2576-2592.
33. Li T, Jin M, Wang H, Zhang W, Yuan Z, Wei C. Whole-genome scanning for selection signatures reveals candidate genes associated with growth and tail length in sheep. Animals. 2024;14:687.
34. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841-842.
35. Taheri S, Javadmanesh A, Zerehdaran S. Identification of selective sweep and associated QTL traits in Iranian Ovis aries and Ovis orientalis populations. Front Genet. 2024;15:1414717.
36. Taheri S, Zerehdaran S, Javadmanesh A. Investigation of critical genes and quantitative trait loci related to economic traits in broiler chicken genome using protein-protein interaction network. Poult Sci J. 2025;13.
37. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43:447-D452.

38. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498-2504.
39. Doncheva NT, Assenov Y, Domingues FS, Albrecht M. Topological analysis and interactive visualization of biological networks and protein structures. Nat Protoc. 2012;7:670-685.
40. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics. 2003;4:2.
41. Yang L, He T, Xiong F, Chen X, Fan X, Jin S, et al. Identification of key genes and pathways associated with feed efficiency of native chickens based on transcriptome data via bioinformatics analysis. BMC Genomics. 2020;21:292.
42. Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics. 2020;36:2628-2629.
43. Fonseca I, Trindade F, Santos M, Leite-Moreira A, Moreira-Gonçalves D, Vitorino R, et al. Bioinformatics tools for drug repurposing: a tutorial using heart failure as a case study. J Mol Cell Cardiol Plus. 2025;13:100460.
44. Chen ZH, Xu YX, Xie XL, Wang DF, Aguilar-Gómez D, Liu GJ, et al. Whole-genome sequence analysis unveils different origins of European and Asiatic mouflon and domestication-related genes in sheep. Commun Biol. 2021;4:1307.
45. Ciani E, Mastrangelo S, Da Silva A, Marroni F, Ferenčaković M, Ajmone-Marsan P, et al. On the origin of European sheep as revealed by the diversity of the Balkan breeds and by optimizing population-genetic analysis tools. Genet Sel Evol. 2020;52:25.

46. Cheng H, Zhang Z, Wen J, Lenstra JA, Heller R, Cai Y, et al. Long divergent haplotypes introgressed from wild sheep are associated with distinct morphological and adaptive characteristics in domestic sheep. PLoS Genet. 2023;19:1010615.
47. Li X, Yang JI, Shen M, Xie XL, Liu GJ, Xu YX, et al. Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. Nat Commun. 2020;11:2815.
48. Boman IA, Klemetsdal G, Nafstad O, Blichfeldt T, Våge DI. Impact of two myostatin (MSTN) mutations on weight gain and lamb carcass classification in Norwegian White Sheep (Ovis aries). Genet Sel Evol. 2010;42:4.
49. Clop A, Marcq F, Takeda H, Pirottin D, Tordoir X, Bibé B, et al. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nat Genet. 2006;38:813-818.
50. Xu Y, Gao C, Wang H, Cao C, Zhang X, Sun B, et al. The origin of the double-muscle mutation in the MSTN gene of sheep and its effects on growth traits in the Hu× East Friesian hybrid population. Front Vet Sci. 2026;13:1791780.
51. Yang LG, March ZM, Stephenson RA, Narayan PS. Apolipoprotein E in lipid metabolism and neurodegenerative disease. Trends Endocrinol Metab. 2023;34:430-445.
52. An L, Pan Y, Yuan M, Wen Z, Qiao L, Wang W, et al. Full-length transcriptome and gene expression analysis of different Ovis aries adipose tissues reveals transcript variants involved in lipid biosynthesis. Animals. 2023;14:7.

53. Fu X, Zhao L, Tian H, Zhang D, Zhang Y, Zhao Y, et al. Genome-wide association and transcriptome studies identify candidate genes regulating perirenal fat deposition in sheep. Genomics. 2025;111182.
54. Francisco A, Alves SP, Portugal PV, Pires VMR, Dentinho MT, Alfaia CM, et al. Effect of feeding lambs with a tanniferous shrub (rockrose) and a vegetable oil blend on fatty acid composition of meat lipids. Animal. 2016;10:2061-2073.
55. Dervishi E, González-Calvo L, Blanco M, Joy M, Sarto P, Martin-Hernández R, et al. Gene expression and fatty acid profiling in longissimus thoracis muscle, subcutaneous fat, and liver of light lambs in response to concentrate or alfalfa grazing. Front Genet. 2019;10:1070.
56. Luo R, Zheng Z, Yang C, Zhang X, Cheng L, Su G, et al. Comparative transcriptome analysis provides insights into the polyunsaturated fatty acid synthesis regulation of fat-1 transgenic sheep. Int J Mol Sci. 2020;21:1121.
57. Wang B, Yang L, Luo Y, Su R, Su L, Zhao L, et al. Effects of feeding regimens on meat quality, fatty acid composition and metabolism as related to gene expression in Chinese Sunit sheep. Small Rumin Res. 2018;169:127-133.
58. Wang K, Niu Z, Tang S, Han B, Xu J, Ma Y, et al. GWAS and selection signature analyses of whole-genome sequencing data identify genes associated with body size in Duolang sheep. Front Vet Sci. 2026;13:1869513.
59. Cassmann ED, Moore SJ, Smith JD, Greenlee JJ. Sheep with the homozygous lysine-171 prion protein genotype are resistant to classical scrapie after experimental oronasal inoculation. Vet Pathol. 2019;56:409-417.

