

## RESEARCH ARTICLE

# Genomic comparisons of Persian Kurdish, Persian Arabian and American Thoroughbred horse populations

Navid Yousefi-Mashouf<sup>1</sup>\*, Hassan Mehrabani-Yeganeh<sup>2†</sup>, Ardeshir Nejati-Javaremi<sup>2‡</sup>, Ernest Bailey<sup>1</sup>, Jessica L. Petersen<sup>3</sup>\*

**1** MH Gluck Equine Research Center, University of Kentucky, Lexington, KY, United States of America, **2** Department of Animal Sciences, University of Tehran, Karaj, Iran, **3** Department of Animal Science, University of Nebraska-Lincoln, Lincoln, NE, United States of America

\* These authors contributed equally to this work.

† These authors also contributed equally to this work.

\* [navid.yousefi@uky.edu](mailto:navid.yousefi@uky.edu) (NYM); [jessica.petersen@unl.edu](mailto:jessica.petersen@unl.edu) (JLP)



## OPEN ACCESS

**Citation:** Yousefi-Mashouf N, Mehrabani-Yeganeh H, Nejati-Javaremi A, Bailey E, Petersen JL (2021) Genomic comparisons of Persian Kurdish, Persian Arabian and American Thoroughbred horse populations. *PLoS ONE* 16(2): e0247123. <https://doi.org/10.1371/journal.pone.0247123>

**Editor:** Tzen-Yuh Chiang, National Cheng Kung University, TAIWAN

**Received:** September 4, 2020

**Accepted:** February 1, 2021

**Published:** February 16, 2021

**Copyright:** © 2021 Yousefi-Mashouf et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The full SNP dataset in PED/MAP format has been deposited on the Open Science Framework website under DOI [10.17605/OSF.IO/TR9KW](https://doi.org/10.17605/OSF.IO/TR9KW) which can be accessed by the URL: <https://osf.io/tr9kw/>.

**Funding:** NYM and HMY, The fund was provided by a private company, Ava Tejarat Saba for non-profit purposes. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

The present research aimed to characterize the Persian Kurdish horse population relative to the Persian Arabian and American Thoroughbred populations using genome-wide SNP data. Fifty-eight Kurdish, 38 Persian Arabian and 83 Thoroughbred horses were genotyped across 670,796 markers. After quality control and pruning to eliminate linkage disequilibrium between loci which resulted in 13,554 SNPs in 52 Kurdish, 24 Persian Arabian and 58 Thoroughbred horses, the Kurdish horses were generally distinguished from the Persian Arabian samples by Principal Component Analyses, cluster analyses and calculation of pairwise  $F_{ST}$ . Both Persian breeds were discriminated from the Thoroughbred. Pairwise  $F_{ST}$  between the two Persian samples (0.013) was significantly greater than zero and several fold less than those found between the Thoroughbred and Kurdish (0.052) or Thoroughbred and Persian Arabian (0.057). Cluster analysis assuming three genetic clusters assigned the Kurdish horse and Thoroughbred to distinct clusters (0.942 in cluster 2 and 0.953 in cluster 3 respectively); the Persian Arabian was not in a distinct cluster (0.519 in cluster 1), demonstrating shared ancestry or recent admixture with the Kurdish breed. Diversity as quantified by expected heterozygosity was the highest in the Kurdish horse (0.342), followed by the Persian Arabian (0.328) and the Thoroughbred (0.326). Analysis of Molecular Variance showed that 4.47% of the genetic variation was present among populations ( $P < 0.001$ ). Population-specific inbreeding indices ( $F_{IS}$ ) were not significantly different from zero in any of the populations. Analysis of individual inbreeding based on runs of homozygosity using a larger SNP set suggested greater diversity in both the Kurdish and Persian Arabian than in the Thoroughbred. These results have implications for developing conservation strategies to achieve sound breeding goals while maintaining genetic diversity.

**Competing interests:** NYM and HMY, The fund was provided by a private company, Ava Tejarat Saba for non-profit purposes. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

## Introduction

The Kurdish horse of Iran is one of the five major Iranian horse breeds. The other Iranian horse breeds include Caspian, Turkoman, Persian Arabian (also known as Assil) and Dare-Shouri. Historical literature chronicles a developmental history of more than 2500 years for Kurdish horses, relating them to an ancient, now-extinct population of horses called “Nesayee.” The Nesayee horses have been documented to have served as transportation for the army of Medes tribe, whose realm was congruent with today’s homeland of Kurdish horses (west of Iran) [1]. It is believed that the Nesayee horses brought to the Southwest of Iran also became the ancestors of the modern Persian Arabian horses.

No formal registry exists for the Kurdish horse, however, the breed is a well-known land-race, selected for agility, dressage gaits, mountain riding and resistance to harsh environmental conditions. We previously characterized the Kurdish horse from a phenotypic perspective by establishing the breed standards, which describe the ideal characteristics used as selection criteria by breeders [2].

The Persian Arabian horse originated in the southwestern part of Iran but currently is a geographically neighboring population to the Kurdish horse and one might suspect some admixture or ancestral relationship between these two populations, as also endorsed by the available historical information. Preliminary genetic studies suggested an immediate common ancestor between Kurdish and Persian Arabian populations [3]. In addition, as these breeds occupy overlapping geographical areas, the question has arisen as to whether they are distinct. Hence, the present research aimed to characterize the diversity of and relationships between these Persian breeds using genome-wide single nucleotide polymorphism (SNP) data from Kurdish horses and Persian Arabians as well as data from the more distantly related American Thoroughbred. We hypothesized that the current population of Kurdish horses can be considered as a unique and homogeneous population distinct from Persian Arabian horses from the genomic standpoint.

