

ORIGINAL ARTICLE

Genealogical analyses in open populations: the case of three Arab-derived Spanish horse breeds

I. Cervantes^{1,2}, J.P. Gutiérrez¹, A. Molina², F. Goyache³ & M. Valera⁴

1 Departamento de Producción Animal, Facultad de Veterinaria, Madrid, Spain

2 Departamento de Genética, Universidad de Córdoba, Córdoba, Spain

3 SERIDA-Somió, Gijón (Asturias), Spain

4 Departamento de Ciencias Agro-Forestales, EUITA, Universidad de Sevilla, Sevilla, Spain

Keywords

Effective population size; migration; open populations; pedigree analysis; Spanish Arab horse.

Correspondence

I. Cervantes, Departamento de Producción Animal, Facultad de Veterinaria, Avda. Puerta de Hierro s/n, E-28040-Madrid, Spain.

Tel: +34913943767;

Fax: +34913943767;

E-mail: icervantes@vet.ucm.es

Received: 12 July 2008;

accepted: 17 November 2008

Summary

This research assesses the genetic composition of three Arab-derived Spanish horse breeds as an example to highlight the major shortcomings related to genealogical analyses in open populations and to propose approaches useful to deal with this task. The studbooks of three Spanish Arab (SA)-derived horse breeds, Spanish Anglo-Arab (dAA), Hispano-Arab (dHA) and Spanish Sport Horse (dSSH) and those of their parental breeds SA, Spanish Purebred (SPB) and Thoroughbred (TB), totalling 211 754 individuals, were available. The genealogies of the dAA, dHA and dSSH were analysed not only using the corresponding studbook (breed exclusive dataset) but also including the genealogies of the founders from parental breeds (completed dataset). Coancestry analyses revealed that the present SA-derived populations share more genes with the Arab than with the other parental breeds. Effective population size was computed by accounting for migration rates to obtain an equivalent closed-population effective size ($_{eq}N_e$) of 39.2 for the dAA, 56.3 for dHA and 114.1 for dSSH. The essayed methodologies were useful for characterising populations involving migration. The consequences of the management of the analysed breeds are discussed. The results emphasize the need to include the complete genealogies of the individuals to attain reliable genealogical parameters.

Introduction

The Spanish Arab (SA) horse breed studbook was founded in 1847 by the Spanish Ministry of War (Maxwell 1995). Historically, crossbreeding to obtain individuals with the desired performance characteristics was the major purpose of the Arab horses in Spain (Maxwell 1995). This is the basis of the development of a number of SA-derived horse breeds. Two major Spanish horse breeds, Anglo-Arab (dAA) and Hispano-Arab (dHA) constituting 5.3% of the national equine census (MAPyA, 2003), were formed by continuous cross breeding between SA

horses and Thoroughbred (TB) and Spanish purebred (SPB; Andalusian) horses respectively. Their studbooks were founded respectively in 1884 and 1986, remaining open to individuals with an Arab parent. The Anglo-Arab is one of the most ancient Spanish studbooks. Despite this there is a consensus on the existence of the Hispano-Arab horse population early in the 20th century (Steen 2006), genealogical recording began recently with the foundation of its studbook. This breed in Spain is officially considered as an endangered breed (Spanish regulation RD 1662/1997). Additionally, the development of the Spanish equine industry (Valera *et al.* 2005; Azor

et al. 2007; Cervantes *et al.* 2008a) stimulated by the increasing use of horses for sports and recreation, has led to the creation of a new breed, the Spanish-Sport horse (dSSH; studbook founded in 2002), with some genetic influence from the SA breed. Despite the importance of this scenario, no analysis on the genetic influence of the SA horse on the Arab-derived breeds has been carried out. Furthermore, the problems regarding inbreeding, population size or crossbreeding in SA-derived breeds were unknown until this study.

The genealogical analyses of open populations, allowing the use of individuals from other breeds for reproduction, present particular characteristics that make it difficult to ascertain the real rates of genetic variability in the analysed stock (Hamann & Distl 2008). The analysis of populations receiving genetic contributions from other populations over time needs to account for migration rates to assess their genetic variability in terms of inbreeding and effective size (Falconer & Mackay 1996). Moreover, the effect of the mating policy on the genetic variability of an open population cannot be straightforwardly assessed via the computation of the effective population size because this parameter is influenced by rates of migration from external populations.

In this respect, in order to ascertain the genetic influence of the source populations in terms of founder contributions, one needs to trace genealogies back using pedigrees from pure studbooks of origin to identify founders with certainty. However, the impact of ignoring those genealogies from pure breeds is usually unknown (Hamann & Distl 2008).

The aim of this research was to assess the genetic composition of three SA-derived Spanish horse breeds and the relationships and the differentiation between them using genealogical information, addressing the consequences of ignoring the parental pedigrees, as well as developing and essaying methodologies to manage populations submitted for migration. To deal with these objectives the complete Spanish horse studbooks for the SA, SPB, TB, dHA, dAA and dSSH horse breeds were used. The consequences of the obtained results for the management of the analysed breeds are discussed.

Material and methods

Pedigree data

Information registered in the studbooks (from their foundation to December 2004) of the Spanish Anglo-Arab, Hispano-Arab and Spanish Sport horse

breeds, as well as the full genealogies of the SA, SPB and Spanish TB horse breeds, were obtained from the official web site of the Spanish horse studbook (Ministry of Defense: <http://www.librogenealogico.com>), and used in this analysis. The total number of animals was: 8289 for dAA, 3394 for dHA, 7099 for dSSH, 18 880 for SA, 140 629 for SPB and 33 463 for TB. The records of animals born per period of 5 years from the dAA, dHA and dSSH studbooks are summarized in Table 1.

Note that a given SA individual can act as parent in each of the dAA, dHA and dSSH studbooks. Furthermore, dSSH is a composite breed that includes in its studbook any individual participating in sports competitions and that is not registered in any other studbook and, therefore, a given SA, SPB, TB, dAA or dHA individual can act as parent in the dSSH studbook. The available information was edited to avoid redundancies and to make the identification of the same individual uniform regardless of the breed exclusive database in which it is recorded and to identify with certainty the breed of each individual.

Throughout the manuscript the dAA, dHA and dSSH breeds will be referred to as 'derived' populations, while the SA, SPB and TB breeds will be considered 'parental' populations.