60. Rodrigues S, Moreira G, Santos-Silva S, Gomes-Gonçalves S, Aires Pereira M, Baptista A, et al. Mapping PRNP polymorphisms in Portuguese Serra da Estrela ovine populations: insights into scrapie susceptibility and farm animal improvement. Animals. 2025;15:2750.
61. Salman M, Souf IB, Khnissi S, Odeh AB, Halaweh W, M'Hamdi N. Identification and characterization of novel PRNP haplotypes in the Assaf sheep breed from Palestine. Small Rumin Res. 2025;107618.
62. Cao YH, Xu SS, Shen M, Chen ZH, Gao L, Lv FH, et al. Historical introgression from wild relatives enhanced climatic adaptation and resistance to pneumonia in sheep. Mol Biol Evol. 2021;38:838-855.
63. Khalkhali-Evrigh R, Hedayat N, Seyedsharifi R, Shakouri M, Ponnampalam EN. Genomic evidence of improved fertility and adaptation in Iranian domestic sheep attributed to introgression from Asiatic Mouflon and urial. Sci Rep. 2025;15:1185.
64. Albrechtsen A, Nielsen FC, Nielsen R. Ascertainment biases in SNP chips affect measures of population divergence. Mol Biol Evol. 2010;27:2534-47.
65. Lachance J, Tishkoff SA. SNP ascertainment bias in population genetic analyses: why it is important, and how to correct it. BioEssays. 2013;35:780-6.
66. Yurtman E, Özer O, Yüncü E, Dağtaş ND, Koptekin D, Çakan YG, et al. Archaeogenetic analysis of Neolithic sheep from Anatolia suggests a complex demographic history since domestication. Commun Biol. 2021;4:1279.
67. Kaptan D, Atağ G, Vural KB, Morell Miranda P, Akbaba A, Yüncü E, et al. The population history of domestic sheep revealed by paleogenomes. Mol Biol Evol. 2024;41:158.

68. Argyriadou A, Michailidou S, Vouraki S, Tsartsianidou V, Triantafyllidis A, Gelasakis A, et al. A genome-wide association study reveals novel SNP markers associated with resilience traits in two Mediterranean dairy sheep breeds. Front Genet. 2023;14:1294573.
69. Musumeci G, Castrogiovanni P, Szychlinska MA, Aiello FC, Vecchio GM, Salvatorelli L, et al. Mammary gland: from embryogenesis to adult life. Acta Histochem. 2015;117:379-385.
70. Xiao X, Ji Z, Wang T, Zhu D, Li Z, Xu X, et al. Investigation of high fecundity genes by nanopore sequencing in sheep (Ovis aries) pituitary. BMC Genomics. 2025;26:564.
71. Xie Y, Li X, Liang H, Chu M, Cao G, Jiang Y. Integrated multiomic profiling of tail adipose tissue highlights novel genes, lipids, and metabolites involved in tail fat deposition in sheep. BMC Genomics. 2025;26:212.
72. Sánchez-Pozos K, Ortiz-López MG, Peña-Espinoza BI, de los Ángeles Granados-Silvestre M, Jiménez-Jacinto V, Verleyen J, et al. Whole-exome sequencing in Maya indigenous families: variant in PPP1R3A is associated with type 2 diabetes. Mol Genet Genomics. 2018;293:1205-1216.
73. Hao Z, Zhou H, Hickford JG, Gong H, Wang J, Hu J, et al. Transcriptome profile analysis of mammary gland tissue from two breeds of lactating sheep. Genes. 2019;10:781.
74. Ren W, Badgery W, Ding Y, Guo H, Gao Y, Zhang J. Hepatic transcriptome profile of sheep (Ovis aries) in response to overgrazing: novel genes and pathways revealed. BMC Genet. 2019;20:54.

75. Óvilo C, Benítez R, Fernández A, Isabel B, Núñez Y, Fernández AI, et al. Dietary energy source largely affects tissue fatty acid composition but has minor influence on gene transcription in Iberian pigs. J Anim Sci. 2014;92:939-954.
76. Zhu M, Zhang H, Yang H, Zhao Z, Blair HT, Liang H, et al. Targeting GNAQ in hypothalamic nerve cells to regulate seasonal estrus in sheep. Theriogenology. 2022;181:79-88.
77. Gu B, Wang A, Liu H, Liu X, Jiang H. Validation of targeted relationships of novel circRNA803/lncRNA MSTRG.19726–oar-let-7a–CPEB1 ceRNA networks, key to follicle development in single-litter and multi-litter sheep based on whole-transcriptome sequencing. Int J Mol Sci. 2025;26:5161.
78. Wang W, Pang Z, Zhang S, Yang P, Pan Y, Qiao L, et al. Multi-omics integrated analysis reveals the molecular mechanism of tail fat deposition differences in sheep with different tail types. BMC Genomics. 2025;26:465.

