Iranian Society of Animal Science

Journal of Livestock Science and Technologies

Shahid Bahonar University of Kerman

ISSN: 2322-3553 (Print) ISSN: 2322-374X (Online)

Paper type: Original Research

Comparative analysis of genomic diversity and genetic architecture amongst 14 horse breeds using ultra high-density SNP arrays: Implications for conservation and breeding

Moslem Moghbeli Damane[1,2], Ahmad Ayatollahi Mehrgardi[1], Ali Esmailizadeh[1]*

[1]Department of Animal Science, Faculty of Agriculture, Shahid Bahonar University of kerman, Kerman, Iran
[2]Department of Animal Science, Faculty of Agriculture, University of Jiroft, Jiroft, Iran

[*]Corresponding author,
E-mail address:
m.moghbeli@agr.uk.ac.ir

Received: 22 Nov 2025,
Received in revised form: 26 Jan 2026,
Accepted: 02 Feb 2026,
Published online: 03 Feb 2026,

ORCID
Moslem Moghbeli Damane
0000-0002-1853-8103
Ahmad Ayatollahi Mehrgardi
0000-0001-5278-8459
Ali Esmailizadeh
0000-0003-0986-6639

Abstract Genetic variability is fundamental for the adaptability, health, and long-term sustainability of horse populations. This study sought to provide a comprehensive assessment of genomic diversity, population structure, and selection signatures across 14 horse breeds, encompassing five indigenous Iranian breeds (Iranian Arabian-Asil, Turkmen, Caspian, Kurdish, Dareshuri) as well as nine international breeds (Thoroughbred, Quarter Horse, Belgian, Arabian, Welsh Pony, Standardbred, Icelandic, Lusitano, Morgan), based on ultra-high-density SNP array data. Following quality control and imputation, we analyzed the combined dataset of 1.57 million SNPs across 438 individuals to assess within- and between-breed genetic diversity, population structure, and signatures of selection. Observed heterozygosity (H_o) ranged from 0.1727 (Thoroughbred) to 0.2068 (Welsh), with Iranian breeds generally demonstrating lower allelic richness and effective number of alleles. Pairwise F_{ST} and private allele analyses revealed moderate genetic differentiation, with Thoroughbred and Icelandic being the most divergent, while Iranian breeds revealed high shared ancestry. Population history assessments via Tajima's D and Garza–Williamson M-ratio highlighted varying degrees of past bottlenecks, most evident in Dareshuri and Kurdish populations. Population structure analyses (PCA and PCoA) showed well-defined breed clusters, highlighting both geographic origin and selective breeding history. HapFLK analysis detected 83 candidate genes under selection, associated with neural development, immune response, coat color and traits related to performance or reproduction. These findings emphasize the combined effects of historical selection, demographic events, and breed management on the genomic architecture of horse populations. The results provide applicable information for conservation strategies and breeding programs aimed at maintaining genetic diversity while preserving breed-specific traits, particularly in endangered Iranian and intensively selected international breeds.

Keywords: conservation genetics, genomic diversity, horse breeds, population structure, selection signatures, SNP array

Introduction

Genetic diversity is an important factor in genetic sustainability, as well as for the ability of livestock populations to respond and adapt to changes in their environment, and selection pressure. For a species like the horse, with significant historical, cultural, and economic

importance to human societies, preserving genetic diversity is valuable from both conservation, and breeding-performance considerations (Petersen et al., 2013; Bazvand et al., 2024). The decline in horse populations, uncontrolled admixture with foreign breeds, and the widespread adoption of breeding practices have caused a significant decrease in genetic diversity in many indigenous and working horse breeds, accompanied by a decrease in individual heterozygosity, and autosomal nucleotide diversity (Pi) (Fages et al., 2019).

During the twentieth century, horse breeding experienced substantial changes, leading to a marked decline in the effective population size (Ne) of many draft and indigenous breeds (Leroy et al., 2011; Petersen et al., 2013; Bazvand et al., 2024). Although *in situ* conservation measures have been implemented for some populations, the prioritization of conservation efforts requires a comprehensive understanding of genetic structure and interbreeds relationships (Leroy et al., 2009). Insights derived from such genomic studies can assist breeders and relevant institutions in identifying breeds and populations that require enhanced protection and in preventing the accumulation or fixation of deleterious alleles (Fages et al., 2019).

In previous studies, molecular markers (e.g., microsatellites, mitochondrial DNA) were applied to evaluate genetic diversity in livestock populations (Amirinia et al., 2007; Hedayat Evrigh et al., 2018; Amjadi et al., 2021). The emergence of single nucleotide polymorphism (SNP) markers, particularly in the form of high-density genotyping arrays, has enabled a deeper and more reliable evaluation of genome-wide autosomal genetic diversity (Moravčíková et al., 2024). High-density SNP arrays, due to their extensive genomic coverage, represent key and valuable tools in genomic studies, facilitating rapid and high-throughput genotyping. They are essential for a wide range of genetic analyses, including genomic selection and high-resolution population genomics (Ventura et al., 2020; Xuereb et al., 2023). This technology is not only suitable for evaluating within- and between-breed diversity, but also allows for the identification of genomic regions under natural or artificial selection so-called selection signatures which can elucidate the genetic basis of traits such as athletic performance, disease resistance, and environmental adaptability (Hu et al., 2023).

Previous studies, as reported by Schaefer et al. (2017) and Durward-Akhurst et al. (2021) provided evidence that SNP-based data constitute a powerful tool for analyzing genetic diversity across various horse breeds. Related studies in other livestock species, including cattle, sheep, and pigs have revealed genomic regions under selection, providing evidence for the power of high-density SNP arrays to identify important functional variants (Hu et al., 2023). However, most of these studies have been limited to a small number of breeds or low-density panels, and large-scale patterns of genomic diversity remain insufficiently characterized (Atroshchenko et al., 2023; Abdoli et al., 2024; Bazvand et al., 2024).

In the current study, we used ultra-high density SNP array data to conduct a comparative analysis of genomic diversity and genetic structure within 14 horse breeds. These breeds exhibit considerable variation in historical origin, phenotypic traits, and functional applications. Given that the power of individual methods to detect genomic regions under selection can vary depending on population history, type of selection, and marker density (Chang et al., 2019; Temple et al., 2024), combining multiple approaches such as F_{ST}, hapFLK, and heterozygosity-based methods provides a more robust framework for assessing genetic diversity and identifying candidate regions under selection. The main objective of this study was to conduct a comprehensive comparative genomic analysis of 14 horse breeds, integrating multiple approaches to assess genetic diversity, population structure, and regions under selection, with the ultimate goal of informing sustainable breeding programs and conservation strategies for at-risk horse breeds.

Materials and methods

Dataset and quality control

Two datasets were used in this study. The first dataset consisted of five Iranian breeds [Iranian Arabian (Asil) (n = 71), Turkmen (n = 11), Caspian (n = 7), Kurdish (n = 7), and Dareshuri (n = 5)] that were genotyped using the Axiom™ Equine Genotyping Array 670K (MNEc670k) (Sadeghi et al., 2019). The second dataset comprised of nine international (U.S.) breeds [Thoroughbred (n = 28), Quarter Horse (n = 72), Belgian (n = 22), Arabian (n = 36), Welsh Pony (n = 40), Standardbred (n = 39), Icelandic (n = 18), Lusitano (n = 21), and Morgan (n = 61)] that were genotyped using the Axiom™ Equine Genotyping Array MNEc2M (Momen et al., 2023).

