

Analysis of founders and performance test effects on an autochthonous horse population through pedigree analysis: structure, genetic variability and inbreeding

A. Giontella^{1†}, C. Pieramati¹, M. Silvestrelli¹ and F. M. Sarti²

¹Department of Veterinary Medicine – Sportive Horse Research Center, University of Perugia, via S. Costanzo 4, 06123 Perugia, Italy; ²Department of Agricultural, Food and Environmental Sciences, University of Perugia, Borgo XX Giugno, 74, 06121 Perugia, Italy

(Received 14 November 2017; Accepted 23 April 2018)

The Maremmano is an autochthonous Italian horse breed, which probably descended from the native horses of the Etruscans (VI century B.C.); the Studbook was acknowledged in 1980, and it includes 12 368 horses born from that year up to 2015. The aim of this study was to evaluate the effect of the selection program on the genetic variability of the Maremmano population; the analysis was performed using both the 'Endog v 4.8' program available at http://webs.ucm.es/info/prodanim/html/JIP_Web.htm and in-house software on official pedigree data. Four Reference Populations were considered, and the most important one was the population of the 12 368 Maremmano horses officially registered in the National Studbook. The pedigree completeness of this population was very good because it was more than 90% at the third parental generation and more than 70% at the fifth generation; the pedigree traced back to a maximum of 10.50 generations with an average of 3.30 complete generations and 5.70 equivalent complete generations. The average generation interval was 10.65 ± 4.72 years, with stallions used for longer periods than mares. The intervals ranged from 10.15 ± 4.45 (mother–daughter) to 10.99 ± 4.93 (father–daughter). The effective number of founders (f_e) was 74 and the effective number of ancestors (f_a) was 30 so that the ratio f_e/f_a was 2.47. The founder genome equivalents (f_g) was 13.72 with a ratio f_g/f_e equal to 0.18. The mean of the genetic conservation index was 5.55 ± 3.37 , and it ranged from 0.81 to 21.32. The average inbreeding coefficient was 2.94%, with an increase of 0.1%/year, and the average relatedness coefficient was 5.52%. The effective population size (N_e) computed by an individual increase in inbreeding was 68.1 ± 13.00 ; the N_e on equivalent generations was 42.00, and this value slightly increased to 42.20 when computed by Log regression on equivalent generations. The analysis confirmed the presence of seven traditional male lines. The percentage of Thoroughbred blood in the foals born in 2015 was 20.30% and has increased 0.21%/year since 1980; in particular, it increased more than twice (0.51%/year) until 1993 and afterwards slightly fluctuated. The pedigree analysis confirmed the completeness of genealogical information and the traditional importance that breeders gave to the male lines; although the genetic diversity of Maremmano seemed to be not endangered by the selection program, some effects on the population structure were found and a more scientific approach to genetic conservation should be incorporated in the selection plans.

Keywords: maremmano horse, founders analysis, genetic diversity, effective population size, selection program

Implications

In this study, the performance test effect, genealogical information and the genetic variability of the Maremmano population were analyzed. The results highlight the importance of monitoring the population structure to avoid the risk of a dangerous reduction in genetic diversity. The presence of seven blood lines, which are transmitted as a paternal attribute along with 440 female lines, is an important factor in maintaining genetic variability across the population that can be used to

obtain a totipotent national saddle horse capable of satisfying all possible demands of an international market.

Introduction

Maremma is the Italian territory along the Tyrrhenian coast of Southwestern Tuscany and Northern Latium. Located about 100 to 200 km Northwestern of Rome, this geographical area spreads in two neighboring Italian provinces: Grosseto in the South of Tuscany region, and Viterbo in the North of Latium; there is evidence of horse breeding in Maremma in the Etruscan civilization (VI century B.C.). The wild and difficult

[†] E-mail: andrea.giontella@unipg.it

environment of Maremma shaped the Maremmano which is the most important native saddle horse. In the past centuries in Italy other Mediterranean populations, like Arab and Spanish, carried out raids, wars and invasions, which shaped the Maremmano; therefore, this horse is strongly linked to Italian history and represents an important cultural resource. These aspects, especially from 1861, along with the evolution of Maremmano, have been described by Bonavolontà (Bonavolontà and Silvestrelli, 1989).

During the time, the Maremmano's versatility allowed it to be adapted to the following requested roles: a stock horse, a war horse, a farm horse, and finally a sport horse. The selection of Maremmano, therefore, was influenced by the historical context and market demands rather than by the attention to safeguard its typical traits. The only unchanged condition during the time was the wild environment in which the horse lives.

The Maremmano maintained its typical identity until the first half of the 19th century when it was strongly crossed with the Thoroughbred and the Arabian horse as other breeds around the world; however, this breed has preserved its genetic diversity until now (Petersen *et al.*, 2013). The Maremmano breeders have traditionally given importance to the male-to-male path so that every living horse is officially assigned to one of the following seven lines: Uszero (born in 1920), Ajace (1926), Otello (1927), Ingres (1946), Le Brun (1960), Noris (1961) and Ognon (1964) (Bonavolontà and Silvestrelli, 1989). Otello is considered as the 'true' Maremmano horse; Uszero was a Salerno horse while the other five were Thoroughbreds.

In the past, the main buyers of Maremmano horses were the cavalry and the so-called *butteri*, 'Italian cow-boys' who used to tend cattle; however, the demand for cavalry horses and draught horses suddenly dropped after World War II causing a severe market decline.

During this period, many prestigious farms stopped horse breeding, and a lot of fine mares were dispersed so that in the '60s this ancient Italian native breed was close to extinction.

At the beginning of the '70s, the local breeder associations started looking for horses that showed the original traits of Maremmano such as suitability to hard work, adaptability to severe environmental conditions; in addition, the following morphological traits were taken into consideration: height around 160 cm, ram's head profile, solid legs with hard hooves, and black or bay coat color. In 1979, the Maremmano horse breeders' National Association (Associazione Nazionale Allevatori Cavallo di Razza Maremmana (ANAM)) was founded, and the following year the Studbook was officially appointed by the Italian law.

A 100-day station test was organized to approve the 3-year old Maremmano stallions in 1993: it was the first horse breeding evaluation in Italy; the test became mandatory 3 years later. Starting from 1999, the Maremmano was officially evaluated by 100-day station test together with Italian Saddle Horse. Because of the economic crisis in 2010, the breeders of other horse breeds suspended the stallion testing; on the contrary, the station testing of the Maremmano breed has continued, although the period was reduced to 70 days.