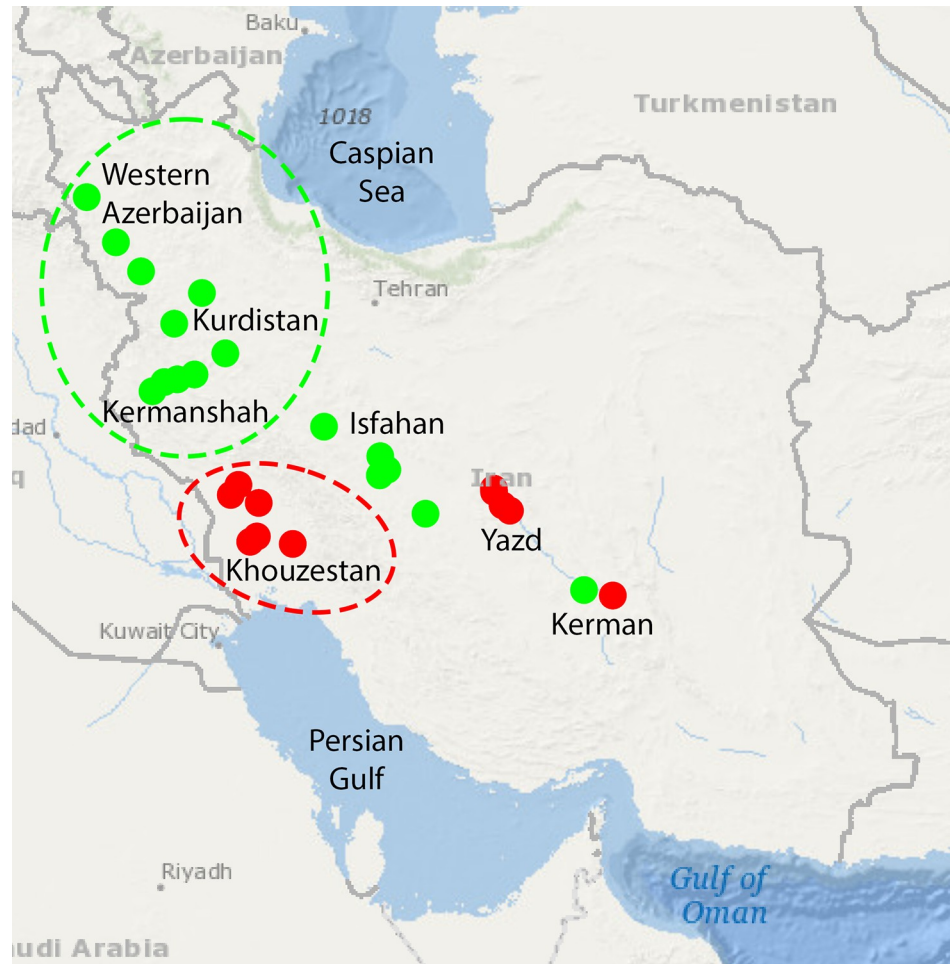
## Materials and methods

### Sampled individuals

We sampled 58 Kurdish horses (43 males and 15 females) distributed over a wide geographic range to ensure the highest level of diversity, including five provinces of Kermanshah, Kurdistan, Western Azerbaijan, Isfahan and Kerman. The 38 Persian Arabian samples (11 males and 27 females), all registered in the Persian Arabian studbook, were obtained in the provinces of Khuzestan, Yazd and Kerman (Fig 1). The Kurdish horses sampled from the central locations (Isfahan and Kerman) did not originate there, rather they (or their ancestors) were imported from the three western provinces (Kermanshah, Kurdistan and Western Azerbaijan). Similarly, the Persian Arabian horses sampled from the central locations (Yazd and Kerman), have their origin from the southwest (Khuzestan). DNA samples were provided from the archive at the University of Kentucky for 83 American Thoroughbred horses (44 males and 39 females). The Thoroughbred horses were randomly sampled from 8 farms in central Kentucky.

### Blood collection and DNA extraction

Blood samples drawn from the jugular vein were collected in 6 ml EDTA vacuum tubes, transported to the lab cooled, and were kept frozen until DNA isolation. DNA extraction was carried out using Phenol-Chloroform protocol [4].



**Fig 1. Sampling locations for Iranian populations.** Green and red points signify sampling locations for Kurdish and Persian Arabian horses, respectively. The regions circled in green and red identify the original homeland of Kurdish and Persian Arabian horses, respectively. All the point locations outside the ovals represent horses that were descendants of, or were themselves imported from the original homelands. Map imported from the USGS National Map open resources.

<https://doi.org/10.1371/journal.pone.0247123.g001>

### Ethical statement

The IACUC committee at the University of Kentucky waived review by the ethics committee in 2010 since the Thoroughbred horse samples were obtained from privately owned horses and provided by the owners. Likewise, the samples from Persian horses came from horses that were privately owned and managed, samples were provided by owners so a formal review of the study protocol for ethical treatment of horses was deemed unnecessary. The Department of Animal Science at the University of Tehran is responsible for evaluations of ethical use of animals.

### Genotyping

DNA samples were submitted to Neogen GeneSeek (Lincoln, Nebraska) for genotyping with the Axiom Equine Genotyping Array (Affymetrix Inc.) which harbors 670,796 SNP markers [5].

