

# Assessment of population structure depending on breeding objectives in Spanish Arabian horse by genealogical and molecular information



J.F. Delgado<sup>a,1</sup>, N. De Andrés<sup>a,1</sup>, M. Valera<sup>b</sup>, J.P. Gutiérrez<sup>a</sup>, I. Cervantes<sup>a,\*</sup>

<sup>a</sup> Department of Animal Production, Veterinary Faculty, Complutense University of Madrid, Avda. Puerta del Hierro s/n, 28040 Madrid, Spain

<sup>b</sup> Department of Agro-Forestal Sciences, ETSIA, University of Sevilla, Ctra. Utrera km 1, 41013 Sevilla, Spain

## ARTICLE INFO

### Article history:

Received 10 December 2013

Received in revised form

11 July 2014

Accepted 17 July 2014

### Keywords:

Equine

Pedigree

Molecular markers

Breeding objective

Genetic variability

## ABSTRACT

The Arabian Horse is one of the most valued breeds in an international and historical context and has been involved in the formation of many other horse breeds. Since 2005, the Spanish Arabian Horse Breeder Association (AECCA) has developed a breeding program aimed at improving both conformation traits and endurance performance. While this selection depends on individual breeders, a population structure might appear by preferential mating within groups of animals according to different objectives. The aims of this study were to determine the differences between Arabian horses bred for different breeding objectives: endurance competitions, morphological shows and other aptitudes and to check if this structure population can be assessed by using genealogical, molecular tools or both. Genealogical and molecular information was randomly obtained from 120 Arabian horses. The animals were classified into three groups according to the breeding goal: morphology, endurance and other aptitudes. Some initial analyses were carried out to study the structure of the sampled animals using genealogical and molecular parameters. An analysis of the genetic structure using both types of information source was performed. Both molecular and genealogical analyses were congruent, and both seemed to be valid when studying the genetic structure of this population. The correlation between coancestries using molecular and pedigree information was 0.60. The differences between the groups were minimum when compared with the genetic structure within groups. Therefore, a horse with a specific breeding objective is not genetically much different regard the rest of the objectives. However, the morphological group appeared as the most separated from the rest, both at a genealogical and molecular level. Regarding the possible impact of the subdivision in the population it can be claimed that no loss of genetic variability is expected in the short-term, because the groups were genetically connected.

© 2014 Elsevier B.V. All rights reserved.

## 1. Introduction

The Arabian Horse is one of the most valued breeds at an international and historical level. The breed's versatility

for sport performance and good conformation has resulted in an important contribution to the formation of other horse breeds. The Spanish Arabian horse is one of the main populations of this breed in Spain, with a census of 14,247 individuals (Magrama, 2011). The Arabian horse was imported to Spain in the middle of the 19th century by the Spanish Ministry of War, while private breeders started importing these horses in the 20th century. The selection

\* Corresponding author. Tel./fax: +34 913943773.

E-mail address: [icervantes@vet.ucm.es](mailto:icervantes@vet.ucm.es) (I. Cervantes).

<sup>1</sup> Both authors contributed equally to this work.

objective of this breed is now twofold: (1) some breeders prefer an animal performing in endurance competitions within other sport disciplines, (2) while other breeders try to select animals to present in morphological shows assessed by expert judges during European Conference of Arab Horse Organization (ECAHO). The selection of the animals has been based on phenotypes and on the experience of breeders. Since 2005, the Spanish Arabian Horse Breeder Association (AECCA) has been developing a breeding program aimed at improving both conformation traits and endurance performance (Valera et al., 2009) and at selecting animals using the breeding values. While this selection depends on individual breeders, a population structure might appear by preferential mating within groups of animals according to different objectives.

The analysis of the genetic structure of a population can be carried out using genealogical or molecular information. Traditionally, the management of genetic variability was done by using genealogical information. But, if the pedigree is absent or incomplete, it would be better to use molecular information to characterize a population. Álvarez et al. (2008), using a Spanish sheep breed as a model, observed that the use of microsatellite markers could be an additional tool for herd management when pedigree is unknown or incomplete. Other authors found similar results in conservation programs (Fernández et al., 2005). Pedigree information is well completed in the Spanish Arabian population, but the parentage control has only been established since 1987 and, moreover, the molecular information provides the additional relatedness between animals appearing as founders in the pedigree. The genetic management of a population might be carried out efficiently and at low cost with pedigree data and molecular markers (such as microsatellites) that might be used to monitor pedigree errors. Nevertheless, the use of pedigree and markers together may be feasible with the development of new methods of wide scale genotyping (Kijas et al., 2009; Meuwissen, 2009) and new generation sequencing (Archibald et al., 2010).

