npg

Hierarchical analysis of genetic structure in Spanish donkey breeds using microsatellite markers

J Aranguren-Méndez[1,2], M Gómez[3] and J Jordana[1]

[1]Unitat de Genètica i Millora Animal, Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, Barcelona, Spain; [2]Universidad del Zulia, Facultad de Ciencias Veterinarias, Departamento de Producción Animal, Maracaibo, Venezuela; [3]Servicio de Ganadería, Diputación Foral de Bizkaia, Avda. Lehendakari Aguirre, Bilbao, Spain

The hierarchical population structure of five, native-Spanish donkey breeds (Andaluza, Catalana, Mallorquina, Encartaciones and Zamorano-Leonesa) has been studied using F-statistics. In addition, nine Moroccan asses and 24 Merens breed horses were included in the analysis. Data came from 15 DNA microsatellites. The analysis shows that Spanish donkeys are substructured at both hierarchical levels studied, among breeds and within breeds (between subpopulations). In the whole population, the deficit of heterozygotes was estimated to be about 21%. The fixation indices corresponding to differences between breeds, sub-populations within breeds, and within subpopulations were estimated to be 6.4%, 3.5% and 3.0%, respectively. The dendrogram obtained shows that the Andaluza and the Moroccan ass form a separate cluster from the northern Spanish breeds (Catalana, Encartaciones, Mallorquina and Zamorana-Leonesa). These groupings coincide with those obtained from historical and archaeological data.
Heredity (2002) **89,** 207–211. doi:10.1038/sj.hdy.6800117

Introduction

The ass (*Equus asinus*), is a herbivorous animal of the order *Perissodactyla*, family *Equidae*. It was domesticated about 6000 years ago, probably in either Egypt or Mesopotamia (Littauer and Crouwel, 1979). In Spain, the development of donkey populations has been influenced by their extensive use for riding and as a beast of burden; it is capable of carrying over 100 kg.

The Spanish donkey breeds Andaluza, Catalana, Mallorquina, Encartaciones and Zamorano-Leonesa have suffered a substantial decrease in population size (Jordana and Folch, 1998) which might create high levels of inbreeding which may result in inbreeding depression, increasing the risk of breed extinction. Currently, the census population sizes of these five breeds are very low, and they are included in the Food and Agricultural Organization of the United Nation's (FAO) list of domestic animals to be preserved (FAO, DAD-IS http://fao.org/dad-is).

The objective of the present study is to characterize the genetic structure of five Spanish donkey populations in danger of extinction by using F-statistic analysis (Wright, 1965; Nei, 1977; Weir and Cockerham, 1984). F-statistics have proven to be a very useful tool in elucidating the pattern and extent of genetic variation residing within and among natural populations of animal and plant species.

Genetic characterisation is the first step in breed conservation programmes and may have implications for future breeding strategies (Bjørnstad *et al*, 2000). Little genetic data from domestic donkeys exist (Breen *et al*, 1994; Bellone *et al*, 1998; Jordana *et al*, 1999, 2001; Aranguren-Méndez *et al*, 2001).

Materials and methods

Breeds studied and DNA collection

Genomic DNA was prepared from whole blood according to standard methods involving lysates of the washed white cells and phenol-cloroform-isoamylalcohol (25:24:1) extraction. The sample size and breeds involved were: 87 Andaluza (AND), 140 Catalana (CAT), 104 Mallorquina (MALL), 74 Encartaciones (ENC) and 108 Zamorano-Leonesa (ZAM). The areas of main distribution of these indigenous Spanish breeds are shown in Figure 1. In addition, nine Moroccan asses were used as members of the *E. asinus africanus,* and 24 horses of the Merens breed (*E. caballus*) were used as an outgroup. The breeds were divided into 16 subpopulations, according to geographical criteria and/or areas of influence of certain breeders (Figure 1).

Microsatellite markers

The 15 equine microsatellite markers chosen for analysis were: AHT4, AHT5, ASB2, HMS1, HMS2, HMS3, HMS5, HMS6, HMS7, HTG4, HTG6, HTG7, HTG10, HTG15 and VHL20. Primer sequences, reaction conditions and the

Correspondence: J Jordana, Unitat de Genètica i Millora Animal, Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, Barcelona, Spain
E-mail: Jordi.Jordana@uab.es

1 - AND 1 (n=33)   9 - ENC 3 (n=29)

2 - AND 2 (n=28)   10 - MALL 1 (n=24)

3 - AND 3 (n=26)   11 - MALL 2 (n=42)

4 - CAT 1 (n=55)   12 - MALL 3 (n=17)

5 - CAT 2 (n=71)   13 - MALL 4 (n=21)

6 - CAT 3 (n=14)   14 - ZAM 1 (n=58)

7 - ENC 1 (n=22)   15 - ZAM 2 (n=22)

8 - ENC 2 (n=23)   16 - ZAM 3 (n=28)

Figure 1 Geographical location of the Spanish donkey subpopulations and their reference codes (*n* is the sample size).

data obtained (available from the authors on request) have been described previously by Aranguren-Méndez *et al* (2001).