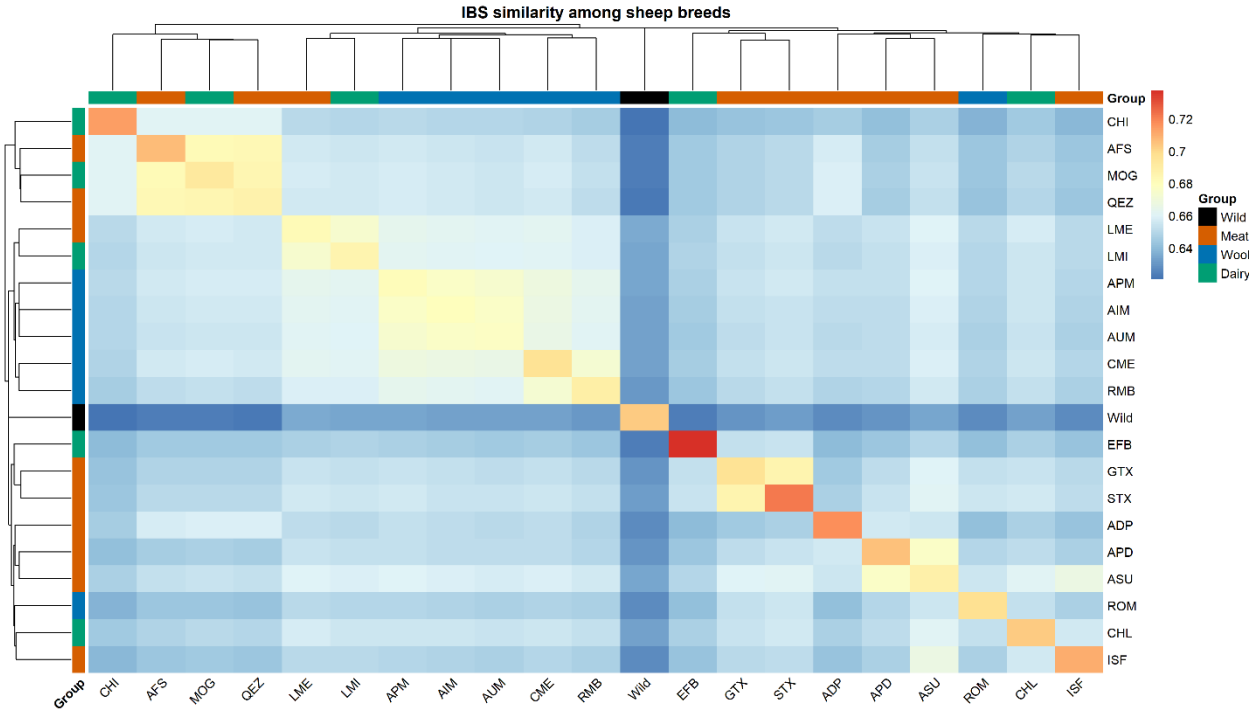


Fig.1. Heatmap plot based on IBS similarity among domestic populations and wild sheep.

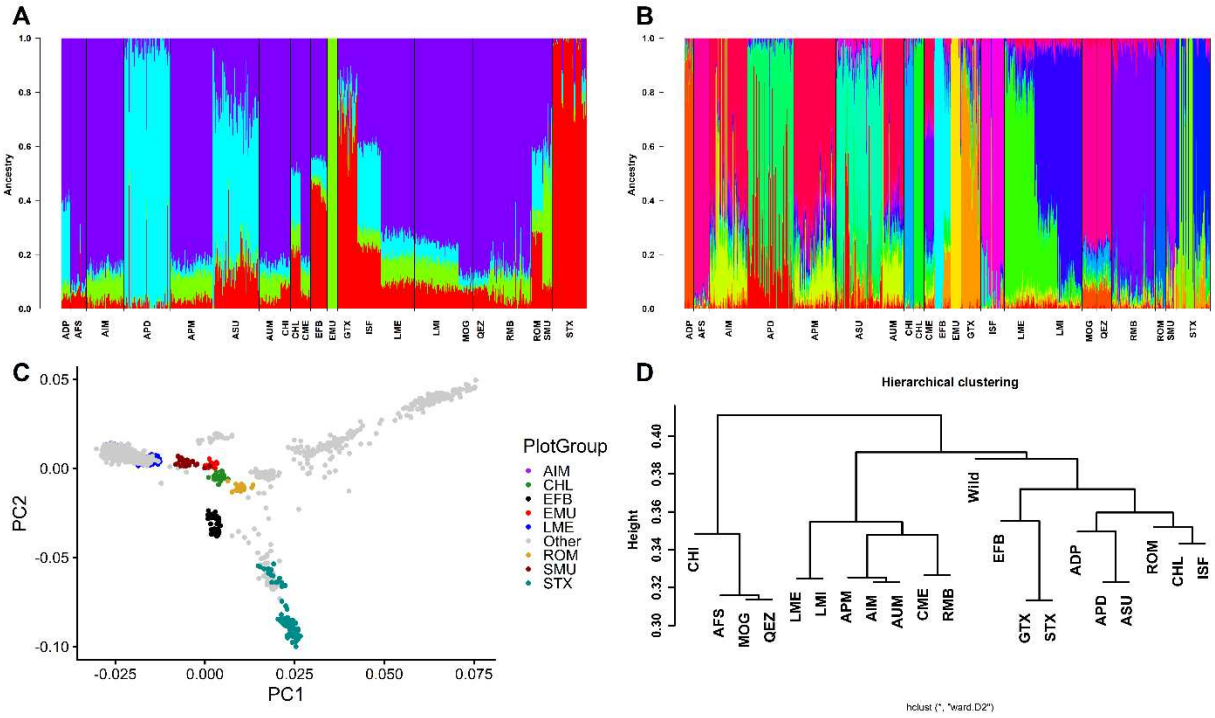


Fig. 2. Population genomic structure and relationships among the 22 sheep populations. (A) ADMIXTURE analysis at K = 4, showing the major ancestral components across the 22 populations. (B) ADMIXTURE analysis at K = 20, providing finer resolution of population-specific ancestry patterns and admixture. (C) Principal component analysis (PCA), illustrating the genomic differentiation and clustering of individuals across the studied populations. (D) Hierarchical clustering of the 22 sheep populations based on genetic distances using the Ward.D2 method.

