doi:10.1111/j.1365-2052.2006.01545.x

Genetic diversity and relationships of Portuguese and other horse breeds based on protein and microsatellite loci variation

C. Luís*, R. Juras†, M. M. Oom* and E. G. Cothran†

*Universidade de Lisboa, Faculdade de Ciências, Departamento de Biologia Animal, Centro de Biologia Ambiental, Edifício C2-Piso 3, Campo Grande, 1749–016 Lisboa, Portugal. †Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX 77843, USA

Summary

There are three native Portuguese horse breeds: Lusitano, Sorraia and Garrano. This study compares diversity patterns of 17 protein and 12 microsatellite markers in these three as well as 30 other breeds to infer relationships among the breeds and to compare levels of polymorphism of these breeds for use in conservation efforts. The Garrano and the Lusitano showed a high level of genetic diversity, similar to that observed for most of the other analysed breeds, while the Sorraia and Friesian breeds showed low levels of variation for both genetic marker types. The combined protein and microsatellite data produced a tree that fit historical records well and with greater confidence levels than those for either data set alone. The combined genetic diversity and relationship information provides important baseline data for future breed conservation efforts, especially for a critically endangered breed such as the Sorraia.

Keywords horse breed history, microsatellites, proteins, genetic variation.

Introduction

The horse has a recognized importance as an animal genetic resource all over the world. Equines have a very old and continuous presence in the Iberian Peninsula since the Pleistocene (Gonzaga 2004). It is known that horses from the Peninsula contributed to the development of many other modern European horses and were later introduced and dispersed throughout the Americas, where they are the founders of numerous New World breeds (e.g. Andrade 1937; Gonzaga 2004; Luís *et al.* 2006). The horse has been used since the Paleolithic in Portugal, where today there are three native horse breeds.

The Lusitano is one of the world's most ancient breeds. It is bred mainly in the Ribatejo and Alentejo regions of Portugal (around 2000 breeding mares), with an increasing number being bred in foreign countries, primarily Brazil, France and México (with another 2000 breeding mares). This breed has had an official studbook since 1967, and its breeding is inextricably tied up with cattle breeding and bullfighting. It shares the same heritage and evolution as the Pura Raza Española of Spain, both being designated as 'Andalusian', 'Iberian' or 'Peninsular' horses, which have been selected over centuries for use in Iberian wars (gineta cavalry tactics). While the Lusitano and Pura Raza Española are very similar in general characteristics and appearance, they are designated as two distinct pure-bred Iberian saddle horse breeds. Differences that exist today are the result of selective breeding for different purposes because of the practice of bullfighting in Portugal and its abandonment in Spain, with the Lusitano considered closer to the original Iberian type (Loch 1986; Oom 1992).

The Sorraia breed is extremely hardy and noted for its ability to survive in very adverse conditions, feeding on pastures frequently flooded in winter and dry during summer. It was recovered by Dr Ruy d'Andrade in 1937 from 12 founders, including five males and seven females. It is a horse with primitive characteristics, such as dun or grullo color, black dorsal stripe, black-tipped ears and striped legs, and is believed to be the ancestor of the southern Iberian saddle horse breeds with a subconvex profile and possibly one of the ancestors of the world's light saddle horses (e.g. Andrade 1937; 1945; Oom *et al.* 2004; Oom & Cothran 1994). Because of the extremely reduced effective size of the population (<200 animals from which around 80 are breeding mares), it is currently considered as a breed in critical maintained risk status, according to criteria established by the Food and Agriculture Organization (FAO

Address for correspondence

C. Luís, DBA/CBA, Faculdade de Ciências da Universidade de Lisboa, Edifício C2-Piso 3, Campo Grande, 1749-016 Lisboa, Portugal.
E-mail: cmluis@fc.ul.pt

Accepted for publication 8 October 2006

1998; DAD-IS Database http://dad.fao.org). The official studbook was published in 2004 (Oom *et al.* 2004).

The Garrano is a pony breed from the northwest mountains of Portugal where it is still bred in a semi-feral regime because of its sturdiness and resilience (the population has around 950 breeding mares and around 2000 individuals). Animals of this breed have a straight, sometimes concave, head profile, bay coat and very dense mane and tail. This general morphological type is represented in Paleolithic paintings from north Iberian caves (Andrade 1938; Oom 1992), suggesting the Garrano may have originated from an ancient lineage very distinct from the other two southern Portuguese native breeds, and thus harbour genetic information relevant to conservation strategies.

Microsatellites and protein loci have been widely used to investigate genetic relationships of horse populations (e.g. Cothran *et al.* 1998; Nozawa *et al.* 1998; Cañon *et al.* 2000; Juras *et al.* 2003; Aberle *et al.* 2004; Solis *et al.* 2005), and can provide an indication of the levels of inter- and intra-breed variability. This study compares diversity patterns of microsatellite and protein markers in several horse breeds. The genetic markers were used to infer relationships of Portuguese and worldwide horse breeds, as well as to compare levels of polymorphism in the analysed breeds. This information will provide a basis for supporting sustainable breed management and conservation decisions.

Material and methods

Sampling and protein analysis

Blood samples were collected from 33 horse breeds (Table 1), sampling across several populations within each breed when possible. The samples were separated into red blood cells (RBC), RBC lysate and serum.

