Journal Pre-proofs



Genomic Diversity and Historical Legacy of Baroque Horses: A Living Cultural Heritage

H. Vostra-Vydrova, N. Moravcikova, R. Kasarda, M. Shihabi, T. Druet, A. Halvonik, B. Hofmanova, P. Dovc, M. Zorc, A. Molina, S. Demyda-Peyrás, M. Valera, M. Kovács, J. Posta, B. Rogić, V. Cubric-Curik, L. Vostry, I. Curik

PII: S1751-7311(26)00144-8
DOI: https://doi.org/10.1016/j.animal.2026.101897
Reference: ANIMAL 101897

To appear in: *Animal*

Received Date: 6 December 2025
Revised Date: 27 June 2026
Accepted Date: 29 June 2026

Please cite this article as: H. Vostra-Vydrova, N. Moravcikova, R. Kasarda, M. Shihabi, T. Druet, A. Halvonik, B. Hofmanova, P. Dovc, M. Zorc, A. Molina, S. Demyda-Peyrás, M. Valera, M. Kovács, J. Posta, B. Rogić, V. Cubric-Curik, L. Vostry, I. Curik, Genomic Diversity and Historical Legacy of Baroque Horses: A Living Cultural Heritage, *Animal* (2026), doi: https://doi.org/10.1016/j.animal.2026.101897

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see: https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article. Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2026 Published by Elsevier B.V. on behalf of The animal Consortium.

Genomic Diversity and Historical Legacy of Baroque Horses: A Living Cultural Heritage

H. Vostra-Vydrova[a], N. Moravcikova[b], R. Kasarda[b], M. Shihabi[c], T. Druet[d], A. Halvonik[b], B. Hofmanova[a], P. Dovc[e], M. Zorc[e], A. Molina[f], S. Demyda-Peyrás[f], M. Valera[g], M. Kovács[h,i], J. Posta[i], B. Rogić[j], V. Cubric-Curik[c], L. Vostry[a], I. Curik[c,k],

[a]Czech University of Life Science Prague, Kamycka 129, 16500 Prague, Czech Republic

[b]Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 94976 Nitra, Slovak Republic

[c]University of Zagreb, Faculty of Agriculture, Svetošimunska cesta 25, 10000 Zagreb, Croatia

[d]Unit of Animal Genomics, GIGA-R & Faculty of Veterinary Medicine, University of Lège, Belgium,

[e]Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, 1000 Ljubljana, Slovenia

[f]Departamento de Genética, Facultad de Veterinaria, Universidad de Córdoba, Córdoba, Spain

[g]Departamento de Agronomía, ETSIA, Universidad de Sevilla, Sevilla, Spain

[h]Doctoral School of Animal Science, University of Debrecen, Böszörményi str 138, 4032, Debrecen, Hungary

[i]Faculty of Agricultural and Food Sciences and Environmental Management, University of Debrecen, Böszörményi str 138, 4032 Debrecen, Hungary

[j]Faculty of Agriculture, University of Banja Luka, Bulevar vojvode Petra Bojovica 1A, 78000, Banja Luka, Bosnia and Herzegovina

[k]Institute of Animal Sciences, Hungarian University of Agriculture and Life Sciences (MATE), Guba S. str. 40, H-7400, Kaposvár, Hungary

Corresponding author: Lubos Vostry – vostry@af.czu.cz

Abstract

Baroque horses represent a distinctive group of breeds that emerged mainly between the 16th and 18th centuries, shaped by the demands of European royalty and their courts. Although recognised as separate breeds, they share common ancestry, breeding goals, and historical gene flow. This study aimed to evaluate their genomic diversity, population structure, and admixture patterns. We analysed 1 160 individuals from eight breeds - Lipizzan, Old Kladruby Horse, Pura Raza Española, Lusitano,

Journal Pre-proofs

Friesian, Criollo, Peruvian Paso, and Puerto Rican Paso Fino - using 34 026 Single-Nucleotide Polymorphisms (**SNP**s), with the Croatian Arab horse included as an outgroup. The Lipizzan was further divided into national subpopulations, the two Old Kladruby subpopulations defined by colour (black and grey), and within the Pura Raza Española, the Carthusian strain was treated as a distinct subpopulation. Observed genetic diversity, measured by heterozygosity and haplotype richness, was broadly comparable across breeds, although clear differences were observed between breeds with the lowest values (Friesian) and those with the highest values (Lusitano and Criollo). Genomic inbreeding coefficients, estimated using runs of homozygosity ($\mathbf{F_{ROH>2Mb}}$) and homozygosity-by-descent segments ($\mathbf{F_{HBD}}$), ranged from ~0.06 to ~0.29 and ~0.14 to ~0.27, respectively. These values indicate relatively high levels of inbreeding, which may reflect the closed breeding structure of the populations and their relatively limited effective population sizes. In contrast, most individual inbreeding coefficients ($\mathbf{F_{IS}}$) values were negative, reflecting deliberate avoidance of close-relative matings. Estimates of contemporary effective population size ($\mathbf{N_{eLD}}$) ranged from ~50 to ~240 across breeds, depending on the method. Historical N_{eLD} trajectories revealed two distinct demographic decline patterns: one occurring around 12 generations ago and the other around 22 generations ago. Evidence of recent between-breed gene flow was detected among Iberian breeds and between Croatian Arab and Lusitano, whereas shared historical ancestry among Baroque breeds was evident across populations. These results provide new insights into the genomic architecture of Baroque horses and offer valuable guidance for managing their genetic diversity.

Keywords: genetic diversity, admixture, conservation, genetic diversity loss, genomics

Implications

Understanding genomic diversity is essential for conservation and sustainable management of Baroque horses. Baroque horses are a culturally significant group of breeds whose long history is mirrored in their genomic architecture. Our analysis revealed clear genetic differentiation among breeds, variable levels of genomic inbreeding, and signatures of historical admixture. These findings highlight both the uniqueness and vulnerability of Baroque horses. Understanding their diversity and structure provides guidance for sustainable breeding strategies and supports targeted conservation efforts. Combined with ex situ conservation approaches, such knowledge may facilitate the regeneration of these rare breeds while safeguarding their genetic legacy for future generations.

Introduction

A primary objective in the breeding strategy for small and closed populations is to maintain the highest possible level of genetic diversity. Classical horse breeds are often managed as closed populations, and previous studies have demonstrated that breeding such populations without effective control of inbreeding can result in a loss of genetic variability and inbreeding depression. The loss of genetic variability and the accumulation of inbreeding may negatively impact reproduction traits and fitness (Laseca et al., 2022; Müller-Unterberg et al., 2017; Vostrý et al., 2021), and productive

(Todd et al., 2018; Vostra-Vydrova et al., 2020) and morphological traits (Gómez et al., 2009; Vostrý et al., 2018). Baroque horses breeds represent an important group of warmblood breeds that originated mainly in the 16th–18th centuries. The Pura Raza Española, developed on the Iberian Peninsula from Iberian horses with influences from the Arab and Berber breeds, has contributed to the formation of many other equine breeds both in the American continent and in Europe, such as Lusitano, Friesian and Old Kaldruby horse (Ripoles-Lobo et al., 2023). Within this breed, a distinctive lineage known as the Carthusian strain has been maintained since the 19th century through deliberate selection and inbreeding, which is distinguished by certain morphological differences (Poyato-Bonilla et al., 2022). The closely related Lusitano of Portugal shares the same genetic background but has retained stronger working traits, and its breeding is inextricably tied up with cattle breeding and bullfighting (Luís et al. 2007). In parallel, two other major Baroque breeds - the Lipizzan and the Old Kladruby horse - were developed within the Habsburg Monarchy. The Lipizzan and Old Kladruby trace their origins to Old Spanish horses (primarily the Pura Raza Española) and Old Italian horses (notably the Neapolitan) (Zechner et al., 2002; Kasarda et al., 2016). The close genetic ties between the Lipizzan and the Old Kladruby are evident in their shared foundations. These historical links are consistent with molecular evidence indicating close affinity and substructure between Old Kladruby and Lipizzan populations (Kasarda et al., 2016). Breeding goals for Lipizzan horses are defined by state studs in Austria, Croatia, Hungary, Italy, Romania, Slovakia, and Slovenia, countries that once formed part of the Austro-Hungarian Empire (Zechner et al., 2002).

The influence of Old Spanish horses also spread to Northern Europe, where the originally heavier Friesian horse gained its Baroque elegance in the 16th and 17th centuries through crossbreeding with Pura Raza Española stallions. After World War I, most horse breeds experienced a significant decline in numbers of individuals, and the Old Kladruby horse nearly became extinct. During the regeneration process of the Old Kladruby horse (1934–1974), Lipizzan, Lusitano, and Friesian stallions were used (Bílek, 1957). In small, closed populations, such as Baroque horse breeds, inbreeding can lead to a substantial loss of genetic variability. This reduction may have serious consequences for long-term population viability and, in extreme cases, can even result in extinction (Amos and Balmford 2001). This risk is evident in the Friesian horse. Genome-wide Single-Nucleotide Polymorphisms (**SNP**) analyses show that Friesians have high levels of autozygosity and long runs of homozygosity (Schurink et al., 2019). However, other analysed breeds also exhibited higher levels of autozygosity, including the Lipizzan (Grilz-Seger et al., 2019), Pura Raza Española, and Cartusian horses (Poyato-Bonilla et al., 2021).

A fundamental prerequisite for the preservation of these unique breeds is not only the maintenance of their genetic diversity but also the analysis of their mutual admixture, which can provide essential information for a possible future regeneration process. In addition, the increasingly available use of genomic SNP chips or whole-genome sequencing allows the study of direct genetic diversity, which is not affected by incomplete or erroneous pedigree records. As a matter of fact in these breeds, comparisons between pedigree-based and genomic inbreeding coefficients reveal meaningful discrepancies (Luštrek et al., 2025). Several microsatellite-based studies have described genetic relationships among Baroque breeds (Achmann et al., 2004; Kasarda et al., 2016; Zechner et al., 2004; Luís et al. 2007), whereas genome-wide analyses based on high-density SNP data remain scarce. Such datasets allow direct estimation of key genomic parameters, including runs of homozygosity, identity-by-

descent segments, linkage disequilibrium–based effective population size, and fine-scale population structure. Despite these advantages, a unified genome-wide comparative analysis encompassing multiple Baroque breeds has not yet been undertaken.

We hypothesized that, despite their shared historical origins, Baroque horse breeds have undergone distinct genomic trajectories due to breeding isolation and different selection pressures, leading to variable levels of genetic diversity and admixture.

The main objective of this study was to provide a comprehensive genomic assessment of Baroque horse breeds using SNP data by (i) evaluating genetic variability within and between subpopulations, (ii) estimating genomic inbreeding as well as contemporary and historical effective population size, and (iii) analysing patterns of genetic admixture and gene flow among populations. By integrating these approaches, this study aims to generate information relevant for conservation management and for potential future regeneration strategies of endangered Baroque horse breeds.

Material and methods

This study used genome-wide SNP data from five Baroque horse breeds in Europe: Lipizzan (**Lip**), Old Kladruby horse (**OKH**), Friesian (**FriCZ**), Pura Raza Española (**PRE**), and Lusitano (**Lus**). To better understand gene flow and admixture among them, the selected breeds were further subdivided into subpopulations. The individuals of the Old Kladruby breed are currently split according to coat colour: grey (**OKHg**) and black (**OKHb**), which are bred separately. The individuals of the Lipizzan breed are structured according to country of origin: Bosnia and Herzegovina (**LipBA**), Croatia (**LipHR**), Czechia (**LipCZ**), Hungary (**LipHU**), Slovakia (**LipSK**) and Slovenia (**LipSI**). Together with Austria, Italy, and Romania (for which samples were unavailable), these countries form the core regions of Lipizzan horse breeding. The Cartujano or Carthusian strain (**Cart**), regarded as the oldest and purest lineage within the Pura Raza Española, was analysed separately from the main population of the breed. To enable broader comparisons of genetic diversity and admixture, additional South American breeds derived from the historical Iberian horse were included: Peruvian Paso (**Peru**; Petterson et al., 2013) and Puerto Rican Paso Fino (**PRPF**; Petterson et al., 2013), along with the Criollo (**Cri**). In addition, the Arab horse, represented by data from the Croatian Arab (**ArabHR**, Raguz et al., 2023), was used as an outgroup. Given the known influence of Arab horses on the development of Baroque breeds, historical gene flow between them is to be expected. In this study, DNA was extracted from hair root samples collected from 1 049 horses representing multiple breeds. Genotyping was performed using several SNP chip platforms, primarily the GGP Equine 70k (ranging from 55k to 70k markers, versions 1–5) and the Illumina Equine 80k, by a commercial laboratory following the manufacturer's protocols. The number of genotyped animals met the minimum threshold of 20 individuals required for reliable estimation of allele frequencies for biallelic markers such as SNPs (Hein et al., 2004). Alongside our own dataset, publicly available data from 112 individuals from three breeds (Peru and PRPF from Petterson et al., 2013; and ArabHR from Raguz et al., 2023) were also included in the analysis. The number of animals genotyped is listed in Table 1. Prior to merging, all datasets were aligned to the EquCab3.0 reference genome assembly to ensure consistency in genomic

coordinates. To maintain strand consistency, ambiguous SNPs (A/T and C/G), which could not be unambiguously verified between platforms, were excluded to prevent potential strand assignment errors. SNP datasets originating from different platforms were then merged based on common genomic positions and SNP identifiers. The proportion of SNPs retained after merging ranged between 48.5 % and 68.3 % per platform. After merging, the dataset comprised 1 221 individuals. Only autosomal SNPs with known chromosomal positions were retained to avoid sex-related biases. Quality control was first applied by removing individuals with call rates below 0.9. Genomic relationships were then estimated using PLINK v1.9 (Purcell et al., 2007), and individuals with relationship coefficients greater than 0.25 were excluded. After the removal of low-quality and related individuals, 1 160 individuals remained for subsequent analyses. Three datasets were subsequently generated. Data1 included SNPs filtered for call rates below 0.9 and comprised 34 026 SNPs. Data2 was processed using the same quality control criteria as Data1, with the additional removal of non-informative SNPs with a minor allele frequency below 0.05, resulting in 28 875 SNPs. Data3 was derived from Data2 by further pruning for linkage disequilibrium using a threshold of $R^2 < 0.01$, following the procedure described by Petersen et al. (2013), and contained 8 276 SNPs.

