

Genomic insights into population structure, local ancestry and within-breed relationships in the Norik of Muran horse

KAROLÍNA PÁLEŠOVÁ¹ , NINA MORAVČÍKOVÁ¹ , ADRIÁN HALVONÍK¹ ,
EVA MIXTAJOVÁ² , GÁBOR MÉSZÁROS³ , LUBOŠ VOSTRÝ⁴ , RADOVAN KASARDA^{1*} 

¹*Institute of Nutrition and Genomics, Slovak University of Agriculture in Nitra, Nitra, Slovak Republic*

²*Project Management Office, Slovak University of Agriculture in Nitra, Nitra, Slovak Republic*

³*Institute of Livestock Sciences, University of Natural Resources and Life Sciences, Vienna, Austria*

⁴*Department of Genetics and Breeding, Czech University of Life Sciences Prague, Prague, Czech Republic*

*Corresponding author: radovan.kasarda@uniag.sk

Citation: Pálešová K., Moravčíková N., Halvoník A., Mixtajová E., Mészáros G., Vostrý L., Kasarda R. (2026): Genomic insights into population structure, local ancestry and within-breed relationships in the Norik of Muran horse. Czech J. Anim. Sci., 71: 456–473.

Abstract: The Norik of Muran is a local Slovak draught horse with historical connections to Central European heavy horses. This study characterised its genomic structure in relation to related Noriker breeds and other reference breeds through analyses of local ancestry, haplotype sharing and recent family structure. The dataset included Norik of Muran, Austrian Noriker, Silesian Noriker, Belgian Draught, Haflinger, Hucul and Norwegian Fjord, with Shagya Arabian and Quarter Horse as outgroups. Genotypes of 664 individuals were used, with 37 875 autosomal markers retained after quality control. Population structure analyses used a balanced dataset of 180 individuals (20 per breed) with reduced within-breed relatedness. Pairwise F_{ST} was lowest between Norik of Muran and Silesian Noriker (0.01), followed by Austrian Noriker (0.02). Local ancestry inference assigned the largest proportions of the Norik of Muran genome to Silesian Noriker (40.61%) and Austrian Noriker (39.61%), and a smaller proportion to Belgian Draught (9.23%). Haplotype sharing was also highest with Silesian and Austrian Noriker, from which Norik of Muran copied on average 108.38 and 108.14 haplotype chunks, respectively. Reconstruction of recent family structure in 125 Norik of Muran individuals identified 76 parent-offspring relationships, nine paternal and 10 maternal half-sibling groups, the largest paternal group including eight offspring. The results indicate close relationships with Silesian and Austrian Noriker, heterogeneous local ancestry and clearly detectable recent family structure. This information may support mating decisions in conservation breeding programmes, particularly those aimed at limiting repeated use of closely related animals, balancing sire contributions and preserving family-level diversity.

Keywords: admixture; autochthonous breed; conservation genomics; genetic diversity; haplotype sharing

Supported by the Slovak Research and Development Agency (APVV-20-0161), the Scientific Agency of the Ministry of Education, Research, Development and Youth of the Slovak Republic (VEGA 1/0316/25), and the Cultural and Educational Agency of the Ministry of Education, Research, Development and Youth of the Slovak Republic (KEGA 027SPU-4/2025). The contribution of Luboš Vostrý to this work was supported by the Ministry of Education, Youth and Sports of the Czech Republic (QL25020022).

© The authors. This work is licensed under a Creative Commons Attribution 4.0 International (CC-BY 4.0) license.

<https://doi.org/10.17221/75/2026-CJAS>

Local livestock breeds represent important animal genetic resources because they preserve unique combinations of alleles shaped by their origin, breeding history, local adaptation and long-term management (Oldenbroek 2021). In small or endangered populations, the loss of genetic diversity can reduce adaptive potential and increase the risk of inbreeding depression, making genomic monitoring important for conservation-oriented breeding decisions (Kardos et al. 2016).

Modern breeds have been shaped by strong artificial selection, studbook formation and breed-specific breeding goals (Petersen et al. 2013). Within such breeding populations, close relatives, half-sibling groups and unequal parental contributions can influence future inbreeding rates, especially when a small number of breeding animals contribute disproportionately to the next generation (Huisman 2017). At the same time, breed boundaries in horses are not always clearly separated, especially when populations share recent ancestry, historical introgression or common breeding origins (Gmel et al. 2024). This is important for local and regional breeds, where conservation value may be linked not only to complete genetic distinctiveness, but also to a specific combination of shared ancestry, demographic history and within-breed structure (Nogueira et al. 2022).

The Norik of Muran is a local Slovak draught horse and an important component of local animal genetic diversity. The breed was established at the Veľká Lúka stud near Muráň in the 1950s, with breeding aimed at producing a strong and manoeuvrable working horse suitable for forestry, agriculture and military purposes (Kasarda et al. 2019; Moravcikova et al. 2019). Its formation involved a heterogeneous foundation population, including Hucul and local heavy horses crossed with Noriker and Belgian Draught horses, followed by later contributions from Fjord, Haflinger, cold-blooded mares from Czechia, Moravia and Silesia and Hungarian Coldblood during subsequent stages of breed development (Moravcikova et al. 2019). The breed has subsequently been maintained within a closed breeding programme with specific breeding objectives resulting in a distinct phenotype, whereas the Austrian Noriker has been regarded as a potential controlled source of genetic diversity due to its historical relationship with the Norik of Muran (Halo et al. 2018). This history indicates that the genomic structure of the Norik

of Muran should be evaluated in the context of related Noriker breeds. Previous studies of the Norik of Muran have mainly focused on pedigree-based population structure, genetic diversity and inbreeding. Kasarda et al. (2019) evaluated pedigree-based and genomic inbreeding in the Norik of Muran using pedigree records and single nucleotide polymorphism (SNP) data. They reported low inbreeding estimates in the current generation, with pedigree inbreeding of 0.23% and genomic inbreeding of 0.11%. Moravcikova et al. (2019) further characterised the breed at the genomic level by analysing runs of homozygosity, which were detected in all analysed animals and covered, on average, 13.17% of the autosomal genome represented by SNP loci. Together, these studies provided important information on inbreeding and homozygosity within the Norik of Muran, but the broader genomic relationships of the breed with related horse populations remained insufficiently described.

In this study, we therefore aimed to characterise the genomic structure of the Norik of Muran as a local Slovak breed in comparison with related but separately recognised Noriker breeds and other reference horse breeds. We further assessed its local ancestry, haplotype sharing patterns and recent family structure within the breed. Together, these analyses provided a comparative framework for evaluation of the genomic position of the Norik of Muran, the distribution of ancestry patterns across its genome, and the organisation of recent family structure within the breed.

MATERIAL AND METHODS

Source data, breed composition and initial quality control. The dataset comprised nine horse breeds selected to represent the genetic background of the Norik of Muran and to enable comparisons to other populations. The dataset included Austrian Noriker (AN, $n = 184$), Belgian Draught (BD, $n = 30$), Norwegian Fjord (NF, $n = 20$), Haflinger (HF, $n = 91$), Hucul (HC, $n = 32$), Norik of Muran (MN, $n = 120$), Shagya Arabian (SA, $n = 47$), Silesian Noriker (SN, $n = 100$), and Quarter Horse (QH, $n = 40$). The breeds AN, BD, NF, HF, HC, and SN were selected to include populations with potential historical or biological relevance to the MN. In contrast, SA and QH were included as outgroup breeds.

Genotype data were derived from three SNP array platforms. MN, HC, SA, and SN were genotyped using the Illumina EquineSNP70 BeadChip. HF and AN were genotyped using the Axiom Equine Genotyping Array (670K). In addition, publicly available genotype data (Petersen et al. 2013) for QH, BD, and NF were obtained using the Illumina EquineSNP50 BeadChip.

Only autosomal SNPs shared across all breeds and genotyping platforms were retained for downstream analyses ($N = 37\,896$). Standard QC filtering was applied to the genotype data in PLINK v1.9 (Chang et al. 2015) using the options `--mind 0.10`, `--geno 0.10` and `--maf 0.01`, excluding individuals with more than 10% missing genotypes, SNPs with more than 10% missing data and markers with a minor allele frequency below 0.01. After filtering, the full dataset included 37 875 SNPs and 664 individuals.

Selection of unrelated individuals. To obtain a balanced dataset for the main comparative genomic analyses, 20 individuals were selected from each breed using the dataset after initial quality control. The aim was to keep the same number of animals per breed and to reduce the influence of closely related individuals. A realised genomic relationship matrix was calculated in PLINK v1.9 (Chang et al. 2015) using the `--make-rel` square option, following the definition of the relationship matrix by Yang et al. (2011). The resulting matrix was used as the basis for the within-breed selection procedure.

For each individual, global relatedness was calculated as the mean genomic relationship to all other individuals from the same breed. During the iterative selection, pairwise relatedness was calculated as the mean genomic relationship between each remaining candidate and the individuals already selected from the same breed. The detailed scoring algorithm, including the scaling of relatedness measures and the initial selection step, is provided in Electronic Supplementary Material (ESM) S1. The resulting balanced dataset comprised 180 individuals across nine breeds.

Preparation of datasets for genomic analyses. The balanced dataset of 180 individuals and 37 875 autosomal SNPs was used as the starting point for the preparation of datasets required by individual genomic analyses. First, SNPs with any missing genotypes were removed in PLINK v1.9 (Chang et al. 2015) using the `--geno 0` option. This

reduced the dataset from 37 875 to 23 773 autosomal SNPs.