To increase the SNP density, genotypes from the 670K array were imputed to 2M density using Beagle v5.5 (Browning et al., 2018), with the MNEc2M dataset serving as the reference panel and an effective population size (Ne) of 1000. The combine dataset consisted of an ultra-high-density integrated panel combining the imputed 670K and original MNEc2M data.

Quality control was carried out using PLINK v1.90 (Chang et al., 2015), excluding SNPs with a minor allele frequency (MAF) < 0.01, deviations from Hardy–Weinberg equilibrium (P-value < 1×10^{-6}), or a call rate ≤ 80%, and retaining only biallelic SNPs on 31 autosomes. After quality control, the combine dataset contained 1,568,289 SNPs from 438 horses, representing the same 14 breeds.

Within-breed genetic diversity

Observed and expected heterozygosity (H_o, H_e)

Observed (H_o) and expected (H_e) heterozygosities were calculated to assess within-breed genetic diversity. Basic diversity indices were calculated in PLINK v1.9. Individual observed heterozygosity was derived using the --het function, while expected heterozygosity was calculated in R as $H_e = 2p(1 - p)$, where p represents the minor allele frequency at each SNP.

Allelic richness and effective number of alleles

Given that haplotype-based analyses capture the combined variation of linked genomic regions and more accurately reflect true allelic richness than single-SNP data, allelic richness (AR) and the effective number of alleles were therefore estimated from haplotype-based data to evaluate within-breed genetic diversity. Haplotype blocks were delineated across all autosomes by applying PLINK v1.9 based on the criteria of Gabriel et al. (2002), with a maximum block length of 200 kb.

Nucleotide diversity (Pi)

Nucleotide diversity (Pi) denotes the mean number of nucleotide differences per site between any two sequences randomly sampled from a population (Nei and Li, 1979). Evaluating nucleotide diversity provides valuable information about population differentiation and can be utilized to infer the demographic history and past effective population size of a species (Yu et al., 2004). Nucleotide diversity (Pi) was calculated in sliding windows of 50 kb with a step size of 25 kb using the --window-pi command in VCFtools v0.1.17 (Danecek et al., 2011). The mean Pi value for each breed was then obtained by averaging Pi values across all genomic windows.

Between-breed genetic differentiation

Pairwise F_{ST} value

Pairwise genetic differentiation among the 14 horse breeds was estimated using Weir and Cockerham's (1984) F_{ST} implemented in VCFtools v0.1.17. Breed-specific sample lists were generated and used in all pairwise comparisons (91 combinations) with the --weir-fst-pop option.

Private alleles analysis (private allele count and frequency)

Private alleles (PAs), described as alleles occurring exclusively within a single breed with an allelic frequency greater than 0.1% (Machmoum et al., 2020), were identified across all 14 horse breeds. Alleles present in only one breed were classified as private, and their counts were summarized per breed.

Demographic history and bottleneck detection

Tajima's D statistic

Tajima's D was calculated separately for each breed using VCFtools v0.1.17. Breed-specific individual lists were extracted from the VCF file headers, and the analysis was performed with the --TajimaD option in non-overlapping 50 kb sliding windows across the autosomes.

Garza–Williamson M-ratio

Historical reductions in genetic diversity were assessed by applying the Garza–Williamson M-ratio (Garza and Williamson, 2001) for each breed, using R (R Core Team, 2023). Because SNPs are biallelic, absolute M-ratio values were typically close to 1, which limits their interpretative power. To enhance the informativeness of this metric, haplotype-based analyses were subsequently conducted, allowing a more robust assessment of historical bottlenecks.

Population structure and genetic relationships

Principal Component Analysis (PCA) and Principal Coordinates Analysis (PCoA)

Principal component analysis (PCA) was performed to investigate the major axes of genetic variation among individuals from the 14 horse breeds. High-density SNP genotypes in VCF format were converted to GDS using the SNPRelate R package (Zheng et al., 2012) to enable efficient computation. To minimize the effect of linkage disequilibrium (LD), SNPs were pruned using an LD threshold of $r^2 < 0.2$ (Santos et al., 2021; Yordanov et al., 2025), retaining approximately independent loci for analysis. Following LD pruning, a total of 303,931 SNPs were retained from the initial dataset for downstream analyses. PCA was then conducted on the pruned SNP set to explore the overall genomic structure and relationships among individuals. The resulting principal components were visualized in R, with samples color-coded by breed.

Principal coordinates analysis (PCoA) was performed to further evaluate genetic relationships among the horse breeds based on pairwise genetic distances. Using the same GDS dataset, pairwise identity-by-state (IBS) similarity matrices were computed in SNPRelate, and genetic distances were obtained as 1 – IBS (Cortes-Rodriguez et al., 2019; Ospina et al., 2024). PCoA was then carried out on the resulting distance matrix to visualize breed-level and individual-level clustering patterns. Additionally, mean pairwise genetic distances among breeds were summarized and visualized as a heatmap to highlight genomic divergence at the population level.

Analysis using hapFLK

To identify genomic regions potentially subjected to selection across the studied horse breeds, we employed the hapFLK v1.4 program to detect signatures of selection through haplotype differentiation among hierarchically structured populations, following the approach of Fariello et al. (2013). High-density SNP

genotype data were first phased using Beagle v5.4, and the resulting haplotypes were used as input for the hapFLK analyses. All 14 breeds were included, with the Icelandic breed defined as the outgroup based on pairwise F_{ST} estimates and PCA-derived genetic distances. hapFLK analyses were conducted separately for each autosome using K = 5, 10, 15, 20, and 25 haplotype clusters to evaluate model stability. Each run involved 20 expectation–maximization (EM) iterations for fitting the linkage disequilibrium model (–nfit = 20). A Reynolds distance matrix based on 10,000 SNPs was computed to account for population kinship structure during the analysis.

Results

Within-breed genetic diversity

Within-breed genetic diversity indices including observed heterozygosity (H_o), expected heterozygosity (H_e), allelic richness (AR), effective number of alleles (NE), mean minor allele frequency (MAF), and nucleotide diversity (Pi) were estimated for all 14 horse breeds (Table 1). Observed heterozygosity ranged from 0.1727 in Thoroughbred to 0.2068 in Welsh (overall mean = 0.193), showing a similar pattern to expected heterozygosity, which ranged from 0.1677 in Kurdish to 0.2092 in Quarter Horse. In Iranian breeds such as Iranian Arabian (Asil) and Caspian, the lowest heterozygosity values were observed, and similarly, among international breeds, Thoroughbred, Standardbred, and Arabian also showed low levels of heterozygosity.

MAF values varied across breeds, with the lowest estimates in Quarter Horse (0.1462) and Morgan (0.1487), and the highest in Dareshuri (0.2029) and Kurdish (0.2014). Allelic richness also varied markedly: Quarter Horse (9.14) and Morgan (8.80) showed the highest AR, whereas Dareshuri (2.85) and Kurdish (3.02) had the lowest values.

Effective number of alleles followed the same trend, reflecting larger genetic variation in worldwide breeds. Nucleotide diversity showed a narrow range across breeds (0.000123-0.000147), with higher values again observed in Quarter Horse and Morgan compared with most native Iranian breeds.