When animals are selected for different breeding goals inside the same population, this could lead to a subdivision and consequently a loss of genetic variability within the subpopulation. Except in the case of extremely numerous livestock populations, subdividing a breed to pursue different objectives causes a substantial loss of efficiency within each sub-set of individual animals. It is preferable, first, to agree on the overall objectives across the breed and then, to try to maintain the variability in the breeding objectives in such a way that the animals are valuable for different users (Verrier, 2011).

Four ancient genetic lines were identified in the formation of the breed, but the presence of a subdivision, due to this fact, is today scarce (Cervantes et al., 2008a). Despite there not being a predefined mating system to attain animals for different performances, in recent years some Spanish Arabian horse breeders have been specializing in breeding morphological show animals (“bred for beauty”) and others have specialized in breeding for sport performance (“bred for endurance”) based on phenotypes. But, since the breeding program was approved for the Spanish Arabian horse the main objective has been to

breed animals with functional conformation. The previous situation needs to be evaluated because the mating system developed by the breeders might have produced a strong differentiation between animals inside breeding objective.

The aims of this study were to determine the differences between Arabian horses bred for different breeding objectives: endurance competitions, morphological shows and other aptitude, to check whether this population structure can be assessed by using genealogical tools, molecular tools or both, and to study if promoting the creation of specialized lines could be the right course of action.

## 2. Material and methods

Genealogical and molecular information was obtained from 120 pure Spanish Arabian horses (50 females, 48 males and 22 geldings), with ages between 3 and 14 years. They were classified in three groups: 45 with morphological aptitude, 49 with endurance aptitude and 26 with other aptitudes. The sample was randomly chosen and was previously used to carry out a morphometric analysis (Cervantes et al., 2009). These animals were bred before the breeding value estimation establishment, and then only phenotypic selection was carried out on them. Classification of the animals within groups was carried out after sampling based on the aptitude in which the horses were bred for, i.e. the breeding goal. For the “morphological” aptitude classification the participating animals were evaluated for their beauty, correctness of legs, Arabic type and basic gaits in morphological shows organized by ECAHO. Participants in endurance races were classified as animals bred for “endurance” aptitude; this is a sport competition where horses have to race over long distances (even 200 km and more) on natural tracks undergoing several veterinary health checks during the race. The third level of the breeding goal was “other” aptitudes including individuals without a clear aptitude, e.g. used for recreational activities.

Some initial analyses were carried out to study the structure of the sampled animals using genealogical and molecular parameters. Finally, an analysis of genetic structure using both type of information source was performed.

### 2.1. Genealogical analyses

The genealogical analyses were based on available pedigree information; the genealogy was traced back to the farer known ancestors for the 120 genotyped animals with a mean of 7.7 equivalent generations known, and a maximum of 13.2. The total number of animals in the pedigree was 1326. Animals with different objectives were fully connected via genealogical information. The following parameters were computed:

*Number of equivalent complete generations (t)* in the pedigree was computed as the sum of  $(1/2)^n$ , where  $n$  is the number of generations separating the individual to each known ancestor (Boichard et al., 1997).

*Effective number of founders (f<sub>e</sub>)*. This parameter is the reciprocal of the probability that two genes drawn at random in the studied population originate from the same

founder (James, 1972) and it is computed from the genetic contribution of founders to the descendant gene pool of the population (Lacy, 1989).

*Effective number of ancestors ( $f_a$ )*. To compute this parameter, the ancestors explaining a percentage of population higher than their parents were identified, and only their marginal contribution that was not explained by other ancestors previously chosen, was considered. This parameter complements the information offered by the effective number of founders accounting for the losses of genetic variability produced by the unbalanced use of reproductive individuals producing bottlenecks (Boichard et al., 1997).

*Number of founder genome equivalents ( $f_g$ )*. Defined as the number of founders that would be expected to produce the same genetic diversity as in the population under study if the founders were equally represented and no loss of alleles occurred (Lacy, 1989). This was computed as the inverse of twice the average coancestry of the individuals within the population (Caballero and Toro, 2000).

*Inbreeding coefficient ( $F$ )* defined as the probability that an individual has two identical genes by descent (Malécot, 1948).

*Average relatedness coefficient (AR) of each individual*, defined as the probability that an allele randomly chosen from the whole population belongs to a given animal (Goyache et al., 2003; Gutiérrez et al., 2003).