For local ancestry inference and haplotype sharing analysis, genetic map information was then assigned to the resulting marker set. A genome-averaged recombination map for the domestic horse, aligned to the EquCab2 reference genome and derived from Beeson et al. (2019), was used to obtain genetic positions and recombination rates. Of the 23 773 SNPs, 23 323 were directly matched to the recombination map, and 414 additional SNPs were assigned genetic positions by interpolation between neighbouring mapped markers. Together, the directly matched and interpolated SNPs accounted for 23 737 SNPs with assigned genetic map information. The remaining 36 SNPs could not be assigned genetic positions because suitable corresponding or neighbouring markers were not available in the reference map and were excluded. Genotypes in the final marker set were phased separately for each chromosome using SHAPEIT2 (Delaneau et al. 2013) and the corresponding chromosome-specific genetic map. For principal component analysis (PCA) and global ancestry analysis, the dataset without missing genotypes was further filtered for linkage disequilibrium in PLINK v1.9 (Chang et al. 2015) using the `--indep-pairwise 50 5 0.2` option, corresponding to a sliding window of 50 SNPs, a step size of 5 SNPs and an r^2 threshold of 0.2. The resulting linkage disequilibrium pruned dataset contained 17 303 SNPs.

A separate dataset was prepared for family structure reconstruction within the MN using all available genotyped individuals from the breed. Quality control was performed in PLINK v1.9 (Chang et al. 2015) using the same filtering criteria as for the full dataset. After quality control, the MN dataset contained 125 individuals and 55 426 autosomal SNPs. For pedigree reconstruction, SNPs with a minor allele frequency below 0.05 and SNPs with missing genotypes were removed in PLINK v1.9 (Chang et al. 2015) using the `--maf 0.05` and `--geno 0` options. Linkage disequilibrium pruning was then performed in PLINK v1.9 (Chang et al. 2015) using the `--indep-pairwise 50 5 0.2` option, which corresponds to a sliding window of 50 SNPs, a step size of 5 SNPs and an r^2 threshold of 0.2. From the resulting marker set, 700 SNPs were selected by chromosome proportional sampling to obtain an informative marker subset distributed across

<https://doi.org/10.17221/75/2026-CJAS>

the genome. Within each chromosome, SNPs were selected from evenly spaced genomic intervals, preferentially retaining markers with the highest minor allele frequency to maintain genome-wide coverage.

The complete workflow of dataset preparation is summarised in Figure 1. ESM Table S1 provides an overview of the datasets used in the study, including the filtering or processing steps, number of individuals, breed composition, number of retained SNPs, and mean and median distances between neighbouring SNPs. Unless stated otherwise, graphical outputs derived from these analyses were prepared in R using *ggplot2* (Wickham 2016).

Between-breed population structure analyses. The PCA was performed in PLINK v1.9 (Chang et al. 2015) using the *--pca* 10 option. The first 10 principal components were calculated to allow

inspection of additional axes of genetic variation, while the first two components were visualised because they captured the major pattern of differentiation among individuals. The analysis was first performed using the balanced dataset of 180 individuals from nine breeds. Because AN, MN, SN and BD clustered closely in the overall analysis, a second PCA was performed using only these four breeds to provide a more detailed view of their genetic structure. For visualisation of the four-breed PCA, one extreme AN outlier was excluded before PCA calculation to avoid excessive compression of the remaining individuals in the plot. This exclusion was restricted to the four-breed PCA visualisation; the individual was retained in the nine-breed PCA and in all other analyses.

Genetic differentiation among breeds was assessed using Wright's fixation index (F_{ST}), calcu-

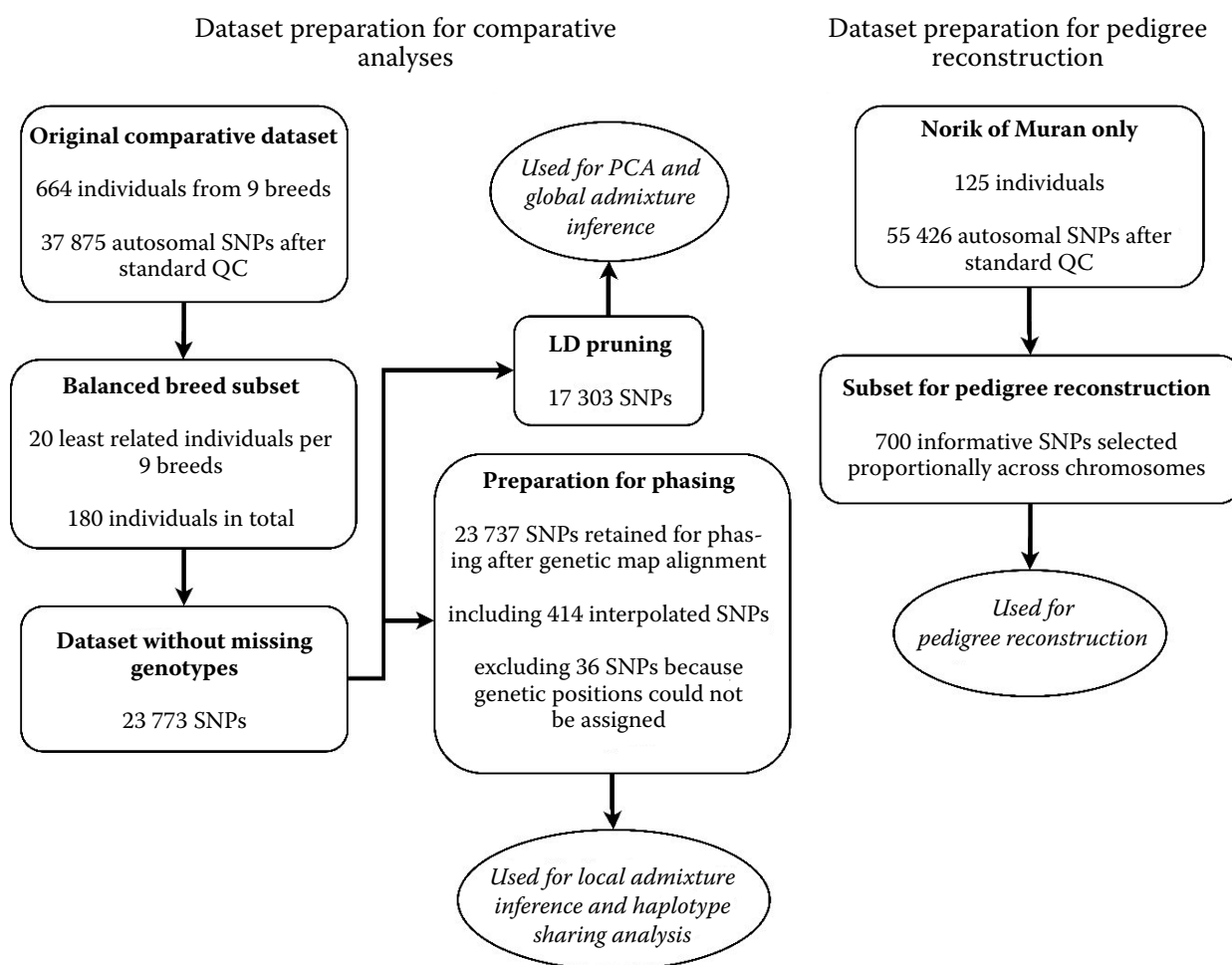


Figure 1. Workflow of dataset preparation for genomic analyses and pedigree reconstruction

LD = linkage disequilibrium; PCA = principal component analysis; QC = quality control; SNP = single nucleotide polymorphism

lated for all breed pairs in PLINK v1.9 (Chang et al. 2015) using the *--fst* option. Pairwise F_{ST} estimates were visualised as a heatmap in R.

Genetic relationships among individuals were assessed using identity by state (IBS) distances calculated in PLINK v1.9 (Chang et al. 2015) with the *--distance square 1-ibs* option. The resulting individual distance matrix was visualised as a heatmap in R. At the breed level, Nei's genetic distances were calculated in R using the *StAMPP* package (Pembleton et al. 2013). A NeighborNet network was constructed from the Nei distance matrix using the *phangorn* package (Schliep 2011).

Population structure was further explored using the NetView approach (Neuditschko et al. 2012), which represents individuals as nodes connected according to genetic similarity, thereby enabling the detection of fine-scale population structure through network topology. Pairwise genetic distances based on IBS were calculated in PLINK v1.9 (Chang et al. 2015) and used to construct networks based on mutual k -nearest neighbours (mkNN), which retain reciprocal nearest-neighbour relationships between genetically similar individuals and thereby preserve local genetic structure within the network. Representative networks were visualised for $k = 5, 10$ and 15 to capture different levels of network connectivity. The parameter exploration used to select these values is described in ESM S1 and ESM Figure S1.

Global admixture analysis. Ancestry proportions were estimated using ADMIXTURE (Alexander et al. 2009), a model-based maximum-likelihood clustering algorithm that estimates each individual's ancestry proportions from K hypothetical ancestral populations. Analyses were performed for a range of ancestral populations ($K = 1$ – 20), with five independent runs conducted for each K value. The most appropriate number of ancestral populations was evaluated using StructureSelector (Li and Liu 2018) based on the lowest cross-validation error. Results from multiple runs were aligned and compared in pong (Behr et al. 2016) to assess the consistency of clustering patterns across replicates and identify representative clustering solutions for each K before visualisation.

Local ancestry inference. Ancestry along the genome was inferred using RFMix v1 (Maples et al. 2013), a discriminative modelling approach based on random forests and conditional random fields that assigns local ancestry along phased chromo-

somes by comparing target haplotypes to reference population haplotypes. MN was defined as the target population, and the remaining breeds (AN, SN, BD, HC, HE, NE, QH and SA) were used as reference populations. The reference breeds were used to represent possible sources of shared ancestry, not direct historical ancestors of MN. Analyses were conducted separately for each chromosome using phased genotype data, resulting in ancestry estimates at the SNP level.