## Data analysis

SNP & Variation Suite version 8 (Golden Helix, Inc., Bozeman, MT, [www.goldenhelix.com](http://www.goldenhelix.com)) software [6] was used for basic quality control of the genotype data, in which the markers were disqualified if the Minor Allele Frequency (MAF) was  $< 0.05$  and per-SNP Call Rate  $< 0.95$ . All the markers on the X chromosome or with unknown genomic location (categorized as CHR\_UN) were removed from the dataset. To avoid inclusion of closely-related individuals, the dataset was pruned for Identity by Descent (IBD), such that between each pair of individuals with  $IBD > 0.2$ , the one that had more genetic relationship with the other individuals in the sample set was removed. Lastly, SNPs that were in LD across samples were also removed, pruning for an LD threshold of  $r^2 = 0.25$ , considering 100 SNP windows and moving 25 SNPs per set (LD computation method: CHM). As the unequal number of samples among populations could bias LD pruning towards the populations with higher number of samples, we randomly selected 24 individuals from each population and performed LD pruning on the 72-individual subset (24×3), then applied the selected markers to the whole population. To analyze runs of homozygosity, however, LD pruning was not applied although markers with a genotyping rate  $< 0.95$  and the loci on the X chromosome or contigs unassigned chromosomes were removed.

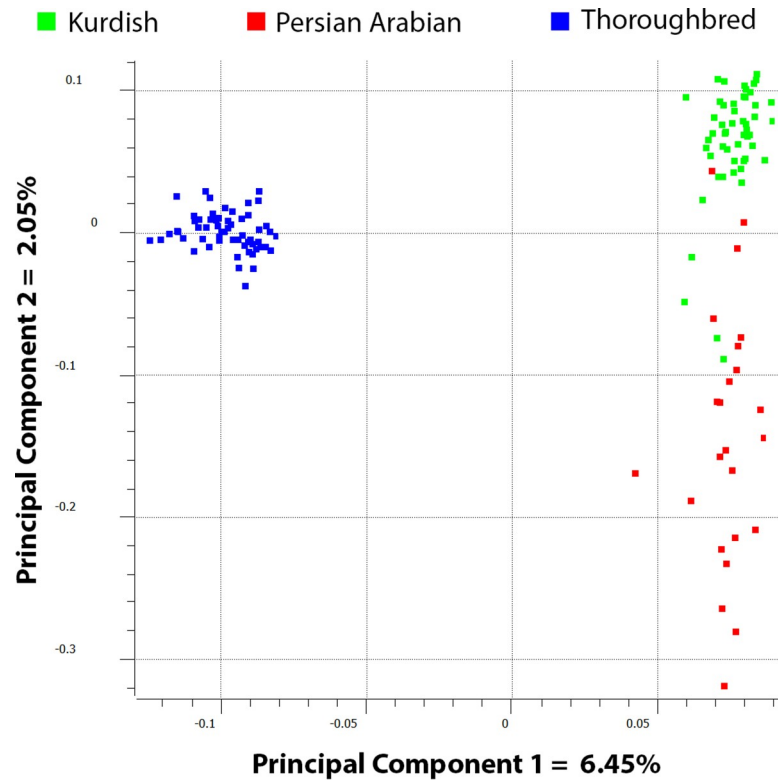
Principal Component Analysis was carried out in the SNP & Variation Suite v8 (SVS). Pairwise  $F_{ST}$  values between breeds were calculated using SVS and Arlequin version 3.5 [7]. Arlequin was also used to obtain expected heterozygosity ( $H_E$ ) and population-specific inbreeding ( $F_{IS}$ ) values as well as to perform an analysis of molecular variance (AMOVA). Average inbreeding coefficients for each population were calculated using SVS v8 and PLINK 1.07 [8]. To calculate runs of homozygosity (ROH) and obtain ROH-based inbreeding coefficients for each individual, the R package detectRuns was used [9] with parameters: windowSize = 15, threshold = 0.1, minSNP = 15, maxOppWindow = 1, maxMissWindow = 1, maxGap = 1000000, minLengthBps = 250000, minDensity = 1/10000. To quantify ROH of various lengths (0 to  $\geq 48$ Mb), 20 individuals from each sample were randomly selected in three iterations with the mean and standard deviation in counts/class calculated for each sample. Clustering of breeds into genetic groups was examined using the STRUCTURE program version 2.3.4 assuming K values of 1 to 5, replicating the analysis of each K value five times. The STRUCTURE algorithm assumed the admixture model and correlated allele frequencies. Burn-in iterations of 10,000, 25,000, 100,000, 150,000 and 300,000 reps were tested along with different MCMC repetitions of 100,000, 200,000, 250,000 and 600,000 to confirm convergence. Also, to see if the Thoroughbred samples biased the clustering patterns of Kurdish and Persian Arabian groups, we ran STRUCTURE without the Thoroughbred samples at  $K = 2$ . To determine the optimal value of K using the Evanno method [10], the online program STRUCTURE Harvester [11] was employed.

## Results

After data pruning, the final dataset included 13,554 SNPs for a total of 134 individuals out of the 180 horses genotyped. The IBD filtering left 50 out of 58 Kurdish, 24 out of 38 Persian Arabian and 58 out of 83 Thoroughbred horses. For ROH, after removing SNPs with genotyping rate  $< 95\%$ , 262,390 autosomal SNPs were included in the analysis.