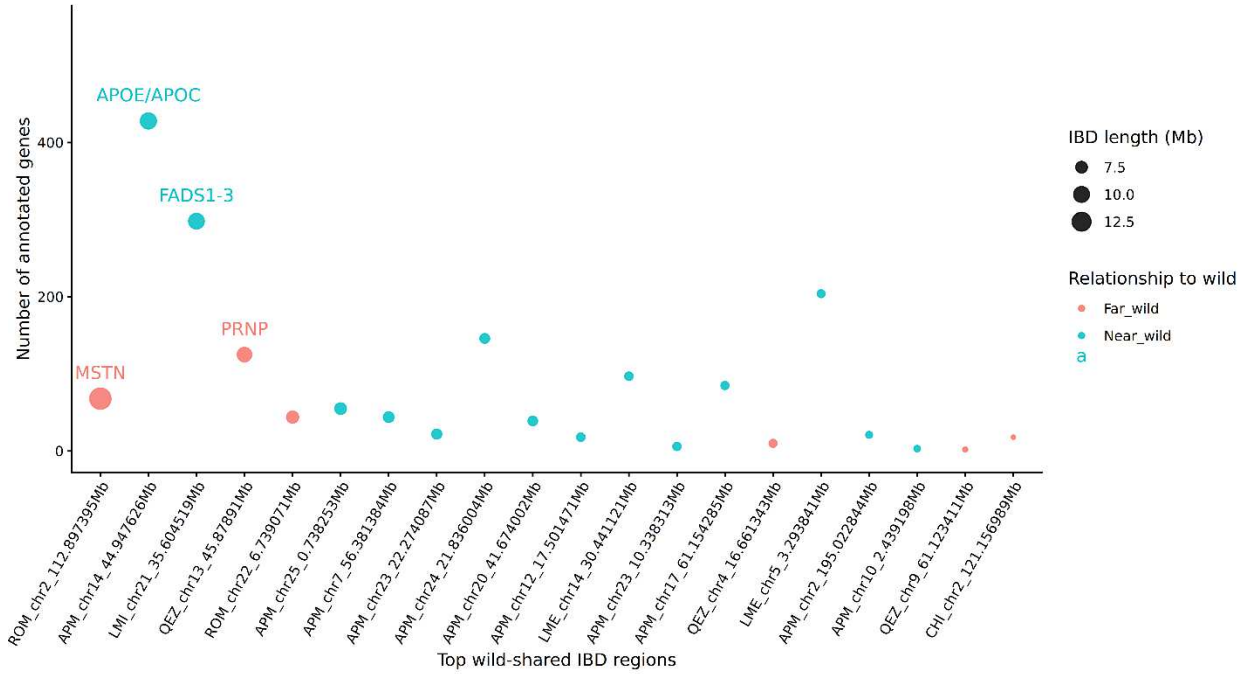
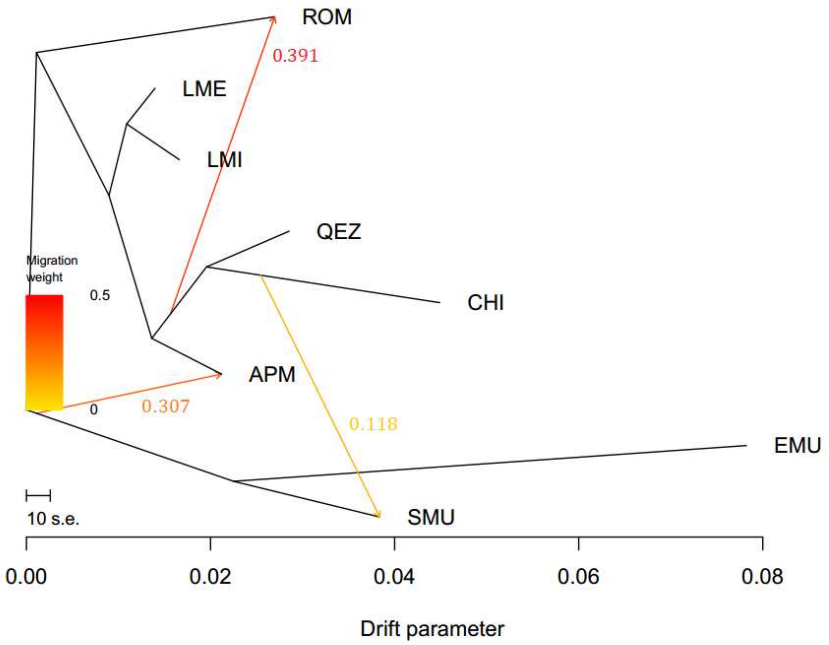


Fig. 3. Top 20 wild-shared IBD regions between domestic sheep breeds and wild sheep, ranked by genomic length. Point size indicates the length of each IBD region (Mb), point position represents the number of annotated genes within the region, and colors distinguish Near-wild and Far-wild breeds. Breed chromosome combinations are shown on the x-axis.



1259

1260 **Fig.4** Maximum-likelihood population graph inferred using TreeMix (m=3).

1261

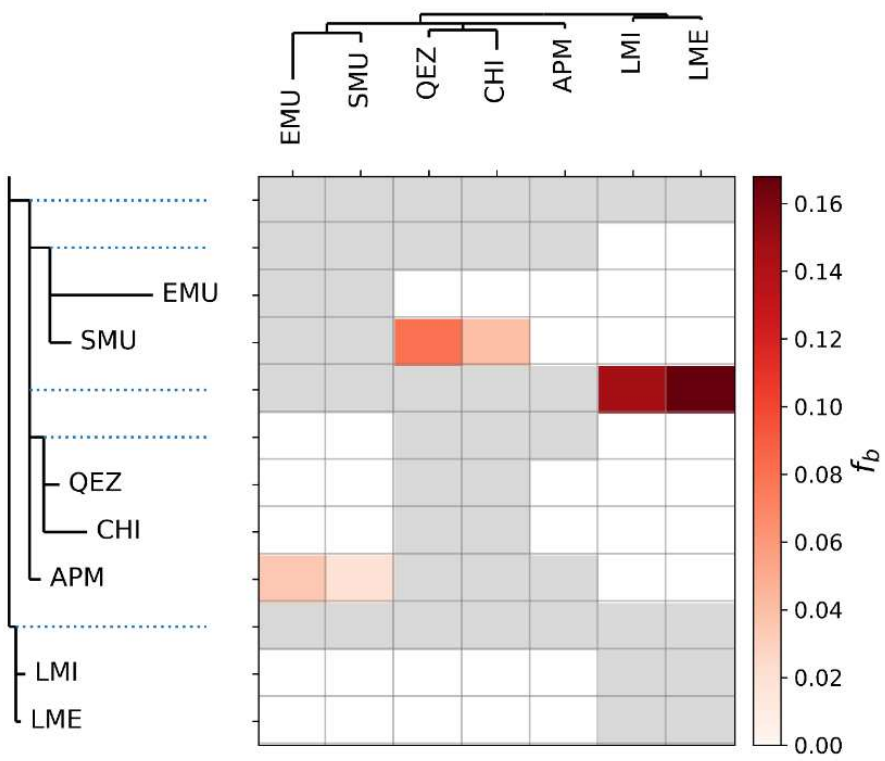


Fig. 5 F-branch analysis illustrating introgression signals among wild and domestic sheep populations based on the inferred population topology.

