

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

The authors are fully responsible for both the content and the formal aspects of the electronic supplementary material. No editorial adjustments were made.

Electronic Supplementary Material (ESM)

ESM S1. Material and Methods

Table S1. Overview of datasets used for genomic analyses and pedigree reconstruction

Table S2. Summary of global and pairwise genomic relationship matrix values within breeds after the selection of representative individuals

Table S3. Genome-wide and chromosome-specific inferred local ancestry proportions (%) in Norik of Muran

Table S4. QTL overlaps normalised by the genomic length of dominant ancestry segments in the Norik of Muran genome

Table S5. Breed-level haplotype sharing inferred from mean chunk counts and chunk lengths

Figure S1. Variation in the number of detected communities across nearest neighbour values (k) inferred using the NetView approach

Figure S2. Population structure inferred by ADMIXTURE across K values

Figure S3. Cross-validation error identifying $K = 7$ as the optimal clustering solution

Figure S4. Individual-level haplotype-based clustering

Figure S5. Sizes of maternal (A) and paternal (B) half-sibling groups in the Norik of Muran population

ESM S1. MATERIAL AND METHODS

Detailed scaling and scoring algorithm for selection of minimally related individuals. The within-breed selection was performed iteratively. The first individual was selected according to the lowest scaled global relatedness, because no selected subset was available at this step. From the second step onwards, global and pairwise relatedness values were scaled to a 0 to 1 range among the remaining candidates and summed into one score:

$$\text{Score}_i = \frac{G_i - G_{\min}}{G_{\max} - G_{\min}} + \frac{P_i - P_{\min}}{P_{\max} - P_{\min}} \quad (1)$$

where:

- G_i – the mean genomic relatedness of an individual i to all other individuals within the breed;
- P_i – the mean genomic relatedness of an individual i to the already selected individuals;
- $G_{\min}, G_{\max}, P_{\min}, P_{\max}$ – the minimum and maximum values among the remaining candidate individuals at each selection step. The candidate with the lowest score was selected, and the procedure was repeated until 20 individuals per breed were retained.

NetView parameter exploration. The analysis was performed across a range of nearest-neighbour values ($k = 1$ to 60) to evaluate the effect of k on network structure and the number of detected communities (ESM Figure S1). Community detection was assessed using three algorithms implemented in NetView, Fast Greedy, Infomap and Walktrap. Because all three algorithms showed a comparable decline in the number of detected communities with increasing k , representative networks were visualised for selected values of k (5, 10 and 15), chosen to capture different levels of network connectivity across the evaluated range.

ChromoPainter and fineSTRUCTURE analysis. ChromoPainter model parameters were estimated before the final linked painting analysis using 10 expectation-maximisation iterations, with all individuals included and 10% of markers used for parameter estimation. The estimated parameters were subsequently retained for the final linked painting analysis.

For breed-level summarisation of the ChromoPainter chunk count matrix, values were averaged across recipient and donor breed combinations. Off-diagonal values represented haplotype sharing between individuals from different breeds, whereas diagonal values represented haplotype sharing among individuals from the same breed. For within-breed comparisons, self-copying values, corresponding to cases in which the recipient and donor were the same individual, were excluded when calculating the mean number of copied haplotype chunks. Breed-level and individual-level coancestry matrices were visualised using the observed range of copied haplotype chunks. The individual-level matrix was ordered according to the fineSTRUCTURE clustering output, with breed assignment shown as a colour annotation.

The ChromoPainter coancestry matrix was analysed in fineSTRUCTURE to infer haplotype-based clustering among individuals. Two independent Markov chain Monte Carlo (MCMC) runs were performed to assess the stability of the inferred clustering. The runs produced consistent major clustering patterns and were retained for tree building. The resulting tree was visualised separately to provide a detailed representation of haplotype-based clustering. Support values were extracted for internal branching points, corresponding to splits between groups of individuals or clusters, with values closer to one indicating stronger support for the respective split.