*Coancestry coefficient ( $f$ )* computed as the probability that two alleles randomly sampled from two individuals are copies from an allele of a shared ancestor.

*Effective population size*, computed using both the individual increase in inbreeding ( $c$ ) (Gutiérrez et al., 2008, 2009; Cervantes et al., 2008b) and the increase in pairwise coancestry ( $\bar{N}_{ec}$ ) (Cervantes et al., 2011a). The individual increase in inbreeding was defined as  $\Delta F_i = 1 - t_i^{-1} \sqrt{1 - F_i}$ , with  $t_i$  as the number of equivalent complete generations and  $F_i$  as the inbreeding coefficient of an individual  $i$ . The increase in coancestry between any pair of individuals  $j$  and  $k$ , can be computed as  $\Delta c_{jk} = 1 - (t_j + t_k)^{-2} \sqrt{1 - c_{jk}}$ , where  $c_{jk}$  is the inbreeding of a descendant from both, and  $t_j$  and  $t_k$  are the number of equivalent complete generations for the parents. Both parameters take into account the exclusion of self-fertilization. By averaging the individual increase in inbreeding and the increase in pairwise coancestry for all pairs of individuals in a reference subpopulation, effective population size was estimated based on inbreeding  $\bar{N}_e = 1/2\Delta F$  and in coancestries  $\bar{N}_{ec} = 1/2\Delta C$ . The ratio between both effective sizes ( $\bar{N}_{ec}/\bar{N}_e$ ) was also computed in order to ascertain the presence of population structure.

## 2.2. Molecular analyses

To develop the molecular analysis, the DNA was extracted from a sample of the hair root of each individual. These animals were genotyped for 16 microsatellite molecular markers (AHT4, AHT5, ASB17, ASB23, CA425, HMS1, HMS2, HMS3, HMS6, HMS7, HTG10, HTG4, HTG6, HTG7, VHL20 and ASB2) recommended for paternity tests and

individual identification by the International Society for Animal Genetics (ISAG).

Parameters related to the genetic variability and allelic diversity were obtained such as:

*Observed heterozygosity ( $H_o$ )*, computed as the proportion of observed heterozygotes regarding the total number of tested individuals.

*Expected heterozygosity ( $H_e$ )*, computed as the frequency of heterozygotes expected under random mating. We use the following expression:  $H_e = 1 - \sum p_i^2$ , with  $p_i$  the allele frequency (Nei, 1987). Standard error of both heterozygosities were computed using bootstrapping (Gutiérrez et al., 2005).

*Effective allelic number* computed as the inverse of square sum of allelic frequencies (Selander, 1976).

*The rarefacted mean number of alleles* was computed to correct the simple average number of alleles per population by the sample size (Hurlbert, 1971).

*Molecular coancestry* between two individuals  $i$  and  $j$  at a given locus computed using the Caballero and Toro (2002) rules.

To evaluate the concordance of both information resources, the Pearson correlation between coancestries using molecular and pedigree information was also computed.

## 2.3. Population structure analyses

The following parameters were computed with both molecular and genealogical information:

*F-statistics ( $F_{ST}$ ,  $F_{IS}$ ; Wright, 1978)*.  $F_{ST}$  is the average inbreeding of the sub-population relative to the whole population,  $F_{IS}$  is the inbreeding coefficient of the individual relative to its own subpopulation and  $F_{IT}$  is the inbreeding coefficient of the individual relative to the entire population (Falconer and Mackay, 1996). These parameters were computed following Caballero and Toro (2000, 2002) as:

$F_{ST} = (\bar{f} - \bar{f}) / (1 - \bar{f})$ ,  $F_{IS} = (\bar{F} - \bar{f}) / (1 - \bar{f})$  and  $F_{IT} = (\bar{F} - \bar{f}) / (1 - \bar{f})$  where  $\bar{f}$  and  $\bar{F}$  are the mean coancestry and the inbreeding coefficient for the subpopulation. The  $F$  statistics were computed using the same expression for molecular and genealogical approach but, notice that  $\bar{F}$  in the case of the molecular parameters is not the same as genealogical inbreeding, defined as the probability that an individual has two identical alleles by descent (Malécot, 1948), but the homozygosity, referred to the identity by state (Caballero and Toro, 2002).

The  $F$ -statistics were standardized by the sample size when assessed from genealogical information (Bartolomé et al., 2010). For molecular analyses a measure of uncertainty was obtained via bootstrapping (Simianer, 2002; Baumung et al., 2006).

