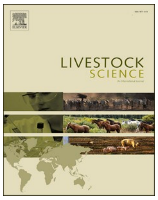
Livestock Science 277 (2023) 105343

Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci



Genomic characterization and population structure of Croatian Arabian horse

Nikola Raguz[a], Nidal Korabi[b,*], Boris Lukić[a,*], Ivana Drzaic[c], Lubos Vostry[d], Nina Moravcikova[e], Ino Curik[c], Radovan Kasarda[e], Vlatka Cubric-Curik[c]

[a] Faculty of Agrobiotechnical Sciences Osijek, J.J. Strossmayer University of Osijek, Vladimira Preloga 1, 31000 Osijek, Croatia
[b] Ministry of Agriculture, Service for Horse Breeding and Domestic Animals Gene Bank, Ilica 101, 10000 Zagreb, Croatia
[c] Department of Animal Science, University of Zagreb Faculty of Agriculture, Svetošimunska cesta 25, 10000 Zagreb, Croatia
[d] Czech University of Life Sciences Prague, Kamýcká129, Prague 165 00, Czech Republic
[e] Faculty of Agrobiology and Food Resources, Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 94976 Nitra, Slovakia

HIGHLIGHTS

- This study analyses Croatian Arabs with genomic data for the first time.
- Croatian Arab horse is genetically close to the other Arab populations.
- Specific genetic background typical for the Balkan environment is expected.
- High ancient and low present inbreeding levels were detected.
- Small effective population size.

ARTICLE INFO

Keywords:
Croatian Arabian horse
Genomic inbreeding
Population structure
Genetic admixture
Effective population size

ABSTRACT

The long history of Arabian horse breeding in Croatia spans several centuries and was strongly influenced by the breeding of Arabian horses in Bosnia and Herzegovina. Thus, the modern breeding of the Croatian Arabian horse is closely connected with the horse breeding of the Borike Stud, founded in 1895 in the mountainous region of Bosnia. Our main goal was to study the specifics of the Croatian Arabian horse (ARABCRO), the influence of the "Borike" breeding and the possible influence of other horse breeds. Therefore, we analyzed the population structure and admixture of the ARABCRO population (62 horses) together with 538 horses of Arabian and other origin (530 publicly available genotypes and 8 newly genotyped Shagya horses from Croatia) using high-throughput genomic data (GeneSeek® Genomic Profiler™ Equine SNP BeadChip 70 K). Our analyses (PCA, Neighbour Network, F_{ST} and STRUCTURE) revealed that ARABCRO is genetically closely related to the other Arabian horse breeds (populations) with an average F_{ST} of 0.09 whereas (the F_{ST} between ARABCRO and the Kladrub, Lipizzaner or Thoroughbred breeds, for example, is over 0.14) and thus can be considered a member of a large "Arabian horse metapopulation" (ARABMETA). At the same time, we have shown that ARABCRO is a distinct population with a specific position within ARABMETA, which can be quantified by genomic analyses. We also estimated the conservation status of ARABCRO, i.e., genomic inbreeding level (F_{ROH}) and effective population size (Ne) given the small census size of the breed (428 horses in 2021). Estimated genomic inbreeding levels were high ($F_{ROH>2Mb} = 0.136$, $F_{ROH>4Mb} = 0.100$), especially those estimated for recent inbreeding ($F_{ROH>8Mb} = 0.071$ and $F_{ROH>16Mb} = 0.039$), indicating intentional mating practices of very close relatives. The estimated effective population size was small, ranging from Ne = 47 to Ne = 67 (95% CI) and was significantly reduced from Ne = 500 (approximately) from 9 to 12 generations back (generation interval about 10 years). The obtained results will be used in further breeding of ARABCRO, especially with regard to the current conservation status, e.g. by controlled mating and "blood refreshment" with other Arabian horses.

---

* Corresponding authors.
E-mail addresses: nidal.korabi@mps.hr (N. Korabi), blukic@fazos.hr (B. Lukić).

https://doi.org/10.1016/j.livsci.2023.105343
Received 11 May 2023; Received in revised form 18 August 2023; Accepted 21 September 2023

1. Introduction

In Croatia, the existence of Arabian horses is known for centuries. Ancient records from more than 600 years ago testify their use in warfare purposes (Imamovic, 1997). Also, there are some evidences of 300 years of continuous breeding at the stud farm in Djakovo from the beginning of the 16th century (Žiga et al., 2008). Its population size varied depending on the given circumstances of a given time and social events. The first systematic breeding of Arabian horses on the territory of ex-Yugoslavia was started at the stud farm Borike in 1895, which unfortunately ended its existence in 2019. The first breeding goal was to improve the Bosnian Mountain horse using a sufficient number of Arabian stallions, but through time there was also a need for the purebred Arabian horses, that should become more adapted to the mountainous Bosnian environment (Telalbašić et al., 2009). The majority of the breeding animals in Borike came from the four different sources. The first source comprised of the Arabian horses imported from Syria, Saudi Arabia, Yemen and Iraq. The second comprised of the animals from Hungary (stud farm in Babolna), third from Poland (stud farm in Slawuta), and fourth from Romania (stud farm in Radautz) (Hrasnica, 1957). From the Bosnian population, Croatian population was formed in the early ’70s of the previous century at the stud farm Višnjica. According to Korabi (2012), the most prominent imported stallions with great influence on the Arabian population in Bosnia and consequently on the Croatian Arabian population were 17 Ilderim from 1915 to 1917 and from 1927 to 1930, 24 Lenkoran from 1926 to 1929, and 19 Shagya X-13 from 1919 to 1929. Nowadays in Croatia, the Arabian-bred horse is bred by the private owners only. First genetic analysis of Arabian-bred dam lines from Višnjica stud were done with mtDNA in thesis of Korabi (2012). Based on the detected haplotypes and sequence analyses, results indicate that the Arabian-bred horses from Croatia have a heterogeneous origin, which was confirmed by the part of the shared nucleotide sequences with other horse breeds, and consequently their shared origin. The founders of this population were the animals from the stud farm Borike, especially two mares from the line El Hafi and Hamdani and one stallion from the line Gazal. During the ‘80s, the private breeders started to buy animals from Višnjica. Due to all these circumstances, a special type of Arabian-bred horse was formed with characteristics adapted to the local environment (typical for the Balkan) and breeding purposes.

Since Arabian-bred horses in Croatia are bred in the same gene pool for more than 40 years, it therefore represents interesting population for identifying specific genetic population structure. Genomic analysis (characterization) of the breed or population can be of great benefit in revealing the conservation status, genetic background, and complex evolutionary history of horse populations around the world, as shown by a number of recent studies (Cosgrove et al., 2020; Grilz-Seger et al., 2019; Kvist et al., 2019; Schurink et al., 2019; Jasielczuk et al., 2020).

Genetic structure, in both species and breeds, represents fundamental basis for adaptation and breeding. It is important to document the relative levels of genetic diversity within and between the breeds/populations, so as to provide useful information for breeding and conservation programs. The availability of low-cost genome-wide SNP data for horses (70k SNPs) in the last decade enables identification of population genomic parameters with high accuracy and also the possibility to compare data across different populations.

Our main objective was genomic characterization of the Croatian Arabian horse (ARABCRO) with emphasis on its relatedness with other Arabian horses, influence of “Borike” breeding and possible influence of other horse breeds using high-throughput genomic data. In our analyses, we also used publicly available genomic information for several Arabian horse populations (Cosgrove et al., 2020) as well as genomic information of other horse breeds sampled in Slovakia (Hucul, Lipizzan and Shagya horses) and in the Czech Republic (Lipizzan, Kladrub and Noriker horses). We also estimated the conservation status of ARABCRO, i.e. genomic inbreeding level (F_{ROH}) and effective population size (Ne) given the small census size of the breed (428 horses in 2021).