Statistical analyses

Population structure was analysed by means of *F*-statistics using Weir and Cockerham's methods (1984), implemented in the FSTAT computer programme (Goudet, 2000). A hierarchical F_{ST} analysis was performed to determine the amount of variance attributable to subpopulation substructure (Wright, 1978; Johannesen and Loeschcke, 1996). The hierarchical analysis was carried out using analysis of molecular variance (AMOVA) in the ARLEQUIN 2000 package (Schneider *et al*, 2000). AMOVA yields estimations of population structure at different levels of the specified hierarchy.

In order to test the significance of the different statistics for the null hypothesis of no differentiation at the corresponding level, permutation and resampling tests (jackknifing and bootstrapping) were carried out as implemented in the previous programmes.

The Reynolds genetic distance, D_R, (Reynolds *et al*, 1983), a measure based on F_{ST} values [$D_R = -\ln(1-F_{ST})$], with neighbour-joining (NJ) clustering (Saitou and Nei, 1987), was used to construct a dendrogram of breed relationships, using the PHYLIP package (Felsenstein, 1995). An unrooted consensus tree, evaluated by 1000 bootstrap replicates, was obtained. The method of Slatkin (1993), implemented in GENEPOP, was used to assess the genetically effective migration rate ($M = N_e m$, the average number of migrants exchanged per generation).

Results

Thirteen of 15 equine microsatellites investigated amplified well and were polymorphic in the donkey, except for locus ASB2, which failed to amplify, and HMS1, which was monomorphic (165 bp) in all breeds. F_{ST}, F_{IT} and F_{IS} values, computed over all breeds and loci, were obtained. Levels of apparent breed differentiation were considerable and multilocus F_{ST} values indicate that around 4.1% of the total genetic variation could be explained by breed differences, the remaining 95.9% corresponding to differences among individuals. Genetic differentiation among breeds was highly significant ($P < 0.001$) for all loci. On average, individuals within breeds had a 17.8% ($P < 0.001$) deficit of heterozygotes, whereas the total population had a 21.1% ($P < 0.001$) deficit of heterozygotes.

Values for the *F*-statistics of the Spanish donkey populations at all hierarchical levels are presented in Table 1. The degree of genetic differentiation among subpopulations was highly significant for all breeds, varying from 1.3% for the Encartaciones breed to 5.8% for the Andaluza breed ($P < 0.001$).

The hierarchical analysis of the Spanish donkey populations (Table 2) revealed that, as expected, most of the differentiation occurs among breeds with respect to the total population rather than among subpopulations within breeds, and within subpopulations; 0.064 *vs* 0.035, and 0.030, respectively. The differentiation among breeds in the hierarchical analysis was 6.4%, more than that obtained without using a hierarchical analysis (4.1%).

Table 3 shows the total inbreeding estimate ($F \cong F_{IT}$) per breed, evaluated by hierarchical analysis. The F_{IT} average, obtained from jackknifing over loci, ranged from 0.112 ± 0.049 to 0.232 ± 0.058, for the Mallorquina and Andaluza breeds, respectively.

The interbreed genetic distance, or F_{ST} estimates, below the diagonal, and gene flow ($N_e m$) above the diagonal, between pairs of Spanish donkey breeds are shown in Table 4. After 1000 permutations, performed with FSTAT, all F_{ST} calculated by pairwise breed combinations were significantly different from zero ($P < 0.001$). Least differentiation was detected between the Andaluza and Zamorano-Leonesa breeds and Encartaciones-Mallorquina breeds ($F_{ST} = 0.031$), and the most divergence was observed between the Andaluza and Catalana breeds ($F_{ST} = 0.058$). $N_e m$ represents the number of effective migrants exchanged per generation. The $N_e m$ values for pairs of breeds varied from 4.16 to 7.88 for the AND-CAT pair and the MALL-ENC pair, respectively.