The genetic structure was also analyzed using STRUCTURE software (Pritchard et al., 2000). The best  $K$ -value, corresponding with the number of subpopulation, was calculated from  $\Delta K$ , and based on second order rate changes for the likelihood with respect to  $K$  using equation  $\Delta K = m[L \cdot K] / s[L(K)]$  (Evanno et al., 2005). Also the rate of posterior probability of the data given  $L$ ,  $\Pr(X/K)$ , the

$\ln P(D) (L(K))$  given by the software was used as criteria. Twenty (20) runs were done per each  $k$  (from  $K=2$  to  $K=5$ ).

Parameters assessed from molecular information were obtained using the software MOLKIN 2.0 (Gutiérrez et al., 2005) and those attained from pedigree were computed with ENDOG v.4.8 (Gutiérrez and Goyache, 2005).

### 3. Results and discussion

Since the breeding program was approved for this breed, the global objective of breeding animals with functional conformation was defined by the managers of the breed. However, breeding is in practice carried out by local farmers who could have promoted, or not, the creation of specialized lines using phenotypes. An analysis aimed at discovering a structure for breeding objectives (endurance competitions, morphological shows and other aptitude) was conducted by using a sample of animals previously analyzed from a morphometric perspective (Cervantes et al., 2009). They are assumed to represent the whole Spanish Arabian population in which pedigree analyses were possible (Cervantes et al., 2008a). However, other analyses involving laboratory methods are only feasible with small sample sizes. Here both types of analyses, using genealogical and molecular tools, were carried out on the same animals and a comparison between methodologies was also possible.

#### 3.1. Genealogical analysis

The previous genealogical analysis was a study of the structure of the samples as a representation of the breed. Since the samples were taken at random before the animals were classified within groups, the sample sizes were expected to represent the proportion of the animals

in the breed. Therefore, it seemed that the breed was made up of the following objectives: 37.5% “Morphology”, 40.8% “Endurance” and 21.7% “Other”.

Table 1 shows the main genealogical parameters characterizing the genetic variability in the three groups and in the total samples. All groups showed similar pedigree knowledge between 7.2 and 8 average number of complete equivalent generations. The total number of founders was 178 in the morphology group followed by the value found for the “other” group (146) and the lowest value was for the endurance group (123). The effective number of founders ranged between 25 (“endurance”) and 39 (“morphology”).

Regarding the ancestors they had similar numbers, 60 for the “morphology” and 57 for “endurance” groups but only 40 ancestors were identified for the “other” group. The effective number of ancestors, ranged between 11 (“endurance” and “other”) and 16 (“morphology”), with the highest to the morphology group. The number of founder genome equivalent also showed this loss of genetic diversity, ranging between 7.2 (“morphology”) and 4.7 (“endurance” and “other”). Note that some of these parameters could not be representative of the whole population, since this depends on the sample size. For example, parameters regarding founder genome equivalents could have underestimated the genetic diversity if descendants of some founders with low representation were not sampled at random. Here the samples were taken at random using horses spread throughout Spain and when compared with an analysis of the whole population (Cervantes et al., 2008a), values of  $f_a$  and  $f_e$  for morphology group were more similar to those found in this previous study; 39.5 for  $f_e$  and 13 for  $f_a$  for animals born between 1995 and 2004.

The ratios  $f_e/f_a$  in every group were similar and around 2.3–2.4 lower that the ratio found in the whole population

**Table 1**

Genealogical and molecular parameters characterizing the genetic variability in the Arabian Horse population considering different animal groups for their breeding objectives and for the total population.

