J. Anim. Breed. Genet. 118 (2001), 57–63
© 2001 Blackwell Wissenschafts-Verlag, Berlin
ISSN 0931–2668

Ms. received: 25.4.2000

SHORT COMMUNICATION

Departament de Patologia i de Producció Animals, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain

Microsatellite analysis of genetic diversity in the Catalonian donkey breed

By J. JORDANA, P. FOLCH and J. A. ARANGUREN

Summary

A total of 111 individuals (79 females and 32 males) of the endangered Catalonian donkey breed were analysed by using a commercial equine paternity polymerase chain reaction typing kit. Eleven of the 12 horse microsatellites were amplified when using donkey's DNA. One locus, ASB2, did not amplify in any sample. The allele range for HTG4 overlapped with HMS7 and their results could not be interpreted. The mean number of alleles detected per locus was 7.7 (±1.0), being the average of genetic diversity detected in the population, and measured as the unbiased average expected heterozygosity (H_e), of 0.712 ± 0.038. All the analysed loci showed disagreement with Hardy–Weinberg proportions except HTG7 (from permutation tests). The within-population-inbreeding estimate was highly significant ($p < 0.001$) and equal to 15.4% (as measured by F_{IS}-statistic). The cumulative exclusion probability was 0.999. The deficit of heterozygotes can be partly explained by a population subdivision effect ($F_{ST} = 3.5\%$), although the main factor that has provoked this lack of heterozygotes can be attributed to consanguinity.

Zusammenfassung

Analyse der genetischen Diversität der Katalanischen Eselpopulation mittels Mikrosatelliten

Insgesamt wurden 111 Tiere (79 weibliche und 32 männliche) der vom Aussterben bedrohten Katalanischen Eselrasse mit Hilfe eines kommerziell erhältlichen PCR-Kits zur Abstammungskontrolle beim Pferd untersucht. Elf der zwölf equinen Mikrosatelliten konnten auch beim Esel amplifiziert werden, mit dem Mikrosatellitenmarker ASB2 konnte bei keiner der verwendeten Eselproben ein Produkt amplifiziert werden. Die Marker HTG4 und HMS7 wiesen die gleichen Allelgrößen auf, so dass diese beiden Mikrosatellitenmarker nicht ausgewertet werden konnten. Pro Locus wurden durchschnittlich 7,7 (±1,0) Allele identifiziert, was der durchschnittlichen genetischen Diversität in der Population entspricht und als unverzerrte durchschnittlich erwartete Heterozygotie (H_e) von 0,712 ± 0,038 gemessen wurde. Die untersuchten Loci entsprachen mit Ausnahme des Mikrosatelliten HTG7 (aus dem Permutationstest) nicht dem Hardy–Weinberg Gleichgewicht. Die geschätzte Inzucht innerhalb der Population war hoch signifikant ($p < 0,001$) und entsprach dem Wert von 15,4% wie er auch in der F_{IS} Statistik berechnet wurde. Die gemeinsame Ausschlusswahrscheinlichkeit betrug 0,999. Der Mangel an heterozygoten Tieren kann teilweise durch einen Populationsteilungseffekt ($F_{ST} = 3,5\%$) erklärt werden, obwohl der Hauptfaktor, der diesen Mangel an heterozygoten Tieren hervorruft, der Verwandtschaftsgrad innerhalb der Population ist.

Resumen

Mediante la utilización de un kit comercial para caballos (PE Applied Biosystems), se caracterizó genéticamente la Raza Asnal Catalana, población en peligro de extinción, a partir del análisis de 111 individuos de ambos sexos (79 hembras y 32 machos). Once de los doce microsatélites equinos amplificaron perfectamente utilizando DNA de asno, con la excepción del locus ASB2 que no amplificó para ninguna muestra. Los resultados de HTG4 no pudieron ser interpretados ya que su rango de alelos se solapaba con HMS7. El número medio de alelos detectados por locus fue de 7,7 (±1,0), siendo el

