

# Influence of foreign breeds on the genetic structure of the Spanish Sport Horse population

E. Bartolomé <sup>a,\*</sup>, I. Cervantes <sup>b</sup>, M. Valera <sup>a</sup>, J.P. Gutiérrez <sup>b</sup>

<sup>a</sup> Departamento de Ciencias Agroforestales, ETSIA, Universidad de Sevilla, Ctra. Utrera, km1, 41013 Sevilla, Spain

<sup>b</sup> Departamento de Producción Animal, Facultad de Veterinaria, Universidad Complutense de Madrid, Avda. Puerta de Hierro s/n, 28040 Madrid, Spain

## ARTICLE INFO

### Article history:

Received 1 December 2010

Received in revised form 27 June 2011

Accepted 28 June 2011

### Keywords:

Equine

Pedigree analysis

Open population

Founder genome equivalents

Genetic structure

## ABSTRACT

The Spanish Sport Horse (SSH) is a recent breed, founded in 2002 in the search for a breed with a first-rate performance at any of the Olympic equestrian disciplines (dressage, show jumping and/or eventing). The aim of this research was to describe the genetic structure of the SSH and to assess the influence of foreign breeds in its genetic structure. The whole pedigree included 34,017 animals. Pedigree analyses were carried out for the registered SSH after the Stud Book was closed (for a 'reference population' with 4289 individuals). The consequences of a new Stud Book regulation were considered in three subpopulations, including as a reference population those SSH animals with both SSH parents or with at least one SSH parent (father or mother). In order to study the influence of the foreign sport breeds, 9 genetic lines were assessed according to their country of origin. Demographic parameters, level of inbreeding, effective population size, probability of gene origin and genetic structure of the SSH population were assessed using ENDOG v4.8. In order to compare the evolution of the genetic variability present within the different genetic lines, the resultant founder genome equivalent ( $f_g$ ) after 10 generations was calculated for each genetic line and for each subpopulation. The reference population showed a completeness level of over 85% until the third parental generation with a generation interval of 10.8 years and an inbreeding coefficient of 0.66%, with an average relatedness of 0.16%. The probability of gene origin showed that the SSH breed has its origins mainly in German sport horse breeds. The evolution of the genetic lines  $f_g$  showed that 'Mixed', 'Others' and 'SPB' genetic lines proportionally decreased the most over ten generations (22.8%, 23.0% and 17.2%, respectively), whereas the evolution of the subpopulations  $f_g$  showed a higher loss of genetic variability in the reference population and in the subpopulation with at least one SSH mother. The ratio between the effective size computed using paired increases in coancestries and the effective size based on individual increases in inbreeding for the reference population was quite high (4.6), showing a strong subdivision that could be an effect of the three breeding goals proposed in the SSH Breeding Program. This value also increased considerably when the three subpopulations studied (both SSH parents, at least the father SSH or at least the mother SSH) were considered (to 7.3, 5.6 and 5.8, respectively). These results suggested that it is not yet advisable to apply restricting policies in this population, as this could lead to a reduction in the genetic diversity.

© 2011 Elsevier B.V. All rights reserved.

## 1. Introduction

Recent developments in the Spanish equine industry during the 21st century have been stimulated by the increasing use of

horses for sports and recreation (Azor et al., 2007; Cervantes et al., 2008a; Valera et al., 2005), leading to the formation of new breeds, such as the Spanish Sport Horse (SSH), which has arisen from a general demand for animals with a good aptitude and performance in sports. However, this breed was designed in a similar way to other European sport horse breeds, such as Dutch Warmblood, Hanoverian or Oldenburger (Janssens,

\* Corresponding author. Tel.: +34 954487748; fax: +34 954486436.

E-mail address: [ebartolome@us.es](mailto:ebartolome@us.es) (E. Bartolomé).

2008), solely with a view to the horse's ability for sports performance.

The SSH is a recent breed, founded in 2002 in the search for a breed with a first-rate performance at any of the Olympic equestrian disciplines (dressage, show jumping and/or eventing) (Cervantes et al., 2009). The SSH population represents 5.4% of the national equine census in Spain (Valera et al., 2009). When the SSH Official Stud Book was created, an animal could be registered in the Foundational Registry (and thus be identified as an SSH animal) up until 2004, if it came from any of the “permitted” cross-breeds (established by Spanish laws APA/3318/2002 and APA/1646/2004) and was born between 1992 and 1998. Since the Foundational Registry was closed in 2004, an animal was considered to be an SSH either when both parents were registered as SSH, when only one of them was an SSH with the other parent from any of the “permitted” breeds included in these laws or when both parents belonged to any of these “permitted” breeds (even when both parents were from foreign breeds, their offspring could be inscribed as SSH at the owner's request if they were not inscribed in any other official Stud Book). Furthermore, an animal from a foreign breed could be inscribed in the Auxiliary Registry of the SSH Stud Book as a Foreign SSH when it was crossed with an SSH and its descendants were inscribed as SSH in the Principal Registry of the SSH Stud Book.

The breed is managed by the Spanish Sport Horse Breeders Association (ANCADES, [www.ancades.com](http://www.ancades.com)) and its Breeding Program was officially approved in 2005. Its main goal is “to obtain a horse with a good functional conformation, temperament and health, able to attain a high performance at either national or international sports events in which it participates” (Valera et al., 2009). Despite the fact that no breed standard has been described yet, SSH breeders looked forward to certain characteristics of conformation, movement and temperament related to sports performance, through the mating with other sport horse breeds that would enhance their show jumping, dressage or eventing abilities (Bartolomé et al., 2008).

Hellsten et al. (2009) pointed out that the influence of certain foreign horse breed populations over a composite breed may have an impact on its sports performance, enhancing or decreasing its genetic progress. Regarding the SSH, the importation of foreign genetic material (from foreign sport horse breeds) has always been present and the Stud Book is still open (through the Auxiliary Registry), giving the breeders the opportunity to choose the foreign stallions that best contribute to the improvement of their horses' performance, as regards their physical, physiological and/or genealogical characteristics. Furthermore, previous studies have revealed morphological differences in the SSH regarding their specialization in sports performance either in dressage, show jumping or eventing disciplines (Bartolomé et al., 2008), highlighting the fact that specifically-directed selection efforts had been taking place.