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

Table S1. Overview of datasets used for genomic analyses and pedigree reconstruction

Dataset	Filtering/processing	<i>N</i>	Breeds	SNPs	Mean SNP distance (kb)	Median SNP distance (kb)
Full comparative dataset after standard QC	--mind 0.10; --geno 0.10; --maf 0.01	664	AN, BD, HF, HC, NF, MN, QH, SA, SN	37 875	58.94	33.54
Balanced subset	Selection of 20 least related individuals			37 875	58.94	33.54
Dataset without missing genotypes	--geno 0	180	AN, BD, HF, HC, NF, MN, QH, SA, SN	23 773	93.78	54.62
Dataset prepared for phasing	Alignment to recombination map			23 737	93.81	54.63
LD pruned dataset	--indep-pairwise 50 5 0.2			17 303	128.61	85.32
Norik of Muran dataset after standard QC	--mind 0.10; --geno 0.10; --maf 0.01	125	MN	55 426	40.93	31.68
Subset for pedigree reconstruction	--geno 0; --maf 0.05 --indep-pairwise 50 5 0.2			700	3 226.78	3 151.30

AN = Austrian Noriker; BD = Belgian Draught; HC = Hucul; HF = Haflinger; kb = kilobase; LD = linkage disequilibrium; MN = Norik of Muran; *N* = number of individuals in dataset; NF = Norwegian Fjord; QC = quality control; QH = Quarter Horse; SA = Shagya Arabian; SN = Silesian Noriker; SNP = single nucleotide polymorphism

Table S2. Summary of global and pairwise genomic relationship matrix values within breeds after the selection of representative individuals

Breed	<i>N</i> total	<i>N</i> selected	Global relatedness		Pairwise relatedness	
			$\bar{x} \pm \text{SD}$	max	$\bar{x} \pm \text{SD}$	max
AN	184	20	0.048 ± 0.032	0.065	0.029 ± 0.042	0.180
BD	30	20	0.140 ± 0.010	0.153	0.134 ± 0.026	0.258
HF	91	20	0.146 ± 0.050	0.189	0.123 ± 0.071	0.469
HC	32	20	0.213 ± 0.013	0.230	0.202 ± 0.118	0.701
NF	20	20	0.252 ± 0.010	0.273	0.252 ± 0.039	0.394
MN	120	20	0.040 ± 0.009	0.051	0.027 ± 0.043	0.254
QH	40	20	0.249 ± 0.012	0.265	0.234 ± 0.032	0.408
SA	47	20	0.324 ± 0.014	0.349	0.310 ± 0.081	0.653
SN	100	20	0.061 ± 0.009	0.078	0.053 ± 0.042	0.257

AN = Austrian Noriker; BD = Belgian Draught; HC = Hucul; HF = Haflinger; max = maximum value of relatedness per breed; MN = Norik of Muran; NF = Norwegian Fjord; QH = Quarter Horse; SA = Shagya Arabian; SN = Silesian Noriker; $\bar{x}$ = mean, SD = standard deviation

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

Table S3. Genome-wide and chromosome-specific inferred local ancestry proportions (%) in Norik of Muran

Chromosome	AN	BD	HC	HF	NF	QH	SA	SN
Whole genome	39.61	9.23	3.47	2.31	2.04	1.76	0.97	40.61
1	43.16	9.47	4.61	3.22	0.99	0.82	1.69	36.03
2	43.84	6.39	1.70	2.32	2.42	0.56	1.16	41.61
3	32.27	10.83	2.56	0.52	1.53	3.46	0.66	48.17
4	31.00	9.40	0.49	1.21	0.28	2.52	1.81	53.29
5	40.98	8.68	2.23	2.34	1.55	0.73	0.25	43.24
6	40.76	10.88	5.07	1.80	3.17	1.36	0.31	36.65
7	34.20	5.09	1.24	3.03	5.14	2.97	0.13	48.20
8	47.82	9.06	4.59	0.88	0.32	1.96	0.98	34.39
9	47.42	3.39	9.54	4.01	0.77	0.76	0.20	33.92
10	33.90	11.21	3.69	1.64	1.80	3.71	0.29	43.74
11	35.95	3.82	2.33	3.76	1.97	0.09	0.42	51.66
12	49.28	7.28	0.81	1.45	12.54	1.50	0.57	26.58
13	27.09	10.01	4.35	4.24	0.49	1.36	1.36	51.10
14	36.11	11.53	0.58	3.97	0.20	4.00	0.28	43.33
15	40.57	7.56	1.88	1.66	1.47	0.66	0.79	45.40
16	41.44	10.93	2.08	0.62	2.52	0.64	1.65	40.14
17	45.87	8.73	4.77	4.16	0.92	1.12	0.23	34.20
18	35.95	5.11	6.33	1.35	2.70	1.38	4.66	42.52
19	38.21	5.14	4.21	1.58	3.14	5.20	0.96	41.57
20	28.67	27.62	6.24	3.33	2.88	0.71	0.82	29.73
21	46.00	11.27	1.46	2.26	2.14	0.72	0.16	35.98
22	45.46	2.12	1.20	3.52	0.22	2.00	0.34	45.14
23	49.07	4.61	4.41	1.90	5.25	1.43	1.50	31.83
24	47.65	10.15	2.72	1.81	4.83	1.45	1.42	29.96
25	41.20	7.16	10.13	0.99	0.75	2.95	0.00	36.82
26	38.42	13.41	6.66	3.80	1.86	3.82	1.74	30.30
27	35.09	8.35	2.81	1.54	6.02	0.95	0.33	44.90
28	37.38	8.26	5.97	0.24	0.88	4.65	1.36	41.26
29	40.05	20.49	2.53	2.95	0.62	0.35	2.25	30.78
30	18.21	25.78	0.90	2.95	2.72	0.35	0.29	48.81
31	40.47	13.20	1.99	5.06	0.45	2.57	1.65	34.60