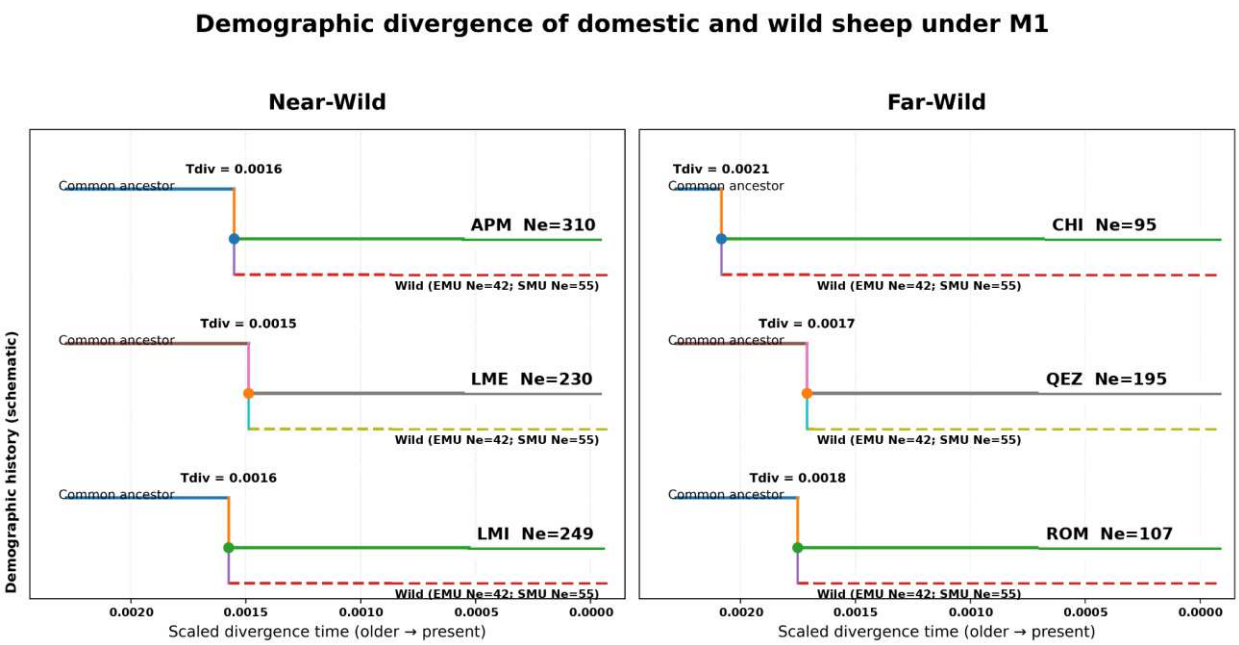


Fig. 6 Demographic divergence of domestic sheep from wild sheep using the split-without-migration model (M1). It illustrates the common ancestral population before the split and the divergence points (Tdiv) between each domestic breed and the wild lineage. Solid horizontal branches represent domestic breeds, while dashed branches indicate the pooled wild lineage (EMU + SMU). Larger Tdiv values signify older inferred divergence.

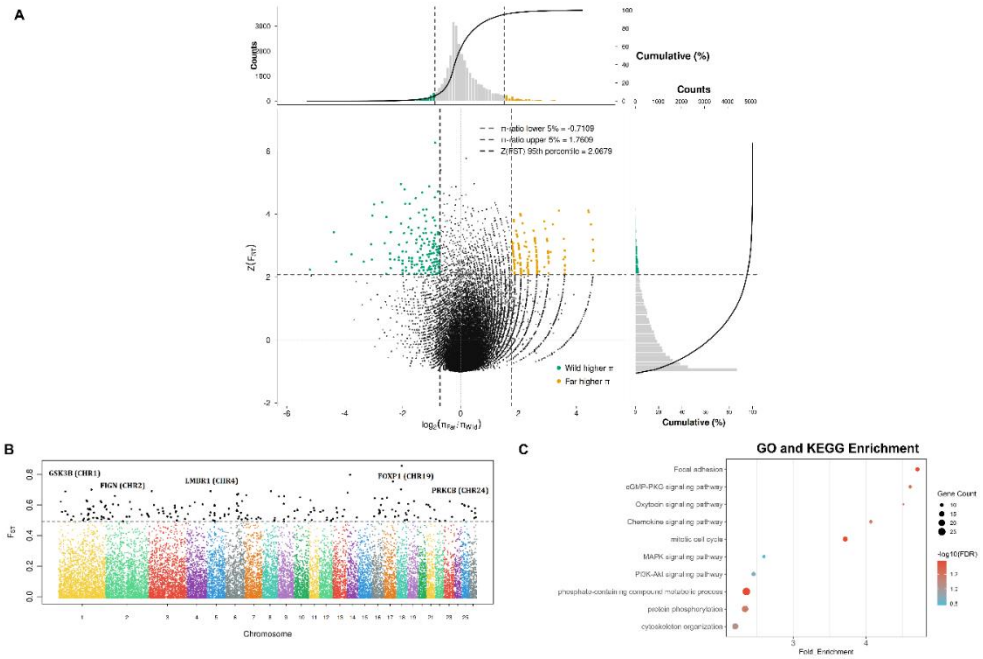
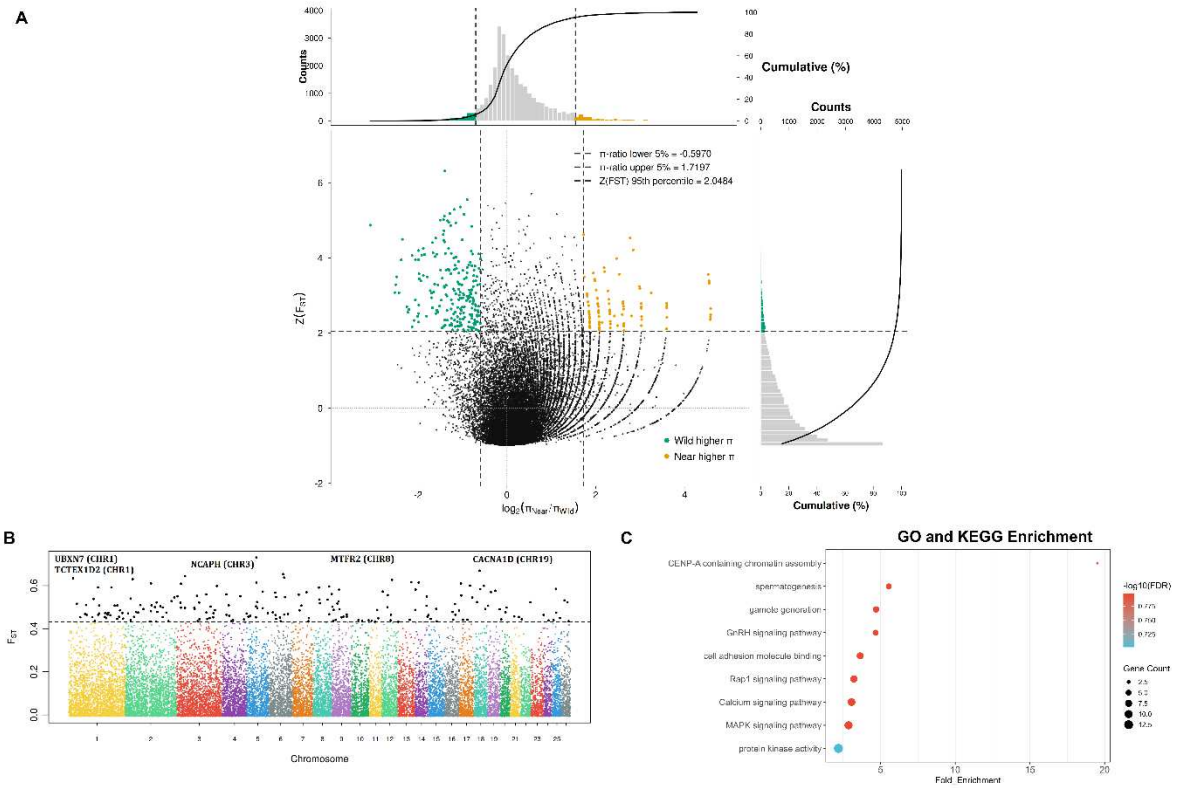