### Principal Component Analysis (PCA)

The first Principal Component explained 6.45% of the variance, which discriminated the Thoroughbred from both Iranian breeds. The second PC, capturing 2.05% of the variance, showed divergence between the Kurdish and Persian Arabian samples (Fig 2).



**Fig 2. Plot of principal components 1 versus 2 for the 134 horse representing 3 breeds.**

<https://doi.org/10.1371/journal.pone.0247123.g002>

The plot also showed that Thoroughbred horses formed a tight cluster. The Kurdish horses clustered in a manner similar to the Thoroughbred, while the Persian Arabian horses had a wider distribution across PC2, where the Kurdish and Persian Arabian clusters were adjacent to one another, with several individuals lying in overlapping regions.

### $F_{ST}$

Pairwise  $F_{ST}$  values were all significantly greater than zero with the least divergence observed between the Kurdish horse and Persian Arabian (Table 1).

### Population specific inbreeding

Although greatest in the Kurdish horses, no  $F_{IS}$  value for any sample was significantly greater than zero (Table 2).

**Table 1. Pairwise  $F_{ST}$  values between breed groups.**

Population Pair	$F_{ST}$	$P$ Value
Kurdish–Persian Arabian	0.013	<0.0001
Kurdish–Thoroughbred	0.052	<0.0001
Thoroughbred–Persian Arabian	0.057	<0.0001

All of the  $P$  values were significant ( $P < 0.05$ ).

<https://doi.org/10.1371/journal.pone.0247123.t001>

**Table 2. Population-specific  $F_{IS}$  values.**

Population	$F_{IS}$	$P$ (Rand $F_{IS} \geq$ Obs $F_{IS}$ )
Kurdish	0.005	0.386
Persian Arabian	-0.018	0.620
Thoroughbred	-0.008	0.636

<https://doi.org/10.1371/journal.pone.0247123.t002>

### Analysis of Molecular Variance (AMOVA)

AMOVA identified 4.47% of the variation present among populations ( $P < 0.001$ ), -0.41% among individuals within populations ( $P = 0.585$ ), and 95.94% within individuals ( $P = 0.065$ ). By eliminating the Thoroughbred from the analysis, the among-population variation remained significant ( $P < 0.001$ ), explaining 1.29% of the variation between the Kurdish and Persian Arabian samples.

### Runs of homozygosity analysis

The Thoroughbred samples had a greater number of longer runs of homozygosity than the Persian and Kurdish horses (S1 Table). The Kurdish horse had fewer ROH  $\geq$  6Mb than either the Persian Arabian or Thoroughbred, with no ROH longer than 24Mb.

The mean individual inbreeding coefficient for the Kurdish horse was less than that of the Persian Arabian (Fig 3), both of which were less than that of the Thoroughbred. Variation among samples was greatest in the Iranian samples, and notably in the Persian Arabian.

The violin plot identifies three outliers in the Persian Arabian group as well as two in the Kurdish population. The same individuals were identified as outliers by PLINK calculations of individual inbreeding. The pedigree of the three outlier Persian Arabian horses available from the studbook confirmed presence of common ancestors in both their paternal and maternal lines and conformed with their higher inbreeding values. For the two outlier Kurdish horses, however, limited pedigree information was available and the extent of shared ancestry could not be determined. Lastly, the correlation between the  $F_{ROH}$  and the inbreeding values ( $F$ ) calculated in PLINK was 0.765.

### Expected heterozygosity ( $H_E$ )

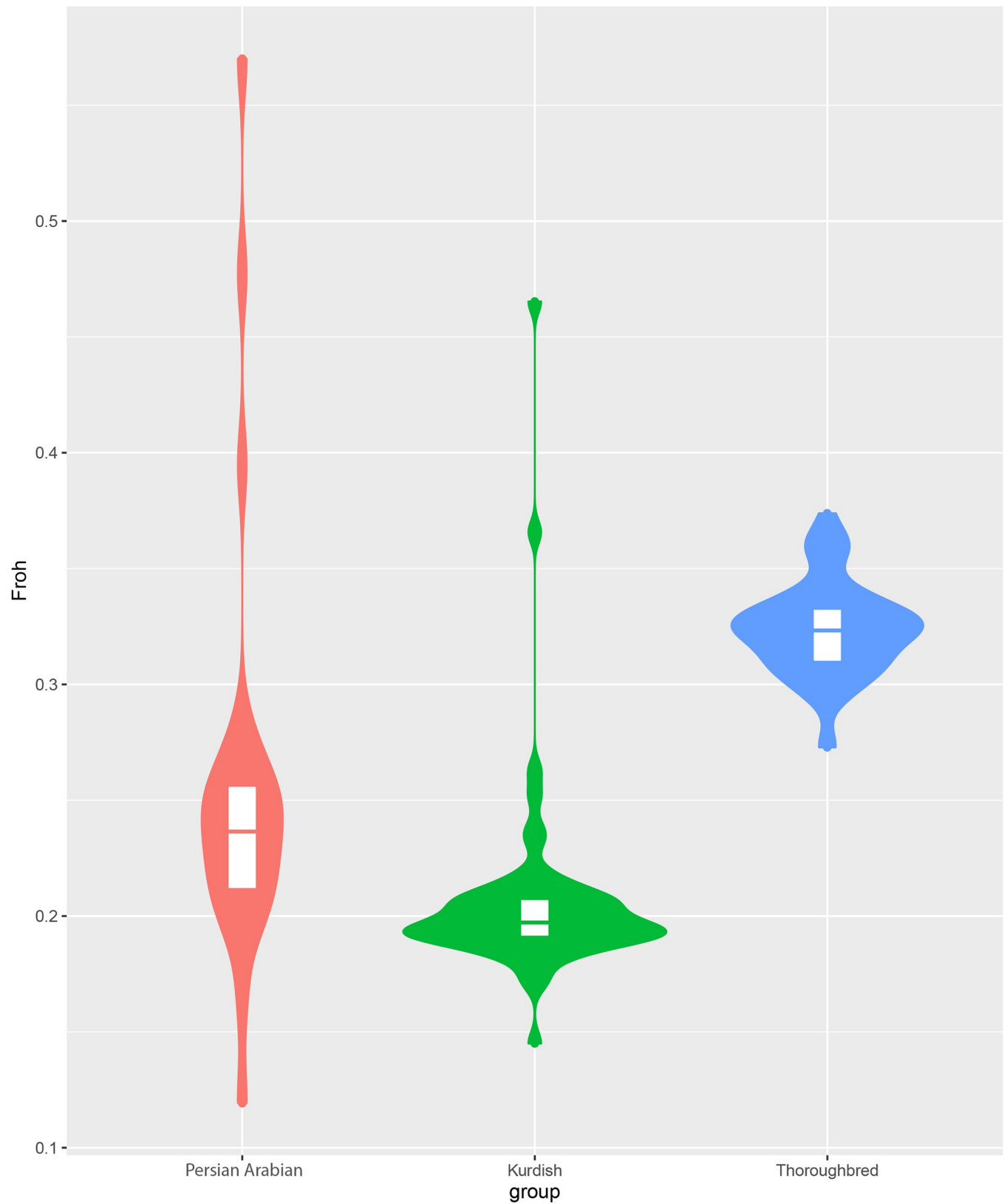
Expected heterozygosity was greatest in the Kurdish, followed by the Persian Arabian and Thoroughbred (Table 3).

