Journal of Heredity 2011:102(6):753–758
doi:10.1093/jhered/esr091
Advance Access publication September 13, 2011

Genetic Analysis of Sicilian Autochthonous Horse Breeds Using Nuclear and Mitochondrial DNA Markers

Anna Maria Guastella, Antonio Zuccaro, Andrea Criscione, Donata Marletta, and Salvatore Bordonaro

From the DISPA Sezione di Scienze delle Produzioni Animali, Università degli Studi di Catania, via Valdisavoia, 5, 95123 Catania, Italy.

Address correspondence to D. Marletta at the address above, or e-mail: d.marletta@unict.it.

Genetic diversity and relationship among 3 Sicilian horse breeds were investigated using 16 microsatellite markers and a 397-bp length mitochondrial D-loop sequence. The analysis of autosomal DNA was performed on 191 horses (80 Siciliano [SIC], 61 Sanfratellano [SAN], and 50 Sicilian Oriental Purebred [SOP]). SIC and SAN breeds were notably higher in genetic variability than the SOP. Genetic distances and cluster analysis showed a close relationship between SIC and SAN breeds, as expected according to the breeds' history. Sequencing of hypervariable mitochondrial DNA region was performed on a subset of 60 mares (20 for each breed). Overall, 20 haplotypes with 31 polymorphic sites were identified. A higher haplotype diversity was detected in SIC and SAN breeds, with 13 and 11 haplotypes respectively, whereas only one haplotype was found in SOP. These were compared with 118 sequences from GenBank. BLAST showed that 17 of the 20 haplotypes had been reported previously in other breeds. One haplotype, found in SIC, traces back to a Bronze Age archaeological site (Inner Mongolia). The 3 Sicilian breeds are now rare, and 2 of them are officially endangered. Our results represent a valuable tool for management strategies as well as for conservation purposes.

Key words: *conservation, D-loop, genetic diversity, microsatellites, mitochondrial DNA, Sicilian horses*

Sicily detains an ancient tradition of horse breeding which goes back to Greek domination (600 BC). Until 16th century, 2 different types of horses were mainly reared in Sicily: the Asiatic and the North African. The crossbreed between these 2 genetic types gave origin to the "Sicilian breed" (Chiari 1901). Thanks to its beauty and harmony of shapes, it was considered an admirable example, thus at the end of 15th Century, Leonardo da Vinci chose a noble Sicilian breed horse (the Ciciliano horse of Galeazzo Sanseverino) as a model for his studies on horse body proportions. At the beginning of 20th century, Sicilian breed still maintained its excellent reputation and was one of the few recognized Italian horse breeds (Diffloth 1923).

Nowadays, about 30 000 horses are counted in the island but only the 8% is represented by autochthonous horses, named Siciliano (SIC), Sanfratellano (SAN), and Sicilian Oriental Purebred (SOP).

SIC (400 heads) is an heterogeneous and largely unmanaged population, reared in extensive system. This population, not yet officially recognized, represents one of the most authentic Sicilian equine genetic heritage. SAN (1600 heads) and SOP (150 heads) are considered at risk of extinction (for more details, see Zuccaro et al. 2008).

Genetic characterization of the endangered breeds is a compelling prerequisite for preservation and management strategies. Mitochondrial DNA (mtDNA) was used to investigate the origins and the genetic relationships of several horse breeds (Royo et al. 2005; Aberle et al. 2007; Glazewska 2010); combined with historical information, mtDNA was able to evaluate maternal inheritance and to identify founder mares in some breeds (Bowling et al. 2000; Hill et al. 2002).

Microsatellite markers (short tandem repeats) have been widely used in horse studies, also associated with mtDNA information (Kakoi et al. 2007; Pérez-Gutiérrez et al. 2008), in order to quantify genetic variation and to address conservation programmes (Marletta et al. 2006; Thirstrup et al. 2008). In this study, genetic variability has been assessed in 3 Sicilian indigenous horse breeds. The use of molecular information supplied by nuclear and mtDNA markers was aimed to provide suitable strategies for management and conservation.

Materials and Methods

Sampling and DNA Extraction

DNA was extracted from blood samples (10 ml in K3-EDTA tubes) collected all over Sicily from 191 minimally related Sicilian horses: 80 SIC, 61 SAN, and 50 SOP (Figure 1). A subset of 60 maternally unrelated mares (20 per each breed) was submitted to the mtDNA analysis.