U.S. Copyright Clearance Center Code Statement: 0931–2668/2001/1801–0057 $15.00/0

promedio de diversidad genética detectado en la población, y medida como promedio no sesgado de heterocigosidad esperada (H_e) de 0,712 ± 0,038. Todos los loci analizados mostraron desequilibrio con las proporciones Hardy–Weinberg excepto el locus HTG7 (a partir de tests permutacionales). La estimación de consanguinidad intra población fue altamente significativa ($p < 0,001$), siendo su valor del 15,4% (medida mediante el estadístico F_{IS}). La probabilidad de exclusión acumulada fue de 0,999. El déficit de heterocigotos puede ser parcialmente explicado por un efecto de subdivisión de la población ($F_{ST} = 3.5\%$), aunque el principal factor que ha provocado esta pérdida de heterocigotos deba ser atribuido a la consanguinidad.

Introduction

The Catalonian donkey breed is a population in danger of extinction (at the last census there were slightly more than 100 animals, one-third of which were males), located in several Pyrenean and pre-Pyrenean regions of the Catalonian area of north-east Spain. Following the guidelines proposed by the Food And Agriculture Organisation Of The United Nations (FAO), this breed was characterized morphologically (Folch and Jordana 1997), haematologically (Folch et al. 1997), by clinical biochemical parameters (Jordana et al. 1998) and demographically (Folch and Jordana 1998), in order to carry out the '*Programme of Conservation and Maintenance of Animal Genetic Resources*' in this population (Jordana and Folch 1998).

In a previous study, this breed was genetically characterized (Jordana et al. 1999) by using seven biochemical polymorphism (protein markers) and the analysis of 12 short tandem repeat loci (STRs or microsatellites) by resolution using 10% polyacrylamide gel electrophoresis (PAGE) and ethidium bromide staining. However, the StockMarks for Horses Equine Paternity PCR Typing Kit combines the advantages of polymerase chain reaction (PCR)-based tests and the informativeness of microsatellites to provide an automated approach to genotyping individuals. It is based on 12 microsatellite loci, the primers of which are fluorescently dye-labelled.

Identification of polymorphic markers will allow us to describe levels of genetic variability, estimate the degree of inbreeding, parentage verification, and identify the most heterozygous individuals in the population in order to arrange the best matings for retaining the maximum ancestral genetic variability, in order to carry out the programme of conservation in this breed.

Materials and methods

This study was carried out with 111 individuals of both sexes (79 females and 32 males). We used PCR to amplify Donkey DNA using StockMarks for Horses Kit Reagents (PE Applied Biosystems, Foster City, CA) which includes 12 previously reported loci: ASB2 (Breen et al. 1997); AHT4, AHT5 (Binns et al. 1995); HMS2, HMS3, HMS6, HMS7 (Guerin et al. 1994); HTG4, HTG6 (Ellegren et al. 1992); HTG7, HTG10 (Marklund et al. 1994) and VHL20 (Van Haeringen et al. 1994).

All amplification reactions were performed with 20 ng donkey DNA in two multiplex PCR (eight-plex and four-plex), both in a final volume of 15 μl. These reactions were performed with reagents supplied in the kit and according to manufacturers instructions.

After PCR, aliquots of the multiplex products for each animal were combined to run in a single lane. Samples were analysed using the 310 DNA Sequencer with GENESCAN analysis software (PE Applied Biosystems, Foster City, CA), and ROX internal size as a size standard. This allows reliable computer-aided analysis for fragment sizing, and genotyping.

Statistical analyses

Alleles frequencies and mean heterozygosity values were obtained using the BIOSYS-1 computer program (Swofford and Selander 1989). Tests of genotype frequencies for

deviations from Hardy–Weinberg equilibrium were carried out using the exact tests of the GENEPOP computer program (RAYMOND and ROUSSET 1995), using the Markov chain method (GUO and THOMPSON 1992).

Using the methods of WEIR and COCKERHAM (1984), as implemented in the FSTAT computer program (GOUDET 1995), the *f*-statistic value for each locus was calculated. This statistic is analogous to WRIGHT's (1965; 1978) F_{IS}-statistic; i.e. it measures the deficit or the excess of heterozygotes which could exist in the Catalonian donkey breed. Significance was determined from permutation tests with the sequential Bonferroni procedure (HOCHBERG 1988) applied over all loci (alleles were permuted within the population).