Table 1. Within-breed genomic diversity parameters in 14 horse breeds

Breed	n	H_O[1]	H_E[2]	NE[3]	AR[4]	MAF[5]	Pi[6]
Arabian	36	0.1851	0.1883	3.88	6.04	0.1594	0.0001335
Belgian	22	0.1965	0.1944	3.69	5.26	0.1604	0.0001391
Caspian	7	0.1888	0.1790	2.87	3.37	0.1911	0.0001351
Dareshuri	5	0.1916	0.1736	2.55	2.85	0.2029	0.0001353
Icelandic	18	0.1939	0.1882	3.44	4.70	0.1723	0.0001354
Kurdish	7	0.1963	0.1677	2.53	3.02	0.2014	0.0001266
Lusitano	21	0.1947	0.1960	3.89	5.42	0.1651	0.0001404
Morgan	61	0.2016	0.2089	5.04	8.80	0.1487	0.0001472
Iranian Arabian (Asil)	71	0.1797	0.1812	4.16	7.25	0.1495	0.0001277
Quarter Horse	72	0.2067	0.2092	5.06	9.14	0.1462	0.0001472
Standardbred	39	0.1857	0.1934	4.10	6.75	0.1619	0.0001370
Thoroughbred	28	0.1727	0.1735	3.14	4.54	0.1753	0.0001235
Turkmen	11	0.1952	0.1857	3.37	4.23	0.1754	0.0001362
Welsh	40	0.2068	0.2063	4.58	7.12	0.1480	0.0001460

[1]Observed heterozygosity; [2]Expected heterozygosity; [3]Effective number of alleles; [4]Allelic richness; [5]Minor allele frequency; [6]Nucleotide diversity

Genetic differentiation and private alleles

Pairwise F_{ST} estimates revealed a generally moderate level of genetic differentiation among the 14 studied horse breeds (Figure 1). The mean F_{ST} values ranged from 0.013 (Dareshuri-Turkemen) to 0.124 (Thoroughbred-Icelandic), while the weighted F_{ST} values showed a similar pattern, ranging from 0.017 (Dareshuri-Iranian Arabian (Asil)) to 0.182 (Thoroughbred-Icelandic). These results indicate that the Thoroughbred and Icelandic breeds are the most genetically distinct pair, whereas Dareshuri and Turkemen exhibit the closest genetic relationship.

Overall, lower F_{ST} values among several Iranian breeds (Caspian, Dareshuri, Kurdish, Turkemen, and Iranian Arabian-(Asil)) suggest relative genetic similarity, likely reflecting shared ancestry, while higher values between these breeds and international breeds (such as Thoroughbred and Standardbred) reflect long-term divergence and breed-specific selection histories.

The analysis of private alleles showed strong variation among breeds. Morgan (2,835), Standardbred (1,163), and Welsh (729) had the highest numbers of private alleles (Figure 1). No private alleles were detected in the Iranian breeds; however, this observation likely reflects limitations of the SNP bead-chip used, which was not specifically designed for these populations and may not capture breed-specific variants. Consistent with this pattern, F_{ST} values were generally lower among Iranian breeds compared to the other breeds.

Historical Demography and Genetic Bottleneck Signatures

Genome-wide Tajima's D values were calculated in non-overlapping 50 kb windows for each of the 14 horse breeds to assess historical demographic dynamics

(Figure 2). The mean Tajima's D values were generally positive across breeds, ranging from 0.085 in Dareshuri to 0.901 in Thoroughbred, suggesting an overall excess of intermediate-frequency alleles consistent with population contraction or balancing selection. The Thoroughbred, Standardbred, Iranian Arabian (Asil), and Quarter Horse showed the highest mean Tajima's D values (> 0.7), indicating stronger signals of reduced recent effective population size, possibly linked to intensive selective breeding and closed studbook practices. In contrast, Caspian, Dareshuri, and Turkemen horses exhibited lower Tajima's D values (< 0.25), suggesting relatively more stable demographic histories or weaker bottleneck effects. The distribution of Tajima's D values was wide in all breeds (min ≈ −2.6 to max ≈ 4.3), reflecting regional variation across the genome in selective and demographic pressures.

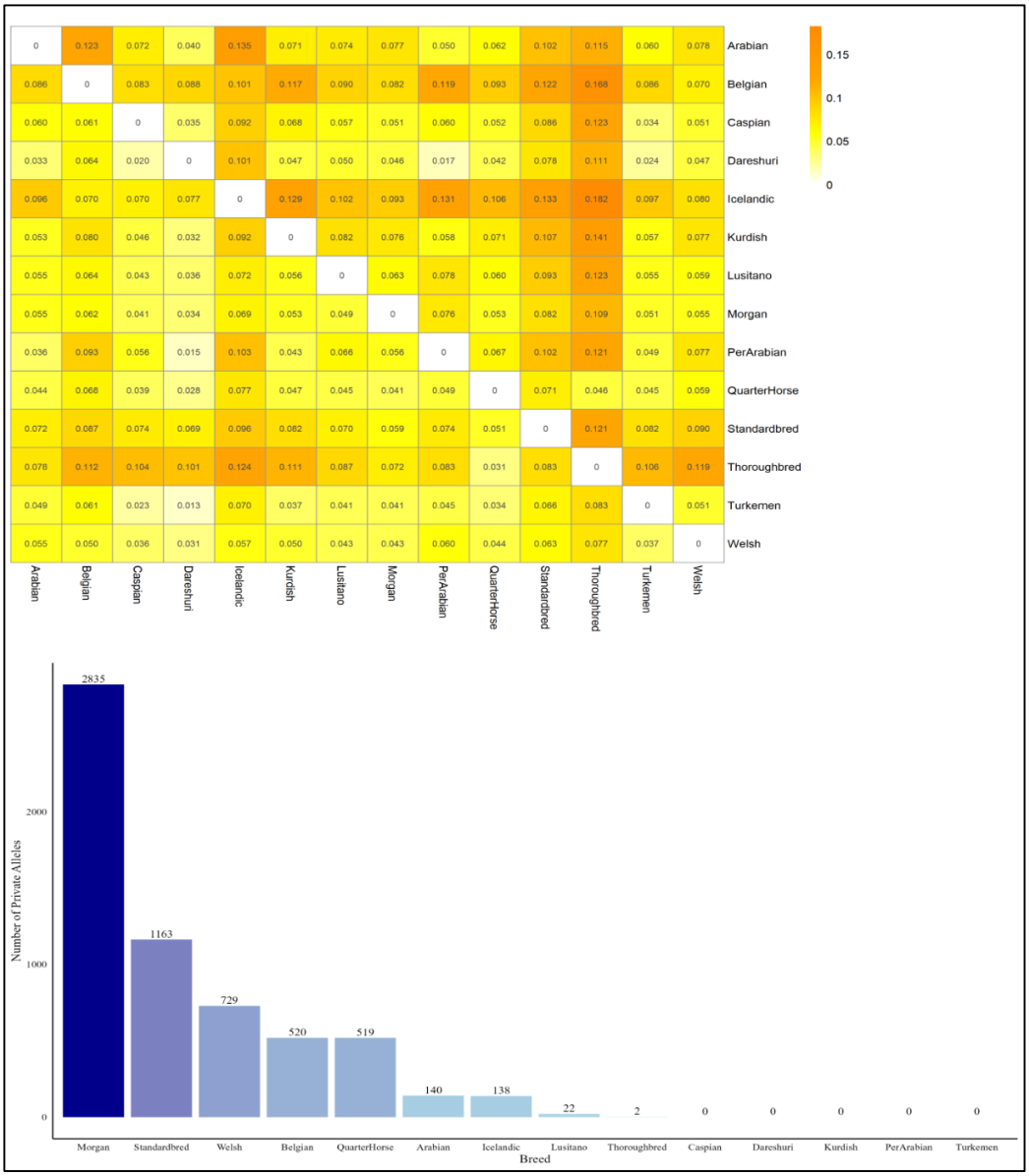

Figure 1. Comparative visualization of pairwise genetic differentiation among horse breeds (upper panel: weighted F_{ST} above and mean F_{ST} below the diagonal) and the distribution of private alleles per breed (lower panel)

Estimates of the Garza–Williamson M-ratio ranged from 0.711 in Dareshuri to 0.832 in Quarter Horse, with most breeds exhibiting mean M values between 0.75 and 0.82. These relatively high ratios are consistent with the expectation for SNP-based data but still suggest moderate historical reductions in genetic variation, particularly in Dareshuri, Kurdish, and Caspian populations. The highest M-ratio values observed in Morgan, Quarter Horse, and Iranian Arabian (Asil) breeds may reflect either recent recovery following

historical bottlenecks or maintenance of higher allelic diversity through larger contemporary effective population sizes. Overall, the combined Tajima's D and M-ratio results indicate that while most breeds show genomic evidence of past demographic contractions, the magnitude and genomic extent of these events vary substantially among breeds.