Table 2 shows the number of individuals acting as parents in the breed exclusive datasets (those directly obtained from the SA-derived studbooks) of those derived by parental horse breeds. A total of 21.4% of individuals which acted as parents were SA and 26.3% were TB in the dAA studbook; a total of 28.3% for SA and 42.2% for SPB individuals acted as parents in the dHA studbook and finally, a total of 4.4% SA, 19.1% SPB, 10.3% TB, 5.3% dAA, 1.6% dHA and 36.6% of individuals from other breeds acted as parents in the dSSH studbook.

Within pedigrees, a group of animals approaching the last generation were defined as 'reference populations' within each population. For this purpose, the average generation length (GL), defined as the average age of parents at the birth of their offspring kept for reproduction, was computed for each derived population using the program ENDOG (Gutiérrez & Goyache 2005). According to average GLs (12.0 ± 0.13 years for dAA, 10.1 ± 0.16 years for dHA and 11.4 ± 0.21 years for dSSH), the reference populations finally included the animals born in the last 12 years for the dAA breed, 10 years for dHA and 11 years for dSSH.

Table 2 also gives the number of stallions and mares acting as parents of the reference populations. In dAA the highest proportion of mares (46.8%)

Table 1 Number of available records included in the Spanish Anglo-Arab (dAA), Hispano-Arab (dHA) and Spanish-Sport horse (dSSH) studbooks per 5 year periods

	<1925	1925–1929	1930–1934	1935–1939	1940–1944	1945–1949	1950–1954	1955–1959	1960–1964	1965–1969	1970–1974	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999	2000–2004	total
dAA	445	79	83	43	111	245	383	261	429	436	515	680	588	503	820	1286	1382	8289
dHA							3		1	7	28	91	214	393	874	921	862	3394
dSSH									6	2	13	24	149	226	624	1618	4437	7099

were dAA, while for stallions these were SA (44.7%); in dHA, the highest proportion of mares (50.4%) were dHA and for stallions these were SPB (66.3%); with respect to dSSH, the highest proportion of mares were dSSH (36.7%) and for stallions these were SPB (39.9%).

The pedigree information summarized in Table 2 was used to include all the available genealogies of the parents identified from the SA, SPB and TB studbooks in the dAA, dHA and dSSH datasets. The datasets were completed with parental genealogies given that, in the breed exclusive datasets, the genealogies were truncated where an ancestor from a parental breed appeared. After including information from the parental studbooks the available records totalled 16 304, 10 605 and 21 826 respectively, for dAA, dHA and dSSH. Throughout the manuscript, the information contained in dAA, dHA and dSSH studbooks without recording 'parental genealogies', will be referred to as 'breed exclusive datasets' and after including the information from the parental studbooks they will be referred to as 'completed datasets'. Note that, regardless of whether the analysed dataset was breed exclusive or completed, those individuals with both parents unknown were considered as founders following Moureaux *et al.* (1996), Valera *et al.* (2005) and Cervantes *et al.* (2008a). As a consequence, most individuals that were considered founders in the breed exclusive datasets were not considered as such in the completed datasets. Note also that the number of animals acting as parents, in the reference populations are defined the same before and after completing the genealogies. Additionally, according to Lacy (1989) and Cervantes *et al.* (2008a), when only one parent is known, the unknown parent was considered a founder.

The contributions of founders within breed of origin were summed up to calculate the influence of the different pure breeds. When a founder appeared assigned to its own studbook instead a parental breed, it was assumed that it had a parental composition equal to the average contribution of each parental population to the corresponding derived breed.

Statistical analyses

A number of population and genetic parameters were computed on the reference populations using the program ENDOG (current version v4.5; Gutiérrez & Goyache 2005).

The pedigree completeness level was computed for both the breed exclusive and the completed datasets

Table 2 Number of individuals acting as parents in the breed exclusive datasets of the Spanish Anglo-Arab (dAA), Hispano-Arab (dHA) and Spanish-Sport horse (dSSH) studbooks, by parental horse breed

Parental Breed	Whole pedigrees			Reference populations					
	dAA	dHA	dSSH	dAA		dHA		dSSH	
				Stallion	Mare	Stallion	Mare	Stallion	Mare
Spanish Arab (SA)	1016 (21.4)	644 (28.3)	255 (4.4)	346 (44.7)	225 (15.1)	182 (25.5)	269 (30.3)	171 (9.7)	47 (1.5)
Spanish Purebred (SPB)		959 (42.2)	1102 (19.1)			474 (66.3)	172 (19.3)	703 (39.9)	278 (8.9)
Thoroughbred (TB)	1253 (26.3)		595 (10.3)	166 (21.5)	567 (38.1)			156 (8.8)	315 (10.1)
Anglo-Arab (dAA)	2486 (52.3)		307 (5.3)	261 (33.8)	698 (46.8)			103 (5.8)	149 (4.8)
Hispano-Arab (dHA)		669 (29.5)	91 (1.6)			59 (8.2)	448 (50.4)	26 (1.5)	61 (2.0)
Spanish-Sport (dSSH)			1301 (22.6)					203 (11.5)	1144 (36.7)
Others ^a			2110 (36.6)					400 (22.7)	1122 (36.0)

Whole pedigrees included all animals in the dataset and reference populations included animals born in the last 12 years for the dAA breed, 10 years for dHA and 11 years for dSSH. In brackets, % with respect to the total parents, total stallions and total mares, respectively.

^aAll parents belonging to non-listed breeds.

as the proportion of ancestors known per parental generation (MacCluer *et al.* 1983). Also, the number of equivalent to discrete generations (t) for each individual in a pedigree was computed as the sum of $(\frac{1}{2})^n$, where n is the number of generations separating the individual to each known ancestor (Boichard *et al.* 1997).

The inbreeding coefficient (F), defined as the probability that two alleles at a randomly chosen locus are identical by descent (Malécot 1948), and the average relatedness coefficient (AR), 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), were computed for each individual included in the breed exclusive and the completed datasets. Note that the AR coefficient of a founder means its genetic contribution to the population. For each reference population, these coefficients can be summed up for the founders belonging to a parental breed to ascertain the relative contributions from parental breeds.

The probability of gene origin was characterized by computing the following parameters: (i) effective number of founders (f_c), which is the reciprocal of the probability that two alleles drawn at random in the studied population originate from the same founder (James 1972) and this is computed from the genetic contribution of founders to the descendant gene pool of the population (Lacy 1989); (ii) effective number of ancestors (f_a), defined as the minimum number of ancestors, not necessarily founders, explaining the complete genetic diversity of a population (Boichard *et al.* 1997). Parameter f_a does not fully account for gene loss by drift from the ancestors to a reference population but complements the

information offered by f_c accounting for the losses of genetic variability produced by the unbalanced use of reproductive individuals producing bottlenecks (Gutiérrez *et al.* 2005); (iii) the founder genome equivalents (Ballou & Lacy 1995), defined as the theoretically expected number of founders that would be required to provide the genetic diversity in the actual population if the founders were equally represented and had lost no alleles, was obtained by the inverse of twice the average coancestry of the individuals within the population (Caballero & Toro 2000).