Polymorphic information content (PIC) for each microsatellite loci was calculated according to BOTSTEIN et al. (1980), and probability of exclusion (PE) (JAMIESON 1994) was determined for all systems.

Results

Table 1 shows the amplification characteristics, allele numbers detected, sizes (bp) in each loci, as well as heterozygosity, PIC and PE values for the 10 loci analysed. The ASB2 locus did not amplify for any individual, and HTG4 could not be interpreted due to allele sizes overlapping with those of HMS7. Table 2 shows the allele frequencies obtained for each loci. The average allele number detected per locus was 7.7 (±1.0), ranging between 3 for HMS6 and 12 for AHT5 and HTG7.

All loci, except HMS6 and HTG7, showed disagreement with Hardy–Weinberg proportions, with a significant heterozygotes deficit-exact Hardy–Weinberg test, using the Markov chain method-(GUO and THOMPSON 1992).

The unbiased average expected heterozygosity (H_e; NEI 1978) was 0.712 (± 0.038), values ranging from 0.513 (HMS6) to 0.842 (AHT5). Microsatellites which have PIC values higher than 0.5 are considered highly informative (BOTSTEIN et al. 1980), so with the exception of HMS6 (PIC = 0.40) and HMS7 (PIC = 0.47), we can consider these loci as very informative.

Table 1. **Amplification characteristics, allele number detected, statistics of genetic variation (H_o = observed heterozygosity, H_e = expected heterozygosity, PIC = polymorphic information content) and exclusion probabilities (PE) for the 10 analysed loci**

Locus	Dye label	Size range (bp)	Number of alleles	H_o	H_e^*	PIC	PE
AHT4	FAM	128–158	9	0.667	0.735	0.70	0.53
AHT5	JOE	126–152	12	0.721	0.842	0.82	0.69
HMS2	TAMRA	229–245	8	0.578	0.668	0.62	0.44
HMS3	TAMRA	152–170	7	0.604	0.690	0.64	0.44
HMS6	JOE	155–165	3	0.441	0.513	0.40	0.21
HMS7	FAM	165–173	5	0.369	0.517	0.47	0.29
HTG10	TAMRA	87–107	10	0.745	0.801	0.78	0.63
HTG6	JOE	76–92	7	0.582	0.808	0.77	0.61
HTG7	TAMRA	134–158	12	0.802	0.829	0.81	0.67
VHL20	FAM	75–91	4	0.514	0.713	0.65	0.44
Mean alleles per locus			7.7 (1.0)				
Mean heterozygosity				0.602 (0.043)	0.712 (0.038)		
Cumulative exclusion probability							0.9993

*Unbiased estimate (NEI 1978)
Standard errors in parentheses

Table 2. **Allele sizes (in bp) and allele frequencies (Freq) for 10 microsatellite loci in Catalonian donkey breed**

Locus	Size	Freq	Locus	Size	Freq
AHT4	128	0.023	HMS7	165	0.036
($N = 111$)	136	0.005	($N = 111$)	167	0.194
	138	0.059		169	0.662
	148	0.437		171	0.090
	150	0.032		173	0.018
	152	0.081			
	154	0.225	HTG6	76	0.282
	156	0.122	($N = 110$)	78	0.214
	158	0.018		80	0.027
				82	0.059
AHT5	126	0.054		84	0.150
($N = 111$)	128	0.063		90	0.073
	132	0.144		92	0.195
	134	0.036			
	136	0.045	HTG7	134	0.068
	140	0.032	($N = 111$)	136	0.338
	142	0.009		138	0.027
	144	0.027		142	0.050
	146	0.081		144	0.077
	148	0.131		146	0.041
	150	0.320		148	0.180
	152	0.059		150	0.032
				152	0.023
HMS2	229	0.009		154	0.018
($N = 109$)	231	0.041		156	0.059
	233	0.037		158	0.090
	235	0.234			
	237	0.514	HTG10	87	0.073
	239	0.101	($N = 110$)	91	0.059
	243	0.055		93	0.155
	245	0.009		95	0.005
				97	0.055
HMS3	152	0.036		99	0.091
($N = 111$)	158	0.009		101	0.382
	162	0.072		103	0.091
	164	0.009		105	0.032
	166	0.441		107	0.059
	168	0.311			
	170	0.122	VHL20	75	0.347
			($N = 111$)	77	0.288
HMS6	155	0.550		89	0.081
($N = 111$)	157	0.432		91	0.284
	165	0.018			