	Morphology	Endurance	Other	Total
Number of analyzed individuals	45	49	26	120
Number of founders	178	123	146	221
Effective number of founders	39	25	27	30
Total number of ancestors	60	57	40	88
Number of ancestors explaining 50% of the genetic variability	7	4	4	5
Effective number of ancestors	16	11	11	13
Number of founder genome equivalents	7.2	4.7	4.7	6.2
Realized Effective size based on inbreeding	41.3	30.6	36.4	35.1
Realized Effective size based on coancestry	59.3	39.6	42.2	48.5
Ratio $\bar{N}_{ec}/\bar{N}_e$	1.4	1.3	1.2	1.4
Average inbreeding (%)	7.7	11	9.1	9.4
Average coancestry (%)	7.0	10.6	10.7	8.1
Average number of equivalent complete generations	7.2	8.0	7.8	7.7
Observed Heterozygosity ( $H_o$ )	0.529 ± 0.012	0.537 ± 0.010	0.529 ± 0.015	0.533 ± 0.015
Expected Heterozygosity ( $H_e$ )	0.641 ± 0.014	0.622 ± 0.014	0.615 ± 0.020	0.636 ± 0.009
Effective Number of alleles ( $N_{EA}$ )	3.374	3.259	3.177	3.414
Rarefacted mean number of alleles ( $a_g$ )	6.063	6.500	5.500	7.750
$F_{IS}$ (NS)	0.00773	0.00473	−0.01752	0.01419 ( $F_{IT}$ )
$F_{IS}$ (S)	0.00625	0.00217	−0.00990	0.01233 ( $F_{IT}$ )
$F_{IS}$ (mol)	0.170 ± 0.033	0.150 ± 0.028	0.136 ± 0.039	0.166 ± 0.020 ( $F_{IT}$ )

NS: Genealogical  $F_{IS}$  not standardized by size samples; S: Genealogical  $F_{IS}$  standardized by size samples; mol: Molecular  $F_{IS}$ .



after standardization, which mainly affected the lowest sample sizes. A decrease was observed in  $F_{IT}$  value.  $F_{ST}$  values also had changes after the standardization making the morphological group more distant. The standardized  $F_{ST}$  values ranged between 0.00724 and 0.01123. Although the genetic differentiation is very low, showing that the creation of specialized lines is not producing a subdivision yet. The morphological group is the most distant of the three groups.

### 3.2. Molecular analysis

The values for observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ) and effective allelic number ( $N_{EA}$ ) and rarefacted mean number of alleles ( $a_g$ ) are presented in Table 1 for each group and for the total population.  $H_o$  values ranged between 0.529 and 0.537 and  $H_e$  values between 0.615 and 0.641 where the difference between groups was insignificant. The difference between  $H_o$  and  $H_e$  values suggested the existence of population structure within groups. This difference was reported in other horse populations (Achmann et al., 2004; Azor et al., 2007). For both heterozygosities the values were higher than those found in the Polish Arabian (Glazewska et al., 2004) and lower than in other Arabian Populations (Khanshour et al., 2013) and lower than in other Spanish horse breeds in which the  $H_o$  values ranged between 0.618 for Andalusian Horse and 0.712 for Menorquina horse (Azor et al., 2007). Whereas  $H_e$  values ranged between 0.604 for Mallorquina and 0.725 for Menorquina horse (Azor et al., 2007). Also the values were lower than those found in the Lipizzan horse (Achmann et al., 2004).  $N_{EA}$  ranged between 3.177 and 3.374, and  $a_g$  (between 5.500 and 6.500) being similar across groups. Both parameters characterize the low or high genetic polymorphism richness of the markers used. Although values were very similar in the three groups, the “morphological” group always presented slightly higher values for most of all these parameters, and therefore showing that it retained more genetic variability. Regarding the total population,  $H_o$  was, as expected, the weighted mean of those in the groups according to the animals sampled for each of them. Under no population structure depending on the breeding objective,  $H_e$  might also have

to be intermediate but it was higher surrounding the higher value obtained in the groups revealing the presence of this population structure; even much lower than that present within groups. This is a consequence of the rebalancing of the allelic frequencies when all the groups are gathered into a unique one. This hypothesis of the higher global genetic diversity was endorsed by the observed high increase in  $N_{EA}$  and  $a_g$  parameters.

Table 1 shows the  $F_{IS}$  and Table 2 shows the  $F_{ST}$  parameter (below the diagonal line).  $F_{IS}$  values were higher than those found with genealogical data and positive, showing again that the inbreeding is higher than coancestry within groups (internal structure besides the genetic structure under study). Otherwise,  $F_{ST}$  values were low:  $F_{ST}$  ranged between 0.00378 and 0.01272, showing a scarce genetic differentiation between the created subpopulations compared with the internal structured shown inside the group. Note that this type of markers and the low number of them could not detect drift in specific genomic regions, selective sweeps among groups or increasing or decreasing allele frequencies for specific trait. Whereas it is possible with postgenomic era beadchips, but markers used here are freely available because the parentage control is performed routinely and the results could be useful for breeders. As was demonstrated before in the genealogical analysis, we can also observe with this analysis that, even though the differences between the groups are small and no subdivision is already present because of the specialized lines, the morphological group is slightly the most separated. Regarding the analysis using structure the best K parameter equivalent to the number of subpopulation was 3 (Fig. 1). However, the suggested groups were far from that concerning breeding goals. Also note that the highest  $\Delta K$  value is lower than those found in Evanno et al. (2005) revealing a much weaker structure in the population. Fig. 2 presents the clustering outcomes performed at  $K=3$  (average of 20 replicates). Each color represents one cluster and the length of the colored segment shows individual's estimated proportion of membership in that cluster. All animals seem to be a mix of the three possible breeding goal suggesting that the possible structure of the population is not due to the phenotypic selection made until now. Moreover, in the morphology