Cervantes et al. (2009) revealed that the genealogical analyses of open populations, permitting the use of individuals from other breeds for reproduction, presented particular characteristics that made it difficult to ascertain the real rates of genetic variability in the analyzed stock. The analysis of populations receiving genetic contributions from

other populations over time needs to take into account migration rates in order to assess their genetic variability in terms of inbreeding and effective size (Falconer and Mackay, 1996). Moreover, the effect of the mating policy on the genetic variability of an open population could also determine the impact of foreign stallions on the genetic exchange between populations (Koenen et al., 2004) and differences in the mating preferences of foreign breeds could lead to changes in the SSH gene pool.

The aim of this research was to describe the genetic structure of the SSH as a recent breed and to assess the influence of foreign breeds in the genetic structure of the SSH population.

## 2. Material and methods

The pedigree information was gathered from the Stud Book provided by ANCADES. We analyzed the horses registered in it, from its foundation in 2002 to December 2009. The whole pedigree included 34,017 animals (13,356 males and 20,661 females), only 11,659 of which (44.5% males and 55.5% females) were registered as SSH. Pedigree analyses were carried out for a group made up of the SSH animals born between 2005 and 2009, in order to include the registered individuals after the Foundational Registry of the SSH Stud Book was closed (4289 individuals), which we refer to as the ‘reference population’.

In order to obtain the possible change in the genetic structure and the effective size of the SSH population, the consequences of a hypothetical new Stud Book regulation, in which only individuals SSH could act as parents, were also considered in three probable scenarios, referred to as ‘subpopulations’: including as reference population those SSH animals with both parents of SSH breed and SSH animals with at least one SSH parent (father or mother). 1984 SSH animals from the reference population did not have SSH breed animals as parents and thus were not included in these probable scenarios.

Regarding the composite nature of the SSH, many different sport horse breeds were included in the Stud Book as relatives. In order to study the influence of these breeds on the reference population, 9 groups referred to as “genetic lines” were assessed according to their origin; an SSH belonged to one of them when more than 50% of its ancestors were of a breed in this line. Thus, 4 lines were created according to their country of origin (‘Germany’ for Holsteiner, Hanoverian, Westphalian, Oldenburger and Trakehner; ‘Netherlands’ for Dutch Warmblood, Belgian Warmblood, Danish Warmblood and Zangersheide; ‘France’ for the Selle Français; and ‘Portugal’ for Lusitano); 3 genetic lines were from Spanish breeds registered in the Spanish Horse Breeds Official Catalogue – Spanish regulation RD 2129/2008 – (‘Spanish Purebred’ – SPB; ‘Spanish Arab’ – SA and ‘Thoroughbred’ – TH); one genetic line, termed ‘Others’, that included SSH with ancestors of minority foreign sport horse breeds (animals registered as Foreign SSH on the Stud Book were also grouped here); and the last genetic line, referred to as ‘Mixed’, that included SSH formed by several different breeds contributing less than 50% to the animal's genetic profile.

In order to identify the quantity of founders and ancestors from different breeds that appeared in the SSH pedigree,

those founders and ancestors were grouped similarly to genetic lines and referred to as 'founders' groups' ('German' for Holsteiner, Hanoverian, Westphalian, Oldenburger and Trakehner; 'Netherland' defined as before; 'Selle Français'; 'Lusitano'; 'SPB' for the Spanish Purebred; 'SA' for the Spanish Arab; 'TH' for the Thoroughbred; 'MFB' for the Minority Foreign Breeds (animals registered as Foreign SSH in the Stud Book were also grouped here)). Note that the first classification in genetic lines is assessed within the reference population and this second one refers to their ancestors and founders.

The following pedigree analyses were computed using the program ENDOG (v4.8, Gutiérrez and Goyache 2005).

### 2.1. Demographic parameters

Circumstances affecting the genetic history of the SSH population, were assessed with the following analyses:

The *pedigree completeness level* was assessed as the proportion of ancestors known per generation for each offspring (MacCluer et al., 1983). The *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).

The *average generation interval* ( $GI$ ) was calculated as the average age of parents at the birth of their offspring kept for reproduction (James, 1977) and was computed for the reference population.

### 2.2. Inbreeding, average relatedness coefficient and effective population sizes

In order to assess the level of genetic variability in the SSH population and in the subpopulations studied, the following parameters were analyzed:

- *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).
- *Effective population size*, computed using both the individual increase in inbreeding ( $\bar{N}_e$ ) (Cervantes et al., 2008b; Gutiérrez et al., 2008, 2009) and the increase in pairwise coancestry ( $\bar{N}_{ec}$ ) (Cervantes et al., 2011). The individual increase in inbreeding was defined as  $\Delta F_i = 1 - \sqrt[t_i]{1 - F_i}$ , 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  $\Delta c_{jk} = 1 - \sqrt[\frac{t_j+t_k}{2}]{1 - c_{jk}}$ , 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, we can estimate an effective population size based on inbreeding

$\bar{N}_e = \frac{1}{2\Delta F}$  or in coancestries  $\bar{N}_{ec} = \frac{1}{2\Delta C}$ . The ratio between both effective sizes  $\left(\frac{\bar{N}_{ec}}{\bar{N}_e}\right)$  will be also computed in order to ascertain the presence of population structure.

### 2.3. Probability of gene origin

The genetic representation of the founder population in the SSH population and in the 3 subpopulations studied, was highlighted with the following parameters:

- *Effective number of founders* ( $f_e$ ). This parameter is the reciprocal of the probability that two genes drawn at random from the studied population originate from the same founder (James, 1972) and is computed from the genetic contribution of founders to the descendant gene pool of the population (Lacy, 1989). Following Moureaux et al. (1996), ancestors with both parents unknown in the available database were considered founders. If one parent was known, the unknown parent was considered a founder (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, thus producing bottlenecks. The genetic contributions of the ancestors have been averaged by *lustra* (periods of 5 years) of birth of the registered individuals in order to identify changes in the mating preferences.
- *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). It was computed as the inverse of twice the average coancestry of the individuals within the population ( $f_g = \frac{1}{2C}$ ) (Caballero and Toro, 2000).