N is the total number of samples that has amplified successfully for each locus.

Moreover, the average PE of the 10 markers was 0.49, values ranging from 0.21 (HMS6) to 0.69 (AHT5), the cumulative exclusion probability being 0.999.

Table 3 shows the within-population-inbreeding estimate ($f \cong F_{IS}$). All loci showed values differing significantly from zero, except HTG7, as well as the estimated average ($p < 0.001$). This F_{IS} (f) average, obtained from jackknifing over loci, was equal to 0.154 (±0.030). From a bootstrap analysis the true value of the F_{IS}-statistic, with a 95% confidence interval, would range from 0.102 to 0.212.

Table 3. **Within-population-inbreeding estimate ($f \cong F_{IS}$) in the Catalonian donkey breed**

Locus	$F_{IS} \cong f$
AHT4	0.093**
AHT5	0.144***
HMS2	0.135**
HMS3	0.126***
HMS6	0.140*
HMS7	0.286***
HTG10	0.070*
HTG6	0.281***
HTG7	0.032
VHL20	0.280***
Mean estimate[a]	0.154 (0.030)***

[a]Mean estimate from jackknife over loci. Standard deviation in parentheses.
*p < 0.05, **p < 0.01, ***p < 0.001, from permutation tests in FSTAT program.

Discussion

Few genetic studies have been performed in donkey populations. BREEN et al. (1994), using only eight individuals, verified that a set of 13 microsatellite loci, isolated from the domestic horse (*Equus caballus*) amplified satisfactorily in donkeys.

In a previous work (JORDANA et al. 1999), we genetically characterized the Catalonian donkey breed, using seven biochemical polymorphisms and 12 microsatellite loci. These loci were however, analysed using 10% PAGE and ethidium bromide staining.

Fluorescently dye-labelled primers and the resolution of the amplified products on a Applied Biosystems 310 DNA Sequencer with GENESCAN Analysis software, were essential in detecting a large genetic variability in the Catalonian donkeys in comparison with our previous work.

In the above-mentioned paper, only four out of seven protein markers, and three out of 12 microsatellite loci analysed were polymorphic. Five microsatellites were in common in both studies (HMS3, HMS6, HMS7, HTG6, VHL20), and the average number of alleles detected per locus increased from 2.7 ± 0.7 (JORDANA et al. 1999) to 7.7 ± 1.0 (this article). The average expected heterozygosity (H_e) increased from 0.546 (± 0.049) to 0.712 (± 0.038), and the PIC values as well as the cumulative exclusion probability (PE) values improved significantly. Combining the four protein markers and the three polymorphic microsatellites, the global PE value was only 82.9%, but now it has increased to 99.93%. The StockMarks equine typing kit is therefore, an effective tool in donkey parentage verification.

This study also significantly altered the within-population-inbreeding estimate ($f \cong F_{IS}$). JORDANA et al. (1999) obtained an average F_{IS} value of –0.086 (±0.097) using microsatellites, and a value of –0.079 (±0.050) when using protein markers. For each locus, no value was significantly different from zero. The mean f estimates were not significant either, and the authors postulated the hypothesis that, inbreeding, if it occurs, would be negligible. FOLCH and JORDANA (1998), from the genealogical Catalonian donkey breed data (1979–96 period), obtained an inbreeding estimate of 6% for the actual breeding population. Nevertheless, they suggested that this value was an underestimate, and it would actually be higher due to the scarce pedigree knowledge. The pedigree thoroughness was found to be very incomplete up to the fifth generation of ancestors because the proportion of known ancestors was less than 20%.