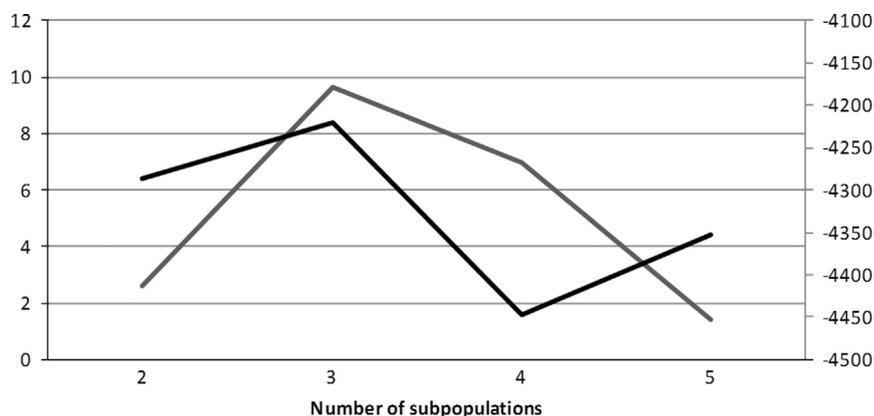


Fig. 1. Representation of  $\Delta K$  calculated based on mean  $L(K)$  (black line, right axis) and second order rate changes for the likelihood with respect to  $K$  using equation  $\Delta K = m[L''K]/s[L(K)]$  (grey line, left axis).

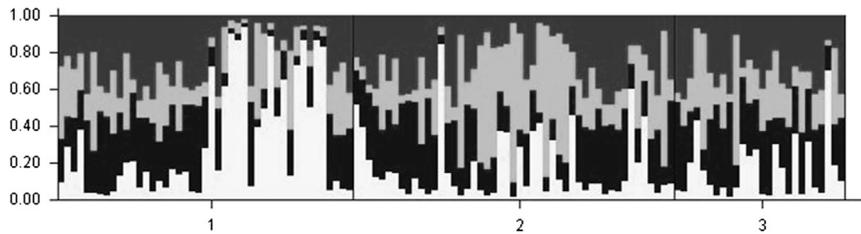


Fig. 2. Graphical representation of the clustering outcomes suggested by the STRUCTURE analysis performed at  $K=3$  (average 20 replicates).

group a light subdivision seems to be present. To confirm that no subdivision is already present due to the specialized lines, Principal Components analyses were carried out trying to identify a sub-structure in the population according to breeding objectives. However, no molecular nor genealogical coancestries showed such genetic structure, but again the most differentiated was the morphological group. The Structure results should be interpreted with caution because of the low number of markers used in the analysis and the high level of mixing in the animals included in this sample. In this sense, seven of the ten most contributing ancestors in each group are common ancestors. Even though we were not able to confirm this hypothesis, the  $K$  value could indicate that another subdivision different from that studied here could be present partially due for example to the four international lines established in the breed (Spanish, Egyptian, Polish and Russian; [Cervantes, 2008](#)).

Both molecular and genealogical analyses were congruent, and both seemed to be valid when studying the genetic structure of this population. All this despite the correlation between genealogical and molecular coancestries was 0.60 which was only moderate.

Finally, both molecular and genealogical analyses would lead us to conclude that differences between the groups assessed here for breeding goals were minimum when compared with the genetic structure within groups. Therefore, a horse with a specific breeding objective is not genetically much differentiated regarding the rest of the objectives. However, the morphological group was the most separated from the rest, both at a genealogical and molecular level. These results were also found when studying the three subpopulations at a morphometric level ([Cervantes et al., 2009](#)). Regarding the possible impact of the subdivision in the population it can be said that no loss of genetic variability is expected in the short-term, because the lines are connected and share some individuals, but some other reason not studied here (e.g. the four international lines mentioned above) seems to exist that tends to subdivide the Arabian population. But more studies are necessary to confirm this and to identify the cause of the subdivision. On the other hand, since the animals are not so specialized, the creation of full lines could improve the genetic progress proposed in the Breeding Program.