Conservation status

Diversity and Haplotype Richness

To reduce the ascertainment bias of the SNPs chips (Gautier et al., 2010), short haplotypes were used instead of single SNPs, as recommended by Simčič et al. (2015), Pogorevc et al. (2024) and Vostry et al. (2024). Therefore, the SNP data were divided into non-overlapping blocks of four SNP genotypes (block) with an inter-marker distance of less than 50 kb for neighbouring SNPs, with the maximum length of each block being less than 150 kb. A total of 3 231 blocks were used for further analysis. These blocks act as multi-allelic markers and allow a better analysis of genetic diversity than bi-allelic SNPs. The haplotype blocks and SNPs were used to assess genetic diversity based on a comparison of observed ($\mathbf{H_O}$) and expected heterozygosity ($\mathbf{H_E}$) and individual inbreeding coefficients ($\mathbf{F_{IS}}$). The F_{IS} indicates deviation from Hardy-Weinberg equilibrium. Negative values are usually the result of outcrossing or avoidance of inbreeding, whereas positive values indicate mating of close relatives (Wright, 1965). All these parameters were calculated using the R package PopGenReport v3.1 (Adamack and Gruber, 2014). The estimated haplotype rarefied richness (**HRR**), used here instead of allelic richness for multiallelic loci, was calculated following Kalinowski (2004) using ADZE software (Szpiech et al., 2008) to account for differences in sample size among breeds. For HRR the number of samples per subpopulations was standardised to the smallest number of allele copies across all sampling sites (2n = 40, twenty diploid individuals) as uneven sample size impact the number of distinct alleles observed (Kalinowski, 2004).

Genomic Inbreeding

Genomic inbreeding was estimated using the „identity-by-descent“ concept, which is widely recognised as valid, was employed as the best method for estimating genomic inbreeding. Two distinct approaches were used to identity-by-descent regions: i) the rule-based and ii) the model-based approaches.

i) The rule-based approach identified Runs-of-Homozygosity (**ROH**s) using the detectRUNS R package (Biscarini et al., 2018). ROHs were detected individually for each individual using the empirical consecutive approach with following criteria: minimum ROH length was set to 2 Mb, maximum gap and maximum density parameters were disregarded. All detected ROHs were categorised by length into four classes: >2 Mb, 2–4 Mb, 4–8 Mb, and >8 Mb, which correspond approximately to common ancestors occurring 25–12, 12–6, and fewer than 6 generations ago, respectively (Curik et al. 2014). To account for genotyping errors, the number of allowed heterozygous and missing genotypes within ROHs was adjusted for each class, following the recommendations of Ferenčaković et al. (2013). According to Ferenčáković et al. (2013), the analysis of shorter ROH segments when using a middle density SNP panel (<50k SNP) leads to a systematic overestimation of ROH values. A dataset processed with the same quality control as in other analyses, but without minor allele frequency filtering, was used for this analysis. The rule-based genomic inbreeding coefficient ($\mathbf{F_{ROH}}$) was calculated as the proportion of the autosomal genome covered by ROHs, using the formula $F_{ROH} = L_{ROH}/ L_{AUTOSOME}$, where L_{ROH} is the length of all ROH in the genome of the individual and $L_{AUTOSOME}$ is the specific length of the autosomal genome, covering 31 autosomal chromosome pairs.

ii) The model-based approach (Druet and Gautier, 2017) detected homozygous-by-descent (**HBD**) segments using the RZooROH package v0.4.1 in R (Bertrand et al., 2019). Based on Data2, we ran a layer model (Druet and Gautier, 2022) with HBD classes set to 'Interval' and the krates parameters set to {3, 6, 12, 25, 50, 100, 250, 500, 1 000}. In this model, the number of generations to the common ancestors contributing to HBD class c is approximately between G_{C-1} and G_C (where G_C refers to the parameter provided with 'krates' when = 'Intervals' and with G_0 set to 1). See Forneris et al. (2025) for more details.

Longer ROH or lower G_c values correspond to longer HBD segments, indicating more recent common ancestry Curik et al. (2014). Therefore, we divided rule-based genomic inbreeding coefficient (F_{ROH}) and model-based genomic inbreeding ($\mathbf{F_{HBD}}$) into three groups: current ($F_{ROH>8Mb}$ or $F_{HBD=1-6}$), moderate ($F_{ROH\ 4-8Mb}$ or $F_{HBD=6-12}$), and distant ($F_{ROH\ 2-4Mb}$ or $F_{HBD=12-100}$). The last three HBD classes ($F_{HBD=100-250}$, $F_{HBD=250-500}$ and $F_{HBD=500-1\ 000}$) were not considered in the 'Results and Discussion' sections as they correspond to HBD associated with more distant ancestors that is less relevant for our current purpose. In addition, their identifications are less reliable due to the lower number of SNPs per segment. Based on model-based approaches the overall F_{HBD} was calculated as the proportion of the genome assigned to HBD classes with $G_c \leq 100$. Based on a rule-based approach, the total genomic inbreeding coefficient (F_{ROH}) was considered using ROH segments > 2Mb. Two approaches were used because most previous studies in horses have estimated genomic inbreeding using rule-based ROH methods, thereby allowing direct comparison with the existing literature and ensuring broader comparability for researchers using either ROH-based or HBD-based approaches, the latter of which may become the preferred method in future studies. In addition, the more recent model-based HBD approach offers several methodological advantages, including lower sensitivity to marker density, greater robustness in the presence of overlapping long autozygous segments, improved precision for detecting more remote autozygosity represented by shorter ROH segments (Druet and Gautier, 2022), and, according to our internal simulations, lower sensitivity to recombination events.

Current and Historical Effective Population Sizes

Two approaches were used to estimate the contemporary effective population size based on linkage disequilibrium (N_{eLD}). In this study, we have labelled all our estimates as N_{eLD} to indicate that the estimates are based on gamete/linkage disequilibrium. The N_{eLD} is the size of an idealised population, often considered a Wright-Fisher population, that exhibits the same degree of genetic drift and change in gamete/linkage disequilibrium as the population under consideration. Further information on N_{eLD} and potential biases can be found in Waples et al. (2014), Ryman et al. (2019), and Waples (2021). The first approach applied to Data2, implemented in NeEstimator v2 (Do et al., 2014), is based on gametic linkage disequilibrium. Waples and Do (2008) describe the method in detail and the 95% confidence interval for this approach was estimated using the jackknife procedure. The second approach applied to Data1 relies on the theoretical relationship between linkage disequilibrium and effective population size, as described by Hill (1981), and is implemented using a more complex modelling framework in the software GONE2 (Santiago et al., 2025).

In addition to contemporary N_{eLD}, historical N_{eLD} size applied in Data2 was estimated for each population and subpopulation. Historical N_{eLD} was estimated using the GONE2 software, which applies a genetic algorithm to derive the historical series of N_{eLD} that best minimise the sum of squared differences between observed D^2 values (average squared correlations between allele frequencies at two loci, weighted by their variance) across bins and those predicted under alternative demographic trajectories (Saura et al., 2021). Historical N_{eLD} estimates extended back 40 generations, corresponding to approximately up to 400 years, assuming an average generation interval of 10 years in Baroque horse breeds (Vostra-Vydrova et al., 2016; Pjontek et al., 2014; Steensma et al., 2024). This timeframe covers the period associated with the origin and development of Baroque horse breeds in Europe.

Relatedness, Population Structure, and Admixture

Relatedness

The phylogenetic relatedness among the 16 analysed Baroque horse subpopulations together with Arab horse was assessed using a Neighbor-Net network inferred from Reynolds's pairwise genetic distances (Reynolds et al., 1983), which were calculated with the StAMPP R package (Pembleton et al., 2013) and visualised using SplitsTree5 (Huson and Bryant, 2006). In addition, relatedness between analysed baroque horse populations was explored using variational autoencoders (**VAE**), an unsupervised machine learning method based on neural networks, where feed-forward architectures were applied for both encoder and decoder stages to regenerate the data and visualise the genetic structure in a latent space. VAEs consist of two interconnected deep neural networks designed to effectively reduce the high-dimensional complexity of genomic data: the encoder compresses genomic input into a probability distribution within the latent space, while the decoder reconstructs the input from these latent coordinates (Kingma and Welling, 2013). Compared with other dimensionality-reduction methods developed for large genomic datasets (Battey et al., 2021), the algorithm implemented in POPVAE (https://github.com/kr-colab/popvae) produces embeddings of lower dimensionality than principal component analysis while preserving the geometry of genetic relationships. POPVAE was run using parameters (latent dimension = 2, batch size = 32, max_epochs = 100, training/validation split =

0.9), which were selected based on the mean validation loss across different parameter settings and four independent replicates with different random seeds. Pairwise Procrustes similarity between replicates was moderate (≈0.70), indicating sufficient stability of the latent embeddings. Only the representative replicate with the lowest validation loss is described in the results. In addition, population structure was tested using NetView (Neuditschko et al., 2012) to construct k-nearest neighbour networks. To better focus on the genomic relatedness of Baroque horse subpopulations, further VAE and Neighbor-Net analyses included other local Central European breeds and publicly available horse genotype data, as described by Petersen et al. (2013) and Cosgrove et al. (2020).

Population Structure

Population structuring into genetic groups was examined using NeuralADMIXTURE (Mantes et al., 2023) assuming values of K ranging from 1 to 18. NeuralADMIXTURE follows the same modelling assumptions as ADMIXTURE (Alexander et al., 2009) but reformulates the task as a neural-network-based autoencoder. It is an interpretable autoencoder with two main components: an encoder, composed of two linear layers with a Gaussian error linear unit activation followed by a soft-max activation, which projects a genotype sequence onto a vector of fractional ancestry assignments for each individual (Q membership matrix); and a decoder, implemented as a single linear layer with weights restricted between 0 and 1, yielding an interpretable projection matrix that learns the cluster centroids or, equivalently, the average allele frequency at each site for each ancestral population. NeuralADMIXTURE results were visualised using the web-based StructureSelector tool (Li and Liu, 2018). The optimal value of K for explaining genetic diversity in the analysed dataset was determined by estimators MedMedK (Puechmaille, 2016) implemented in web-based StructureSelector tool (Li and Liu, 2018). To verify the number of genetic groups we employed hierarchical iterative rounds of NeuralAMIXTURE analyses following Coulon et al. (2008). Prior to analyses of genetic differentiation within and among populations, analysis was applied to Data3, following the procedure described by Petersen et al. (2013), retaining 8 276 SNPs. Finally, the results were visualised using the web-based StructureSelector tool (Li and Liu, 2018). To account for the potential effect of unequal sample sizes on population structure estimates, the NeuralAMIXTURE analyses was repeated using a random selection of 50 individuals from each large population (LipBI, LipSI, OKHb, OKHg, and ArabHR), while smaller subpopulations with fewer than 50 individuals were not subsampled.

Admixture Quantified by Effective Number of Parental Clusters

The NeuralAdmixture software can be used to estimate the admixture coefficient (Q membership matrix) for each individual and provides information on the membership of the observed clusters within a particular breed or population. To distil the complexity of admixture into a single, informative metric for each breed or individual, Vostry et al. (2024) introduced a novel measure, termed effective number of parental clusters (**ENPC**). The ENPC serves as a quantitative indicator that captures the extent to which parental clusters contribute uniformly and effectively to the formation of a given breed or individual. It is calculated by deriving the inverse of the squared sum of all parental cluster proportions that contribute to the genetic composition of the breed or individual in question. For example, if only two parental clusters contribute 0.5 each to a breed, the resulting ENPC would be 2.0. In a more complex scenario where

parental clusters A, B, C, D and E contribute to a breed with proportions of 0.5, 0.3, 0.1, 0.1 and 0.0 respectively, the resulting ENPC would be 2.8. This implies that 2.8 parental clusters, each contributing equally, characterise the genetic composition of that particular breed. Notably, when ENPC is equal to 1.0, this means that the breed is 100% purebred, as only one parental cluster contributes exclusively to its genome.

Gene Flow Between Breeds

Contemporary gene flow among Baroque horse breeds was assessed by estimating migration rates (**m**) using the assignment test implemented in BayesAss v3 (Wilson and Rannala, 2003) applied to Data3. Ten independent replicates with different random seeds were run, each consisting of up to 16 million Markov chain Monte Carlo iterations, with the first 10 million iterations discarded as burn-in and sampling every 100 iterations, to improve the reliability of parameter estimates. The mixing parameters – delta allele frequency, delta migration rate, and delta inbreeding coefficient - were each set to 0.1. Acceptance rates for the three proposal types ranged between 25% and 45%. Effective sample sizes calculated from the posterior chains exceeded 2 700 for all parameters, indicating adequate mixing and convergence. Our choices of burn-in length and sampling scheme follow those used in previous BayesAss applications (Faubet et al. 2007), and the criteria for assessing chain mixing (Effective sample sizes > 2 700) are based on standard Bayesian Markov chain Monte Carlo diagnostics (Gelman et al. 2013).

The resulting log files were evaluated using Tracer (Rambaut et al., 2018), and Bayesian deviance was computed following the R script of Meirmans (2014). Migration estimates from the three runs with the lowest Bayesian deviance were merged to generate a posterior distribution of migration rates, from which the median values and 95% credibility intervals were calculated. Stability between independent BayesAss replicates is presented in Supplementary Fig. S1. A full overview of the results is provided in Supplementary Table S1.

Results

Conservation status

Diversity and Haplotype Richness

Our analysis encompassed all Baroque horse breeds and, notably, provides, to our knowledge, the first high-density SNP-based population study of the Old Kladruby horse. Table 1 presents the parameters of genetic diversity (H_O, H_E, F_{IS}, HRR) based on haplotype blocks. Haplotype block heterozygosity (H_O and H_E) among the analysed Baroque horse breeds ranged from 0.455 (FriCZ) to 0.617 (Cri) for H_O, and from 0.469 (FriCZ) to 0.649 (Lus) for H_E. Between-breed differences in heterozygosity were evident, with FriCZ at the lower end and Lus/Cri at the upper end. The ArabHR horse displayed H_O values comparable to Baroque horses, but generally higher H_E values, apart from three breeds (Cri, Peru, Lus). Both heterozygosity estimates (H_O and H_E) derived from haplotype blocks were higher - up to twice as high - than estimates based on SNP data (Supplementary Table S2). This reflects the fact that haplotype blocks are more similar to STR markers, which typically show higher H_O and H_E compared with SNP markers. Accordingly, we interpret primarily the between-breed contrasts, using SNP-based estimates (Supplementary Table S2) as a robustness check.

Although the differences in heterozygosity values estimated by different approaches were considerable, heterozygosity estimates based on haplotype blocks and SNPs followed the same trends between the two approaches and within breed, which was confirmed by a very high linear correlation (0.97). A similar trend was observed in a study of genetic diversity in dairy goat breeds (Vostry et al., 2024), where considerable differences between these approaches were also reported, with a comparable correlation. These differences are consistent with the biallelic nature of SNP markers.

Haplotype Rarefied Richness (HRR, standardised to 2n = 40, where n is the number of individuals in the population with the smallest sample size) ranged from 2.82 (FriCZ) to 4.20 (Lus). The ArabHR showed HRR values falling in the middle range of all Baroque breeds. Central European breeds (Lip and OKH) displayed HRR values similar to those of Iberian breeds (PRE, Lus, and Cart) as well as South American breeds (Cri, Peru, and PRPF).