2. Material and methods

2.1. Data and analyzed population description

DNA was isolated from the blood of 62 Croatian Arabian horses (ARABCRO) according to the procedure described in Korabi (2012). In contrast, DNA was obtained from hair follicles of 29 Hucul horses from Slovakia (HUCUSVK), 24 Kladruber horses from the Czech Republic (KLADCZE), 28 Lipizzan horses from the Czech Republic and Slovakia (LIPICZE and LIPISVK), 24 Norik horses from the Czech Republic (NORICZE), and 48 Shagya horses from Croatia and Slovakia (SHAGCRO and SHAGSVK). We also included genomic data from different Arabian populations in our analyses published in Cosgrove et al. (2020), namely 77 horses from Egypt (ARABEGY), nine horses from Iran (ARABIRN), 11 horses from Poland (ARABPOL), 12 horses from Saudi Arabia (ARABSAU), and 259 Arabian horses of different origins from Europe and the USA (ARABMLO). In addition, the genotypes of 17 Thoroughbred horses from the United Kingdom (THORGBR) were taken from Cosgrove et al. (2020).

In Fig. 1a world map showing the samples used in this study is represented.

2.2. Genomic analysis, quality control and phasing

Croatian samples were genotyped using the GeneSeek® Genomic Profiler™ Equine SNP BeadChip 70 K (73,860 SNPs). The SNPs from sex chromosomes were excluded from the analysis. Quality control was performed in PLINK (Purcell et al., 2007). 4715 SNP-s were removed due to missing genotype data, 795 SNP-s were removed due to minor allele threshold (–maf 0.05). Only one horse was removed due to missing genotype data (–mind 0.23). Total genotype call rate in remaining samples was 0.87. Overall, the final dataset consisted of 600 animals and 31,390 SNPs. In the estimation of population structure and effective population size we also used phased genotypes, while phasing was performed by using Shapeit2 software (Delaneau et al., 2013).

2.3. Genomic relatedness

The principal component analysis (PCA) was performed using PLINK (Purcell et al., 2007) and ggplot package in R software (R Core Team, 2021).

Estimation of pairwise genetic differentiation between all analyzed populations was based on F_{ST} fixation coefficients (Weir and Cockerham, 1984), averaged over all available SNPs in the genome and calculated using GenePop version 4.7.0 (Rousset, 2008). In addition, we analysed genetic divergence among horse populations based on Nei’s distances (Nei, 1972), which accounts for both mutations and genetic drift, and finally visualized it by the NeighbourJoining tree (NJ) using SplitsTree4 software (Huson and Bryant, 2006). Nei genetic distances were calculated in R using package stAMPP (Pembleton et al., 2013).

2.4. Unsupervised analysis of population structure and admixture

Population genetic structure and admixture (membership in clusters) were inferred using the algorithm implemented in the program STRUCTURE (Pritchard et al., 2000). To reduce the ascertainment bias of the SNP chip our analysis was based on short haplotypes representing multiallelic markers as recommended by Simčič et al. (2015). Thus, after phasing, our genotypes were divided into non-overlapping haplotypes consisting of three consecutive SNPs with an inter-marker distance of less than 50 kb between adjacent SNPs (the maximum length of each haplotype was less than 100 kb). Subsequently, the analysis was performed on the total number of 1513 haplotypes representing 1513 multiallelic markers, which can better describe population differences caused by genetic drift (causing clustering) compared to biallelic SNPs. The analysis was performed with an admixture and correlated allele

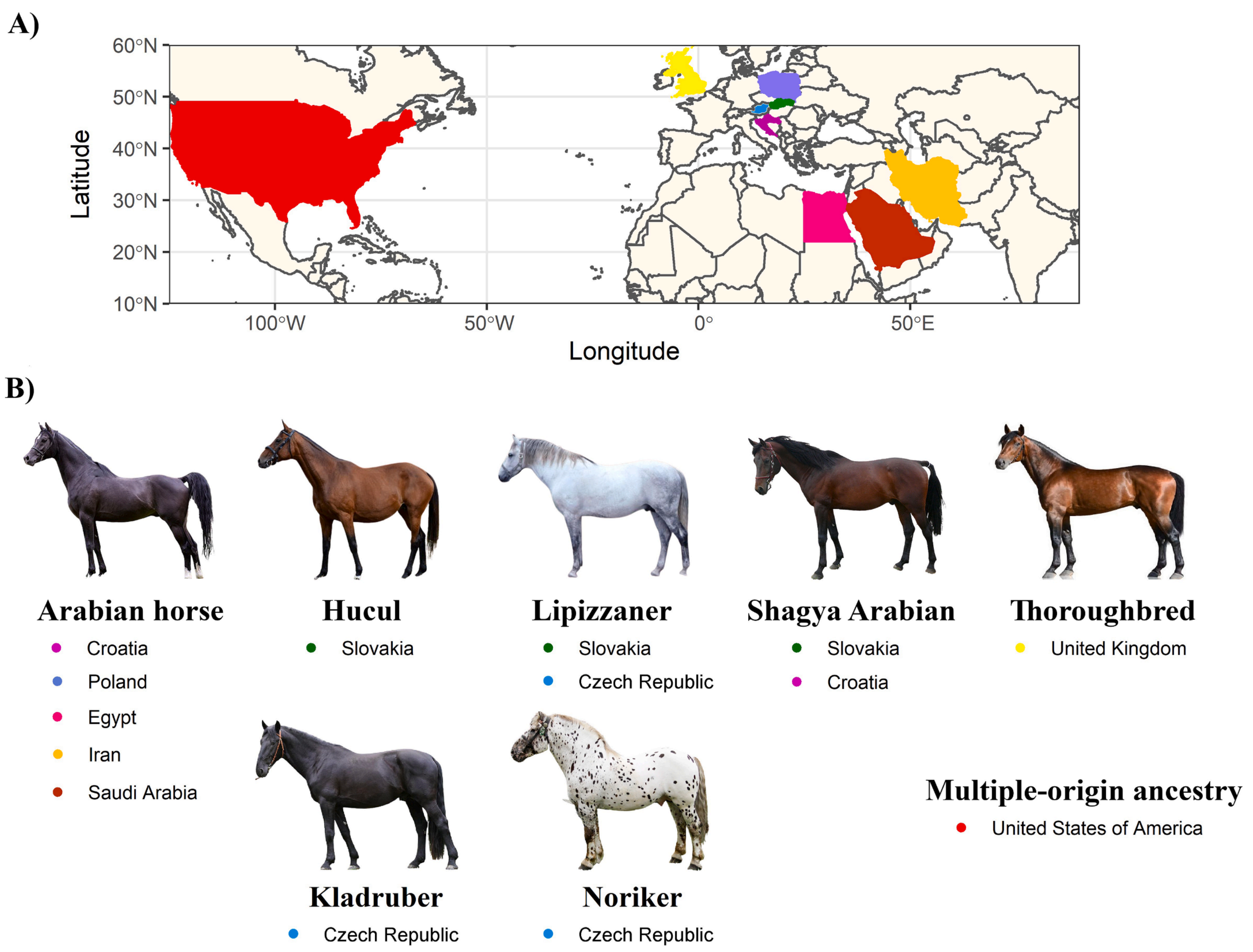


Fig. 1. Map of sampling locations for analysed horses.
The map was created using R software (R Core Team 2021). It illustrates the geographic locations where samples were collected. Each country of origin is presented with a different color.

frequency model, using 10^5 iterations with a burn-in period of 10^4. Runs were repeated 10 times for each assumed K (1–16). The most likely number of clusters was determined using the ΔK method (Evanno et al., 2005) and recommendations implemented in StructureSelector software (Li and Liu, 2018), along with visualization of population structure and admixture.