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

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

Table S4. QTL overlaps normalised by the genomic length of dominant ancestry segments in the Norik of Muran genome

Ancestry source	Dominant segments (N)	Total segment length		SNPs in segments (N)	QTL overlaps (N)	QTL overlap density	
		Mb	cM			Mb	cM
AN	66	1 101.16	1 321.44	11 985	1 194	1.08	0.90
SN	68	1 089.27	1 296.04	11 734	1 172	1.08	0.90
BD	1	0.95	1.49	18	2	2.10	1.34

cM = centimorgan; Mb = Megabase; QTL = quantitative trait locus; SNP = single nucleotide polymorphism

Table S5. Breed-level haplotype sharing inferred from mean chunk counts and chunk lengths

R/D	$\bar{x}$	AN	BD	HF	HC	NF	MN	QH	SA	SN
AN	count	117.41	99.72	107.23	96.29	96.28	106.80	96.15	94.41	103.55
	length	36.86	28.70	32.15	27.38	27.30	31.73	27.63	26.86	30.34
BD	count	100.08	126.93	99.72	98.44	99.48	102.38	95.94	95.28	103.40
	length	28.79	41.38	28.65	28.05	28.31	29.61	27.39	27.02	29.95
HF	count	106.10	98.39	126.72	95.66	95.17	99.25	96.29	94.12	99.11
	length	31.90	28.37	42.08	27.34	27.09	28.78	27.96	26.97	28.73
HC	count	95.87	97.70	96.24	134.76	98.50	97.14	97.47	97.07	97.10
	length	27.31	27.91	27.47	47.29	28.11	27.70	28.13	27.86	27.68
NF	count	96.39	99.28	96.21	99.04	139.98	98.17	95.61	95.61	98.14
	length	27.34	28.28	27.31	28.21	48.29	27.91	27.22	27.04	27.90
MN	count	108.14	103.15	101.36	98.56	99.07	115.18	96.49	95.66	108.38
	length	32.05	29.76	29.22	27.96	28.07	35.28	27.47	27.05	32.00
QH	count	96.21	95.76	97.41	98.16	95.66	95.60	130.83	104.21	95.51
	length	27.67	27.35	28.19	28.29	27.25	27.31	44.88	31.08	27.32
SA	count	92.76	93.42	93.40	95.88	93.90	93.17	102.24	143.11	93.31
	length	26.52	26.63	26.79	27.60	26.69	26.57	30.62	51.60	26.65
SN	count	103.81	103.29	100.39	97.67	98.24	107.34	95.58	94.97	119.52
	length	30.39	29.93	29.01	27.78	27.91	31.76	27.33	26.98	37.89

AN = Austrian Noriker; BD = Belgian Draught; D = donor breed; HC = Hucul; HF = Haflinger; MN = Norik of Muran; NF = Norwegian Fjord; QH = Quarter Horse; R = recipient breed; SA = Shagya Arabian; SN = Silesian Noriker; $\bar{x}$ = mean shared chunk count/length (cM)