The Wright's inbreeding coefficient (F_{IS}), a key population parameter, divided the analysed breeds into two groups. Central European breeds (Lip and OKH), except for LipHR, showed negative F_{IS} values, suggesting outbreeding. In contrast, the second group, consisting of Iberian and South American breeds, apart from PRPF, reported positive F_{IS} values, suggesting more frequent mating among closely related individuals. Given the small magnitudes, we report SE and departures from zero in Table 1.

Genomic Inbreeding

Genomic inbreeding based on identity-by-descent concept was estimated using two different approaches: the rule-based and the model-based approaches. Genomic inbreeding estimated from the rule-based approach (ROH > 2 Mb) showed lower values compared to model-based approach (Table 1 and Fig. 1). This difference reflects the conceptual distinctions between the two methods. On average, genomic inbreeding was moderate (median $F_{ROH>2Mb} \approx 0.123$ of $F_{HBD} \approx 0.167$), with marked between-breed differences (Table 1). A more detailed overview of genomic inbreeding coefficients, including their distributions and partitioning by expected autozygosity origin (from recent to distant common ancestors), is shown in Fig. 1. Thus, Fig. 1A shows that Central-European breeds (Lipizzan and Old Kladruby) showed intermediate F_{HBD}/F_{ROH} levels and narrower dispersion (smaller Inter quartile ranges) than Iberian and South-American breeds, reflecting their historical development. The highest genomic inbreeding was observed in FriCZ, followed by Cart, a lineage representing a more closely related subset of PRE. The lowest values were detected in the South American breeds Criollo (Cri) and Peruvian Paso (Peru). Fig. 1B and 1C further show that a substantial fraction of autozygosity derives from recent ($F_{ROH>8Mb}$ or $F_{HBD\leq6}$) and intermediate ($F_{ROH\ 4\text{-}8Mb}$ or $F_{HBD=12–25}$) HBD classes. The largest contributions from recent/intermediate classes were seen in FriCZ, Cart and both OKH colour strains, whereas Cri and Peru showed the smallest recent contributions.

Current and Historical Effective Population size

Table 2 presents estimates of contemporary N_{eLD} with confidence interval for 16 analysed horse breeds with two different approaches implemented in NeEstimator v2 and GONE2 software. Estimates of N_{eLD} obtained with GONE2 were higher in most estimates than those from NeEstimator v2, with significant differences in 5 of the 16

breeds where the 95% confidence intervals did not overlap. We treat non-overlapping 95% CIs as strong evidence of difference among approaches but overlapping CIs do not preclude meaningful discrepancies; methods differ in model assumptions and information used. In all 16 populations analysed, the upper 95% CI of Ne_{LD} estimates exceeded the FAO (1998) critical threshold of 50 effective animals, while the expected values themselves were above this threshold in 13 populations. For management targets, benchmarks of Ne ≈ 100 (medium-term) and Ne ≥ 500 (long-term adaptive potential) are widely cited. Estimates of N_{eLD} tended to be lower in Central European Baroque subpopulations (Lipizzan and Old Kladrub horse) than in other Baroque breeds and the Arab horse (Table 2).

Our analyses of historical effective population size (Fig. 2) revealed a marked decline between approximately 13 and 8 generations ago (≈1890–1945 assuming a 10-year generation interval) in half of the breeds examined, including the Iberian and South American breeds as well as the Old Kladruby. In contrast, the remaining populations - mainly Lipizzan subpopulations, together with the Friesian, Lusitano and Cartujano - showed evidence of a pronounced decline between 24 and 19 generations ago (≈1780–1830). Notably, the historical trends in effective population size across the analysed populations closely mirrored the patterns of genomic inbreeding inferred from F_{HBD} values (Fig. 1C). For example, in the Lipizzan horse subpopulations, Cart and PRPF, a significant increase in ancestral inbreeding was observed between 24 and 12 generations ago. In contrast, in OKHb, OKHg and Lus, the highest increase in ancestral inbreeding occurred between 12 and 4 generations ago. These increases in ancestral inbreeding correspond to significant decreases in the historical N_{eLD} of these breeds. However, this relationship requires more systematic investigation.

Relationship Among Parameters Defining Conservation Status

Relatedness

The relationships among the parameters defining conservation status - genomic inbreeding coefficients (F_{ROH} and F_{HBD}), Wright's inbreeding coefficient (F_{IS}), and effective population size (Ne_{LD}) - are summarised in Fig. 3. This visualisation highlights the dynamics of each population with respect to effective population size, realised inbreeding, and non-random mating within populations, and thereby provides guidance on the effectiveness of diversity management. Intermediate genomic inbreeding coefficients (F_{ROH}/F_{HBD}) were observed in the Central European Baroque breeds (Lip and OKH), largely reflecting their low effective population sizes. Nevertheless, their negative F_{IS} values indicate deliberate avoidance of close-relative matings, consistent with careful diversity management (Table 1, Fig. 3). In contrast, some populations with moderate effective population sizes, such as PRPF, but with F_{IS} values close to zero, showed comparable or only slightly lower F_{ROH} values. The highest ROH/HBD inbreeding levels were recorded in the Friesian (FriCZ) and Cartujano (Cart) populations, both of which displayed positive F_{IS} values. This demonstrates how insufficient avoidance of close-relative matings can lead to high realised inbreeding despite larger effective population sizes. The predominance of recent and intermediate genomic inbreeding in FriCZ and Cart further underscores suboptimal diversity management (Fig. 1C). Across breeds, F_{IS} tended to correlate positively with F_{HBD} ($r = 0.15$), and N_{eLD} (GONE2) negatively with the recent F_{HBD} component ($F_{HBD\leq6}$) ($r = -0.49$).

Genetic Relatedness and Population Structure

Various methods were used to determine relatedness and population structure. Relatedness results obtained using the VAE, Neighbor-Net, and NetView methods are shown in Fig. 4 and 5, as well as Supplementary Fig. S2. To achieve a more comprehensive assessment of relatedness, additional publicly available horse breed data were incorporated into the analysis (Peterson et al., 2013; Cosgrove et al., 2020) (Supplementary Fig. S3 and S4). The VAE and NetView methods reveal relatedness at both the individual and breed levels, whereas the Neighbor-Net method depicts only relationships between breeds. The VAE analysis clearly divided the analyzed populations according to breed. The first axis separated Lipizzan horses from Old Kladruby horses, while the second axis captured within-breed structure in both breeds, separating the colour variants within Old Kladruby horses and the subpopulations within Lipizzan horses. Within Lipizzan horses, the second axis distinguished the LipSI and LipBA subpopulations from the LipCZ, LipSK, LipHU, and LipHR subpopulations. Iberian breeds occupied an intermediate position in the plot and, together with ArabHR, were located between the Lipizzan and Old Kladruby subpopulations. The FriCZ breed was clearly separated from all other breeds. With the addition of other publicly available horse breeds (Petersen et al., 2013 and Cosgrove et al., 2020), the VAE analysis clearly separated breeds according to their origin. The first axis distinguished warm-blooded and Baroque riding breeds from cold-blooded and carriage breeds. The second axis separated Baroque horses together with Arab horses from other riding and working horses. Notably, Iberian breeds (PRE, Cart, and Lus) occupied an intermediate position in the latent space of the plot. The analysed subpopulations of Lip and OKH formed distinct, compact metapopulations, with both colour variants of OKH showing greater genetic divergence than the individual subpopulations of Lip. The NetView analysis (Supplementary Fig. S2) divided the monitored populations of baroque horses into two separate clusters. The first cluster included all subpopulations of Lip, and the second cluster included all other breeds. Once again, the Lipizzan horse subpopulation showed greater similarity than the OKH colour variants.

The phylogenetic network (Neighbor-Net) based on Reynolds' pairwise genetic distances, constructed from predefined breeds, complements the VAE results (Fig. 5). The Neighbor-Net analysis clearly grouped Baroque breeds into two clusters. The first cluster comprised South American and Iberian breeds together with FriCZ and ArabHR, whereas the second cluster included Lipizzan and Old Kladruby subpopulations. Notably, Old Kladruby horses were positioned closer to the first cluster than the Lipizzan subpopulations. With the addition of other publicly available horse breeds (Petersen et al., 2013 and Cosgrove et al., 2020), Neighbour-Net distinctly grouped the Baroque breeds into three clusters: South American, Iberian, and Central European Baroque horses. The intermediate position of the Lipizzan subpopulations between the grey and black variants of the Old Kladruby horse reflects the historical regeneration process of the breed. During the regeneration programme, each colour variety was restored using different founder breeds. Specifically, grey Old Kladruby horses were regenerated using Lipizzan and Lusitano stallions, whereas black Old Kladruby horses were regenerated using black Lipizzan and Friesian stallions (Bílek, 1957). This distinct use of founder breeds explains the observed molecular patterns: the grey variety shows closer genetic affinity to Iberian-related populations together with Lipizzans, while the black variety occupies an intermediate position between Friesian and Lipizzan horses. Baroque horses occupied a position between Arab and draft horse breeds.

The population relatedness of the 15 analysed Baroque horse breeds, together with Arab breed is presented in the form of pairwise Wright's fixation index (F_{ST}) values, along with the mean F_{ST} (**MF_{ST}**) value for each breed (Supplementary Fig. S5 and Table S3). Estimated F_{ST} values ranged from 0.020 (LipCZ and LipSK) to 0.132 (FriCZ and PRPF), with an overall mean F_{ST} of 0.073. The most genetically distant population was FriCZ, with the highest average F_{ST} of 0.111, whereas the most central breeds were LipHU, Lus, and PRE, all showing the lowest average F_{ST} of 0.057. The lowest F_{ST} values were generally observed within groups of populations, such as among Lipizzan subpopulations. Pairwise F_{ST} between OKH colour variants was relatively high, comparable to - or exceeding - F_{ST} values between each OKH variant and Lipizzan subpopulations, consistent with colour-based substructuring and drift.

Population Structure

Population structure and admixture levels were assessed using NeuralADMIXTURE, an algorithm based on a neural network autoencoder that provides faster computation with fast computation. Based on these analyses, we selected K = 4, 7 and 10 as representative solutions for interpretation (Fig. 6B). K = 4 captures the major genetic structure with the broad differentiation among populations. K = 7 was chosen based on the first plateau observed in the MedMed K curve, reflects intermediate structuring of subpopulations. Finally, K = 10 was identified as the most suitable level for our dataset, corresponding to the highest peak in the MedMed K and revealing a clear and detailed partitioning of the populations. K=10 was also verified as the most suitable level using hierarchical iterative analysis according to Coulon et al. (2008) (Supplementary Fig. S6). At K = 4, the breeds separated clearly into major groups: one cluster comprised Lipizzan subpopulations, the second and third distinguished the two-colour variants of the Old Kladruby, and the fourth combined the Iberian and South American breeds together with the Friesian (FriCZ) and Arab. At this level, genetic admixture was minimal, with LipSI showing the lowest degree. Note that K reflects model complexity rather than the true number of ancestral populations.

At K = 7, the ArabHR and FriCZ breeds separated from the Iberian and South American groups and each formed distinct clusters, likely reflecting their different breeding practices and selection histories. At the same level, the Lipizzan subpopulations further subdivided, with LipSI becoming more distinct. This separation of Lipizzan subpopulations corresponds to two distinct breeding strategies: one emphasising classical dressage and riding (Austria, Italy, Slovenia) and the other oriented towards carriage driving (Croatia, Hungary, Slovakia), as reported by Achmann et al. (2024).

At K = 10, the PRE and Cart separated from the Lus and South American breeds, and the LipBA subpopulation diverged from the other Lipizzans. The K = 10 revealed a shared ancestry component between the ArabHR and Lusitano. The OKH subpopulations remained well differentiated, consistent with colour-based substructuring and drift, while at K = 10 the Cartujano (Cart) also emerged as genetically distinct from the PRE. The South American breeds consistently showed a high degree of similarity to one another. By randomly selecting 50 individuals from large vulture populations to reduce the potential influence of uneven sample sizes on population structure, the same population grouping was achieved as for all genotyped individuals (Supplementary Fig. S7). The result of the NeuralADMIXTURE analysis can be considered consistent.

The results of NeuralADMIXTURE at K = 10 were further used to evaluate the degree of admixture in individual breeds using the ENPC metric (Table 1). The highest ENPC values were observed in LipHR (ENPC = 3.19), Lus (ENPC = 2.82), and OKHg (ENPC = 2.45), while the lowest was found in FriCZ (ENPC = 1.00), and Cart (ENPC = 1.06). The remaining breeds showed relatively similar ENPC values, ranging from 1.0 to 2.0. These ENPC estimates are consistent with the admixture patterns inferred from NeuralADMIXTURE. The median ENPC across all breeds was 1.47. ENPC was computed at K=10 as ENPC=1/Σ_k q_k^2 (breed-level ENPC = mean of individual ENPCs).

Gene Flow Between Breeds

The analysis of current gene flow among the 16 analysed breeds (Fig. 7) complements the admixture patterns identified with NeuralADMIXTURE. Runs showed acceptable mixing (proposed acceptance 25–45%) and Effective sample sizes > 2 700 across parameters (Supplementary Table S1 and Fig. S1). We considered migration statistically supported when the 95% credible interval (**CI**) excluded 0; among those, we report as biologically relevant only rates whose 95% CI was entirely above 0.01. (Supplementary Table S1). Recent between-breed migration was detected only among Iberian breeds, whereas high levels of within-breed migration were observed among Lipizzan and Old Kladruby subpopulations, with no detectable migration between Lipizzan and Old Kladruby breeds. A weak signal of migration between PRPF and Cri (m = 0.018) was observed, but its 95% CI included 0, hence not considered supported under our criteria. The non-significant migration among South American breeds indicates that the admixture detected by NeuralADMIXTURE reflects historical rather than ongoing processes. A supported signal of recent gene flow was inferred between Arab and Lusitano, consistent with the shared-ancestry component seen in NeuralADMIXTURE. High levels of recent migration were also observed among subpopulations of Lipizzan and Old Kladruby horses. The strongest migration was recorded between LipCZ and LipSK (m = 0.088), followed by migration into LipSI from LipHU and LipHR (m = 0.084 and 0.071, respectively). Among the Old Kladruby horse colour variants, gene flow from the grey to the black variety was stronger than in the opposite direction (m = 0.043 vs. 0.022). Gene flow between the colour variants of the Old Kladruby horse is primarily driven by the autosomal dominant inheritance of the grey coat colour allele over the black allele, together with the presence of heterozygous individuals in the population. The reverse gene flow resulted from past breeding practices aimed at increasing genetic variability in the grey variant through controlled mating with black individuals, which facilitated genetic exchange between the two colour variants. Between Cart and PRE, migration was detected only from Cart to PRE (m = 0.065) and not in the opposite direction (Supplementary Table S1), underscoring the closed and highly conserved breeding of the Cartujano strain.