Standard immunological procedures involving haemaglutination and complement-mediated haemolysis (Stormont & Suzuki 1964; Stormont *et al.* 1964) were used to detect variation of red cell alloantigens at seven blood group loci (EAA, EAC, EAD, EAK, EAP, EAQ and EAU).

Starch gel electrophoresis, polyacrylamide gel electrophoresis and isoelectric focusing were used to detect variation at 10 serum and RBC lysate protein loci (Braend 1973; Sandberg 1974; Juneja *et al.* 1978; Pollitt & Bell 1980; Braend & Johansen 1983; Henney *et al.* 1994). The biochemical protein loci analysed were alpha-1-beta glycoprotein (A1B), albumin (ALB), serum carboxylesterase (ES), vitamin D binding protein (GC), glucosephosphate isomerase (GPI), haemoglobin alpha (HBA), phosphogluconate dehydrogenase (PGD), phosphoglucomutase (PGM), protease inhibitor (PI) and transferrin (TF). Nomenclature for variants was in accordance with internationally standardized usage for horses (Bowling & Clark 1985; Bowling & Ryder 1987), except for variants at some loci (ALB, 6-PGD, PI and TF), which have not yet received international recognition.

DNA extraction and microsatellite analysis

DNA was extracted from blood using the modified method described by Gustincich *et al.* (1991) and using the high salt extraction procedure (Montgomery & Sise 1990). The DNA typing panel consisted of 12 microsatellites: *ASB2* (Breen *et al.* 1997), *AHT4*, *AHT5* (Binns *et al.* 1995), *HMS2*, *HMS3*, *HMS6*, *HMS7* (Guérin *et al.* 1994), *HTG4*, *HTG6* (Ellegren *et al.* 1992), *HTG7*, *HTG10* (Marklund *et al.* 1994) and *VHL20* (Van Haeringen *et al.* 1994).

Fragment sizes of microsatellite alleles were determined using the computer software programs STRAND (Hughes 2000) and RFLPSCAN 3.1 (Scanalytics, CPS Inc., Rockville, MD, USA).

Statistical analysis

Frequencies of alleles at blood group loci were calculated by the allocation method (Andersson 1985). Blood group frequency data were used for the calculation of genetic distance and phylogenetic analyses. Genetic variation measures are not reported because genotypes could not be accurately determined.

Genetic variability for the microsatellite and protein loci was measured by estimating mean number of alleles per locus (MNA), as well as observed (H_o) and unbiased expected (H_e) heterozygosities (Nei 1978) with MICROSATELLITE TOOLKIT version 3.1 (Park 2001). Descriptive statistics were obtained with the GDA 1.0 (d15) computer program (Lewis & Zaykin 2000). Departures from Hardy–Weinberg equilibrium (HWE) were tested using GENEPOP 3.4 (Raymond & Rousset 1995). To account for multiple simultaneous tests, the sequential Bonferroni procedure (Holm 1979; Rice 1989) was applied.

The chord distance, Dc (Cavalli-Sforza & Edwards 1967), is one of the best measures to evaluate the distances of populations with intermediate divergence times, as represented by horse breeds worldwide. However, along with Dc, the distance of Nei's *et al.* (1983) Da has also been considered very efficient in obtaining the correct tree topology of closely related populations (Takezaki & Nei 1996). Therefore, we estimated both Dc and Da distances for the protein and microsatellite data separately using the Neighbour-Joining clustering method. These calculations were performed with POPULATIONS 1.2.28 software (written by Olivier Langella) and the trees obtained were visualized in TREEVIEW (Page 1996). The genetic relationships of breeds, using the combined data from microsatellites, proteins and blood groups, were obtained by applying the restricted maximum likelihood (REML) analysis to the Dc distance using PHYLIP software (Felsenstein 1993), with the Przewalski horse being used as an outgroup. The distance matrices obtained from using both distance estimators were statistically compared by use of the Mantel test (Mantel 1967), performed with the NTSYS-pc package version 1.80 (Rohlf 1993; Exeter Software, Setauket, NY, USA).

Table 1 Summary statistics for horse breeds used in microsatellite and proteins marker analysis.