In order to compare the evolution of the genetic variability present within the different genetic lines, the resultant  $f_g$ , after each of the first hypothetical 10 generations had elapsed, was calculated for each genetic line and for each subpopulation studied, standardizing by the size of the line/subpopulation and assuming no changes in their sizes and breeding policy over the generations. To achieve these proposals, the coancestry of the formula  $f_g$  for each discrete generation (from 1 to 10) was calculated as:

$$C = \frac{((N * S_t) + (N * (N-1) * C_t))}{N^2},$$

where  $N$  was the subpopulation size,  $C_t$  was the coancestry for a given discrete generation and  $S_t$  was the mean self-coancestry parameter.  $C_t$  was calculated as  $1 - \left(1 - \frac{1}{2\bar{N}_{ec}}\right)^t$ , where  $\bar{N}_{ec}$  was the effective population size calculated by increase in coancestry and  $t$  was the given number of discrete generations.  $S_t$  was defined as  $\frac{1}{2}(1 + F_t)$ , where  $F_t$  was the

inbreeding for generation  $t$  and was calculated as  $1 - \left(1 - \frac{1}{2N_e}\right)^{t-1}$ , with  $N_e$  as the effective population size calculated by individual increase in inbreeding.

#### 2.4. Genetic structure

The genetic structure of the SSH population and of the subpopulations studied was assessed by two parameters:

- *Coefficient of coancestry* ( $f_{ij}$ ). Defined as the probability that two gametes taken at random, one from each of the two individuals compared, carry alleles that are identical by descent (Falconer and Mackay, 1996). For this study, SSH genetic lines were compared. The within-genetic lines coancestry ( $f_{ii}$ ) and the between-genetic lines coancestry matrix ( $f_{ij}$ ) were computed, averaging all pairwise coancestry coefficients of the individuals belonging, respectively, to a given genetic line  $i$  or to two different genetic lines  $i$  and  $j$ .
- *Nei's minimum distance* ( $D_m$ ). Computed following Caballero and Toro (2000, 2002) as  $D_m = [(f_{ii} + f_{jj})/2] - f_{ij}$ , where  $f_{ii}$  and  $f_{jj}$  were the average coancestry within a genetic line  $i$  or  $j$  and  $f_{ij}$  was the coancestry between two genetic lines  $i$  and  $j$ .

### 3. Results

#### 3.1. Demographic analyses

The quality of the pedigree information of the individuals included in both the entire population and the reference population is illustrated in Fig. 1. The completeness level for the whole SSH population at first generation is 71.2% and drops to values below 30% after the fourth generation. The reference population showed a completeness level of over 85% until the third parental generation. The number of equivalent complete generations ( $t$ ) and the average generation interval ( $GI$ ) showed a value of 4.29 and 10.8 (respectively), both for the reference population.

#### 3.2. Inbreeding, average relatedness coefficient and effective population sizes

The inbreeding coefficient ( $F$ ) and average relatedness ( $AR$ ) for the reference population were 0.66% and 0.16%, respectively. Due to the scarce pedigree knowledge, 79.6% of this reference population appeared to show an  $F$  equal to 0, whereas only 0.3% of them showed an  $AR$  value close to 0 (i.e.  $<0.0001$ ). Despite the multiple genetic origins of this breed, and taking into consideration that in an ideal population with random matings,  $F$  would be half the  $AR$ , the evolution of these parameters (Fig. 2) showed that  $F$  attained values much larger than half the  $AR$ , accounting for quite an important subdivision in this population. Nevertheless, the highest mean  $F$  value reached in this population was low (0.67% for the period 2000–2004), ranging from 0% to 0.67%, while the  $AR$  remained fairly stable, ranging from 0.04% to 0.16%.

In order to describe the genetic diversity in this population, the effective population size ( $N_e$ ; Falconer and Mackay, 1996) was computed for the reference population, for three subpopulations of mating policies in the SSH Stud Book and for the 9 genetic lines analyzed, following two different computation approaches, by the increase of coancestry ( $\bar{N}_{ec}$ ) and by the individual increase of inbreeding ( $\bar{N}_e$ ) (Table 1).  $\bar{N}_{ec}$  was considerably higher than  $\bar{N}_e$ , both for the reference population and for the three subpopulations, highlighting higher increases in inbreeding than in coancestry, which fits with the results shown previously in Fig. 2. When compared with results obtained for the reference population,  $\bar{N}_{ec}$  decreased the most (64.3%) when only SSH animals were used as parents, falling from 1046.3 to 373.9 individuals; 58.4% when at least the father was SSH (decreasing to 435.7) and just 23.0% when at least the mother was SSH (decreasing to 806.0). As regards to  $\bar{N}_e$ , the decrease was greater, with a fall of 77.2% (to 51.4) in the first subpopulation, 65.5% (to 77.8) in the second and just 38.1% (to 139.8) in the third. Otherwise, these results were quite expectable, because, with such population reductions, their origins would also be reduced.

$\bar{N}_e$  and  $\bar{N}_{ec}$  were also computed for every genetic line considered. The 'Others' genetic line showed the highest

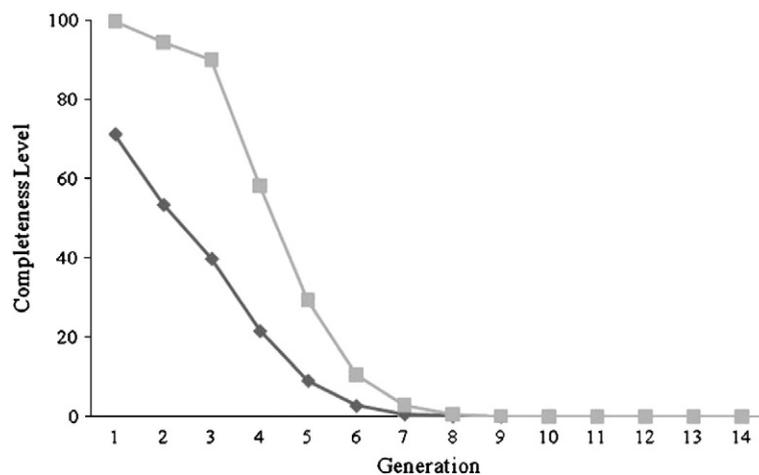


Fig. 1. Average percentage of ancestors known per generation, (generation 1 = parents, generation 2 = grandparents, etc.) in the Spanish Sport Horse. Information is provided for the whole SSH population (black squares) and for the reference population (gray squares).