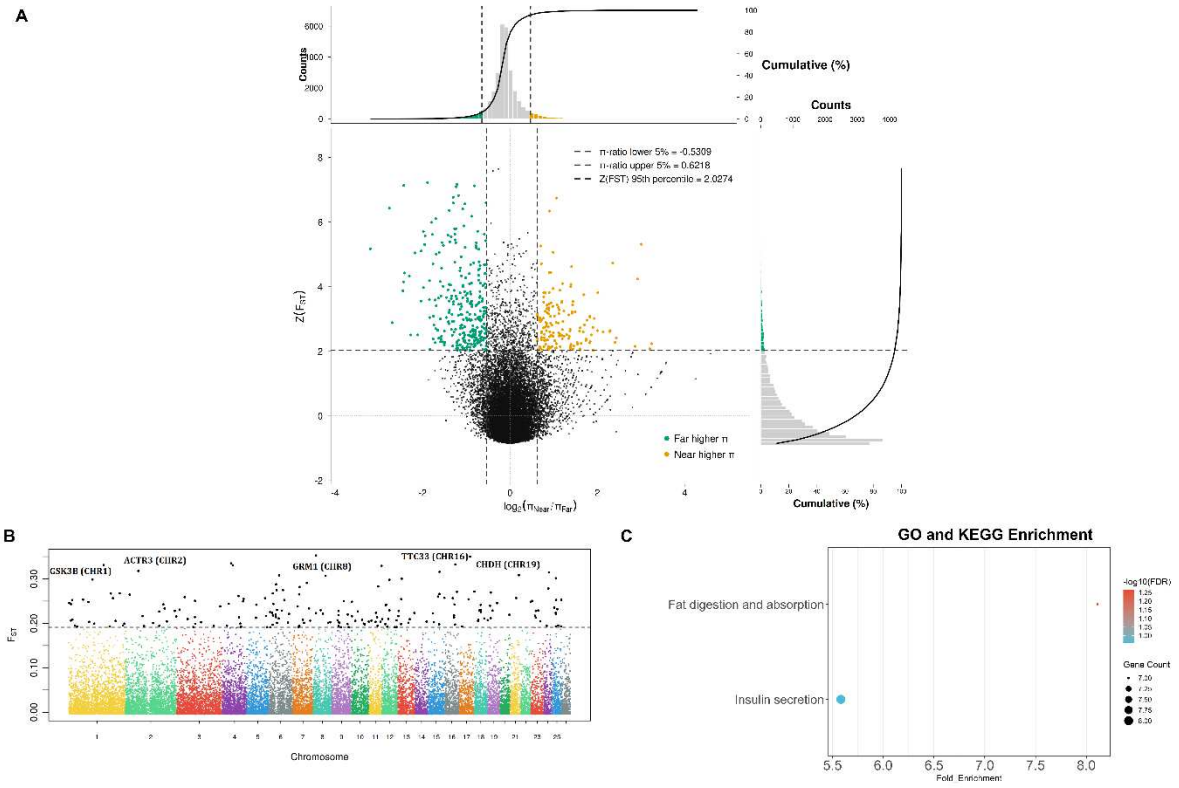
Fig. 7 Genome-wide selection scan comparing far-wild and wild populations. (A) Distribution of π ratios and F_{ST} values, which are calculated in 100 kb windows sliding in 10 kb steps. Data points located to the right of the vertical dashed lines (corresponding to the 5% right tail of the empirical π ratio distribution) and above the horizontal dashed line (the 5% right tail of the empirical F_{ST} distribution) were identified as selected regions (green and orange points). (B) Manhattan plot of F_{ST} values (far-wild vs. wild). The dashed gray line represents the threshold of the top 5% F_{ST} values. (C) GO enrichment analysis of candidate genes under selection between far-wild and wild.