	Microsatellites					Proteins				
Breeds	*n*	H_e	H_o	MNA	FixAl	*n*	H_e	H_o	MNA	FixAl
Portugal										
Garrano	37	0.779 ± 0.020	0.745 ± 0.021	7.17 ± 1.53	–	72	0.379 ± 0.073	0.388 ± 0.019	4.30 ± 3.13	–
Lusitano	70	0.721 ± 0.041	0.677 ± 0.017	6.33 ± 1.83	–	245	0.357 ± 0.096	0.360 ± 0.010	4.50 ± 3.98	–
Sorraia	60	0.506 ± 0.072	0.487 ± 0.019	3.83 ± 1.34	HTG7	46	0.347 ± 0.089	0.359 ± 0.022	2.40 ± 1.43	GC, PGM
Asia										
Arabian	95	0.690 ± 0.037	0.624 ± 0.016	6.58 ± 2.27	–	120	0.329 ± 0.083	0.308 ± 0.013	3.80 ± 3.01	–
Akhal-Teke	84	0.715 ± 0.029	0.668 ± 0.016	7.58 ± 1.73	–	85	0.360 ± 0.074	0.384 ± 0.017	3.30 ± 2.26	–
Caspian Pony	147	0.757 ± 0.026	0.746 ± 0.011	7.75 ± 1.96	–	94	0.361 ± 0.092	0.383 ± 0.016	4.20 ± 3.74	–
Europe										
Andalusian	33	0.727 ± 0.031	0.693 ± 0.023	5.67 ± 0.89	–	140	0.363 ± 0.080	0.348 ± 0.013	4.40 ± 3.81	–
Conemmara	69	0.745 ± 0.027	0.768 ± 0.015	6.50 ± 1.83	–	66	0.452 ± 0.083	0.453 ± 0.019	4.40 ± 4.01	–
Dales Pony	43	0.662 ± 0.040	0.715 ± 0.020	5.58 ± 1.31	–	42	0.380 ± 0.091	0.348 ± 0.023	3.50 ± 2.68	GPI
Exmoor	98	0.609 ± 0.032	0.601 ± 0.014	5.25 ± 1.14	–	102	0.470 ± 0.065	0.451 ± 0.016	3.60 ± 2.55	–
Fell Pony	53	0.731 ± 0.030	0.782 ± 0.018	6.42 ± 1.38	–	42	0.427 ± 0.081	0.445 ± 0.024	3.70 ± 2.83	PGD
Friesian	159	0.466 ± 0.053	0.454 ± 0.013	4.50 ± 1.45	–	313	0.307 ± 0.073	0.307 ± 0.008	2.90 ± 1.66	A1B, PGD
Hanoverian	28	0.767 ± 0.019	0.698 ± 0.027	6.58 ± 1.51	–	41	0.371 ± 0.085	0.379 ± 0.025	4.60 ± 3.72	–
Haflinger	341	0.641 ± 0.058	0.630 ± 0.009	6.25 ± 2.05	–	158	0.414 ± 0.092	0.407 ± 0.012	3.90 ± 3.60	PGD
Holsteiner	27	0.705 ± 0.028	0.733 ± 0.028	5.42 ± 1.93	–	15	0.372 ± 0.101	0.373 ± 0.039	3.00 ± 2.36	GC, PGM, GPI
Irish Draught	47	0.772 ± 0.020	0.766 ± 0.018	7.08 ± 1.62	–	63	0.409 ± 0.087	0.399 ± 0.020	3.90 ± 3.48	–
Lippizzan	40	0.724 ± 0.015	0.700 ± 0.021	6.33 ± 1.44	–	140	0.323 ± 0.094	0.326 ± 0.013	3.80 ± 3.29	GPI
Shetland Pony	36	0.672 ± 0.028	0.671 ± 0.023	6.00 ± 1.65	–	22	0.324 ± 0.094	0.306 ± 0.033	2.60 ± 1.58	PGD, PGM, GPI
Suffolk Punch	66	0.724 ± 0.028	0.679 ± 0.017	6.42 ± 1.51	–	125	0.434 ± 0.069	0.437 ± 0.014	3.70 ± 3.20	–
Thoroughbred	175	0.695 ± 0.031	0.674 ± 0.011	6.25 ± 1.91	–	144	0.292 ± 0.090	0.300 ± 0.012	3.00 ± 2.54	PGM, GPI
North America										
American Saddlebred	63	0.728 ± 0.043	0.730 ± 0.017	7.25 ± 1.82	–	162	0.404 ± 0.089	0.389 ± 0.012	4.60 ± 4.01	–
Morgan Horse	60	0.740 ± 0.026	0.712 ± 0.018	7.25 ± 2.18	–	56	0.405 ± 0.073	0.397 ± 0.021	4.20 ± 3.79	–
Quarter Horse	41	0.752 ± 0.026	0.717 ± 0.022	7.00 ± 1.65	–	190	0.394 ± 0.084	0.393 ± 0.011	5.00 ± 4.22	–
Rocky Mountain	45	0.749 ± 0.024	0.727 ± 0.019	6.75 ± 1.29	–	132	0.372 ± 0.085	0.350 ± 0.013	4.30 ± 3.86	–
Standardbred	31	0.739 ± 0.021	0.733 ± 0.027	6.00 ± 1.95	–	211	0.395 ± 0.070	0.381 ± 0.011	4.20 ± 3.99	–
South America										
Brazillian Criollo	43	0.742 ± 0.021	0.761 ± 0.021	6.58 ± 1.56	–	47	0.438 ± 0.099	0.422 ± 0.023	4.50 ± 3.89	–
Chilean Criollo	30	0.726 ± 0.020	0.703 ± 0.024	5.67 ± 1.15	–	173	0.384 ± 0.097	0.372 ± 0.012	4.80 ± 4.76	–
Colombian Paso Fino	30	0.740 ± 0.022	0.750 ± 0.023	6.75 ± 1.22	–	41	0.337 ± 0.091	0.339 ± 0.023	3.70 ± 3.13	GC
Chilote	30	0.756 ± 0.023	0.718 ± 0.024	7.50 ± 1.38	–	58	0.396 ± 0.092	0.391 ± 0.020	3.70 ± 2.87	GC
Campolina	30	0.725 ± 0.037	0.713 ± 0.024	6.42 ± 1.38	–	106	0.404 ± 0.094	0.409 ± 0.015	4.80 ± 4.44	–
Peruvian Paso Fino	38	0.738 ± 0.035	0.739 ± 0.021	6.83 ± 1.40	–	141	0.446 ± 0.090	0.450 ± 0.013	4.60 ± 4.38	–
Pantaneiro	28	0.741 ± 0.026	0.764 ± 0.025	6.50 ± 1.51	–	102	0.381 ± 0.091	0.381 ± 0.015	4.10 ± 3.70	–
Puerto Rican Paso Fino	80	0.714 ± 0.020	0.668 ± 0.017	7.67 ± 1.37	–	61	0.427 ± 0.084	0.472 ± 0.020	4.60 ± 4.43	–

