Animal Genetics, 2000, **31**, 39–48

The genetic structure of Spanish Celtic horse breeds inferred from microsatellite data

J Cañon, M L Checa, C Carleos, J L Vega-Pla, M Vallejo, S Dunner

Summary

Partition of the genetic variability, genetic structure and relationships among seven Spanish Celtic horse breeds were studied using PCR amplification of 13 microsatellites on 481 random individuals. In addition, 60 thoroughbred horses were included. The average observed heterozygosity and the mean number of alleles were higher for the Atlantic horse breeds than for the Balearic Islands breeds. Only eight percentage of the total genetic variability could be attributed to differences among breeds (mean $F_{ST} \approx 0{\cdot}08$; $P < 0{\cdot}01$). Atlantic breeds clearly form a separate cluster from the Balearic Islands breeds and among the former only two form a clear clustering, while the rest of Atlantic breeds (Jaca Navarra, Caballo Gallego and Pottoka) are not consistently differentiated. Multivariate analysis showed that Asturcon populations, Losina and Balearic Islands breeds are clearly separated from each other and from the rest of the breeds. In addition to this, the use of the microsatellites proved to be useful for breed assignment.

Keywords: equine Celtic breeds, microsatellite, genetic diversity, genetic structure

Introduction

Ancient horse populations almost disappeared from the North Iberian Peninsula and Europe during the Mesolithic period (Middle Stone Age) but were reintroduced by the Celts after the VIIIth century BC as a domestic animal. The Celts passed through the Pyrenees and some generations later arrived in North-west Iberia (Spain and Portugal). A trade with Ireland and Brittany was then probably established. Most extant Celtic horse populations are semiferal local breeds distributed across Western Europe (Iberian Peninsula, Italy, France, Brittany, Scotland, Ireland).

Mediterranean breeds (Mallorquina and Menorquina), located in the Balearic Islands, are connected with Celtic horses through a primitive variety of Catalonian horse, now extinct (Martinez *et al*. 1996), and most authors agree with the idea of introgession between an original Celtic population and populations greatly influenced by African genes (Andalusian or Spanish breed and the Thoroughbred) (Aran 1949).

Celtic horses have morphological characteristics considered to be 'primitive', not very different from those in the ancestral population and with small differences between breeds. Celtic horses are below average sized (1·2–1·4 m), except for Balearic Islands horses, which are medium sized (1·5–1·6 m). Although a wide range of coat colours can be found within some populations, most animals are black pigmented: blacks, browns or bays. In some breeds, e.g. the Asturcon, only black animals are accepted for inclusion in the studbook. In some breeds the use of animals for riding is increasing, due to a general gentleness and, in the Asturcon breed due to the pacing, in which both legs of the same side are extended together at the same time.

Celtic horse populations have not been exposed to reproductive technology or other modern breeding tools used in selective breeding, so male and female gene flow between breeds has been very limited, with individual dispersal only at a local level. Establishment of male lineages can also be expected as a consequence of the breeding system used. On the other hand, the creation of organised studbooks is, in many of the breeds, very recent so genetic introgression between breeds has been frequent.

Reproductive isolation, a consequence of local use and management, reduces effective population size, contributing to genetic subdivision that could be detected through drift-based measures based on variation observed at microsatellite loci. In conservation genetics the main objective is to preserve variability within populations under the hypothesis of correlation between genetic variation and the population's viability. In theory, fragmenting populations into subpopulations (local pony breeds) could play an important role in maintaining genetic variation, as it reduces the loss of alleles.

J Cañon
M L Checa
M Vallejo
S Dunner
Laboratorio de Genetica, Facultad de Veterinaria, Universidad Complutense de Madrid, 28040 Madrid, Spain
C Carleos
Departamento de Estadistica e Investigacion Operativa, Facultad de Matematicas, Universidad de Oviedo, 33071 Oviedo, Spain
J L Vega-Pla
Laboratorio de Grupos Sanguineos, Cria Caballar, 14071 Cordoba, Spain

Correspondence: Javier Cañon.

Accepted 26 October 1999

In this paper we investigate the degree of genetic divergence between local populations of Celtic horses, most with extremely small population size, by examination of the spatial distribution of microsatellite variation in ten pony populations from seven local breeds. This analysis includes genetic subdivision and migration rate under the island model, the distribution within and between breeds of the observed genetic variation, phylogenetic analysis of individual animals and breed assignment from microsatellite allele frequencies.

Materials and methods

Sampling of populations

Fresh blood collected in a preserving buffer (APS = Anticoagulant Preservative Solution) (Arctander 1988) was taken from individuals from well defined geographical areas and chosen at random without consideration of the relationship between animals (Table 1). The geographical distribution of these populations is shown in Fig. 1.

DNA extraction and PCR amplification

DNA was extracted following the 'salting out' procedure (Miller *et al.* 1988). Primers and Polymerase Chain Reactions (PCR) conditions are described in Table 2. PCR products were separated by electrophoresis in 8% polyacrylamide gels under denaturating conditions, followed by silver staining (Bassam *et al.* 1991). Allele size was scored against known samples used as standards on every gel.