#### Conflict of interest statement

All author's declare that there are no known conflicts of interest associated with this publication.

#### Acknowledgments

The authors wish to thank the Spanish Arabian Horse Breeder Association and to the ‘Subsecretaría de Agricultura, Alimentación y Medio Ambiente’ of the Spanish Ministry of Agriculture, Food and the Environment for their support.

#### 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, 285–292.
- Álvarez, I., Royo, L.J., Gutiérrez, J.P., Fernández, I., Arranz, J.J., Goyache, F., 2008. Relationship between genealogical and microsatellite information characterising losses of genetic variability: empirical evidence from the rare Xalda sheep breed. *Livest. Sci.* 115, 80–88.
- Archibald, A.L., Cockett, N.E., Dalrymple, B.P., Faraut, T., Kijas, J.W., Maddox, J.F., McEwan, J.C., Hutton Oddy, V., Raadsma, H.W., Wade, C., Wang, J., Wang, W., Xun, X., 2010. The sheep genome reference sequence: a work in progress. *Anim. Genet.* 41, 449–453.
- Azor, P.J., Valera, M., Gómez, M.D., Goyache, F., Molina, A., 2007. Genetic characterization of the Spanish Trotter horse breed using microsatellite markers. *Genet. Mol. Biol.* 30, 37–42.
- Bartolomé, E., Goyache, F., Molina, A., Cervantes, I., Valera, M., Gutiérrez, J.P., 2010. Pedigree estimation of the (sub) population contribution to the total gene diversity: the horse coat colour case. *Animal* 4, 867–875.
- Baumung, R., Cubric-Curik, V., Schwend, K., Achmann, R., Sölkner, J., 2006. Genetic characterisation and breed assignment in Austrian sheep breeds using microsatellite marker information. *J. Anim. Breed. Genet.* 123, 265–271.
- Boichard, D., Maignel, L., Verrier, E., 1997. The value of using probabilities of gene origin to measure genetic variability in a population. *Genet. Sel. Evol.* 29, 5–23.
- Caballero, A., Toro, M.A., 2000. Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genet. Res.* 75, 331–343.
- Caballero, A., Toro, M.A., 2002. Analysis of genetic diversity for the management of conserved subdivided populations. *Conserv. Genet.* 3, 289–299.
- Cervantes, I., Molina, A., Goyache, F., Gutiérrez, J.P., Valera, M., 2008a. Population history and genetic variability in the Spanish Arab Horse assessed via pedigree analysis. *Livest. Sci.* 113, 24–33.
- Cervantes, I., Goyache, F., Molina, A., Valera, M., Gutiérrez, J.P., 2008b. Application of individual increase in inbreeding to estimate realised effective sizes from real pedigrees. *J. Anim. Breed. Genet.* 125, 301–310.
- Cervantes, I., 2008. Estructura genética del caballo de Pura Raza Árabe español y su influencia en razas derivadas: aplicación de nuevas metodologías en el cálculo del tamaño efectivo (Doctoral thesis). Complutense University of Madrid.
- Cervantes, I., Baumung, R., Molina, A., Druml, T., Gutiérrez, J.P., Sölkner, J., Valera, M., 2009. Size and shape analysis of morphofunctional traits in the Spanish Arab horse. *Livest. Sci.* 125, 43–49.
- Cervantes, I., Goyache, F., Molina, A., Valera, M., Gutiérrez, J.P., 2011a. Estimation of effective population size from the rate of coancestry in pedigreed Populations. *J. Anim. Breed. Genet.* 128, 56–63.
- Cervantes, I., Pastor, J.M., Gutiérrez, J.P., Goyache, F., Molina, A., 2011b. Computing Effective Population Size from molecular data: the case of three rare Spanish ruminant populations. *Livest. Sci.* 138, 202–206.