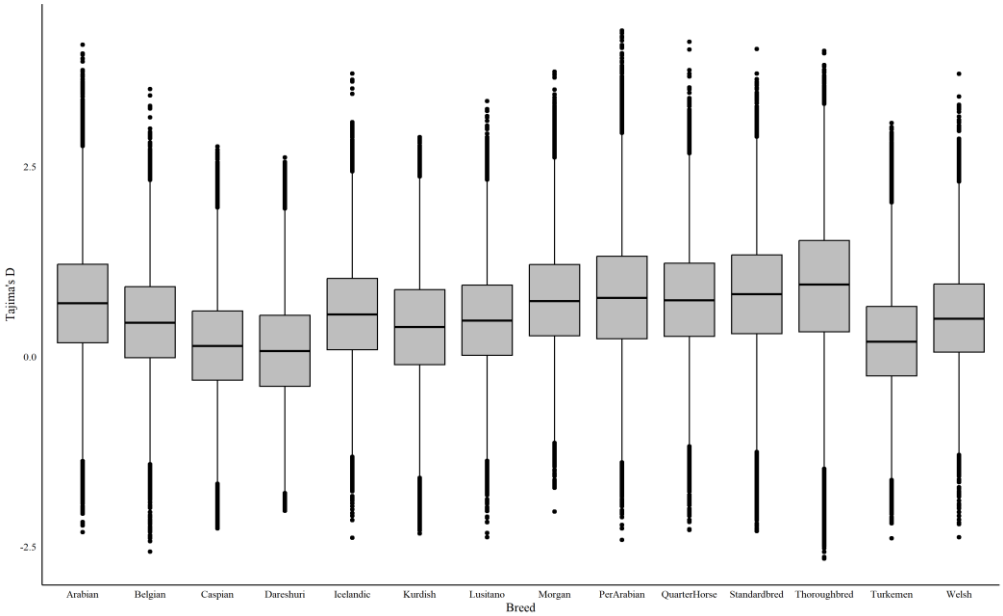

Figure 2. Distribution of Tajima's D amongst 14 horse breeds

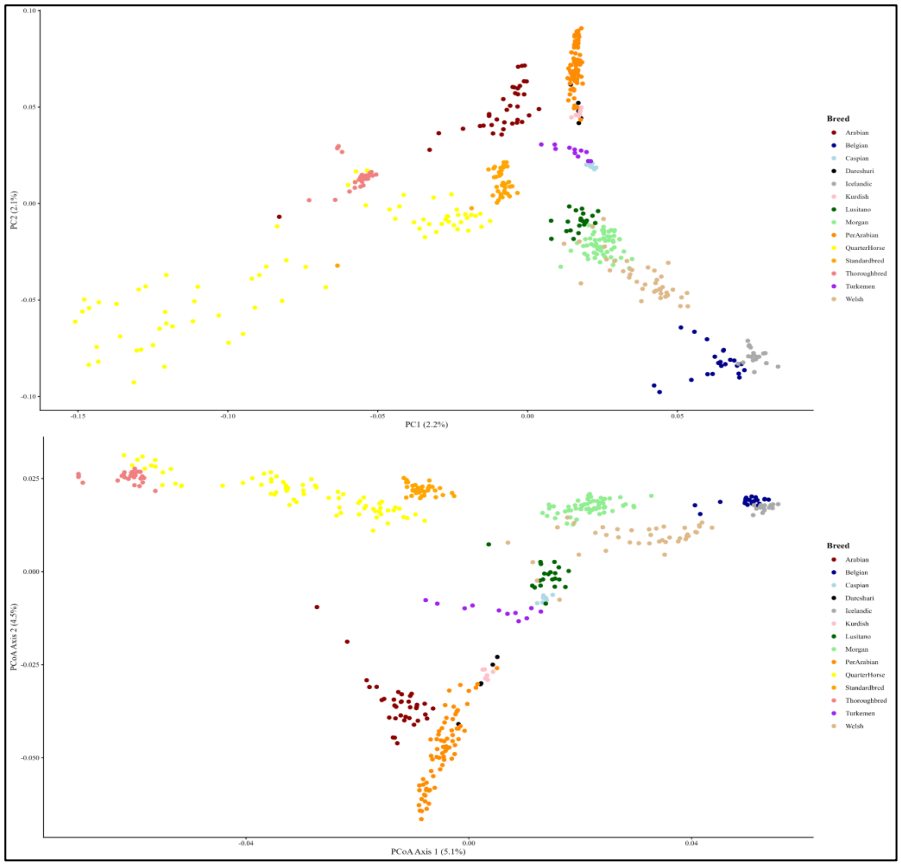

Figure 3. Principal component analysis (PCA, top) and principal coordinate analysis (PCoA, bottom) showing population structure and genetic relationships amongst 14 horse breeds

Population structure and genetic relationships

Principal component analysis (PCA) revealed clear patterns of genomic differentiation among the 14 horse breeds (Figure 3). The first two principal components explained 2.2% (PC1) and 2.1% (PC2) of the total genetic variation. Several breeds, including Belgian and Icelandic, formed distinct clusters, indicating strong genetic differentiation. Thoroughbred, Iranian Arabian (Asil), and Standardbred were closely positioned but still separated, reflecting moderate genetic similarity possibly due to shared breeding histories. Conversely, breeds such as Quarter Horse and Morgan displayed some overlap with other breeds, suggesting either admixture or incomplete lineage sorting. Overall, the PCA confirms that most breeds maintain breed-specific genomic signatures, while a few shows evidence of historical gene flow or related ancestry.

PCoA based on pairwise genetic distances (1 – IBS) revealed clear genetic separation among the horse breeds (Figure 3). The first two axes explained 5.1% and 4.5% of the total variation. Distinct clusters corresponded to individual breeds, with Arabian, Standardbred, and Quarter Horse showing tight groupings, indicating within-breed uniformity. Some overlap was observed among Welsh and Lusitano, suggesting partial shared ancestry. Caspian and Dareshuri clustered closely, while Turkmen and Thoroughbred occupied intermediate positions, reflecting mixed genetic backgrounds. These patterns align with the phylogenetic and structure analyses, confirming breed-level genetic differentiation.