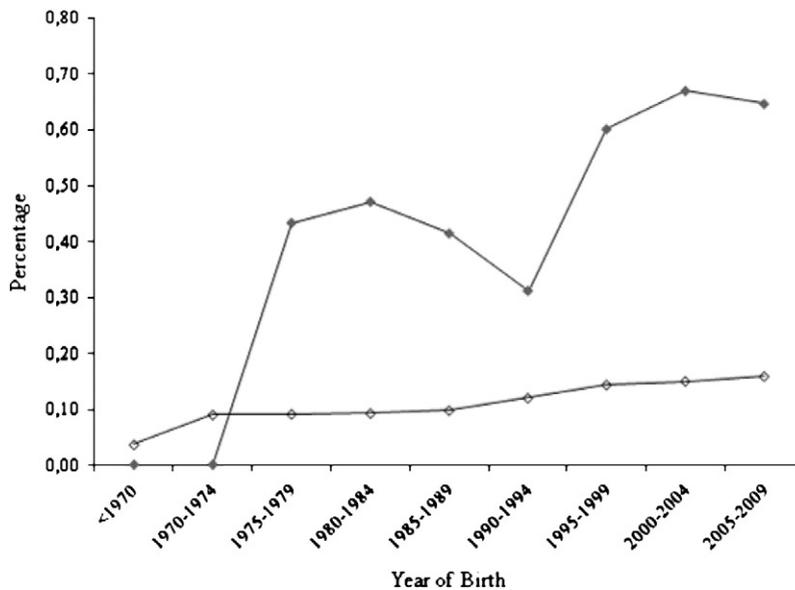


Fig. 2. Evolution of individual inbreeding coefficient (in black squares) and average relatedness coefficient (in open squares), on the SSH population (in percentage).

effective population size calculated by the increase of coancestries (1031.4), whereas the ‘Thoroughbred’ line showed the highest effective population size calculated from the increase of inbreeding (415.9), the ‘Netherlands’ line showed the lowest value for  $\bar{N}_e$  (125.3) and the ‘Portugal’ line showed the lowest for  $\bar{N}_{ec}$  (104.0).

The  $\frac{\bar{N}_{ec}}{\bar{N}_e}$  ratio was 4.6 for the reference population, increasing to 7.3 when both parents were SSH, to 5.6 when at least the father, and to 5.8 when at least the mother, was SSH.

### 3.3. Probability of gene origin

Table 2 shows parameters characterizing the genetic variability of the population in terms of gene origin for the

Table 1

Total number of animals ( $N$ ), effective population size calculated by coancestry ( $\bar{N}_{ec}$ ) and by inbreeding ( $\bar{N}_e$ ) and their ratio ( $\frac{\bar{N}_{ec}}{\bar{N}_e}$ ), calculated for the reference population, for three subpopulations of the SSH Stud Book regarding different mating policies and for every genetic line.

	$N$	$\bar{N}_{ec}$ ( $\pm$ s.d.)	$\bar{N}_e$ ( $\pm$ s.d.)	Ratio $\frac{\bar{N}_{ec}}{\bar{N}_e}$
Reference Population	4289	1046.3 $\pm$ 8.2	225.8 $\pm$ 72.4	4.6
<i>Stud Book subpopulations</i>				
Both parents SSH	481	373.9 $\pm$ 6.2	51.4 $\pm$ 16.4	7.3
Father SSH	797	435.7 $\pm$ 6.1	77.8 $\pm$ 25.3	5.6
Mother SSH	1989	806.0 $\pm$ 7.9	139.8 $\pm$ 45.5	5.8
<i>Genetic lines</i>				
SPB	609	577.2 $\pm$ 6.9	185.2 $\pm$ 51.6	3.1
TH	465	326.3 $\pm$ 7.1	415.9 $\pm$ 149.4	0.8
France	99	232.5 $\pm$ 3.6	184.0 $\pm$ 41.0	1.3
Others	694	1031.4 $\pm$ 12.4	269.9 $\pm$ 129.3	3.8
SA	228	268.2 $\pm$ 4.4	116.2 $\pm$ 27.4	2.3
Germany	386	220.2 $\pm$ 4.2	252.3 $\pm$ 45.9	0.9
Netherlands	101	125.3 $\pm$ 4.1	346.2 $\pm$ 69.7	0.4
Portugal	109	125.5 $\pm$ 4.6	104.0 $\pm$ 50.7	1.2
Mixed	1598	480.5 $\pm$ 5.2	240.6 $\pm$ 67.7	2.0

reference population and for the subpopulations considered, as regards different mating policies. A total of 6772 founders were identified for the reference population, decreasing to 1585 if only SSH animals were used as parents, and to 3469 and 3748 when at least one of the parents was SSH, father or mother, respectively. The effective number of founders was 14.2% (963) of the total founders for the reference population, 23.9% (380) if both parents were SSH and 13.5% (467) or 19.1% (714) if only one parent was SSH (father or mother, respectively). The effective number of ancestors was 407 for the reference population, which represented 12.6% of the ancestors, thus accounting for 100% of the genetic variability (3230 ancestors). This value decreased to almost half (196) when only SSH animals with at least one SSH father were analyzed, and to a quarter (122) when those animals had both parents SSH. The founder genome equivalents parameter measured how many founder genes were maintained in

Table 2

Parameters characterizing the probability of gene origin for the reference population and for three subpopulations of the SSH Stud Book.