Local ancestry patterns were summarised across individuals and visualised as a karyogram. For each SNP position, the dominant ancestry component was determined based on the highest inferred ancestry frequency across individuals within the MN population. Consecutive SNPs assigned to the same dominant ancestry were grouped into continuous ancestry segments along chromosomes. Each segment retained both physical coordinates, based on SNP positions in bp, and genetic coordinates, based on the corresponding cM positions from the recombination map. To reduce short, fragmented ancestry transitions, segments shorter than 1 cM were iteratively merged with adjacent segments using genetic length as the smoothing criterion. The final merged segments retained their updated physical and genetic start and end positions. Overlap with quantitative trait loci (QTL) regions from the Animal QTLdb database (Hu et al. 2019) was then assessed using physical genomic coordinates aligned to the EquCab2 reference genome. Overlaps between dominant ancestry segments and QTL regions from Animal QTLdb were summarised for each inferred ancestry source and visualised as bar plots. All overlapping database entries were retained, so the summaries reflect the total number of available QTL annotations and are not limited to unique QTL identifiers. Because ancestry sources differed in the total length of dominant segments, overlap counts were also expressed per Mb and per cM of assigned segment length. These summaries were used only to provide descriptive functional context and were not treated as formal QTL enrichment tests.

Haplotype sharing analysis. Patterns of haplotype sharing and haplotype-based population structure were analysed using the linked ChromoPainter/fineSTRUCTURE v2 pipeline (Lawson et al. 2012). The phased haplotype data were analysed in linked mode using an all-to-all painting design, in which all individuals were included as both donors and

<https://doi.org/10.17221/75/2026-CJAS>

recipients, with the corresponding recombination maps used to account for linkage among markers. ChromoPainter model parameters were estimated using the expectation-maximisation procedure implemented in the fineSTRUCTURE pipeline before the final painting step.

The ChromoPainter chunk count matrix was used as the primary coancestry matrix. This matrix represents the expected number of contiguous haplotype chunks copied by recipient individuals from donor individuals under the ChromoPainter copying model. Breed-level haplotype sharing was summarised by averaging chunk count values across recipient and donor breed combinations. Self-copying values were excluded when calculating mean within-breed haplotype sharing. The ChromoPainter coancestry matrix was subsequently analysed in fineSTRUCTURE to infer haplotype-based clustering, with the resulting clustering used for tree construction. Technical details of the ChromoPainter parameter estimation and processing and visualisation of the coancestry matrices and fineSTRUCTURE clustering and tree construction are provided in ESM S1.

Within-breed structure of Norik of Muran. To investigate recent within-breed structure in MN, pedigree reconstruction was performed using all available genotyped individuals from the breed. Although pedigree records exist for the breed, the complete official pedigree corresponding to the genotyped individuals was not available for the present study; therefore, pedigree reconstruction was used to explore recent within-breed family structure from the available genomic data. The resulting reconstruction was intended to describe recent family relationships rather than to represent a complete multigenerational pedigree. Parentage relationships and sibship structure were inferred using the *sequoia* package in R (Huisman 2017).

The 700 SNP subset prepared for pedigree reconstruction was used as input, following the *sequoia* recommendation to use an informative set of approximately independent markers for pedigree reconstruction. Because genotyping errors can affect parentage assignment, several assumed genotyping error rates were tested (0.001, 0.003, 0.005, 0.010 and 0.020). The same parent-offspring assignments were obtained across all tested values, and an error rate of 0.005 was retained for the final analysis.

Individual metadata included available sex information for genotyped individuals. Among the

genotyped individuals, 75 were mares, and 50 were stallions. Because birth year information was unavailable, birth year was set as missing for all individuals. Genotyped individuals were anonymised as MN_001 to MN_125, whereas inferred non-genotyped parents were labelled as Dummy_1, Dummy_2 and so forth.

The inferred pedigree structure was subsequently used to identify paternal lines, maternal families and half-sibling groups within the population. Results were visualised using anonymised pedigree network graphs and bar plots summarising reconstructed paternal and maternal half-sibling groups and parental contributions within the population.

RESULTS

Relatedness filtering. After selection, global and pairwise genomic relatedness differed among breeds (ESM Table S2). The lowest mean global relatedness was observed in MN (0.040 ± 0.009), AN (0.048 ± 0.032) and SN (0.061 ± 0.009), whereas the highest values were observed in SA (0.324 ± 0.014), NF (0.252 ± 0.010) and QH (0.249 ± 0.012). Mean pairwise relatedness showed a similar pattern, with the lowest values in MN (0.027 ± 0.043), AN (0.029 ± 0.042) and SN (0.053 ± 0.042), and the highest values in SA (0.310 ± 0.081), NF (0.252 ± 0.039) and QH (0.234 ± 0.032). The highest maximum pairwise values were observed in HC (0.701) and SA (0.653).

Between-breed population structure. The PCA of the balanced dataset revealed clear genetic structure among the analysed breeds (Figure 2A). The first two principal components, which explained the largest proportions of variation, accounted for 23.00% (PC1) and 14.52% (PC2) of the total genetic variation. In the nine-breed PCA, MN did not form a separate cluster but overlapped with SN and was positioned close to AN. BD was separated from MN, SN and AN mainly along PC2. Among the remaining breeds, PC1 mainly separated SA and QH from the other breeds, whereas PC2 contributed to the separation of NF and HF. HC also formed a distinct cluster, separated from the remaining breeds along both components.

The four-breed PCA of AN, BD, MN and SN provided a more detailed view of the genetic structure among these closely related breeds (Figure 2B). The first two principal components explained 19.64%

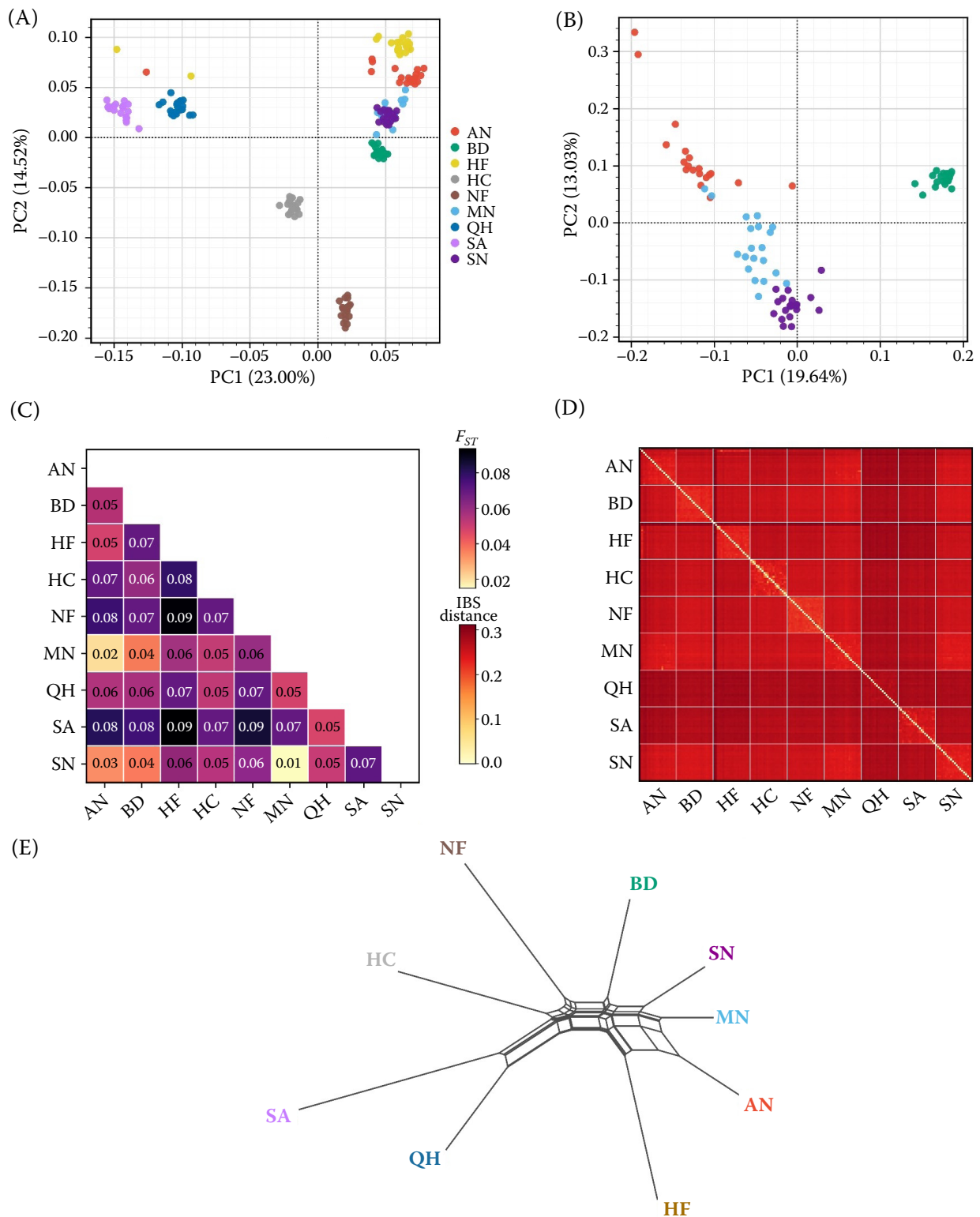


Figure 2. Genomic structure, differentiation, and relationships among nine horse breeds visualised using (A) PCA of all breeds, (B) PCA of closely related breeds, (C) pairwise F_{ST} heatmap, (D) individual-level IBS heatmap, (E) NeighborNet network based on Nei's genetic distances

AN = Austrian Noriker; BD = Belgian Draught; F_{ST} = Wright's fixation index; HC = Hucul; HF = Haflinger; IBS = identity by state; MN = Norik of Muran; NF = Norwegian Fjord; PCA = principal component analysis; QH = Quarter Horse; SA = Shagya Arabian; SN = Silesian Noriker

<https://doi.org/10.17221/75/2026-CJAS>

and 13.03% of the genetic variation, respectively. BD showed the clearest separation from the other three breeds, mainly along PC1. The AN also showed differentiation from the remaining breeds, although partial overlap was observed between AN and MN. The MN occupied an intermediate position between AN and SN and showed greater within-breed dispersion than SN. The SN formed a more compact cluster and was mainly separated from AN and MN along PC2.