A better inbreeding estimate has been obtained from this work. The mean estimate F_{IS} was 0.154 ± 0.030 ($p < 0.001$), showing for all the loci, except HTG7, values significantly different from zero (Table 3). However, the main cause of the lack of heterozygotes (15.4%) can be attributed to inbreeding. It is well known that inbreeding affects all or most loci in a similar way, as a deficit of heterozygotes, and in the present work, a deficit for all markers analysed has been detected. We believe, however, that this inbreeding value could be slightly overestimated.

We must remember that there are other factors that can also cause a lack of heterozygotes in a population (NEI 1987). First, the locus can be under selection, the 'genetic hitchhiking' effect, with some morphological or productive traits of selective interest. Second, 'null alleles' (nonamplifying alleles) may be present which lead to a false observation of excess homozygotes. Third, the presence of population substructure may lead to Wahlund's effect.

The selection influence could not be proved because production data were not available. Despite the fact that pedigrees were available for analysis, it was not possible to demonstrate the presence of null alleles (usually caused by a mutation in the primer-binding site leading to an allele that will not amplify). This was due to the scarce pedigree knowledge that could have allowed us to examine the heredity of such alleles. Nevertheless, we can not disregard a certain, although slight, effect of these alleles in the observed deficit of heterozygotes.

The analysis of a large number of loci increases the power of detecting population substructure, because each locus will contain an independent history of the population depending on the amount of random drift, mutation, and migration that has occurred.

Because of that, the population was divided into three subpopulations (SP1 = 50, SP2 = 47, and SP3 = 14 individuals) according to geographic distribution criteria, breeding policy and scarce genetic flow. *F*-statistics were computed by using the methods of WEIR and COCKERHAM (1984), as implemented in the computer program FSTAT, and the significance of *F*-statistics estimates was determined from permutation tests.

For each locus F_{IS} (f), F_{IT} (F) and F_{ST} (θ) values were calculated from jackknife over populations (data not shown). The mean estimates, from jackknife over loci, were: $f = 0.135 \pm 0.031$; $p < 0.001$ (within-population-inbreeding estimate), $F = 0.165 \pm 0.030$; $p < 0.001$ (total inbreeding estimate), and $\theta = 0.035 \pm 0.006$; $p < 0.001$ (measure of population differentiation). A significant θ-value shows us that there is a certain degree of population differentiation in this breed (3.5%). In such a statistic (θ), only two loci (HMS2 and HMS6) were not statistically different from zero. This population substructure can be the origin of some of the differences in allelic frequency among these subpopulations for some loci (data not shown), and this has caused the entire population to have a net deficiency of heterozygotes even if Hardy–Weinberg equilibrium exists within each subpopulation (Wahlund effect). In Table 3 it can be observed that all loci, except HTG7, showed very significant disagreement with Hardy–Weinberg equilibrium, in the sense of a loss of heterozygotes, for the entire population; whereas two, three and seven loci were in Hardy–Weinberg equilibrium agreement for subpopulations SP1, SP2 and SP3, respectively.

To conclude, we can say that the average inbreeding estimate in the Catalonian donkey breed, is notably higher than the values obtained by JORDANA et al. (1999) and by FOLCH and JORDANA (1998). However, part of the 15.4% value of deficit of heterozygotes can also be attributed to the subpopulation structure in this endangered breed.

Acknowledgements

The authors wish to thank the Departament d'Agricultura, Ramaderia i Pesca of the Generalitat de Catalunya and the CICYT (project AGF98–0503), which financed this study; likewise, the AFRAC

association for their helpful co-operation and assistance during the sample collection. Also, we would like to thank JULIE SMITH for the English revision.

References

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. Brit. Vet. J. **151:** 9–15.

BOTSTEIN, D.; WHITE, R. L.; SKOLNICK, M.; DAVIS, R. W., 1980: Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. **32:** 314–331.

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.