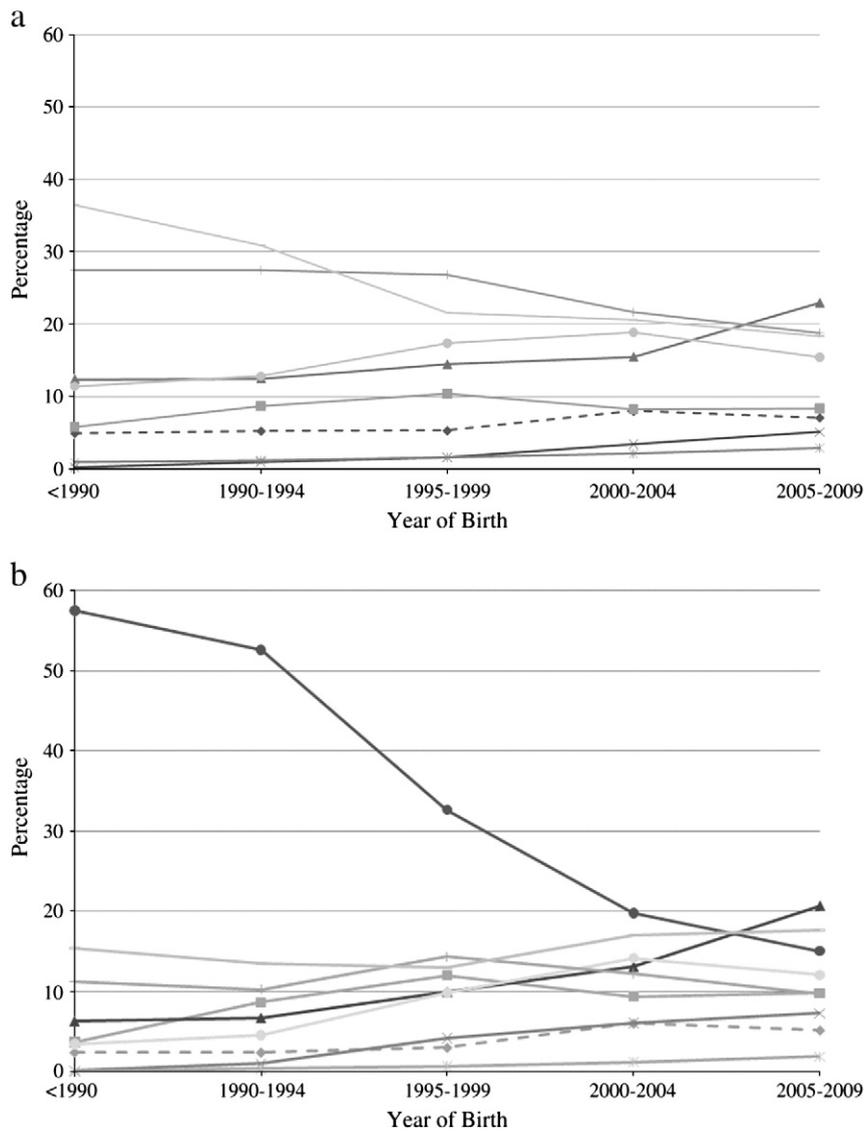
	Reference population	Stud Book subpopulations		
		Both parents SSH	Father SSH	Mother SSH
Total number of animals in the population	4289	481	797	1989
Total number of founders	6772	1585	3469	3748
Effective number of founders	963	380	467	714
Effective number of ancestors	407	122	154	288
Founder genome equivalents	254	74	94	181
Ancestors explaining 50% of the genetic variability	248	53	96	132
Ancestors explaining 100% of the genetic variability	3230	122	196	196

the population for a given locus and how balanced their frequency was in the offspring of the reference population (James, 1972; Lacy, 1989; Olsen et al., 2010). The reference population showed a value of 254.3 for this parameter. When different mating policies were taken into consideration, the  $f_g$  value decreased to 73.9, when both parents were SSH, to 94.4 when at least the father was SSH and to 180.6 when at least the mother was. Regarding the  $f_g$  value of the genetic lines, 'Others' (230.5), 'SPB' (116.3) and 'Mixed' (110.2) showed the highest values, whereas 'Netherlands' (26.7), 'Portugal' (27.6) and 'France' (36.3) showed the lowest.

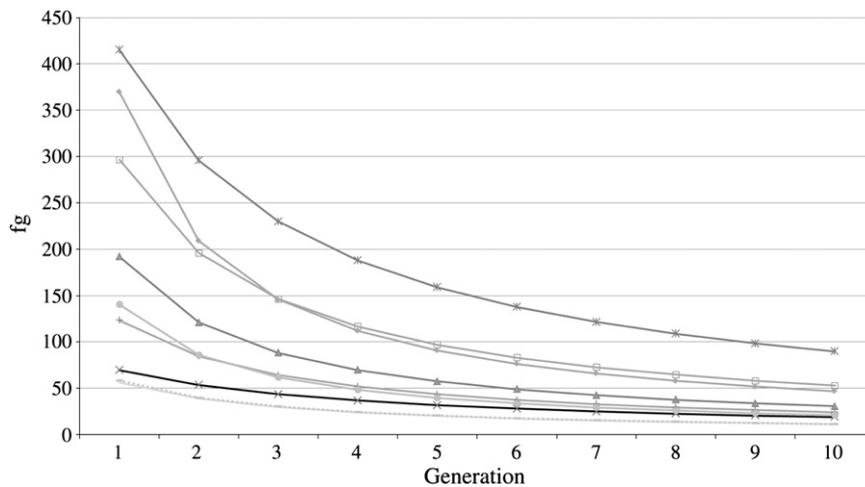
In order to illustrate the genetic history of the SSH breed, an indirect measure of the changes of breeders' preferences over time was obtained by averaging the genetic contribution

of their founders and ancestors in Fig. 3 by date of birth of the registered animals. In the case of the founders (Plot A), the 'MFB' founders' group showed the highest influence at the beginning, decreasing from 36% to 18% of influence. The 'TH' founders' group had the second greatest influence on the SSH until the period 2005–2009, where the influence of the 'Germany' founders' group increased. As for ancestors (Plot B), the 'SSH' breed showed the highest influence (58%) during the period before 1990, decreasing to 15% in the last period, below the 'Germany' (21%) and 'MFB' (18%) founders' groups.

Fig. 4 shows the evolution of this parameter ( $f_g$ ) per generation, for each genetic line considered. The 'Others' genetic line showed the highest  $f_g$  for all generations, with values that ranged from 415.1 for the first generation to 89.9



**Fig. 3.** Evolution of the contribution of the predefined founders breeds of the SSH to the reference population (plot a) and evolution of the contribution of the Boichard et al.'s (1997) ancestors by year of birth of the registered SSH individuals (plot b). The following founders' groups are represented: Spanish Sport Horse (●), Spanish Arab (◆), Selle Français (■), Germany (▲), Netherland (×), Lusitano (✱), Spanish Purebred (○), Thoroughbred (⊕) and Minority Foreign Breeds (—).



**Fig. 4.** Evolution of founder genome equivalents ( $f_g$ ) per generation for each genetic line. The following genetic lines are represented: Spanish Arab (—), France (—), Germany (—), Mixed (—), Netherlands (—), Portugal (—), Spanish Purebred (—), Thoroughbred (—) and Others (—).