Candidate genes under selection amongst 14 horse breeds

To identify genomic regions potentially under selection across the 14 horse breeds, hapFLK analysis was performed using K=10 haplotype clusters (Figure 4). A total of 83 genes were detected across multiple chromosomes, with the highest density on chromosomes 7 and 11, particularly involving keratin and olfactory receptor clusters (Table 2; Additional file). Key genes included *ZNF536* (chromosome 10), associated with neural development and behavior; *CD300E* and *CD300LB* (chromosome 11), involved in immune response; keratin genes *KRT15*, *KRT33B*, *KRT38*, *KRT34*, *KRTAP16-1* (chromosome 11), related to coat and hair structure; *RBFOX1* (chromosome 13), influencing neuronal development; *KIT* (chromosome 3), linked to coat color, fertility, and athletic performance; *CIRBP* and *CBARP* (chromosome 7), associated with cold stress tolerance and physical resilience; and *NDUFA10* (chromosome 6), related to energy metabolism and endurance. These findings suggest that selection in these horse populations has targeted genes contributing to neural development, behavior, coat characteristics, immune response, reproduction, physical resilience, cold tolerance, and performance traits. The analysis included five Iranian breeds (Iranian Arabian-Asil, Turkmen, Caspian, Kurdish, and Dareshuri) and nine international breeds (Thoroughbred, Quarter Horse, Belgian, Arabian, Welsh Pony, Standardbred, Icelandic, Lusitano, and Morgan), highlighting both shared and breed-specific genomic regions under selection that may contribute to phenotypic differentiation.

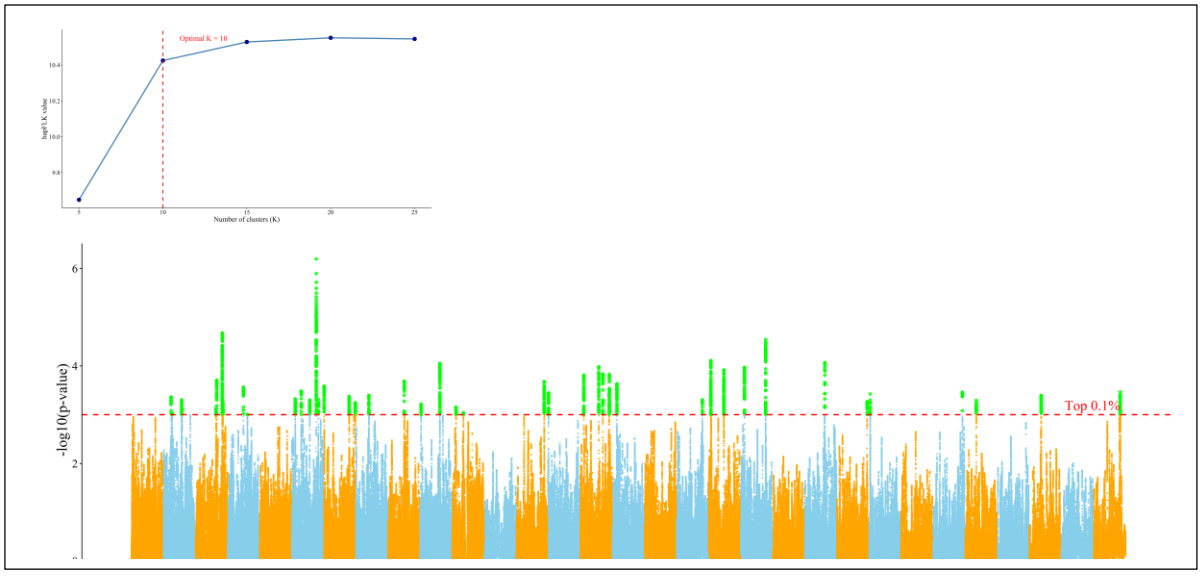

Figure 4. HapFLK Analysis of 14 Horse Breeds: Elbow plot for optimal K (top) and Manhattan plot of Top 0.1% signals (bottom)

Discussion

The estimated heterozygosity was 0.17–0.21 and generally smaller than the values previously reported for several genetically diverse horse populations characterized by broader founder contributions and less intensive directional selection. For instance, some Arabian horse populations showed higher heterozygosity (Cosgrove et al., 2020; Khanshour et al., 2013), while our estimate for Thoroughbred was smaller than the values reported by Petersen et al. (2013). The relatively reduced heterozygosity observed in Thoroughbred, Iranian Arabian (Asil), and Caspian horses is consistent with the influence of unequal founder contributions, intense directional selection, and selective breeding, which are known to reduce both the heterozygosity and allelic richness (Pjontek et al., 2012; Petersen et al., 2013). In the present study, He and Ho values for Iranian breeds were lower than those previously reported, ranging from 0.1677 to 0.1857 for He and from 0.1797 to 0.1963 for Ho. Bazvand et al. (2024) reported higher He (0.3457–0.3460) and Ho (0.2907–0.3458) values for Arabian, Caspian, Kurdish, and Turkmen horses; Moradi and Khaltabadi Farahani (2018) found He = 0.27–0.30 across five Asian breeds; and Sadeghi et al. (2019) reported He = 0.45 and Ho = 0.43 in 71 Arabian horses. Similarly, Almarzook et al. (2017) reported He = 0.30–0.31 and Ho = 0.30–0.32 in Syrian Arab horses. These data suggest that the reduced heterozygosity observed in our Iranian breeds likely reflects strong artificial selection and/or population isolation, leading to limited gene flow and loss of genetic diversity (Moradi and Khaltabadi Farahani, 2018).

A clearer contrast emerged between international and native Iranian breeds when considering allelic richness (AR), minor allele frequency (MAF), and the effective number of alleles (NE). Despite showing moderate heterozygosity, the Iranian breeds, particularly Caspian, Dareshuri, and Kurdish, displayed markedly reduced AR and Ne values, which is consistent with long-term isolation, small effective population sizes, and historical founder events. Similar trends were documented previously. Microsatellite analyses found high diversity in cosmopolitan breeds such as Quarter Horse (0.847) and Thoroughbred (0.877) (Vázquez-Armijo et al., 2017), while Cozzi et al. (2018) noted moderate but variable diversity in Akhal-Teke, shaped by demographic contractions and breeding practices.

Patterns of MAF further emphasized these differences. Very low MAF in Quarter Horse and Morgan aligns with expectations for strong directional selection, which reduces minor allele frequencies through allele fixation (Vitti et al., 2013). Conversely, the relatively high MAF in Dareshuri and Kurdish suggests retention of ancestral polymorphisms under weaker selection or limited artificial breeding pressure (Charlesworth, 2009). High AR in large cosmopolitan breeds such as Quarter Horse and Morgan reflects their large, admixed breeding populations, whereas the extremely low AR observed in Dareshuri and Kurdish supports scenarios of historical bottlenecks and long-term isolation patterns consistent with the conclusions of Greenbaum et al. (2014). Nucleotide diversity (Pi) exhibited limited variation across breeds; however, international breeds generally showed slightly higher values compared to native Iranian breeds, mirroring differences in demographic history and degrees of admixture.

Pairwise F_{ST} and private allele analyses revealed moderate genetic differentiation among the 14 horse breeds, reflecting both shared ancestry and selective divergence. The low F_{ST} values observed among Caspian, Dareshuri, Kurdish, Iranian Arabian (Asil), and Turkemen breeds indicate strong genetic relatedness and suggest historical gene flow within Iranian lineages. Additionally, the marked divergence between Thoroughbred and Icelandic horses reflects their contrasting breeding histories, demographic trajectories, and selection intensities (Weir and Cockerham, 1984; Weir and Goude, 2017).