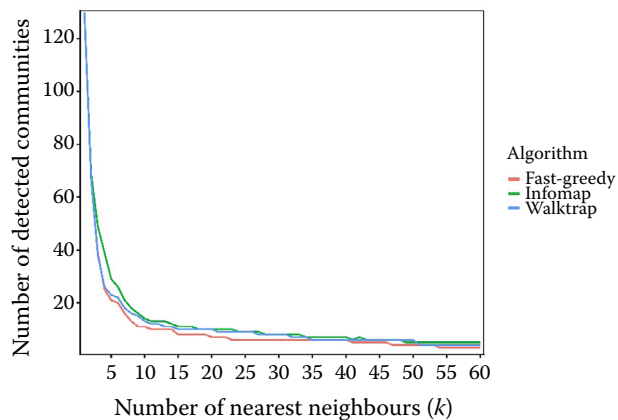


Figure S1. Variation in the number of detected communities across nearest neighbour values (k) inferred using the NetView approach

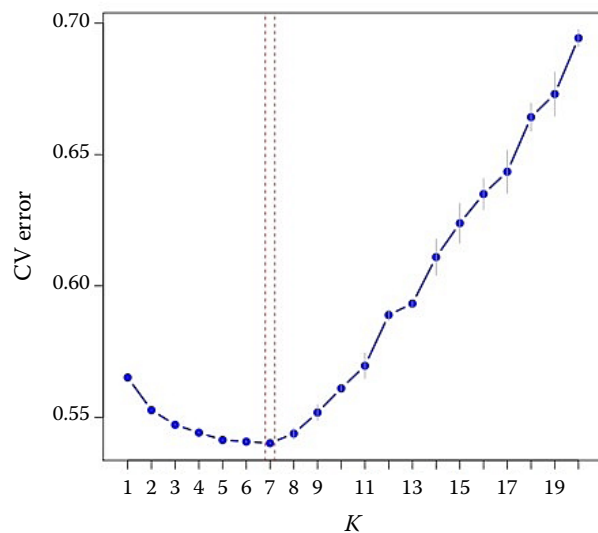


Figure S3. Cross-validation error identifying $K = 7$ as the optimal clustering solution