In addition, Maremmano mares were the first horses to be tested in Italy; in fact, ANAM has organized a 30-day station test since 1996; this test became compulsory 3 years later to appoint 3-year-old 'elite mares.' The results from station testing of both sexes have greatly helped breeders in planning matings.

The aim of this study is to estimate inbreeding and genetic variability of the Maremmano horse breed from genealogical data, to check the contribution of male ancestors on the actual population, and to evaluate the effects of the performance test on the genetic structure and conservation of this important ethnological resource.

Material and methods

Data

The data were obtained from National Horse Breeders Association of Maremmano Breed (ANAM) and from Sport Horse Research Centre (University of Perugia) archives. To improve the quality of the animal registration program, ANAM controlled breeders' declarations by verifying every year the parentage of some new colts by microsatellite markers: because of the possibility of this control, the reliability of the genealogical data in the last generations should be good; however, these data are not sufficient to support a pedigree analysis.

Pedigree completeness

The completeness of pedigree information for each individual was based on: (a) the *number of fully traced generations*, (defined as the generations, g , separating the individual from the furthest generation where all the 2^g ancestors are known); (b) the *maximum number of generations traced* (the number of generations separating the individual from its furthest ancestor); (c) the *equivalent complete generations* (computed as the sum of $(1/2)^n$ where n is the number of generations separating the individual from each known ancestor) (Boichard *et al.*, 1997).

Genetic variability

The genetic variability of the population was identified by the following parameters: (a) *individual inbreeding (F)* (Wright, 1931) defined as the probability that an individual has two identical alleles by descent computed according to Meuwissen and Luo (1992); (b) *average relatedness coefficients (AR)* (Goyache *et al.*, 2003; Gutiérrez *et al.*, 2003) defined as the probability that an allele randomly chosen from the whole population belongs to a given animal. In the founders AR values show their genetic contribution to the entire pedigree (Gutiérrez *et al.*, 2008); (c) *effective number of founders (f_e)* is the number of equally contributing founders that are expected to produce the same genetic diversity of the studied population, it is computed using the AR coefficients of founder individuals; (d) *effective number of ancestors (f_a)* is the minimum number of ancestors, not necessarily founders, explaining the whole genetic diversity of a population. This parameter complements the effective number of founders information in accounting for the bottleneck produced by the

unbalanced use of reproductive individuals (Boichard *et al.*, 1997); (e) *the ratio between the effective number of founders and the effective number of ancestors* (f_d/f_a) indicates if the analyzed population is affected by bottleneck; in fact, a bottleneck effect occurs if f_a is clearly lower than f_e (ratio $f_d/f_a > 1$) (Boichard *et al.*, 1997); (f) *founder genome equivalents* (f_g) (Ballou and Lacy, 1995) can be defined as the number of founders that would be expected to produce the same genetic diversity as in the studied population if the founders are equally represented and no alleles lost occurs (Caballero and Toro, 2000); (g) *the ratio between the founders genome equivalent and the effective number of founders* (f_g/f_e) indicates if the analyzed population is affected by genetic drift (Boichard *et al.*, 1997); (h) the *Genetic Conservation Index (GCI)* (Alderson, 1992) computed as the genetic contributions of all the identified founders. The trend of GCI over time was performed with the regression coefficient of the GCI of each animal on the year of birth (1980–2015) using the *lm* function in *R* software (R Core team of the Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>).

Population structure

The structure of the population was studied by calculating: (a) *F-statistics* (Wright, 1978) from genealogical information following Caballero and Toro (2000; 2002); (b) the *effective population size* (N_e) computed in three ways: via individual increase in inbreeding (Gutiérrez, *et al.*, 2008), via regression, and via Log regression on equivalent generations; (c) the ratio $N_e/2$ that provides information on the occurrence of bottlenecks in the population (in a population where genetic drift has stabilized f_e should be close to $N_e/2$) (Caballero and Toro, 2000); (d) the *generation intervals*, defined as the average age of parents at the birth of their reproducing progeny, as well as the average age of parents at the birth of their progeny used whether for reproduction or not (James, 1977). Both parameters are computed for the four pathways (father–son, father–daughter, mother–son and mother–daughter); (e) the *genetic importance of the geographic area* computed according to Gutiérrez and Goyache (2005), in which regions or provinces were considered instead of herds.

Reference populations

For an efficient pedigree analysis, several ‘reference population’ have to be defined which are constituted by the animals potentially contributing to the next generation (Siderites *et al.*, 2013). For this purpose, in this study the following four reference populations (REF) were considered:

- (1) **REF1** defined by the pedigree file provided by ANAM historical archives that included 15 875 horses from 1920’s to 31 December 2015.
- (2) **REF2** represented by 14 271 Maremmano horses after the exclusion of animals having a Thoroughbred parent.
- (3) **REF3** included 12 368 animals officially registered in the ANAM Studbook.
- (4) **REF4** included 5705 living horses (1532 living parents).

The REF1 is very important to reconstruct the pedigree and breed’s history, and REF2 is considered the population of the ‘typical Maremmano horses,’ while REF3 and 4 are relevant for the genetic selection of the actual breed.

A similar approach in horse (several populations) was considered by other authors (Pjontek *et al.*, 2012; Siderites *et al.*, 2013; Vostrá-Vydrová *et al.*, 2016).

The demographic and genetic parameters were calculated using ‘Endog v 4.8’ software (Gutiérrez and Goyache, 2005), the percentage of Thoroughbred blood in each horse of the population was carried out using an in-house Fortran software and graphed using Microsoft Excel.

Results

Data

In the 1st year of its official activity, the ANAM Studbook registered 203 foals (REF3), and the number did not change in the following 2 years. Then, the number increased to 498 in 1991, reaching its maximum in 1997 (504 foals). Subsequently, the number of foals decreased to 311 in 2000 and remained stable until 2008, when it decreased to 214 by 2015 because of the economic crisis. The number registered in 2015 was almost the same as the number registered when the Studbook first started.

In the first 3 years of the Studbook, more than a half of the foals were used as parents; in the years after that, the number of foals increased, however, the number of stallions and mares selected each year remained unchanged until the introduction of performance testing. This selection method caused a reduction in the number of parents, therefore, a more stable selection intensity; in addition, the number of foals decreased as indicated by the similar trends (Supplementary Figure S1).

The percentage of Thoroughbred blood (Figure 1) strongly influenced the choice of candidates at the performance test; in fact, in the newly born colts, this value was 20.75 ± 1.05 , whereas in the 421 tested males, this value was 21.96 ± 1.81 . This percentage further increased to 23.20 ± 2.43 in the 152 approved stallions. The values of the approved horses were greater than those of the candidates in 17 out of 23 performance tests. The Thoroughbred blood percentage was lower in the female; in fact, the value of newly born fillies was 20.59 ± 0.62 , whereas in the 785 tested mares it increased to 21.97 ± 1.31 . However, only in 5 out of 19 years did the elite mares have a Thoroughbred blood percentage higher than the tested mares so that the 360 elite mares showed a lower value of 21.46 ± 1.67 . As a result of the selection system, approved stallions had a higher value than the newly born animals in 19 out of 23 years, and elite mares in 12 out of 19. Moreover, it should be pointed out that the percentage of Thoroughbred blood in this population has increased 0.21%/year since 1980; however, the increase doubled to 0.51% in 1993 and remained stable in the following years.

Pedigree completeness

The completeness of the pedigree information (number of fully traced generations) is shown in Figure 2. The REF4 had

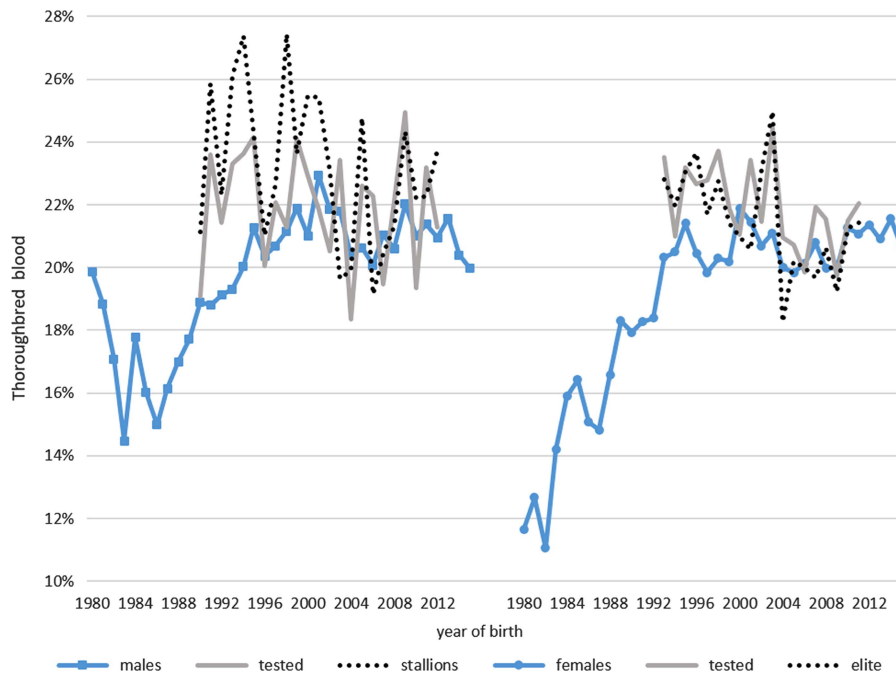


Figure 1 Percentage of the Thoroughbred blood of the Reference population 3 (12 368 Maremmano horses officially registered in the National Studbook) per year of birth and sex (all males, males tested in the performance test, and approved stallions; all females, females tested in performance test and elite mares).

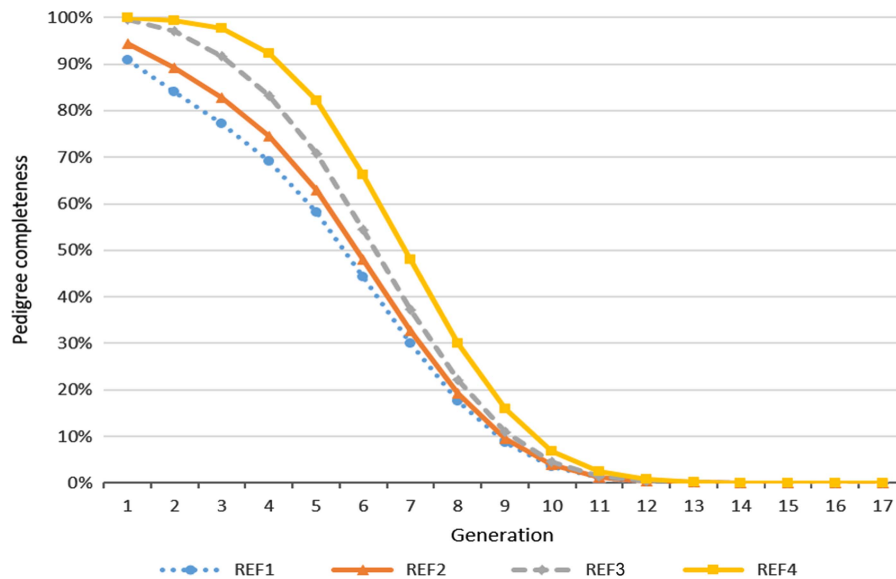


Figure 2 Pedigree completeness per generation of REF1 (Reference population of 15 875 horses in the complete pedigree), REF2 (Reference population of 14 271 Maremmano horses after the exclusion of animals having a Thoroughbred parent), REF3 (Reference population 12 368 Maremmano horses officially registered in the National Studbook), REF4 (Reference population of 5705 living horses).

very good values of completeness: 99.90%, 99.40%, 97.70%, 92.30%, and 82.20% from the first to the fifth parental generation. On the other hand, REF3 had obviously lower percentages of pedigree completeness (99.50%, 97.20%, 91.60%, 83.20% and 70.90%, respectively), and REF2 and REF1 had further decreased values. All populations showed a similar pattern of completeness by generation, and both these values and the values of the oldest records can be considered satisfactory.

As shown in Figure 3, the paternal pathways seem to be more complete than the maternal pathways, and this

difference has been decreasing in time. In fact, the difference in completeness between parents was below 1% in both REF3 and REF4; in the grandparent's generation, the difference increased to 3.50% in the Studbook (REF3), whereas it was still below 1% in living Maremmano horses (REF4). In the grand-grandparents, the difference was 3.60% in the living animal population (REF4), although it was more than 11% of the Studbook population (REF3).

In the REF2 population, the mean of the maximum number of generations traced was 9.4; the mean of the complete generations was 3.0, and the mean of equivalent complete

Performance test effect on a horse population

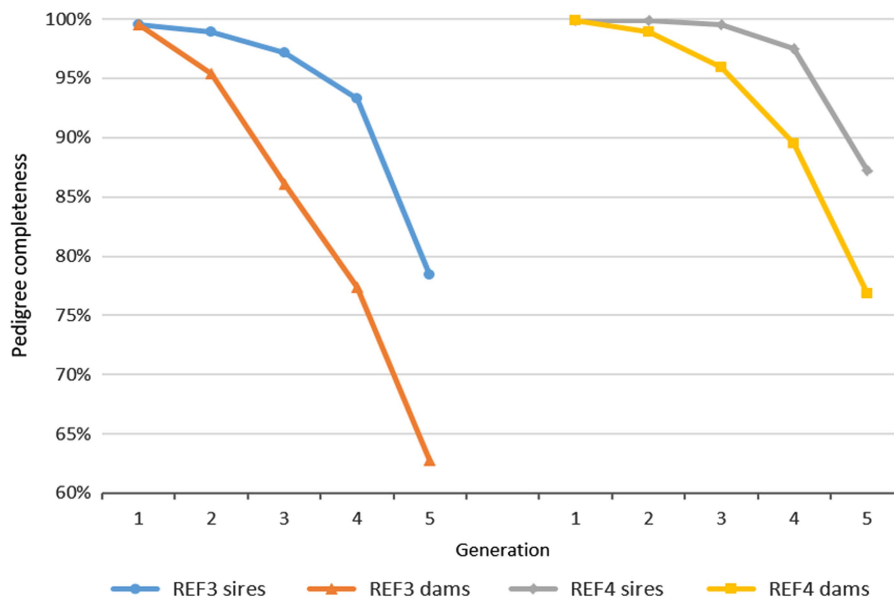


Figure 3 Pedigree completeness per generation of the paternal and maternal pathways in the REF3 (Reference population 12 368 Maremmano horses officially registered in the National Studbook) and REF4 (Reference population of 5705 living horses).

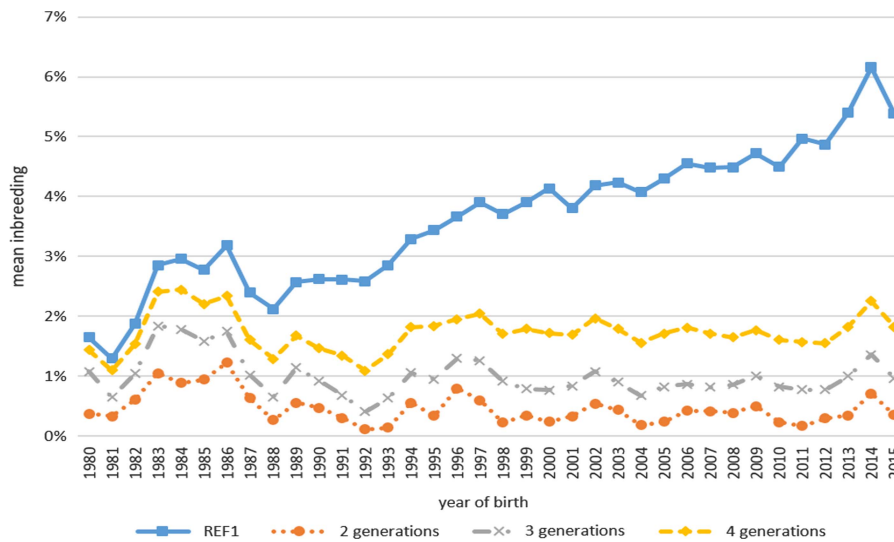


Figure 4 Mean inbreeding in REF1 (Reference population of 15 875 horses in the complete pedigree) Maremmano horses per year of birth.

generations 5.2. In the REF3 population, these values increased to 10.5, 3.3 and 5.7, whereas in REF4 they were 11.6, 3.8 and 6.4, respectively.

Genetic variability

The average inbreeding (F) in the whole population (REF1) was 2.90% and ranged from a minimum of 1.29% (1981) to a maximum of 6.16% (2014) as shown in Figure 4; this value increases on average 0.10%/year although it decreased to 5.38% in 2015. The average inbreeding computed with 2, 3 and 4 generations was 0.46%, 0.99% and 1.73%, respectively. The values of the average inbreeding in the three other reference populations were as follows: 2.63% in REF2, 2.94% in REF3 and 3.55% in the REF4.

The mean of the AR in the REF1 population was 5.16 % and ranged from a minimum of 0.01% to a maximum of

14.61%. In the three other populations, this value was 2.01% in REF2, 5.52% in REF3, and 6.13% of the living animals (REF4).

The effective number of founders (f_e) in the four studied populations was 84 (REF1), 87 (REF2), 74 (REF3) and 64 (REF4); however, only 13 individuals can explain 50% of the genetic variability of the breed in REF1, 11% in REF2 and REF3 and 8% in REF4.

The effective number of ancestors (f_a) was 35 in REF1, 36 in REF2, 30 in REF3 and 25 in REF4.

The ratio f_e/f_a was 2.40 in REF1, 2.42 in REF2, 2.47 in REF3 and 2.56 in REF4. The founder genome equivalents (f_g) were 19.37, 16.79, 13.72 and 11.06 in REF1, 2, 3 and 4 populations, respectively.

The ratio f_g/f_e was 0.23 in REF1, 0.19 in REF2, 0.18 in REF3 and 0.17 in REF4.

The GCI index was calculated from the genetic contributions of all the identified founders of the population, and it is assumed that every horse that has a more balanced quantity of genes from a large number of founders has a higher GCI value (McManus *et al.*, 2013).

In the REF3 population, the mean of GCI was 5.55 ± 3.37 and ranged from a minimum of 0.81 to a maximum of 21.32; the regression coefficient estimated on the animals born from 1980 to 2015 was equal to 0.034 ± 0.009 ($P < 0.001$). This index computed by sex was equal to 6.03 ± 3.31 in the males and 5.22 ± 3.37 in the females. In the official Maremmano Studbook population (REF3), an increase of GCI was noted along the observed period both in males and females; in fact, the values of the regression coefficients on the year of birth are 0.028 ± 0.008 ($P < 0.01$) and 0.034 ± 0.009 ($P < 0.001$), respectively.

In order to ascertain if GCI was effectively influenced by the selection system, this parameter was calculated in the tested parents and their progeny (Figure 5). The 'new colts' (foals born after the beginning of stallion performance test) had a GCI value of 6.23 ± 0.28 , and this value decreased to 5.98 ± 0.87 in tested males. It must be observed that the GCI value further decreased to 5.56 ± 1.36 in the approved males: in fact, the group of approved males had a lower GCI than the candidates in 15 out of 23 performance tests. The GCI reduction was not present in the females path; in fact, the newly born fillies and the tested mares had very similar values (6.03 ± 0.32 and 6.11 ± 0.55 , respectively). The elite group had a further increase of the GCI value to 6.16 ± 0.83 , whereas the elite mares had lower GCI than the tested mares only in 10 out of 19 performance tests.

Table 1 shows the percentage of the seven founders represented in REF3, REF4, living parents (1 532 horses), tested males and females, as well as stallions and elite mares. In the REF3 and the living parents, Ajace presented the highest gene percentage (18.1% and 18.6% respectively). However, the most prevalent founder in REF4 (16.8%), tested males (27.0%), mares (19.6%), stallions (25.85%) and elite mares (20.9%)

was Otello followed by Ajace, Ussero, and Ingres. Le Brun, Ognon and Noris were scarcely represented. According to the number of descendants, the breeders call the first four stallions 'the major lines' and the last three ones 'the minor lines.' Because there is no living stallion connected through the paternal bloodline to Le Brun, Noris or Ognon, these minor lines can be considered extinct. In fact, when the Studbook was officially founded, the blood contribution in each of the three main lines was near 2.5% (Supplementary Figure S2), and it increased to 4% in the following years. Ajace has been in the third place since 1994, and Ingres has been in the first place since 2003. Ussero was the lowest contributor amongst the main lines: in fact, his blood percentage ranged between 1% and 2%. The blood percentage in each of the three minor lines (Le Brun, Noris and Ognon) was always below 1%.

Population structure

The effective population size (N_e) computed by an individual increase in inbreeding was 68.1 ± 13.0 in REF3 population

Table 1 Percentage of founders represented in horses of different subsets according to their paternal line

	Paternal line						
	Otello	Ajace	Ussero	Ingres	Le Brun	Ognon	Noris
REF3	13.6	18.1	6.0	5.5	0.7	0.3	0.2
REF4	16.8	15.5	6.2	8.2	0.7	0.4	0.2
Living parents	16.7	18.6	6.3	10.8	0.8	0.3	0.2
Tested males	27.0	15.0	4.6	8.6	0.3	0.5	0.3
Stallions	25.8	16.1	3.6	10.2	0.5	0.6	0.1
Tested mares	19.6	14.3	6.9	9.4	0.6	0.5	0.2
Elite mares	20.9	14.5	7.2	9.2	0.5	0.4	0.2

REF3 = Reference population (12 368 animals officially registered in the Associazione Nazionale Allevatori Cavallo di Razza Maremmana Studbook); REF4 = Reference population 4 (5705 living horses).

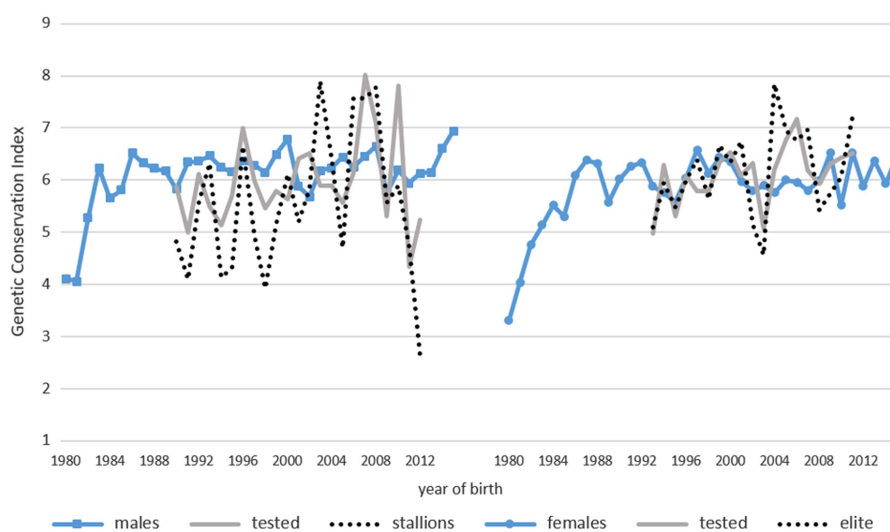


Figure 5 Genetic Conservation Index of the Reference population 3 (12 368 Maremmano horses officially registered in the National Studbook) per year of birth and sex (all males, males tested in the performance test and approved stallions; all females, females tested in performance test and elite mares).

and 71.2 ± 12.5 in the REF4. The effective population size on equivalent generations was 42.0 in the REF3 and 36.6 in REF4; however, these values slightly changed to 42.2 and 36.9, respectively, when computed by Log regression on equivalent generations. The ratio $N_e/2$ (where N_e is computed by individual increase in inbreeding) was equal to 34 in REF3 and 36 in REF4.

The average age (Table 2) was 10.08 ± 4.50 years, ranging from 9.81 ± 4.32 (mother–daughter) to 10.40 ± 4.71 (father–daughter), and the average age in sires was longer than in dams because stallions are used for longer periods than mares. The generation interval showed rather higher values than the average age expected.

As for the geographical area (Table 3), Tuscany seems to be the most important because 48.7% of the foals (6 953) were born in this region, and 74.3% of their fathers were born in the same area. Only 2 378 foals (16.7%) were born out of the origin area although 85.7% of their fathers still came from Tuscany or Latium. Discussing this in more detail and taking into account only the most significant origin provinces, it can be observed that the stallions from Grosseto and Viterbo respectively produced 39.0% and 25.0% of all Maremmano foals. The *F*-statistics were calculated to evaluate the within-population genetic differences between Latium, Tuscany and other regions, and the values were as

follows: $F_{ST} = 0.001568$, $F_{IT} = -0.000208$ and $F_{IS} = -0.001778$. If the provinces of Viterbo and Grosseto were considered instead of Latium and Tuscany, the values were as follows: $F_{ST} = 0.002789$, $F_{IT} = -0.000208$ and $F_{IS} = -0.003006$. The increase of the F_{ST} was expected because the two provinces are more separated than the two regions, whereas the decrease of F_{IS} indicates that the breeders are more aware of the inbreeding problems in these limited areas.

Discussion

Pedigree completeness

Some authors have studied horse breeds with similar consistency as Maremmano. In the Spanish Arab horse the completeness of the pedigree was higher than 90% until the sixth generation (Cervantes *et al.*, 2008). In the Austrian draft horse (Noriker), 90% of ancestors were known up to the 9th generation (Druml *et al.*, 2009). In the German Paint Horse, the completeness in the second generation was respectively 99.9%, 99.9% and 99.6% for the RP 1 (reference population 1 registered animals born in Germany between the years 2000 and 2009), RP 2 (reference population 1 with a manually improved pedigree) and RP 3 (third reference population including 679 horses based on animals born in Germany between 1990 and 1999) (Siderits *et al.*, 2013). In the Pantaneiro, the number of pedigrees known over the generations increased mainly between the 4th and 5th generation (96%) (McManus *et al.*, 2013). In Turkish Arab horse, the pedigree completeness was 100% in both first and second generations although it decreased markedly after the seventh generation (Duru, 2017).

In larger horse populations, the estimated completeness was slightly different. In the Andalusian horse, this parameter was higher than 90% considering the first five generations of ancestors (Valera *et al.*, 2005), whereas in the Brazilian Sport horse (Medeiros *et al.*, 2014), the completeness for all animals born until 1996 was equal to 99.80, 78.60, 28.10 and 6.70% from first to fourth parental generation, respectively. In the same breed and the same generations, the completeness was equal to 99.99, 98.30, 69.90 and 37.8 in more recent pedigrees (1997–2012).

Finally, Zechner *et al.* (2002) reported that in Lipizzano (3867 animals) at the first five generations, the pedigrees were virtually complete, and even in the 10th generation 90% of the ancestors were known. Pjontek *et al.* (2012) studied some endangered horse populations bred in Slovakia and found that the first four generations of pedigrees were virtually complete in Lipizzan (162 horses) and Shagya Arabian (171 horses), whereas the proportion of the known ancestors dropped to <50% after 11 generations in the Lipizzan, 10 generations in the Shagya Arabian, and seven generations in the Hucul (158 horses). In the Old Klauer horse (7971 horses), 100% of the complete pedigree was known until the sixth generation (Vostrá-Vydrová *et al.*, 2016).

Genetic variability

As already mentioned, in the Maremmano the value of average inbreeding in the REF1 population was 2.90% and

Table 2 Average age and generation interval and (years) of parents at the birth of their offspring for the REF3 Maremmano horses

	<i>n</i>	Interval \pm SD	MSE
Average age			
Father–son	280	9.82 ± 4.36	0.26
Father–daughter	2464	10.40 ± 4.71	0.28
Mother–son	280	9.91 ± 4.14	0.25
Mother–daughter	2464	9.81 ± 4.32	0.26
Overall	5488	10.08 ± 4.50	0.06
Generation interval			
Father–son	5533	11.14 ± 4.93	0.07
Father–daughter	6771	10.99 ± 4.93	0.07
Mother–son	5532	10.34 ± 4.50	0.06
Mother–daughter	6782	10.15 ± 4.45	0.06
Overall	24 618	10.65 ± 4.72	0.03

REF3 = reference population (12 368 animals officially registered in the Associazione Nazionale Allevatori Cavallo di Razza Maremmana Studbook).

Table 3 Distribution of REF2 Maremmano foals and use of stallions by geographic area

	Other				
	Latium	Tuscany	regions	Viterbo	Grosseto
Number of foals born in the area	4939	6953	2378	3499	5614
Percentage of father from same area	40.3	74.3	14.3	41.6	65.0

REF2 = reference population 2 (14 271 Maremmano horses after the exclusion of animals having an Thoroughbred parent).

increased to 3.55% in REF4. It is well known that this value is very influential in pedigree completeness and quality because there is a positive correlation between these two parameters (Valera *et al.*, 2005; Druml *et al.*, 2009). In pedigrees with similar size, the average inbreeding reached 7.0% in Spanish Arab Horse (Cervantes *et al.*, 2008), 5.1% in Austrian Noriker (Druml *et al.*, 2009), 4.6% in Turkish Arab (Duru, 2017) and 4% in Pantaneiro (McManus *et al.*, 2013), whereas it was still 0.15% in Brazilian Sport Horse (Medeiros *et al.*, 2014). Therefore, in the four studied Maremmano REF populations, the average inbreeding coefficients seem to be generally lower than in other horse populations. As previously discussed in deeper pedigrees, this value was more stable and ranged between 8% and 9% (Valera *et al.*, 2005; Vicente *et al.*, 2012), whereas in small populations, it ranged from a minimum of 0.35% (Siderits *et al.*, 2013) to a maximum of 13.1% (Vostrá-Vydrová *et al.*, 2016).

In animal breeding, AR can be used to preserve the genetic pool; as a matter of fact, the use for reproduction of animals with the lowest AR can minimize the inbreeding because they are expected to balance the contribution of the founders in the population.

In Maremmano, the AR values were much lower than the values reported for the Andalusian horse (12.20%) (Valera *et al.*, 2005) and the Turkish Arab (9.50%) (Duru, 2017). These results are related to the inbreeding coefficients mentioned above; in fact, 93.40% of the animals registered in Andalusian Studbook and 94.20% of those registered in Turkish Arab Horse Herdbook had an inbreeding coefficient >0 , whereas in the whole Maremmano pedigree this percentage dropped to 69.90%.

The pedigree analysis of Maremmano showed the presence of 440 founder mares, and 14 of them were born before World War II. This situation had already been observed in a previous mitochondrial analysis (Lancioni *et al.*, 2011) whose results were in agreement with the Studbook data.

The bottleneck (ratio f_b/f_a) in Maremmano (2.40 to 2.50) was higher than that observed in Spanish Arab (0.98), Turkish Arab (1.81), Pantaneiro (all = 1.05; recent = 1.10) and Lusitano (2.34) reference populations (Cervantes *et al.*, 2008; Vicente *et al.*, 2012; McManus *et al.*, 2013; Duru, 2017). The progressive reduction in genetic variability was confirmed by genetic drift (f_b/f_e ratio). Because of this situation, the selection plans should also consider the maintaining of genetic variability; however, a higher bottleneck was estimated in Austrian Noriker (4.00) and in the population of Paint German horse (RP1: 2.69, RP2: 5.17, RP3: 4.26) (Druml *et al.*, 2009; Siderits *et al.*, 2013).

The GCI is useful to estimate the effective number of founders in a pedigree, so animals with a high value of GCI are crucial for the breed conservation; in fact, the ideal individual would receive an equal contribution from all the founder ancestors. Due to lack of pedigree information in Maremmano, this index (5.55) is lower than the index in Lusitano (53 411 horses) where it reaches 9.5 ± 2.73 (Vicente *et al.*, 2012). On the other hand, the Maremmano GCI is >1.33 value that was observed in Pantaneiro horse (10 441 animals) (McManus *et al.*, 2013).

The trend of GCI in our population has increased over time in both males and females. However, in recent years, this index has rapidly decreased in stallions, whereas it has increased in the elite females. This situation is certainly due to the use of Thoroughbred blood in males to obtain remarkable sporting aptitude rapidly; whereas, in the elite female, the increase of GCI values is due to their selection plan that strongly preserves the founder's genes. This fact is also confirmed by the variability in maternal genetic contributions (Achilli *et al.*, 2012; Cardinali *et al.*, 2016), which was higher when the Maremmano horse breed was officially founded (1980) and was maintained over the years, balancing the decrease that was observed in the 'stallions' GCI. In this way, it is possible to maintain two different horse lines in the breed: 'sports' and 'traditional.'

These trends were shown by the positive value of the regression coefficient of GCI on the year of birth computed in the Maremmano REF2 which revealed an increase over the time. In Lusitano (Vicente *et al.*, 2012), an opposite trend (-0.009747 ± 0.00150) was observed; in this breed, the founders representation was, in fact, progressively lost.

The good results in Maremmano selection and conservation are probably due to both the completeness of pedigree data and the suitable breeding program that takes into account the founders genes.

As previously reported, there were four 'major' and three 'minor' founders lines in Maremmano; although the breeders focus their choice on these male lines, it should be noted that there is a different percentage of founders in each of these lines. Besides being the less represented in the population, the three minor line founders also have lower percentages than the overall founders. The blood percentage of Ingres in the population is fourfold the percentage of Ussero. When a breeder selects a sire from the Ussero line, there is only 1% of its genes, whereas in a sire from the Ingres line there is 4% of genes from Ingres; if breeders select an elite mare from the Ussero line, she preserves 7.2% of the Maremmano founders' genes, and this percentage is just a little higher (9.2%) in a mare from the Ingres line. In a stallion from the Otello line, which is probably the most famous and wanted the line, there is still 25% of Maremmano founders' genes.

The geographical genetics analysis by the Wright's F shows that Viterbo and especially Grosseto are actually the 'centers' of the breed although the value of F_{ST} indicates that there is no difference between the two subpopulations. Furthermore, the negative values of F_{IT} and F_{IS} pointed out that probably the breeders choose mating systems that contain the inbreeding.

Population structure

The values of N_e computed via individual increase in inbreeding in the REF3 and REF4 Maremmano populations (68.10 to 71.20) were close to the values estimated in Turkish Arab (74.40) (Duru, 2017) and were higher than the values reported by Vicente *et al.* (2012) in Lusitano (41.24); however, they were lower than the values of the Brazilian Sport horse (205.78) (McManus *et al.*, 2013). In addition, the

N_e calculated on equivalent generations (42.00 to 36.60) and on Log regression (42.20 to 36.90) were also higher than those estimated in Brazilian Sport Horse (25.06 to 22.77) (Medeiros *et al.*, 2014). The values of $N_e/2$ in the REF3 and REF4 Maremmano populations (34–36) together with the effective number of founders (87 to 64) and the ratio f_e/f_a (2.42, 2.56) confirmed, as already reported that a bottleneck in this breed is occurring.

The average generation intervals estimated on the four parent-offspring pathways were close to those of Andalusian horse (10.11), Lusitano (10.28) and American Quarter Horse (10.5). (Valera *et al.*, 2005; Vicente *et al.*, 2012 Petersen *et al.*, 2014). Lower generation intervals were observed in Pantaneiro horse (8.20 years) and Slovak Sport Pony (9.96) (Pjontek *et al.*, 2012; McManus *et al.*, 2013). Higher values of generation intervals were observed in the Hucul horse (11.14), Carthusians (11.50), Lipizzaner (11.61), Shagya Arabian horse (12.27) and Turkish Arab (12.40) (Pjontek *et al.*, 2012; Valera *et al.*, 2005; Duru, 2017).

Conclusions

This study confirms the completeness of genealogical information and the traditional importance that breeders gave to the male lines; these two factors had a key role in helping the Maremmano horse to preserve much of its genetic diversity. In the first decade of ANAM activity, the goal of the association was the conservation of the breed, and it seems to have been achieved. In the following years the goal changed, and a more intensive selection took place by means of performance testing. The genetic diversity of Maremmano was not endangered by intensive selection although there have been some effects on the population structure. In our opinion, the traditional methods of preserving genetic diversity are no longer sufficient, and a more scientific approach should be incorporated in the selection plans.

Acknowledgments

The authors would like to express their special appreciation and thanks to ANAM – Maremmano Horse Breeders Association for providing pedigree data and to two anonymous referees for providing constructive input.

Declaration of interest

The authors have no conflict of interest to declare.

Ethics statement

The approval of the work by an ethics committee was not required by Italian legislation.

Software and data repository resources

The Maremmano horse pedigree dataset is the property of Associazione Nazionale Allevatori Cavallo di Razza Maremmana, via Canova 13/a – 58100 Grosseto (Italy) www.anamcavallomaremmano.com

Supplementary material

To view supplementary material for this article, please visit <https://doi.org/10.1017/S1751731118001180>

References

- Achilli A, Olivieri A, Soares P, Lancioni H, Kashani BH, Perego UA, Nergadze SG, Carossa V, Santagostino M, Capomaccio S, Felicetti M, Al-Achkar W, Penedo MCT, Verini-Supplizi A, Houshmand M, Woodward SR, Semino O, Silvestrelli M, Giulotto E, Pereira L, Bandelt HJ and Torroni A 2012. Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proceedings of the National Academy of Sciences* 109, 2449–2454.
- Alderson L 1992. A system to maximize the maintenance of genetic variability in small populations. In *Genetic conservation of domestic livestock* (ed. L Alderson and I Bodó), pp 18–29. CABI Publishing, Wallingford, UK.
- Ballou JD and Lacy RC 1995. Identifying genetically important individuals for management of genetic variation in pedigreed populations. In *Population management for survival and recovery: analytical methods and strategies in small population management* (ed. JD Ballou, M Gilpin and TJ Foose), pp 76–111. Columbia University Press, New York, NY, USA.
- Boichard D, Maignel L and Verrier É 1997. The value of using probabilities of gene origin to measure genetic variability in a population. *Genetics Selection Evolution* 29, 5–23.
- Bonavolontà G and Silvestrelli M 1989. *Il Maremmano*. Bramante edizioni Equestri, Milano, Italy.
- Caballero A and Toro MA 2000. Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genetics research* 75, 331–343.
- Cardinali I, Lancioni H, Giontella A, Capodiferro MR, Capomaccio S, Buttazzoni L, Biggio GP, Cherchi R, Albertini E, Olivieri A, Cappelli K, Achilli A and Silvestrelli M 2016. An Overview of Ten Italian Horse Breeds through Mitochondrial DNA. *Plos One* 11, 4.
- Cervantes I, Molina A, Goyache F, Gutierrez JP and Valera M 2008. Population history and genetic variability in the Spanish Arab Horse assessed via pedigree analysis. *Livestock Science*. 113, 24–33.
- Druml T, Baumung R and Sölkner J 2009. Pedigree analysis in the Austrian Noriker draught horse: genetic diversity and the impact of breeding for coat colour on population structure. *Journal of Animal Breeding and Genetics* 126, 348–356.
- Duru S 2017. Pedigree analysis of the Turkish Arab horse population: structure, inbreeding and genetic variability. *Animal*. 9, 1449–1456.
- Goyache F, Gutiérrez JP, Fernández I, Gómez E, Álvarez I, Díez J and Royo LJ 2003. Using pedigree information to monitor genetic variability of endangered populations: the Xalda sheep breed of Asturias as an example. *Journal of Animal Breeding and Genetics* 120, 95–103.
- Gutiérrez JP and Goyache F 2005. A note on ENDOG: a computer program for analysing pedigree information. *Journal of Animal Breeding and Genetics* 122, 172–176.
- Gutiérrez JP, Altarriba J, Díaz C, Quintanilla AR, Cañón J and Piedrafita J 2003. Genetic analysis of eight Spanish beef cattle breeds. *Genetics Selection Evolution* 35, 43–64.
- Gutiérrez JP, Cervantes I, Molina A, Valera M and Goyache F 2008. Individual increase in inbreeding allows estimating realised effective sizes from pedigrees. *Genetics Selection Evolution* 40, 359–378.
- James JW 1977. A note on selection differentials and generation length when generations overlap. *Animal Production*. 24, 109–112.
- Lancioni H, Capomaccio S, Felicetti M, Cappelli K, Silvestrelli M, Verini Supplizi A and Achilli A 2011. Horse Mitochondrial Genome Analyses to Define the Origin of Ancient Human Populations. In *Proceeding of the joint AGI-SIBV-SIGA Meeting*, 19–22th September 2017, Assisi, Italy, 2A.61.
- McManus C, Santos SA, Lima Dallago BS, Rezende Paiva S, Saraiva Martins RF, Braccini Neto J, Rocha Marques P and Pinto de Abreu UG 2013. Evaluation of conservation program for the Pantaneiro horse in Brazil. *Revista Brasileira de Zootecnia* 42, 404–413.
- Medeiros BR, Bertoli CD, Garbade P and McManus CM 2014. Brazilian Sport Horse: pedigree analysis of the Brasileiro de Hipismo breed. *Italian Journal of Animal Science* 13, 3146.
- Meuwissen T and Luo Z 1992. Computing inbreeding coefficients in large populations. *Genetics Selection Evolution* 24, 305–313.

- Petersen JL, Mickelson JR, Cleary KD and McCue ME 2014. The American Quarter Horse: population structure and relationship to the Thoroughbred. *Journal of Heredity* 105, 148–162.
- Petersen JL, Mickelson JR, Cothran EG, Andersson LS, Axelsson J, Bailey E, Bannasch D, Binns MM, Borges AS, Brama P, da Câmara Machado A, Distl O, Felicetti M, Fox-Clipsham L, Graves KT, Guérin G, Haase B, Hasegawa T, Hemmann K, Hill EW, Leeb T, Lindgren G, Lohi H, Lopes MS, McGivney BS, Mikko S, Orr N, Penedo MCT, Piercy RJ, Raekallio M, Rieder S, Røed KH, Silvestrelli M, Swinburne J, Tozaki T, Vaudin M, Wade CM and McCue ME 2013. Genetic diversity in the modern horse illustrated from genome-wide SNP Data. *Public Library of Science One* 8, e54997.
- Pjontek J, Kadlečík O, Kasarda R and Horný M 2012. Pedigree analysis in four Slovak endangered horse breeds. *Czech Journal of Animal Science* 57, 54–64.
- Siderits M, Baumung R and Fuerst-Waltl B 2013. Pedigree analysis in the German Paint Horse: genetic variability and the influence of pedigree quality. *Livestock Science*. 151, 152–157.
- Valera M, Molina A, Gutiérrez JP, Gómez J and Goyache F 2005. Pedigree Analysis in the Andalusian horse: population structure, genetic variability and influence of the Carthusian strain. *Livestock Production Science* 95, 57–66.
- Vicente AA, Carolino N and Gama LT 2012. Genetic diversity in the Lusitano horse breed assessed by pedigree analysis. *Livestock Science*. 148, 16–25.
- Vostrá-Vydrová H, Vostrý L, Hofmanová B, Krupa E and Zavadilová L 2016. Pedigree analysis of the endangered Old Kladruber horse population. *Livestock Science*. 185, 17–23.
- Wright S 1931. Evolution in Mendelian population. *Genetics* 16, 97–159.
- Wright S 1978. Evolution and the genetics of populations. Volume 4: variability within and among natural populations. University of Chicago Press, Chicago, IL, USA.
- Zechner P, Sölkner J, Bodo I, Druml T, Baumung R, Achmann R, Marti E, Habe F and Brem G 2002. Analysis of diversity and population structure in the Lipizzan horse breed based on pedigree information. *Livestock Production Science* 77, 137–146.