2.5. ROH based genomic inbreeding

Genomic inbreeding coefficients were calculated with the detectRUNS package in R (Biscarini et al., 2019) using the consecutive ROH (Runs of Homozygosity) approach, which is a window-free approach that directly scans the genome SNP by SNP (some comparisons with other approaches can be found in Curik et al., 2014). According to Ferenčaković et al. (2013), ROH were called when 15 or more consecutive homozygous SNPs were present at a density of at least 1 SNP every 100 kb, with gaps of no more than 1000 kb between them. The parameters for the genomic runs were set as follows: maximum gap length was 100 kb, minimum run length was 250 kb, maximum number of opposite genotypes in the run was 0, maximum number of missing genotypes in the run was 0. The genomic ROH-based inbreeding coefficients for each individual (F_{ROH}) were calculated for each individual by dividing the sum of the length of all ROH considered autozygous (depending on ROH length >2 Mb, >4 Mb, >8 Mb, and >16 Mb) by the total length of the autosomal genome covered by SNPs, as described by McQuillan et al. (2008). In this way, we were able to calculate a range of inbreeding coefficients for each individual ($F_{ROH>2Mb}$, $F_{ROH>4Mb}$, $F_{ROH>8Mb}$, and $F_{ROH>16Mb}$), depending on the distance to the expected common ancestor contributing to the observed "autozygosity" (ROH length), as described in Ferenčaković et al. (2013).

2.6. Current and historical gametic/linkage effective population size

Estimation of current and historical effective population size (Ne_{LD}) was based on the observed gamete and linkage disequilibrium spectrum (LD) and their relationship to effective population size (Hill, 1981; Sved, 1971; Hayes et al., 2003). Here, Ne_{LD} was considered to be the size of an idealized Wright-Fisher population that exhibits the same degree of change in gamete and linkage disequilibrium due to genetic drift as the population under consideration. Estimates were obtained using the software GONE (Santiago et al., 2020), which uses genetic algorithms (Mitchell, 1998) to derive the historical series of effective population size by minimizing the squared differences between the observed averaged squared correlations between two loci allele frequencies, weighted by their variance (D^2) of bins, and those predicted at corresponding different demographic trajectories (Santiago et al., 2020; Saura et al., 2021). Although the software GONE has only recently been developed, it is already widely used to estimate current and historical effective population size in managed (Saura et al., 2021), livestock (Magnier et al., 2022; Drzaic et al., 2022; Vostry et al., 2023), wild (Pacheco et al., 2022), and experimental (Novo et al., 2023) populations with complex demographic histories. GONE was performed using standard input parameters for phased and unphased genomic datasets, in both cases with three MAF constraints (none, <0.01, <0.05, and <0.10) and with approximated recombination rate, where 1 cM equals 1 Mb.

3. Results

3.1. Genomic relatedness

Genetic relatedness of the breeds (populations) studied (600 horses), represented by the first two principal components explaining 36.7% of the total variation, is shown in Fig. 2. The smallest cluster, consisting only of THORGBR horses, was clearly separated from the non-Arabian and Arabian clusters, while at the same time high genetic diversity was observed within the THORGBR cluster, especially along PC2.

Based on the first two principal components, the non-Arabian cluster included five populations (LIPISVK, LIPICZE, KLADCZE, NORICZE and HUCUSVK), although the NORICZE horses could ultimately be considered a separate cluster that did not overlap with other nearby populations. As expected, the cluster of six populations of Arabian origin (ARABCRO, ARABEGY, ARABIRN, ARABPOL, ARABSAU, and ARABMLO) and two populations of Shagya horses (SHAGCRO and SHAGSVK) was the largest cluster and formed a large metapopulation of Arabian horses (ARABMETA). Arabian horses from Egypt (ARABEGY) were the most distant group within this cluster and overlapped only with the ARABMLO horses. Other Arabian horse populations were much more closely intertwined, with the exception of the ARABPOL horses, which exhibited a reasonable degree of estrangement. The ARABCRO horses were quite widely dispersed, especially along PC2 and overlapping with the ARABMLO horses. The proximity to horses from the Near East (ARABIRN and ARABSAU) and to Shagya horses from Croatia and Slovakia proves that there were close historical links between them during population formation. The highest genetic diversity was found in the ARABMLO horses, which was to be expected given their diverse origins.

The estimated overall mean F_{ST} was 0.10, whereas pairwise values ranged between 0.02 (between LIPICZE and LIPISVK), as these two populations belong to a well-defined Lipizzan breed known for its metapopulation structure (Achmann et al., 2004), and 0.22 (between THORGBR and ARABEGY) (Table 1). In general, greater differentiation (larger F_{ST} differences) was observed between ARABMETA subpopulations and non-Arabian breeds (populations) and between THORGBR and all other breeds (populations). Estimated F_{ST} values between ARABMETA subpopulations ranged from 0.02 to 0.17, with the highest values between ARABSAU and other Arabian horses, while less genetic differentiation was observed between Arabian populations with a closer genetic history. With an average F_{ST} of 0.09, ARABCRO was genetically closely related to the other Arabian horse breeds (populations), while the F_{ST} between ARABCRO and, for example, the

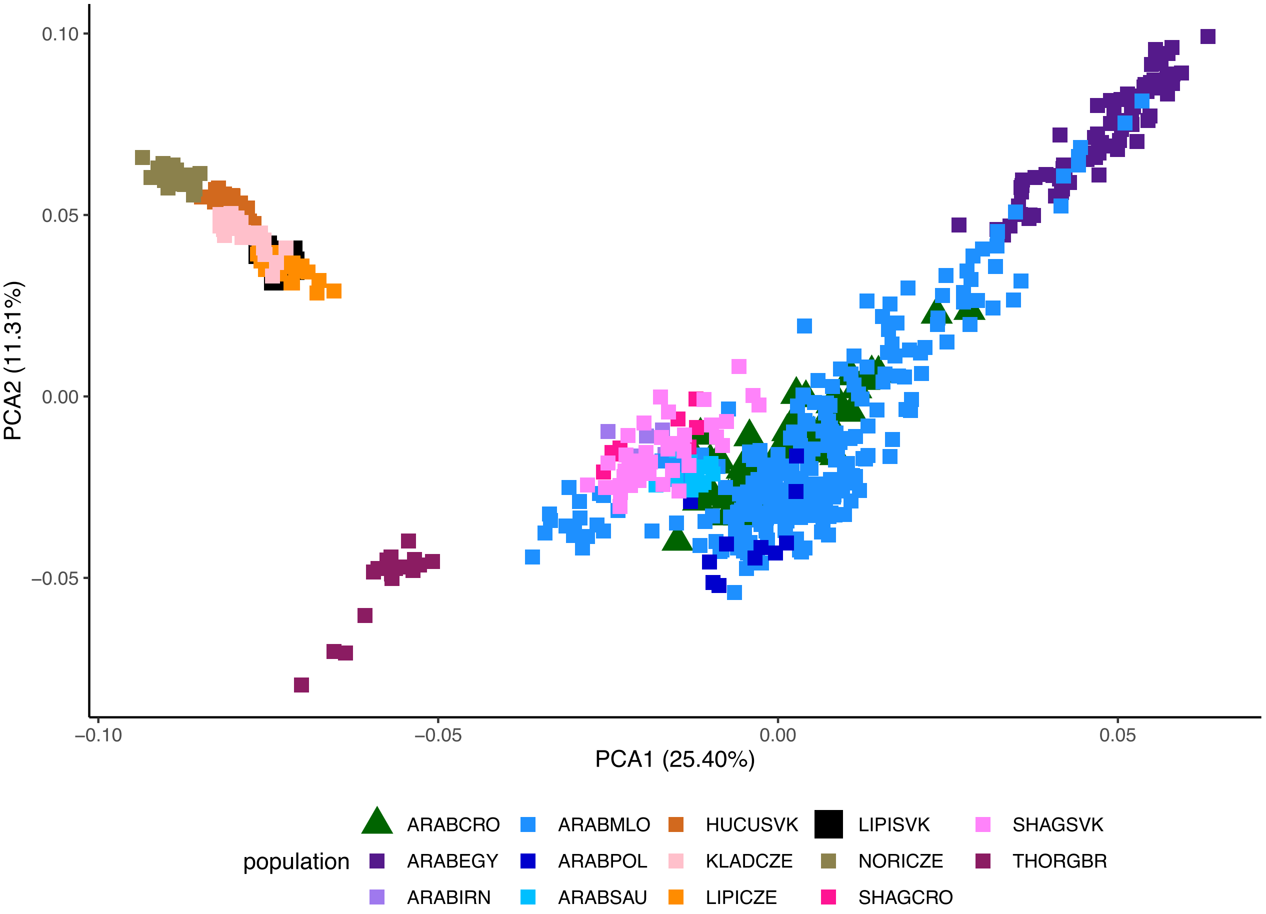


Fig. 2. PCA plot for analysed horse breeds.
The principal component analysis (PCA) was performed using PLINK (Purcell et al., 2007) and ggplot package in R software (R Core Team 2021). The plot is presented with the three main clusters according to PC1 and PC2. Each breed is presented with different color and different letter coding.

Table 1
Population differentiation among 14 horse breeds (populations) based on genome-wide F_{ST} estimates.

Breed/population	AC	AE	AI	AM	AP	AS	HS	LS	LC	NC	CC	SC	SS	MF_{ST}
ARABCRO														**0.12**
ARABEGY	0.11													**0.16**
ARABIRN	0.08	0.13												**0.10**
ARABMLO	0.05	0.06	0.04											**0.08**
ARABPOL	0.08	0.13	0.07	0.02										**0.11**
ARABSAU	0.12	0.17	0.10	0.08	0.12									**0.15**
HUCUSVK	0.15	0.21	0.13	0.12	0.15	0.18								**0.14**
LIPISVK	0.14	0.21	0.12	0.11	0.14	0.18	0.13							**0.13**
LIPICZE	0.14	0.20	0.11	0.10	0.13	0.16	0.12	**0.02**						**0.12**
NORICZE.	0.16	0.22	0.14	0.13	0.16	0.20	0.10	0.14	0.12					**0.15**
KLADCZE	0.15	0.20	0.13	0.11	0.14	0.18	0.12	0.11	0.09	0.12				**0.14**
SHAGCRO	0.07	0.12	0.05	0.03	0.05	0.10	0.12	0.11	0.11	0.13	0.11			**0.09**
SHAGSVK	0.08	0.13	0.07	0.05	0.07	0.12	0.14	0.13	0.12	0.14	0.13	0.04		**0.11**
THORGBR	0.16	**0.22**	0.14	0.13	0.15	0.19	0.19	0.17	0.18	0.20	0.18	0.13	0.16	**0.17**

MF_{ST} is the mean of pairwise comparisons of each breed (population) with all other horse breeds (populations).

Kladrub, Lipizzaner or Thoroughbred breeds was above 0.14.

Genetic differentiation among 14 breeds (populations) was represented using a rooted neighbour-joining tree based on genetic distances defined by Nei (1972). The genetic uniqueness of ARABCRO was further supported by its separation from the other breeds studied (Fig. 3). ARABEGY was the most distant population, while ARABCRO was very close to ARABMLO and beyond to other populations of Arabian horses. A separate group of horses, including Lipizzaners, Kladrubers, Norikers, and Huculs, was on the opposite side of the tree and formed their own cluster according to the results of PCA. At the same time, THORGBR was confirmed as an outgroup.

3.2. Population structure and admixture

Population structure and admixture were estimated using the algorithm implemented in the software STRUCTURE, which can reveal "hidden structures" without determining a priori the membership of individual breed or populations (clusters). According to the

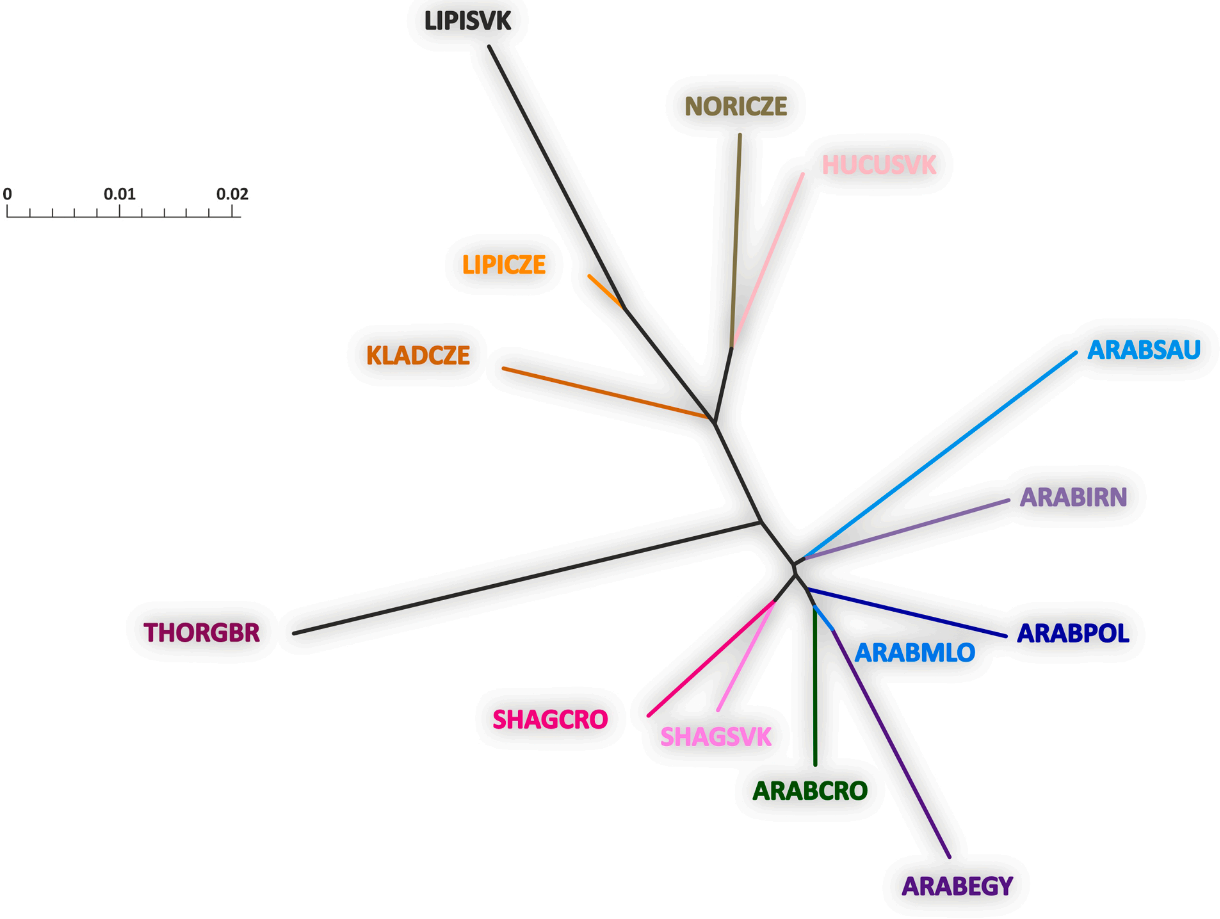


Fig. 3. Neighbour network based on the Nei genetic distances illustrating genetic relatedness of 14 horse breeds.
The analysed genetic divergence among horse populations based on Nei's distances (Nei, 1972) is shown, which accounts for both mutations and genetic drift, and finally visualized by the NeighbourJoining tree (NJ) using SplitsTree4 software (Huson and Bryant, 2006). Nei genetic distances were calculated in R using package stAMPP (Pembleton et al., 2013). Each breed is presented with different color and different letter coding.

recommendations (Prichard et al. 2000; Falush et al., 2007), the most likely K is the one where Ln Pr(G|K) is maximized or where the smallest value of K captures the main structure in the data. In our analyzes, this difference was very small up to $K = 11$ (Supplementary Figure S1), making it difficult to decide which is the optimal number of clusters in the data set (the most likely parameter K) (Supplementary Figure S2). Furthermore, we analyzed the rate of change of Ln Pr(G|K) between successive K values. In the presence of a hierarchical population structure, this approach could reveal a more subtle structure within populations, as suggested by Evanno et al. (2005). In our data set (see Supplementary Figure S3), several peaks were observed at $K = 2$, $K = 10$, and $K = 13$, indicating the presence of a complex hierarchical structure. Therefore, we decided to show in Fig. 4 only the results of $K = 2$, $K = 4$ (the first subdivision in which ARABCRO appears as a separate cluster), and $K = 10$, since K reveals the best sub-structuring in the dataset. Thus, at $K = 2$, the horses studied were divided into two groups, one more specific to Arabian horses, although showing purity only in ARABEGY, and the other characteristic of non-Arabian horses such as Huculs, Kladruber, Lipizzan, Noriker, and Thoroughbreds. At $K = 4$, ARABCRO was clustered as a distinct population, indicating its specific position within ARABMETA, previously demonstrated by other analyses (PCA, pairwise F_{ST}, and Neighbour Network). Finally, the most likely genomic population structure and admixture of the studied horses was shown at $K = 10$ (see Fig. 4). As expected, the greatest admixture was observed in ARABMLO, while some clustering was also observed in ARABEGY and ARABPOL. Of course, common clustering was observed between SHAGCRO and SHAGSVK (with higher estimated admixture in SHAGCRO) and between LIPICZE, LIPISVK, and KLADCZE, confirming their common breeding history. Some common "cluster components" were also observed between HUCUSVK and NORICZE, which was also expected.

3.3. Inbreeding level and effective population size

The distributions of ROH-based genomic inbreeding coefficients as a function of ROH length ($F_{ROH>2Mb}$, $F_{ROH>4Mb}$, $F_{ROH>8Mb}$, and $F_{ROH>16Mb}$) in ARABCRO are shown in Fig. 5. The average inbreeding observed in ARABCRO was 0.136 for $F_{ROH>2Mb}$ and 0.100 for $F_{ROH>4Mb}$. The observed inbreeding value for $F_{ROH>2Mb}$ was higher than the estimated inbreeding value (0.101) in the Polish Arabian population (Szmatola et al., 2022), but lower than the estimated (0.225) in the Persian Arabian population (Sadeghi et al., 2019). However, the observed genomic inbreeding is considered high because much lower ROH inbreeding levels have been estimated in other breeds. For example, the estimated ROH inbreeding levels in five non-Arabian horse breeds (Hucul, Malopolski, Polish Konik, Sokolski, and Szumski) presented in Szmatola et al. (2022) ranged from 0.053 to 0.084 for $F_{ROH>2Mb}$ and from 0.033 to 0.081 for $F_{ROH>4Mb}$. Extremely high recent genomic inbreeding ($F_{ROH>8Mb} = 0.071$ and $F_{ROH>16Mb} = 0.039$) was observed in ARABCRO, indicating intentional mating of close relatives. For example, Szmatola et al. (2022) observed slightly lower ROH inbreeding in Polish Konik ($F_{ROH>8Mb} = 0.063$ and $F_{ROH>16Mb} = 0.035$), whereas close inbreeding in other horses ranged from 0.018 to 0.054 for $F_{ROH>8Mb}$ and from 0.007 to 0.028 for $F_{ROH>16Mb}$. In addition, a large percentage of ARABCRO horses (23.6%) had a $F_{ROH>16Mb}$ value greater than 0.0613, which is likely to occur in offspring of half siblings. The estimated average inbreeding coefficient in the ARABCRO pedigree was 0.087 (with complete generation equivalent of 11.97), while the estimated average recent inbreeding coefficient (with a pedigree limited to five recent generations) was also very high (0.064), indicating frequent intentional matings of close relatives (Korabi 2012).

The estimated current effective population size (Ne_{LD}) varied between 44 and 56, depending on the data analyzed (unphased versus phased) and the MAF parameters used in the software GONE (none, <0.01, <0.05, and <0.10). Although the differences were not significant (overlap at 95% CI), higher estimates were obtained in the analyses with phased genotypes (between 52 and 56) than with unphased genotypes (between 44 and 46). In both cases, although the difference in estimated Ne_{LD} was very small (2 and 4 individuals in the “unphased” and "phased” estimates, respectively), the estimated effective population size increased proportionally with the increase in MAF (from none to 0.10). The obtained estimates were slightly higher but in agreement with

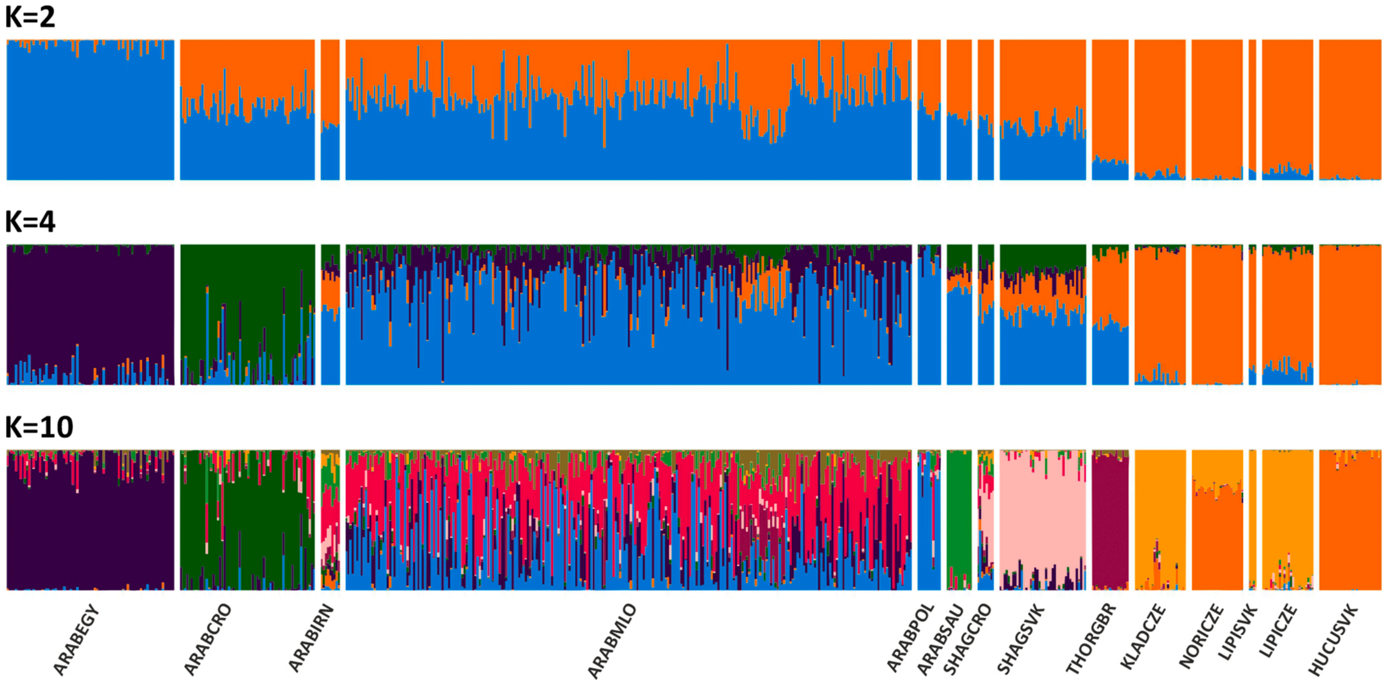


Fig. 4. Population structure and admixture estimated using the algorithm implemented in the software STRUCTURE.
Graphical representation of unsupervised population structure and admixture using the STRUCTURE algorithm for 600 horses representing 14 a priory defined breeds (populations). Each horse is represented by a single vertical line broken into K colored segments whose length is proportional to the estimated membership of the inferred cluster, illustrating the presence of admixture.

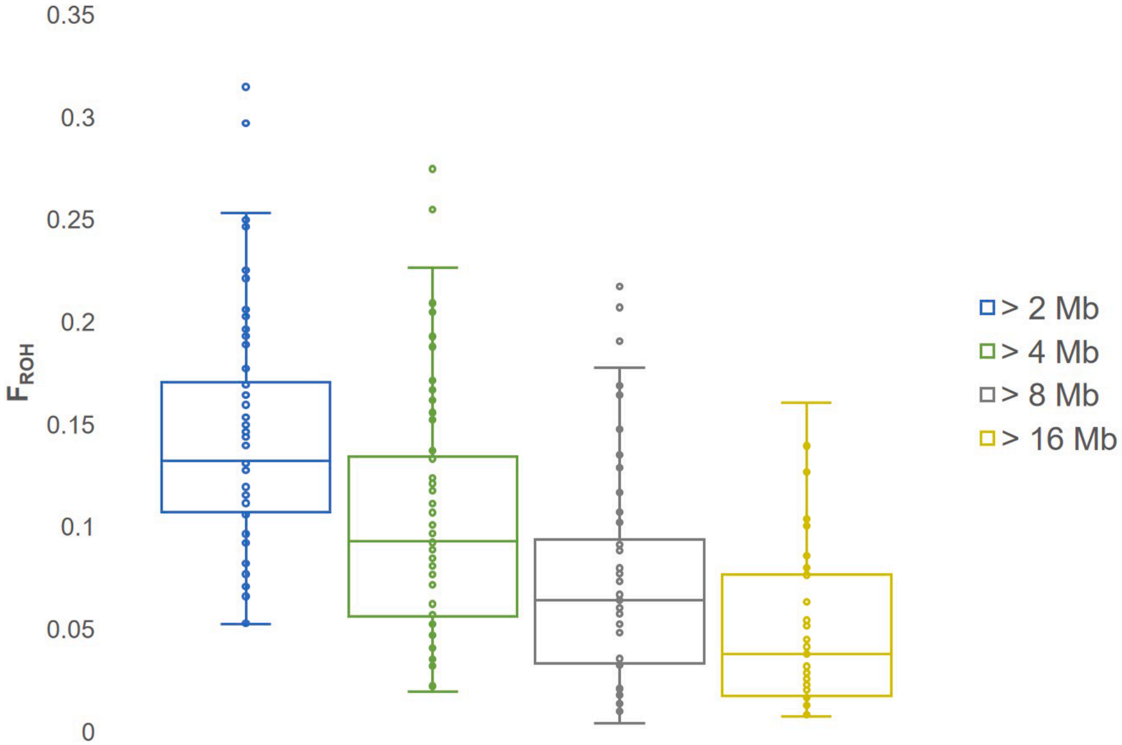


Fig. 5. Box-plot illustration presenting distributions of the ROH based genomic inbreeding coefficients (F_{ROH}>2Mb, F_{ROH}>4Mb, F_{ROH}>8Mb, and F_{ROH}>16Mb) in Croatian Arab horses.
Genomic inbreeding was analysed using ROHs (Runs of Homozygosity), and it was performed using detectRUNS package in R (Biscarini et al. 2019) using consecutive approach, a window-free approach which directly scans the genome SNP by SNP. F_{ROH} classes were specified according to the ROH length: >2 Mb, >4 Mb, >8 Mb and >16 Mb. The box plots represent the mean values and standard deviations of the respective F_{ROH} class.

effective population size (60) of ARABCRO estimated from the informative pedigree in Korabi (2012) using approach described in Gutiérrez et al. (2008, 2009).

The Ne_{LD} estimated for ARABCRO was small (about 50), so we were also interested in historical estimates of effective population size, because it is not the same whether the small Ne_{LD} is "stable" or a consequence of a constant or sudden decline. Here, we estimated the trend in effective population size (Ne_{LD}) over the last 40 generations using different options ("phased" versus "unphased" genotypes and different MAF adjustments).

As shown in Fig. 6, the small Ne_{LD} of ARABCRO was stable over the last seven to nine generations, while the tenfold decrease in Ne_{LD} from 500 to 50 (approximations) was observed nine to 12 generations back.

The use of phased or unphased genotypes affected the historical estimates of effective population size, as the lag of the estimated trend shifted back several generations when estimates were calculated using the phased genotypes, whereas negligible differences were observed for trends based on different MAF options (Fig. 6). Considering that the generation interval in ARABCRO is about 10 years (10.25 according to the calculation in Korabi, 2012), the sudden decline occurred 90 to 120 years ago or between 1903 and 1933, which is consistent with the exponential global use of mechanization. The low Ne_{LD}, ranging from 32 to 62, was observed in several draft horse populations (Druml et al., 2007), which was explained by the significant decline in the census population size of draft horses in the geographic area of the former Austro-Hungarian Empire and Germany from four million draft horses

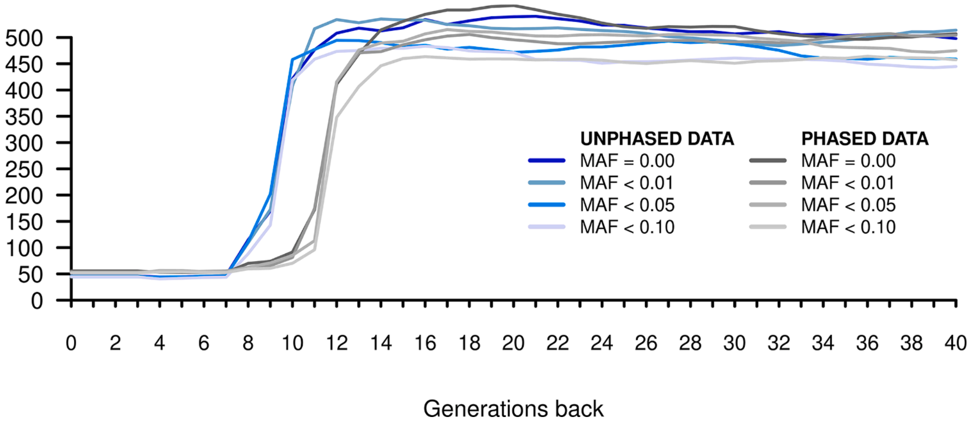


Fig. 6. Effective population size (Ne_{LD}) using phased and unphased data and MAF exclusions
Trend representing estimates of the effective population size (Ne_{LD}) of the Croatian Arabian horse over the last 40 generations with different data inputs (phased and unphased genotypes) and MAF exclusions (none, <0.01, <0.05, and <0.10).

and crosses in 1876 to 20,000 in 2007.

4. Discussion

In this study, our aim was to analyze population structure, admixture and genomic characterization ("positioning") of ARABCRO horses. It is generally small population and highly underestimated by the breeders, because during the last centuries in Croatia (tradition related with Austrian-Hungarian breeding practices) main focus was given to the Lipizzaner breed. Nowadays, it is used mainly by the recreational ridders and professionals in endurance and gallop racing sport.

As we described, our knowledge and the historical record of this breed led us to other conclusions and raised the question of why the observed results didn't meet our expectations (the presence of the "background blood" of other breeds). Historically, Arabs had a specific role in the Balkans as a famous and cosmopolitan breed with clear and distinctive purposes, therefore in the breeding context they were mainly used within warmblood breeding to improve local horse breeds i.e. Bosnian mountain horse, to become lighter horse with better riding ability traits, and for the sport purposes (gallop racing and endurance). However, Pure Arab horses haven't completely adjusted to the breeding preferences in this area (as Borike are situated at the higher altitudes), so the breeders decided to introduce a half-bred Arab mares, which became more suitable in the present environment, with more reliable and stronger characteristics (Hrasnica, 1957). Shortly after these half-bred Arabians were bred in Stud Farm Borike, occasionally the Pure Arabian stallions were also used. Because of these facts, we expected that the ARABCRO will be genetically more distant than the world-wide Arabs we analysed here. It appears that the breeding was much more controlled than our expectations, and that the pure Arabs were dominated in breeding. These practices were detected in our analysis of principal components, where the ARABCRO were situated within the cluster of other i.e. in a ARABMETA ("Arab horse metapopulation"), This finding additionally confirms the results from Cosgrove et al. (2020), which showed that the substantial variation is present in the worldwide Arab population, mainly as a result of closely controlled breeding within countries over the past 200 years. In our scenario on Croatian population, this horse represents subgroup with specific breeding purposes typical for the Balkan environment, but it is still the Arabian-bred horse, a member of ARABMETA. Admixture analysis also confirmed these findings. In the ADMIXTURE plots on cluster assignments for $K = 4$ and higher, ARABCRO clustered as a separate subgroup and shared ancestry with other Arabian horses, and to a lesser extent with Shagya.

Genomic inbreeding analysis provided additional insight into the conservation status of ARABCRO. The advantage of genomic inbreeding calculations using ROHs is the possibility to draw conclusions about breeding history (recent or distant inbreeding), as recent inbreeding is associated with the long ROH segments, while ancient inbreeding is associated with short ROH segments. Inbreeding estimated with ROHs on a genome-wide level proved to be high (ARABCRO was 0.136 for $F_{ROH>2Mb}$ and 0.100 for $F_{ROH>4Mb}$), which is common for Arabian horses (Supplementary Figure S4), but much higher than in other horse breeds, for example, draft horses. Unfortunately, extremely high recent inbreeding has been observed in ARABCRO ($F_{ROH>8Mb} = 0.071$ and $F_{ROH>16Mb} = 0.039$), which could lead to the expression of deleterious mutations. The most realistic explanation for the observed high inbreeding is the mating of closely related animals to achieve desirable phenotypes in the offspring. This practice should be discontinued and systematic planning of mating with a view to avoiding inbreeding should be initiated.

To gain further insight into the demographic history of ARABCRO, we estimated current and historical effective population size (Ne) using the pattern of gametic/linkage disequilibrium (LD). As shown in Fig. 6, the Ne for the last eight generations was approximately 50 animals, which is alarming and indicates the need for a breeding conservation strategy. A dramatic decline in effective population size occurred about nine to 12 generations ago, from 500 to 50 animals, which historically corresponds to the first half of the 20th century, given a generation interval of 10 years. At that time, horse breeding in the Balkans and throughout Europe suffered a sharp decline in population sizes, probably due to the events of the war and the introduction of agricultural mechanization.

Additionally, the level of genetic differentiation based on Wright's statistics (Wright, 1978) between populations were estimated, as the genetically similar populations tend to have lower F_{ST} values, whereas more genetically diverse have higher F_{ST} values. Pairwise F_{ST} values (Table 1) as a reflection of allele frequencies between analysed horse populations, shows highest population similarity between the ARABCRO and ARABMLO (0.05). On the contrary, the most divergent populations indicated with the highest F_{ST} values were the ARABCRO and THORGBR, which is in accordance to the PCA and STRUCTURE analysis.

5. Conclusion

With respect to the geographic origin and historical background of ARABCRO, this study provides the first insights into the genomic characterization and conservation status of ARABCRO. PCA, STRUCTURE, F_{ST} and Neighbour network analyses yielded results characterizing ARABCRO as a specific population within the Arabian horse metapopulation (ARABMETA), as demonstrated in this study. Genomic inbreeding analysis revealed higher ancient inbreeding characteristic of purebred Arabian horses, but also extremely high recent inbreeding that needs to be controlled. The current effective population size was low (about 50) but stable over the last eight generations, whereas a significant decrease in effective population size was observed nine to 12 generations ago. Thus, our results suggest that the revised breeding strategy should be used judiciously in terms of controlled matings and increasing genetic diversity. In this sense, one possibility is to open the herdbook to other Arabian horses.

CRediT authorship contribution statement

This study was conceived by Raguz, N., Korabi, N., Cubric-Curik, V., Lukic, B. and Curik, I. **Raguz, N.:** Investigation, Software, Visualization, Writing. **Korabi, N.:** Conceptualization, Validation. **Lukic, B.:** Software, Methodology, Validation, Writing. **Drzaic, I.:** Software, Visualization, Molecular analysis. **Vostry, L.:** Investigation, Software, Visualization. **Moravcikova, N.:** Samples providing, Molecular analysis. **Curik, I.:** Conceptualization, Methodology, Writing, Supervision. **Kasarda, R.:** Samples providing, Molecular analysis. **Cubric-Curik, V.:** Conceptualization, Samples providing, Molecular analysis, Supervision. All authors read and approved the manuscript.

Declaration of Competing Interest

None.

Acknowledgements

This study was supported by Croatian Science Foundation project IP-2018-01-8708 Application of NGS methods in the assessment of genomic variability in ruminants – ANAGRAMS; KK.01.1.1.06.0002 – LazAg and by the Ministry of Agriculture of the Czech Republic, Project No. QK1910156.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.livsci.2023.105343.

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. Anim. Genet. 35 (4), 285–292. https://doi.org/10.1111/j.1365-2052.2004.01157.x.

Biscarini, F., Cozzi, P., Gaspa, G., Marras. G., 2019. Detect runs of homozygosity and runs of heterozygosity in diploid genomes. R package version 0.9.5. Accessed May 10, 2019. https://cran.r-project.org/web/packages/detectRUNS/.

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. Sci. Rep. 10 (1), 9702. https://doi.org/10.1038/s41598-020-66232-1.

Curik, I., Ferenčaković, M., Sölkner, J., 2014. Inbreeding and runs of homozygosity: a possible solution to an old problem. Livest. Sci. 166, 26–34. https://doi.org/10.1016/j.livsci.2014.05.034.

Delaneau, O., Zagury, J.F., Marchini, J., 2013. Improved whole-chromosome phasing for disease and population genetic studies. Nat. Methods 10, 5–6. https://doi.org/10.1038/nmeth.2307.

Druml, T., Curik, I., Baumung, R., Aberle, K., Distl, O., Sölkner, J., 2007. Individual-based assessment of population structure and admixture in Austrian, Croatian and German draught horses. Heredity (Edinb) 98 (2), 114–122. https://doi.org/10.1038/sj.hdy.6800910.