A1B, alpha-1-beta glycoprotein; GC, vitamin D binding protein; GPI, glucosephosphate isomerase; PGD, phosphogluconate dehydrogenase; PGM, phosphoglucomutase; *n*, sample size; Ho, observed heterozygosity; He, expected heterozygosity; MNA, average number of alleles per locus; FixAl, loci that show fixed alleles.

Results

Levels of variation

A total of 194 alleles were detected across 10 protein markers (69 alleles) and 12 microsatellite markers (125 alleles). Allele frequencies are shown in Tables S1 & S2. The average number of alleles per locus was 10.4 for microsatellites (range: 7–13) and 6.9 for proteins (range: 2–23). The least polymorphic microsatellite was *HTG7* (0.560), and the highest genetic diversity (0.774) was found at the *VHL20* microsatellite. Of the protein loci, the lowest diversity was found for A1B (0.147) and the highest for PI (0.790) (Table S3).

Statistics for each horse breed are presented in Table 1. For the microsatellite loci tested across populations, the lowest observed heterozygosity (0.454) and gene diversity (0.466) were found in the Friesian, while the highest observed heterozygosity (0.782) and highest gene diversity (0.779) were found in the Fell Pony and the Garrano respectively. The lowest variability (MNA = 3.83) was found in Sorraia whereas the Caspian breed had the highest (7.75).

Regarding the protein loci tested across populations (Table 1), the lowest observed heterozygosity (0.300) and gene diversity (0.292) were found in the Thoroughbred, while the Puerto Rican PasoFino presented the highest observed heterozygosity (0.472) and the Exmoor the highest gene diversity (0.470). The Sorraia had the lowest MNA (2.40) and the Quarter Horse the highest (5.00). Only the Sorraia breed showed a non-polymorphic microsatellite locus, while 11 breeds were monomorphic for at least one of five protein loci (Table 1).

Of the Portuguese breeds, the Garrano showed the higher variability for most of the parameters considered (except for the protein MNA, where the Lusitano had the highest value), while the Sorraia showed the lowest variability for both microsatellite and protein loci.

When pooled across microsatellite loci, four populations (Caspian, Friesian, Haflinger and Suffolk Punch) had highly significant deviations from HWE ($P < 0.001$). There were no deviations from HWE for protein loci ($P > 0.05$).

Genetic distances and breed relationships

Distance matrices obtained from microsatellite and protein data correlated fairly well for both Da ($r = 0.791$ and t-test = 5.35) and Dc ($r = 0.760$ and t-test = 5.27). Trees obtained for each of the distances (Da and Dc) and markers analysed (microsatellite, proteins and blood groups) showed some cluster differences (data not shown). However, there were some consistent groups in all trees: (i) American Saddlebred, Rocky Mountain, Standardbred and Morgan Horse; (ii) Friesian, Dales Pony and Fell Pony; (iii) Lusitano and Andalusian; and (iv) Irish Draught, Quarter Horse, Hanoverian, Thoroughbred and Holstein.

The tree obtained from the application of REML to the Dc using microsatellite, protein and blood group data is shown in Fig. 1. Seven major groups and one subgroup are identified in this tree, indicated in the figure by letters A–H.

Discussion

The Garrano and the Lusitano Portuguese breeds show a high level of genetic diversity similar to that observed for most of the other breeds analysed in this and other studies (e.g. Cañon *et al.* 2000; Aberle *et al.* 2004; Solis *et al.* 2005). Hence, there is no immediate danger in the conservation status of the Lusitano and Garrano breeds (Luís *et al.* 2002b; Morais *et al.* 2005). The Sorraia horse showed low levels of microsatellite variation, a fact that has been reported in several previous works using microsatellites and other genetic markers (Oom & Cothran 1994; Luís *et al.* 2002a,b; Aberle *et al.* 2004; Luís *et al.* 2005). The Sorraia breed suffers from a severe founder effect as it descends from only 12 founders, and this combined with a small effective size associated with genetic drift account for this breed's reduced variability. However, it is interesting to note that the Sorraia breed shows a low allelic diversity for protein loci but a heterozygosity level near that observed for the other breeds analysed in this study. The difference in the Sorraia's relative heterozygosity levels for microsatellites and biochemical loci is likely because of the nature of variation at these types of loci. Microsatellites have a large number of alleles, many at low frequency, whereas protein loci have relatively fewer alleles. According to Nei *et al.* (1975), allele numbers are usually reduced faster than heterozygosity during inbreeding or a bottleneck period, as seen for the protein loci. For the microsatellites, the Sorraia shows a very low MNA (Table 1) and also, very low heterozygosity as compared with all but the Friesian breed. This likely is because of the loss of rare alleles and their contribution to heterozygosity.