Discussion

The conservation or management of genetic variability in livestock is a challenging task, particularly for breeds with small population sizes and historical importance, such as the Baroque horse breeds of Central Europe. These breeds have been closed populations for long period, meaning that only individuals belonging to the breed can be used for reproduction, which makes inbreeding increase unavoidable and leads to a gradual loss of genetic variability. While several studies have compared the genetic similarity of some Baroque breeds using STR markers (e.g., Achmann et al., 2004;

Kasarda et al., 2016; Grilz-Seger et al., 2019), this study provides the first broad overview of a large group of Baroque breeds based on SNP markers. The values of genetic variability parameters – such as expected and observed heterozygosity and F_{IS} – estimated from haplotype blocks (Table 1) were up to twice as high as those derived from biallelic SNP markers (Supplementary Table S2), consistent with previous findings that haplotype blocks capture greater variability (Simčič et al., 2015; Vostrý et al., 2024). Moreover, the heterozygosity values based on haplotype blocks were comparable to those from microsatellite analyses (Achmann et al., 2004; Kasarda et al., 2016; Vostry et al., 2011), reflecting their methodological similarity (approximately one microsatellite roughly corresponds to a four SNP haplotype blocks). In contrast, estimates from biallelic SNP markers aligned well with earlier SNP-based studies of horses (Petersen et al., 2013; Cosgrove et al., 2020; Criscione et al., 2022). Negative F_{IS} values indicate a preference for outbreeding or avoidance of close-relative mating, whereas positive F_{IS} values reflect inbreeding; in this study, negative F_{IS} values were observed within Central European subpopulations (Lip and OKH). Despite breeders' efforts to promote outbreeding, differences in breeding strategies among Lipizzan subpopulations (Achmann et al., 2004) and relatively high migration between subpopulations have not precluded elevated levels of genomic autozygosity in Lipizzan. Similarly, a comparable pattern was observed in the Old Kladruby horse. When considering genealogical data, the rise in homozygosity across most populations, despite negative F_{IS} values, reflects the closed nature of these breeds and their limited effective number of ancestors ($\boldsymbol{f_a}$ = 21 or 11 for PRE and Cart, Poyato-Bonilla et al., 2022; f_a = 36 for LipSK, Pjontek et al., 2012; f_a = 13 and 9 for OKHg and OKHb, Vostra-Vydrova et al., 2016). This loss of genetic diversity is further evident in the relatively high genomic inbreeding values (Table 1 or Fig. 1) detected in all populations, with the highest recorded in Friesian horses bred in Czechia, consistent with Schurink et al. (2019), who reported a value of 0.22. F_{ROH} values in other breeds also align with previous studies, such as Lipizzan horses (Grilz-Seger et al., 2019) and PRE (Poyato-Bonilla et al., 2022). Corresponding average F_{ROH} values as for the analysed breeds were also estimated for Quarter Horse racing line $F_{ROH>1Mb}$ = 0.19 (Santos et al., 2023). A similar value of inbreeding coefficient to that in our analysis was also estimated by Karlau et al. (2025) using genealogical data (F_x = 0.09). Notably, despite their small global population size (~2 000 individuals), both OKH colour variants showed F_{ROH} values comparable to other Baroque breeds, likely due to the avoidance of close inbreeding, as supported by negative F_{IS} values. Differences between F_{HBD} and F_{ROH} results stem from methodological variation. In this study, medium-density SNP chips were used, which allow accurate estimation of ROH longer than 2 Mb (Ferenčaković et al., 2013). Although shorter HBD segments were not considered in this study, the model-based approach is capable of detecting HBD associated with such segments (Druet and Gautier, 2017). We therefore emphasise F_{HBD} for quantifying overall autozygosity.

The effective population size plays a key role in genetic improvement programs for domestic animals, as it directly influences genetic diversity, control of inbreeding, selection response, adaptability, and the long-term sustainability of populations. Estimating current and historical effective population size based on linkage disequilibrium (N_{eLD}) is the most widely used genomic approach. Differences between the two methods implemented in NeEstimator v2 and GONE2, as observed in this study, have also been reported in previous analyses of various livestock species, such as sheep (Mészáros et al., 2025) and goats (Vostrý et al., 2024). NeEstimator v2

consistently produced often, but not uniformly, lower N_{eLD} estimates than GONE2. Estimates of Ne are highly sensitive to the methodological approaches used. For example, Peterson et al. (2013) estimated significantly higher values of N_e for the analysed breeds using the methodological approach of Weir and Hill (1980). The simulations from several studies (Santiago et al., 2020; Reid and Pinsky, 2022; Novo et al., 2023) suggest that results from GONE and then GONE2 are generally more accurate. However, these simulations did not account for all empirical complexities, such as admixture or population structure, which are known to bias N_{eLD} estimates (Santiago et al., 2020). Because admixed individuals were confirmed in the present dataset, some of the estimates must be interpreted with caution. Despite the closed character of these populations and their relatively high genomic inbreeding, N_{eLD} estimates for all breeds exceeded the FAO-recommended critical threshold of 50 effective individuals for sustainable populations (FAO, 1998), with the only exception being the OKHb population (CI = 46-49). Comparable Ne estimates were also obtained from pedigree data for both colour variant of OKH (Ne = 62 and 45; Vostra-Vydrova et al., 2016), while other breeds showed a wider range of pedigree-based values, from 35 to 77 for PRE and Cart (Poyato-Bonilla et al., 2022) to 117 for LipSK (Pjontek et al., 2014). Since genomic data reflect the realized genetic makeup of individuals, genomic N_{eLD} is generally considered more accurate than pedigree-based estimates. On the other hand, the N_{eLD} estimates based on SNP data obtained by Poyato-Bonilla et al. (2022) for PRE and Cart were significantly different compared to the current analysis. The difference could be due to the use of the SNeP software (Barbato et al., 2015) in Poyato-Bonilla et al. (2022), which tends to underestimate the Ne estimates compared to GONE2.

Some caution is warranted in interpreting these results. First, all analysed breeds had Ne below 500, a threshold often cited for long-term adaptive sustainability (Franklin, 1980). Second, our estimates did not consider the Lipizzan as a metapopulation encompassing all subpopulations, nor the Old Kladruby as a single entity across its colour variants. Estimating N_{eLD} for subdivided populations is more complex and would likely yield higher values depending on migration structure (Whitlock and Barton, 1997; Ryman et al., 2019). We regard the estimation of metapopulation effective size in domestic animals - particularly in cross-border breeds - as an important but largely neglected issue that warrants more systematic investigation in a separate study.

Analysis of historical N_{eLD} development revealed two distinct patterns across breeds. The first group, consisting primarily of Lipizzan subpopulations, Friesian, Cartustion and Lusitano, experienced a marked population declines around 22 generations ago, while the second group, including all other monitored populations, underwent a similar reduction about 12 generations ago. Assuming a generation interval of approximately 10 years in Baroque horses (Pjontek et al., 2014; Vostra-Vydrova et al., 2016; Steensma et al., 2024), these population declines correspond to major historical disruptions: for most breeds, the decline occurred after World War I (~12 generations ago), when horse populations dropped sharply, whereas in the Lipizzan breed it can be traced to the late 18th and early 19th centuries, during the Napoleonic Wars (~22 generations ago), which had a profound effect on their subsequent development. Although the FriCZ population represents a specific subgroup and may not fully reflect the entire Friesian breed, the historical N_{eLD} of FriCZ showed a trend similar to that observed in the analysed Lipizzan subpopulations. In FriCZ, a marked decline in effective population size occurred approximately 24

generations ago and was more pronounced than in the other analysed breeds. However, as Waples and England (2011) and Novo et al. (2023) showed that current N_{eLD} estimates based on linkage disequilibrium are not significantly affected by admixture, the FriCZ population was retained for the estimation of current N_{eLD}.

In this study, the relationships among Baroque horse breeds were examined using multiple approaches (F_{ST}, VAE, NeuralADMIXTURE, Neighbor-Net, and NetView). The analysed Baroque breeds formed homogeneous clusters without significant overlap either within the VAE graphical output or with other horse breeds (Fig. 4). The VAE embedding separated breeds into coherent clusters; while axes lack direct biological interpretability, Iberian breeds (PRE, Cart, Lus) occupied an intermediate position in the latent space. The higher genetic variability observed between the two OKH colour variants reflects their distinct regeneration histories. The higher genetic variability observed between the two OKH colour variants reflects their distinct regeneration histories. Both colour varieties of OKH are bred separately with controlled gene flow, which occurs mainly from the grey to the black subpopulation when two grey heterozygous individuals (Gg) at the grey locus produce a black offspring through Mendelian segregation (probability 0.25). In recent history, black individuals have also been crossed with grey ones to a limited extent to increase genetic diversity. The separation of the OKH subpopulations has historical origins: when the Old Kladruby horse nearly became extinct between World War I and World War II, a regeneration program was initiated. This program was carried out between 1934 and 1974 and involved the use of Lusitano and grey Lipizzan stallions for the grey variety and black Lipizzan and Friesian stallions for the black variety (Bílek 1957). The use of different breeds in the regeneration process significantly contributed to the higher genetic variability within and between the colour variants of the Old Kladruby horse. The application of VAE in this context highlights its utility as a powerful tool in population genomics, enabling the detection of hidden nonlinear structures and providing deeper insights into relatedness and genetic diversity within and between populations (Battey et al., 2021). Reynolds distances placed Baroque breeds distinct from other groups (Fig. 5). The colour variants of OKH were split by Lip cluster. The historical process of OKH regeneration also explains the insertion of Lipizzan subpopulations among the colour variants of OKH and the closeness of OKHg to Iberian breeds and OKHb to the Friesian breed in the Neighbor-Net analysis.

Analysis of admixture patterns revealed clear differences among Lipizzan subpopulations. The most pronounced separation was observed between LipSI and the other subpopulations, as well as between LipBA and primarily LipHU, LipSK, and LipCZ (K=10). These patterns largely reflect geopolitical influences. As reported by Achmann et al. (2004), LipSI is primarily associated with the Austrian and Italian Lipizzan population and shows limited genetic exchange with other groups. The ancestors of LipCZ individuals derives mainly from LipSK and LipHU, although genetic admixing with LipHR remains evident due to exchanges during the Yugoslav era. As noted by Machová et al. (2023), such subpopulation differences within Lip or OKH should not be interpreted as evidence that they represent separate breeds, since the K value is derived from the number of different events such population declines, populations splits and merging's or gene flow between subpopulations. Historical contributions from Iberian breeds, especially PRE, are still detectable in both OKH and Lip, consistent with their documented role in breed formation. Evidence of historical mixing between Lip and OKH is also clear, reflecting frequent gene flow between these two breeds during the Habsburg monarchy. In addition, Arab admixture in Lipizzans is

well documented, as Arab were historically introduced to improve riding qualities, a practice also applied in the Lusitano. A shared ancestry component between Arab and the Friesian subset (FriCZ) was visible at low K (≤8). Historical connections between Friesian and Iberian breeds are documented. However, at K=10, the Friesian separated as a distinct breed, and this signal of admixture was no longer observed.

Recent gene flow was analysed using BayesAss; however, this method has limitations, especially when estimating long-term migration links, which prevents the investigation of historical interactions. Overall, it can be said that the current migration rates derived between Baroque horse breeds provide insight into the observed degree of admixture and complements the NeuralADMIXTURE approach. However, caution should be exercised when interpreting these estimates, especially when model assumptions are challenged. Estimating gene flow using BayesAss is particularly sensitive when F_{ST} is very low (e.g., <0.02). In our dataset, several breed pairs showed much higher differentiation (up to F_{ST}≈0.13), for which migration estimates are generally more reliable. This value was not exceeded between any pair of breeds. Even when model assumptions are violated, accurate estimates of migration rates exist when F_{ST} estimates exceed 0.100, as evident in several pair estimates in this study. Rigorous and comprehensive analyses, including model selection from 3 optimal replicates out of 10, mitigated convergence and other problems associated with BayesAss estimation in this study (Faubet et al., 2007; Meirmans, 2014).

The analysed breeds of Baroque horses were relatively homogeneous, mostly with ENPC values below 2. The calculation of the ENPC follows the logic of the definition of the effective number of alleles proposed by Kimura and Crow (1964), which is defined here as the effective number of parental clusters that contribute equally to a breed (population). For example, an ENPC value of 4.0 would correspond to an admixture of four breeds (populations) each contributing 25% of the genes, while an ENPC value of 2.0 would correspond to a breed (population) in which two breeds each contribute 50% of the genes. In this study, we found that ENPC is an appropriate metric that quantifies admixture with a single number. We consider that ENPC is useful to explain the variations of middle inbreeding ($F_{ROH4\text{-}8Mb}$), as the negative linear regression coefficient of $F_{ROH4\text{-}8Mb}$ on ENPC showed a statistical trend (P=0.09), with a remarkable R^2=0.19.

In the event of an unexpected catastrophe, such as the outbreak of a severe disease or extreme environmental conditions, the results suggest a high degree of genetic compatibility between Iberian and South American breeds. This indicates that Iberian horse populations (PRE, Cart, Cri, Lus) could potentially be regenerated among themselves or, if necessary, with support from South American breeds, and vice versa. The same applies to Lipizzan subpopulations, which appear capable of mutual regeneration. In the case of a threat to the Old Kladruby horse, a regeneration process similar to that successfully implemented in the past could potentially be repeated.

Conclusion

The genetic variability of 15 Baroque horse breeds was evaluated using medium-density SNP chips and compared with the Croatian Arab breed used as an outgroup. Signals of reduced genetic diversity, reflected by genomic inbreeding and

NeLD, were observed across breeds. All breeds showed genetic distinctiveness, highlighting their uniqueness. Marked genetic differences were also detected among Lipizzan, as well as between the colour variants of the Old Kladruby horse. These differences, consistent across VAE/NetView, Neighbor-Net and NeuralADMIXTURE, are particularly valuable for the long-term preservation of genetic diversity - especially in Old Kladruby horse, a breed with very small census size in this dataset. Overall, the level of genetic diversity in the studied Baroque breeds was comparable to that reported for other horse breeds with similar histories and census sizes. Likewise, the patterns of genetic similarity and admixture corresponded closely with results from pedigree analyses and microsatellite-based studies. The results of the present study can provide a strong basis for improving the conservation process of Baroque horses, especially those with smaller populations, and, together with ex-situ conservation programs, for possible regeneration programs for these breeds.