Microsatellite Amplification and Analysis

A set of 16 STRs (HTG6, HTG10, VHL20, HTG7, HTG4, AHT5, AHT4, HMS3, HMS6, HMS7, HMS2, ASB2, HMS8, HTG15, HMS1, and HMS5) was used for genotyping by an ABI PRISM 377 genetic analyzer (Applied Biosystems, Foster City, CA). The main parameters of genetic variability were obtained by running the software Molkin v.3.0 (Gutiérrez et al. 2005), which was also implemented to assess the contribution of each breed to the total genetic diversity according to the methods proposed by Caballero and Toro (2002) and Petit et al. (1998).

In order to assess the distribution of genetic variability within and among breeds, FSTAT ver.2.9.3 software (Goudet 2001) was employed to estimate F-statistics (FIT, FIS, and FST), including population pair-wise FST; statistical significance was inferred by randomization-based methods and corrected by Bonferroni method to account for multiple comparisons. The model-based approach proposed by Falush et al. (2003) in the software STRUCTURE 2.2 was used to assess the genomic clustering of the sample. The admixture model was implemented to infer the population structure using no prior information (500 000 burn-ins, 500 000 iterations). The range of possible clusters (K) tested was from 1 to 10 (10 runs for each K).

MtDNA Amplification and Sequences Analysis

Mitochondrial control region (D-loop) was amplified according to Cozzi et al. (2004) using primers designed according to the horse sequence (Acc Num X79547). The amplicons, of 397-bp length, were purified and sequenced using the BigDye Terminator v1.1 Kit, on an ABI PRISM 310 DNA Sequencer (Applied Biosystems).

The aligned sequences were edited in MEGA 4 (Tamura et al. 2007) in order to identify the polymorphic sites and accomplish the BLAST search (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Arlequin ver.3.1 (Excoffier et al. 2005) was used to achieve the analysis of molecular variance (AMOVA) within the Sicilian breeds (Slatkin linearized indexes ΦST calculation) as well as to evaluate haplotype diversity (h), nucleotide diversity (πn), and the mean number of pair-wise differences (π) of Sicilian sample and 118 sequences (397-bp length) acquired from the literature.

Maximum parsimony network (Polzin and Daneschmand 2003) was constructed using the median-joining method (Bandelt et al. 1999). The analysis, performed by NETWORK ver.4.5.0.1 (http//:www.flexus-engineering.com), included sequences belonging to several European and geographically distant breeds.

Results

Microsatellites

Notable autosomal polymorphism (Table 1) and a relatively moderate gene flow among breeds (FST index 5.7%; $P < 0.001$) were observed. SIC was the most heterogeneous population in our sampling, SOP showed the lowest variability. FIS inferred per breed was always close to zero. A total of 26 breed-specific alleles (14 in SOP, 9 in SIC, and 3 in SAN) were detected, always at a frequency lower than 0.010 (data not shown). The pair-wise FST values were significant ($P < 0.01$), and a notably higher differentiation was observed between SAN and SOP (10.3%), followed by the between SIC and SOP FST (5.9%), whereas SIC and SAN accounted for the least genotypic diversity (2.6%). Reynolds distance (DR) between SAN and SOP (0.112) was the highest detected, whereas the closest breeds resulted SIC and SAN (0.031).

Clustering analysis inferred 3 ancestral subdivisions in the whole sample; genome structure of SAN and SOP are clearly referable to distinct clusters with a probability of 83.6% and 95.1% each; SIC grouped mainly in a third cluster with an estimated membership of 71.4%.

According to a conservation program, aimed at preserving the maximum amount of total gene diversity (GD_T) and the maximum amount of allele diversity (AR_T), each breed showed favorable contribution to the variability of the whole set, except for SAN when we took into account the method based on the rarefacted number of alleles proposed by Petit et al. (1998).

Table 1 Sample size, mean number of alleles per locus (A_n), allelic richness (A_r), observed and expected heterozygosity (H_o, H_e), F_{IS} values; within breed (GD_W), between breed (GD_B), and total contribution to the gene diversity (GD_T) according to Caballero and Toro (2002), within breed (AR_W), between breed (AR_B), and total contribution to the allelic richness (AR_T) according to Petit et al. (1998) in 3 Sicilian autochthonous horse breeds

Breeds	Size	A_n	A_r	H_o	H_e	F_{IS}	GD_W/GD_B	GD_T	AR_W/AR_B	AR_T
SIC	80	7.00	6.77	0.725	0.740	0.027	−2.975/1.384	−1.591	4.061/0.255	4.316
SAN	61	6.25	6.16	0.710	0.702	−0.003	0.497/−1.072	−0.575	−0.864/−0.313	−1.177
SOP	50	5.88	5.88	0.696	0.670	−0.029	1.884/−2.698	−0.814	−3.197/8.187	4.989

Mitochondrial Sequences

Twenty haplotypes and 31 polymorphic sites (7.8% on average) were identified (Table 2). Considering that only one haplotype (U) was found in SOP, this breed was excluded from the molecular diversity analysis. SIC and SAN counted 9 and 6 exclusive haplotypes, respectively. SAN shares 4 haplotypes with SIC (C, D, H, and O) and the haplotype U with SOP (Table 3). The analysis of molecular diversity, reported in Table 3, appointed SIC more variable than SAN: In terms of polymorphic sites and mean number of pairwise differences, SIC displayed the highest variability even when compared with the reference GenBank sequences.

AMOVA analysis, implemented only among the Sicilian breeds by using K2-parameter distances, yielded not significant variation between SAN and SIC ($\Phi ST = 0.046$; $P > 0.05$).

BLAST search showed that the 20 haplotypes found in the Sicilian sample overlapped with all the GenBank sequences but for C, P, and S. Haplotype B, found in SIC, was identified exclusively in a Bronze Age archaeological site (Inner Mongolia; DQ900929); this finding is noteworthy because in our knowledge this haplotype was not found in any other living horse. Moreover, 2 haplotypes (K and L) observed in SAN can be considered rare in a wide context because they were discovered only in few horse breeds. The median-joining network (Figure 2), constructed according to Jansen et al. (2002), showed a widespread distribution of Sicilian autochthonous haplotypes occupying both central and peripheral positions.

Discussion

Notwithstanding the old and prestigious equestrian tradition, in the last 50 years, the number of horses reared in Sicily has been suffering a rapid and severe decline, which mainly affected local resources. Despite the risk of extinction faces all the Sicilian native breeds, the genetic analysis revealed a notable molecular variability, comparable to that reported in Maremmano (Felicetti et al. 2010), Danish native (Thirstrup et al. 2008), and South European horses (Solis et al. 2005; Marletta et al. 2006). The SIC breed showed the highest genetic variation, probably because it is still an unmanaged population. On the contrary, SOP was the least variable breed, although 14 privates alleles were detected; this high diversity might be explained by: 1) the thorough selection of this breed, started on a group of imported Oriental horses (Balbo 1995) and 2) the employment of only Arab stallions.

The highly significant overall FST for the whole data set (5.7%) was lower than in other European population studies with values never below 8% (Glowatzki-Mullis et al. 2005; Luis et al. 2007) but higher than in 4 Basque-Navarrese semiferal native horse breeds (Solis et al. 2005). The significant estimates of between-population FST indicate a relatively low gene flow between SOP and the 2 other Sicilian breeds studied, probably due to reproductive isolation. The common breeding management which characterized SIC and SAN in the past might explain the low genetic differentiation (FST 2.6%) also

Table 2 Nucleotide substitutions in 20 D-loop haplotypes (397-bp fragment) detected in 3 Sicilian autochthonous horse breeds; haplotype nomenclature according to Jansen et al (2002) is reported in brackets

	Position of nucleotide substitutions relative to GenBank X79547 reference sequence																															
Haplotype	15494	15495	15496	15526	15534	15538	15540	15542	15574	15585	15596	15597	15598	15601	15602	15603	15604	15615	15616	15617	15635	15649	15650	15659	15666	15698	15703	15709	15718	15720	15726	15740
X79547	T	T	A	T	C	A	A	C	G	G	A	A	T	T	C	T	G	A	A	T	C	A	A	T	G	T	T	C	C	G	G	A
A (C1)		C													T					C				C						A		
B		C													T		A							C			C			A	A	G
C		C												C	T								G							A		
D		C						T		A		G			T						T		G		A		C			A		
E		C								A													G		A					A		
G		C				G					G				T								G					T		A		
H		C																					G		A					A		
I		C											C		T			G	G								C			A		
K		C								A					T								G							A		
L		C													T											A	C			A		
M		C													T								G							A		
N		C								A					T					C				C						A		
O		C		C			G			A					T							G							T	A		
P		C		C			G			A												G							T	A		
R (A4)	C	C	G		T											C						G								A		
S	C	C	G		T				A	A					T	C	A					G								A		
T (D2)	C	C	G		T										T	C						G								A		
U	C	C	G		T					A					T	C	A					G								A		
V	C	C	G		T					A					T	C						G								A		
Z (A6)		C																														

Table 3 Haplotype (h) and nucleotide diversity (π_n) with standard deviation, mean number of pair-wise differences (π), polymorphic sites (p.s.), and number of haplotypes (Hn) observed in SIC, SAN, and SOP (number of individuals in brackets) and 4 cosmopolitan horse breeds selected from GenBank database

Breed	h ± SD	π_n ± SD	π	p.s.	Hn	Sicilian breeds haplotypes/GenBank sequences
SOP	—	—	—	—	1	$U_{(20)}$
SIC	0.93 ± 0.04	0.029 ± 0.016	7.112	28	13	$A_{(2)}$ $B_{(1)}$ $C_{(1)}$ $D_{(5)}$ $E_{(3)}$ $G_{(1)}$ $H_{(1)}$ $I_{(1)}$ $O_{(1)}$ $R_{(1)}$ $S_{(1)}$ $T_{(1)}$ $Z_{(1)}$
SAN	0.92 ± 0.04	0.026 ± 0.014	6.503	20	11	$C_{(1)}$ $D_{(1)}$ $H_{(2)}$ $K_{(1)}$ $L_{(2)}$ $M_{(1)}$ $N_{(2)}$ $O_{(1)}$ $P_{(3)}$ $U_{(5)}$ $V_{(1)}$
THO	1.00 ± 0.02	0.026 ± 0.015	6.420	24	19	AF481305, AF481323
ARA	0.96 ± 0.01	0.027 ± 0.014	6.657	27	63	AF132568, AF132594
BAR	1.00 ± 0.04	0.026 ± 0.015	6.492	18	10	AJ413658, AJ413671
AND	1.00 ± 0.01	0.027 ± 0.015	6.615	23	26	AY997165, AY997168, AF516509, AF516511, AY805645, AY805664

Thoroughbred (THO), Arab (ARA), Barb (BAR), and Andalusian (AND).

strengthened by the short-term breed differentiation (DR distances), which showed SIC closely related to SAN.

According to both the approaches dealing with the conservation priorities, SIC and SOP highlighted the highest contributions in terms of within breed variability and between breed divergence. In the contest of conservation criteria, the methods proposed by Caballero and Toro (2002) and Petit et al. (1998) to preserve a suitable rate of diversity, pointed out SIC and SOP as the breeds of choice. This outcome probably refers to the absence of selection schemes in SIC population which led to a relevant heterozygosity and underlines the distinctiveness of SOP breed highlighted by autosomal markers and in agreement with the introduction of Arab stallions in the selection schemes of Sicilian Oriental horses.

Even if mitochondrial lineages are not a powerful tool to identify horse breeds, a detailed analysis of mtDNA haplotypes can provide relevant information about the history and the genetic variation of maternal lines in autochthonous populations (Aberle et al. 2007; Kakoi et al. 2007).

A moderate genetic base of Sicilian maternal lines, compared with other horse breeds (Cothran et al. 2005; Pérez-Gutiérrez et al. 2008), was observed and might be partially explained by the lack of polymorphism in SOP. However, it is noteworthy that only 4 of the 17 most frequent mtDNA haplotypes (C1, A4, A6, and D2) reported by Jansen et al. (2002) were identified in Sicilian autochthonous breeds (A, R, Z, and T, respectively), revealing a broad diversification of these breeds at the worldwide scale.

Overall, Sicilian haplotypes are shared or found closely related with all kind of horses: draught and riding breeds, Celtic ponies, even with those that are rather isolated (Pottoka, Icelandic pony) or geographically distant such as Tuva, Yumman, and Cheju. The presence of exclusive (C, P, and S) and rare (B) haplotypes suggests a possible geographic segregation of the SIC and SAN horses which probably spring from 4 common and ancestral maternal lineages. These lineages might be referable to the ancient Sicilian breed considering that they were found only in SIC and SAN so far.

The unique haplotype (U) identified in SOP corresponds to the haplotype A16 belonging to the "Dafina" matrilineal line founder of the Keilan el Krush strain (Bowling et al. 2000) and overlaps with GenBank sequences belonging

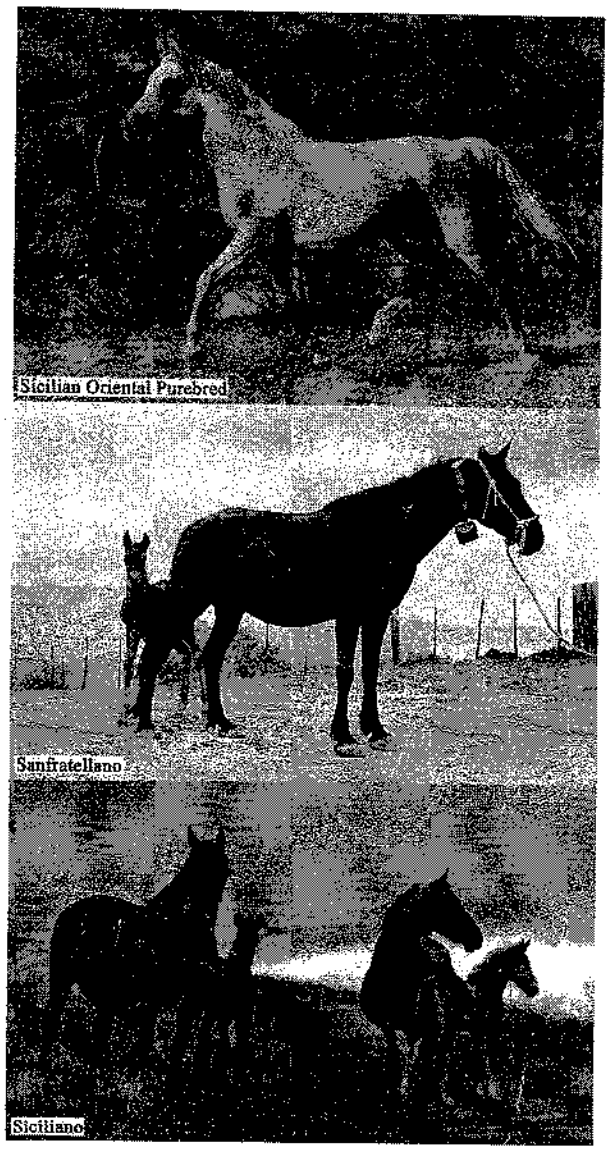

Figure 1. The 3 Sicilian native horse breeds analyzed.

European ponies; haplotypes R and Z match with clusters A4 and A6, respectively; finally, haplotype T corresponds to cluster D2. Throughout the results shown, a concrete association of Sicilian haplotypes to a geographic area or specific breed must be rejected. These outcomes obtained from nuclear (STRs) and mitochondrial (D-loop) markers are consistent and in agreement with historical backgrounds. Both proved the clear genetic differentiation between SOP and the 2 other Sicilian autochthonous breeds, confirming the accuracy of the Stud-Book genealogies.

The current SIC and SAN breeds seem to originate from a unique horse type, the ancient Sicilian breed, as consequence of different management strategies occurred during Centuries. This could explain why SIC and SAN have never been mentioned, as different genetic entities, before the half of the 20th century. The Sicilian endangered horses should be maintained and preserved even though diverse approaches seem to suggest different priorities. The presence of peculiar and rare female lineages, which lays emphasis on the conservation of these breeds as notable reservoirs of genetic biodiversity, needs to be taken into account for improving management strategies.

Funding

"Regione Siciliana—Assessorato Regionale delle Risorse Agricole e Alimentari" and "Istituto di Incremento Ippico per la Sicilia."

References

Aberle KS, Hamann H, Drögemüller C, Distl O. 2007. Phylogenetic relationships of German heavy draught horse breeds inferred from mitochondrial DNA D-loop variation. J Anim Breed Genet. 124:94–100.

Balbo SM. 1995. L'influenza dell'Arabo-Orientale sul cavallo Siciliano. In: Savier M, editor. L'ASIL Arabo—Il cavallo nobile d'Arabia. Spoleto Perugia (Italy): R&R Editrice. p. 169–175.

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 16:37–48.

Bowling AT, Del Valle A, Bowling M. 2000. A pedigree-based study of mitochondrial D-loop DNA sequence variation among Arabian horses. Anim Genet. 31:1–7.

Caballero A, Toro MA. 2002. Analysis of genetic diversity for the management of conserved subdivided populations. Conserv Genet. 3:289–299.

Chiari E. 1901. Trattato di Ippologia. Torino (Italy): UTET.

Cothran EG, Juras R, Macijauskiene V. 2005. Mitochondrial DNA D-loop sequence variation among 5 maternal lines of the Zemaitukai horse breed. Genet Mol Biol. 28:677–681.

Cozzi MC, Strillacci MG, Valiati P, Bighignoli B, Cancedda M, Zanotti M. 2004. Mitochondrial D-loop sequence variation among Italian horse breeds. Genet Sel Evol. 36:663–672.

Diffloth P. 1923. Italie: Races Italiennes. In: Wery G, editor. Encyclopédie Agricole. Zootechnie. Races Chavalines. 5th ed. Paris: Librairie J.B. Bailliere et Fils. p. 158–159.

Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online. 1:47–50.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 164:1567–1587.

Felicetti M, Lopes S, Verini-Supplizi A, da Ca?mara Machado A, Silvestrelli M, Mendonc D, Distl O. 2010. Genetic diversity in the Maremmano horse and its relationship with other European horse breeds. Anim Genet. 41(Suppl 2):53–55.

Glazewska I. 2010. Speculations on the origin of the Arabian horse breed. Livestock Sci. 129:49–55.

Glowatzki-Mullis ML, Muntwyler J, Pfister W, Marti E, Rieder S, Poncet PA, Gaillard C. 2005. Genetic diversity among horse populations with a special focus on the Franches-Montagnes breed. Anim Genet. 37:33–39.

Goudet J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3) [Internet]. [cited 2011 Jun]. Available from: URL http://www2.unil.ch/popgen/softwares/fstat.htm.

Gutiérrez JP, Royo LJ, Álvarez I, Goyache F. 2005. MolKin v2.0: a computer program for genetic analysis of populations using molecular coancestry information. J Hered. 96:718–721.

Hill EW, Bradley DG, Al-Barody M, Ertugrul O, Splan RK, Zakharov I, Cunningham EP. 2002. History and integrity of Thoroughbred dam lines revealed in equine mtDNA variation. Anim Genet. 33:287–294.

Jansen T, Forster P, Levine MA, Oelke H, Hurles M, Renfrew C, Weber J, Olek K. 2002. Mitochondrial DNA and the origins of the domestic horse. Proc Natl Acad Sci U S A. 99:10905–10910.

Kakoi H, Tozaki T, Gawahara H. 2007. Molecular analysis using mitochondrial DNA and microsatellites to infer the formation process of Japanese native horse population. Biochem Genet. 45:375–393.

Luis C, Cothran EG, Oom MM. 2007. Inbreeding and genetic structure in the endangered Sorraia horse breed: implication for its conservation and management. J Hered. 98:232–237.

Marletta D, Tupac-Yupanqui I, Bordonaro S, Garcia D, Guastella AM, Criscione A, Cañón J, Dunner S. 2006. Analysis of genetic diversity and the determination of relationships among western Mediterranean horse breeds using microsatellite markers. J Anim Breed Genet. 123:315–325.

Pérez-Gutiérrez LM, De La Peña A, Arena P. 2008. Genetic analysis of Hispano-Breton horse. Anim Genet. 39:506–514.

Petit RJ, El Mousadik A, Ponds O. 1998. Identifying populations for conservation on the basis of genetic markers. Conserv Biol. 12:844–855.

Polzin T, Daneschmand SV. 2003. On Steiner trees and minimum spanning trees in hypergraphs. Oper Res Lett. 31:12–20.

Royo LJ, Álvarez I, Beja-Pereira A, Molina A, Fernández I, Jordana J, Gómez E, Gutiérrez JP, Goyache F. 2005. The origins of the Iberian horses assessed via mitochondrial DNA. J Hered. 96:663–669.

Solis A, Jugo BM, Mériaux JC, Iriondo M, Mazón LI, Aguirre AI, Vicario A, Estomba A. 2005. Genetic diversity within and among four South European native horse breed on microsatellite DNA analysis: applications for conservation. J Hered. 96:670–678.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 24:1596–1599.

Thirstrup JP, Pertoldi C, Loeschcke V. 2008. Genetic analysis, breed assignment and conservation priorities of three native Danish horse breeds. Anim Genet. 39:496–505.

Zuccaro A, Bordonaro S, Criscione A, Guastella AM, Perrotta G, Blasi M, D'Urso G, Marletta D. 2008. Genetic diversity and admixture analysis of Sanfratellano and three other Italian horse breeds assessed by microsatellite markers. Animal. 7:991–998.

Received March 7, 2011; Revised July 25, 2011; Accepted August 5, 2011

Corresponding Editor: Ernest Bailey