### Cluster analysis

For cluster analyses, burn-in and MCMC iterations of 150,000 and 250,000 (respectively) produced consistent results and converged at the highest value of K examined ( $K = 5$ ). Of the five K values (i.e. the number of subpopulations hypothesized to exist within the entire sample set) tested in this analysis,  $K = 2$  and  $K = 3$  were most informative and were further scrutinized in more detail. Little change was observed when 4 or 5 clusters were considered. The Evanno method identified  $K = 2$  as the best fit for these data (S2 Table).

Assuming two clusters, the Thoroughbreds were assigned to a single cluster and the Persian horses (Kurdish + Persian Arabian) to the other (Table 4, Fig 4).

Assuming three genetic clusters ( $K = 3$ ), the Kurdish and Thoroughbred horses were each assigned to distinct clusters (Fig 5). The third genetic cluster was found primarily in Persian Arabians, which still showed shared ancestry with the Kurdish horses (S3 Table).



**Fig 3. Violin plot of mean, quartiles and the frequency (the width of the plot) of the ROH-based inbreeding coefficient ( $F_{ROH}$ ) for each breed group.**

<https://doi.org/10.1371/journal.pone.0247123.g003>

**Table 3. Average expected heterozygosity values for each breed group.**

Population	Expected Heterozygosity
Kurdish	0.342
Persian Arabian	0.328
Thoroughbred	0.326
Total	0.341

<https://doi.org/10.1371/journal.pone.0247123.t003>

Lastly, running the STRUCTURE program excluding Thoroughbred samples from the dataset (at  $K = 2$ ) returned similar clustering values for Kurdish and Persian Arabian groups to the original  $K = 3$  results (S4 Table).

For a comparison of outliers/overlapping individuals between Kurdish and Persian Arabian samples, we first identified the individuals in the PC plot that were positioned between the two main breed clusters and ranked them in the order of being closest to the opposing population (based on their PC2 values). Then we examined the individual Structure output for each of the individuals that were intermediate between the two breeds and ranked them in the order of having the highest membership in the opposing cluster. The number and ranking of the outlier/overlapping individuals were identical between the two analyses (6 Persian Arabian and 4 Kurdish individuals). All horses that appear intermediate between the two sets of Persian horse samples were assigned more strongly to the cluster representing the other breed.

## Discussion

Several genetic studies have incorporated horses from Iranian populations [12–19], this is the first, however, to have focused specifically on the Persian Kurdish horse. With respect to our hypothesis, we found evidence that the Kurdish horse population is distinct from but shares ancestry with the Persian Arabians. Additionally, although some individual horses have genomic evidence of inbreeding, overall both Persian breeds appear to be more diverse than the sample of American Thoroughbreds to which they were compared. As our Thoroughbred sample set was limited to a small number of farms in Kentucky, the Thoroughbred analyses in this report may not be generalizable to the whole population of Thoroughbred, but were intended to serve as a comparison to the two Iranian populations. Overall, the results of the present study serve as baseline data characterizing the diversity and genetic make-up of these breeds to allow the development of conservation strategies to achieve sound breeding goals while maintaining genetic diversity of these breeds.