The Friesian has experienced a severe bottleneck in recent times with the number of breeding stallions reduced to just three after World War II (Hendricks 1995). This pattern is reflected in the fairly concordant results obtained for microsatellites and proteins, both indicating reduced number of alleles and low levels of heterozygosity. Differences in breeding practices for Friesian and Sorraia after the bottleneck may account for the differences in respective MNA and H_o levels as revealed by both types of markers.

The significant deviations from HWE for microsatellites in several breeds may be because of genetic substructure resulting from different local selection strategies across studs. Sampling procedures across unrecognized substructures also could be a cause for some of the observed deviations from HWE. The higher number of alleles in microsatellites, together with the large differences in selection pressures between mares and stallions, could lead to

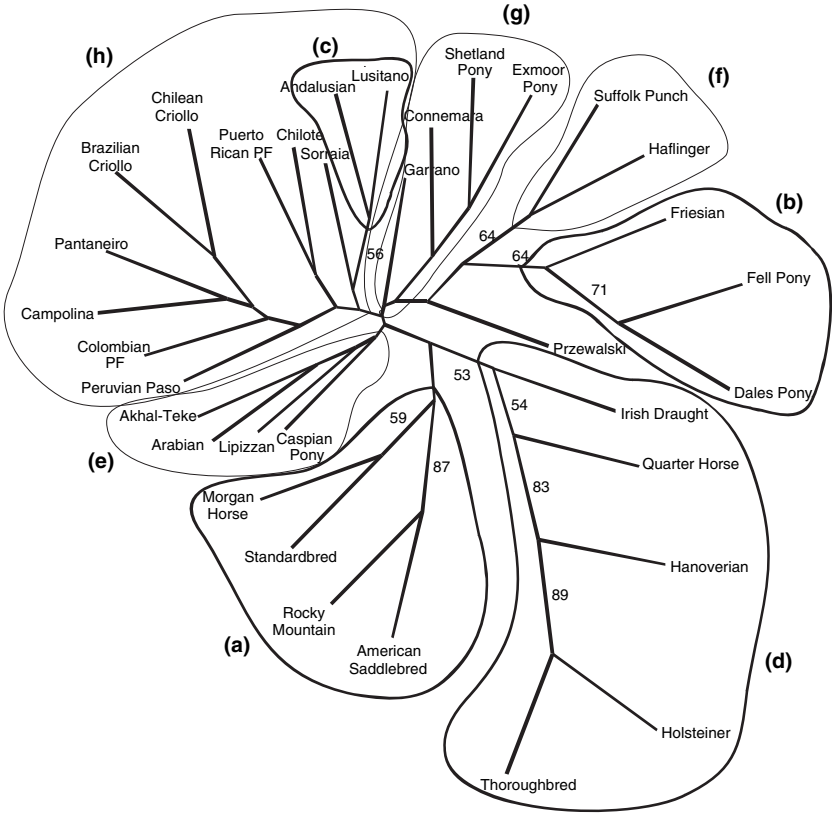

Figure 1 Restricted maximum likelihood network for the 33 horse populations based on the chord distance measure estimated with three genetic markers (microsatellites, proteins and blood groups). Major groups and one subgroup are identified by letters from a to h. Clusters with bootstrap values over 50 are indicated. Thick lines surround four breed groups found consistently across different distances and clustering methods.

greater differences among subgroups within a breed which could account for more deviations from HWE in microsatellites than in proteins. In addition, the possible occurrence of non-amplifying or null alleles could have lead to false observation of homozygotes. However, for all breeds where deviations were found, parentage testing did not give evidence of null alleles.

Four groups were found consistently across different distances and clustering methods (Fig. 1, thick lines). Group A was formed by four North American horse breeds, the Rocky Mountain, the American Saddlebred, the Standardbred and the Morgan Horse. According to Hendricks (1995) the Rocky Mountain is considered a close relative to the American Saddlebred. The Morgan Horse had influence in the American Saddlebred, and the Massachusetts Narragansett Pacer was important in the establishment of the Standardbred and American Saddlebred breeds.

Group B clusters two very close breeds, the Dales Pony and the Fell Pony, with the Friesian, which have no apparent relationship. However, according to the English author Anthony Dent there is historical evidence that the Friesian influenced the Fell Pony and the Dales Pony (Hendricks 1995).

Group C is formed by the Iberian Lusitano and Andalusian breeds from Portugal and Spain respectively. These two breeds only differ by the recent selection criteria used in both countries (Oom 1992; Bowling & Ruvinsky 2000). The Lusitano has been mostly selected in Portugal for functional purposes, while the Andalusian has been selected in Spain for its general conformation (Andrade 1987).

Group D clusters the Irish Draught, Quarter Horse, Hanoverian, Thoroughbred and Holstein. The common link between all these breeds is the influence they had from the Thoroughbred (Hendricks 1995; Sponenberg 1996). The Holstein experienced many introductions from Thoroughbred in the 19th century and especially since World War II (Hendricks 1995). Thoroughbreds greatly influenced the Irish Draught throughout its establishment, as it was affirmed that crossing the Irish Draught to Thoroughbreds produced excellent hunters and competition horses (Hendricks 1995). Quarter horse also had infusion of a good deal of Thoroughbred blood, particularly in racing lines.