No private alleles were identified for the Iranian breeds using the current SNP bead-chip. This observation should not be interpreted as evidence of reduced genetic diversity, but rather as a methodological limitation, as the SNP bead-chip was not specifically designed for Iranian horse populations and may not capture breed-specific variants. In addition, historical gene flow, limited sample sizes, and shared ancestry among Iranian breeds may further reduce the probability of detecting private alleles. Separate analyses focused exclusively on Iranian breeds may provide more accurate estimates of private alleles. This interpretation is also consistent with their low genetic differentiation (low F_{ST} values).

Conversely, higher levels of differentiation and the large numbers of private alleles detected in Morgan, Standardbred, and Welsh horses highlight their substantial genetic distinctiveness, likely shaped by closed breeding practices and long-term selection for specific performance traits. The exceptionally high number of private alleles in the Morgan breed further underscores its unique genetic composition and prolonged isolation. Similar patterns have been reported in intensively selected or closed horse populations, where long-term artificial selection and restricted gene flow have led to reduced genetic diversity and increased differentiation (Durward-Akhurst et al., 2021; Esdaile et al., 2022; Hou et al., 2023).

Positive Tajima's D values across most breeds indicate an excess of intermediate-frequency alleles, consistent with historical population contractions or balancing selection (Tajima, 1989). The highest values observed in Thoroughbred, Standardbred, and Quarter Horse likely reflect strong artificial selection and restricted breeding, whereas lower values in Dareshuri and Caspian horses may suggest comparatively more stable demographic histories. It should be noted that the Caspian breed has undergone a substantial reduction in census size over the past decades, which is reflected in

relatively low effective population size (Ne) estimates compared with other Iranian breeds (Bazvand et al., 2024; Mousavi et al., 2023). The relatively low positive Tajima's D for Caspian horses in our genome-wide analysis may be influenced by factors such as limited sample size, genome-wide averaging, and SNP array design favoring common variants, which can moderate the detectable effect of recent bottlenecks. Similarly, M-ratio estimates (0.71–0.83) indicate moderate historical bottlenecks, particularly in Dareshuri and Kurdish breeds (Garza and Williamson, 2001). Overall, these results reveal varying degrees of past demographic reductions among breeds, reflecting differences in management history, breeding intensity, and population size dynamics (Luikart et al., 1998).

The PCA and PCoA results collectively demonstrate a clear genetic structure among the 14 horse breeds, consistent with their known breeding histories and geographic origins. The formation of distinct clusters, particularly for Belgian, Icelandic, and Arabian breeds, highlights strong genetic differentiation and limited recent admixture. Moderate proximity among Thoroughbred, Iranian Arabian (Asil), and Standardbred suggests partial shared ancestry, likely reflecting historical crossbreeding in performance lines (Momen et al., 2022; Yordanov et al., 2025). Overlaps observed among breeds such as Welsh, Lusitano, Quarter Horse, and Morgan may indicate gene flow or incomplete lineage sorting. Overall, these findings confirm that most breeds retain distinct genomic signatures, while a few exhibit patterns of historical connectivity within the broader equine gene pool (Arefnejad et al., 2024).

The hapFLK analysis revealed distinct genomic regions under selection across the 14 horse breeds, reflecting both shared and breed-specific adaptive pressures. These patterns indicate that historical selection, driven by environmental adaptation, performance, and Artificial selection, has shaped the genetic architecture of each breed. Iranian breeds, alongside international populations, show unique signatures that likely correspond to local adaptation and breed-specific traits. The results highlighted the importance of considering selection footprints in conservation and breeding programs, as they provide insights into maintaining genetic diversity while preserving desirable phenotypic characteristics. Understanding these selective patterns can guide strategies to optimize breeding decisions, protect vulnerable breeds, and sustain the long-term genetic health of horse populations (Momen et al., 2022; Bazvand et al., 2025).

Limitations and future directions

A potential limitation of this study is the integration of genotype data obtained from SNP arrays with different densities. Although combining datasets may introduce bias due to differences in array density or population sampling, we applied stringent quality control, removed closely related individuals, performed LD pruning, and conducted comprehensive population structure analyses (PCA, PCoA, hapFLK) to ensure that the main genomic patterns remained robust. A parallel evaluation comparing lower- and higher-density datasets also confirmed that key metrics, including heterozygosity, F_{ST}, admixture, and population clustering, were consistent and reliable.

Nevertheless, potential bias cannot be completely ruled out. For future studies, we recommend analyzing datasets separately before merging and comparing results across datasets to avoid biases arising from data integration. This approach can enhance the accuracy and reliability of genomic findings and help ensure accurate assessment of genetic diversity and population structure.

Conclusion

This study offers a detailed genomic view of diversity, structure, and selection across 14 horse breeds, highlighting clear differences between indigenous Iranian and international populations. Iranian breeds generally show lower heterozygosity, allelic richness, and effective allele numbers, likely reflecting long-term isolation, small effective population sizes, and historical bottlenecks. In contrast, breeds such as Quarter Horse and Morgan maintain higher genetic diversity, probably due to larger, admixed populations and intensive breeding programs. Analyses of pairwise F_{ST} and private alleles reveal both shared ancestral variation within Iranian breeds and pronounced differentiation in modern Western breeds, underscoring the influence of artificial selection, closed breeding systems, and geographic separation. Demographic assessments using Tajima's D and M-ratio further support historical reductions in population size, with varying intensity across breeds. Population structure analyses identify distinct breed-specific clusters while also showing evidence of gene flow or incomplete lineage sorting in some populations. HapFLK analyses highlight genomic regions under selection, pinpointing genes linked to performance (*NDUFA10*; chr6), behavior (*ZNF536*; chr10), coat traits (*KIT*; chr3), immune function (*C1QA*; chr2), and environmental adaptation (*SESN2*; chr2). Together, these findings emphasize the value of integrating multiple genomic measures to capture the complex interplay of selection, genetic drift, and admixture shaping horse populations. The insights gained provide a strong foundation for sustainable breeding programs, prioritization of at-risk breeds, and preservation of genetic diversity, all while maintaining key phenotypic traits that contribute to the cultural, historical, and economic significance of horses worldwide.

Conflict of interest

The authors declare that they have no competing interests.

References

Almarzook, S., Reissmann, M., Arends, D., Brockmann, G.A., 2017. Genetic diversity of Syrian Arabian horses. *Animal Genetics* 48(4), 486-489.

Amirinia, C., Seyedabadi, H., Banabazi, M.H., Kamali, M.A., 2007. Bottleneck study and genetic structure of Iranian Caspian horse population using microsatellites. *Pakistan Journal of Biological Sciences* 10(9), 1540-1543.

Amjadi, M.A., Yeganeh, H.M., Sadeghi, M., Abbas Raza, S.H., Yang, J., Najafabadi, H.A., Batool, U., Shoorei, H., Abdelnour, S.A., Ahmed, J.Z., 2021. Microsatellite analysis of genetic diversity and population structure of the Iranian Kurdish horse. *Journal of Equine Veterinary Science* 98, 103358.

Arefnejad, B., Zeinalabedini, M., Talebi, R., Mardi, M., Ghaffari, M.R., Vahidi, M.F., Nekouei, M.K., Szmatoła T, Salekdeh, G.H., 2024. Unveiling the population genetic structure of Iranian horse breeds by whole-genome resequencing analysis. *Mammalian Genome* 35, 201-227.

Atroshchenko, M., Dementieva, N., Shcherbakov, Y., Nikolaeva, O., Azovtseva, A., Ryabova, A., Nikitkina, E., Makhmutova, O., Datsyshin, A., Zakharov, V., Zaitsev, A., 2023. The genetic diversity of horse native breeds in Russia. *Genes* 14(12), 2148.