Ethical approval

Not applicable

Data and model availability statement

None of the data was deposited in an official repository. These datasets were used under a material transfer agreement, and hence, are not publicly available. The datasets used and analysed during the current study are available upon reasonable request to the authors who provide the data: PRE – A. Molina, Cart – M. Valera, Cri - S. Demyda-Peyrás, LipSK – R. Kasarda, LipHR – V. Cubric-Curik and the rest to L. Vostry.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author used Grammarly and ChatGPT to check grammar, spelling and sentence clarity were improved to enhance readability and ensure correct use of English. After using these applications, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Author ORCIDs

H Vostra-Vydrova: https://orcid.org/0000-0002-3443-964X

N. Moravcikova: https://orcid.org/0000-0003-1898-8718

R. Kasarda: https://orcid.org/0000-0002-2723-3192

M. Shihabi: https://orcid.org/0000-0001-7290-1957

Journal Pre-proofs

T. Druet: https://orcid.org/0000-0003-1637-1706

A. Halvonik: https://orcid.org/0000-0003-4849-0830

B. Hofmanova: https://orcid.org/0000-0001-5266-3170

P. Dovc: https://orcid.org/0000-0002-7418-0054

M. Zorc: https://orcid.org/0000-0003-3330-7909

A. Molina: https://orcid.org/0000-0002-5571-7258

S. Demyda-Peyrás: https://orcid.org/0000-0003-3286-2441

M. Valera: https://orcid.org/0000-0003-1742-550X

J. Posta: https://orcid.org/0000-0002-7208-1896

M. Kovács: https://orcid.org/0009-0009-9944-0367

B. Rogić: https://orcid.org/0000-0002-7808-6268

V. Cubric-Curik: https://orcid.org/0000-0002-2223-5957

L. Vostry: https://orcid.org/0000-0001-9778-980X

I. Curik: https://orcid.org/0000-0001-7090-1654

Declaration of interest

None

Acknowledgements

Thanks to National Stud Kladurby nad Labem in Czechia and National stud Topolcanky in Slovakia for providing samples for analysis.

Financial support statement

H.V.-V., B.H. and L.V. acknowledge support from the Czech Science Foundation (GACR 24-14325L), N.M., R.K. and A.H. acknowledge support from the Slovak Research and Development Agency (Bratislava, Slovakia; APVV-20-0161) and by the Ministry of Education, Research, Development and Youth of the Slovak Republic (grant no. 1/0316/25), V.C.-C was supported by the Croatian Science Foundation under project number HRZZ IP-2022-10-8926, I.C. was supported by the Flagship Research Groups Programme of the Hungarian University of Agriculture and Life Sciences and S.D.-P. acknowledge support from the AEI, Spain (CNS2024-154334).

Journal Pre-proofs

References

Achmann, R., Curik, I., Dovc, P., Kavar, T., Bodo, I., Habe, F., Marti, E., Sölkner, J., Brem, G., 2004. Microsatellite diversity, population subdivision and gene flow in the Lipizzan horse. Animal Genetics 35, 285–292. doi:10.1111/j.1365-2052.2004.01157.x

Adamack, A.T., Gruber, B., 2014. PopGenReport: simplifying basic population genetic analyses in R. Methods in Ecology and Evolution 5, 384–387. doi:10.1111/2041-210X.12158

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19, 1655–1664. doi:10.1101/gr.094052.109

Amos, W., Balmford, A., 2001. When does conservation genetics matter? Heredity 87, 257–265. doi:10.1046/j.1365-2540.2001.00940.x

Battey, C.J., Coffing, G.C., Kern, A.D., 2021. Visualizing population structure with variational autoencoders. G3 Genes|Genomes|Genetics 11, jkaa036. doi:10.1093/g3journal/jkaa036

Bílek, F., 1957. The Old Kladrub Horse. In Speciální zootechnika II. Chov koní. (ed. Koubek, K., Bulánek, J.). Státní zemědělské nakladatelství, Praha, CZ, pp. 74-104 (in Czech).

Bertrand, A.R., Kadri, N.K., Flori, L., Gautier, M., Druet, T., 2019. RZooRoH: An R package to characterize individual genomic autozygosity and identify homozygous-by-descent segments. Methods in Ecology and Evolution 10, 860–866. doi:10.1111/2041-210X.13167

Biscarini, F., P. Cozzi, G. Gaspa, and G. Marras. 2019. Detect runs of homozygosity and runs of heterozygosity in diploid genomes. R package version 0.9.5. Accessed 20 June 2019. https: //cran .r -project.org/ web/ packages/ detectRUNS/

Chen, C., Zhu, B., Tang, X., Chen, B., Liu, M., Gao, N., Li, S., Gu, J., 2023. Genome-Wide Assessment of Runs of Homozygosity by Whole-Genome Sequencing in Diverse Horse Breeds Worldwide. Genes 14, 1211. doi:10.3390/genes14061211

Coulon, A., Fitzpatrick, J.W., Bowman, R., Stith, B.M., Makarewich, C.A., Stenzler, L.M., Lovette, I.J., 2008. Congruent population structure inferred from dispersal behaviour and intensive genetic surveys of the threatened Florida scrub-jay (Aphelocoma cœrulescens). Molecular Ecology 17, 1685–1701. doi:10.1111/j.1365-294X.2008.03705.x

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. doi:10.1038/s41598-020-66232-1

Criscione, A., Mastrangelo, S., D'Alessandro, E., Tumino, S., Di Gerlando, R., Zumbo, A., Marletta, D., Bordonaro, S., 2022. Genome-wide survey on three local horse populations with a focus on runs of homozygosity pattern. Journal of Animal Breeding and Genetics 139, 540–555. doi:10.1111/jbg.12680

Do, C., Waples, R.S., Peel, D., Macbeth, G.M., Tillett, B.J., Ovenden, J.R., 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Molecular Ecology Resources 14, 209–214. doi:10.1111/1755-0998.12157

Dominguez Mantes, A., Mas Montserrat, D., Bustamante, C.D., Giró-i-Nieto, X., Ioannidis, A.G., 2023. Neural ADMIXTURE for rapid genomic clustering. Nature Computational Science 3, 621–629. doi:10.1038/s43588-023-00482-7

Druet, T., Gautier, M., 2022. A hidden Markov model to estimate homozygous-by-descent probabilities associated with nested layers of ancestors. Theoretical Population Biology 145, 38–51. doi:10.1016/j.tpb.2022.03.001

Druet, T., Gautier, M., 2017. A model-based approach to characterize individual inbreeding at both global and local genomic scales. Molecular Ecology 26, 5820–5841. doi:10.1111/mec.14324

FAO, 1998. Secondary guidelines for development of national farm animal genetic resources management plans: animal recording for medium input production environment. FAO, Rome, Italy.

Faubet, P., Waples, R.S., Gaggiotti, O.E., 2007. Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. Molecular Ecology 16, 1149–1166. doi:10.1111/j.1365-294X.2007.03218.x

Ferenčaković, M., Sölkner, J., Curik, I., 2013. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. Genetics Selection Evolution 45, 42. doi:10.1186/1297-9686-45-42

Forneris N.S., Faux P., Gautier M., Druet, T. Updating the RZooRoH package for the analysis of inbreeding, identity-by-descent and relatedness from genomic data. bioRxiv 2025.07.08.663647; doi: https://doi.org/10.1101/2025.07.08.663647 (Posted 08 July 2025)

Gautier, M., Laloë, D., Moazami-Goudarzi, K., 2010. Insights into the Genetic History of French Cattle from Dense SNP Data on 47 Worldwide Breeds. PLoS ONE 5, e13038. doi:10.1371/journal.pone.0013038

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian Data Analysis, 3rd edition. CRC Press, Boca Raton, FL, USA.

Gómez, M.D., Valera, M., Molina, A., Gutiérrez, J.P., Goyache, F., 2009. Assessment of inbreeding depression for body measurements in Spanish Purebred

(Andalusian) horses. Livestock Science 122, 149–155. doi:10.1016/j.livsci.2008.08.007

Grilz-Seger, G., Druml, T., Neuditschko, M., Dobretsberger, M., Horna, M., Brem, G., 2019. High-resolution population structure and runs of homozygosity reveal the genetic architecture of complex traits in the Lipizzan horse. BMC Genomics 20, 174. doi:10.1186/s12864-019-5564-x

Grilz-Seger, G., Mesarič, M., Cotman, M., Neuditschko, M., Druml, T., Brem, G., 2018. Runs of Homozygosity and Population History of Three Horse Breeds With Small Population Size. Journal of Equine Veterinary Science 71, 27–34. doi:10.1016/j.jevs.2018.09.004

Hein, J., M. H. Schierup, and C. Wiuf. 2004. Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory. Oxford University Press, Oxford, UK.

Hill, W.G., 1981. Estimation of effective population size from data on linked genes. Advances in Applied Probability 13, 4–4. doi:10.2307/1426455

Huson, D.H., Bryant, D., 2006. Application of Phylogenetic Networks in Evolutionary Studies. Molecular Biology and Evolution 23, 254–267. doi:10.1093/molbev/msj030

Kalinowski, S.T., 2004. Counting Alleles with Rarefaction: Private Alleles and Hierarchical Sampling Designs. Conservation Genetics 5, 539–543. doi:10.1023/B:COGE.0000041021.91777.1a

Karlau, A., Azcona, F., Molina, A., Trigo, P., Sánchez-Serrano, J.P., Demyda-Peyrás, S. 2025. Exploring the Genetic Link Between Coat Colour and Morphological Traits: The Case of Peruano de Paso Horse. Animals 15, 2720. https://doi.org/10.3390/ani15182720

Kasarda, R., Vostrý, L., Moravčíková, N., Vostrá-Vydrová, H., Dovč, P., Kadlečík, O., 2016. Detailed insight into genetic diversity of the Old Kladruby horse substructure in comparison to the Lipizzan breed. Acta Agriculturae Scandinavica, Section A - Animal Science 66, 67–74. doi:10.1080/09064702.2016.1249400

Kingma, D.P., Welling, M., 2013. Auto-Encoding Variational Bayes. arXiv preprint arXvi: 1312.6114. doi:10.48550/ARXIV.1312.6114 (Posted 20 December 2013).

Laseca. N., Molina, A., Ramón, M., Valera, M., Azcona, F., Encina, A., Demyda-Peyrás, S., 2022. Fine-Scale Analysis of Runs of Homozygosity Islands Affecting Fertility in Mares. Frontiers in Veterinary Science 17, 754028. doi:10.3389/fvets.2022.754028.

Lencz, T., Lambert, C., DeRosse, P., Burdick, K.E., Morgan, T.V., Kane, J.M., Kucherlapati, R., Malhotra, A.K., 2007. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. Proceedings of the National Academy of Sciences 104, 19942–19947. doi:10.1073/pnas.0710021104

Journal Pre-proofs

Li, Y., Liu, J., 2018. StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. Molecular Ecology Resources 18, 176–177. doi:10.1111/1755-0998.12719

Luís, C., Juras, R., Oom M.M., Cotham, E.G., 2007. Genetic diversity and relationships of Portuguese and other horse breeds based on protein and microsatellite loci variation. Animal Genetic 38, 20-27. doi: 10.1111/j.1365-2052.2006.01545.x

Luštrek, B., Šimon, M., Turk, K., Bogičević, S., Potočnik, K., 2025. Comparing Genomic and Pedigree Inbreeding Coefficients in the Slovenian Lipizzan Horse as a Case Study for Small Closed Populations. Animals 15, 2774. doi:10.3390/ani15192774

Machová, K., Marina, H., Arranz, J.J., Pelayo, R., Rychtářová, J., Milerski, M., Vostrý, L., Suárez-Vega, A., 2023. Genetic diversity of two native sheep breeds by genome-wide analysis of single nucleotide polymorphisms. Animal 17, 100690. doi:10.1016/j.animal.2022.100690

Meirmans, P.G., 2014. Nonconvergence in Bayesian estimation of migration rates. Molecular Ecology Resources 14, 726–733. doi:10.1111/1755-0998.12216

Mészáros, G., Fürst-Waltl, B., Sölkner, J., Vostrá-Vydrová, H., Calta, J., Đokić, M., Marković, B., Stanojević, D., Bogdanovic, V., Moravčíková, N., Chalupková, M., Halvoník, A., Kasarda, R., Vostrý, L., 2025. Genomic diversity characteristics of sheep populations in the Czech Republic, Slovakia, Serbia and Montenegro. Czech Journal of Animal Science 70, 234–241. doi:10.17221/69/2025-CJAS

Müller-Unterberg, M., Wallmann, S., Distl, O., 2017. Effects of inbreeding and other systematic effects on fertility of Black Forest Draught horses in Germany. Acta Veterinaria Scandinavica 59, 70. doi:10.1186/s13028-017-0338-4

Novo, I., Pérez-Pereira, N., Santiago, E., Quesada, H., Caballero, A., 2023. An empirical test of the estimation of historical effective population size using *Drosophila melanogaster*. Molecular Ecology Resources 23, 1632–1640. doi:10.1111/1755-0998.13837

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., M. Wade, C., McCue, M.E., 2013. Genetic Diversity in the Modern Horse Illustrated from Genome-Wide SNP Data. PLoS ONE 8, e54997. doi:10.1371/journal.pone.0054997

Pembleton, L.W., Cogan, N.O.I., Forster, J.W., 2013. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Molecular Ecology Resources 13, 946–952. doi:10.1111/1755-0998.12129

Pjontek, J., Kadlečík, O., Kasarda, R., Horný, M., 2012. Pedigree analysis in four Slovak endangered horse breeds. Czech Journal of Animal Science 57, 54–64. doi:10.17221/5132-CJAS

Pogorevc, N., Dotsev, A., Upadhyay, M., Sandoval-Castellanos, E., Hannemann, E., Simčič, M., Antoniou, A., Papachristou, D., Koutsouli, P., Rahmatalla, S., Brockmann, G., Sölkner, J., Burger, P., Lymberakis, P., Poulakakis, N., Bizelis, I., Zinovieva, N., Horvat, S., Medugorac, I., 2023. Whole-genome SNP genotyping unveils ancestral and recent introgression in wild and domestic goats. Molecular Ecology 33, e17190. doi:10.1111/mec.17190

Poyato-Bonilla, J., Laseca, N., Demyda-Peyrás, S., Molina, A., Valera, M., 2022. 500 years of breeding in the *Carthusian Strain* of Pura Raza Español horse: An evolutional analysis using genealogical and genomic data. Journal of Animal Breeding and Genetics 139, 84–99. doi:10.1111/jbg.12641

Puechmaille, SJ. 2016. The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. Molecular Ecology Resources 16, 608-627. doi: 10.1111/1755-0998.12512

Raguz, N., Korabi, N., Lukić, B., Drzaic, I., Vostry, L., Moravcikova, N., Curik, I., Kasarda, R., Cubric-Curik, V., 2023. Genomic characterization and population structure of Croatian Arabian horse. Livestock Science 277, 105343. doi:10.1016/j.livsci.2023.105343

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Systematic Biology 67, 901–904. doi:10.1093/sysbio/syy032

Reid, B.N., Pinsky, M.L., 2022. Simulation-Based Evaluation of Methods, Data Types, and Temporal Sampling Schemes for Detecting Recent Population Declines. Integrative and Comparative Biology 62, 1849–1863. doi:10.1093/icb/icac144

Reynolds, J., Weir, B.S., Cockerham, C.C., 1983. Estimation of the Coancestry Coefficient: Basis for a Short-Term Genetic Distance. Genetics 105, 767–779. doi:10.1093/genetics/105.3.767