## Analyses of population structure

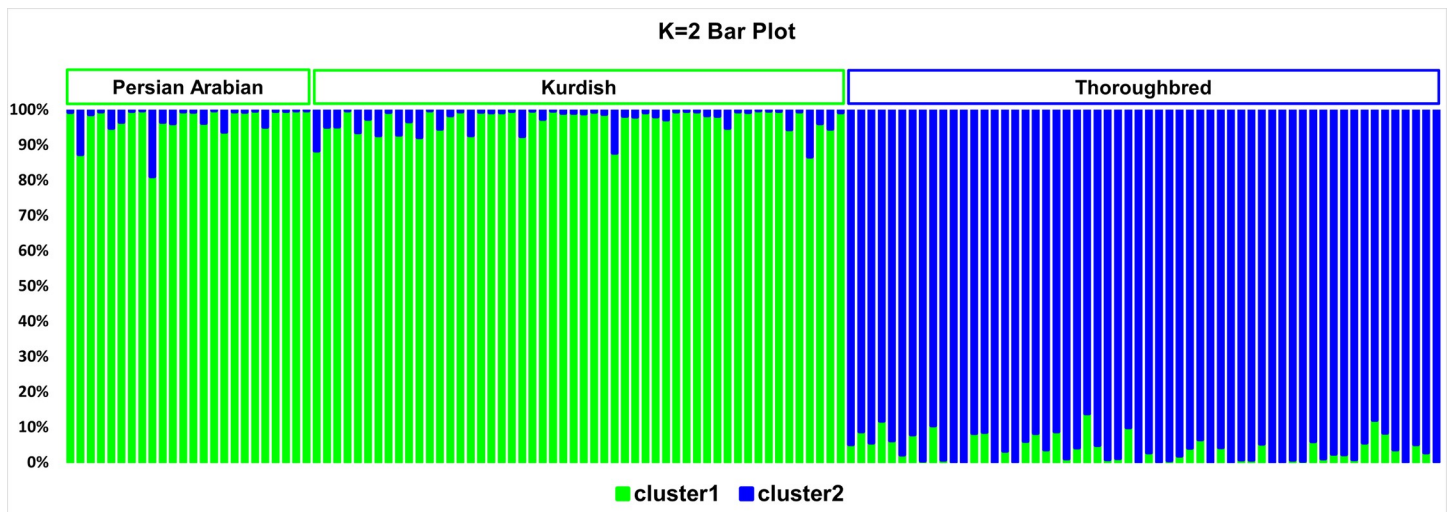
Our hypothesis that the two Persian breeds are distinct was supported by significant pairwise  $F_{ST}$  values and AMOVA analyses support the distinction of the Kurdish and Persian Arabian samples. As visualized by the PC plot and cluster analyses, however, there is evidence of gene flow between the two breeds. There are two explanations for this observation. The first is that Persian Arabians might have originated from the Kurdish horse and thus Arabians maintain

**Table 4. Average proportion of membership of each pre-defined population in each of the 2 clusters at  $K = 2$ .**

Population	Cluster 1	Cluster 2
Kurdish	0.971	0.029
Persian Arabian	0.971	0.029
Thoroughbred	0.038	0.962

<https://doi.org/10.1371/journal.pone.0247123.t004>

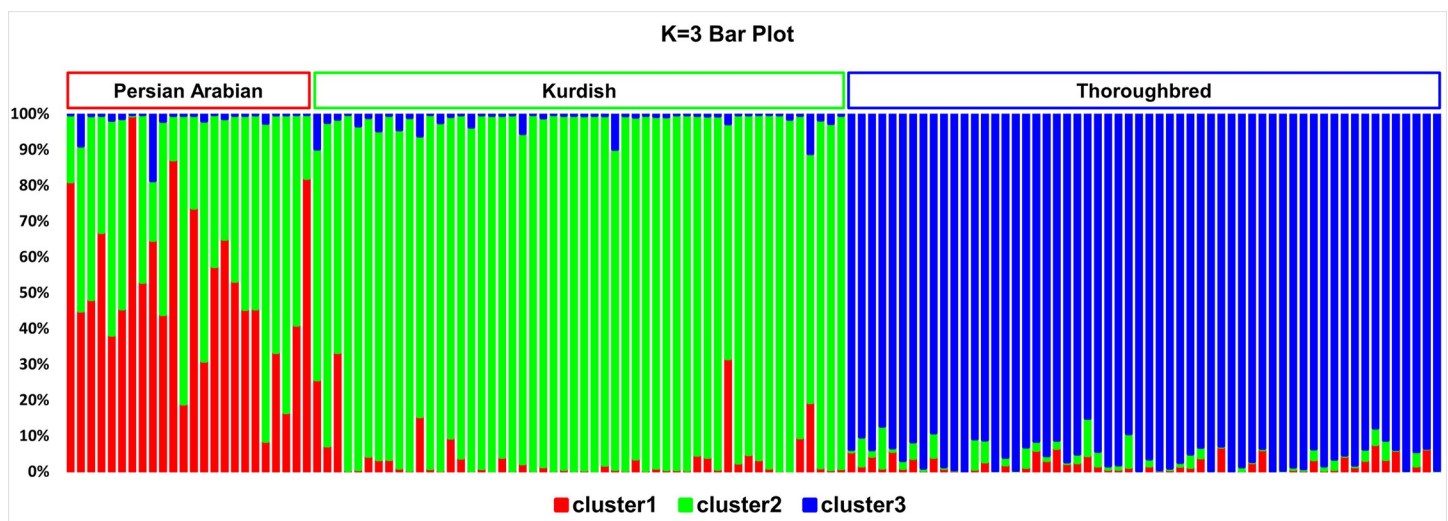




**Fig 4. Bar plot of the K = 2 results.** The green color designates cluster 1 (which mostly harbored Kurdish and Persian Arabian horses) and the blue color signifies cluster 2 (which mostly contained Thoroughbred horses). Each individual is represented by a single vertical line broken into K colored segments, with lengths proportional to each of the K inferred clusters.