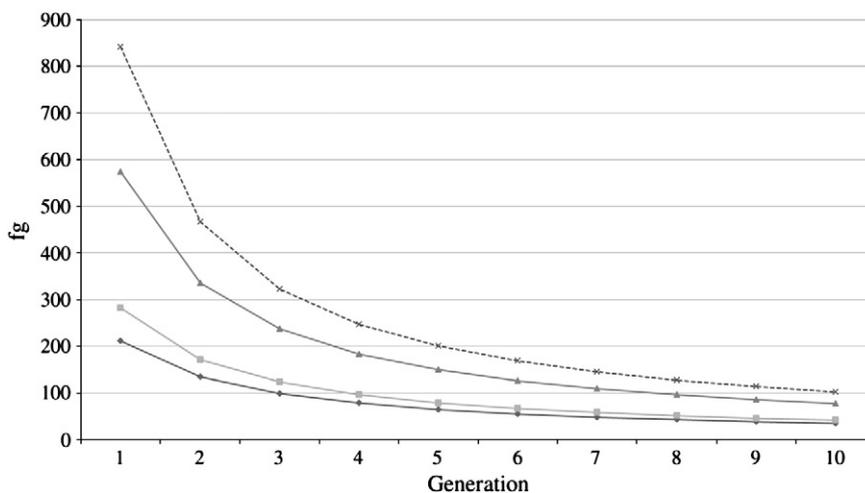
for the tenth. It was followed by the 'Mixed' and 'SPB' genetic lines, with values ranging from 369.6 and 296.6 to 46.9 and 52.9, respectively. Furthermore, all the genetic lines showed a homogeneous decreasing tendency, with only the 'Mixed' line crossing trajectory with 'SPB' in the third generation, highlighting a greater loss of variability for the 'Mixed' than for the 'SPB' genetic lines. Otherwise, the 'Netherlands' and 'Portugal' genetic lines showed a gentler decline in evolution, highlighting a smaller loss of genetic variability.

Evolution of  $f_g$  was also checked for the reference population and for the 3 subpopulations studied (Fig. 5), with the reference population showing the highest  $f_g$  values (from 841.3 in the first generation to 102.3 in the tenth) and the most restricted subpopulation (with both parents SSH) showing the lowest (from 210.6 to 34.7). The other subpopulations (with at least the father SSH and at least the mother SSH) showed values in

between, ranging from 281.9 and 573.8 (respectively) for the first generation, to 41.4 and 77.6 (respectively) for the tenth.

### 3.4. Genetic structure

Table 3 shows Nei's minimum distance ( $D_m$ ) values and the pairwise coancestry coefficients ( $f_{ij}$ ) between genetic lines of the reference population ( $\times 1000$ ). The 'Netherlands' and 'Portugal' genetic lines showed the highest genetic differentiation ( $D_m = 18.3$ ), followed by the 'France' and 'Portugal' ( $D_m = 15.9$ ) or 'France' and 'Netherlands' ( $D_m = 15.3$ ) genetic lines, whereas 'Mixed' and 'Others' showed the lowest genetic differentiation ( $D_m = 2.3$ ). Within-genetic line coancestry values varied between 18.7 ('Netherlands') and 2.9 ('Others'), while between-genetic



**Fig. 5.** Evolution of founder genome equivalents ( $f_g$ ) per generation for the reference population (---) and for the three subpopulations studied: with both SSH parents (—), with at least an SSH father (—) and with at least an SSH mother (—).

**Table 3**

Parameters describing the SSH population ( $\times 1000$ ). Diagonal and below: average coancestry within genetic lines and among genetic lines (respectively). Above diagonal: Nei's minimum distance between genetic lines.

	Ger	Mix	TH	SPB	SSH	SA	Oth	Neth	Fr	Port
Ger	10.07	3.17	7.20	6.97	6.53	9.79	5.51	12.83	9.96	13.86
Mix	4.13	4.54	4.15	3.92	4.03	6.25	2.31	10.07	5.89	10.58
TH	1.14	1.42	6.60	4.73	5.62	7.21	4.03	11.94	9.19	12.12
SPB	0.22	0.50	0.72	4.30	4.35	6.52	2.81	11.05	8.82	10.72
SSH	1.37	1.10	0.55	0.66	5.73	7.24	3.73	11.90	9.03	11.68
SA	0.33	1.09	1.18	0.71	0.70	10.16	5.72	13.56	11.55	13.42
Oth	0.96	1.39	0.70	0.77	0.57	0.79	2.87	10.25	7.00	10.32
Neth	1.56	1.55	0.71	0.45	0.32	0.87	0.54	18.70	15.33	18.30
Fr	1.97	3.26	1.00	0.22	0.72	0.41	1.32	0.90	13.77	15.88
Port	0.23	0.73	0.22	0.47	0.23	0.70	0.16	0.10	0.05	18.09

Ger = Germany; Mix = Mixed; TH = Thoroughbred; SPB = Spanish Purebred; SSH = Spanish Sport Horse; SA = Spanish Arab; Oth = Others; Neth = Netherlands; Fr = France; Port = Portugal.

lines coancestry showed values that ranged from 4.1 for 'Mixed' and 'Germany' to 0.05 for 'Portugal' and 'France'.

## 4. Discussion

### 4.1. Genetic variability

Parameters derived from probabilities of gene origin revealed that the SSH population had a very high genetic variability. The importation of foreign genetic material has always taken place in this breed as the Stud Book is still open, as has occurred with other sport horse breeds (Cervantes et al., 2009; Thorén et al., 2009). However, when we used animals with both SSH parents as the reference population, the total and effective number of founders and ancestors decreased considerably as expected. On the other hand, if they allowed mating with animals from other breeds, with at least one of them an SSH animal, the effects were not so marked.

The sum of the genetic contributions of the founder individuals to the population, according to their breeds grouped by country of origin, would make it possible to ascertain major sources of genes in the SSH population. Our study showed that despite the fact that most of the founders belonged to the 'MFB' group, the 'German Breeds' showed the highest founder genetic contribution to the SSH population. This could be due to the breeders preferring German breeds of horses for the improvement of the SSH sport performance, considering the well-known genetically proven sports ability of these breeds (Koenen, et al. 2004; Thorén, 2008). Focusing on the founder genome equivalents parameter, again the values shown for the SSH were much higher (254.3) than those shown by other horse breeds, such as the Noriker draught horse with 10.6 (Druml et al., 2009), the Lippizzan with 6.0 (Zechner et al., 2002), the Spanish Anglo-Arab with 50.0 or the Hispano-Arab with 18.0 founder genome equivalents (Cervantes et al., 2009). Large founder genome equivalent values found here were mainly linked to a limited degree of pedigree knowledge. Unaccounted relationships between animals coming from the same origin will lead, however, to an immeasurable decrease of the genetic diversity, in terms of identity by state.