Ripoles-Lobo, M., Perdo-gonzálesz, D.I., Azor, P.J. Valera, M., 2023. Evaluation of potential effects and genetic parameters in conformational limb defects in Pura Raza Espanola horses. Italian Journal of Animal Science 22, 407–417. https://doi.org/10.1080/1828051X.2023.2206419

Ryman, N., Laikre, L., Hössjer, O., 2019. Do estimates of contemporary effective population size tell us what we want to know? Molecular Ecology 28, 1904–1918. doi:10.1111/mec.15027

Santiago, E., Köpke, C., Caballero, A., 2025. Accounting for population structure and data quality in demographic inference with linkage disequilibrium methods. Nature Communications 16, 6054. doi:10.1038/s41467-025-61378-w

Santiago, E., Novo, I., Pardiñas, A.F., Saura, M., Wang, J., Caballero, A., 2020. Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium. Molecular Biology and Evolution 37, 3642–3653. doi:10.1093/molbev/msaa169

Santander, C.G., Refoyo Martinez, A., Meisner, J., 2024. Faster model-based estimation of ancestry proportions. Peer Community Journal 4, e115. doi:10.24072/pcjournal.503

Santos, W.B., Pereira, C.B., Maiorano, A.M., Arce, C.D.S., Baldassini, W.A., Pereira, G.L., Chardulo, L.A.L., Neto, O.R.M., Oliveira, H.N., Curi, R.A., 2023. Genomic inbreeding estimation, runs of homozygosity, and heterozygosity-enriched regions uncover signals of selection in the Quarter Horse racing line. Journal of Animal Breeding and Genetics 140, 583–595. doi:10.1111/jbg.12812

Saura, M., Caballero, A., Santiago, E., Fernández, A., Morales-González, E., Fernández, J., Cabaleiro, S., Millán, A., Martínez, P., Palaiokostas, C., Kocour, M., Aslam, M.L., Houston, R.D., Prchal, M., Bargelloni, L., Tzokas, K., Haffray, P., Bruant, J.-S., Villanueva, B., 2021. Estimates of recent and historical effective population size in turbot, seabream, seabass and carp selective breeding programmes. Genetics Selection Evolution 53, 85. doi:10.1186/s12711-021-00680-9

Simčič, M., Smetko, A., Sölkner, J., Seichter, D., Gorjanc, G., Kompan, D., Medugorac, I., 2015. Recovery of Native Genetic Background in Admixed Populations Using Haplotypes, Phenotypes, and Pedigree Information – Using Cika Cattle as a Case Breed. PLoS ONE 10, e0123253. doi:10.1371/journal.pone.0123253

Steensma, M.J., Doekes, H.P., Pook, T., Derks, M.F.L., Bakker, N., Ducro, B.J., 2024. Evaluation of breeding strategies to reduce the inbreeding rate in the Friesian horse population: Looking back and moving forward. Journal of Animal Breeding and Genetics 141, 668–684. doi:10.1111/jbg.12872

Schurink, A., Shrestha, M., Eriksson, S., Bosse, M., Bovenhuis, H., Back, W., Johansson, A.M., Ducro, B.J., 2019. The Genomic Makeup of Nine Horse Populations Sampled in the Netherlands. Genes 10, 480. doi:10.3390/genes10060480

Szpiech, Z.A., Jakobsson, M., Rosenberg, N.A., 2008. ADZE: a rarefaction approach for counting alleles private to combinations of populations. Bioinformatics 24, 2498–2504. doi:10.1093/bioinformatics/btn478

Todd, E.T., Ho, S.Y.W., Thomson, P.C., Ang, R.A., Velie, B.D., Hamilton, N.A., 2018. Founder-specific inbreeding depression affects racing performance in Thoroughbred horses. Scientific Reports 8, 6167. doi:10.1038/s41598-018-24663-x

van Eldik, P., van der Waaij, E.H., Ducro, B., Kooper, A.W., Stout, T.A.E., Colenbrander, B., 2006. Possible negative effects of inbreeding on semen quality in Shetland pony stallions. Theriogenology 65, 1159–1170. doi:10.1016/j.theriogenology.2005.08.001

Journal Pre-proofs

Vostra-Vydrova, H., Vostry, L., Hofmanova, B., Krupa, E., Zavadilova, L., 2016. Pedigree analysis of the endangered Old Kladruby horse population. Livestock Science 185, 17–23. doi:10.1016/j.livsci.2016.01.001

Vostra-Vydrova, H., Hofmanova, B., Moravcikova, N., Rychtarova, J., Kasarda, R., Machova, K., Brzakova, M., Vostry, L., 2020. Genetic diversity, admixture and the effect of inbreeding on milk performance in two autochthonous goat breeds. Livestock Science 240, 104163. doi:10.1016/j.livsci.2020.104163

Vostry, L., Vostra-Vydrova, H., Moravcikova, N., Kasarda, R., Margetin, M., Rychtarova, J., Drzaic, I., Shihabi, M., Cubric-Curik, V., Sölkner, J., Curik, I., 2024. Genomic analysis of conservation status, population structure, and admixture in local Czech and Slovak dairy goat breeds. Journal of Dairy Science 107, 8205–8222. doi:10.3168/jds.2023-24607

Vostry, L., Vostra-Vydrova, H., Citek, J., Gorjanc, G., Curik, I., 2021. Association of inbreeding and regional equine leucocyte antigen homozygosity with the prevalence of insect bite hypersensitivity in Old Kladruby horse. Animal Genetics 52, 422–430. doi:10.1111/age.13075

Vostry, L., Milerski, M., Schmidova, J., Vostra-Vydrova, H., 2018. Genetic diversity and effect of inbreeding on litter size of the Romanov sheep. Small Ruminant Research 168, 25–31. doi:10.1016/j.smallrumres.2018.09.004

Vostry, L., Kracikova, O., Hofmanova, B., Czernekova, V., Kott, T., Pribyl, J., 2011. Intra-line and inter-line genetic diversity in sire lines of the Old Kladruby horse based on microsatellite analysis of DNA. Czech Journal of Animal Science 56, 163–175. doi:10.17221/1437-CJAS

Waples, R.S., 2021. Relative Precision of the Sibship and LD Methods for Estimating Effective Population Size With Genomics-Scale Datasets. Journal of Heredity 112, 535–539. doi:10.1093/jhered/esab042

Waples, R.S., Antao, T., Luikart, G., 2014. Effects of Overlapping Generations on Linkage Disequilibrium Estimates of Effective Population Size. Genetics 197, 769–780. doi:10.1534/genetics.114.164822

Waples, R.S., England, P.R., 2011. Estimating Contemporary Effective Population Size on the Basis of Linkage Disequilibrium in the Face of Migration. Genetics 189, 633–644. doi:10.1534/genetics.111.132233

Waples, R.S., Do, C., 2008. LDNE : a program for estimating effective population size from data on linkage disequilibrium. Molecular Ecology Resources 8, 753–756. doi:10.1111/j.1755-0998.2007.02061.x

Wilson, G.A., Rannala, B., 2003. Bayesian Inference of Recent Migration Rates Using Multilocus Genotypes. Genetics 163, 1177–1191. doi:10.1093/genetics/163.3.1177

Weir, B.S., Hill, W.G., 1980. Effect of Mating Structure on Variation in Linkage Disequlibrium. Genetics 95, 477–488. doi:10.1093/genetics/95.2.477

Whitlock, M.C., Barton, N.H., 1997. The Effective Size of a Subdivided Population. Genetics 146, 427–441. doi:10.1093/genetics/146.1.427

Wright, S., 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19, 395-420. doi:10.1111/j.1558-5646.1965.tb01731.x

Zechner, P., Sölkner, J., Bodo, I., Druml, T., Baumung, R., Achmann, R., Marti, E., Habe, F., Brem, G., 2002. Analysis of diversity and population structure in the Lipizzan horse breed based on pedigree information. Livestock Production Science 77, 137–146. doi:10.1016/S0301-6226(02)00079-9

Table 1. Genomic observed (H_O) and expected (H_E) heterozygosity, Wright's inbreeding coefficient (F_{IS}), haplotype rarefied richness (HRR for 2n = 40), genomic inbreeding based on rule-based consecutive Runs-of-Homozygosity ($F_{ROH\ >2Mb}$) and model-based (F_{HBD}) approach and effective number of parental clusters (ENPC), all presented with SE, for baroque horse breeds.

Breed	n	H_O	H_E	HRR	F_{IS}	$F_{ROH>2Mb}$	F_{HBD}	ENCP
LipBA	42	0.570±0.001	0.544±0.000	3.13±0.08	-0.019±0.000	0.137±0.005	0.160±0.005	1.24
LipSI	239	0.568±0.000	0.557±0.000	3.28±0.09	-0.048±0.000	0.137±0.002	0.164±0.002	1.30
LipHR	48	0.577±0.000	0.581±0.000	3.52±0.10	0.007±0.000	0.128±0.006	0.150±0.005	3.19

Journal Pre-proofs

LipHU	46	0.586±0.000	0.578±0.000	3.50±0.10	-0.014±0.000	0.109±0.004	0.131±0.005	1.97
LipSK	64	0.580±0.000	0.563±0.000	3.33±0.09	-0.029±0.000	0.118±0.004	0.147±0.004	1.71
LipCZ	24	0.582±0.001	0.570±0.001	3.42±0.09	-0.019±0.001	0.109±0.006	0.141±0.006	1.57
OKHg	290	0.575±0.000	0.560±0.000	3.33±0.09	-0.027±0.000	0.136±0.003	0.150±0.002	2.45
OKHb	119	0.564±0.000	0.548±0.000	3.23±0.09	-0.029±0.000	0.152±0.004	0.166±0.004	1.91
FriCZ	24	0.455±0.001	0.469±0.001	2.82±0.07	0.026±0.001	0.285±0.008	0.266±0.006	1.00
Lus	43	0.606±0.000	0.640±0.000	4.20±0.13	0.049±0.000	0.100±0.007	0.129±0.007	2.82
PRE	40	0.599±0.000	0.609±0.000	3.73±0.11	0.016±0.000	0.099±0.002	0.130±0.006	1.55
Cart	30	0.547±0.001	0.566±0.001	3.42±0.10	0.032±0.001	0.178±0.013	0.199±0.010	1.06
Cri	40	0.617±0.000	0.627±0.000	3.92±0.12	0.017±0.000	0.068±0.009	0.101±0.007	1.40
Peru	21	0.615±0.001	0.620±0.001	3.90±0.12	0.032±0.001	0.063±0.009	0.094±0.009	1.96
PRPF	20	0.581±0.001	0.581±0.001	3.65±0.11	-0.001±0.001	0.108±0.015	0.126±0.013	1.33
ArabHR	71	0.579±0.000	0.609±0.000	3.78±0.12	0.049±0.000	0.128±0.006	0.182±0.007	1.27
Median		0.580	0.578	3.50	0.007	0.123	0.152	1.47

Abbreviations: LipBA - Lipizzan (Bosnia and Herzegovina), LipSI - Lipizzan (Slovenia), LipHR - Lipizzan (Croatia), LipHU – Lipizzan (Hungary), LipSk - Lipizzan (Slovakia), LipCZ - Lipizzan (Czechia), OKHg - Old Kladruby horse (grey), OKHb - Old Kladruby horse (black), FriCZ - Friesian (Czehia), PRE - Pura Raza Española, Cart - Cartujanos, Lus - Lusitano, Cri - Crillos, Peru - Peruvian Paso, PRPF - Puerto Rican Paso Fino, ArabHR -Arabian (Croatia)

Journal Pre-proofs

Journal Pre-proofs

Table 2. Recent genomic effective population size with estimated 95% Confidence Interval (in parentheses) for 16 analysed horse breeds.

Breed	NeEstimator v2	GONE2
LipBA	28 (23 – 34)	49 (40-59)*
LipSI	59 (54 – 64)	75 (70-79)*
LipHR	91 (66 – 141)	106 (85-130)

LipHU	156 (124 – 207)	136 (108-171)
LipSK	33 (28 – 40)	68 (59-80)*
LipCZ	24 (15-46)	59 (44-78)
OKHg	42 (39 – 45)	47 (46-49)*
OKHb	46 (40 – 53)	46 (42-51)
FriCZ	62 (36 -179)	70 (52-96)
Lus	75 (54 – 120)	130 (98-174)
PRE	315 (227 – 505)	127 (99-162)
Cart	89 (61 – 165)	99 (74-131)
Cri	195 (110 – 296)	243 (181-324)
PERU	281 (173 - 713)	193 (124-303)
PRPF	124 (85 - 223)	114 (75-171)
ArabHR	38 (3 - 47)	56 (48-64)*

Where: NeEstimator v2 softare estimates; Do et al. (2014), GONE2 softare estimates; Santiago et al. (2025), Abbreviations: LipBA - Lipizzan (Bosnia and Herzegovina), LipSI - Lipizzan (Slovenia), LipHR - Lipizzan (Croatia), LipHU – Lipizzan (Hungary), LipSk - Lipizzan (Slovakia), LipCZ - Lipizzan (Czechia), OKHg - Old Kladruby horse (grey), OKHb - Old Kladruby horse (black), FriCZ - Friesian (Czehia), PRE - Pura Raza Española, Cart - Cartujanos, Lus - Lusitano, Cri - Crillos, Peru - Peruvian Paso, PRPF - Puerto Rican Paso Fino, ArabHR -Arabian (Croatia)

Journal Pre-proofs

Figure captions

Fig. 1. Distribution of genomic inbreeding coefficients estimated using rule- and model-based approaches for 15 Baroque and one Arab horse breeds. A) Violine plot of rule-based genomic inbreeding coefficient ($F_{ROH>2Mb}$; left hand site) and model-based genomic inbreeding ($\mathbf{F_{HBD}}$; right hand site). B) and C) Stacked-bar presentation of the expected autozygosity origin. Included breeds: Lipizaner subpopulations splited based on country of origin: LipBA (Bosnia and Herzegovina), LipSI (Slovenia), LipHR (Croatia), LipHU (Hungary), LipSK (Slovakia), LipCZ (Czechia); Old Kladruby subpopulation splited base on color varianty: OKHg (grey), OKHb (black); Friesian horse (FriCZ); Iberian breeds: Lusitano (Lus), Pura Raza Española (PRE), Cartujanos (Cart); South American breeds: Criollos (Cri), Peruvian Paso (Peru), Puerto Rican Paso Fino (PRFP) and Arab horse (ArabHR).