Figure 2 shows a NJ tree based on Reynolds genetic distance (data not shown) relating the five Spanish donkey breeds studied, the Moroccan ass, and the Merens breed used as outgroup. The numbers at the nodes are bootstrapping values for 1000 replicates of the 13 loci genotyped.

Discussion

Genetic differentiation among Spanish donkey breeds exists. All loci contribute to this differentiation with F_{ST} values being moderately low and similar for all systems studied, but very significant ($P < 0.001$). However, it is clear that most of the total genetic variation corresponds to differences among individuals (95.9%) and only less

Table 1 Genetic structure of the Spanish donkey breeds through an analysis of their subpopulations

Breeds	$F_{IS} \cong f$	$F_{IT} \cong F$	$F_{ST} \cong \theta$
Andaluza	0.184 (0.058)*** †(0.084–0.306)	0.232 (0.058)*** (0.132–0.352)	0.058 (0.009)*** (0.040–0.075)
Catalana	0.189 (0.053)*** (0.102–0.303)	0.214 (0.052)*** (0.127–0.324)	0.031 (0.007)*** (0.018–0.044)
Mallorquina	0.091 (0.045)*** (0.016–0.193)	0.112 (0.049)*** (0.032–0.221)	0.022 (0.007)*** (0.010–0.038)
Encartaciones	0.118 (0.040)*** (0.053–0.210)	0.130 (0.041)*** (0.062–0.224)	0.013 (0.006)*** (0.003–0.026)
Zamorano-Leonesa	0.193 (0.053)*** (0.099–0.307)	0.223 (0.061)*** (0.116–0.353)	0.037 (0.014)*** (0.014–0.068)

f = within-population inbreeding estimate; F = total inbreeding estimate; θ = measure of population differentiation.
Standard deviation in parentheses estimate from jackknife over loci.
***$P < 0.001$, from permutation tests in FSTAT programme.
†95% Confidence Interval.

Table 2 Hierarchical *F*-statistics and variance components for Spanish donkey breeds with two subdivision levels

Sources of variation	Variance component	Percentage of variation	Fixation indices (%)
Among breeds	0.133	2.96	6.4
Among subpopulations within breeds	0.153	3.41	3.5
Within subpopulations	4.208	93.63	3.0

Table 3 Total breed inbreeding estimated ($F \cong F_{IT}$) by hierarchical analysis in the Spanish donkey breeds
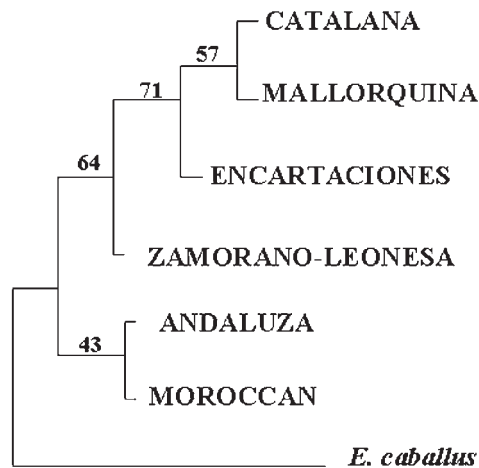
Locus	Andaluza	Catalana	Mallorquina	Encartaciones	Zam-Leonesa
AHT4	0.147***	0.129**	−0.038	0.146**	0.151**
AHT5	0.080**	0.220***	0.138***	0.098*	0.253***
HMS2	0.250***	0.106*	0.195***	0.065	0.223***
HMS3	0.144**	0.117*	−0.085	0.017	−0.030
HMS5	−0.095	−0.037	−0.105	0.133	−0.015
HMS6	0.479***	0.271***	0.033	0.141*	0.315***
HMS7	0.661***	0.357***	0.336***	0.478***	0.577***
HTG4	0.695***	0.823***	0.782***	0.604***	0.944***
HTG6	0.177***	0.364***	0.052	0.153***	0.206***
HTG7	0.224***	0.067**	0.048	0.082	0.246***
HTG10	0.144**	0.125***	0.076**	−0.013	0.139***
HTG15	0.138**	−0.000	0.036	0.093	0.111*
VHL20	−0.015	0.193***	0.031	−0.061	−0.105
Mean	0.232*** (0.058)	0.214*** (0.052)	0.112*** (0.049)	0.130*** (0.041)	0.223*** (0.058)

*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ from permutation tests in FSTAT programme.

Table 4 F_{ST} estimates (below the diagonal) as a measure of genetic distance between donkey breeds and the number of effective migrants per generation (N_em) (above the diagonal)

	Andaluza	Catalana	Mallorquina	Encartaciones	Zam-Leonesa
Andaluza	–	4.167	4.762	6.532	7.797
Catalana	0.058	–	6.134	6.585	4.842
Mallorquina	0.051	0.040	–	7.883	7.118
Encartaciones	0.037	0.037	0.031	–	7.235
Zamorano-Leonesa	0.031	0.050	0.034	0.034	–

|— 0.01 —|

Figure 2 Neighbour-joining dendrogram showing the genetic relationships among the Spanish donkey breeds, inferred through microsatellite data. This tree is based on Reynolds' genetic distance. The numbers at the nodes are values for 1000 bootstrap resamplings of the 13 loci genotyped.

than 5% is due to differences among breeds. These values of total genetic differentiation (F_{ST}) among breeds are close to those found in other domestic species, for example: among river buffalo breeds ($F_{ST} = 0.038$, Barker et al, 1997), among Spanish horse breeds ($F_{ST} = 0.078$, Cañon et al, 2000), though they are slightly lower than those found in Norwegian horse breeds ($F_{ST} = 0.12$, Bjørnstad et al, 2000), in European cattle breeds ($F_{ST} = 0.11$, MacHugh et al, 1998; Kantanen et al, 2000) and among Spanish dogs ($F_{ST} = 0.099$, Jordana et al, 1992).

The results from the hierarchical analysis further show that populations of Spanish donkeys are substructured at different levels (Table 2). The differentiation within breeds is only half of that between breeds (3.5% and 6.4%, respectively). This value (within breeds) was similar to that reported for other organisms: Spanish dogs (Jordana et al, 1992), black-tailed dogs (Chesser, 1983) and house mice (Selander and Kaufman, 1975).

The total inbreeding estimate ($F \cong F_{IT}$, Table 3) showed a significant deficit of heterozygotes ($P < 0.001$) for all breeds. Consanguinity, which is produced by mating between relatives, can be one cause of a deficit of heterozygotes, but this deficit affects all or most of the loci in a similar way. The AND, CAT and ZAM breeds showed 10 or 11 loci, of the total of 13, with a significant deficit of heterozygotes, which might thus result from consanguinity (Table 3). Although in these breeds high and significant values of F_{ST} (Table 1) (0.058, 0.031 and 0.037, respectively), also suggest that a significant subpopulation structure (a Wahlund effect) exists. On the other hand, MALL and ENC breeds showed significant deficits of heterozygotes at only five and six loci, respectively, these deficits being harder to explain by consanguinity.

We must remember that there are other factors that can also cause a lack of heterozygotes in the population (Nei, 1987). First, 'null alleles' (non-amplifying alleles) may be present and lead to a false observation of excess homozy-

gotes. Secondly, the presence of population substructure within the breed may lead to a Wahlund effect.

Reproductive substructure within the breed is a feasible explanation for the high deficit of heterozygotes observed in the Spanish donkey populations. Nevertheless, two loci exist, specifically HTG4 and HMS7, which show a very significant deficit of heterozygotes in all breeds ($P < 0.001$). The most coherent interpretation to explain this deficit is the possible presence, not detected, of 'null alleles' in these populations. Previous reports have indicated the occurrence of a 'null allele' present in the HMS7 locus in the Quarter-horse equine breed (Bozzini et al, 1997).

The dendrogram's topology (Figure 2) is very similar to that obtained by Aranguren-Mèndez et al (2001) using the D_A genetic distance (Nei, 1987) and the NJ algorithm. In this work, the four breeds of black coat from the north of Spain (CAT, ENC, MALL and ZAM) form a closed cluster (64% support), supporting the hypothesis of a common ancestral past from E. asinus europeus.

In this paper, unlike Aranguren-Mèndez et al (2001), the African origin of the Andaluza breed is clarified, which forms a cluster with the Moroccan ass, a genuine representative of the ancestral trunk of E. asinus africanus. This supports the thesis earlier stated by other authors about its African origin (Aparicio, 1960; Epstein, 1984; Sotillo and Serrano, 1985). However, the low bootstrap value for this cluster (43% support) reflects the instability of the topology, so additional studies to confirm this hypothesis are necessary.

Further investigations involving more European and African donkey populations, as well as the analysis of genetic variability of mtDNA, could look for possible introgression of mtDNA haplotypes of African origin into these populations.

Acknowledgements

References

Aparicio G (1960). Zootecnia Especial. Etnología compendiada: Imprenta Moderna Córdoba.

Aranguren-Méndez JA, Jordana J, Gomez M (2001). Genetic diversity in Spanish donkey breeds using microsatellite DNA markers. Genet Sel Evol **33**: 433–442.

Barker JSF, Moore SS, Hetzel DJS, Evans S, Tan SG, Byrne K (1997). Genetic diversity of Asian water buffalo (Bubalus bubalis): microsatellite variation and a comparison with protein-coding loci. Anim Genet **28**: 103–115.

Bellone RR, Cothran EG, Ketchum MS (1998). Genetic variation in the rare donkey breed, Baudet du Poitou. Anim Genet **29**: (Suppl. 1): 17.

Bjørnstad G, Gunby E, Røed KH (2000). Genetic structure of Norwegian horse breeds. J Anim Breed Genet **117**: 307–317.

Bozzini M, Eggleston-Stott ML, Del Valle A, Ziegle JS (1997). Addition of degenerate primer in multiplex PCRs corrects the

equine microsatellite locus HMS7 null allele in the Quarter horse. Abstracts P50. Plant & Animal Genome V Conference. San Diego, CA.

Breen M, Downs P, Irvin Z, Bell K (1994). Intrageneric amplification of horse microsatellite markers with emphasis on the Przewalski's horse (*E. przewalskii*). *Anim Genet* **25**: 401–405.

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

Chesser RK (1983). Genetic variability within and among populations of the black-tailed prairie dog. *Evolution* **37**: 320–331.

Epstein H (1984). Ass, mule and onager. In: Mason IL (ed) *Evolution of Domesticated Animals*, Longman: London and New York, pp 174–184.

Felsenstein J (1995). *PHYLIP (Phylogeny inference package) version 3.57c*. Department of Genetics, University of Eashington, Seattle.

Goudet J (2000). *FSTAT Ver. 2.9.1. Computer package for PCs*. Institute of Ecology, Biology Building, UNIL, CH-1015 Lausanne, Switzerland.

Johannesen J, Loeschcke V (1996). A hierarchical analysis of genetic structure and variability in patchily distributed coexisting *Chiastocheta* species (Diptera: Anthomyiidae). *Heredity* **76**: 437–448.

Jordana J, Piedrafita J, Sanchez A, Puig P (1992). Comparative F statistics analysis of the genetic structure of ten Spanish dog breeds. *J Hered* **83**: 367–374.

Jordana J, Folch P (1998). The Catalonian donkey breed: program of conservation and improvement of an endangered breed. *Arch Zootec* **47**: 403–409.

Jordana J, Folch P, Sanchez A (1999). Genetic variation (protein markers and microsatellites) in endangered Catalonian donkeys. *Bioch Syst Ecol* **27**: 791–798.

Jordana J, Folch P, Aranguren JA (2001). Microsatellite analysis of genetic diversity in the Catalonian donkey breed. *J Anim Breed Genet* **118**: 57–63.

Kantanen J, Olsaker I, Holm LE, Lien S, Vilkki J, Brusgaard K, et al (2000). Genetic diversity and population structure of 20 North European cattle breeds. *J Hered* **91**: 446–457.

Littauer MA, Crouwel JH (1979). *Wheeled Vehicles and Ridden Animals in the Ancient Near East*. Brill: Leiden and Köln.

Machugh DE, Loftus RT, Cunningham P, Bradley DG (1998). Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers. *Anim Genet* **29**: 333–340.

Nei M (1977). F-statistics and analysis of gene diversity in subdivided populations. *Ann Hum Genet* **41**: 225–233.

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

Reynolds J, Weir BS, Cockerham CC (1983). Estimation of the coancestry coefficient: Basis for a short-term genetic distance. *Genetics* **105**: 767–769.

Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.

Schneider S, Roessli D, Excoffier L (2000). *Arlequin ver. 2000: a software for population genetics data analysis*. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Selander RB, Kaufman DW (1975). Genetic structure of populations of the brown snail (*Helix apersa*): I. Microgeographic variation. *Evolution* **29**: 385–401.

Slatkin M (1993). Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**: 264–279.

Sotillo JL, Serrano V (1985). Producción Animal 1, Etnología Zootécnica, Tebar-Flores: Madrid.

Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.

Wright S (1965). Interpretation of population structure by F-statistics with special regard to system of mating. *Evolution* **19**: 395–420.

Wright S (1978). *Evolution and the genetics of populations: Vol. 4. Variability within and among natural populations*. University of Chicago Press: Chicago.