1282

1283 **Fig. 8** Genome-wide selection scan comparing near-wild and wild populations. (A) Distribution of π
1284 ratios and F_{ST} values, which are calculated in 100 kb windows sliding in 10 kb steps. Data points
1285 located to the right of the vertical dashed lines (corresponding to the 5% right tail of the empirical
1286 π ratio distribution) and above the horizontal dashed line (the 5% right tail of the empirical F_{ST}
1287 distribution) were identified as selected regions (green and orange points). (B) Manhattan plot of
1288 F_{ST} values (near-wild vs. wild). The dashed gray line represents the threshold of the top 5% F_{ST}
1289 values. (C) GO enrichment analysis of candidate genes under selection between near-wild and wild.

1290



1291

1292 **Fig. 9** Genome-wide selection scan comparing near-wild and far-wild populations. (A) Distribution
1293 of π ratios and F_{ST} values, which are calculated in 100 kb windows sliding in 10 kb steps. Data
1294 points located to the right of the vertical dashed lines (corresponding to the 5% right tail of the
1295 empirical π ratio distribution) and above the horizontal dashed line (the 5% right tail of the
1296 empirical F_{ST} distribution) were identified as selected regions (green and orange points). (B)
1297 Manhattan plot of F_{ST} values (near-wild vs. far-wild). The dashed gray line represents the threshold
1298 of the top 5% F_{ST} values. (C) GO enrichment analysis of candidate genes under selection between
1299 near-wild and far-wild.

1300

1301

1302

Table 1. Summary of genotype quality control and SNP filtering in different populations.

Quality control	Populations
Total number of animals	1225
Total number of SNPs	50384
Removed animals with GCR< 85%	0
Removed SNP_S with GCR[1] < 90%	13616
Removed SNPs with MAF[2] < 5%	2444
Remained SNPs	34324
Remained animals	1225

1. GCR: Genotype call rate, 2. MAF: Minor allele frequency, 3. HWD: Hardy-Weinberg deviation

Table 2. Genome-wide IBD sharing between wild sheep populations and domestic breeds

Breed	Sample	N-IBD	T-IBD-length (Mb)	M-IBD-length (Mb)	>4 Mb seg	>8 Mb seg	EMU sharing (Mb)	SMU sharing (Mb)	EMU contribution (%)	SMU contribution (%)	IBD Mb/indiv
APM	98	423	1769.24	4.18	210	0	1335.51	433.72	75.5	24.5	18.1
QEZ	35	397	1489.32	3.75	137	10	869.37	619.96	58.4	41.6	42.6
LME	78	163	597.34	3.66	67	0	418.86	178.49	70.1	29.9	7.7
ROM	24	135	553.5	4.1	43	1	316.82	236.67	57.3	42.7	23.1
LMI	103	124	488.54	3.94	60	1	288.46	200.08	59	41	4.7
CHI	23	91	326.58	3.59	32	0	241.01	85.57	73.8	26.2	14.2

Table 3. Significant D-statistic and f4-ratio signals indicating gene flow among wild and domestic sheep populations

P1	P2	P3	D-statistic	Z-score	P-value	f4-ratio
CHI	APM	EMU	0.0140	4.31	1.62×10^{-5}	0.0297
CHI	APM	SMU	0.0051	1.99	0.0461	0.0236
EMU	APM	LME	0.0098	5.02	5.16×10^{-7}	0.1669
EMU	APM	LMI	0.0122	5.49	4.05×10^{-8}	0.1651
QEZ	APM	EMU	0.0193	8.22	$<2.3\times10^{-16}$	0.0404
SMU	APM	LME	0.0098	6.18	6.25×10^{-10}	0.1691
SMU	APM	LMI	0.0091	6.22	4.95×10^{-10}	0.1300
QEZ	APM	SMU	0.0088	3.96	7.60×10^{-5}	0.0400
EMU	CHI	LME	0.0097	3.56	3.68×10^{-4}	0.1686
EMU	CHI	LMI	0.0126	3.91	9.24×10^{-5}	0.1725
EMU	SMU	CHI	0.0101	3.12	1.83×10^{-3}	0.0413
SMU	CHI	LME	0.0098	4.94	7.70×10^{-7}	0.1708
SMU	CHI	LMI	0.0096	4.78	1.73×10^{-6}	0.1377
EMU	QEZ	LME	0.0098	4.08	4.52×10^{-5}	0.1711
EMU	QEZ	LMI	0.0128	4.41	1.04×10^{-5}	0.1769
EMU	SMU	QEZ	0.0118	4.69	2.80×10^{-6}	0.0803
LME	LMI	QEZ	0.0023	2.09	0.0368	0.0211
SMU	QEZ	LME	0.0099	5.24	1.63×10^{-7}	0.1733
SMU	QEZ	LMI	0.0098	4.55	5.45×10^{-6}	0.1424

Note: Significant results were defined as $P < 0.05$. Positive D-statistics indicate excess allele sharing consistent with historical introgression. The largest D-statistic was observed for QEZ–APM–EMU, whereas the highest f4-ratio among the significant tests was observed for EMU–QEZ–LMI.

Additional Files

Additional File 1 Table S1

Format: DOCX

Title: Sheep breeds included in the study according to production type.

Description: Summary of the 22 sheep populations analyzed in this study, detailing their production types, population codes, and sample sizes utilized for genome-wide analyses.

Additional File 1 Figure S1

Format: DOCX

Title: Shared IBD haplotype length between domestic sheep breeds and the EMU and SMU wild lineages.