Fig. 2. Estimated historical effective population size (N_{eLD}) of Baroque horse breeds (populations) and Arab horse (ArabHR) up to 40 generations ago. Shown are Lipizzan subpopulations defined by national breeding programs (green text): LipBA (Bosnia and Herzegovina), LipSI (Slovenia), LipHR (Croatia), LipHU (Hungary), LipSK (Slovakia), LipCZ (Czechia); The Old Kladruby is represented by two subpopulations reflecting different coat colour: OKHg (grey), OKHb (black). Additional breeds include the Friesian horse (FriCZ); Iberian breeds - Lusitano (Lus), Pura Raza Española (PRE), Cartujanos (Cart); South American breeds - Criollos (Cri), Peruvian Paso (Peru) and Puerto Rican Paso Fino (PRFP); and Arab horse (ArabHR).

Journal Pre-proofs

Fig. 3. Three-dimensional bubble plot of the relationship between rule-based genomic inbreeding coefficient ($F_{ROH>2Mb}$) and model-based genomic inbreeding ($\mathbf{F_{HBD}}$), Wrigh's inbreeding coefficient (F_{IS}) a gamete/linkage effective population size (N_{eLD}, proportional to bubble size) (GONE2) in Baroque horse. Shades of green represent subpopulation of Lippizaner horse, shades of red represent colour variant of Old Kladruby, shades of blue represent Iberian and South American breeds together with Friesian and yellow represent Arab horse. Shown are Lipizzan subpopulations defined by national breeding programs (green text): LipBA (Bosnia and Herzegovina), LipSI (Slovenia), LipHR (Croatia), LipHU (Hungary), LipSK (Slovakia), LipCZ (Czechia); The Old Kladruby is represented by two subpopulations reflecting different breeding regeneration programs: OKHg (grey), OKHb (black). Additional breeds include the Friesian horse (FriCZ); Iberian breeds - Lusitano (Lus), Pura Raza Española (PRE), Cartujanos (Cart); South American breeds - Criollos (Cri/CCR), Peruvian Paso (Peru) and Puerto Rican Paso Fino (PRFP); and Arab horse (ArabHR).

Fig. 4. Relatedness between 15 Baroque horse populations and Arabian horse breed. POPVAE-generated (default settings) clustering based on variational auto encoders (VAE) and presented in a scatter plot to illustrate population genetic variation in the reduced dimensionality of latent variables. The Lipizzan are represented by national breeding programs: LipBA (Bosnia and Herzegovina), LipSI (Slovenia), LipHR (Croatia), LipHU (Hungary), LipSK (Slovakia), LipCZ (Czechia); The Old Kladruby is represented by two subpopulations reflecting different coat colour: OKHg (grey), OKHb (black). Additional breeds include the Friesian horse (FriCZ); Iberian breeds - Lusitano (Lus), Pura Raza Española (PRE), Cartujanos (Cart); South American breeds – Criollos (Cri), Peruvian Paso (Peru) and Puerto Rican Paso Fino (PRFP); and Arab horse (ArabHR).

Fig. 5. Relatedness between 15 Baroque horse populations and Arabian horse breed, visualised with a phylogenetic network (Neighbor-Net) inferred from pairwise Reynolds' genetic distances. Shown are Lipizzan subpopulations according to country of origin: LipBA (Bosnia and Herzegovina), LipSI (Slovenia), LipHR (Croatia), LipHU (Hungary), LipSK (Slovakia), LipCZ (Czechia); The Old Kladruby is represented by two subpopulations reflecting different coat colour: OKHg (grey), OKHb (black). Additional breeds include the Friesian horse (FriCZ); Iberian breeds - Lusitano (Lus), Pura Raza Española (PRE), Cartujanos (Cart); South American breeds – Criollos (Cri), Peruvian Paso (Peru) and Puerto Rican Paso Fino (PRFP); and Arab horse (ArabHR).

Fig. 6. Population structure and admixture of 16 analysed baroque horse breeds together with Arab horse. Selection of representative number of clusters (K) based on plotting MedMed K (Puechmaile, 2016) values (A) and graphical representation of Bayesian unsupervised clustering (K = 4, K= 8, and K = 10) of 740 individual baroque

horse breeds together with Arab horse, performed with the algorithm NeuralADMIXTURE and visualised with the software CLUMPAK (B). Each horse is represented by a single vertical line, divided into K coloured segments whose length is proportional to the estimated membership of the inferred cluster to illustrate the presence of admixture. Shown are Lipizzan subpopulations according to country of origin: Bosnia and Herzegovina (LipBA), Slovenia (LipSI), Hungary (LipHU), Croatia (LipHR), Slovakia (LipSK), Czechia (LipCZ); Old Kladruby horse colour variety: grey (OKHg), black (OKHb). Iberian Breeds: Pura Raza Española (PRE), Cartujanos (Cart), Lusitano (Lus); South Amarika: Criollos (Cri), Peruvian Paso (Peru), Puerto Rican Paso Fino (PRPF); And Friesian (FriCZ) and Arab (ArabHR).

Fig. 7. Estimated relative migration pattern between 15 Baroque horse populations together with Arab horse. The most intense migrations (>1%) are indicated with gene flow directions by black dashed arrows. Lipizzan subpopulation according to country of origin (green): Bosnia and Herzegovina (LipBA), Slovenia (LipSI), Hungary (LipHU), Croatia (LipHR), Slovakia (LipSK), Czechia (LipCZ); Old Kladruby horse colour variety (red): grey (OKHg), black (OKHb). Iberian Breeds (blue): Pura Raza Española (PRE), Cartujanos (Cart), Lusitano (Lus); South Amarika (blue): Criollos (Cri), Peruvian Paso (Peru), Puerto Rican Paso Fino (PRPF). And Friesian (FriCZ) - dark blue, and Arab (ArabHR) - yellow.

Highlights

- Genomic diversity of Baroque horse breeds was characterised using SNP data.

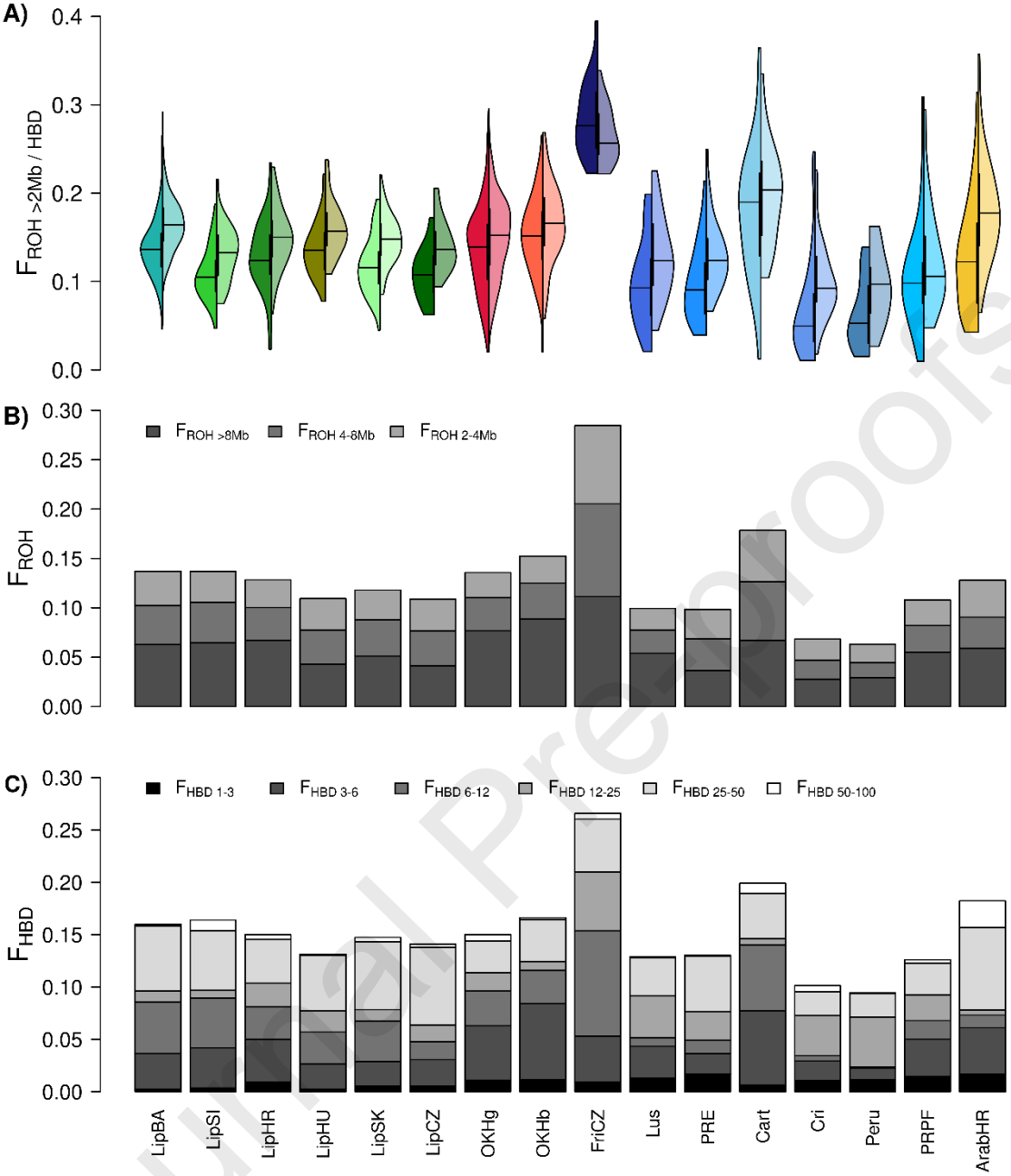
- Baroque breeds show historical admixture without current gene flow.

- Moderate to recent genomic inbreeding levels were identified across breeds.

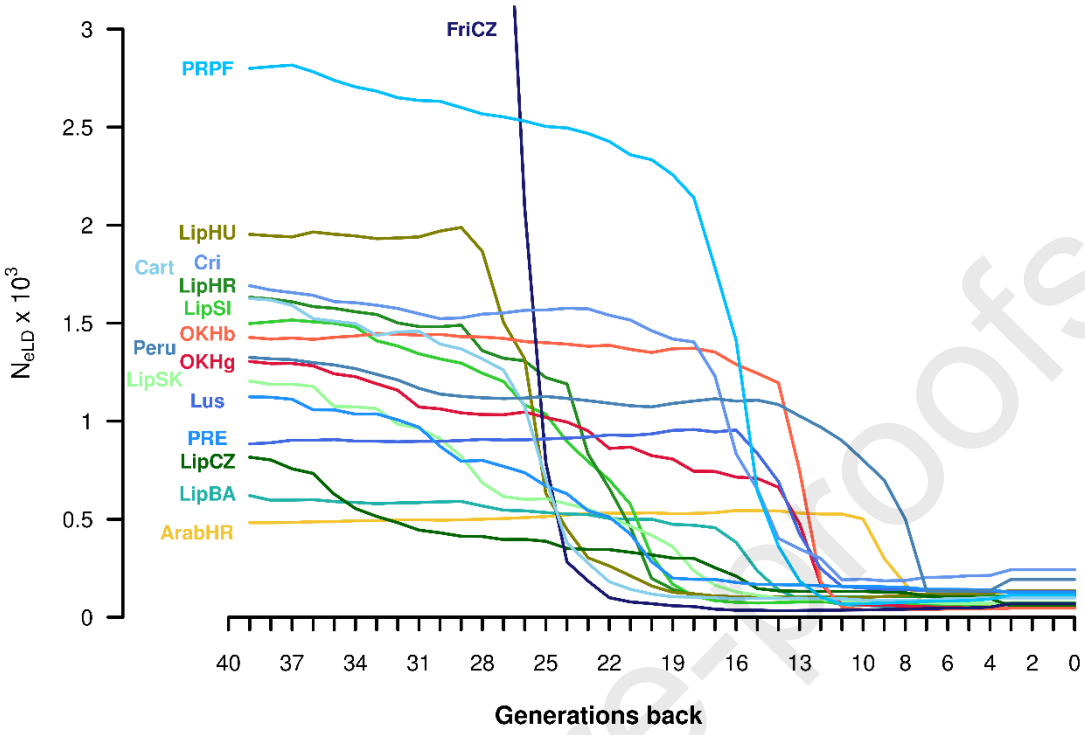
- Several severe historical declines in effective population size were detected.

- Loss of genetic diversity threatens the long-term sustainability of some breeds.