Bazvand, B., Rashidi, A., Zandi, M.B., Moradi, M.H., Rostamzadeh, J., Mahmoudi, P., 2025. Population structure and signature of selection for body size in some Iranian and exotic horse breeds. *Journal of Equine Veterinary Science* 152, 105658.

Bazvand, B., Rashidi, A., Zandi, M.B., Moradi, M.H., Rostamzadeh, J., 2024. Genome-wide analysis of population structure, effective population size and inbreeding in Iranian and exotic horses. *PLoS One* 19(3), e0299109.

Browning, B.L., Zhou, Y., Browning, S.R., 2018. A one-penny imputed genome from next-generation reference panels. *American Journal of Human Genetics* 103(3), 338-348.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7.

Chang, L.Y., Toghiani, S., Hay, E.H., Aggrey, S.E., Rekaya, R., 2019. A weighted genomic relationship matrix based on fixation index (F_{ST}) prioritized SNPs for genomic selection. *Genes* 10(11), 922.

Charlesworth, B., 2009. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* 10(3), 195-205.

Cortes-Rodriguez, N., Campana, M.G., Berry, L., Faegre, S., Derrickson, S.R., Ha, R.R., Dikow, R.B., Rutz, C., Fleischer, R.C., 2019. Population genomics and structure of the critically endangered Mariana Crow (Corvus kubaryi). *Genes* 10(3), 187.

Cosgrove, E.J., Sadeghi, R., Schlamp, F., Holl, H.M., Moradi-Shahrbabak, M., Miraei-Ashtiani, S.R., Abdalla, S., Shykind, B., Troedsson, M., Stefaniuk-Szmukier, M., Prabhu, A., Bucca, S., Bugno-Poniewierska, M., Wallner, B., Malek, J., Miller, D.C., Clark, A.G., Antczak, D.F., Brooks, S.A., 2020. Genome diversity and the origin of the Arabian horse. *Scientific Reports* 10, 9702.

Cozzi, M.C., Strillacci, M.G., Valiati, P., Rogliano, E., Bagnato, A., Longeri, M., 2018. Genetic variability of Akhal-Teke horses bred in Italy. *PeerJ* 6, e4889.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000 genomes project analysis group, 2011. The variant call format and VCFtools. *Bioinformatics* 27(15), 2156-2158.

Durward-Akhurst, S.A., Schaefer, R.J., Grantham, B., Carey, W.K., Mickelson, J.R., McCue, M.E., 2021. Genetic variation and the distribution of variant types in the horse. *Frontiers in Genetics* 12, 758366.

Esdaile, E., Avila, F., Bellone, R.R., 2022. Analysis of genetic diversity in the American Standardbred horse utilizing short tandem repeats and single nucleotide polymorphisms. *Journal of Heredity* 113(3), 238-247.

Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., McCrory Constantz, C., Gamba, C., Al-Rasheid, K.A.S., Albizuri, S., Alfarhan, A.H., Allentoft, M., Alquraishi, S., Anthony, D., Baimukhanov, N., Barrett, J.H., Bayarsaikhan, J., Benecke, N., Bernáldez-Sánchez, E., Berrocal-Rangel, L., Biglari, F., Boessenkool, S., Boldgiv, B., Brem, G., Brown, D., Burger, J., Crubézy, E., Daugnora, L., Davoudi, H., de Barros Damgaard, P., de Los Ángeles de Chorro Y de Villa-Ceballos. M., Deschler-Erb. S., Detry. C., Dill. N., do Mar Oom. M., Dohr. A., Ellingvåg. S., Erdenebaatar. D., Fathi. H., Felkel. S., Fernández-Rodríguez. C., García-Viñas. E., Germonpré. M., Granado. J.D., Hallsson. J.H., Hemmer. H., Hofreiter. M., Kasparov. A., Khasanov. M., Khazaeli. R., Kosintsev. P., Kristiansen. K., Kubatbek. T., Kuderna. L., Kuznetsov. P., Laleh. H., Leonard. J.A., Lhuillier. J., Liesau von Lettow-Vorbeck. C., Logvin. A., Lõugas. L., Ludwig. A., Luis. C., Arruda. A.M., Marques-Bonet. T., Matoso Silva. R., Merz. V., Mijiddorj. E., Miller. B.K., Monchalov. O., Mohaseb. F.A., Morales. A., Nieto-Espinet. A., Nistelberger. H., Onar. V., Pálsdóttir. A.H., Pitulko. V., Pitskhelauri. K., Pruvost. M., Rajic Sikanjic. P., Rapan Papeša. A., Roslyakova. N., Sardari. A., Sauer. E., Schafberg. R., Scheu. A., Schibler. J., Schlumbaum. A., Serrand. N., Serres-Armero. A., Shapiro. B., Sheikhi Seno. S., Shevnina. I., Shidrang. S., Southon. J., Star. B.,

Sykes. N., Taheri. K., Taylor. W., Teegen. W.R., Trbojević Vukičević. T., Trixl. S., Tumen. D.,

Undrakhbold. S., Usmanova. E., Vahdati. A., Valenzuela-Lamas. S., Viegas. C., Wallner. B., Weinstock. J., Zaibert. V., Clavel. B., Lepetz. S., Mashkour. M., Helgason. A., Stefánsson. K., Barrey. E., Willerslev. E., Outram. A.K., Librado. P., Orlando, L., 2019. Tracking five millennia of horse management with extensive ancient genome time series. *Cell* 177(6), 1419-1435.e31.

Fariello, M.I., Boitard, S., Naya, H., SanCristobal, M., Servin, B., 2013. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics* 193(3), 929-941.

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S.N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E.S., Daly, M.J., Altshuler, D., 2002. The structure of haplotype blocks in the human genome. *Science* 296(5576), 2225-2229.

Garza, J.C., Williamson, E.G., 2001. Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* 10(2), 305-318.

Greenbaum, G., Templeton, A.R., Zarmi, Y., Bar-David, S., 2014. Allelic richness following population founding events: a stochastic modeling framework incorporating gene flow and genetic drift. *PLoS One* 9(12), e115203.

Hedayat Evrigh, N., Omri, M., Boustan, A., Seyedsharifi, R., Vahedi, V., 2018. Genetic diversity and structure of Iranian horses' population based on mitochondrial markers. *Journal of Equine Veterinary Science* 64, 107-111.

Hou, L., Sulayman, A., Zeng, Y., Zhou, L., Aimaier, A., Kader, A., Shi, L., 2025. Analysis of genetic diversity and race genetic structure of major horse breeds in Xinjiang, China. *Animals* 15(18), 2690.

Hu, M., Jiang, H., Lai, W., Shi, L., Yi, W., Sun, H., Chen, C., Yuan, B., Yan, S., Zhang, J., 2023. Assessing genomic diversity and signatures of selection in Chinese Red Steppe cattle using high-density SNP array. *Animals* 13(10), 1717.

Khanshour, A., Conant, E., Juras, R., Cothran, E.G., 2013. Microsatellite analysis of genetic diversity and population structure of Arabian horse populations. *Journal of Heredity* 104(3), 386-398.

Leroy, G., Callède, L., Verrier, E., Mériaux, J.C., Ricard, A., Danchin-Burge, C., Rognon, X., 2009. Genetic diversity of a large set of horse breeds raised in France assessed by microsatellite polymorphism. *Genetics Selection Evolution* 41(1), 31.