However,  $f_g$  decreased considerably when the Stud Book subpopulations were taken into consideration and only SSH

animals could be used as parents for the offspring, due to the fact that it depends on population size, due to the greater or lower weight of self-coancestries in the mean within-subpopulation coancestry (Bartolomé et al., 2010; Druml et al., 2009; Honda et al., 2006). For this, assuming there is no change in breeding policy, we calculated the  $f_g$  value for 10 pre-established generations, for every genetic line and for the three subpopulations as a measure of genetic variability loss, standardizing by the size of genetic line/subpopulation. The mean coancestry is affected by the subpopulation size, and self-coancestries will have more importance in small populations. As the subpopulation size decreases, the subpopulation coancestry tends to increase, thus producing a lower  $f_g$ . However, in our study, after standardizing by population size, our results showed that, despite the fact that 'Mixed', 'Others' and 'SPB' genetic lines had the highest subpopulation sizes (1598, 694 and 609, respectively), their  $f_g$  values proportionally decreased the most in ten generations. These results suggest that an internal subdivision in these genetic lines overlapped the lesser influence of self-coancestry in large populations. In the case of the 'Mixed' and 'Others' genetic lines, this was probably due to the different horse breeds that go to make them up, emphasizing the different geographical origins of the minority breeds that made up the 'Others' genetic line, which would explain the higher internal subdivision. In the case of the 'SPB' genetic line, it was probably due to the low gene flow existing between SPB stud farms and to the preferential matings (Valera et al., 2005), which would probably have enhanced a subdivision within this genetic line.

When accounting for the  $f_g$  evolution of the subpopulations studied (Fig. 5), those with the biggest census (the reference population – 4289 and the subpopulation with at least a SSH mother – 1989), decreased the most. On the other hand, the smallest subpopulations showed a slight loss of genetic diversity. However, a strong subdivision and a small genetic diversity ( $f_g$  of 34.7) would be observed in the SSH population if the Stud Book was closed and only SSH animals were used for breeding. Despite the fact that negative performance and productivity effects have been found in other horse and cattle breeds due to crossbreeding (Boichard et al., 1997; Dubois and Ricard, 2007; Hagger, 2005), in the case of the SSH population, crossbreeding with other breeds (either foreign or national breeds) should be maintained, keeping the Stud Book still open.

#### 4.2. Demographic analyses

The Spanish Sport Horse is a recent breed created using individuals from national and foreign breeds, in the search for a breed with a first-rate performance at any of the Olympic equestrian disciplines (dressage, show jumping and/or eventing). This has led to a large, assorted gene pool in its population. However, our results showed certain preferences for the breeds selected for mating with SSH mares, such as German foreign breeds (Holsteiner, Hanoverian, Westphalian, Oldenburger or Trakehner). This could be due to the large sport horse population present in Germany and their traditional role as sport horse exporters (Koenen et al., 2004), together with the preferences of SSH breeders, who tend to select high performance international sport horse breeds.

In order to highlight the quality of the pedigree information reported by the SSH individuals, the completeness level and the number of equivalent generations of the reference SSH population were assessed, showing similar values to the Trakehner Horse (Teegen et al., 2009) and lower than other sport horse breeds (Cervantes et al., 2008a; Hamann and Distl, 2008; Valera et al., 2005).

The generation intervals computed for the SSH reference population were consistent with those reported before for other sport horse breeds with longer pedigrees such as the Thoroughbred (GI = 10.5; Langlois, 1982), the Spanish Purebred (GI = 10.1; Valera et al., 2005) or the French Arab Horse (GI = 11.8; Moureaux et al., 1996).

#### 4.3. Inbreeding coefficients and genetic structure

Inbreeding and AR coefficients were both very low (<1%), as can be expected in a recent breed with a multiple gene origin and a pedigree not fully-completed (Cervantes et al., 2009; Moureaux et al., 1996). The mean *F* coefficient dramatically outgrew the AR coefficient over the years, showing that a subdivision was occurring in the population. However, the *F* value has undergone a slight decrease over the last few years (Fig. 2), showing that SSH breeders are trying to avoid matings with relatives.

The ratio  $\frac{N_{ec}}{N_e}$  (Cervantes et al., 2011) (Table 1), showed the 'Others' (with 3.8) and the 'SPB' (with 3.1) as the genetic lines showing the highest ratio values. These results supported the existence of preferential matings within the 'SPB' line (Cervantes et al., 2009; Valera et al., 2005) and the different geographical origins among the minority breeds that made up the 'Others' line. Thus, the ratio was 4.6 for the reference population and increased in the subpopulations studied. Compared with other horse breeds, such as the Spanish Purebred (1.47), the Spanish Arab (1.5), the Thoroughbred (1.14), the Anglo-Arab (1.61) or the Hispano-Arab (1.09) (Cervantes et al., 2011), this value was extremely high, showing a high degree of subdivision in this population. This could be an effect of the three breeding goals proposed in the Breeding Program (dressage, show jumping and eventing).

Regarding Nei's Minimum Distances ( $D_m$ ), 'Netherlands' and 'Portugal' showed the highest genetic differentiation, whereas the 'Others', 'Mixed' and 'SPB' genetic lines showed the lowest, probably due to the fact that these groups shared more ancestors (probably SPB ancestors). However, coances-

tries showed the percentage of genes shared between subpopulations, which were quite low in general, and the highest values appeared between 'Mixed' and 'German' (4.13). This could be due to the fact that most of the breeds that contributed less than 50% to the SSH from the 'Mixed' group were from German sport horse breeds, highlighting again the preference of SSH breeders for these breeds.

Therefore, as the SSH pedigree contains animals from several different breeds, the genetic variability shown in the reference population was fairly high. Nevertheless, our study showed that changes in breeding policies, in order to obtain some sport characteristics in the breed, would lead into a decrease in the genetic variability of the population, with 'Mixed', 'Others' and 'SPB', the genetic lines that would show the greatest decrease.