Statistical analysis

Allele frequencies (available from the authors on request) were obtained by direct counting and unbiased estimates for expected (He) and observed (H_o) heterozygosity, and the number of alleles were computed using BIOSYS-1 (Swofford & Selander 1989). Hardy–Weinberg equilibrium (HWE) was tested: (a) for each locus/population combination by an exact test using Guo & Thompson's (1992) Markov chain Monte Carlo algorithm implemented in the GENEPOP package version 3·1 (Raymond & Rousset 1995a) and (b) across loci and populations using Fisher's method

$$\chi^2 = -2\left(\sum_{i=1}^{r} \ln P_i\right)$$

where r is the total number of loci across breeds or the total number of breeds across loci, and P_i is the likelihood ratio for the i-th locus or breed (Raymond & Rousset 1995b).

Two measures of similarity between individual animals were used: (a) proportion of alleles shared over loci (Bowcock *et al.* 1994), and (b) Dice's (1945) coefficient after scoring '1' for each band (allele) present and '0' for each band not present. Individuals were clustered according to the genetic distances previously computed using the upgma algorithm performed by the SAHN program in the PC version of NTSYS (Rohlf 1988).

The classical estimators for differentiation between populations, F_{ST} (Wright 1965), θ (Weir & Cockerham 1984) and G_{ST} (Nei 1973), were considered most appropriate for this analysis because genetic drift is assumed to be the main factor in genetic differentiation among

Table 1. Populations of Spanish Celtic horse breeds surveyed

Breed	Geographical origin	Number of animals sampled	Population size	Existence of Studbook (year)
Asturcon		119	750	Yes (1981)
	ICONA	61		
	Borines	40		
	LaVita	18		
Caballo Gallego		72	10 000–100 000	No
	Pontevedra	29		
	La Coruña	43		
Losina	Burgos	65	200	No
Pottoka	País Vasco	51	< 1000	Yes (1995)
Jaca Navarra	Navarra	122	250	No
Mallorquina	Palma de Mallorca (Balearic Islands)	20	200	Yes (1993)
Menorquina	Menorca (Balearic Islands)	31	1000	Yes (1993)
Thoroughbred*		60		

*Breed used as reference.

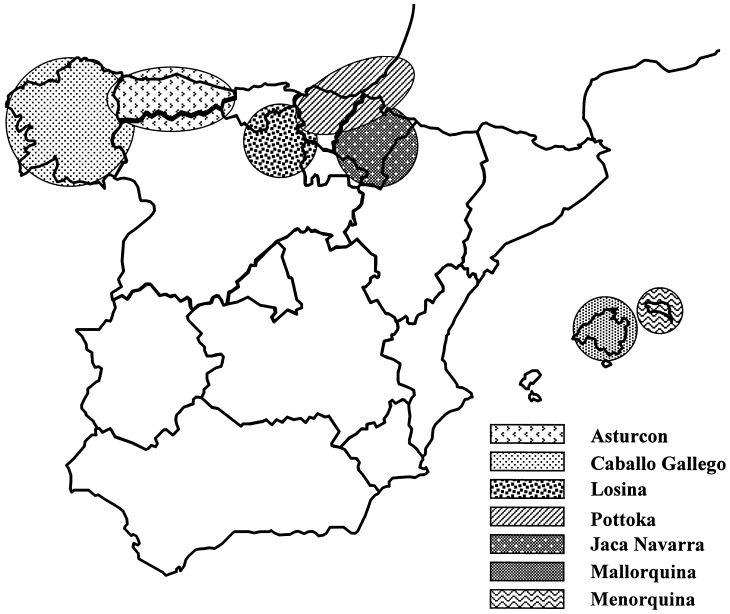

Fig. 1. Geographical location of seven Spanish Celtic horse breeds.

closely related populations or for short-term evolution (Reynolds *et al.* 1983; Weir 1990; Takezaki & Nei 1996). Wright's (1965) F_{IS} and F_{IT} indices were also estimated for each population using FSTAT program (Goudet 1995) and their statistical significance tested using permutation tests. Cluster analysis using the upgma algorithm (Sneath & Sokal 1973) implemented in the DISPAN program (Ota 1993) was applied to the Nei *et al.* (1983) genetic distances.

Table 2. References, primer sequences, observed heterozygosity (direct count), expected heterozygosity and experimental parameters for 13 microsatellite markers

Locus	Ref.	Primer sequences (5′-3′)	T_m°C	$MgCl_2$ (mM)	Cycles	Size range (bp)	Ho	He
HTG4	Ellegren *et al.* (1992)	CTATCTCAGTCTTGATTGCAGGAC						
		CTCCCTCCCTCCCTCTGTTCTC	55	2·5	30	129–141	0·69	0·71
HTG6	Ellegren *et al.* (1992)	CCTGCTTGGAGGCTGTGATAAGAT						
		GTTCACTGAATGTCAAATTCTGCT	55	2·5	30	84–106	0·57	0·64
HTG7	Marklund *et al.* (1994)	CCTGAAGCAGAACATCCCTCCTTG						
		ATAAAGTGTCTGGGCAGAGCTGCT	55	2·5	30	118–128	0·66	0·70
HTG8	Marklund *et al.* (1994)	CAGGCCGTAGATGACTACCAATGA						
		TTTTCAGAGTTAATTGGTATCACA	55	2·5	30	176–192	0·68	0·75
HTG10	Marklund *et al.* (1994)	CAATTCCCGCCCCACCCCCGGCA						
		TTTTTATTCTGATCTGTCACATTT	55	2·5	30	94–114	0·76	0·84
HMS2	Guérin *et al.* (1994)	ACGGTGGCAACTGCCAAGGAAG						
		CTTGCAGTCGAATGTGTATTAAATG	60	2·5	35	216–238	0·73	0·83
HMS3	Guérin *et al.* (1994)	CCAACTCTTTGTCACATAACAAGA						
		CCATCCTCACTTTTTCACTTTGTT	60	2·5	30	150–172	0·58	0·76
HMS6	Guérin *et al.* (1994)	GAAGCTGCCAGTATTCAACCATTG						
		CTCCATCTTGTGAAGTGTAACTCA	60	2·5	35	159–173	0·66	0·78
HMS7	Guérin *et al.* (1994)	CAGGAAACTCATGTTGATACCATC						
		TGTTGTTGAAACATACCTTGACTGT	60	2·5	30	170–186	0·80	0·81
VHL20	Van Haeringen *et al.* (1994)	CAAGTCCTCTTACTTGAAGACTAG						
		AACTCAGGGAGAATCTTCCTCAG	60	2·5	30	87–105	0·82	0·85
ASB2	Breen *et al.* (1997)	CCTTCCTGTAGTTTAAGCTTCTG						
		CACAACTGAGTTCTCTGATAGG	60	2·5	30	154–188	0·81	0·85
AHT4	Binns *et al.* (1995)	AACCGCCTGAGCAAGGAAGT						
		CCCAGAGAGTTTACCCT	60	2	30	148–164	0·82	0·83
AHT5	Binns *et al.* (1995)	ACGGACACATCCCTGCCTGC						
		GCAGGCTAAGGGGGCTCAGC	60	1·5	30	128–142	0·76	0·80

Allele frequencies of breeds at all loci were used as variables to cluster the breeds spatially using correspondence analysis (Lebart *et al.* 1984) which uses Chi-square distances to judge proximity among them. The data from individual genotypes were prepared by scoring a '0' if a particular band was not present, a '1' if it was present in one copy and '2' if it was homozygous. The first three major factors were plotted on a three-dimensional diagram for the 10 populations.

The assignment of an individual to a breed was studied. The maximum likelihood discriminant rule was used and consists in classifying an anonymous animal i in the breed

$$r \in \{r_1, \mathrm{K}\ , r_n\}$$

for which the conditional probability

$$\Pr[i|r]$$

is maximum. Additional information such as *a priori* probabilities or loss functions may be accounted for by means of Bayes' discriminant rule (Buchanan *et al.* 1994). Let

$$\hat{P}_{r,l,a}$$

be the frequency of the allele a of the locus l in the breed r. Then

$$\Pr[i|\ r] = \prod_l h(i,l)\hat{P}_{r,l,a_{i,l,1}}\hat{P}_{r,l,a_{i,l,2}}$$

where $a_{i,l,1}$ and $a_{i,l,2}$ are the alleles of individual i at locus l.

Results

Levels of variation and HWE

A total of 120 alleles were detected across the 13 loci analysed. The mean number of alleles (MNA) observed in different populations and the total mean number of alleles are shown in Table 3. In order to avoid the positive correlation effect (Pearson Correlation = 0·92, $P < 0{\cdot}01$) existing between number of alleles and sample size, a random sample of 18 animals (the smallest sample size gathered, corresponding to La Vita population) was drawn from each population. This process was repeated 1000 times to provide bootstrap confidence intervals for the MNA when the same sample size is considered for all the breeds (Table 3).

Observed and expected heterozygosities per breed ranged from 0·694 and 0·677–0·752 and 0·77, respectively (Table 3). Out of the total of the 143 HWE tests only 5 gave significant deviations at the 1% level. When results were pooled across breeds, three microsatellites (*HTG10*, *HMS3*, *AHT4*) gave a significant deviation ($P < 0{\cdot}01$)

Table 3. Summary statistics for horse breeds used in microsatellite marker analysis of population structure showing geographical location, sample size, observed (Ho) and expected (He) heterozygosity, average number of alleles per locus, and heterozygote deficiency (F_{IS})

Breed	Origin of the samples	*N*	Ho	He	MNA (full sample)†	MNA (uniform sample)‡	F_{IS}
Asturcon	Icona*	61	0·735 ±0·041	0·684 ±0·032	5·4 (0·4)	5·1 (0·34)	− 0·075
	Borines	40	0·729 ±0·038	0·699 ±0·030	6·0 (0·4)	5·3 (0·35)	− 0·045
	LaVita	18	0·765 ±0·037	0·742 ±0·019	5·8 (0·4)	5·8 (0·00)	− 0·031
	Mean of the three subpopulations	119	0·738 ±0·031	0·729 ±0·024	7·1 (0·4)	5·7 (0·37)	− 0·012
Caballo Gallego	La Coruña	43	0·700 ±0·026	0·750 ±0·023	7·2 (0·5)	6·4 (0·43)	0·066
	Pontevedra	29	0·724 ±0·016	0·760 ±0·017	7·1 (0·4)	6·7 (0·34)	0·048
	Mean of the two subpopulations	72	0·710 ±0·020	0·760 ±0·021	7·8 (0·4)	6·6 (0·43)	0·066
Losina	Burgos	66	0·715 ±0·037	0·702 ±0·032	6·8 (0·4)	5·7 (0·42)	− 0·019
Pottoka	Pais Vasco	51	0·741 ±0·021	0·770 ±0·020	7·0 (0·5)	6·7 (0·34)	0·039
Jaca Navarra	Navarra	122	0·703 ±0·046	0·728 ±0·031	7·8 (0·5)	6·0 (0·49)	0·034
Menorquina	Menorca	31	0·752 ±0·036	0·726 ±0·030	6·2 (0·5)	5·8 (0·34)	− 0·036
Mallorquina	Mallorca	20	0·700 ±0·028	0·748 ±0·023	6·1 (0·5)	6·0 (0·18)	0·066
Thoroughbred	España	60	0·694 ±0·039	0·677 ±0·030	5·2 (0·3)	4·6 (0·26)	− 0·025
Total	8 breeds	541	0·718 ±0·025	0·780 ±0·018	9·2 (0·5)		

*Group of animals managed by the 'Consejería de Agricultura' of Asturias.
†Includes all the animals in each population sampled.
‡Random samples of 18 individuals were drawn repeatedly for each population except for LaVita in which only 18 animals were available. Values in brackets represent 95% bootstrap confidence intervals.

Table 4. F statistic estimates and their significances by locus

	F_{IS}	F_{ST}	G_{ST}	θ	F_{IT}
HTG4	– 0·015	0·060†	0·062	0·061	0·046*
HTG6	0·025	0·108†	0·099	0·104	0·130†
HTG7	– 0·007	0·082†	0·094	0·081	0·076†
HTG8	0·024	0·079†	0·077	0·081	0·101†
HTG10	0·033	0·081†	0·065	0·076	0·111†
HMS2	0·041	0·080†	0·099	0·087	0·118†
HMS3	0·158 †	0·112†	0·109	0·114	0·253†
HMS6	0·019	0·153†	0·163	0·159	0·169†
HMS7	– 0·036	0·056†	0·057	0·055	0·022
VHL20	– 0·015	0·058†	0·070	0·060	0·043†
ASB2	0·005	0·042†	0·053	0·043	0·046†
AHT4	– 0·033	0·041†	0·041	0·041	0·009
AHT5	– 0·004	0·061†	0·070	0·061	0·057†
Mean	0·014 *	0·078†	0·081†	0·078†	0·090†

*$P < 0·05$.
†$P < 0·01$.

and when pooled across loci, three populations (Losina breed and two populations of the Caballo Gallego breed) gave significant ($P < 0·01$) deviations. The main factor that may have caused such a deviation in the Losina breed is probably its very small effective population size (< 10, data not shown), while in the Caballo Gallego breed, deviation from HWE is probably related to the sampling procedure as samples were collected in two different years and animals may have been representatives of the four geographic clusters recognised by some authors (Sanchez *et al*. 1996).

F statistics

The G_{ST}, θ and F_{ST} values for each locus are very close and are shown with the F_{IT} and F_{IS} values in Table 4. Levels of apparent breed differentiation were considerable and multilocus F_{ST} values indicate that around 8% of the total genetic variation was explained by breeds differences, the remaining 92% corresponding to differences among individuals. Genetic differentiation among breeds was highly significant ($P < 0·01$) for all loci. On average, breeds had a 1·4% ($P < 0·05$) deficit of heterozygotes, whereas the total population had a 9% ($P < 0·01$) deficit of heterozygotes. Table 5 presents F_{ST} values when breeds are considered in couples. Genic differentiation values among breeds range from 2·6% for the Pottoka-Gallego pair to 15·0% for the Asturcon-Thoroughbred pair. All values were different from 0 ($P < 0·01$). Values above the diagonal in Table 5 represent the estimated number of individuals exchanged between populations per generation (Nm, where N is the total effective number of animals and m the migration rate) which balances the diversifying effect of the genetic drift.

Figure 2 shows a UPGMA tree constructed from the pairwise distances between 263 individuals, 25 animals taken at random for each population described in the Table 1 except for La Vita and Mallorquina populations from which all individuals were used. Similarity matrices computed by Dice's (1945) and Bowcock *et al*. (1994) methods were highly

Table 5. F_{ST} estimates (below diagonal) as a measure of genetic distance between horse breeds and the number of effective migrants per generation (Nm) (above diagonal) in balance with genetic drift (Wright 1969) ($F_{ST} = 1(4Nm + 1)$

	Asturcon	Gallego	Losina	Pottoka	Jaca Navarra	Mallorquina	Menorquina	Thoroughbred
Asturcon		5·49	3·51	4·20	3·66	2·25	2·17	1·42
Gallego	0·044		5·36	9·52	7·05	3·62	5·05	1·81
Losina	0·067	0·045		3·94	4·17	1·83	2·17	1·35
Pottoka	0·056	0·026	0·060		5·93	5·07	6·26	2·23
Jaca Navarra	0·064	0·034	0·057	0·041		2·18	2·93	1·49
Mallorquina	0·100	0·065	0·120	0·047	0·103		3·60	1·88
Menorquina	0·103	0·047	0·103	0·038	0·079	0·065		2·02
Thoroughbred	0·150	0·121	0·157	0·100	0·144	0·118	0·110	

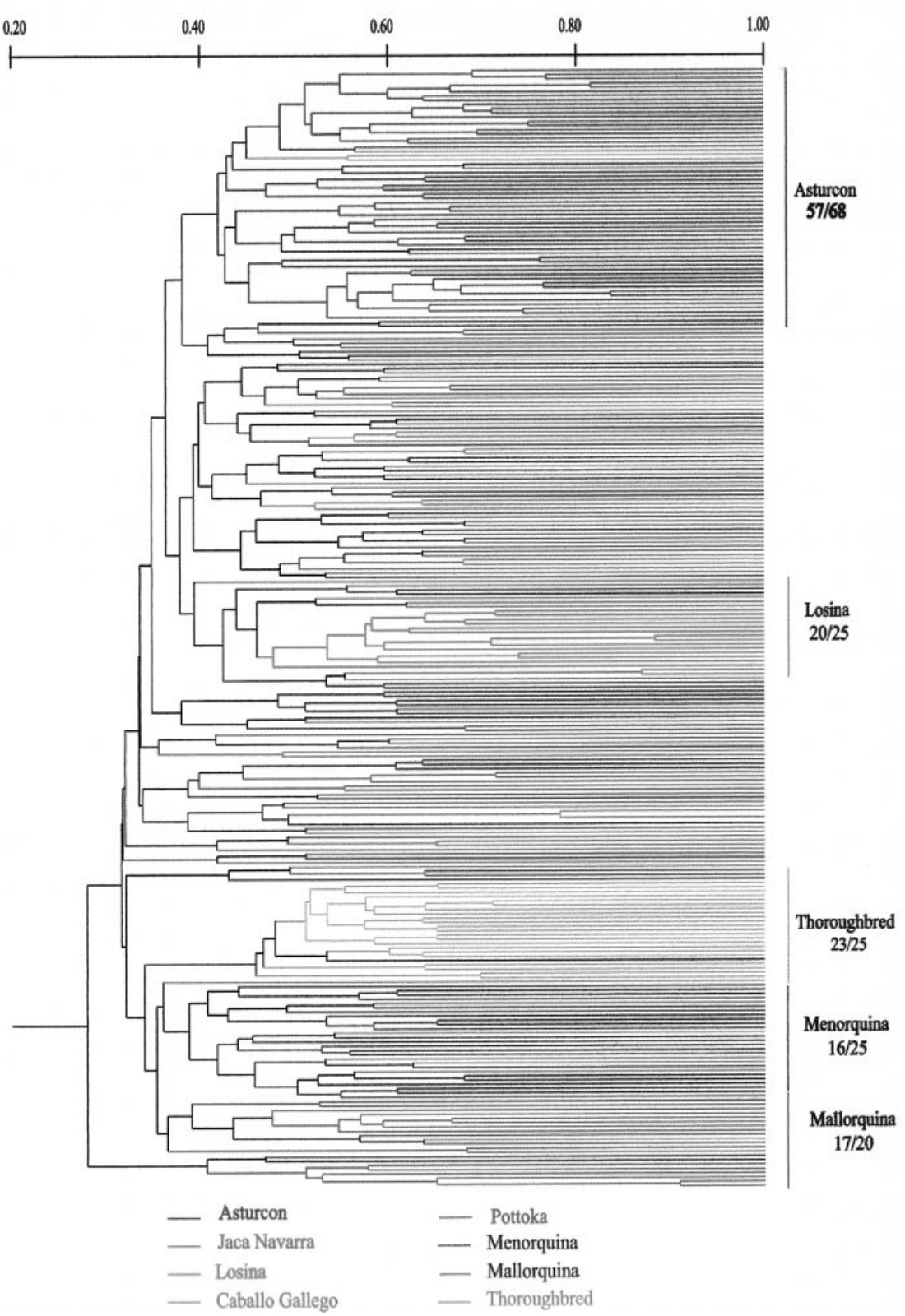

Fig. 2. UPGMA dendrogram constructed from the pairwise distances inferred from microsatellite data between 263 individuals from eight horse breeds. Numbers to the right indicate the fraction of individuals from the breed found in a cluster.

correlated ($r = 0{\cdot}98$, Mantel test, $P < 0{\cdot}01$), so only the tree constructed using the former is shown. Goodness of fit (see Rohlf & Sokal 1981) of the clustering to data set was acceptable ($r = 0{\cdot}82$) although only two Atlantic breeds and the two Balearic Islands breeds showed a clear level of clustering. Only Asturcon and Losina formed discrete clusters,

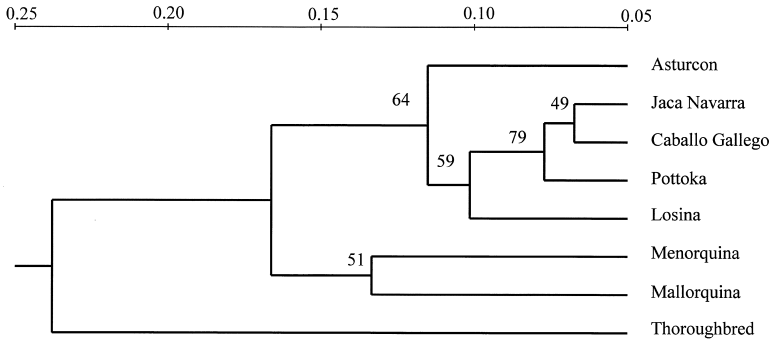

Fig. 3. UPGMA dendrograms showing the genetic relationships among the sampled horse breeds, inferred from microsatellite data. Tree is based on Nei *et al.* (1983) genetic distances. The numbers at the nodes are values for 1000 bootstrap resamplings of the 13 loci genotyped.

while the distribution of animals from the other Atlantic breeds revealed a very low degree of breed structure, with the Pottoka and Galician breeds showing a more highly fragmented pattern of clustering.

A UPGMA tree based on Nei *et al.* (1983) genetic distances relating the eight populations studied is presented in Fig. 3. The numbers at the nodes are bootstrapping values for 1000 bootstrap resamplings of the 13 loci genotyped.

Multivariate correspondence analysis

The first three principal factors of the correspondence analysis are plotted in Fig. 4. Examination of this figure reveals a clear separation of the Asturcon populations from the rest of the Atlantic breeds. Balearic Islands (Mediterranean) breeds cluster towards the top right quadrant of the plot. Most of the variation in Asturcon populations is explained by first factor, while factor 3 is the most important to discriminate Losina breed (Fig. 4).

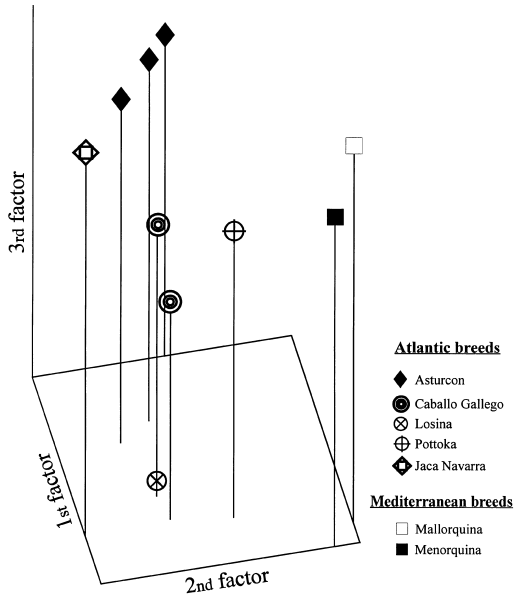

Fig. 4. Correspondence analysis of allele frequencies from 13 microsatellite loci typed in 10 populations from seven Spanish Celtic horse breeds.

Breed assignment

Results for the assignment of animals to populations using 13 microsatellites are presented in Table 6.

Discussion

The significant between-population F_{ST} estimates indicate a relatively low gene flow between the breeds studied, probably due to reproductive isolation. The F_{ST} value is mostly a consequence of the distance between Atlantic and Mediterranean Celtic breeds. The mean genetic distance (F_{ST}) among Atlantic breeds is 0·049 and between Atlantic and Mediterranean breeds 0·081, values commonly observed between conspecific populations. In the context of the conservation and maintenance of genetic variability, migration values (Nm) can be interpreted as the upper limit of the number of migrants per generation, which would allow the maintenance of the observed genetic differentiation between the breeds. For example, an introgression rate of nine individuals per generation between Pottoka and Galician breeds would maintain the estimated degree of genetic differentiation between these breeds. Similarly, when we compare the Balearic Islands breeds (Mallorquina and Menorquina) with the Thoroughbred, a gene flow between them and the Thoroughbred greater than only two individuals per generation could constitute a real threat for both Mediterranean breeds. It must be emphasised that such a strategy would accept a greater introgression rate between genetically closer populations than between more divergent ones.

Table 6. Breed assignment using 13 microsatellites and the maximum likelihood classification rule for the Atlantic breeds

	Asturcon	Gallego	Losina	Pottoka	Jaca Navarra	% of errors
Asturcon	**96·9**	0·9	1·0	0·4	0·8	3·1
Gallego	4·5	**77·4**	5·5	4·4	8·2	22·6
Losina	0·7	1·3	**96·3**	0·5	1·1	3·7
Pottoka	2·1	3·4	2·7	**88·4**	3·4	11·6
Jaca Navarra	1·2	2·5	1·7	1·3	**93·3**	6·7

Eight percentage of the total genetic variation is due to breed differentiation, a value close to that found in other domestic species, e.g. 10% in European cattle breeds (MacHugh *et al.* 1998), 9·9% in dogs (Jordana *et al.* 1992), though slightly lower than that found in goats 17% (Saitbekova *et al.* 1999) and humans 10–20% (Cavalli-Sforza *et al.* 1994).

The significant deficit of heterozygotes observed (F_{IS} = 0·014, $P < 0·05$) may not be an inbreeding effect, since the deficit is attributable to a single marker, *HMS3*. The F_{IS} negative values observed in some of the populations may be explained by the Wahlund effect.

Topology of the dendrograms in Figures 2 and 3 show a similar pattern: (a) Celtic horses are first divided in two clusters: Atlantic (Asturcon, Losina, Jaca Navarra, Caballo Gallego, and Pottoka breeds) and Mediterranean (Mallorquina and Menorquina breeds) (b) Mediterranean breeds are closer to the Thoroughbred than Atlantic horses (c) within the Atlantic breeds, only Losina and Asturcon breeds show a clear degree of clustering while Jaca Navarra, Gallego and Pottoka breeds split into a number of clusters across the dendrogram.

From the examination of the dendrogram constructed from band-sharing distances (Fig. 2) and of the spatial distribution of allele frequency among the ten populations studied (Fig. 4) an unclear population structure is apparent. This lack of population structure of the semiferal Celtic horses is probable due to the absence of closed breeds within which strong selection is being carried out. A similar origin for all breeds, a hypothetical Cantabric–Pyrenean trunk and very similar morphological characteristics allowed reciprocal introgression among populations, but not enough to neutralise the genetic drift consequence of the isolation by the physical geography where the breeds are located. Although the Spanish Celtic breeds have been long ago officially recognised as breeds, the fact that studbooks have been created recently has prevented these breeds existing as discrete populations (Table 1), allowing easier genetic exchange. As a first consequence, the topology of the tree (Fig. 2) is not robust, showing relatively low values of bootstrapping for some branches (values not shown). Mediterranean breeds clearly form a distinct cluster, probably because of geographic distances from Atlantic populations, together with the influence of Arabian horse blood. Among the Atlantic breeds, the topology of the Asturcon clade is the most robust (84% of the animals cluster within the breed); the major effort to preserve this breed during the last decades (official studbook was created in 1981) and the major bottleneck suffered by this breed at the beginning of this century have probably contributed to this pattern. The Losina breed, which also shows a clear grouping (80% of animals group together), is a special population since today most animals are effectively descendants from a single herd.

Breeds are mainly artefacts classically based on morphological differences and tightly related to geographical locations, in such a way that different names can be assigned to very closely related populations located in different administrative areas. Reproductive isolation by geographic barriers or by socio-political considerations leads to a within population genetic drift process that will cause the genetic differentiation between populations detected by the use of neutral molecular markers. Morphological differences between populations, which are frequently negligible and a consequence of artificial selection, are not taken into account by neutral molecular markers.

To establish a conservation program of genetic resources, molecular markers could also be important not only to test whether an animal or a small set of animals belongs to an endangered breed in order to add or not it to the gene pool, but also to estimate relatedness between individuals when pedigrees are unknown (Blouin *et al.* 1996), as, for example, occurs in the Caballo Gallego or Losina breeds. Demographic history information is also of great interest for conservation purposes (Milligan *et al.* 1994; Dunner *et al.* 1998).

There has recently been growing interest in the use of a set of alleles in an anonymous

sample to identify the source population (Shriver *et al.* 1997; MacHugh *et al.* 1998). Most procedures use the Kullback & Leibler (1951) divergence concept, which is a measure of the difference between two distributions, or more precisely, on their asymptotic distributions. The results presented in Table 6 demonstrate the possibilities of using highly polymorphic microsatellites for assigning breed identities to anonymous equine samples as had been previously shown for cattle (MacHugh *et al.* 1998), sheep (Buchanan *et al.* 1994) and humans (Shriver *et al.* 1997).

This study contributes to the knowledge of the genetic structure and molecular characterisation of small populations, many of them in potential threat of extinction. It also shows how microsatellites can be used to establish the genetic relationships between populations providing reasonable statistical power for breed assignment, regardless of whether they are closely related or not, allowing their future management to be based on greater knowledge of genetic structuring and relationships between populations.

Acknowledgements

We thank R. De Juana (President of the Asociacion Española de Criadores de Caballos de raza Losina) who provided Losina breed samples, the ITG Ganadero (sección equina) who provided samples of the Jaca Navarra and J. L. Benedito who kindly provided samples of the Caballo Gallego. We gratefully acknowledge D. García and J.P. Gutiérrez for their statistical and computational advice and M. A. García-Atance for her genotyping contribution. This work received the financial support of the Comisión Interministerial de Ciencia y Tecnología (CICYT): (Grant no. AGF95–064), ACPRA (Asociación de Criadores de Ponis de Raza Asturcon) and Caja Asturias. The authors thank the referees for their comments.

References

Aran S. (1949) *Caballos, Mulos, Asnos*. Gráficas Yagües, Madrid, Spain.

Arctander P. (1988) Comparative studies of avian DNA by restriction fragment length polymorphism analysis: convenient procedures on blood samples from live birds. *Journal of Ornitology* **129,** 205–16.

Bassam B.J., Caetano-Anollés G. & Gresshoff P.M. (1991) Fast and sensitive silver staining of DNA in polyacrilamide gels. *Analytical Biochemistry* **80,** 81–4.

Binns M.M., Holmes N.G., Holliman A. & Scott, A.M. (1995) The identification of polymorphic microsatellite loci in the horse and their use in thoroughbred parentage testing. *British Veterinary Journal* **151,** 9–15.

Blouin M.S., Parsons M., Lacaille V. & Lotz S. (1996) Use of microsatellite loci to classify individuals by relatedness. *Molecular Ecology* **5,** 393–401.

Bowcock A.M., Ruiz-Linares A., Tomfohrode J., Minch E., Kidd J.R. & Cavalli-Sforza L.L. (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368,** 455–7.

Breen H., Downs P., Irwin Z. & Bell K. (1994). Intrageneric amplification of horse microsatellite markers with emphasis on the Przewalski's horse (*E. przewalskii*). *Animal Genetics* **25**, 401–405.

Buchanan F.C., Adams L.J., Littlejohn R.P., Maddox J.F. & Crawford A.M. (1994) Determination of evolutionary relationships among sheep breeds using microsatellites. *Genomics* **22,** 397–403.

Cavalli-Sforza L.L., Menozzi P. & Piazza A. (1994) *The History and Geography of Human Genes*. Princeton University Press., Princeton, NJ. USA.

Dice J.R. (1945). Measures of the amount of ecologic association between species. *Ecology* **26,** 297–302.

Dunner S., Checa M.L., Gutierrez J.P., Martin J.P. & Cañon J. (1998) Genetic analysis and management in small populations, the Asturcon pony as an example. *Genetics, Selection and Evolution* **30,** 397–405.

Ellegren H., Johansson M., Sandberg K. & Andersson L. (1992) Cloning of highly polymorphic microsatellites in the horse. *Animal Genetics* **23,** 133–42.

Goudet J. (1995) FSTAT, Version 1.2, a computer program to calculate F-statistics. *Journal of Heredity* **86,** 485–6.

Guérin G., Bertaud M. & Amigues Y. (1994) Characterization of seven new horse microsatellites: HMS15 and HMS20. *Animal Genetics* **25,** 62.

Guo S.W. & Thompson E.A. (1992) Performing the exact test of Hardy–Weinberg proportions for multiple alleles. *Biometrics* **48,** 361–72.

Jordana J., Piedrafita J., Sanchez A. & Puig P. (1992) Comparative F statistics analysis of the genetic structure of ten Spanish dog breeds. *Journal of Heredity* **83,** 367–74.

Kullback S. & Leibler A. (1951) On information and sufficiency. *Annals of Mathematica Statistics* **22,** 79–86.

Lebart L., Morineau A. & Warwick K.M. (1984). *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices*. John Wiley and Sons, New York.

MacHugh D.E., Loftus R.T., Cunningham P. & Bradley D.G. (1998) Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers. *Animal Genetics* **29,** 333–40.

Marklund S., Ellegren H., Eriksson S., Sandberg K. & Andersson L. (1994) Parentage testing and linkage analysis in the horse using a set of highly polymorphic microsatellites. *Animal Genetics* **25,** 19–23.

Martinez J.M., Valera M. & Molina A. (1996) El caballo Losino. *Animal Genetic Resources Information* **19,** 17–27.

Miller S.A., Dykes D.O. & Poleskcy H.T. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research* **16,** 1215.

Milligan B.G., Leebens-Mack J. & Strand A.E. (1994) Conservation genetics: beyond the maintenance of marker diversity. *Molecular Ecology* **3,** 423–35.

Nei M. (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the USA* **70,** 3321–3.

Nei M., Tajima F. & Tateno Y. (1983) Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular and Evolution* **19,** 153–70.

Ota T. (1993) *DISPAN*: *Genetic Distance and Phylogenetic Analysis*. Pennsylvania State University, University Park, USA.

Raymond M. & Rousset F. (1995a) GENEPOP, Version 1.2, population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86,** 248–9.

Raymond M. & Rousset F. (1995b) An exact test for population differentiation. *Evolution* **49,** 1280–3.

Reynolds J., Weir B.S. & Cockerham C.C. (1983) Estimation of the coancestry coefficient: Basis for a short-term genetic distance. *Genetics* **105,** 767–79.

Rohlf F.J. (1988) *NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System*. Exeter Software. Heritage Lane. Setauket, New York, USA.

Rohlf F.J. & Sokal R.R. (1981). Comparing numerical taxonomic studies. *Systematic Zoology* **30,** 459–90.

Saitbekova N., Gaillard C., Obexer-Ruff G. & Dolf G. (1999) Genetic diversity in Swis goat breeds based on microsatellite analysis. *Animal Genetics* **30,** 36–41.

Sanchez L., Iglesias A., Fernandez A. & Viana J.L. (1996) Caballo gallego de monte (poney gallego) *Animal Genetic Resources Information* **19,** 51–64.

Shriver M.D., Smith M.W., Jin L., Marcini A., Akey J.M., Deka R. & Ferrell R.E. (1997) Ethnic-Affiliation Estimation by use of population-specific DNA markers. *American Journal of Human Genetics* **60,** 957–64.

Sneath P.H.A. & Sokal R.R. (1973) *Numerical Taxonomy*. W.H. Freeman., San Francisco, CA.

Swofford D.L. & Selander R.B. (1989) *BIOSYS-1. A Computer Program for the Analysis of Allelic Variaton in Population Genetics and Biochemical Systematics (Release 1.7)*. University of Illinois, Urbana, Champaign, USA.

Takezaki N. & Nei M. (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* **144,** 389–99.

Van Haeringen H., Bowling A.T., Stott M.L., Lenstra J.A. & Zwaagstra K.A. (1994) A highly polymorphic horse microsatellite locus: VHL20. *Animal Genetics* **25,** 207.

Weir B.S. (1990) *Genetic Data Analysis*. Sinauer Associates, Sunderland, Massachusetts, USA.

Weir B.S. & Cockerham C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38,** 1358–70.

Wright S. (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* **19,** 395–420.

Wright S. (1969) *The Theory of Gene Frequencies: Evolution and the Genetics of Populations, Vol. 2*. Chicago University Press, Chicago, USA.