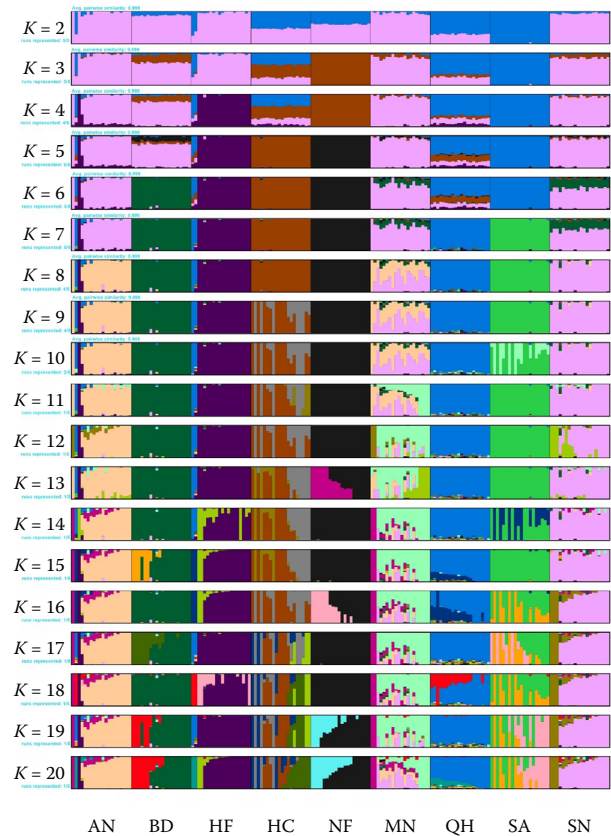


Figure S2. Population structure inferred by ADMIXTURE across K values

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

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

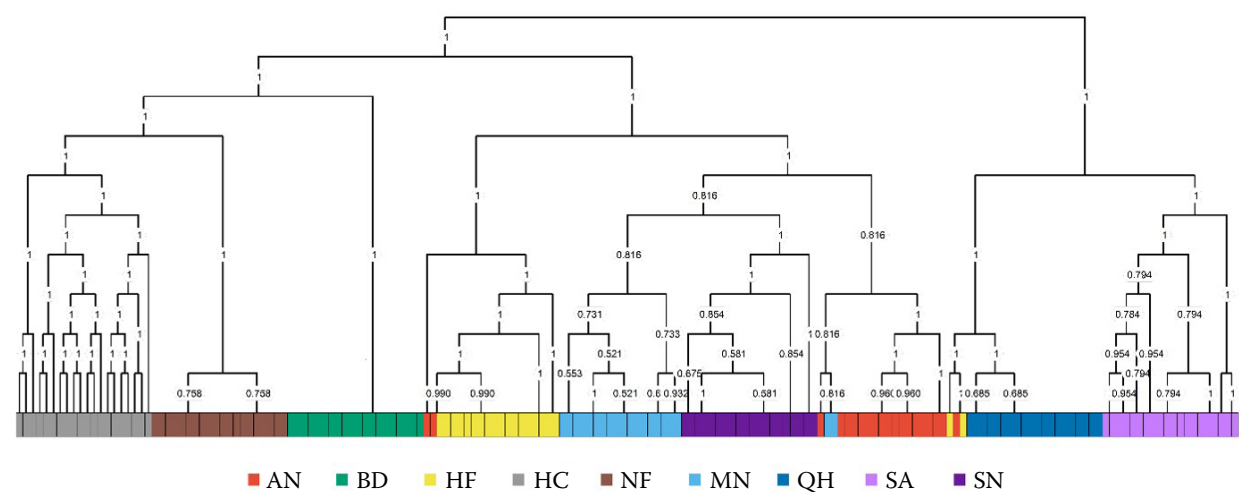


Figure S4. Individual-level haplotype-based clustering
AN = Austrian Noriker; BD = Belgian Draught; HC = Hucul; HF = Haflinger; MN = Norik of Muran; NF = Norwegian Fjord; QH = Quarter Horse; SA = Shagya Arabian; SN = Silesian Noriker

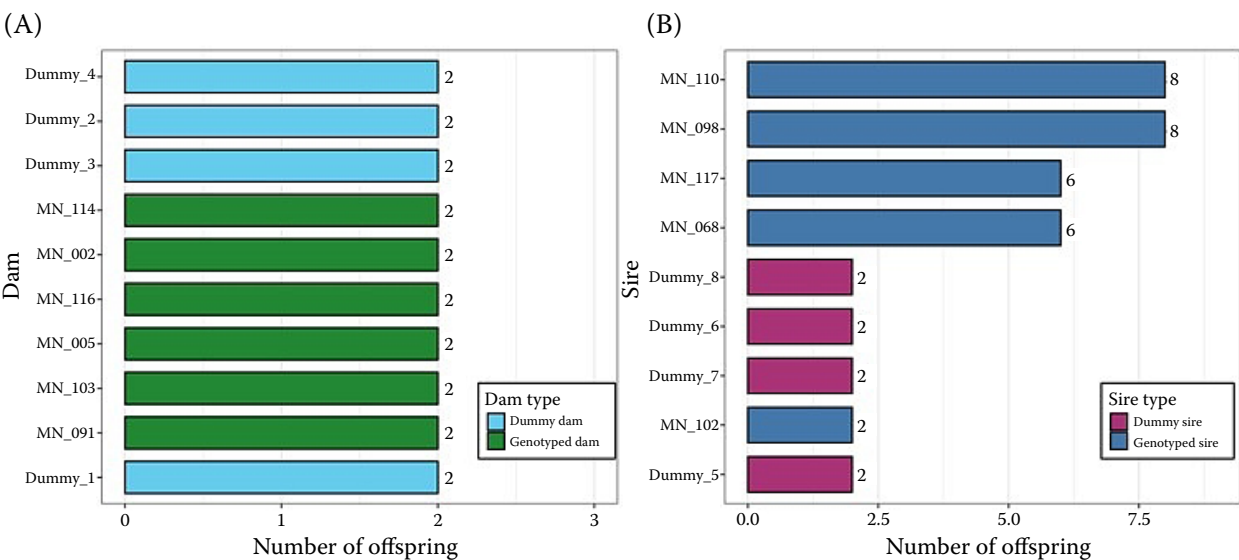


Figure S5. Sizes of maternal (A) and paternal (B) half-sibling groups in the Norik of Muran population