Drzaic, I., Curik, I., Lukic, B., Shihabi, M., Li, M.H., Kantanen, J., Mastrangelo, S., Ciani, E., Lenstra, J.A., Cubric-Curik, V, 2022. High-density genomic characterization of native Croatian sheep breeds. Front. Genet. 13, 940736 https://doi.org/10.3389/fgene.2022.940736.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol. Ecol. 14, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x.

Falush, D., Stephens, M., Pritchard, J.K., 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol. Ecol. Notes. 7 (4), 574–578. https://doi.org/10.1111/j.1471-8286.2007.01758.x.

Ferenčaković, M., Sölkner, J., Curik, I., 2013. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. Genet. Sel. Evol. 45, 42. https://doi.org/10.1186/1297-9686-45-42.

Grilz-Seger, G., Druml, T., Neuditschko, M., Mesarič, M., Cotman, M., Brem, G., 2019. Analysis of ROH patterns in the Noriker horse breed reveals signatures of selection for coat color and body size. Anim. Genet. 50, 334–346. https://doi.org/10.1111/age.12797.

Gutiérrez, J.P., Cervantes, I., Goyache, F., 2009. Improving the estimation of realized effective population sizes in farm animals. J. Anim. Breed. Genet. 126 (4), 327–332. https://doi.org/10.1111/j.1439-0388.2009.00810.x.

Gutiérrez, J.P., Cervantes, I., Molina, A., Valera, M., Goyache, F., 2008. Individual increase in inbreeding allows estimating effective sizes from pedigrees. Genet. Sel. Evol. 40 (4), 359–378. https://doi.org/10.1186/1297-9686-40-4-359.

Hayes, B.J., Visscher, P.M., McPartlan, H.C., Goddard, M.E., 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Res. 13 (4), 635–643. https://doi.org/10.1101/gr.387103.

Hill, W.G., 1981. Estimation of effective population size from data on linkage disequilibrium. Genet. Res., Camb. 38, 209–216. https://doi.org/10.1017/S0016672300020553.

Hrasnica, F., 1957. Uzgojna analiza arapskog rasplodnog materijala ergele Borike. Veterinaria 1–14, 2-3.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267. https://doi.org/10.1093/molbev/msj030.

Imamović, M., 1997. Historija Bošnjaka, BZK „Preporod. Sarajevo.

Jasielczuk, I., Gurgula, A., Szmatoła, T., A, Semik-Gurgul, E., Pawlina-Tyszko, K., Stefaniuk-Szmukier, M., Polak, G., Tomczyk-Wrona, I., Bugno-Poniewierska, M., 2020. Linkage disequilibrium, haplotype blocks and historical effective population size in Arabian horses and selected Polish native horse breeds. Liv. Sci. 239, 104095 https://doi.org/10.1016/j.livsci.2020.104095.

Korabi, N., 2012. Genetic and Morphological Analysis of Traditional Arab and Purebred Arab and Shagya horses in Croatia. Faculty of Agriculture Zagreb, University of, Zagreb, Croatia. PhD Thesis.

Kvist, L., Niskanen, M., Mannermaa, K., Wutke, S., Aspi, J., 2019. Genetic variability and history of a native Finnish horse breed. Genet. Sel. Evol. 51 (1), 35. https://doi.org/10.1186/s12711-019-0480-8.

Li, Y.L., Liu, J.X., 2018. StructureSelector: a web-based software to select and visualize the optimal number of clusters using multiple methods. Mol. Ecol. Resour. 18, 176–177. https://doi.org/10.1111/1755-0998.12719.

Magnier, J., Druet, T., Naves, M., Ouvrard, M., Raoul, S., Janelle, J., Moazami-Goudarzi, K., Lesnoff, M., Tillard, E., Gautier, M., Flori, L., 2022. The genetic history of Mayotte and Madagascar cattle breeds mirrors the complex pattern of human exchanges in Western Indian Ocean. G3 Genes|Genomes|Genetics 12 (4). https://doi.org/10.1093/g3journal/jkac029 jkac029.

McQuillan, R., Leutenegger, A.L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., et al., 2008. Runs of homozygosity in european populations. Am. J. Hum. Genet. 83, 359–372. https://doi.org/10.1016/j.ajhg.2008.08.007.

Mitchell, M., 1998. An Introduction to Genetic Algorithms. MIT Press, Cambridge, MA, USA.

Nei, M., 1972. Genetic distance between populations. Am. Nat. 106, 283–292. https://doi.org/10.1086/282771.

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. Mol. Ecol. Resour. 00, 1–9. https://doi.org/10.1111/1755-0998.13837.

Pacheco, C., Stronen, A.V., Jędrzejewska, B., Plis, K., Okhlopkov, I.M., Mamaev, N.V., Drovetski, S.V., Godinho, R., 2022. Demography and evolutionary history of grey wolf populations around the Bering Strait. Mol. Ecol. 31 (18), 4851–4865. https://doi.org/10.1111/mec.16613.

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. Mol. Ecol. Resour. 13, 946–952. https://doi.org/10.1111/1755-0998.12129.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155 (2), 945–959. https://doi.org/10.1093/genetics/155.2.945.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., et al., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575. https://doi.org/10.1086/519795.

R Core Team, 2021. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Rousset, F., 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol. Ecol. Resour. 8, 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x.

Sadeghi, R., Moradi-Shahrbabak, M., Ashtiani, Seyed Reza Miraei, Schlamp, F., Cosgrove, E.J., Antczak, D.F., 2019. Genetic diversity of Persian Arabian Horses and their relationship to other native Iranian Horse Breeds. J. Hered. 110 (2), 173–182. https://doi.org/10.1093/jhered/esy061.

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. Mol. Biol. Evol. 37, 3642–3653. https://doi.org/10.1093/molbev/msaa169.

Saura, M., Caballero, A., Santiago, E., Fernandez, A., Morales-Gonzalez, E., Fernandez, J., Cabaleiro, S., Millan, A., Martanez, 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. Genet. Sel. Evol. 53, 85. https://doi.org/10.1186/s12711-021-00680-9.

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 (Basel) 10 (6), 480. https://doi.org/10.3390/genes10060480.

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 (4), e0123253. https://doi.org/10.1371/journal.pone.0123253.

Sved, J.A., 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theor. Popul. Biol. 2, 125–141. https://doi.org/10.1016/0040-5809(71)90011-6.

Szmatoła, T., Gurgul, A., Jasielczuk, I., Oclon, E., Ropka-Molik, K., Stefaniuk-Szmukier, M., Polak, G., Tomczyk-Wrona, I., Bugno-Poniewierska, M., 2022. Assessment and distribution of runs of homozygosity in horse breeds representing different utility types. Animals 12 (23), 3293. https://doi.org/10.3390/ani12233293.

Telalbašić, R., Žiga, E., Rahmanović, A., 2009. Historijat ergele„ Borike" i uzgoj arapskog konja od 1895. do 2007. Godine. Stočarstvo 63 (1), 65–76.

Vostry, L., Vostra-Vydrova, H., Moravcikova, N., Kasarda, R., Cubric-Curik, V., Brzakova, M., Solkner, J., Shihabi, M., Moreno, J.A.H., Spehar, M., Curik, I., 2023. Genomic diversity and population structure of the Czech Holstein cattle. Liv. Sci. 273, 105261 https://doi.org/10.1016/j.livsci.2023.105261.

Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population-structure. Evolution (N Y) 38, 1358–1370. https://doi.org/10.1111/j.1558-5646.1984. tb05657.x.

Wright, S., 1978. Evolution and the Genetics of Populations. University of Chicago Press, Chicago. Vol. 4.

Žiga, E., Telalbašić, R., Rahmanović, A., 2008. Borički Arapski Konj, Zbornik Radova. TKD Šahinpašić, Sarajevo.