Besides the consistent groups obtained throughout the use of different distances and markers, there are also other interesting groups that emerge from the joint use of all three types of genetic markers. One of the clusters included the Caspian Pony, Arabian, Akhal-Teke and Lipizzaner breeds (Fig. 1e). The first three breeds are part of the oriental-type horses, so this clustering would be expected based upon their origin in the Middle East and Asian steppes, and it is in accordance with results previously obtained by Cothran & Luís (2005). The presence of the Lipizzan in this cluster might be related to Arabian horse influence during the formation of that breed (Dovč *et al.* 2004). The clustering between the Suffolk Punch and the Haflinger (Fig. 1f) reflects their cold-blood grouping. Another interesting group

is the one formed by several pony breeds, Garrano, Connemara, Shetland Pony and Exmoor Pony (Fig. 1g), all of which share a Celtic connection (e.g. Cañon *et al.* 2000; Oom & Cothran 1996).

All the South American breeds belong to a group that also clusters with the Iberian Andalusian, Lusitano and Sorraia breeds (Fig. 1h). The introduction of horses in the New World by Spanish and Portuguese explorers and colonists is well documented (e.g. Andrade 1954; Bort 2004). During the course of history the South American breeds likely had introduction of blood from breeds other than Iberian breeds. However, genetic evidence based upon the use of mtDNA (Mirol *et al.* 2002; Luís *et al.* 2006) and the close relationship found in this work is in accordance with greater influence from the Spanish and Portuguese horses in the New World breeds.

Because of its low variability, the Sorraia usually separated from the other breeds when only one type of marker was used, as seen also in Oom & Cothran (1994); Juras *et al.* (2003); Cothran & Luís (2005) and Morais *et al.* (2005). However, when we used all three types of markers the Sorraia clustered with the other Iberian and Iberian-influenced breeds. Indeed the Sorraia is considered a primitive horse and is believed to be the primary ancestor of the Iberian horses and therefore an ancestor of light and saddle horses. This theory has not been proven in previous studies using mitochondrial DNA caused by the reduced number of matrilineal lines found for the Sorraia (Jansen *et al.* 2002; Luís *et al.* 2002a; Lopes *et al.* 2005; Royo *et al.* 2005). However, new results from mtDNA analysis (Luís *et al.* 2006) revealed a lost haplotype in the Sorraia breed, which is included in the mtDNA haplogroup recognized for having high frequency in Iberian horses (Lopes *et al.* 2005; Royo *et al.* 2005; Luís *et al.* 2006), and the results obtained here indicate closeness of the Sorraia with the other Iberian breeds, which fits historical documentation.

The results obtained here can be used for advising breed management for conservation purposes, first because they provide comparable measures of diversity among different breeds, showing which ones are threatened with extinction because of progressive loss of genetic variation, and, second, because they inform breeders of the relationships between breeds thus allowing better prioritization of breeds for conservation and for possible crossbreeding of animals from close breeds to overcome unavoidable loss of genetic diversity. Based upon the known demographic history of the Sorraia horse, low variation would be expected, and the data gathered in this study clearly supports the expectations concerning extensive depletion of this breed's genetic variation. The information can be used in the establishment of a breeding programme designed to avoid inbreeding and to maximize gene diversity in order to minimize further loss of genetic variation.

Genetic distance estimates will vary greatly according to the marker used and the recent demographic history of the breed. For example, a severely bottlenecked breed, such has the Sorraia horse, if analysed with microsatellites, will show a large genetic distance from other breeds primarily caused by founder effect, genetic drift and consequently greater loss of genetic diversity relative to other breeds, as emphasized by Hedrick (1999). Therefore, the inclusion of less-variable markers such has proteins can help increase the connection between statistical and biological significance. The fitting of the major groupings of breeds with historically known relationships, as well as the high bootstrap values obtained in this study by the joint use of three types of markers, confirms the statement made by Bruford *et al.* (2003) that different marker types (both coding and non-coding) should be used to assay variation in order to effectively document and conserve diversity within and between breeds.

Acknowledgements

Data were collected while E.G. Cothran was in the Department of Veterinary Science of the University of Kentucky. We thank the staff of the Equine Parentage Testing and Research lab of the Department of Veterinary Science of the University of Kentucky for their work in genotyping. We would like to thank Dra. Maria Portas from the SNC and Eng. João Ralão from the APSL for information provided regarding census data for the Garrano and Lusitano breeds, respectively. C. Luís was supported by a PhD grant (SFRH/BD/3318/2000) from the Portuguese Foundation for Science and Technology (FCT/MCT).

References

Aberle K.S., Hamann 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. *Animal Genetics* **35,** 270–7.

Andersson L. (1985) The estimation of blood group gene frequencies: a note on the allocation method. *Animal Blood Groups Biochemical Genetics* **16,** 1–7.

Andrade R.d' (1937) Les chevaux du Sorraia. *Comptes Rendues XII Congrès International de Zoologie*, 2368–70.

Andrade R.d' (1938) Garranos. *Boletim Pecuário* **6,** 281–92.

Andrade R.d' (1945) O Cavalo do Sorraia. *Boletim Pecuário* **XIII,** 1–13.

Andrade R.d' (1954) *Alredor del Caballo Español*. Colección de Estúdios, Lisboa, Portugal.

Andrade F.d' (1987) Origines et evolution des races chevalines Portuguaises. In: *Abstracts of the 38th Annual Meeting of the EAAP* Vol. II, pp. 1326–7. Lisbon, Portugal.