<https://doi.org/10.1371/journal.pone.0247123.g004>

genomic ties to the Kurdish horses resulting from that founding event, while the Kurdish population has independently evolved. The alternative explanation, considering the Kurdish population may have originated from the Persian Arabian, is based upon the Persian Arabian being a diverse and large breed, relative to the Kurdish horse. A founder effect could have led to the isolation of a portion of this diversity into a new genetic group, that led to, or integrated with Kurdish horses. Gene flow from the Persian Arabian to the Kurdish horse may have resulted from recent population admixture. At the time when the first studbooks for Persian Arabian horses were established around 1976, there was no official breed registry for Kurdish horses; it may be that horse owners with a Kurdish-Persian Arab crossbred or even a Kurdish horse,



**Fig 5. Bar plot of the K = 3 results.** The blue color designates cluster 3 (which mostly harbored Thoroughbred horses), the green color signifies cluster 2 (which mostly contained Kurdish horses and covered a part of Persian Arabian's genome), and the red color represents cluster 1, which is attributable to Persian Arabian. Each individual is represented by a single vertical line broken into K colored segments, with lengths proportional to each of the K inferred clusters.

<https://doi.org/10.1371/journal.pone.0247123.g005>

preferred to have their horses registered with a breed registry, rather than leaving it officially unknown. This would lead to the presence of Kurdish characteristics among Persian Arabian horse populations. Nevertheless, a caveat of cluster analyses and measures of relationships is that directionality of gene flow is not defined; regardless, these data support gene flow between the two Persian populations.

Cluster and PC analyses both suggested that the Kurdish horses are genetically more homogeneous than the Persian Arabians. The observation of higher diversity in the Arabian population as compared to the Thoroughbred is in agreement with reports by Khanshour *et al.* (2013) [12], Sadeghi *et al.* (2018) [15] and is concordant with the latest population study on Arabian horses, which identified a high degree of genetic variation and complex ancestry of those horses from the Middle East, including the Persian group [20]. Diversity of the Persian Arabians in this study is supported by the evidence of gene flow with the Kurdish horse in cluster analysis. Given the observation that the inbreeding at the population level is not significantly different from zero, it may be interpretable that adverse effects of inbreeding depression is less of a concern in the studied populations and there is still a good wealth of genetic diversity which can be used towards breeding if proper management is applied. This is further supported by the analysis of runs of homozygosity. Despite their relatively small population size and likely ascertainment bias of SNP loci favoring detection of rare variants in the Thoroughbred, the Thoroughbred showed a greater proportion of long (>6 Mb) runs of homozygosity, which were less numerous in the Persian Arabian and absent in the Kurdish horse.

### Analyses of individual diversity

With a small population size and lack of formal record keeping, it was of interest to calculate individual estimates of inbreeding based upon both heterozygosity (PLINK) and runs of homozygosity (detectRuns). Although gross calculations of average individual inbreeding present Persian Arabians to be more inbred than Kurdish and Thoroughbred horses, the overall inbreeding level of the Persian Arabian population appears to be driven by the presence of three outliers. Averaging the inbreeding coefficients of the Persian Arabian samples without those outliers, brings the mean inbreeding of the Persian Arabian population lower than that of the Thoroughbred and Kurdish populations. Nevertheless, inbreeding has always been an apparent issue among Persian Arabian horse breeders, as the majority of Persian Arabian lines are descendants of two famous stallions, namely Haddad and Samarghand. Despite considerable inbreeding at the individual level, however, the population has still maintained a good diversity, as implied from their heterozygosity values. Although, their high genetic diversity ensures a healthy population from the genetic standpoint, it is advisable to the breeders to relieve individual inbreeding in their breeding practices, to avoid negative consequences of inbreeding depression. The results also provide another piece of evidence that the Kurdish horse is relatively diverse. This might be due to the fact that the population of Kurdish horses has been mostly shaped by natural selection over a long time rather than human artificial selection.

### Conclusions

- There is evidence the Kurdish horse population forms a distinct genetic cluster with some individuals showing mixed ancestry.
- There is evidence of gene flow between the Persian Arabian and Kurdish horses.

- The overall diversity parameters in the Kurdish horses resembled that of Thoroughbreds. Kurdish horses are a smaller population and it may be appropriate to develop conservation strategies and sound breeding goals to maintain their genetic diversity.

## Supporting information

**S1 Table. Summary of the total number of runs of homozygosity by each size class.** Due to unequal sample size, three iterations, each of 20 randomly selected individuals was evaluated. Given is the mean and standard deviation (in parenthesis) of the replicates.

(DOCX)

**S2 Table. Results on the comparisons of  $K = 1$  to 5 tested by STRUCTURE Harvester.** The highlighted row belongs to the  $K$  value ( $= 2$ ) that maximizes Delta  $K$  per the Evanno method of determining the best fit for the data.

(DOCX)

**S3 Table. Average proportion of membership of each pre-defined population in each of the 3 clusters at  $K = 3$ .**

(DOCX)

**S4 Table. Average proportion of membership of each pre-defined population (excluding the Thoroughbred samples from the dataset) in each of the 2 clusters at  $K = 2$ .**

(DOCX)

## Acknowledgments

We would like to express our best appreciations for financial support from Ava Tejarat Saba Co. and for critical help and advice from Mr. Mohammad Farzanehfar, Dr. Mohammad Moradi Shahrabak, and Yekrun Horse Industry Development Co. Ltd. We also appreciate cooperation of the horse owners, faculty members and graduate students at the University of Tehran and the University of Kentucky.