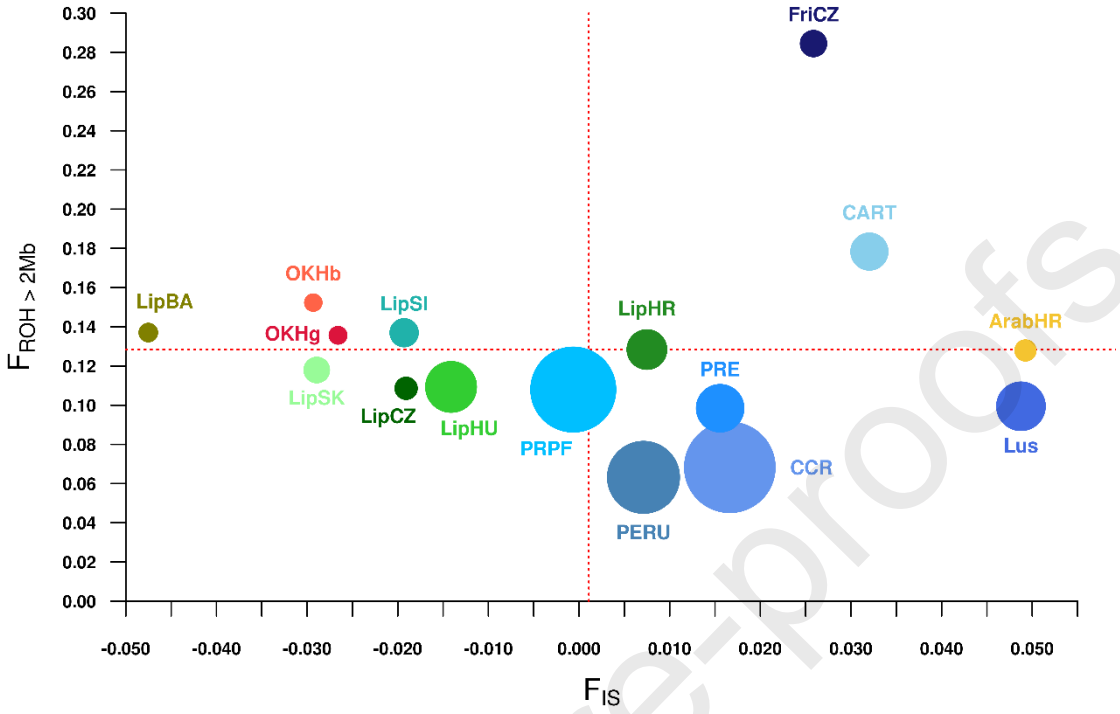
Journal Pre-proofs

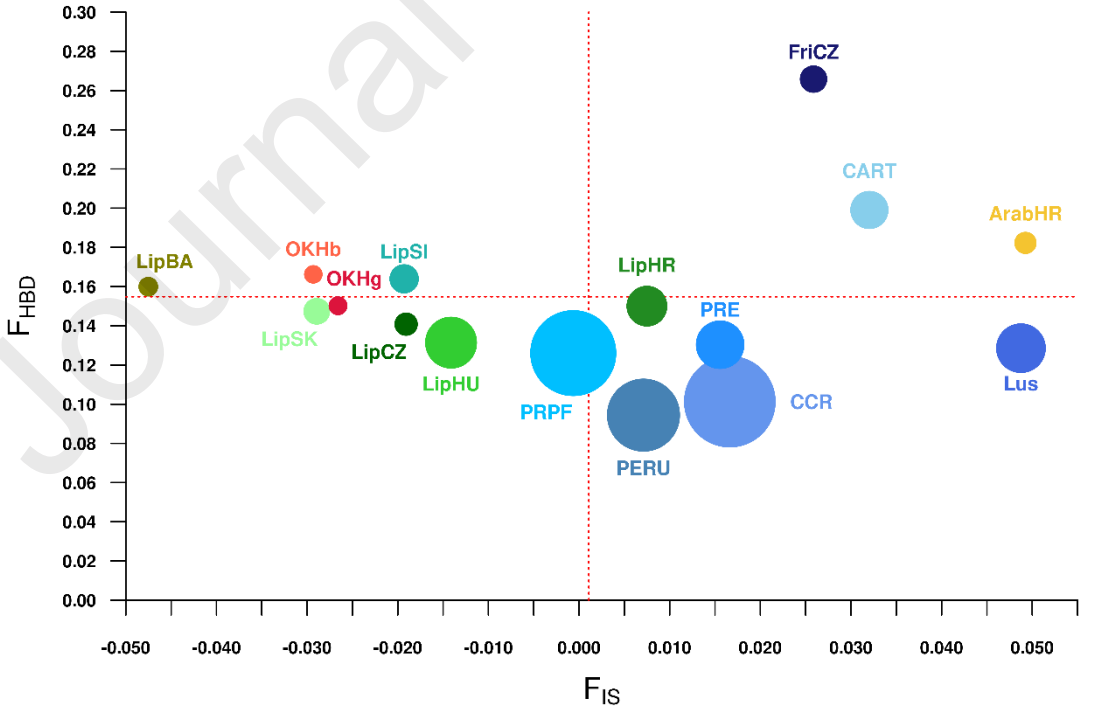


Journal Pre-proofs

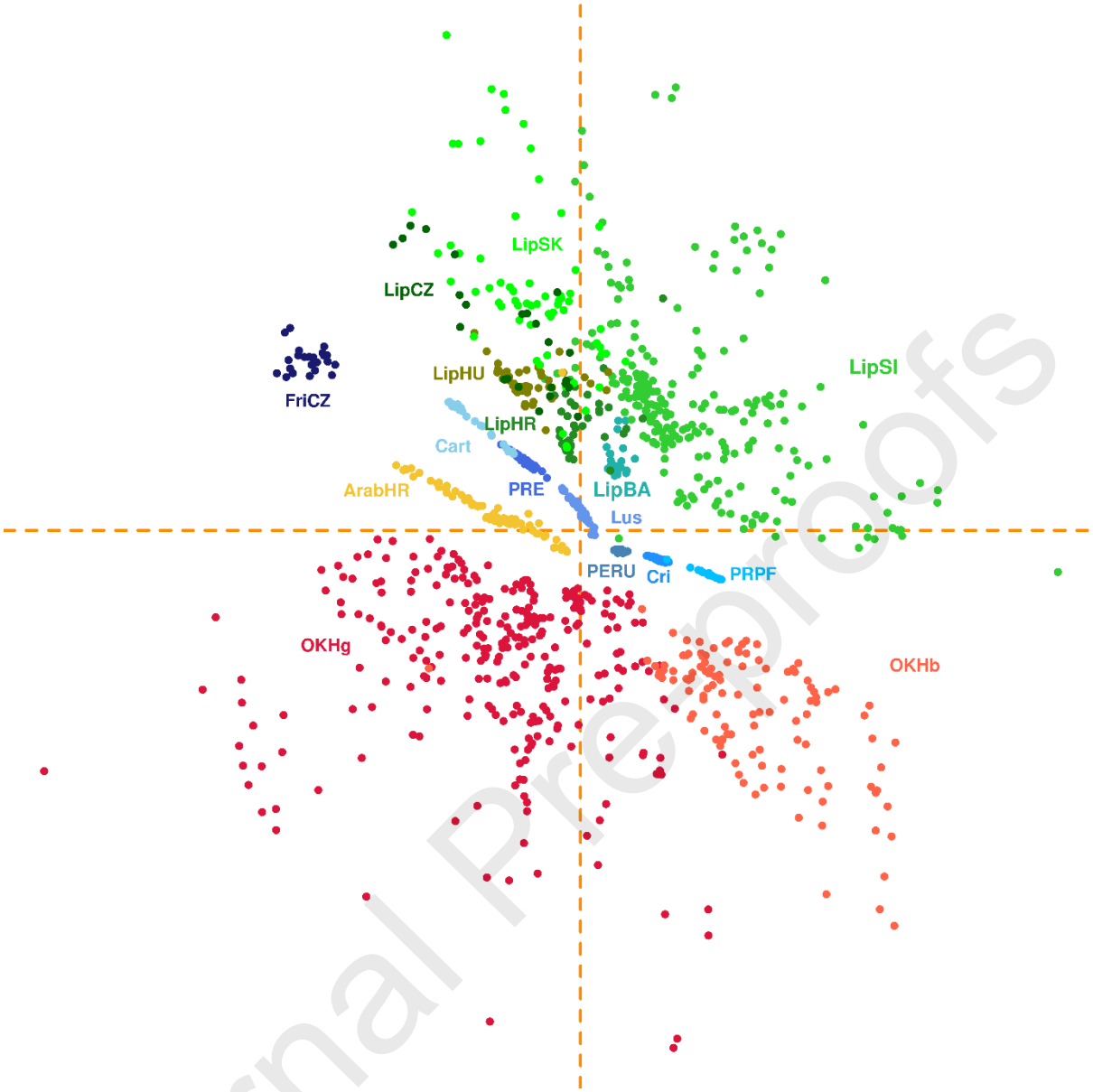


Journal Pre-proofs

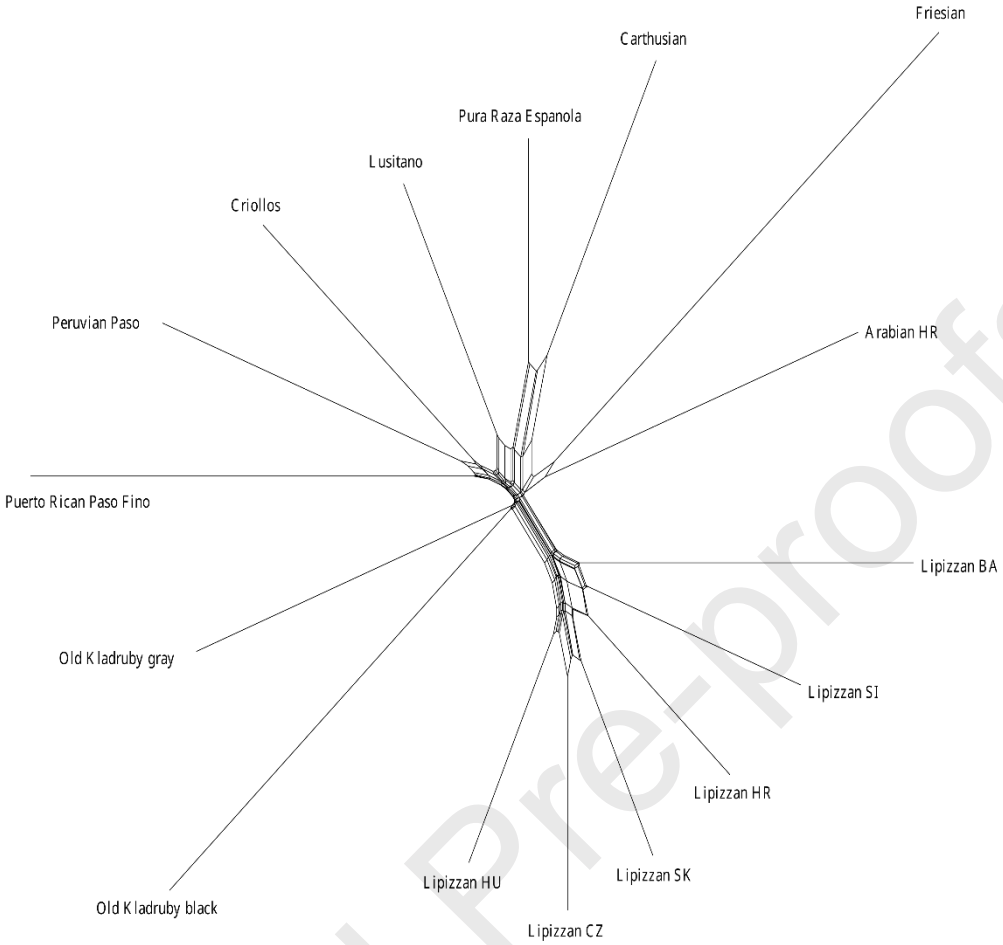




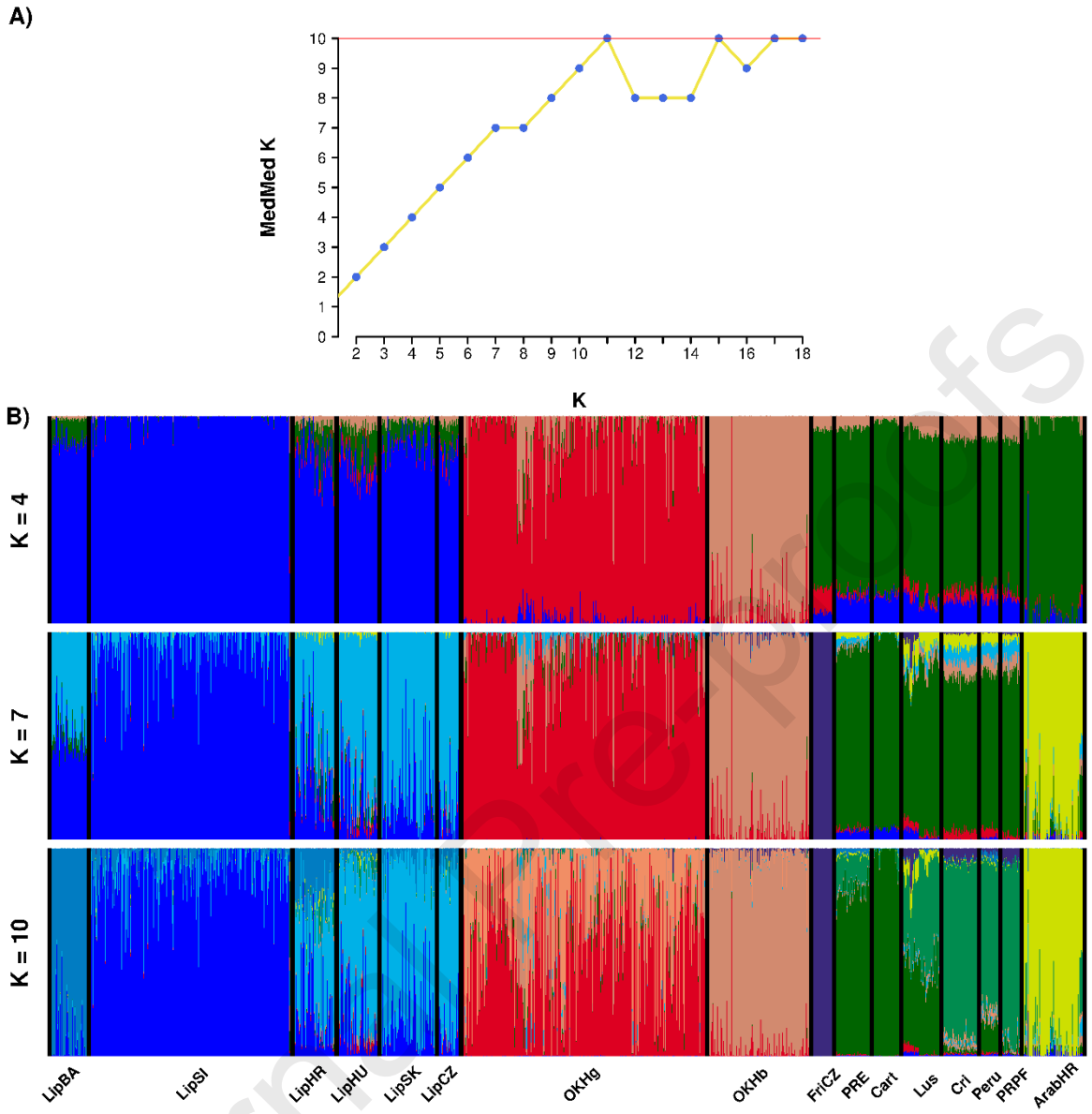
Journal Pre-proofs



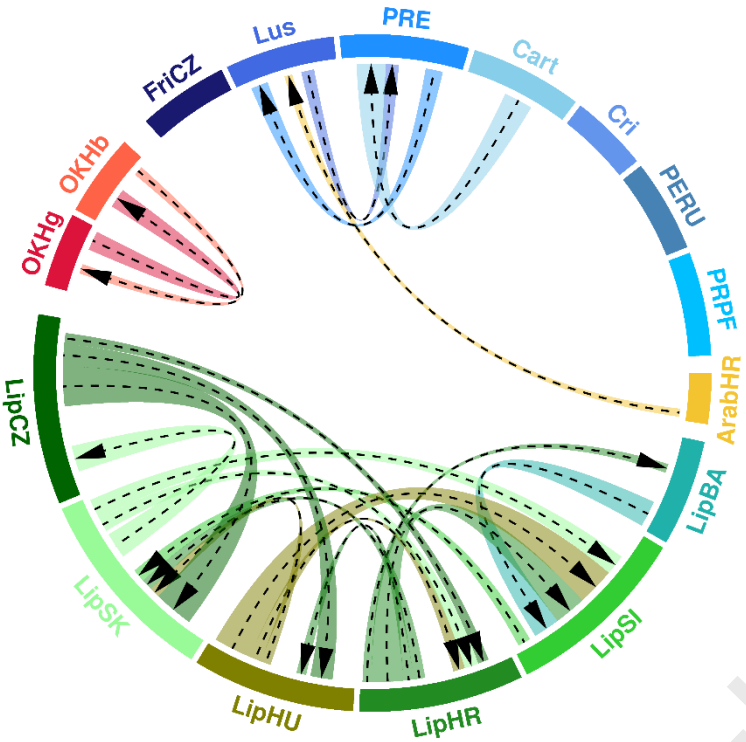
Journal Pre-proofs



Journal Pre-proofs



Journal Pre-proofs



Journal Pre-proofs