#### 5. Conclusion

These results highlighted that the SSH breed mainly originates from German sport horse breeds such as the Hostenier, Hanoverian, Westphalian, Oldenburger or Trakehener, thus demonstrating the breeders' preferences for the aptitudes of these foreign breeds.

The SSH population is highly subdivided, due to the fact that several breeds take part in its pedigree and that the breeders have mating preferences for selecting some animals/breeds over others to mate with their SSH animals, in their quest for the best sports performance.

Applying restrictive policies in this population might produce a decrease in the genetic diversity of the SSH population.

Further research is needed to assess how and to what extent this multi-bred composition of the SSH influences their sport ability.

#### Acknowledgments

The authors wish to thank the National Association of Spanish Sport Horse Breeders (ANCADES) for providing the genealogical database of this breed for this study.

#### References

- 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 (1), 37–42.
- Bartolomé, E., Gómez, M.D., Cervantes, I., Romero, F., Molina, A., Valera, M., 2008. Design of a linear type trait system for morphological evaluation of Spanish Sport Horse breed: preliminary results. Poster. 59th Annual Meeting of the European Association for Animal Production, Vilnius (Lithuania). August 24–27th.
- 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 (6), 867–875.
- 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. 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 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 realized effective sizes from real pedigrees. *J. Anim. Breed. Genet.* 125, 301–310.
- Cervantes, I., Gutiérrez, J.P., Molina, A., Goyache, F., Valera, M., 2009. Genealogical analyses in open populations: the case of three Arab-derived Spanish horse breeds. *J. Anim. Breed. Genet.* 126, 335–347.
- Cervantes, I., Goyache, F., Molina, A., Valera, M., Gutiérrez, J.P., 2011. Estimation of effective population size from the rate of coancestry in pedigreed populations. *J. Anim. Breed. Genet.* 128, 56–63.
- Druml, T., Baumung, R., Solkner, J., 2009. Pedigree analysis in the Austrian Noriker draught horse: genetic diversity and the impact of breeding for coat colour on population structure. *Anim. Breed. Genet.* 126, 348–356.
- Dubois, C., Ricard, A., 2007. Efficiency of past selection of the French Sport Horse: Selle Français breed and suggestions for the future. *Livest. Sci.* 112, 161–171.
- 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., 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., Goyache, F., 2005. A note on ENDOG: a computer program for analysing pedigree information. *J. Anim. Breed. Genet.* 122, 172–176.
- 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., 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 realized effective population sizes in farm animals. *J. Anim. Breed. Genet.* 126, 327–332.
- Hagger, C., 2005. Estimates of genetic diversity in the brown cattle population of Switzerland obtained from pedigree information. *J. Anim. Breed. Genet.* 122, 405–413.
- Hamann, H., Distl, O., 2008. Genetic variability in Hanoverian Warmblood horses using pedigree analysis. *J. Anim. Sci.* 86, 1503–1513.
- Hellsten, E.T., Näsholm, A., Jorjani, H., Strandberg, E., Philipsson, J., 2009. Influence of foreign stallions on the Swedish Warmblood breed and its genetic evaluation. *Livest. Sci.* 121, 207–214.
- Honda, T., Fujii, T., Nomura, T., Mukai, F., 2006. Evaluation of genetic diversity in Japanese Brown cattle population by pedigree analysis. *J. Anim. Breed. Genet.* 123 (3), 172–179.
- James, J., 1972. Computation of genetic contributions from pedigrees. *Theor. Appl. Genet.* 42, 272–273.
- James, J.W., 1977. A note on selection differentials and generation length when generations overlap. *Anim. Prod.* 24, 109–112.
- Janssens, S., 2008. Breeding programs and estimated breeding values [online]. Katholieke Universiteit Leuven. Available from. [http://www.bivw.kuleuven.be/genlog/livgen/chgs\\_is\\_Breedprog.html](http://www.bivw.kuleuven.be/genlog/livgen/chgs_is_Breedprog.html) 2008 [Accessed 23/09/08].
- Koenen, E.P.C., Aldridge, L.I., Philipsson, J., 2004. An overview of breeding objectives for Warmblood sport horses. *Livest. Prod. Sci.* 88, 77–84.
- Lacy, R.C., 1989. Analysis of founder representation in pedigrees: founder equivalent and founder genome equivalents. *Zool. Biol.* 8, 111–123.
- Langlois, B., 1982. Heritability of racing ability in thoroughbreds — a review. *Livest. Prod. Sci.* 7, 591–605.
- MacCluer, J., Boyce, B., Dyke, L., Weitzkamp, D., Pfenning, A., Parsons, C., 1983. Inbreeding and pedigree structure in Standardbred horses. *J. Hered.* 74, 394–399.
- Malécot, G., 1948. *Les Mathématiques de l'Hérédité*. Masson et Cie, Paris.
- 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.
- Olsen, H.F., Klemetsdal, G., Ruane, J., Helfjord, T., 2010. Pedigree structure and genetic variation in the two endangered Norwegian horse breeds: Døle and Nordland/Lyngen. *Acta Agric. Scand. Sect. A* 60, 13–22.
- Teegen, R., Edel, C., Thaller, G., 2009. Population structure of the Trakehner Horse breed. *Animal* 3, 6–15.
- Thorén, E., 2008. *International Sport Horse data for genetic evaluation*. Doctoral Thesis. Swedish University of Agricultural Sciences, Uppsala (Sweden).
- Thorén, E., Näsholm, A., Jorjani, H., Strandberg, E., Philipsson, J., 2009. Influence of foreign stallions on the Swedish Warmblood breed and its genetic evaluation. *Livest. Sci.* 121, 207–214.
- 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.
- Valera, M., Bartolomé, E., Cervantes, I., Gómez, M.D., Azor, P.J., Molina, A., 2009. *Horse Breeding Programs in Spain Managed by MERAGEM Group*. CD-ROM, MERAGEM Research Group 978-84-692-4256-8.
- Zechner, P., Sölkner, J., Bodo, I., Druml, T., Baumung, R., Achmann, R., Marti, E., Habe, F., Brem, G., 2002. Analysis of diversity and population structure in the Lipizzan horse breed based on pedigree information. *Livest. Prod. Sci.* 77, 137–146.