Description: The stacked bars illustrate the total genomic length of shared haplotypes contributed by each wild lineage across six domestic breeds. Variations in the relative proportions of EMU- and SMU-derived sharing indicate differences in wild-associated genomic components among domestic populations.

Additional File 1 Figure S2

Format: DOCX

Title: Top 20 wild-shared IBD regions ranked by genomic length across six domestic sheep breeds.

Description: Ranking of the largest wild-shared IBD segments among six domestic sheep breeds. The top 20 regions are displayed based on their genomic length (Mb). The size of each point indicates the number of genes contained within each region, while the color represents the relationship category with wild lineages.

Additional File 2 Table S2

Format: .xlsx

Title: Assessment of phasing reproducibility and robustness of downstream IBD detection.

Description: Phase-transition concordance indicates the reproducibility of inferred phase transitions between independent Beagle runs. IBD segment counts and cumulative IBD lengths represent the downstream hap-IBD results obtained from each independent

phasing run. Percentage differences quantify the variation between the two replicate analyses.

Additional File 2 Table S3

Format: .xlsx

Title: Top shared IBD regions between domestic and wild sheep populations, with genomic coordinates, length, and annotated genes.

Description: Summary of the key shared IBD regions identified between domestic and wild sheep populations, including their genomic coordinates, region length, and annotated genes.

Additional File 2 Table S4

Format: .xlsx

Title: Evaluation of TreeMix models with increasing numbers of migration events for inference of historical gene flow**.**

Description: Analysis of TreeMix models with varying numbers of migration events, assessed through log-likelihood, AIC, and the relative improvement in model fit, helps determine the optimal number of migration edges for inferring historical gene flow.

Additional File 2 Table S5

Format: .xlsx

Title: Sensitivity of TreeMix migration inference to rooting and SNP ascertainment.

Description: Sensitivity analysis of TreeMix migration inference to alternative rooting schemes and SNP ascertainment, comparing migration weights across analyses to assess the robustness of the inferred migration events and the potential effect of OvineSNP50 ascertainment bias on migration estimates.

Additional File 2 Table S6

Format: .xlsx

Title: Bootstrap assessment of model stability and log-likelihood variation for the TreeMix model with three migration edges (m = 3)**.**

Description: Summary of the bootstrap replicate success and failure, along with log-likelihood variation for the TreeMix model featuring three migration edges (m = 3), is utilized to assess model stability across bootstrap replicates.

Additional File 1 Figure S3

Format: DOCX

Title: D-statistic values for domestic sheep population pairs in relation to EMU and SMU wild sheep lineages.

Description: D-statistic values for population quartets involving EMU and SMU wild sheep lineages are presented. Each point on the graph represents the D-statistic value calculated for a pair of domestic sheep populations. Filled circles indicate significant results ($P < 0.05$), while open circles represent non-significant results ($P \geq 0.05$).

Additional File 2 Table S7

Format: .xlsx

Title: Complete D-statistic, f4-ratio, and allele-sharing results for all tested population quartets**.**

Description: Results of the comprehensive D-statistic, f4-ratio, and allele-sharing analyses for all assessed population quartets are included, along with the relevant statistical estimates and significance levels used to evaluate historical gene flow and introgression.

Additional File 1 Figure S4

Format: DOCX

Title: Genome-wide patterns of local introgression inferred from sliding-window f_d-statistics across domestic sheep populations.

Description: The sliding-window f_d statistics across the sheep genome for various population topologies are illustrated in the plots. They display the genome-wide distribution of local f_d-values for four different topologies: (A) APM–CHI–LME, (B) APM–QEZ–LME, (C) APM–QEZ–ROM, and (D) LMI–QEZ–APM. Dashed lines indicate the selected threshold for identifying genomic windows with elevated f_d-values.

Additional File 2 Table S8

Format: .xlsx

Title: Comparison of demographic models for domestic breed–wild comparisons estimated using moments**.**

Description: M1 represents a two-population split-without-migration model, whereas M2 represents a split model with asymmetric migration. Tdiv denotes the scaled divergence time; m12 and m21 denote the directional migration parameters; and ΔAIC represents the difference in AIC between M2 and M1. Positive ΔAIC values indicate that M2 has lower AIC support than M1.

Additional File 2 Table S9

Format: .xlsx

Title: Current effective population size and divergence time estimates for Near-Wild and Far-Wild domestic sheep breeds relative to the wild populations EMU and SMU**.**

Description: Estimates of the current effective population size and divergence time for Near-Wild and Far-Wild domestic sheep populations, along with the effective population size estimates for the wild populations EMU and SMU.

Additional File 2 Table S10

Format: .xlsx

Title: Top 30 prioritized candidate genes in the Far-Wild vs Wild comparison.

Description: Summary of the top 30 prioritized candidate genes identified in the comparison between Far-Wild and Wild, including their corresponding QTLs and supporting evidence from QTL, PPI, MCODE, and STRING analyses.

Additional File 2 Table S11

Format: .xlsx

Title: Top 30 prioritized candidate genes in the Near-Wild vs Wild comparison.

Description: Summary of the top 30 prioritized candidate genes identified in the comparison between Near-Wild and Wild, including their corresponding QTLs and supporting evidence from QTL, PPI, MCODE, and STRING analyses.

Additional File 2 Table S12

Format: .xlsx

Title: Top 30 prioritized candidate genes in the Far-Wild vs Near-Wild comparison.

Description: Summary of the top 30 prioritized candidate genes identified in the comparison between Far-Wild and Near-Wild, including their corresponding QTLs and supporting evidence from QTL, PPI, MCODE, and STRING analyses.

Additional File 2 Table S13

Format: .xlsx

Title: Pooled Near–Far outlier regions supported by matched production-class pairwise comparisons.

Description: Candidate Near–Far outlier regions supported by pooled and matched pairwise production-class comparisons. Coordinates define the genomic intervals; Pooled_NearFar_windows indicates the number of supporting pooled windows, pairwise_support lists the supporting breed comparisons, and N_genes/Genes report the annotated genes within each region.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile1.docx
- AdditionalFile2.xlsx