## Author Contributions

**Conceptualization:** Navid Yousefi-Mashouf, Hassan Mehrabani-Yeganeh, Ardeshir Nejati-Javaremi, Ernest Bailey, Jessica L. Petersen.

**Formal analysis:** Navid Yousefi-Mashouf, Ernest Bailey, Jessica L. Petersen.

**Funding acquisition:** Hassan Mehrabani-Yeganeh.

**Investigation:** Navid Yousefi-Mashouf, Ernest Bailey, Jessica L. Petersen.

**Methodology:** Navid Yousefi-Mashouf, Ernest Bailey, Jessica L. Petersen.

**Project administration:** Navid Yousefi-Mashouf.

**Resources:** Navid Yousefi-Mashouf.

**Software:** Navid Yousefi-Mashouf, Jessica L. Petersen.

**Supervision:** Navid Yousefi-Mashouf, Hassan Mehrabani-Yeganeh, Ardeshir Nejati-Javaremi, Ernest Bailey, Jessica L. Petersen.

**Validation:** Navid Yousefi-Mashouf, Ernest Bailey, Jessica L. Petersen.

**Visualization:** Navid Yousefi-Mashouf, Hassan Mehrabani-Yeganeh, Ardeshir Nejati-Javaremi, Ernest Bailey, Jessica L. Petersen.

**Writing – original draft:** Navid Yousefi-Mashouf.

**Writing – review & editing:** Navid Yousefi-Mashouf, Hassan Mehrabani-Yeganeh, Ardeshir Nejati-Javaremi, Ernest Bailey, Jessica L. Petersen.

## References

1. Diakonoff IM. The History of Media: Moscow-Leningrad; 1956.
2. Yousefi-Mashouf N, Mehrabani-Yeganeh H, Nejati-Javaremi A, Maloufi F. A Novel Approach to Establish Breed Type and Standards for an Equine Breed: Persian Kurdish Horse. *Journal of Agricultural Science and Technology*. 2020; 22(5).
3. Ovchinnikov IV, Dahms T, Herauf B, McCann B, Juras R, Castaneda C, et al. Genetic diversity and origin of the feral horses in Theodore Roosevelt National Park. 2018; 13(8):e0200795.
4. Sambrook J, Russell DWJCSHP. Purification of nucleic acids by extraction with phenol: chloroform. 2006; 2006(1):pdb. prot4455.
5. Schaefer RJ, Schubert M, Bailey E, Bannasch DL, Barrey E, Bar-Gal GK, et al. Developing a 670k genotyping array to tag~ 2M SNPs across 24 horse breeds. 2017; 18(1):565.
6. Golden Helix I. SNP & Variation Suite™. Bozeman, MT: Golden Helix, Inc. Version 8:[Software]. Available from: <http://www.goldenhelix.com>.
7. Excoffier L, Lischer HEJMer. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. 2010; 10(3):564–7. <https://doi.org/10.1111/j.1755-0998.2010.02847.x> PMID: 21565059
8. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. 2007; 81(3):559–75.
9. Biscarini F, Cozzi P, Gaspa G, Marras G. detectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes. CRAN (The Comprehensive R Archive Network); 2018.
10. Evanno G, Regnaut S, Goudet JJMe. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. 2005; 14(8):2611–20.
11. Earl DAJCGr. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. 2012; 4(2):359–61.
12. Khanshour A, Conant E, Juras R, Cothran EGJJoH. Microsatellite analysis of genetic diversity and population structure of Arabian horse populations. 2013; 104(3):386–98.
13. Petersen JL, Mickelson JR, Cothran EG, Andersson LS, Axelsson J, Bailey E, et al. Genetic diversity in the modern horse illustrated from genome-wide SNP data. 2013; 8(1):e54997.
14. Evrigh NH, Omri M, Boustan A, Seyedsharifi R, Vahedi VJJoEVS. Genetic Diversity and Structure of Iranian Horses' Population Based on Mitochondrial Markers. 2018; 64:107–11.
15. Sadeghi R, Moradi-Shahrbabak M, Miraei Ashtiani SR, Schlamp F, Cosgrove EJ, Antczak DFJJoH. Genetic Diversity of Persian Arabian Horses and Their Relationship to Other Native Iranian Horse Breeds. 2018; 110(2):173–82.
16. Rahimi-Mianji G, Nejati-Javaremi A, Farhadi AJRjog. Genetic diversity, parentage verification, and genetic bottlenecks evaluation in iranian turkmen horse. 2015; 51(9):916–24.
17. Amirinia C, Seyedabadi H, Banabazi MH, Kamali MAJJoBS. Bottleneck Study and Genetic Structure of Iranian Caspian Horse. 2007; 10(9):1540–3.
18. Rafeie F, Amirinia C, Javaremi AN, Mirhoseini SZ, Amirmozafari NJAJoB. A study of patrilineal genetic diversity in Iranian indigenous horse breeds. 2011; 10(75):17347–52. <https://doi.org/10.7150/ijms.8.709> PMID: 22135617
19. Shahsavarani H, Rahimi-Mianji GJJoB. Analysis of genetic diversity and estimation of inbreeding coefficient within Caspian horse population using microsatellite markers. 2010; 9(3).
20. Cosgrove EJ, Sadeghi R, Schlamp F, Holl HM, Moradi-Shahrbabak M, Miraei-Ashtiani SR, et al. Genome diversity and the origin of the Arabian horse. 2020; 10(1):1–13.