Pairwise genetic differentiation among breeds, estimated using F_{ST} , ranged from 0.01 to 0.09 (Figure 2C), indicating generally low levels of genetic differentiation. The lowest F_{ST} value was observed between MN and SN (0.01), followed by AN and MN (0.02), reflecting very low genetic differentiation among these breeds. In contrast, the highest differentiation was observed between HF and NF, HF and SA, and NF and SA (0.09).

The IBS distance matrix showed clear grouping of individuals by breed, with distinct diagonal blocks reflecting higher within-breed similarity and higher distances between breeds (Figure 2D). In the NeighborNet network based on Nei's genetic distances, MN showed the lowest genetic distances to SN (0.016) and AN (0.019), followed by BD (0.026), indicating close genetic relationships between MN and these breeds (Figure 2E). In contrast, higher genetic distances were observed between HF and SA (0.058) and HF and NF (0.051), reflecting higher genetic differentiation among these populations.

Across k values of 5, 10 and 15, network-based clustering revealed a consistent pattern of fine-scale relationships among individuals (Figure 3), with individuals generally grouped according to breed and the resulting clusters remaining largely consistent across different network densities. Individuals from MN and SN formed distinct but interconnected clusters, with additional links observed to AN. In contrast, BD formed a separate cluster without direct connections to these breeds. NF, HF, SA, QH, and HC formed clearly distinct clusters with limited connections to other populations.

Global admixture patterns. Because $K = 1$ represents a single undifferentiated cluster and was included only for model evaluation, graphical results are shown from $K = 2$ onwards (ESM Figure S2). At low K values, the inferred structure reflected broad among-breed differentiation. At $K = 2$, a distinct ancestry component was most evident in QH and SA, whereas the remaining breeds were assigned predominantly to the alternative component. At $K = 3$ and $K = 4$, additional differentiation emerged mainly in HC, HF and NF.

Between $K = 5$ and $K = 8$, several breeds displayed increasingly distinct breed-associated ancestry components, particularly HF, HC, NF, QH, and SA. In contrast, MN showed a more heterogeneous ancestry composition, with shared components mainly involving AN and SN. At $K = 9$, corresponding to the number of analysed breeds, most populations showed a predominant breed-related ancestry component. However, MN did not form a fully dis-

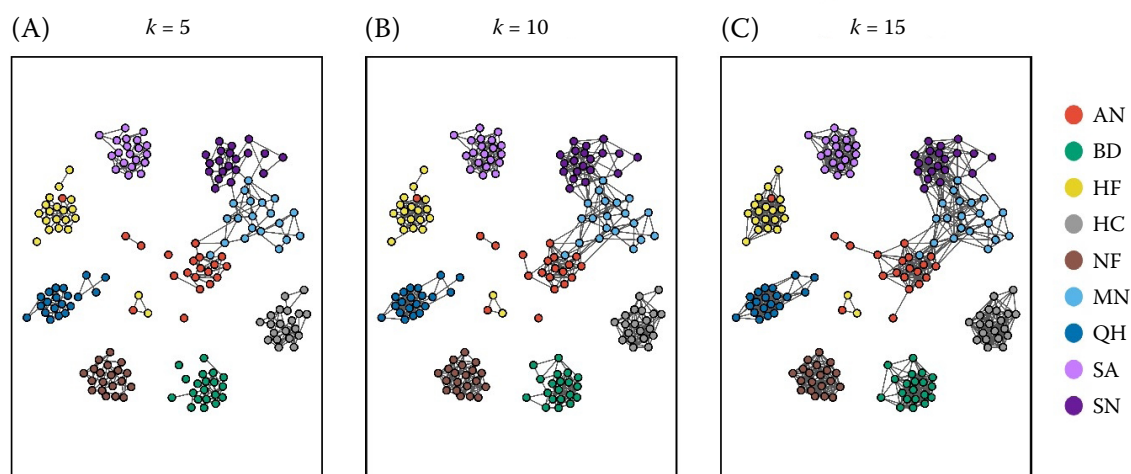


Figure 3. Network-based clustering of individuals from nine breeds using mutual k -nearest neighbour networks at $k = 5, 10$ and 15

AN = Austrian Noriker; BD = Belgian Draught; HC = Hucul; HF = Haflinger; MN = Norik of Muran; NF = Norwegian Fjord; QH = Quarter Horse; SA = Shagya Arabian; SN = Silesian Noriker

tinct and homogeneous group, as its ancestry was divided almost equally between AN and SN, with mean ancestry fractions of 45.5% and 45.4%, respectively. From $K = 10$ onwards, additional structuring was observed within several populations, including MN, SA, QH and AN. At $K = 11$, MN showed a more clearly differentiated ancestry component. However, clustering solutions at higher K values became progressively less stable and increasingly fragmented into multiple modes, indicating reduced agreement among independent runs and lower reproducibility of the inferred ancestry patterns.

The lowest cross-validation error (0.540 17) was observed at $K = 7$, indicating this as the most supported clustering solution (ESM Figure S3). At this K , QH, SA, HC and HF showed largely homogeneous ancestry profiles, whereas MN did not form a separate component and remained mainly assigned to components shared with AN and SN (ESM Figure S2).

Local ancestry of Norik of Muran. Local ancestry inference assigned the largest genome-wide proportions in MN to the SN (40.61%) and AN (39.61%) reference populations, while smaller proportions were assigned to BD (9.23%), HC (3.47%), HF (2.31%), NF (2.04%), QH (1.76%), and SA (0.97%). Detailed genome-wide and chromosome-specific proportions are reported in ESM Table S3. Dominant ancestry patterns across SNP positions showed that AN and SN ancestry components were distributed across all autosomes and represented the predominant ancestry pattern throughout the genome (Figure 4A). Overlap analysis between dominant ancestry segments and QTL regions identified 1 194 overlaps for AN and 1 172 overlaps for SN, whereas two overlaps were detected for BD. When expressed relative to the total length of dominant ancestry segments, AN and SN showed almost identical QTL overlap densities, with 1.08 overlaps per Mb for AN and 1.08 overlaps per Mb for SN, and 0.90 overlaps per cM for both ancestry sources (ESM Table S4).

In AN segments, the highest number of overlaps (Figure 4B,C) was observed for body height QTLs ($n = 340$), followed by alternate gaits QTLs ($n = 168$), guttural pouch tympany QTLs ($n = 116$), and insect bite hypersensitivity QTLs ($n = 99$). A similar pattern was observed in SN segments, where body height QTLs were also the most frequent ($n = 243$), followed by alternate gaits QTLs ($n = 125$), insect bite hypersensitivity QTLs ($n = 112$), and guttural pouch tympany QTLs ($n = 98$).

The proportion assigned to the BD reference population increased markedly on chromosome 20, reaching 27.62% compared with 9.23% genome-wide. Within this chromosome, one dominant ancestry segment assigned to the BD reference population was identified between 44.39 and 45.34 Mb, corresponding to 71.02–72.51 cM. The segment spanned 0.95 Mb and 1.49 cM and was supported by 18 SNPs. The segment overlapped two QTLs annotated for back conformation and female fertility.

Haplotype sharing. Among breed comparisons involving MN, haplotype sharing was strongest with SN and AN (Figure 5A; ESM Table S5). MN copied on average 108.38 chunks from SN and 108.14 chunks from AN, with corresponding mean chunk lengths of 32.00 cM and 32.05 cM, respectively. Reciprocal sharing was also high, with SN and AN copying 107.34 and 106.80 chunks from MN, respectively. Lower values were observed when MN copied from BD (103.15 chunks), HF (101.36), NF (99.07), HC (98.56), QH (96.49) and SA (95.66).

Within-breed haplotype sharing was generally higher than between-breed sharing, as shown by the elevated diagonal values in the breed-level coancestry matrix (Figure 5A; ESM Table S5). Mean chunk counts ranged from 115.18 to 119.52 in MN, AN and SN, compared with 126.72 to 143.11 among the other six breeds.

Mean chunk lengths supported the same pattern, ranging from 35.28 to 37.89 cM in MN, AN and SN compared with 41.38 to 51.60 cM among the other six breeds.

Although the matrix was ordered according to fineSTRUCTURE clustering rather than breed membership, most clusters largely corresponded to recorded breed membership, with compact breed-specific blocks most evident for SA, NF, QH, HC, and BD (Figure 5B). In contrast, MN, AN, and SN showed less clearly separated clustering and higher haplotype sharing among each other, consistent with the breed-level results. The haplotype-based clustering tree showed the same clustering pattern (ESM Figure S4). A total of six individuals were placed outside the main cluster corresponding to their recorded breed, indicating discordance between breed assignment and haplotype-based clustering for these samples. Support values at internal branching points were high for several breed-specific groups, whereas lower support values occurred at some splits involving closely related or partially overlapping clusters.

<https://doi.org/10.17221/75/2026-CJAS>

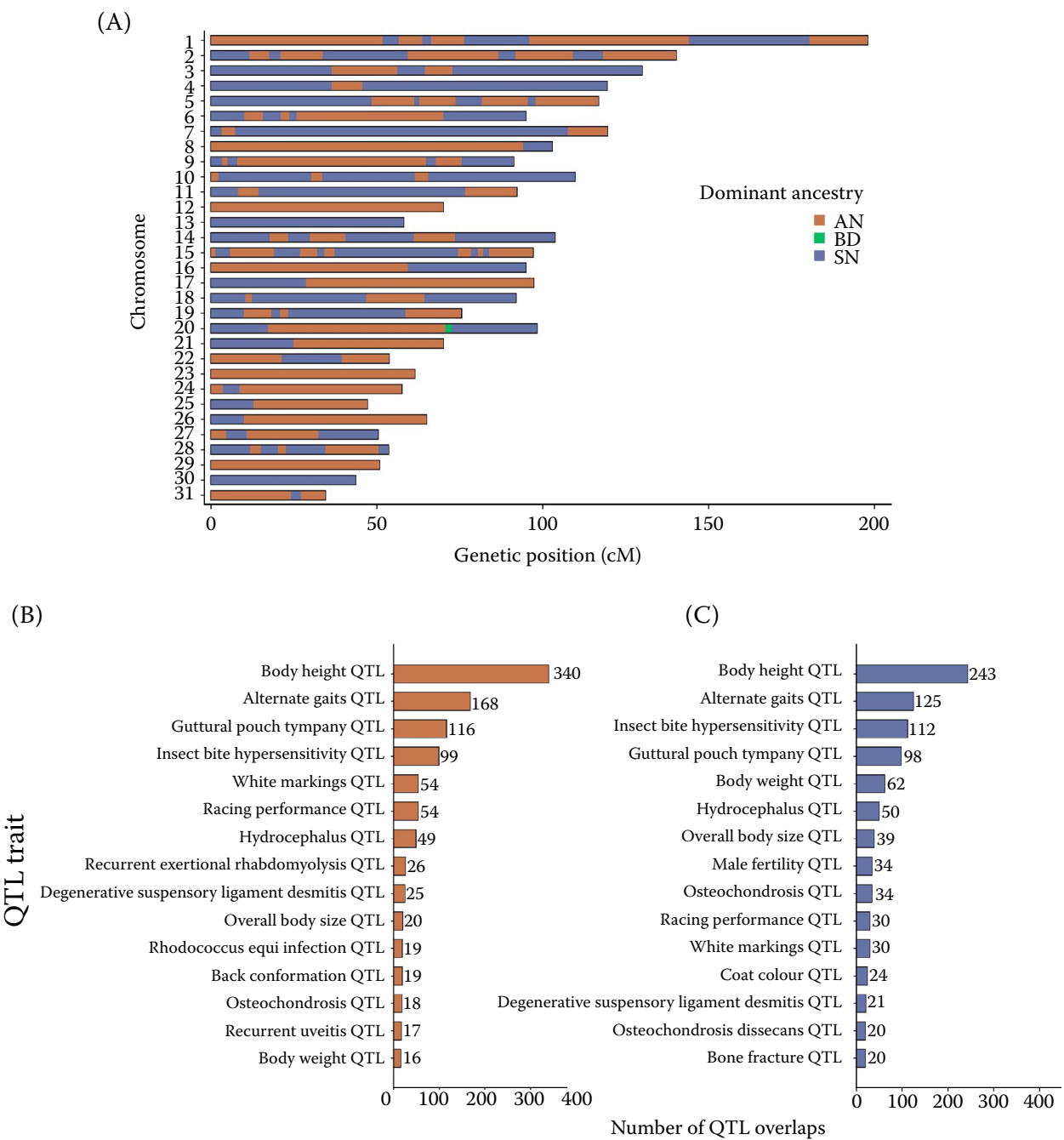


Figure 4. Dominant local ancestry patterns across SNP positions (A) and counts of overlapping QTL regions (B,C) in the MN genome

AN = Austrian Noriker; BD = Belgian Draught; cM = centimorgan; MN = Norik of Muran; QTL = quantitative trait locus; SN = Silesian Noriker; SNP = single nucleotide polymorphism

Within-breed structure of Norik of Muran. The reconstructed recent family structure contained 133 individuals, comprising 125 genotyped individuals and eight inferred non-genotyped dummy parents. Both parents were assigned to 38 genotyped individuals, representing 30.4% of the input individuals, whereas no individuals were assigned

only one parent. In total, 76 parent-offspring relationships were reconstructed, consisting of 38 maternal and 38 paternal relationships. The remaining 87 genotyped individuals had no assigned parents (Figure 6).

A total of 37 unique parents were identified, including 29 genotyped parents and eight inferred

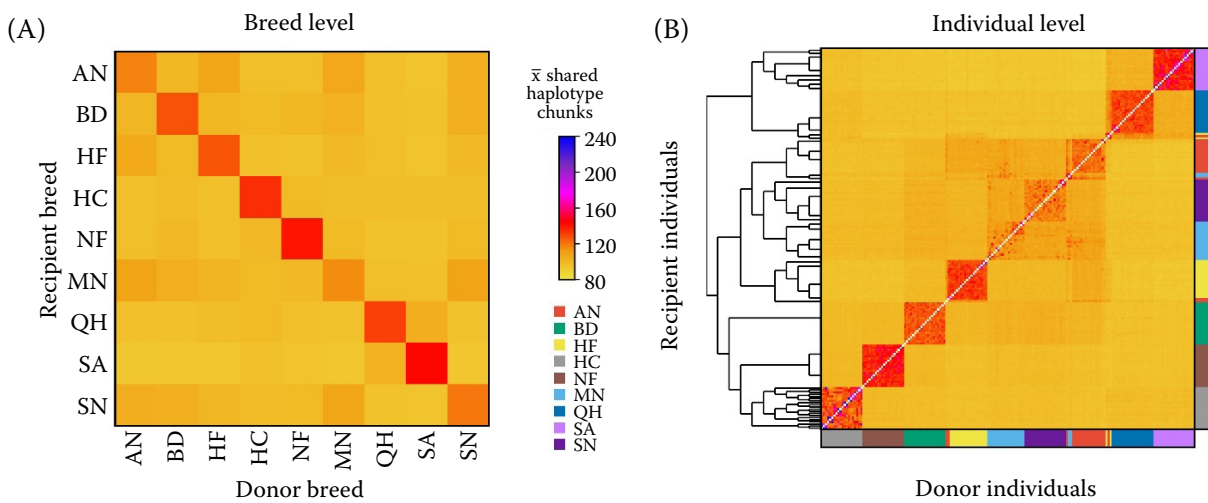


Figure 5. Breed-level haplotype sharing patterns (A) and individual-level haplotype sharing (B)

AN = Austrian Noriker; BD = Belgian Draught; HC = Hucul; HF = Haflinger; MN = Norik of Muran; NF = Norwegian Fjord; QH = Quarter Horse; SA = Shagya Arabian; SN = Silesian Noriker

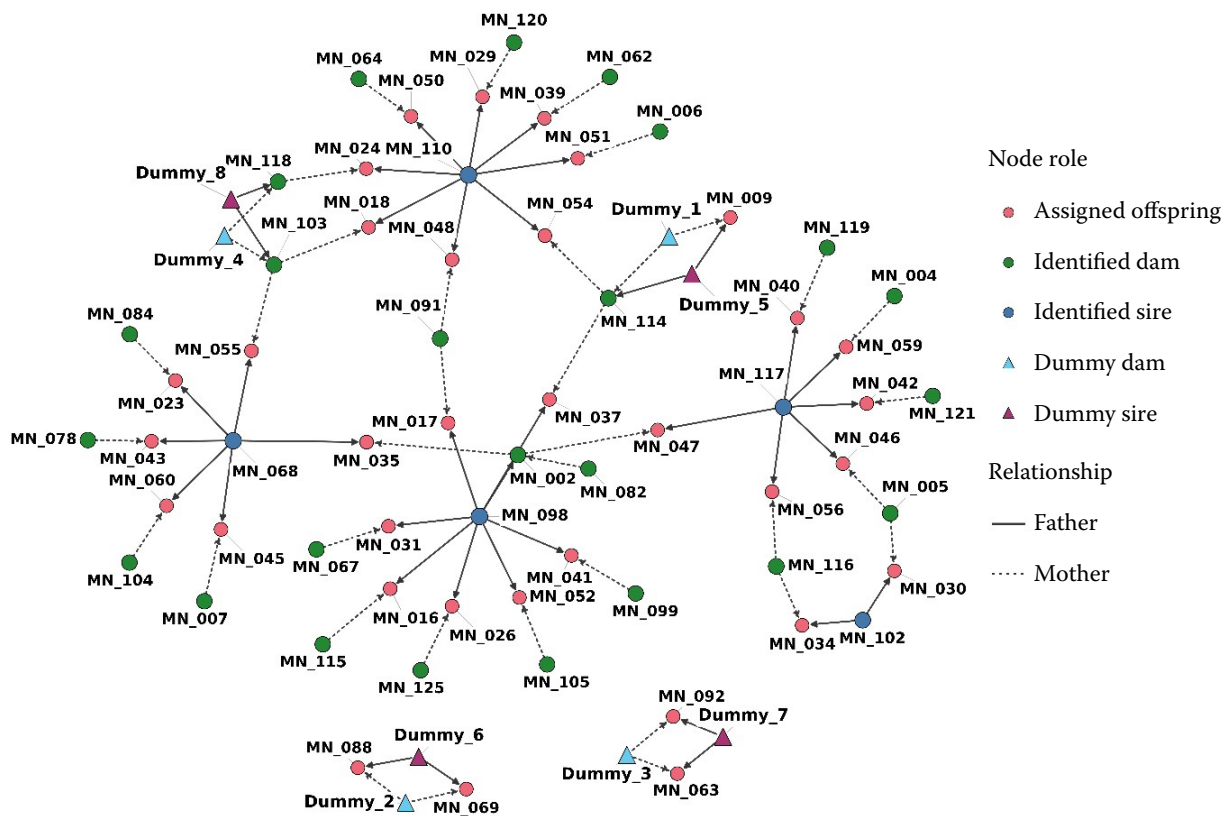


Figure 6. Genomically reconstructed family network of the MN population based on parentage and sibship reconstruction

MN = Norik of Muran

dummy parents. Among the genotyped parents, 24 were dams, and five were sires, whereas the inferred dummy parents consisted of four dummy dams and four dummy sires. The reconstructed

pedigree contained nine paternal and 10 maternal half-sibling groups (ESM Figure S5). Among these, four paternal groups were associated with dummy sires, whereas four maternal groups were

<https://doi.org/10.17221/75/2026-CJAS>

associated with dummy dams. The largest paternal half-sibling group included eight offspring, whereas all maternal half-sibling groups contained two offspring.

DISCUSSION

Dataset preparation for population structure analyses. Filtering related individuals before downstream genomic analyses was necessary to reduce the risk that within-family similarity would bias estimates of genomic relationships, allele frequencies and breed-level differentiation. This is important because unbiased sampling is essential for accurate population-genetic inference, and the overrepresentation of relatives may bias allele-frequency estimates, which form the basis of many downstream diversity and differentiation analyses (Peterman et al. 2016; Ramstetter et al. 2017). Moreover, recent family relationships and broader population structure can both be reflected in allele sharing patterns, making it important to reduce close kinship when the objective is to describe genomic relationships among breeds, not family structure within breeds (Conomos et al. 2016). Therefore, both global and pairwise genomic relatedness derived from the genomic relatedness matrix were considered during individual selection. This approach is consistent with the use of genome-wide marker data to estimate realised genomic relationships among individuals and to minimise the influence of close relatives when pedigree information is incomplete or unavailable (Galla et al. 2020). Although MN is an endangered population (FAO 2022), a larger number of available genotyped animals made it possible to select a subset with relatively low genomic relatedness. However, this reflects the availability of animals for selection and the design of the analytical dataset rather than low relatedness across the entire breed.

A classical recommendation for molecular characterisation of animal genetic resources is to sample 30–50 unrelated individuals per population (Tixier-Boichard et al. 2008). However, in the present study, the number of selected individuals was constrained by the smallest breed group, NF, which included 20 individuals. This number also meets the minimum recommendation that at least 20 animals per breed should be genotyped for reliable estimation of allele frequencies for bial-

lelic markers such as SNPs (Ajmone-Marsan et al. 2023). This sample size is consistent with population-genomic sampling evaluations showing that genetic diversity estimates become relatively stable when more than eight individuals and at least 1 000 SNPs are used (Nazareno et al. 2017). Given that the present dataset retained thousands of SNPs after quality control, selecting 20 individuals per breed with reduced within-breed genomic relatedness represented a practical compromise between sample size, SNP density and the reduction of relatedness-driven bias. Equalising breed sample sizes also improved comparability among breeds and reduced the risk that overrepresented populations would disproportionately influence population-structure inference, a problem previously shown under uneven sampling designs (Puechmaille 2016). Balanced sampling is particularly important for ancestry and admixture analyses because sampling schemes can bias inferred admixture proportions (Toyama et al. 2020). Therefore, the balanced subset of 20 minimally related individuals per breed was not intended to maximise sample size, but to maximise comparability among breeds and reduce the influence of relatedness between animals on genome-wide estimates. This approach represented a compromise between comparability and representativeness, as equal sample sizes facilitated between-breed comparisons but may not fully capture the within-breed variability of each population. This limitation is particularly relevant for breeds represented by relatively few animals in the original dataset, where the selected subset may be more strongly influenced by the particular individuals available and consequently may be less representative of the breed as a whole.

In addition to balanced sampling, LD pruning was applied before PCA and global admixture analysis to reduce marker redundancy and limit the disproportionate influence of highly correlated genomic regions on population structure inference. This follows recommendations for workflows combining PCA and admixture proportion inference, where pruning markers in linkage disequilibrium is used to obtain a more balanced genome-wide representation of genetic variation (Liu et al. 2020).

Despite these steps to improve comparability among breeds, the different SNP array densities used in the source datasets remain a limitation, particularly because the common marker set is constrained by the lower-density platforms. Chessari

et al. (2025) reported high concordance in genetic relationships and broad population structure between approximately 40K SNP and sequence-level imputed datasets in horses, while the higher marker density provided greater resolution for detecting more localised genomic patterns. However, uniformly high-density genomic data are not always readily available for animal genetic resources, as high-density genotyping and whole-genome sequencing remain costly (Crooijmans et al. 2025). In the present study, the shared marker set therefore represented a practical compromise based on the available genotype data.

Genomic position of Norik of Muran among related breeds. The complementary analyses provided consistent evidence of gradual genomic differentiation between MN, SN and AN rather than clear separation. Across allele frequency-based and distance-based analyses, MN consistently showed the closest genomic relationship with SN and AN. The low pairwise differentiation, overlapping PCA position, network connectivity and shared ancestry components indicate that MN shares a close genomic background with the Noriker breeds AN and SN, without forming a fully isolated cluster. This pattern agrees with previous evidence from Central European draught horses, where Austrian Noriker populations and related cold blood populations were grouped within a distinct Noriker-related cluster (Druml et al. 2007). However, this shared genomic background should be viewed in the context of related but separately managed breeds and not as variation within a single cosmopolitan breed.

Vostra-Vydrova et al. (2018) analysed Silesian Noriker, Noriker and Czech Moravian Belgian horses and reported low genetic differentiation and substantial admixture among these historically connected breeds, especially between Silesian Noriker and Noriker. Their results showed that when breed formation has involved shared ancestry, gene flow and related breeding histories, population structure analyses may reveal gradual differentiation rather than clearly separated breed clusters. The absence of a distinct MN cluster should not be interpreted as the absence of breed identity; a comparable pattern in the Franches-Montagnes horse shows that native breeds with documented crossbreeding histories may display fine-scale substructure and high ancestry contributions from introgressed lineages, with admixed individuals explaining much of the genomic variance of the current population

(Gmel et al. 2024). Although less directly comparable, SNP-based ancestry analysis of Theodore Roosevelt National Park horses also illustrates that draft-breed affinities can be recovered in horse populations with complex management histories (Thompson et al. 2024). These genomic patterns are consistent with the documented breeding history of MN, including historical contributions from Noriker-related and other draught horse populations (Kasarda et al. 2019; Moravcikova et al. 2019).

However, the documented breeding history of MN may not fully explain the close genomic similarity observed between MN, AN and SN. In horse populations, past and current breeding practices can maintain genetic similarity between otherwise distinct populations when breeding programmes permit continued genetic exchange (Schurink et al. 2019), while the exchange of genetic material may also be used as a breeding strategy to increase effective population size and maintain genetic diversity in small or structured horse populations (Sole et al. 2018). Genetic exchange between breeding populations can also have lasting effects on present-day genomic relationships, as historical introgression has been shown to contribute to genomic similarity among horse breeds (Capomaccio et al. 2023). Shared breeding goals and historical gene flow can likewise leave persistent genomic similarities even among separately recognised breeds (Vostra-Vydrova et al. 2026). The close genomic relationship among MN, AN and SN was further supported by haplotype sharing analysis, which quantified shared haplotype segments among individuals and breeds at a finer genomic scale. In the present study, the highest between-breed chunk counts and longer shared chunks involving MN were observed with AN and SN, supporting close haplotypic relationships among these breeds, consistent with their documented historical connections. This fine-scale perspective is particularly useful in native and historically connected breeds, where genomic differentiation may be low, and breed boundaries can remain gradual. In Swedish native cattle, haplotype sharing analysis reinforced previous population-structure results while refining the clustering of weakly differentiated mountain cattle breeds, demonstrating the relevance of this approach for closely related livestock populations (Upadhyay et al. 2019). Comparable use has been applied in horse population genomics, where ChromoPainter/fineSTRUCTURE was used along-

<https://doi.org/10.17221/75/2026-CJAS>

side classical structure analyses to describe relationships among breeds through donor–recipient haplotype chunk sharing, allowing recent coancestry to be evaluated at a finer genomic scale than allele-frequency clustering alone (Asadollahpour Nanaei et al. 2020). A similar pattern was reported in Balkan sheep, where fineSTRUCTURE based on haplotype sharing confirmed incomplete differentiation among Pramenka populations and revealed intermingling among several regional breeds (Ciani et al. 2020).

Local ancestry and descriptive annotation of dominant ancestry segments. Local ancestry inference provided chromosome-level resolution of ancestry assignments within MN, making it possible to identify genomic regions assigned predominantly to individual reference populations and to compare these regions with previously reported QTLs. The choice of RFMix followed from comparative benchmarking of local ancestry inference tools, where RFMix and ELAI achieved the highest median accuracy in two-way admixture scenarios depending on population structure, whereas RFMix, MOSAIC and LAMP LD showed the highest median accuracy in three-way admixture scenarios, with RFMix performing best overall in the three-way simulated admixed population (Schubert et al. 2020). However, the reference populations represent present-day breeds rather than the historical populations that contributed to the formation of MN. The inferred ancestry assignments should therefore be interpreted as genomic affinity to the selected reference populations rather than as direct estimates of historical breed contributions.

The QTL overlap analysis provides descriptive functional context for dominant ancestry segments, but it does not test for QTL enrichment or demonstrate functional causality. In both AN and SN segments, the most frequent overlaps involved QTLs annotated for body height. These overlaps indicate that the dominant ancestry segments coincided with genomic regions previously associated with horse body size, but they do not demonstrate a direct effect of AN or SN ancestry on this trait. Body size, particularly withers height, is one of the most extensively studied morphological traits in horses, and previous genome-wide association studies have repeatedly identified the LCORL/NCAPG region as an important locus for withers height and body size variation (Tetens et al. 2013; Vosgerau et al. 2022; Reich et al. 2024).

Similarly, overlaps with alternate gait QTLs provide biological context for these ancestry segments, because this QTL category refers to genomic regions previously associated with locomotor variation in horses. Andersson et al. (2012) identified a premature stop codon in the DMRT3 gene with a major effect on locomotion pattern and alternate gait ability, supporting the relevance of this QTL category. Overlaps with insect bite hypersensitivity and guttural pouch tympany QTLs provide health-related annotations of AN and SN dominant ancestry segments. Insect bite hypersensitivity is a complex allergic skin disease influenced by both genetic and environmental factors, and previous genome-wide association studies have identified several associated genomic regions in horses (Shrestha et al. 2015; Velie et al. 2016; Schurink et al. 2018). Guttural pouch tympany has been described as a hereditary disorder affecting foals during the first months of life, and genome-wide linkage and association analyses identified major loci for this condition on ECA3 in German Warmblood horses and ECA15 in Arabian horses (Metzger et al. 2012).

Within-breed structure in Norik of Muran. The SNP-based reconstruction of recent family relationships revealed that the current MN population contains identifiable paternal and maternal structure. The R package *sequoia* was well suited for this within-breed analysis because it not only assigns direct parent–offspring relationships but also reconstructs half-sibship structure and unsampled parental links, features that were essential for identifying recent paternal and maternal relationships within MN (Huisman 2017).

The presence of paternal and maternal half-sibling groups indicates that the current population is not composed of unrelated individuals but retains recent family structure. However, because birth year information and the complete official pedigree of the genotyped individuals were unavailable, the reconstructed pedigree should be viewed primarily as a reconstruction of recent family structure, not as a complete multigenerational pedigree. The more concentrated paternal structure compared with the maternal structure suggests non-equal sires' contribution to the current population, whereas maternal contributions appear more broadly distributed. This pattern is consistent with horse breeding systems in which a limited number of influential sires can contribute strongly to the next generation, a practice that has been

linked to concerns about declining genetic diversity in horse populations (McGivney et al. 2020).

The reliability of this reconstruction also depended on the use of an informative marker subset and the evaluation of genotyping error. The complete SNP dataset was reduced to an informative subset of approximately independent markers. This followed *sequoia* recommendations to avoid large numbers of redundant markers and to prioritise SNPs with high call rate, low linkage disequilibrium and high minor allele frequency (Huisman 2024). This approach is supported by a recent genomic pedigree study showing that accurate pedigree inference depends on balancing marker quantity and quality, particularly MAF and call rate (Noda et al. 2025). For example, in chickens, a panel of 700 SNPs with MAF = 0.05 achieved perfect pedigree reconstruction accuracy (Zhou et al. 2024). Because the true genotyping error rate was unknown, several assumed error rates were evaluated to account for the possibility that genotyping errors could increase both false-positive and false-negative assignments in pedigree reconstruction; the identical parent-offspring assignments obtained across these settings indicate that the reconstructed MN pedigree was robust to plausible genotyping-error assumptions (Huisman 2024).

From a conservation perspective, these results indicate that information on recent family structure may complement genome-wide diversity estimates when informing MN management. The reconstructed relationships identify animals connected through parent-offspring and half-sibling relationships, which can support mating decisions aimed at avoiding repeated use of closely related breeding animals. A recent study of native horses based on genome-wide SNP data similarly emphasised that information on genetic diversity and inbreeding can support breeding decisions aimed at maintaining genetic diversity and preserving breed-specific characteristics (Asti et al. 2025). In MN, reconstructed family relationships may therefore provide an additional basis for conservation planning by helping to maintain available within-breed variation and reduce the risk of future genetic diversity loss.

CONCLUSION

The main result of this study is that the Norik of Muran is not a genomically isolated population,

but a distinct local breed with a specific genomic profile characterised by close genomic relationships with Silesian Noriker and Austrian Noriker and recent within-breed family structure. Across the analyses, the breed is still retaining a relatively distinct population profile within the comparative dataset. Local ancestry and haplotype-sharing results further showed that these relationships are not limited to broad allele-frequency patterns but are also reflected in chromosome-level ancestry assignments relative to the selected reference populations and fine-scale haplotypic sharing. The reconstruction of recent family relationships identified paternal and maternal half-sibling structure, with a more concentrated paternal contribution, highlighting an important aspect of current breed management. Together, these results indicate that the conservation value of the Norik of Muran lies in its particular combination of Noriker-related genomic affinity, genomic heterogeneity and recent family structure, without requiring complete genetic separation. Conservation management may therefore benefit from considering both the breed-level genomic background and recent family structure of Norik of Muran, particularly by accounting for its close links with Silesian Noriker and Austrian Noriker, and by using the reconstructed family structure to inform decisions aimed at avoiding mating of close relatives, limiting repeated use of the same sire line and preserving several paternal and maternal families in future breeding.

Acknowledgement

The authors thank all breeders, institutions and collaborators who contributed to sample collection and data acquisition for the horse populations analysed in this study.

Conflict of interest

The authors declare no conflict of interest.

REFERENCES

Ajmone-Marsan P, Boettcher PJ, Colli L, Ginja C, Kantanen J, Lenstra JA, editors. Genomic characterization of animal genetic resources - Practical guide. FAO Animal

<https://doi.org/10.17221/75/2026-CJAS>

- Production and Health Guidelines No. 32. Rome: FAO; 2023. 185 p.
- Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009 Sep;19(9):1655-64.
- Andersson LS, Larhammar M, Memic F, Wootz H, Schwachow D, Rubin CJ, Patra K, Arnason T, Wellbring L, Hjalms G, Imsland F, Petersen JL, McCue ME, Mickelson JR, Cothran G, Ahituv N, Roepstorff L, Mikko S, Vallstedt A, Lindgren G, Andersson L, Kullander K. Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. *Nature.* 2012 Aug 30;488(7413):642-6.
- Asadollahpour Nanaei H, Esmailzadeh A, Ayatollahi Mehrgardi A, Han J, Wu DD, Li Y, Zhang YP. Comparative population genomic analysis uncovers novel genomic footprints and genes associated with small body size in Chinese pony. *BMC Genomics.* 2020 Jul 20;21(1):496.
- Asti V, Summer A, Ablondi M, Sartori C, Giontella A, Pilastro V, Mecocci S, Cappelli K, Mancin E, Oian A, Mantovani R, Capomaccio S, Sabbioni A. Selection signatures and inbreeding: Exploring genetic diversity in five native horse breeds. *BMC Vet Res.* 2025 May 16;21(1):346.
- Beeson SK, Mickelson JR, McCue ME. Exploration of fine-scale recombination rate variation in the domestic horse. *Genome Res.* 2019 Oct;29(10):1744-52.
- Behr AA, Liu KZ, Liu-Fang G, Nakka P, Ramachandran S. pong: Fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics.* 2016 Sep 15;32(18):2817-23.
- Capomaccio S, Ablondi M, Colombi D, Sartori C, Giontella A, Cappelli K, Mancin E, Asti V, Mantovani R, Sabbioni A, Silvestrelli M. Exploring the Italian equine gene pool via high-throughput genotyping. *Front Genet.* 2023 Jan 23;14:1099896.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience.* 2015 Feb 25;4:7.
- Chessari G, Reich P, Criscione A, Falker-Gieske C, Mastrangelo S, Tumino S, Bordonaro S, Marletta D, Tetens J. Comparison between SNP array and imputed data to estimate population structure and ROH hotspots in horse breeds. *BMC Genomics.* 2025 Nov 29;26(1):1086.
- Ciani E, Mastrangelo S, Da Silva A, Marroni F, Ferencakovic M, Ajmone-Marsan P, Baird H, Barbato M, Colli L, Delvento C, Dovenski T, Gorjanc G, Hall SJG, Hoda A, Li MH, Markovic B, McEwan J, Moradi MH, Ruiz-Larranaga O, Ruzic-Muslic D, Salamon D, Simcic M, Stepanek O; Econogene Consortium; Sheephapmap Consortium; Curik I, Cubric-Curik V, Lenstra JA. 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 May 14;52(1):25.
- Conomos MP, Reiner AP, Weir BS, Thornton TA. Model-free estimation of recent genetic relatedness. *Am J Hum Genet.* 2016 Jan 7;98(1):127-48.
- Crooijmans RPMA, Gonzalez Prendes R, Colli L, Del Corvo M, Barbato M, Somenzi E, Tosser-Klopp G, Meszaros G, Ajmone-Marsan P, Weigend S, Wallner B, McCue ME, Orlando L, Bradley D, Hiemstra SJ, Schokker D, Peynot N, Stella A, Restoux G, Groenen MAM, Tixier-Boichard M. IMAGE001: A new livestock multispecies SNP array to characterize genomic variation in European livestock gene bank collections. *Anim Genet.* 2025 Oct;56(5):e70039.
- Delaneau O, Howie B, Cox AJ, Zagury JE, Marchini J. Haplotype estimation using sequencing reads. *Am J Hum Genet.* 2013 Oct 3;93(4):687-96.
- Druml T, Curik I, Baumung R, Aberle K, Distl O, Solkner J. Individual-based assessment of population structure and admixture in Austrian, Croatian and German draught horses. *Heredity (Edinb).* 2007 Feb;98(2):114-22.
- FAO. Domestic Animal Diversity Information System (DAD-IS) [Internet]. Rome: FAO; 2022 [cited 2026 May 15]. Available from: www.fao.org/dad-is/en/.
- Galla SJ, Moraga R, Brown L, Cleland S, Hoepfner MP, Maloney RE, Richardson A, Slater L, Santure AW, Steeves TE. A comparison of pedigree, genetic and genomic estimates of relatedness for informing pairing decisions in two critically endangered birds: Implications for conservation breeding programmes worldwide. *Evol Appl.* 2020 Jan 27;13(5):991-1008.
- Gmel AI, Mikko S, Ricard A, Velie BD, Gerber V, Hamilton NA, Neuditschko M. Using high-density SNP data to unravel the origin of the Franches-Montagnes horse breed. *Genet Sel Evol.* 2024 Jul 10;56(1):53.
- Halo M, Mlynekova E, Horna M, Ivancikova M, Hrda A. Evaluation of genetic variability of the breed Norik of Muran according to pedigree information. *Czech J Anim Sci.* 2018;63(5):195-200.
- Hu ZL, Park CA, Reecy JM. Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic Acids Res.* 2019 Jan 8;47(D1):D701-10.
- Huisman J. Pedigree reconstruction from SNP data: Parentage assignment, sibship clustering and beyond. *Mol Ecol Resour.* 2017 Sep;17(5):1009-24.
- Huisman J. Sequoia user guide: Reconstruction of multi-generational pedigrees from SNP data [Internet]. 2024 May 17 [cited 2026 May 15]. Available from: cran.r-project.org/web/packages/sequoia/vignettes/.

- Kardos M, Taylor HR, Ellegren H, Luikart G, Allendorf FW. Genomics advances the study of inbreeding depression in the wild. *Evol Appl*. 2016 Oct 23;9(10):1205-18.
- Kasarda R, Moravcikova N, Kadlecik O, Trakovicka A, Halo M, Candrak J. Level of inbreeding in Norik of Muran horse: Pedigree vs. genomic data. *Acta Univ Agric Silvic Mendeliana Brun*. 2019;67(6):1457-63.
- Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of population structure using dense haplotype data. *PLoS Genet*. 2012 Jan;8(1):e1002453.
- Li YL, Liu JX. StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Mol Ecol Resour*. 2018 Jan;18(1):176-7.
- Liu CC, Shringarpure S, Lange K, Novembre J. Exploring population structure with admixture models and principal component analysis. *Methods Mol Biol*. 2020;2090:67-86.
- Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: A discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet*. 2013 Aug 8; 93(2):278-88.
- McGivney BA, Han H, Corduff LR, Katz LM, Tozaki T, MacHugh DE, Hill EW. Genomic inbreeding trends, influential sire lines and selection in the global Thoroughbred horse population. *Sci Rep*. 2020 Jan 16;10(1):466.
- Metzger J, Ohnesorge B, Distl O. Genome-wide linkage and association analysis identifies major gene loci for guttural pouch tympany in Arabian and German warmblood horses. *PLoS One*. 2012;7(7):e41640.
- Moravcikova N, Kasarda R, Kadlecik O, Trakovicka A, Halo M, Candrak J. Runs of homozygosity as footprints of selection in the Norik of Muran horse genome. *Acta Univ Agric Silvic Mendeliana Brun*. 2019;67(5): 1165-70.
- Nazareno AG, Bemmels JB, Dick CW, Lohmann LG. Minimum sample sizes for population genomics: An empirical study from an Amazonian plant species. *Mol Ecol Resour*. 2017 Nov;17(6):1136-47.
- Neuditschko M, Khatkar MS, Raadsma HW. NetView: A high-definition network-visualization approach to detect fine-scale population structures from genome-wide patterns of variation. *PLoS One*. 2012;7(10):e48375.
- Noda S, Akita T, Ueda R, Katsumura T, Hashiguchi Y, Takeshima H, Sato T. Optimal SNP filtering strategies for pedigree reconstruction: A case study with wild red-spotted masu salmon population. *Popul Ecol*. 2025 Apr; 67(2):125-39.
- Nogueira MB, de Faria DA, Ianella P, Paiva SR, McManus C. Genetic diversity and population structure of locally adapted Brazilian horse breeds assessed using genome-wide single nucleotide polymorphisms. *Livest Sci*. 2022 Oct;264:105071.
- Oldenbroek JK. The use of genomic information for the conservation of animal genetic diversity. *Animals (Basel)*. 2021 Nov 10;11(11):3208.
- Pembleton LW, Cogan NO, Forster JW. StAMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Mol Ecol Resour*. 2013 Sep;13(5):946-52.
- Peterman W, Brocato ER, Semlitsch RD, Eggert LS. Reducing bias in population and landscape genetic inferences: The effects of sampling related individuals and multiple life stages. *PeerJ*. 2016 Mar 14;4:e1813.
- Petersen JL, Mickelson JR, Rendahl AK, Valberg SJ, Andersson LS, Axelsson J, Bailey E, Bannasch D, Binns MM, Borges AS, Brama P, da Camara Machado A, Capomacchio S, Cappelli K, Cothran EG, Distl O, Fox-Clipsham L, Graves KT, Guerin G, Haase B, Hasegawa T, Hemmann K, Hill EW, Leeb T, Lindgren G, Lohi H, Lopes MS, McGivney BA, Mikko S, Orr N, Penedo MC, Piercy RJ, Raekallio M, Rieder S, Roed KH, Swinburne J, Tozaki T, Vaudin M, Wade CM, McCue ME. Genome-wide analysis reveals selection for important traits in domestic horse breeds. *PLoS Genet*. 2013;9(1):e1003211.
- Puechmaille SJ. The program structure does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem. *Mol Ecol Resour*. 2016 May;16(3):608-27.
- Ramstetter MD, Dyer TD, Lehman DM, Curran JE, Dugger R, Blangero J, Mezey JG, Williams AL. Benchmarking relatedness inference methods with genome-wide data from thousands of relatives. *Genetics*. 2017 Sep; 207(1):75-82.
- Reich P, Moller S, Stock KF, Nolte W, von Depka Prondzinski M, Reents R, Kalm E, Kuhn C, Thaller G, Falker-Gieske C, Tetens J. Genomic analyses of withers height and linear conformation traits in German Warmblood horses using imputed sequence-level genotypes. *Genet Sel Evol*. 2024 Jun 13;56(1):45.
- Schliep KP. phangorn: Phylogenetic analysis in R. *Bioinformatics*. 2011 Feb 15;27(4):592-3.
- Schubert R, Andaleon A, Wheeler HE. Comparing local ancestry inference models in populations of two- and three-way admixture. *PeerJ*. 2020 Oct 2;8:e10090.
- Schurink A, da Silva VH, Velie BD, Dibbitts BW, Crooijmans RPMA, Francois L, Janssens S, Stinckens A, Blott S, Buys N, Lindgren G, Ducro BJ. Copy number variations in Friesian horses and genetic risk factors for insect bite hypersensitivity. *BMC Genet*. 2018 Jul 30;19(1):49.
- Schurink A, Shrestha M, Eriksson S, Bosse M, Bovenhuis H, Back W, Johansson AM, Ducro BJ. The genomic makeup of nine horse populations sampled in the Netherlands. *Genes (Basel)*. 2019 Jun 25;10(6):480.

<https://doi.org/10.17221/75/2026-CJAS>

- Shrestha M, Eriksson S, Schurink A, Andersson LS, Sundquist M, Frey R, Brostrom H, Bergstrom T, Ducro B, Lindgren G. Genome-wide association study of insect bite hypersensitivity in Swedish-born Icelandic horses. *J Hered.* 2015 Jul-Aug;106(4):366-74.
- Sole M, Valera M, Fernandez J. Genetic structure and connectivity analysis in a large domestic livestock metapopulation: The case of the Pura Raza Espanol horses. *J Anim Breed Genet.* 2018 Oct;135(6):460-71.
- Tetens J, Widmann P, Kuhn C, Thaller G. A genome-wide association study indicates LCORL/NCAPG as a candidate locus for withers height in German Warmblood horses. *Anim Genet.* 2013 Aug;44(4):467-71.
- Thompson MA, McCann BE, Rhen T, Simmons R. Population genomics provide insight into ancestral relationships and diversity of the feral horses of Theodore Roosevelt National Park. *Ecol Evol.* 2024 Apr 1;14(4):e11197.
- Tixier-Boichard M, Ayalew W, Jianlin H. Inventory, characterization and monitoring. *Anim Genet Resour Inf.* 2008;42:29-47.
- Toyama KS, Crochet PA, Leblois R. Sampling schemes and drift can bias admixture proportions inferred by structure. *Mol Ecol Resour.* 2020 Nov;20(6):1769-85.
- Upadhyay M, Eriksson S, Mikko S, Strandberg E, Stalhammar H, Groenen MAM, Crooijmans RPMA, Andersson G, Johansson AM. Genomic relatedness and diversity of Swedish native cattle breeds. *Genet Sel Evol.* 2019 Oct 2;51(1):56.
- Velie BD, Shrestha M, Francois L, Schurink A, Tesfayonas YG, Stinckens A, Blott S, Ducro BJ, Mikko S, Thomas R, Swinburne JE, Sundqvist M, Eriksson S, Buys N, Lindgren G. Using an inbred horse breed in a high density genome-wide scan for genetic risk factors of insect bite hypersensitivity (IBH). *PLoS One.* 2016 Apr;11(4):e0152966.
- Vosgerau S, Krattenmacher N, Falker-Gieske C, Seidel A, Tetens J, Stock KE, Nolte W, Wobbe M, Blaj I, Reents R, Kuhn C, von Depka Prondzinski M, Kalm E, Thaller G. Genetic and genomic characterization followed by single-step genomic evaluation of withers height in German Warmblood horses. *J Appl Genet.* 2022 May;63(2):369-78.
- Vostra-Vydrova H, Moravcikova N, Kasarda R, Shihabi M, Druet T, Halvonik A, Hofmanova B, Dove P, Zorc M, Molina A, Demyda-Peyras S, Valera M, Kovacs M, Posta J, Rogic B, Cubric-Curik V, Vostry L, Curik I. Genomic diversity and historical legacy of Baroque horses: A living cultural heritage. *Animal.* 2026 Aug;20(8):101897.
- Vostra-Vydrova H, Vostry L, Hofmanova B, Moravcikova N, Vesela Z, Vrtkova I, Novotna A, Kasarda R. Genetic diversity and admixture in three native draught horse breeds assessed using microsatellite markers. *Czech J Anim Sci.* 2018;63(3):85-93.
- Wickham H. *ggplot2: Elegant graphics for data analysis.* Cham: Springer; 2016. 260 p.
- Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011 Jan 7;88(1):76-82.
- Zhou Y, Wang Q, Wang Q, Yan Y, Li G, Wu G, Yang N, Wen C. Pedigree reconstruction based on genotype data in chickens. *Poult Sci.* 2024 Dec;103(12):104327.

Received: June 17, 2026

Accepted: October 12, 2026

Published online: October 27, 2026