- Duchev, Z., Distl, O., Groeneveld, E., 2006. Early warning system for loss of diversity in European livestock breeds. *Archiv. Tierzucht.* 49, 521–531.
- 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.
- Falconer, D.S., Mackay, F.C., 1996. *Introduction to Quantitative Genetics*, 4th edition Longman Group Ltd, England.
- FAO, 1998. Secondary guidelines for the national farm animal genetic resources management plans: management of small populations at risk. FAO, Rome, Italy.
- Fernández, J., Villanueva, B., Pong-Wong, R., Toro, M.A., 2005. Efficiency of the use of molecular markers in conservation programmes. *Genetics* 170, 1313–1321.
- Goyache, F., Gutiérrez, J.P., Fernández, I., Gómez, J., Álvarez, I., Diez, J., Royo, L.J., 2003. Using pedigree information to monitor genetic variability of endangered population: the Xalda sheep breed of Asturias as an example. *J. Anim. Breed. Genet.* 120, 95–103.
- Gutiérrez, J.P., Altarriba, J., Díaz, C., Quintanilla, R., Cañón, J., Piedrafitá, J., 2003. Genetic analysis of eight Spanish beef cattle breeds. *Genet. Sel. Evol.* 35, 43–64.
- Gutiérrez, J.P., Goyache, F., 2005. A note on ENDOG: a computer program for analysing pedigree information. *J. Anim. Breed. Genet.* 122, 172–176.
- Gutiérrez, J.P., Royo, L.J., Álvarez, I., Goyache, F., 2005. Molkin v2.0: a computer program for genetic analysis of populations using molecular coancestry information. *J. Hered.* 96, 718–721.
- 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, 359–378.
- Gutiérrez, J.P., Cervantes, I., Goyache, F., 2009. Improving the estimation of realised effective population sizes in farm animals. *J. Anim. Breed. Genet.* 126, 327–332.
- Hurlbert, S.H., 1971. The non concept of species diversity: a critique and alternative parameters. *Ecology* 52, 577–586.
- James, J., 1972. Computation of genetic contributions from pedigrees. *Theor. Appl. Genet.* 42, 272–273.
- Kijas, J.W., Townley, D., Dalrymple, B.P., Heaton, M.P., Maddox, J.F., McGrath, A., Wilson, P., Ingersoll, R.G., McCulloch, R., McWilliam, S., Tang, D., McEwan, J., Cockett, N., Hutton Oddy, V., Nicholas, F.W., Raadsma, H., 2009. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS One* 4, e4668.
- Khanshour, A., Conant, E., Juras, R., Cothran, E.G., 2013. Microsatellite analysis of genetic diversity and population structure of Arabian Horse populations. *J. Hered.* 104, 386–398.
- Lacy, R.C., 1989. Analysis of founder representation in pedigrees: founder equivalent and founder genome equivalents. *Zoo. Biol.* 8, 111–123.
- Magrama, 2011. El caballo de Pura Raza Árabe. ([http://aplicaciones.magrama.es/arca-webapp/flujos.html?\\_flowId=razaCaballar-flow&\\_flowExecutionKey=e6s1](http://aplicaciones.magrama.es/arca-webapp/flujos.html?_flowId=razaCaballar-flow&_flowExecutionKey=e6s1)).
- Malécot, G., 1948. *Les Mathématiques de l'Hérédité*. Ed. Masson et Cie, Paris.
- Meuwissen, T., 2009. Towards consensus on how to measure neutral genetic diversity? *J. Anim. Breed. Genet.* 126, 333–334.
- Nei, M., 1987. In: *Molecular Evolutionary Genetics* Columbia University Press, New York pp. 512.
- Peel, D., Ovenden, J.R., Peel, S.L., 2004. NEESTIMATOR: Software for Estimating Effective Population Size, Version 1.3. Queensland Government: Department of Primary Industries and Fisheries, Brisbane, Queensland.
- Pritchard, J., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Selander, K., 1976. Genetic variation in natural populations. In: Ayala, F.J. (Ed.), *Molecular Evolution*, Sinauer Associates, Sunderland, Massachusetts, pp. 21–45.
- Simianer, H., 2002. Molekulargenetische Differenzierung verschiedener Rotviehpopulationen. Schriftenreihe des Bundesministeriums für Verbraucherschutz, Ernährung und Landwirtschaft. Heft 493. Landwirtschaftsverlag GmbH. Münster-Hiltrup, Germany.
- Valera, M., Bartolomé, E., Cervantes, I., Gómez, M.D., Azor, P.J., Molina, A., 2009. Horse breeding programs in Spain managed by MERAGEM group. Edited by MERAGEM.
- Verrier, E., 2011. L'amélioration génétique: une activité humaine organisée, au service de l'adaptation des populations animales. 37<sup>ème</sup> Journée de la Recherche Equine Jeudi 24 février 2011.
- Wright, S., 1978. *Evolution and the genetics of populations. Variability within and among natural populations*, 4. University of Chicago Press, Chicago, USA.