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,** 9–16.

Bort D.M. (2004) *La Ganadería Caballar en la Villa de Almonte. Introducción Histórica*. Colección Cuadernos de Almonte, Centro Cultural de la Villa, Ayuntamiento de Almonte, Huelva, Spain.

Bowling A.T. & Clark R.S. (1985) Blood group and protein polymorphism frequencies for seven breeds of horses in the United States. *Animal Blood Groups and Biochemical Genetics* **16,** 93–108.

Bowling A.T. & Ruvinsky A. (2000) *The Genetics of the Horse*. CABI Publishing, UK.

Bowling A.T. & Ryder O.A. (1987) Genetic studies of blood markers in Przewalski's horse. *Heredity* **78,** 75–80.

Braend M. (1973) Genetic variation in equine blood proteins. In: *Proceedings of the 3rd International Conference on Equine Infections Disease* (Ed. by J.T. Bryans, H. Gerber), pp. 394–406. Karger, Basel.

Braend M. & Johansen K.E. (1983) Haemoglobin types in Norwegian horses. *Animal Blood Groups and Biochemical Genetics* **14,** 305–7.

Breen M., Lindgren G., Binns M.M., Norman J., Irvin Z., Bell K., Sandberg K. & Ellegren H. (1997) Genetical and physical assignments of equine microsatellites - first integration of anchored markers in horse genome mapping. *Mammalian Genome* **8,** 267–73.

Bruford M.W., Bradley D.G. & Luikart G. (2003) Genetic analysis reveals complexity of livestock domestication. *Nature Reviews Genetics* **4,** 900–10.

Cañon 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. *Animal Genetics* **31,** 39–48.

Cavalli-Sforza L.L. & Edwards A.W.F. (1967) Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics* **19,** 233–257.

Cothran E.G. & Luís C. (2005) Genetic distance as a tool in the conservation of rare horse breeds. In: *Conservation Genetics of Endangered Horse Breeds* (Ed. by I. Bodo, L. Alderson & B. Langlois), pp. 55–71. Wageningen Academic Publishers, The Netherlands.

Cothran E.G., Santos S.A., Mazza M.C.M., Lear T.L. & Sereno J.R.B. (1998) Genetics of the Pantaneiro horse of the Pantanal region of Brazil. *Genetics and Molecular Biology* **21,** 343–9.

Dovč P., Sušnik S. & Snoj A. (2004) Experience from Lippizan horse and salmonid species endemic to the Adriatic river system. Examples for the application of molecular markers for preservation of biodiversity and management of animal genetic resources. *Journal of Biotechnology* **113,** 43–53.

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

FAO (1998) *Secondary Guidelines for Development of National Farm Animal Genetic Resources Management Plans. Management of Small Populations at Risk. Initiative for Domestic Animal Diversity*. FAO, Rome.

Felsenstein J. (1993) *PHYLIP (Phylogeny Inference Package) version 3.5c*. Department of Genetics, University of Washington, Seattle, USA.

Gonzaga P.G. (2004) *A History of the Horse: The Iberian Horse From Ice Age to Antiquity*. JA Allen & Co. Ltd, London, UK.

Guérin G., Bertaud M. & Amigues Y. (1994) Characterization of seven new horse microsatellites: *HMS1*, *HMS2*, *HMS3*, *HMS5*, *HMS6*, *HMS7*, and *HMS8*. *Animal Genetics* **25,** 62.

Gustincich S., Manfioletti G., Del Sal G. & Schneider C. (1991) A fast method for high-quality genomic DNA extraction from whole human blood. *BioTechniques* **11,** 298–300.

Hedrick P. (1999) Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution* **53,** 313–8.

Hendricks B.L. (1995) *International Encyclopedia of Horse Breeds*. University of Oklahoma Press, Norman, USA.

Henney P.J., Johnson E.L. & Cothran E.G. (1994) A new buffer system for acid PAGE typing of equine protease inhibitor. *Animal Genetics* **9,** 363–4.

Holm S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6,** 65–70.

Hughes S.S. (2000) *Strand Nucleic Acid Analysis Software ver. 1.2.90*. Available at http://www.vgl.ucdavis.edu/informatics/STRand. University of California, Davis, CA, USA.

Jansen T., Forster P., Levine M.A., Oelke H., Hurles M., Renfrew C., Weber J. & Olek K. (2002) Mitochondrial DNA and the origins of the domestic horse. *Proceedings of the National Academy of Sciences USA* **99,** 10905–10.

Juneja R.K., Gahna B. & Sandberg K. (1978) Genetic polymorphism of the Vitamin D binding protein and other post-albumin protein in horse serum. *Animal Blood Groups and Biochemical Genetics* **25,** 29–36.

Juras R., Cothran E.G. & Klimas R. (2003) Genetic analysis of three Lithuanian horse breeds. *Acta Agriculturae Scandinavica* **53,** 180–5.

Lewis P.O. & Zaykin D. (2000) *Genetic Data Analysis: Computer program for the analysis of allelic data, Version 1.0 (d15)*. Free program distributed by the authors over the internet from the GDA Home Page at http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php.

Loch S. (1986) *The Royal Horse of Europe: The Story of the Andalusian and Lusitano*. J A Allen & Co Ltd, UK.