Luikart, G., Sherwin, W.B., Steele, B.M., Allendorf, F.W., 1998. Usefulness of molecular markers for detecting population bottlenecks via monitoring genetic change. *Molecular Ecology* 7(8), 963-974.

Machmoum, M., Boujenane, I., Azelhak, R., Badaoui, B., Petit, D., Piro, M., 2020. Genetic diversity and population structure of Arabian horse populations using microsatellite markers. *Journal of Equine Veterinary Science* 93, 103200.

Momen, M., Brauer, K., Patterson, M.M., Sample, S.J., Binversie, E.E., Davis, B.W., Cothran, E.G., Rosa, G.J.M., Brounts, S.H., Muir, P., 2023. Genetic architecture and polygenic risk score prediction of degenerative suspensory ligament desmitis (DSLD) in the Peruvian horse. *Frontiers in Genetics* 14, 1201628. https://doi.org/10.5061/dryad.cnp5hqc9p.

Momen, M., Brounts, S.H., Binversie, E.E., Sample, S.J., Rosa, G.J.M., Davis, B.W., Muir, P., 2022. Selection signature analyses and genome-wide association reveal genomic hotspot regions that reflect differences between breeds of horse with contrasting risk of degenerative suspensory ligament desmitis. *G3* 12(10), jkac179.

Moradi, M.H., Khalt-Abadi Farahani, A.H., 2018. Estimation of inbreeding values using genomic run of homozygosity and study of evolutionary trend for effective population size in some Asian horse breeds. *Iranian Journal of Animal Science* 49(3), 381-392.

Moravčíková, N., Kasarda, R., Vostra Vydrova, H., Vostry, L., Karásková, B., Candrák, J., Halo, M., 2024. Genomic variability of the MHC region: empirical evidence from five horse breeds. *Livestock Science* 284, 105480.

Mousavi, S.F., Razmkabir, M., Rostamzadeh, J., Seyedabadi, H.R., Naboulsi, R., Petersen, J.L., Lindgren, G., 2023. Genetic diversity and signatures of selection in four indigenous horse breeds of Iran. *Heredity* 131(2), 96-108.

Nei, M., Li, W.H., 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America* 76, 5269-5273.

Ospina, J.A., Lopez-Alvarez, D., Gimode, W., Wenzl, P., Carvajal-Yepes, M., 2024. Genome-wide association study of cassava brown streak disease resistance in cassava germplasm conserved in South America. *Scientific Reports* 14(1), 23141.

Petersen, J.L., Mickelson, J.R., Cothran, E.G., Andersson, L.S., Axelsson, J., Bailey, E., Bannasch, D., Binns, M.M., Borges, A.S., Brama, P., da Câmara Machado, A., Distl, O., Felicetti, M., Fox-Clipsham, L., Graves, K.T., Guérin, G., Haase, B., Hasegawa, T., Hemmann, K., Hill, E.W., Leeb, T., Lindgren, G., Lohi, H., Lopes, M.S., McGivney, B.A., Mikko, S., Orr, N., Penedo, M.C.T., Piercy, R.J., Raekallio, M., Rieder, S., Røed, K.H., Silvestrelli, M., Swinburne, J., Tozaki, T.,

Vaudin, M., Wade, C.M., McCue, M.E., 2013. Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLoS One* 8(1), e54997.

Pjontek, J., Kadlečík, O., Kasarda, R., Horný, M., 2012. Pedigree analysis in four Slovak endangered horse breeds. *Czech Journal of Animal Science* 57(2), 54-64.

R Core Team, 2023. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Sadeghi, R., Moradi-Shahrbabak, M., Miraei Ashtiani, S.R., Schlamp, F., Cosgrove, E.J., Antczak, D.F., 2019. Genetic diversity of Persian Arabian horses and their relationship to other native Iranian horse breeds. *Journal of Heredity* 110(2), 173-182. https://datadryad.org/dataset/doi:10.5061/dryad.54vb7f2.

Santos, W.B., Schettini, G.P., Maiorano, A.M., Bussiman, F.O., Balieiro, J.C.C., Ferraz, G.C., Pereira, G.L., Baldassini, W.A., Neto, O.R.M., Oliveira, H.N., Curi, R.A., 2021. Genome-wide scans for signatures of selection in Mangalarga Marchador horses using high-throughput SNP genotyping. *BMC Genomics* 22(1), 737.

Schaefer, R.J., Schubert, M., Bailey, E., Bannasch, D.L., Barrey, E., Bar-Gal, G.K., Brem, G., Brooks, S.A., Distl, O., Fries, R., Finno, C.J., Gerber, V., Haase, B., Jagannathan, V., Kalbfleisch, T., Leeb, T., Lindgren, G., Lopes, M.S., Mach, N., da Câmara Machado, A., MacLeod, J.N., McCoy, A., Metzger, J., Penedo, C., Polani, S., Rieder, S., Tammen, I., Tetens, J., Thaller, G., Verini-Supplizi, A., Wade, C.M., Wallner, B., Orlando, L., Mickelson, J.R., McCue, M.E., 2017. Developing a 670k genotyping array to tag ~2M SNPs across 24 horse breeds. *BMC Genomics* 18(1), 565.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123(3), 585-595.

Temple, S.D., Waples, R.K., Browning, S.R., 2024. Modeling recent positive selection using identity-by-descent segments. *American Journal of Human Genetics* 111(11), 2510-2529.

Vázquez-Armijo, J.F., Parra-Bracamonte, G.M., Velazquez, M.A., Sifuentes-Rincón, A.M., Tinoco-Jaramillo, J.L., Ambriz-Morales, P., Arellano-Vera, W., Moreno-Medina, V.R., 2017. Diversity and effective population size of four horse breeds from microsatellite DNA markers in South-Central Mexico. *Archives Animal Breeding* 60(2), 137-143.

Ventura, R.V., Brito, L.F., Oliveira, G.A. Jr, Daetwyler, H.D., Schenkel, F.S., Sargolzaei, M., Vandervoort, G., Silva, F.F., Miller, S.P., Carvalho, M.E., Santana. M.H.A., Mattos, E.C., Fonseca, P., Eler, J.P., Sterman, J.B., 2020. A comprehensive comparison of high-density SNP panels and an alternative ultra-high-density panel for genomic analyses in Nellore cattle. *Animal Production Science* 60(3), 333-346.

Vitti, J.J., Grossman, S.R., Sabeti, P.C., 2013. Detecting natural selection in genomic data. *Annual Review of Genetics* 47, 97-120.

Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38(6), 1358-1370.

Weir, B.S., Goudet, J., 2017. Understanding population structure: F_{ST} measures and their interpretation in an evolutionary context. *G3* 7(4), 1495-1501.

Xuereb, A., Nahuelpi, R.M., Normandeau, E., Babin, C., Laporte, M., Mallet, A., Yáñez, J.M., Mallet, M., Bernatchez, L., 2023. Design and validation of a high-density single nucleotide polymorphism array for the Eastern oyster (Crassostrea virginica). *G3* 13(6), jkad071.

Yordanov, G., Yordanov, T., Mehandjyiski, I., Radoslavov, G., Salkova, D., Hristov, P., 2025. Population structure and genetic diversity among Shagya Arabian horse genealogical lineages in Bulgaria based on microsatellite genotyping. *Veterinary Sciences* 12(8), 776.

Yu, N., Jensen-Seaman, M.I., Chemnick, L., Ryder, O., Li, W.H., 2004. Nucleotide diversity in gorillas. *Genetics* 166, 1375-1383.

Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., Weir, B.S., 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28(24), 3326-3328.