Effective population size (N_e) was computed following Gutiérrez *et al.* (2003) approaching the increase in inbreeding (ΔF) as the regression coefficient (b) of the individual inbreeding coefficient over the equivalent discrete generations (t), and considering the corresponding regression coefficient as the increase in inbreeding between two generations, ($b = F_t - F_{t-1} \approx \frac{F_t - F_{t-1}}{1 - F_{t-1}}$), and consequently $N_e = 1/(2b)$.

The migration rate (m) was computed for the breed exclusive dataset as the ratio between the number of animals from the parental populations acting as parents over the whole number of parents (from the parental population or not) identified in the reference population.

Assuming that inbreeding remains stable as a consequence of the balance between migration and drift in open populations, the equivalent to closed-population effective size ($_{eq}N_e$) was approached in the breed exclusive datasets from $F \cong \frac{1}{4m_{eq}N_e + 1}$ (Falconer & Mackay 1996), where F is the average inbreeding in the reference population in which mean inbreeding remains constant and m is the migration rate.

Using this expression, we assume that a particular derived breed behaves as a population under the island migration model (Wright 1931). The parameter $e_{eq}N_e$ can only therefore be computed when no significant regression of F over the year of birth for the animals existed in the reference population and can be interpreted as the N_e corresponding to a population with the same mating policy as the analysed population but without migration. Note that the formula $F \cong \frac{1}{4m_{eq}N_e+1}$ is just a case (assuming that mutation is negligible) of the more general formula $F \cong \frac{1}{4e_{eq}N_e(u+v+m)+1}$ where u and v are the mutation rates in two directions between two alleles at a locus. The expression was developed to study each subpopulation under the island model, assuming no relationships among migrants, and therefore applies for the last purpose to our breed exclusive datasets in which migrants appear as founders. Note again that the $e_{eq}N_e$ would give information on the expected N_e of a population with the same breeding policy as that of the studied breed but without migration, which would allow comparisons of mating policies with other populations via N_e .

Finally, the genetic relationships and differentiation between all reference populations composed by the last generation of individuals belonging to the parental and the derived populations (SA, SPB, TB, dAA, dHA and dSSH) were assessed from a database that included all the records of the six available studbooks. Here the relationship between animals previously considered as migrants is now taken into account. The coefficient of coancestry of two individuals is defined as the probability that two gametes taken at random, one from each, carry alleles that are identical by descent (Falconer & Mackay 1996). The within-breed coancestry (f_{ii}) and the between-breeds coancestry matrix (f_{ij}) were computed averaging all pairwise coancestry coefficients of the individuals belonging, respectively, to a given breed i or to two different breeds i and j . Following Caballero &

Toro (2000, 2002) the between-breeds Nei's minimum distance (D_m) matrix was also computed as $D_m = [(f_{ii} + f_{jj})/2] - f_{ij}$ where f_{ii} and f_{jj} are the average coancestry within two breeds i and j and f_{ij} the coancestry between two breeds i and j .

Results

Pedigree completeness

The quality of the pedigree information of the individuals included in the reference population is illustrated in Figure 1. The inclusion of the available pedigrees obtained from the parental population pedigrees significantly increased the completeness of the genealogies of the dAA, dHA and dSSH breeds. The breed exclusive datasets had a completeness of 51.2% for the dAA breed, 30.3% for dHA and 24.7% for dSSH in the second parental generation, while these figures in the completed pedigrees were, respectively, 98.8%, 91.4% and 59.6%. The highest effect on pedigree knowledge was observed for the dHA breed because of the influence of the SPB genealogies: the completed dataset of the dHA breed has levels of completeness higher than 60% up to the eighth parental generation (Valera *et al.* 2005). The average number of equivalent to complete discrete generation gave similar information. The breed exclusive datasets of the dAA, dHA and dSSH breeds had average number of equivalent to discrete generations (t) of 2.54 for dAA, 1.35 for dHA, and 1.23 for dSSH, while the figures corresponding to the completed datasets were, respectively, 5.60, 7.71 and 3.35.

Probability of gene origin

Parameters characterizing the genetic variability of the identified reference populations in terms of gene origin are given in Table 3. A total of 1942, 1417 and 4966 founders were identified in the breed

Figure 1 Percent of ancestors known per parental generation in Spanish Anglo-Arab (in squares), Hispano-Arab (in circles) and Spanish Sport horse (in triangles), with parental generation 1 that corresponds to parents, 2 that corresponds to grandparents, etc. Results from the breed exclusive datasets are in open symbols, while those from the completed datasets after including information from the parental studbooks are in black symbols.

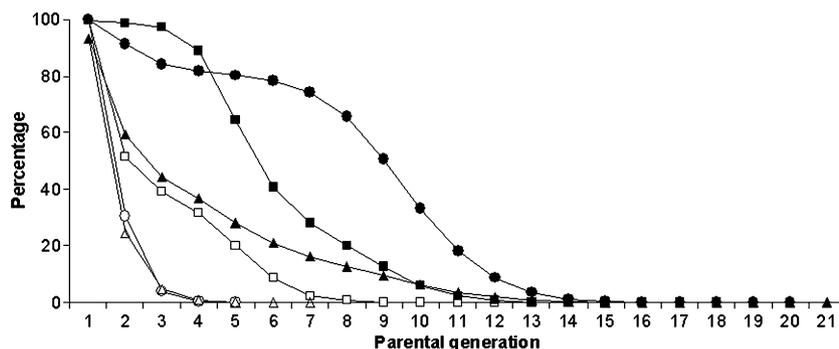


Table 3 Parameters characterising the probability of gene origin of the animals included in the reference populations, animals born in the last 12 years for the Spanish Anglo-Arab, 10 years for Hispano-Arab and 11 years for the Spanish-Sport horse breeds using the breed exclusive and the completed pedigree datasets

	Spanish Anglo Arab		Hispano Arab		Spanish Sport Horse	
	Breed exclusive ^a	Completed ^b	Breed exclusive ^a	Completed ^b	Breed exclusive ^a	Completed ^b
Total number of animals in the reference population	3071	3071	1783	1783	6198	6198
Animals with both parents unknown	1883	2228	1377	670	4467	3889
Total number of founders	1942	2414	1417	860	4966	5655
Effective number of founders	433	210	327	87	762	602
Effective number of ancestors	241	98	311	41	736	436
Founder genomes equivalents	163	50	263	18	655	255
Ancestors explaining 100% of the genetic variability	1569	1202	1215	514	4272	4032
Ancestors explaining 50% of the genetic variability	130	44	203	16	638	284

^aUsing breed exclusive dataset.

^bUsing completed dataset after including information from the parental studbooks.

exclusive datasets of the dAA, dHA and dSSH breeds respectively. However, the complete datasets allowed us to identify a different scenario with 2414 founders for the dAA breed, 860 founders for dHA and 5655 founders for dSSH. The substantial decrease of total number of founders (39%) found in the completed pedigree of dHA, with respect to those identified in the breed exclusive pedigree, would characterize the dHA breed as a product of a limited number of families from the parental breeds thus limiting the genetic variability of the population.

The effective number of founders computed from the completed datasets were 48% for dAA, 27% for dHA and 79% for dSSH of those obtained using the breed exclusive datasets. In turn, the effective number of ancestors computed from the completed

datasets were 41% for dAA, 13% for dHA and 59% for dSSH from those obtained using the breed exclusive datasets. And regarding the founder genome equivalents these values become 31% for dAA, 7% for dHA and 39% for dSSH.

Table 4 details the contributions to the derived breeds of the founders belonging to the paternal breeds as identified in the completed datasets. The contributions of the SA breed to its derived dAA and dHA breeds are lower than 50% and tended to be higher on the mares path. The lowest contribution of the SA founders was found in the dSSH breed (below 10%). In the dSSH breed, the contributions of the founders belonging to the SPB and TB breeds, reached values of roughly 20% each. However, most contributions (higher than 50%) were not from the

Table 4 Contribution of the founders from the paternal populations to the derived Spanish Anglo-Arab (dAA), Hispano-Arab (dHA) and Spanish-Sport horse (dSSH) populations for the whole pedigree (WP, including all animals of the dataset) and for the reference populations (RP, including animals born in the last 12 years for the dAA breed, 10 years for dHA, and 11 years for dSSH) available from the breed exclusive datasets

	Spanish Arab			Thoroughbred			Spanish Purebred			Others		
	Male	Female	Total	Male	Female	Total	Male	Female	Total	Male	Female	Total
dAA												
WP	17.8	22.2	40.0	37.1	22.9	60.0						
RP	24.7	16.7	41.4	30.6	28.0	58.6						
dHA												
WP	20.4	26.2	46.6				33.5	19.9	53.4			
RP	22.8	25.6	48.4				34.7	16.9	51.6			
dSSH												
WP	6.1	2.9	9.0	10.0	9.0	19.0	14.9	6.0	20.9	19.3	31.8	51.1
RP	6.1	2.7	8.8	10.9	8.1	19.0	13.2	4.8	18.0	25.2	29.0	54.2

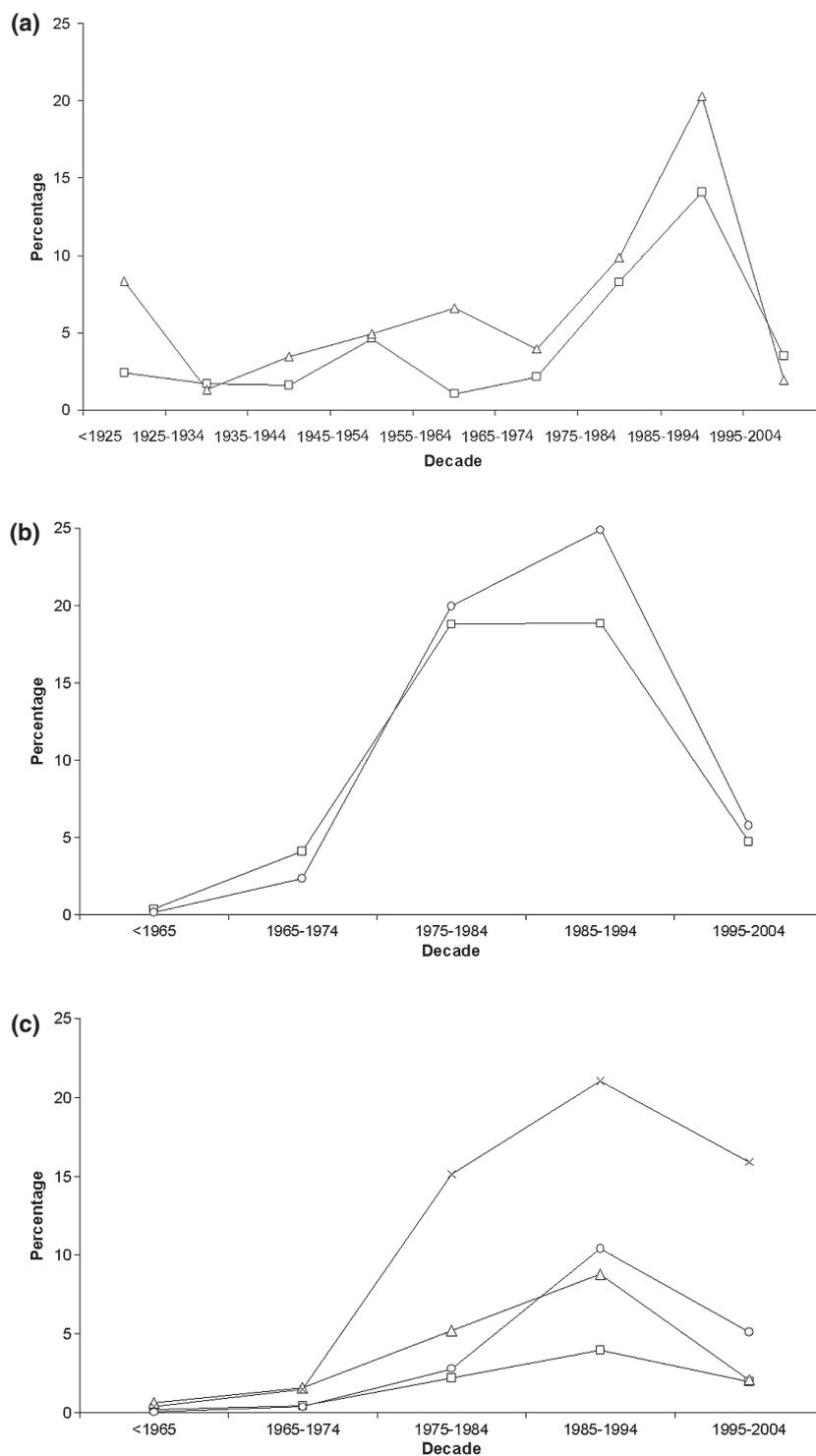


Figure 2 Contribution of the founders from the parental breeds to the reference populations (animals born in the last 12 years for the Spanish Anglo-Arab, 10 years for the Hispano-Arab and 11 years for the Spanish Sport horse) of the Spanish Anglo-Arab breed (Plot A), Hispano-Arab breed (Plot B) and Spanish Sport horse breed (Plot C), by decade of birth of the founders. Contributions of the Spanish Arab (□), Thoroughbred (Δ), Spanish Purebred (○) and from non-listed breeds (x) founders are represented.

populations assumed here to be parental: SA, SPB and TB.

Figure 2 shows the contributions of the founders of the parental breeds to the reference populations defined in the dAA, dHA and dSSH pedigrees by

date of birth of the founders in order to illustrate the genetic history of the derived breeds. The SA influence on both the dAA and dHA breeds was always lower than that from TBs and SPBs respectively. This was basically due to founders born between 1975

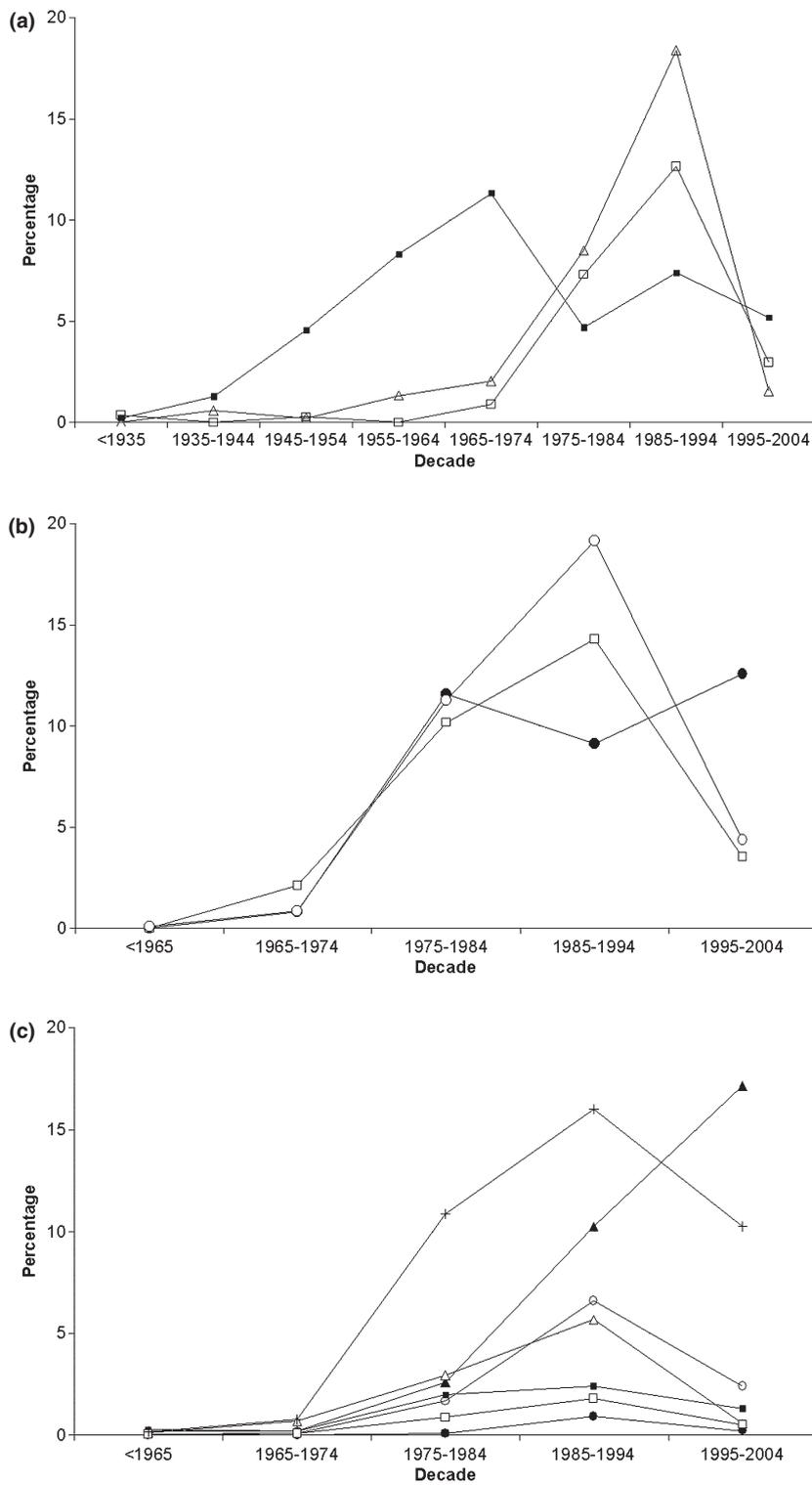


Figure 3 Contribution of the Boichard *et al.*'s (1997) ancestors to the reference populations (animals born in the last 12 years for the Spanish Anglo-Arab, 10 years for the Hispano-Arab and 11 years for the Spanish Sport horse) of the Spanish Anglo-Arab breed (Plot A), Hispano-Arab breed (Plot B) and Spanish-Sport horse breed (Plot C) by decade of birth of the ancestors. Contributions of the Spanish Arab (□), Thoroughbred (Δ), Spanish Purebred (○), Spanish Anglo-Arab (■), Hispano-Arab (●), Spanish-Sport horse (▲) and from non-listed breeds (x) ancestors are represented.

and 1994 (Plot 2A and 2B). The SA contribution to the dSSH breed (Plot 2C) was low and remained constant over the period analysed (over 2%). However, major founder contributions to the dSSH

breeds came from breeds that were not included among the three parental populations analysed here.

Figure 3 shows the contributions of ancestors to the reference populations defined in the dAA (Plot

Table 5 Effective population size computed from regression coefficient, migration rate (m) and equivalent to closed-population effective size ($_{eq}N_e$) computed for the reference populations (animals born in the last 12 years for the dAA breed, 10 years for dHA and 11 years for dSSH) using the breed exclusive and the completed pedigrees

	Spanish Anglo-Arab		Hispano-Arab		Spanish-Sport Horse	
	Breed exclusive dataset ^a	Completed dataset ^b	Breed exclusive dataset ^a	Completed dataset ^b	Breed exclusive dataset ^a	Completed dataset ^b
N_e	64.5	109.5	16.2	-	47.0	135.5
m	0.58		0.68		0.72	
$_{eq}N_e$	39.2		56.3		114.1	

^aUsing breed exclusive dataset.

^bUsing completed dataset after including information from the parental studbooks.

3A), dHA (Plot 3B) and dSSH (Plot 3C) pedigrees by date of birth of the ancestors. In dAA major genetic contributions came from ancestors belonging to their own derived breed (42.9%), with the contributions of the TB ancestors being higher (32.6%) than that from the SA ancestors (24.5%). In the dHA breed the ancestors belonging to the SA and SPB breeds gave similar genetic contributions (30.2% and 35.7% respectively). The contributions of parental ancestors to both the dAA and the dHA population were basically given by individuals born during the period 1985–1994. In dSSH, the genetic contributions of the ancestors belonging to the SA were stable and very low (3.3%). In this breed, major contributions came from non-listed parental breeds and, increasingly, from dSSH ancestors (30.4%).

Inbreeding, average relatedness and effective population sizes

Inbreeding values of the individuals included in the reference populations were computed using both the breed exclusive and the completed pedigrees. Mean inbreeding values for the breed exclusive pedigrees were 1.09% for dAA, 0.64% for dHA and 0.30% for dSSH, while these figures were, respectively, 1.41%, 1.91% and 0.66% for the completed pedigrees. The mean AR coefficients, characterising the genetic identity of the individuals included in the reference populations of the dAA, dHA and dSSH breeds were 0.38%, 0.22% and 0.40% for the breed exclusive pedigrees and 1.21%, 4.62% and 0.41% for the completed pedigrees.

Table 5 gives the effective populations sizes computed from regression of the individual inbreeding coefficient over the equivalent to discrete generations in the reference population both in breed exclusive and completed datasets. The N_e values computed from regression coefficients were higher for the completed datasets in dAA and dSSH (109.5 versus 64.5 and 135.5 versus 47.0 respec-

tively). N_e computed from the dHA breed exclusive dataset was 16.2, while that computed from the complete dataset had a meaningless negative value due to a decrease in the inbreeding coefficient in the last generation.

After checking that there was no significant regression of F over the year of birth for the animals included in the reference population to ensure that there was no trend in inbreeding, the equivalent to closed-population effective size ($_{eq}N_e$) was computed using the breed exclusive pedigrees, in order to compare the results to closed populations in terms of breeding policy. These values are shown in Table 5 jointly with the migration rates in each breed. The N_e values obtained from regression coefficients in completed datasets were higher than those computed for $_{eq}N_e$. After accounting for migration rates, the computed $_{eq}N_e$ values were 39.2 for dAA, 56.3 for dHA, and 114.1 for dSSH.

Between-populations differentiation

Table 6 gives the Nei's minimum distance (D_m) values and pairwise coancestry coefficients (f_{ij}) between the animals born in the last decade of the six analy-

Table 6 Within-breeds coancestry (on diagonal), between-breeds coancestry matrix (above diagonal) and between-breeds Nei's minimum distance matrix (below diagonal) computed using the whole available pedigrees from the individuals born in the last decade in the Spanish Purebred (SPB), Spanish Arab (SA), Spanish Thoroughbred (TB), Spanish Anglo-Arab (dAA), Hispano-Arab (dHA) and Spanish-Sport horse (dSSH)

	SA	SPB	TB	dAA	dHA	dSSH
SA	0.065	0.000	0.000	0.023	0.032	0.005
SPB	0.071	0.076	0.000	0.000	0.028	0.010
TB	0.041	0.035	0.005	0.001	0.000	0.000
dAA	0.021	0.038	0.007	0.010	0.010	0.002
dHA	0.020	0.018	0.016	0.009	0.027	0.006
dSSH	0.035	0.024	0.004	0.005	0.009	0.003

sed breeds: SA, SPB, TB, dAA, dHA and dSSH. The highest differentiation was assessed between the two parental breeds with deeper pedigrees (SA and SPB; $D_m = 0.071$), while the lowest was between the pair TB-dSSH (0.004). With respect to the derived breeds, the SA breed showed distance values varying from 0.020 (SA-dHA) to 0.035 (SA-dSSH). The coancestry assessed between SA and the derived breeds varied between 0.005 for the pair SA-dSSH and 0.032 for the pair SA-dHA. These values are in the range of the coancestry between SPB and each of dHA and dSSH (0.028 and 0.010 respectively) but are higher than those computed between TB and each of dAA and dSSH (0.001 and 0.000 respectively).

Discussion

Genealogical analyses in open populations have major shortcomings that make it difficult to assess the genetic variability of the stock and make inferences useful for the management of the population. Here we apply different approaches to deal with populations involving migration and to assess the genetic composition of three SA-derived Spanish horse breeds. Overall, the analyses carried out here clearly point that the information gathered in the breed exclusive pedigrees is not useful to describe the genetic scenarios of the derived populations. Differences between using the breed exclusive pedigree or the completed genealogy are a realisation of the problems of accounting for more or less deep genealogies. Overall, these figures illustrate how many between-individuals genetic relationships are lost in pedigrees without information on parental breeds. This is true for the dSSH breed regardless of whether all the parental genealogies are available but is especially true for the dHA breed in which a dramatic reduction of the genetic stock can be assessed when the full genealogies of the registered individuals are considered.

One can think such a scenario can be approached using techniques aimed at inferring pedigree information from incomplete pedigrees. There are methods that try to solve this problem based on the uncertain-parentage-matrix (Pérez-Enciso 1995) that considers a probabilistic correction for unknown ancestors. This method has been used very recently in Trakehner horse, but no major differences were found when applied to genealogies with missing ancestries (Teegen *et al.* 2009). Additionally, approaches to recover the inbreeding caused by incomplete pedigrees have been developed by VanRaden (1992), which assumes that animals with

missing parents have inbreeding coefficients equal to the mean of the inbreeding coefficients for animals with known parents born during the same year. Recently, Aguilar & Misztal (2008) have proposed a recursive algorithm to reduce the computing time of VanRaden's method. This approach seemed to work well only when the missing dam information was not more than 10–20% (Lutaaya *et al.* 1999). Therefore, such techniques should not be advantageous in scenarios of open populations but in closed pedigrees with any efficient genealogies recording. Moreover, such approaches do not solve the problem of accurate assessment of the genetic contributions from a parental breed to those derived breeds.

Additionally, the possibilities of using the new information obtained in the present analysis to characterise the breeding policies of the Anglo-Arab, Hispano-Arab and Sport-Spanish horse breeds can be a point of discussion. Historically, the Arab horse has been used to improve the conformation and performance characteristics of other horse populations (Bowling & Ruvinsky 2000a) contributing to a large extent to the formation of breeds such as the TB (Bowling & Ruvinsky 2000b) or the Lipizzan (Zechner *et al.* 2002). The Spanish breeds derived from the Arab horse are: dAA, dHA and dSSH. The introgression of SA individuals into these derived breeds is significant and, in most cases, balanced between stallions and mares (see Table 4). The use of reproductive SA individuals in the derived breeds has been continuous up to the present day and, in fact, the derived breeds cannot be considered as separated genetic entities. Major SA founders and ancestors identified for the derived breeds (especially for dAA and dHA) were born recently between 1975 and 1994 and no unbalanced use of reproductive individuals belonging to the derived breeds was detected until recently (see Figures 2 and 3).

Completeness and gene origin

Any attempt to compare the parameters characterising gene origin of the analysed breeds with the results available in the literature is not straightforward because the latter are computed on closed horse populations with deep studbooks (see Valera *et al.* 2005 and Zechner *et al.* 2002 for reviews). Also, this kind of comparison must be interpreted with caution as these parameters depend on the time from foundation of the studbook and the quality of pedigree data recording. Recently, Cervantes *et al.* (2008a) reported values of total number of founders, effective number of founders and effective number

of ancestors in the SA breed of, respectively, 860, 39.5 and 13. The SA breed is known to have a deep studbook founded in the 19th century that remains open for the inclusion of Arab individuals from foreign countries. However, the figures corresponding to the breeds analysed here are higher than those reported for the SA breed (Cervantes *et al.* 2008a) with the rough exception of the completed dataset of the dHA breed that have a considerably higher pedigree depth than the others. Differences found for the dSSH breed can be explained by the recent foundation of the studbook and the lower quality of the available pedigree data. However, differences between dAA and dHA are mainly due to the higher quality of the genealogies available for the SPB ancestors of dHA individuals.

Inbreeding, Average Relatedness and effective population size

Both the average inbreeding and AR computed for the reference populations defined in the pedigrees analysed here are highly affected by the degree of completeness of the genealogies. Some of the individuals that appeared as a founder in the breed exclusive dataset are heavily inbred. In open populations it seems to be not necessary the control in the breeding policy; a non-inbred offspring will be obtained from two unrelated highly inbred parents. However, mating among progeny in later generations from related founder animals may increase inbreeding in very short time and reduce effective population size. Mean F computed using the completed datasets is from 30% (dAA) to 300% (dHA) higher than that assessed using the breed exclusive datasets. The mean ARs (which is roughly twice the expected average F in the case of random mating, Goyache *et al.* 2003; Gutiérrez *et al.* 2003; Gutiérrez & Goyache 2005) increase dramatically in the breeds with deeper pedigrees (dAA and dHA), thus showing that breeders plan matings in these breeds using related parental individuals, especially in dHA, because of the between-individuals relationship existing in the SPB studbook (Valera *et al.* 2005). This is probably happening because breeders ignore that animals from parental breeds are highly related as they do not have access to the complete dataset.

The N_e computed with the breed exclusive pedigrees had values of 64.5, 16.2 and 47.0 for, respectively, the dAA, dHA and dSSH breeds (Table 5). However, these figures do not address the 'real' N_e because they ignore the relationships between individuals appearing as founders, as well as the breed-

ing policy carried out in the parental breeds. These effective population sizes were computed using a regression approach that has been shown to fit properly with real pedigrees with a stable increase in inbreeding (Gutiérrez *et al.* 2003). However, in the scenario analysed here, the computed values are highly affected by migration and mating policies. This is especially true in the case of short pedigrees in which a small number of individuals with more complete pedigrees are highly inbred, such as in the case of the breed exclusive dataset of the dHA breed which includes only 84 inbred individuals with mean F of 13.7%. Moreover, the regression approach assumes that the reference population is the last and more inbred generation. This is not realistic in the case of migration as highlighted when the completed datasets were analysed. The N_e obtained using regression coefficients with the completed datasets increase both in the dAA and the dSSH breeds. This fact shows that the increase in inbreeding in the parental breeds is lower than in the derived breeds, leading to a global lower increase in inbreeding in the completed datasets. Furthermore, it must be noted that somewhere in the pedigree of each of the individuals of the completed pedigrees, a mating between animals from a different parental breed could occur with an intermediate result of null inbreeding. An example of this fact is observed in the meaningless N_e value computed for dHA when the completed dataset was considered. This is because the parental generation was more inbred than the reference population (6.38% versus 1.91%).

To compare the results obtained here with closed populations similar in size and breeding policy (55.4 ± 4.96 for SPB and 39.1 ± 4.19 for SA; Cervantes *et al.* 2008b), N_e should be computed taking into account migration rates (Falconer & Mackay 1996). The $_{eq}N_e$ given in Table 5 are consistent with the history of the breeds and are from 36% (dAA) to 84% (dSSH) of their respective values computed by regression in completed datasets.

Between-populations differentiation

The between-breeds differentiation assessed via D_m is highly affected by the within-breeds coancestry. Previously unknown ancestors in derived populations with shallower pedigrees are not expected to increase the between-populations coancestry because different breeds do not share common ancestors. However, these ancestors would increase the coancestry within populations. In other words, between-breeds distances are underestimated when

one of them has poor pedigree knowledge. Differentiation between the SA and the Arab-derived breeds analysed here is not negligible because the average within-breeds genetic coancestry exceeds the between-breeds coancestry. This can be also observed in the case of the SPB-derived breeds. However, the very low within-breed coancestry assessed in TBs (0.005) gives lower D_m values than expected. In this respect, the within-breeds coancestry values and, especially, the between-breeds coancestry matrix are more informative on the mating policies and histories of the breeds. Note that coancestries are half the additive genetic relationships that are, in turn, the percentage of genes shared between two individuals or, as here, populations. Taking this into consideration dAA and dHA populations would share a higher percentage of genes (4.6% for dAA and 6.4% for dHA) with the present SA population than with, respectively, the present TB and SPB populations. These results are not consistent with the classical approach used before to assess founder genetic contributions to a population thus showing that the gene origins of dAA and dHA are closer to, respectively, the TB and the SPB breeds than to the SA breed (see Table 4). An alternative explanation is that the TB and SPB individuals used for reproduction in the dAA and dHA breeds are not the same as those used for reproduction within the parental breeds, particularly in the case of TBs. In any case, this analysis clearly shows that estimates of genetic contribution from parental to derived breeds are biased in the case of open studbooks. In this respect, only when studbook information from parental breeds is not available, contrary to what happened in this study, the performance of such estimations could be recommended.

Conclusions

We carried out pedigree analyses in three SA-derived horse breeds. The methodologies used in this work illustrate the difficulties in carrying out genealogical studies in open populations.

The analysis carried out with the exclusive datasets was not enough to characterise the genetic variability of derived breeds, but the analysis made by completed datasets attained better results. The available genetic variability in the SA-derived breeds was poorer than expected in populations subject to continuous migration due to the use for reproduction of a limited number of families from the parental horse breeds. This is reflected by the major decrease assessed for parameters such as effective number of

founders, effective number of ancestors and founder genome equivalents when the completed pedigrees are considered. This is particularly true in the dHA breed which is considered an endangered breed in Spain. From a breeding perspective, the present study emphasizes the need to include the complete genealogies of the individuals from parental breeds used for reproduction in open populations to avoid unobserved loss of genetic variability putting at risk the genetic stocks.

The need to account for migration rates to compute N_e is empirically supported here because the regression approaches tested to compute N_e did not fit populations under migration. Moreover, we point out that ascertaining the genetic contributions from a population to another via founder analyses cannot be consistent with between-populations coancestries and can therefore be questioned. As for between-studbooks connectedness, it has been shown that further studies are needed to assess possible joint evaluation for common traits included in the corresponding selection schemes. The essayed methodologies were shown to be useful to deal with animal populations under migration. Misleading conclusions can be deduced if the whole relationship among migrants is unknown.

Acknowledgements

The authors wish to thank the 'Subdirección General de Medios de Producción Ganaderos' of the Spanish Ministry of Agriculture for their support. We also thank the 'Fondo de Explotación de los Servicios de Cría Caballar y Remonta' of the Spanish Ministry of Defence and the Spanish Arab, Spanish Purebred and Spanish Sport Horse Breeders' Association for their collaboration. This work was partially funded by a grant from the Junta de Andalucía given to I. Cervantes (BOJA, no. 120, 21/06/2004).

References

- Aguilar I., Misztal I. (2008) Technical note: recursive algorithm for inbreeding coefficients assuming nonzero inbreeding of unknown parents. *J. Dairy Sci.*, **91**, 1669–1672.
- 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. Bio.*, **30**, 37–42.
- Ballou J.D., Lacy R.C. (1995) Identifying genetically important individuals for management of genetic variation in pedigreed populations. In: J.D. Ballou,

- M. Gilpin, T.J. Foote (eds), Population Management for Survival and Recovery: Analytical Methods and Strategies in Small Population Management. Columbia University Press, New York, pp. 76–111.
- 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.
- Bowling A.T., Ruvinsky A. (2000a) The genetic resources and their conservation. In: A.T. Bowling, A. Ruvinsky (eds), The Genetic of the Horse. CAB International, UK, pp. 387–410.
- Bowling A.T., Ruvinsky A. (2000b). Genetic aspects of domestication. In: A.T. Bowling, A. Ruvinsky (eds), The Genetic of the Horse. CAB International, UK, pp. 25–52.
- Caballero A., Toro M.A. (2000) Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genet. Res. Camb.*, **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.
- Falconer D.S., Mackay F.C. (1996) Introduction to Quantitative Genetics. Longman Group Ltd, England.
- Goyache F., Gutiérrez J.P., Fernández I., Gómez J., Álvarez I., Díez 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., Goyache F. (2005) A note on ENDOG: a computer program for analysing pedigree pedigree information. *J. Anim. Breed. Genet.*, **122**, 172–176.
- Gutiérrez J.P., Altarriba J., Díaz C., Quintanilla R., Cañón J., Piedrafita J. (2003) Genetic analysis of eight Spanish beef cattle breeds. *Genet. Sel. Evol.*, **35**, 43–64.
- Gutiérrez J.P., Marmí J., Goyache F., Jordana J. (2005) Pedigree information reveals moderate to high levels of inbreeding and a population genetic structure in the Catalanian donkey breed. *J. Anim. Breed. Genet.*, **122**, 378–386.
- Hamann H., Distl O. (2008) Genetic variability in Hanoverian warmblood horses using pedigree analysis. *J. Anim. Sci.*, **86**, 1503–1513.
- James J. (1972) Computation of genetic contributions from pedigrees. *Theor. Appl. Genet.*, **42**, 272–273.
- Lacy R.C. (1989) Analysis of founder representation in pedigrees: founder equivalent and founder genome equivalents. *Zoo. Biol.*, **8**, 111–123.
- Lutaaya B.E., Misztal I., Bertrand J.K., Mabry J.W. (1999) Inbreeding in populations with incomplete pedigrees. *J. Anim. Breed. Genet.*, **116**, 475–480.
- MacCluer J., Boyce B., Buke L., Weitzkamp D., Pfenning A., Parsons C. (1983) Inbreeding and pedigree structure in Standardbred horses. *J. Heredity*, **74**, 394–399.
- Malécot G. (1948) Les Mathématiques de l'Hérédité. Masson et Cie, Paris.
- MAPYA. (2003) Estudio y caracterización genética del sector equino en España. Ministerio de Agricultura Pesca y Alimentación, España. <http://www.mapa.es/app/Equino/documentos/Estudio%20y%20Caracterización%20del%20Sector%20Equino%20en%20España.pdf>
- Maxwell J. (1995) Spanish Arabian Horse Families. Alexander Heriot & Co. Ltd, England.
- Moureaux S., Verrier E., Ricard A., Meriaux J.C. (1996) Genetic variability within French race and riding horse breeds from genealogical data and blood marker polymorphisms. *Genet. Sel. Evol.*, **28**, 83–102.
- Pérez-Enciso M. (1995) Use of the uncertain relationship matrix to compute effective population size. *J. Anim. Breed. Genet.*, **112**, 327–332.
- Steen A.K. (2006) A Glorious Task. A Tale of the Breed. Escandón Impresores, Seville, Spain.
- Teegen R., Edel C., Thaller G. (2009) Population structure of the Trakehner Horse breed. *Animal*, **3**, 6–15.
- Valera M., Molina A., Gutiérrez J.P., Gómez J., Goyache F. (2005) Pedigree analysis in the Andalusian horse: population structure, genetic variability and influence of the Carthusian strain. *Livest. Prod. Sci.*, **95**, 57–66.
- VanRaden P.M. (1992) Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. *J. Dairy Sci.*, **75**, 3136–3144.
- Wright S. (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.
- Zechner P., Sölkner J., Bodo I., Druml T., Baumung R., Achmann R., Marti E., Habe F., Brem G. (2002) Analysis diversity and population structure in the Lipizzan horse bred based on pedigree information. *Livest. Prod. Sci.*, **77**, 137–146.