Lopes M.S., Mendonça D., Cymbron T., Valera M., Costa-Ferreira J. & da Câmara Machado A. (2005) The Lusitano horse maternal lineage based on mitochondrial D-loop sequence variation. *Animal Genetics* **36,** 196–202.

Luís C., Bastos-Silveira C., Cothran E.G. & Oom M.M. (2002a) Variation in the mitochondrial control region sequence between the two maternal lines of the Sorraia horse breed. *Genetics and Molecular Biology* **25,** 309–11.

Luís C., Cothran E.G. & Oom M.M. (2002b) Microsatellites in Portuguese autochthonous horse breeds: usefulness for parentage testing. *Genetics and Molecular Biology* **25,** 131–4.

Luís C., Cothran E.G., Oom M.M. & Bailey E. (2005) Major histocompatibility complex locus DRA polymorphism in the endangered Sorraia horse and related breeds. *Journal of Animal Breeding Genetics* **122,** 69–72.

Luís C., Bastos-Silveira C., Cothran E.G. & Oom M.M. (2006) Iberian origins of New World horse breeds. *Journal of Heredity* **97,** 107–113.

Luís C., Bastos-Silveira C., Cothran E.G. & Oom M.M. (2006) A lost Sorraia maternal lineage found in the Lusitano horse bred. *Journal of Animal Breeding Genetics* **123,** 399–402.

Mantel N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* **27,** 209–20.

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.

Mirol P.M., Peral García P., Vega-Pla J.L. & Dulot F.N. (2002) Phylogenetic relationships of Argentinean Creole horses and other South American and Spanish breeds inferred from mitochondrial DNA sequences. *Animal Genetics* **33,** 356–63.

Montgomery G.W. & Sise J.A. (1990) Extraction of DNA from sheep white blood cells. *New Zealand Journal of Agriculture Research* **33,** 437–41.

Morais J., Oom M.M., Malta-Vacas J. & Luís C. (2005) Genetic structure of an endangered Portuguese semiferal pony breed, the Garrano. *Biochemical Genetics* **43,** 347–64.

Nei M. (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89,** 583–90.

Nei M., Maruyama T. & Chakraborty R. (1975) The bottleneck effect and genetic variability in populations. *Evolution* **29,** 1–10.

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

Nozawa K., Shotake T., Ito S. & Kawamoto Y. (1998) Phylogenetic relationships among Japanese native and alien horses estimated by protein polymorphisms. *Journal of Equine Science* **9,** 53–69.

Oom M.M. (1992) *O Cavalo Lusitano. Uma raça em recuperação*. Ph.D. dissertation, University of Lisbon, Lisbon, Portugal.

Oom M.M. & Cothran E.G. (1994) The genetic variation of an endangered breed: the Sorraia horse. *Animal Genetics* **25,** 15.

Oom M.M. & Cothran E.G. (1996) Preliminary study of genetic variation of a feral pony from Portugal: The Garrano. *Animal Genetics* **27,** 17–42.

Oom M.M., d'Andrade J.L. & Costa-Ferreira J. (2004) *Stud book da raça Sorraia*. Associação Internacional de Criadores Cavalo Sorraia, Portugal.

Page R.D.M. (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12,** 357–8.

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

Pollitt C.C. & Bell T.K. (1980) Protease inhibitor system in horses: classification and detection of a new allele. *Animal Blood Groups and Biochemical Genetics* **11,** 235–44.

Raymond M. & Rousset F. (1995) GENEPOP (Version 3.2a): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86,** 248–9.

Rice W.R. (1989) Analyzing tables of statistical tests. *Evolution* **43,** 223–5.

Rohlf F.J. (1993) *NTSYS-pc: numerical taxonomy and multivariate analysis system v1.80*. Exeter Software, New York, USA.

Royo L.J., Álvarez I., Beja-Pereira A., Molina A., Fernández I., Jordana J., Gómez E., Gutiérrez J.P. & Goyache F. (2005) The origins of the Iberian horses assessed via mitochondrial DNA. *Journal of Heredity* **96,** 663–9.

Sandberg K. (1974) Blood typing of the horse: current status and application to the identification problems. *Proceedings of the 1st World Congress of Genetics Applied to Livestock Production*, 253–65.

Solis A., Jugo B.M., Mériaux J.C., Iriondo M., Mazón L.I., Aguirre A.I., Vicario A. & Estomba A. (2005) Genetic diversity within and among four South European native horse breeds based on microsatellite DNA analysis: implications for conservation. *Journal of Heredity* **96,** 670–8.

Sponenberg D.P. (1996) The proliferation of horse breeds. In: *Horses Through Time* (Ed. by S. Olsen), pp. 153–74. Roberts Rinehart Publishers, Dublin, Ireland.

Stormont C. & Suzuki Y. (1964) Genetic polymorphism of blood groups in horses. *Genetics* **50,** 915–29.

Stormont C., Suzuki Y. & Rhode E.A. (1964) Serology of horse blood groups. *Cornell Veterinarian* **54,** 439–52.

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.

Supplementary Materials

The following supplementary material is available for this article online from http://www.blackwell-synergy.com/doi/full/10.1111/j.1365-2052.2006.01545.x

Table S1 Allele frequencies found in protein loci across the 33 horse breeds.

Table S2 Allele frequencies found in microsatellite loci across 33 horse breeds.

Table S3 Descriptive statistics (by locus) for markers.

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors.