J. Anim. Breed. Genet. ISSN 0931-2668

ORIGINAL ARTICLE

Analysis of genetic diversity and the determination of relationships among western Mediterranean horse breeds using microsatellite markers

D. Marletta[1], I. Tupac-Yupanqui[2], S. Bordonaro[1], D. García[2], A. M. Guastella[1], A. Criscione[3], J. Cañón[2] & S. Dunner[2]

1 DACPA, Sez. di Scienze delle Produzioni Animali, Facoltà di Agraria, Università degli studi di Catania, Catania, Italy
2 Laboratorio de Genetica, Facultad de Veterinaria, Universidad Complutense de Madrid, Madrid, Spain
3 DISTAFA, Facoltà di Agraria, Università degli studi Mediterranea, Polo Universitario di Feodivito, Reggio Calabria, Italy

Keywords
equine breeds; genetic diversity; microsatellites; molecular coancestry; population structure.

Correspondence
Prof. Donata Marletta, DACPA, sezione Scienze delle Produzioni Animali, Facoltà di Agraria, Università degli studi di Catania, via Valdisavoia, 5 - 95123 Catania, Italy. Tel.: +39-95-234478; Fax: +39-95-234345;
E-mail: d.marletta@unict.it

Received: 5 April 2006;
accepted: 10 June 2006

Summary

The distribution of genetic diversity and the genetic relationships among western Mediterranean horse breeds were investigated using microsatellite markers. The examined sample included seven Spanish and three Italian local horse breeds and populations, plus a Spanish Thoroughbred outgroup. The total number of animals examined was 682 (on average 62 animals per breed; range 20–122). The microsatellite marker set analysed provided 128 alleles (10.7 alleles per locus). Within-breed genetic diversity was always high (>0.70), with breeds contributing about 8% of the total genetic variability. The mean molecular coancestry of the entire population examined was 0.205, Losino being the breed that contributed most. In addition to Nei's standard and Reynolds' genetic distances, pair-wise kinship distance and molecular coancestry were estimated. Remarkably similar breed rankings were obtained with all methods. Clustering analysis provided an accurate representation of the current genetic relationships among the breeds. Determining coancestry is useful for analysing genetic diversity distribution between and within breeds and provides a good framework for jointly analysing molecular markers and pedigree information. An integrated analysis was undertaken to obtain information on the population dynamics in western Mediterranean native horse populations, and to better determine conservation priorities.

Introduction

In the Mediterranean area, the gradual replacement of the horse by mechanization in agriculture has led to a severe reduction in the number of surviving breeds. Of the 127 breeds in the Mediterranean region, 23 no longer exist, 59 are endangered, 31 are not at risk and 14 are unknown (Nardone 2000). However, the Mediterranean region is home to a rich diversity of horse breeds.

In Spain, Aparicio (1960) reported two main equine types: the Celtic type of tarpanic origin in the northern and western region of the peninsula, and the Spanish type, descendant from the African Barb horse, in the south and the east. The Spanish Celtic horses, represented by a number of semi-feral local breeds and populations (Jaca Navarra, Asturcon, Caballo Gallego, Losina, Pottoka) with certain 'primitive' traits (such as below average size, different coat colours, etc.) are mainly bred in the north along the

Cantabrian and show a clear genetic differentiation from the Mediterranean breeds, Menorquina and Mallorquina, native to the Balearic islands, which possess African and Arabian genes (Cañón *et al.* 2000).

In Sicily, local ancient horse populations were crossed with horses of diverse origin, including Oriental and African horses (Balbo 1995). Nowadays, three local populations are bred in Sicily. The Sanfratellano breed, of which there are about 2000 head (including a selection nucleus of some 600 animals), is reared in extensive systems in the mountainous north-east of the island. The Sicilian Oriental Purebred, an endangered breed with less than 70 animals remaining, represents the Sicilian selection nucleus of the Arabian horse. A studbook has existed since 1923, but genealogical data are available from 1880. The Sicilian Indigenous is a very heterogeneous population. Largely unmanaged, it probably originated from a primitive strain of Sicilian horse and from the Borbon Real Casa di Ficuzza breed (related to the Neapolitan, Persano and Arabian breeds).

These western Mediterranean populations show some common features: they are vigorous, well adapted to harsh conditions and stress-resistant. The size of most of their populations is small; some are threatened by extinction, and they have not been systematically subjected to reproductive or breeding technologies. As studbooks have only recently been compiled, the possibility of genetic introgression between these populations cannot be ruled out. According to FAO recommendations, determining classic genetic distances using neutral, highly polymorphic microsatellite markers is the method of choice for investigating genetic relationships and breed differentiation. This methodology also provides information for establishing preservation priorities for livestock breeds (Barker 1999). Nevertheless, phylogenetic methods based on genetic distances cannot be exhaustive in their quantification of genetic diversity between closely related groups because they ignore within-group genetic variation (Ruane 1999). Genetic diversity is a function of relatedness between individuals that share one or more ancestors. The level of relatedness can be expressed as a coefficient of kinship, which appears to be of central importance in the definition and measurement of genetic diversity (Eding & Meuwissen 2001). A molecular coancestry-based method has recently been proposed for analysing the genetic diversity in subdivided populations (Caballero & Toro 2002). This uses molecular data to obtain coancestry coefficients that integrate the well-known genetic distances D_S, D_R and F_{ST}, taking into account the allelic frequencies and diversity of supposed founder populations. However, only a few authors have explored this approach for analysing actual data sets (Caballero & Toro 2002; Fabuel *et al.* 2004). A valid alternative is a clustering method that uses individual multilocus genotypes to infer population structure and to assign individuals to theoretical populations (Pritchard *et al.* 2000).

The aim of this article was to assess the genetic diversity in a set of 10 western Mediterranean horse breeds and to determine their relationships by different methods.

Materials and methods

Sample collection

Blood samples were taken from seven Spanish and three Italian local horse breeds and populations, and from a Spanish Thoroughbred outgroup. The entire sample included 682 horses (122 Jaca Navarra, 119 Asturcon, 72 Caballo Gallego, 66 Losina, 51 Pottoka, 31 Menorquina, 20 Mallorquina, 61 Sanfratellano, 50 Sicilian Oriental Purebred, 30 Sicilian Indigenous and 60 Spanish Thoroughbred horses), all from Spain and Sicily (Italy) (Table 1).

Animals for genotyping were chosen from large, well-defined geographical areas (with no consideration of their possible genetic relationships), sampling as many herds as possible in order to obtain a representative sample. In the case of the Sicilian Oriental Purebred, the sample collected ($n = 50$) consisted of nearly the entire studbook-registered population ($n = 64$).

Microsatellite analysis

To search for 12 microsatellite markers (HTG4, HTG6, HTG7, HTG10, HMS2, HMS3, HMS6, HMS7, VHL20, ASB2, AHT4, AHT5), DNA was amplified in three multiplex reactions according to standard protocols using MJ Research PTC200 (MJ Research, Reno, NV, USA) and 9700 ABI Geneamp thermocyclers (Applied Biosystems, Foster City, CA, USA) (see Cañón *et al.* 2000). Fluorescent-labelled polymerase chain reaction (PCR) products were diluted, mixed with an internal size standard, and analysed using ABI PRISM® 3100 and ABI PRISM® 377 genetic analysers, both equipped with Genescan® and Genotyper® software (Applied Biosystems, Foster City, CA, USA). Seven reference DNA samples were used

Table 1 Name, origin, population size, studbook existence, number (n) of samples, expected heterozygosities (He), F_{IS} values, mean number of alleles (MNA), allelic richness (AR) for each of the eleven horse breeds across all 12 microsatellite loci

Breed name	Country	Population size	Existence of Studbook (year)	Sample size	H_e	F_{IS}	MNA	AR
Asturcon	Spain	~1000	Yes (1981)	119	0.733	−0.009	7.17	5.8
Jaca Navarra	Spain	~250	No	122	0.729	0.035*	7.83	6.2
Losino	Spain	<200	No	66	0.704	−0.022	6.83	5.9
Caballo Gallego	Spain	>10 000	No	72	0.761	0.060*	7.92	7.0
Pottoka	Spain	~1000	Yes (1995)	51	0.775	0.042*	7.08	6.6
Menorquin	Spain	1400	Yes (1993)	31	0.721	−0.040	6.25	6.0
Mallorquin	Spain	<200	Yes (1993)	20	0.748	0.054*	6.08	6.1
Sanfratellano	Italy	<700	No	61	0.751	0.019	6.92	6.2
Sicilian Oriental	Italy	<70	Yes (1923)	50	0.702	−0.003	6.67	5.7
Sicilian Indigenous	Italy	~400	No	30	0.803	0.098*	7.92	7.4
Spanish Thoroughbred	Spain			60	0.690	−0.029	5.17	4.7

*Values different from 0 at $p < 0.05$.

to ensure compatibility between the results obtained at the two participating laboratories.

Statistical analysis

Unbiased estimates of gene diversity (expected heterozygosity or Hardy–Weinberg heterozygosity), observed heterozygosity, and the number of alleles per breed (the last two with their associated standard errors), were calculated using the Microsatellite Toolkit (Park 2001). The FSTAT program (Goudet 2001) was used for testing the conformity of genotype frequencies to the Hardy–Weinberg equilibrium. Sequential Bonferroni analysis (Rice 1989) was used to achieve a global type I error of 0.05. FSTAT software was also used to calculate the allelic richness (AR) standardized for variation in sample size. When calculated in this way, the number of alleles between populations with different sample sizes can be compared.

Wright's F coefficients (F_{IT}, F_{IS} and F_{ST}) were calculated using Genetix 4.0 software (Belkhir *et al.* 2001) The significance of the F_{IT} and F_{IS} values were tested by permuting alleles 1000 times within the set of breeds or within each breed, respectively. Significant deviation of F_{ST} from the null hypothesis was tested for by random permutations of genotypes among samples.

Two classical measures of genetic distance were made using Population software (Langella 2002): the standard Nei (D_S) distance, which increases linearly with time, and the Reynolds (D_R) distance, which under a pure genetic drift model (excluding mutations and admixtures) also increases linearly with time. In addition, genetic diversity components between and within breeds were calculated based on the molecular coancestry concept (Caballero & Toro 2002). The molecular coancestry (f) between two individuals (say i and j) at a given locus is expressed as $1/4[I_{11} + I_{12} + I_{21} + I_{22}]$, where $I_{xy} = 1$ when allele x in the first individual and allele y on the same locus in the second individual are identical, and 0 when they are not. For L number of loci, molecular coancestry is obtained by simply dividing by the number of analysed loci

$$f_{ij} = \frac{\sum_{i=1}^{L} f_{ij,l}}{L}.$$

The molecular coancestry of an individual i with itself is the self-coancestry (s_i), which is related to homozygosity (F_i) by the expression $F_i = 2s_i - 1$. Genetic diversity can be partitioned as:

$$(1 - \bar{f}) = (1 - \tilde{s}) + (\tilde{s} - \tilde{f}) + (\tilde{f} - \bar{f})$$

where

$$\bar{f}(\text{the average global coancestry}) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij}}{n^2},$$

$$\tilde{f} = \frac{\sum_{i=1}^{n} f_{ii}}{n}$$

and

$$\tilde{s} = \frac{\sum_{n=1}^{n} s_i}{n},$$

and n the number of breeds.

The average kinship distance for breed i is: $D_{ii} = (s_i - f_{ii})$. The proportion of diversity between individuals for breed i is: $G_i = D_{ii}/(1 - f_{ii})$. The total genetic

diversity $GD_T = (1 - \bar{f})$ is partitioned into three components: the genetic diversity within individuals [$GD_{WI} = (1 - \tilde{s})$], the genetic diversity between individuals [$GD_{BI} = (\tilde{s} - \tilde{f})$], and the genetic diversity between breeds [$GD_{BS} = (\tilde{f} - \bar{f})$].

Finally, the kinship between two individuals i and j was used as a measure of genetic distance among individuals within breeds or among breeds, calculated as $D_{ij} = ([s_i + s_j]/2) - f_{ij}$ (Caballero & Toro 2002), and the proportional contribution of each breed to the estimated global coancestry.

The number of genetic clusters maximizing the likelihood of the molecular data, and the proportion of the individual's genome derived from each inferred theoretical ancestral population or cluster, were estimated using Structure software (Pritchard *et al.* 2000). This model-based clustering program uses a Monte Carlo Markov chain (MCMC) algorithm to assess the presence of a structure underlying the genetic information provided by the genetic markers. A genetic distance between the breeds was defined using the genetic composition output from this program. If $q_k(i)$ is the proportion of the genome of breed i that comes from population k, then the distance between breeds i and j can be defined as (Cañón *et al.* 2006):

$$d_s(i,j) := \sum_{k=1}^{K} |q_k(i) - q_k(j)| \frac{q_k(i) + q_k(j)}{2}.$$

Thus, two breeds are genetically closer when the proportion of their shared genomes is higher and when these have an important weighting in the final genetic composition. It is obvious that $d_S(i,j) \geq 0$, that $d_S(i,j) = d_S(j,i)$, $d_S(i,i) = 0$, and that $d_S(i,j) = 0$ implies that i and j have exactly the same mean genetic composition, which in practice is only likely to happen if i and j are the same population. Distances were averaged over five runs at $K = 10$ and represented on a splits-graph based on the split-decomposition method (Dress *et al.* 1996); this routine is part of the SplitsTree program (Huson & Bryant 2006).

Results

Genetic diversity within breeds

Table 2 shows the number of alleles for each microsatellite marker, which ranged from 6 to 14 (average 10.7). The observed and the expected heterozygosities for each marker are also given. Table 1 shows the average number of alleles and the AR of each breed. The Sicilian Indigenous population showed the highest values for all these variables; the Spanish Thoroughbred showed the lowest. Table 1 also shows the values for the expected heterozygosity under Hardy–Weinberg equilibrium and the measure of the possible discrepancy between this value and the observed heterozygosity (F_{IS}). The negative F_{IS} values of some breeds indicate an excess of heterozygous genotypes with respect to the expected value. The F_{IS} values for the rest of the breeds, even if statistically significant, are not far from 0, indicating that mating is close to panmixia.

Table 3 shows the values for the molecular self-coancestry of each breed (f_{ii}), the average individual self-coancestries (s_i), the within-breed kinship distances (D_{ii}), the average inbreeding (F_i), and the proportion of diversity between individuals(G_i). The table highlights two main points: the similarity of the values for many of these parameters, as well as

Table 2 Number of alleles, range of allele sizes, *F*-statistics (F_{ST}, and F_{IS}), and expected (H_e) and observed (H_o) heterozygosities for each of the 12 microsatellite markers in 11 European horse breeds

Locus	No. of alleles	Range allele sizes	F_{ST}	F_{IS}	H_o	H_e
HTG-4	10	116–138	0.078	−0.011	0.698	0.692
HTG-6	12	82–106	0.138	0.010	0.630	0.634
HTG-7	6	117–127	0.094	−0.018	0.614	0.608
HTG-10	12	86–108	0.093	0.061	0.720	0.798
HMS-2	13	216–240	0.131	0.017	0.749	0.750
HMS-3	10	148–170	0.136	0.142	0.634	0.728
HMS-6	9	152–174	0.135	0.033	0.661	0.689
HMS-7	9	166–182	0.066	−0.036	0.810	0.784
VHL20	13	81–111	0.057	−0.008	0.809	0.798
ASB-2	14	216–250	0.057	0.005	0.777	0.799
AHT-4	11	137–159	0.075	−0.010	0.804	0.808
AHT-5	9	124–140	0.064	0.014	0.768	0.767

Table 3 Molecular self-coancestry (f_{ii}), average of individual self-coancestry (s_i), within-breed kinship distance D_{ii}, inbreeding (F_i), proportion of diversity between individuals (G_i) and proportional contribution of each breed i to the global coancestry of the 10 Mediterranean horse breeds included in the study

	f_{ii}	s_i	D_{ii}	F_i	G_i	Distance within breed	Distance to other breeds	Total
Asturcon	0.271	0.631	0.360	0.261	0.493	0.0271	0.0070	0.0201
Jaca Navarra	0.275	0.648	0.374	0.297	0.515	0.0275	0.0065	0.0210
Losino	0.302	0.640	0.339	0.281	0.485	0.0302	0.0071	0.0230
Caballo Gallego	0.244	0.642	0.398	0.284	0.527	0.0244	0.0050	0.0194
Pottoka	0.233	0.629	0.396	0.258	0.516	0.0233	0.0051	0.0182
Menorquin	0.290	0.625	0.335	0.250	0.472	0.0290	0.0075	0.0216
Mallorquin	0.270	0.646	0.376	0.292	0.515	0.0270	0.0089	0.0182
Sanfratellano	0.255	0.631	0.376	0.262	0.504	0.0255	0.0069	0.0186
Sicilian Oriental	0.305	0.648	0.343	0.296	0.493	0.0305	0.0114	0.0191
Sicilian Indigenous	0.210	0.638	0.427	0.276	0.541	0.0210	0.0065	0.0145

the coincidence in the f_{ii} and F_i values in most breeds. These features, together with G_i values close to 0.5 for many of the breeds, indicate random mating practices. In the Asturcon, Losino, Menorquina and Oriental Purebred breeds, the G_i values were lower than 0.5, showing variability to be preferentially distributed within individuals.

Table 3 also shows the proportional contribution of each breed to the global coancestry of these horses. The Sicilian Oriental Purebred and Losino breeds were found to contribute the most to global coancestry, a consequence of their own coancestry values. However, the total contribution of Losino is higher because of its closer relationships with the other breeds.

Excluding the Spanish Thoroughbred, the values of 0.277 for mean coancestry, 0.268 for average inbreeding, 0.638 for average self-coancestry, and 0.205 for mean coancestry in the whole population, summarize the genetic diversity. The total genetic diversity (0.795) can be distributed into three components: the genetic diversity between breeds ($GD_{BS} = 0.063$), the genetic diversity within individuals ($GD_{WI} = 0.362$), and the genetic diversity between individuals ($GD_{BI} = 0.370$). The sum of the last two components represents the genetic diversity within breeds (0.732). Expressing these values as percentages provides the corresponding F values: $F_{IS} = 0.012$, $F_{ST} = 0.079$ and $F_{IT} = 0.091$.

Genetic distances and clustering

Together, Tables 4 and 5 show the Nei's standard distances, the Reynolds' genetic distances and the pairwise kinship distance and molecular coancestry matrix. The genetic distances for the Spanish Thoroughbred were greater than for any other breed. The Oriental Purebred and Mallorquina breeds were, on average, the most distant breeds, independent of the genetic distance measurement used.

Table 4 Nei's standard D_S (above the diagonal) and Reynolds' D_R genetic distances (below the diagonal) among 10 Mediterranean horse breeds

	Asturcon	Jaca Navarra	Losino	Caballo Gallego	Pottoka	Menorquin	Mallorquin	Sanfratellano	Sicilian Oriental	Sicilian Indigenous
Asturcon		0.224	0.221	0.151	0.227	0.394	0.449	0.368	0.712	0.392
Jaca Navarra	0.068		0.192	0.127	0.164	0.275	0.462	0.358	0.765	0.354
Losino	0.071	0.063		0.160	0.232	0.365	0.520	0.342	0.622	0.360
Caballo Gallego	0.042	0.036	0.048		0.125	0.188	0.312	0.268	0.595	0.319
Pottoka	0.061	0.045	0.067	0.028		0.151	0.243	0.301	0.615	0.281
Menorquin	0.109	0.080	0.110	0.049	0.037		0.278	0.483	0.649	0.484
Mallorquin	0.112	0.116	0.137	0.072	0.053	0.070		0.596	0.790	0.560
Sanfratellano	0.099	0.097	0.099	0.068	0.072	0.122	0.132		0.342	0.152
Sicilian Oriental	0.179	0.189	0.175	0.148	0.148	0.172	0.185	0.098		0.278
Sicilian Indigenous	0.091	0.085	0.093	0.067	0.055	0.107	0.107	0.031	0.073	

Table 5 Pairwise kinship distance (above diagonal) and molecular coancestry (below diagonal) among 10 Mediterranean horse breeds

	Asturcon	Jaca Navarra	Losino	Caballo Gallego	Pottoka	Menorquin	Mallorquin	Sanfratellano	Sicilian Oriental	Sicilian Indigenous
Asturcon		0.421	0.406	0.416	0.430	0.439	0.466	0.448	0.498	0.473
Jaca Navarra	0.218		0.407	0.417	0.424	0.422	0.475	0.453	0.513	0.474
Losino	0.229	0.237		0.410	0.425	0.427	0.473	0.438	0.481	0.463
Caballo Gallego	0.221	0.228	0.231		0.425	0.413	0.456	0.444	0.494	0.475
Pottoka	0.200	0.215	0.210	0.210		0.403	0.441	0.448	0.495	0.466
Menorquin	0.189	0.215	0.205	0.220	0.224		0.423	0.457	0.481	0.479
Mallorquin	0.172	0.172	0.170	0.188	0.197	0.212		0.492	0.516	0.506
Sanfratellano	0.183	0.186	0.198	0.193	0.182	0.171	0.146		0.440	0.434
Sicilian Oriental	0.141	0.135	0.163	0.151	0.144	0.155	0.130	0.200		0.451
Sicilian Indigenous	0.161	0.169	0.176	0.165	0.167	0.152	0.136	0.200	0.192	

The Caballo Gallego breed was the most similar to all the others when classical genetic distances were calculated, and the Losino breed the most similar when coancestry-based genetic distances were used. The similarity of breed ranking obtained with all the genetic distance methods is remarkable (Table 6).

The picture provided by kinship distances, calculated by using the molecular coancestry concept, shows a high level of coancestry and a low genetic distance (D_k) between the Losino and Jaca Navarra breeds, and a very low kinship distance between the Pottoka and Menorquino breeds. These relationships, shown in the dendrogram in Figure 1, show the breeds to group into two main clusters: the Italian and Spanish groups. The length of the branches corresponds to the level of inbreeding; the Sicilian Oriental Purebred, Mallorquino and Menorquino breeds therefore appear to have suffered higher levels of inbreeding than the rest.

This value is the result of assuming no migration between breeds after their reproductive isolation. The information obtained with the Bayesian model-based procedure, however, gives an idea of gene flow between breeds. In fact, a way to see this differentiation (or any kind of phylogeny among the breeds analysed) is to examine the population structure using the clustering model-based method (Pritchard *et al.* 2000), and assuming different K values (i.e. the number of clusters). Using this method, the Italian and Spanish breeds always form clusters of their own, indicating a high degree of isolation (Figure 2). As K increases, new subgroups arise, but

Table 6 Product–moment correlation for distance matrices comparison: molecular coancestry (f), Nei's standard (D_S), Reynolds' (D_R), and kinship distance (D_k)

Distance	D_S	f	D_k
f	−0.891		
D_k	0.902	−0.986	
D_R	0.969	−0.773	0.796

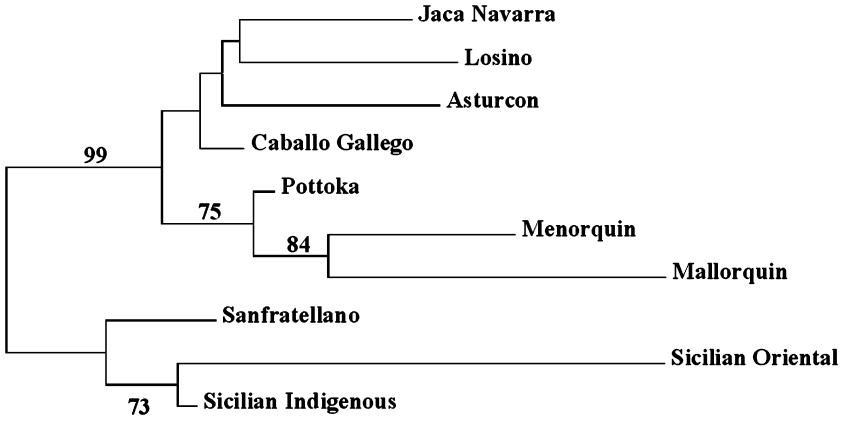

Figure 1 Neighbour-joining tree based on kinship distances for the 10 Mediterranean horse breeds (numbers refer to the fraction of bootstraps, greater than 50%, supporting each branch).

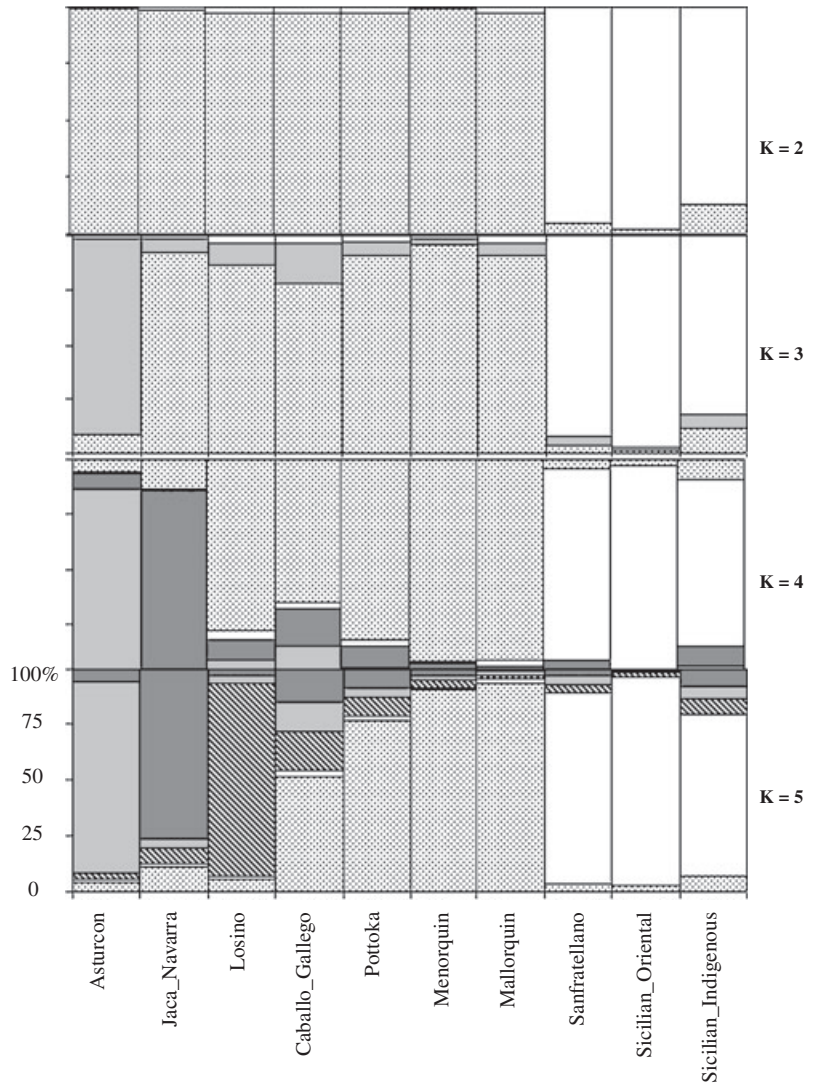

Figure 2 Percentage membership of each breed to each of the possible clusters, assuming $k = 2$ to 5. Each colour or grid represents the percentage membership of each breed, represented by a vertical parallelogram.

Table 7 Proportion of memberships of each breed assigned to each one of the 10 possible clusters

	Inferred clusters									
Breed	1	2	3	4	5	6	7	8	9	10
Asturcon	0.453	0.019	0.344	0.042	0.014	0.016	0.030	0.017	0.007	0.057
Jaca Navarra	0.027	0.252	0.028	0.260	0.028	0.046	0.283	0.030	0.008	0.039
Losino	0.025	0.027	0.028	0.019	0.032	0.029	0.033	0.752	0.012	0.042
Caballo Gallego	0.040	0.038	0.089	0.059	0.069	0.098	0.124	0.079	0.016	0.388
Pottoka	0.028	0.033	0.031	0.062	0.288	0.221	0.073	0.062	0.017	0.186
Menorquin	0.014	0.034	0.019	0.019	0.741	0.078	0.032	0.025	0.008	0.031
Mallorquin	0.012	0.009	0.030	0.007	0.275	0.561	0.016	0.013	0.011	0.066
Sanfratellano	0.022	0.051	0.030	0.017	0.013	0.047	0.040	0.028	0.711	0.042
Sicilian Oriental	0.010	0.007	0.008	0.007	0.017	0.021	0.011	0.016	0.869	0.037
Sicilian Indigenous	0.035	0.064	0.018	0.035	0.018	0.100	0.046	0.049	0.577	0.058

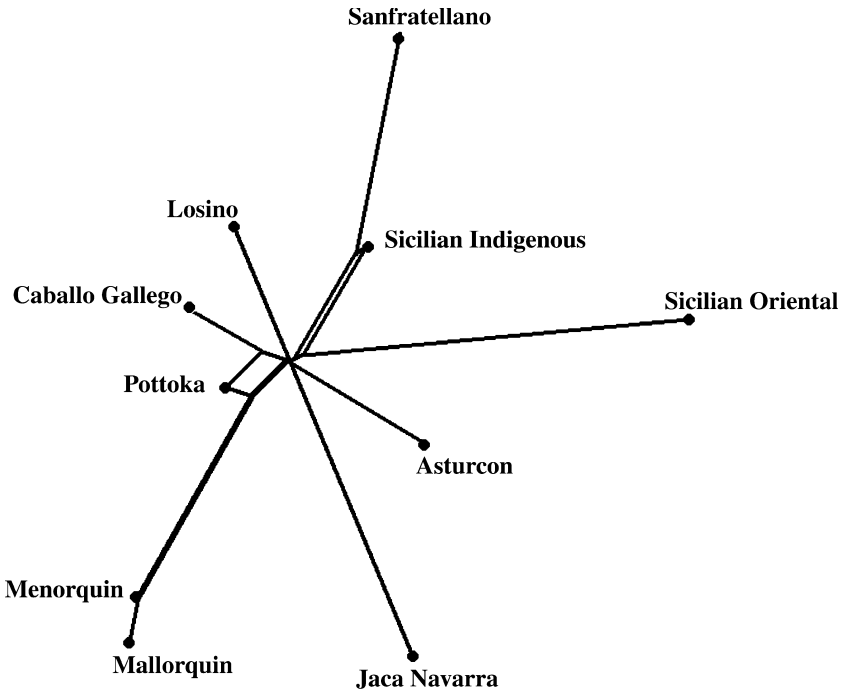

Figure 3 Splits-graph for the 10 Mediterranean horse populations based upon the split-decomposition method and on $q_k(i)$ (see *Materials and methods*).

only because the genomes of the Spanish breeds become distributed across the different clusters. When the number of clusters assumed is the same as the number of breeds (10), the results shown in Table 7 are obtained. It can be seen that the Italian breed genomes group fall within the same cluster, while those of the Spanish breeds, with the exception of those of the Losino and Menorquino breeds [which are mainly assigned to their own cluster (75.2% and 74.1%, respectively)], belong to two or three different clusters. Even so, the horses of the Balearic Islands share a main ancestral population with the Pottoka breed (Table 7). Figure 3 shows the relationships among all breeds based on the proportion of the genome shared across clusters using a splits-graph. The goodness-of-fit estimate for this splits-graph (a measure of the accuracy of the representation of the data set) is particularly good (97.5%). In this type of graph, the sum of the lengths of all the edges along the shortest path from one breed to another is proportional to the actual distance between those breeds in the data set.

Discussion

The total genetic diversity of the analysed horse populations was high (0.795) and mostly (~92%) retained within breeds. The number of alleles observed in these 10 native populations (7.07) was the same as that observed for Lipizzan horse from seven national studs (Achmann *et al.* 2004), greater than that observed in 11 Asian and two European populations (Tozaki *et al.* 2001), in Thoroughbred from four national studs, but one (Cunningham *et al.* 2001) and in other European breeds (Aberle *et al.* 2004; Glowatzki-Mullis *et al.* 2005).The genetic diversity was high (>0.70), except for the Spanish Thoroughbred, and comparable with that of other horse breeds (Wimmers *et al.* 1998; Bjørnstad & Røed 2001; Cunningham *et al.* 2001).

As expected, the more intensive and effective the selection, the less genetic variability. The Spanish Thoroughbred showed the lowest gene diversity values (0.690), followed by the Oriental Purebred (0.702) among the Sicilian populations, and the Losina breed (0.704) among the Spanish ones. Both these breeds have very small population sizes. Heterozygosity in these was higher than that seen in other endangered European and Asiatic breeds (Iriondo *et al.* 2002; Tozaki *et al.* 2003).

It is noteworthy that, despite the small sample size (n = 30), the Sicilian Indigenous population showed the greatest genetic variability, probably because of its wide genetic base and heterogeneity (which is apparent in its morphology). For all the breeds, genetic diversity was consistent with allele richness rank, i.e. it was independent of the sample size.

The coefficient F_{IS}, which indicates the degree of departure from random mating, showed a small deficit of heterozygotes (1.63%) as revealed by a single marker (HMS3). At the population level, the negative F_{IS} value could be explained by the existence of subpopulations which were pooled together to form the present populations (Wahlund effect), and/or by an asymmetrical sex migration that produced an outbreeding effect in the progeny. This would have resulted in an excess of heterozygotes. This is quite noticeable in the Menorquina, Spanish Thoroughbred and Losina breeds, but less clear in the Asturcon and Oriental Purebred breeds. In these populations, F_i is lower than f_{ii}, and G_i is lower than 0.5, showing that genetic variability is preferentially distributed within individuals. The negative values could also be explained as the effect of a breeding strategy that avoids mating between closely related animals.

Positive F_{IS} values mean a significant deficit of heterozygotes. Such values were shown by the Sicilian Indigenous, Caballo Gallego, Pottoka, Mallorquina, and Jaca Navarra breeds. This, plus the related discrepancy between f_{ii} and F_i, might be explained by subdivisions in these breeds, and might even be partially due to samples from genetically different subpopulations being obtained. Furthermore, because of the breeding systems followed, the establishment of male lineages in each breed cannot be ruled out.

The coancestry values seemed to be quite similar for all the tested breeds. The most important data were the high value of f_{ii} and the degree of inbreeding (F_i) in the endangered Oriental Purebred breed. In addition, the Sicilian Indigenous breed showed high D_{ii} and G_i values, and therefore the greatest genetic variation. This is probably a reflection of a still heterogeneous population.

The pairwise genetic distances estimated by D_S and D_R, the methods of choice for assessing relationships among breeds, were virtually the same. The classic genetic distances D_S and D_R are a consequence of the change in the degree of inbreeding after breed separation. Coancestry (f), however, is a measure of similarity, and the kinship distance (D_K) takes into account the allele frequencies in the founder population (Eding & Meuwissen 2001; Eding *et al.* 2002). The correlations between all the genetic distances measured for the studied horse breeds were highly significant. This might be explained by a smaller influence of the allele frequencies in the founder population on kinship distances. This finding highlights the importance of estimating molecular coancestry when trying to explain the genetic distances between breeds.

The strong coancestry and low D_k between the Losino and Jaca Navarra breeds suggest recent differentiation. The reduced kinship distance between the Pottoka and Menorquina breeds indicate that they have received genetically similar migrants rather than having followed a common evolutionary path.

The large genetic distance between the Sicilian and Balearic breeds was unexpected. Both breeds were influenced by the Arabian horse (Cañón *et al.* 2000) and belong to the Mediterranean trunk. Nonetheless, the Oriental Purebred and Mallorquino breeds appear to be the most divergent (in terms of their D_s, D_R and D_k values) and to show the lowest coancestry value (f_{ii}). This could be due to the absence of migration, or reproductive isolation and consequent genetic drift. Nowadays, the Mallorquino and Menorquino breeds are probably the products of the genetic drift caused by their own geographical and reproductive isolation. Cañón *et al.* (2000) reported that the Mallorquino, Menorquino and Losino breeds show a clear clustering among the Iberian horses; this is confirmed in the present work by their high level of molecular coancestry and by the results of the Structure software analysis.

When assuming the same number of clusters and breeds, cluster analysis showed that, on average, most of the genomes of the breeds with small population sizes (characterized by the highest coancestry values, i.e. the Sicilian, Mallorquino, Menorquino and Losino breeds), grouped into a single cluster. The Balearic breeds appear to partially share a cluster, whereas most of the Sicilian horses belong to the same cluster (meaning they derive from a single ancestral population). Moreover, it should be noted that the splits-graph (Figure 3) shows a quite tree-like form (Figure 1), revealing different evolutionary pressures and/or the origin of migrants to the Balearic and Sicilian breeds. This is probably because of the common history of Sicilian horses, all of which were influenced by Oriental breed genes. Despite the short pedigrees recorded, Oriental Stallions were admired for their properties and were bred in Sicily for many centuries. This surely influenced the Sicilian Oriental Purebred and other native populations.

The present paper only considers information provided by neutral markers. These are of interest for estimating population histories and for inferring relationships, but in conservation campaigns, additional genetic variability based on specific loci related with disease resistance or adaptive features, as well as

information on traits of economic impact must be considered jointly. The use of molecular coancestry for measuring genetic diversity is appealing because the interrelation between these parameters and others based on pedigree information is immediately provided. Thus, the integration of both types of information in the same framework is feasible (see Caballero & Toro 2002). Genetic markers are a useful tool for estimating coancestry in the absence of pedigree information – the usual situation when considering unconnected populations or breeds. However, within a breed it is also usual to have pedigree data, which might better help estimate genealogical coancestry.

In the light of the available data for western Mediterranean native horse populations, classical genetic distances tend to overestimate the genetic diversity in inbred populations, in which genetic drift causes extremely different allele frequencies. This contrasts with that seen in large, non-inbred, populations characterized by their even distributions of allele frequencies. Therefore, genetic diversity information, evaluated by integrating within- and between-population analyses may allow conservation priorities to be better established.

Acknowledgements

We are grateful to José L. Vega-Pla who provided samples from some of the Spanish breeds. Financial support was provided by ACPRA, a grant AGF95-064 and by FONDO 6/F Assistenza Tecnica in Zootecnia.

References

Aberle K.S., Haman H., Drögemüller C., Distl O. (2004) Genetic diversity in German draught horse breeds compared with a group of primitive, riding and wild horses by means of microsatellite DNA markers. *Anim. Genet.*, **35**, 270–277.

Achmann R., Curik I., Dovc P., Kavar T., Bodo I., Habe F., Marti E., Solkner J., Brem G. (2004) Microsatellite diversity, population subdivision and gene flow in the Lipizzan horse. *Anim. Gen.*, **35**, 285–292.

Aparicio G. (1960) *Zootecnia Especial*. Imprenta Moderna. Córdoba, Spain.

Balbo S. M. (1995) L'influenza dell'arabo-orientale sul cavallo siciliano. In: Savier M. (ed.), *L'Asil Arabo - Il cavallo nobile d'Arabia*. R&R Editrice, Spoleto (PG), Italy, pp. 169–175.

Barker J.S.F. (1999) Conservation of livestock breed diversity. *Anim. Genet. Res. Inf.*, **25**, 33–43.

Belkhir K., Borsa P., Chikhi L., Raufaste N., Bonhomme F. (2001) *Genetix, logiciel sous Windows TM pour la génétique des populations, Laboratoire Génome, Populations, Interactions, CNRS UPR 9060, Université de Montpellier II, Montpellier (France)*. http://www.genetix.univ-montp2.fr/genetix/genetix.htm

Bjørnstad G., Røed K.H. (2001) Breed demarcation and potential for breed allocation of horses assessed by microsatellite markers. *Anim. Genet.*, **32**, 59–65.

Caballero A., Toro M.A., (2002) Analysis of genetic diversity for the management of conserved subdivided populations. *Conserv. Genet.*, **3**, 289–299.

Cañón J., Checa M.l., Carleos C., Vega-Pla J.l., Vallejo M., Dunner S., (2000) The genetic structure of Spanish Celtic horse breeds inferred from microsatellite data. *Anim. Genet.*, **31**, 39–48.

Cañón J., García D., García-Atance M.A., Obexer-Ruff G., Lenstra J.A., Ajmone-Marsan P., Dunner S., the ECONOGENE Consortium (2006) Geographical partitioning of goat diversity in Europe and the Middle East. *Anim. Genet.*, **37**, 327–334.

Cunningham E.P., Doley J.J., Splan R.K., Bradley D.G. (2001) Microsatellite diversity, pedigree relatedness and the contributions of the founder lineages to thoroughbred horses. *Anim. Gen.*, **32**, 360–364.

Dress A.W.M., Huson D.H., Moulton V. (1996) Analyzing and visualizing sequence and distance data using SplitsTree. *Discrete Appl. Math.*, **71**, 95–109.

Eding H., Meuwissen T.H.E. (2001) Marker-based estimates of between and within population kinships for the conservation of genetic diversity. *J. Anim. Breed. Genet.*, **118**, 141–159.

Eding H., Crooijmans R.P.M.A., Groenen M.A.M., Meuwissen T.H.E. (2002) Assessing the contribution of breeds to genetic diversity in conservation schemes. *Genet. Sel. Evol.*, **34**, 613–633.

Fabuel E., Barragán C., Silió, L., Rodríguez M.C., Toro M.A. (2004) Analysis of genetic diversity and conservation priorities in Iberian pigs based on microsatellite markers. *Heredity*, **93**, 104–113.

Glowatzki-Mullis M.L., Muntwyler J., Pfister W., Marti E., Rieder S., Poncet P.A., Gaillard C. (2005) Genetic diversity among horse populations with a special focus on the Franches-Montagnes breed. *Anim. Gen.*, **37**, 33–39.

Goudet J. (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). http://www2.unil.ch/popgen/softwares/fstat.htm

Huson D.H., Bryant D. (2006) Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.*, **23**, 254–267.

Iriondo M., Jugo B., Rendo F., Solis A., Mazon L.I., Estonba A. (2002) Analysis of genetic diversity of indigenous cattle, horse and sheep breeds from the Basque

country. In:Proceedings of the 7th World Congress on Genetics Applied to Livestock Production. Montpellier, France, 19–23 August, 2002.

Langella O. (2002) Population 1.2.28. Logiciel de génétique des populations. Laborotoire Populations, génétique et évolution, CNRS UPR 9034, Gif-sur-Yvette. http://www.legs.cnrs-gif.fr

Nardone A. (2000) Weather conditions and genetics of breeding systems in the Mediterranean area. In:Proceedings XXXV Simposio Internazionale di Zootecnica. Produzioni animali di qualità ed impatto ambientale nel sistema mediterraneo, Ragusa, Italy, 25 May 2000. pp. 67–91.

Nei M., (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Park S.D.E. (2001) *Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection*, PhD thesis, University of Dublin, Dublin, Ireland.

Pritchard J.K., Stephens M., Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **55**, 945–959.

Rice W.R. (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

Ruane J., (1999) A critical review of the values of genetic distance studies in conservation of animal genetic resources. *J. Anim. Breed. Genet.*, **116**, 767–779.

Tozaki T., Kakoi H., Mashima S., Hirota K., Hasegawa T., Ishida N., Miura N., Choi-Miura N.H., Tomita M. (2001) Population study and validation of paternity testing for Thoroughbred horses by 15 microsatellite loci. *J. Vet. Med. Sci.*, **63**, 1191–1197.

Tozaki T., Takezaki N., Hasegawa T., Ishida N., Kurosawa M., Tomita M., Saitou N., Mukoyama H. (2003) Microsatellite variation in Japanese and Asian horses and their phylogenetic relationship using a European horse outgroup. *J. Hered.*, **94**, 374–380.

Wimmers K., Ponksuksil S., Schmoll F. Hardge T., Hatzipanagiotu A., Weber J., Wostmann S., Olek K., Schellander K. (1998) Efficiency of microsatellite markers of the international standard panel for parentage control in German horse populations. *Zuchtungskunde*, **70**, 233–241.