BREEN, M.; LINDGREN, G.; BINNS, M. M., 1997: Genetical and physical assignments of equine microsatellites – first integration of anchored markers in horse genome mapping. Mamm. Genome **8:** 267–273.

ELLEGREN, H.; JOHANSSON, M.; SANDBERG, K.; ANDERSSON, L., 1992: Cloning of highly polymorphic microsatellites in the horse. Anim. Genet. **23:** 133–142.

FOLCH, P.; JORDANA, J.,1997: Characterization, reference ranges and the influence of gender on morphological parameters of the endangered Catalonian donkey breed. J. Equine Vet. Sci. **17:** 102–111.

FOLCH, P.; JORDANA, J., 1998: Demographic characterization, inbreeding and maintenance of genetic diversity in the endangered Catalonian donkey breed. Genet. Sel. Evol. **30:** 197–203.

FOLCH, P.; JORDANA, J.; CUENCA, C., 1997: Reference ranges and the influence of age and sex on haematological values of the endangered Catalonian donkey breed. Vet. J. **154:** 163–168.

GOUDET, J., 1995: FSTAT (Version 1.2): a computer program to calculate F-statistics. J. Hered. **86:** 385–386.

GUERIN, G.; BERTAUD, M.; AMIGUES, Y., 1994: Characterisation of seven new horse microsatellites. HMS1, HMS2, HMS3, HMS5, HMS6, HMS7, HMS8. Anim. Genet. **25:** 62.

GUO, S. W.; THOMPSON, E. A., 1992: Performing the exact test of Hardy–Weinberg proportions for multiple alleles. Biometrics **48:** 361–372.

HOCHBERG, Y., 1988: A sharper Bonferroni procedure for multiple tests of significance. Biometrika **75:** 800–802.

JAMIESON, A., 1994: The effectiveness of using co-dominant polymorphic allelic series for (1) checking pedigrees and (2) distinguishing full-sib pair members. Anim. Genet. **25:** 37–44.

JORDANA, J.; FOLCH, P., 1998: Programa de conservació i manteniment de recursos genètics animals en la Raća Asinina Catalana. Catalunya Rural i Agraria **42:** 5–8.

JORDANA, J.; FOLCH, P.; CUENCA, C., 1998: Clinical biochemical parameters of the endangered Catalonian donkey breed: normal values and the influence of sex, age and management practices effect. Res. Vet. Sci. **64:** 7–10.

JORDANA, J.; FOLCH, P.; SANCHEZ, A., 1999: Genetic variation (protein markers and microsatellites) in endangered Catalonian donkeys. Biochem. Syst. Ecol. **27:** 791–798.

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. Anim. Genet. **25:** 19–23.

NEI, M., 1978: Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics **89:** 583–590.

NEI, M., 1987: Molecular Evolutionary Genetics. Columbia University Press. New York.

RAYMOND, M.; ROUSSET, F., 1995: GENEPOP (Version 1.2): Population genetics software for exact test and ecumenicism. J. Hered. **86:** 248–249.

SWOFFORD, D. L.; SELANDER, R. B., 1989: BIOSYS-1 User's Manual, Release 1.7. A Computer Program for the Analysis of Allelic Variation in Population Genetics and Biochemical Systematics. Illinois Natural History Survey, Champaign IL.

VAN HAERINGEN, H.; BOWLING, A. T.; STOTT, M. L.; LENSTRA, J. A.; ZWAAGSTRA, K. A., 1994: A highly polymorphic horse microsatellite locus, VHL20. Anim. Genet. **25:** 207.

WEIR, B. S.; COCKERHAM, C. C., 1984: Estimating F-statistics for the analysis of population structure. Evolution **38:** 1358–1370.

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

WRIGHT, S., 1978: Evolution and the Genetics of Populations, Vol. 4. University of Chicago Press, Chicago, IL.

Authors' address: J. JORDANA, P. FOLCH and J. A. ARANGUREN, Unitat de Genètica i Millora Animal, Departament de Patologia i Producció Animals, Facultat de Veterinària, Universitat Autònoma de Barcelona, 08193-Bellaterra, Barcelona, Spain.