GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Mitochondrial DNA lineages of Italian Giara and Sarcidano horses

L. Morelli[1†], A. Useli[1], D. Sanna[1], M. Barbato[1,2], D. Contu[3], M. Pala[4,5], M. Cancedda[6] and P. Francalacci[1]

[1]Dipartimento di Scienze della Natura e del Territorio,
Unità di Zoologia Archeozoologia e Genetica, Università di Sassari,
Sassari, Italy
[2]Cardiff School of Biosciences, Cardiff University, Cardiff, UK
[3]Laboratorio di Immunogenetica, Ospedale Microcitemico,
Cagliari, Italy
[4]Dipartimento di Biologia e Biotecnologie "L. Spallanzani" Università di Pavia,
Pavia, Italy
[5]School of Applied Sciences, University of Huddersfield, Queensgate,
Huddersfield, UK
[6]Dipartimento di Medicina Veterinaria, Università di Sassari, Sassari, Italy

[†]This paper is dedicated to the memory of Laura Morelli who prematurely passed away before the publication.
Corresponding author: A. Useli
E-mail: auseli@uniss.it

Genet. Mol. Res. 13 (4): 8241-8257 (2014)
Received December 18, 2012
Accepted June 14, 2013
Published October 20, 2014
DOI http://dx.doi.org/10.4238/2014.October.20.1

ABSTRACT. Giara and Sarcidano are 2 of the 15 extant native Italian horse breeds with limited dispersal capability that originated from a larger number of individuals. The 2 breeds live in two distinct isolated locations on the island of Sardinia. To determine the genetic structure and evolutionary history of these 2 Sardinian breeds, the first hypervariable segment of the mitochondrial DNA (mtDNA) was sequenced and analyzed in 40 Giara and Sarcidano horses and compared with publicly available mtDNA data from 43 Old World breeds. Four

different analyses, including genetic distance, analysis of molecular variance, haplotype sharing, and clustering methods, were used to study the genetic relationships between the Sardinian and other horse breeds. The analyses yielded similar results, and the F_{ST} values indicated that a high percentage of the total genetic variation was explained by between-breed differences. Consistent with their distinct phenotypes and geographic isolation, the two Sardinian breeds were shown to consist of 2 distinct gene pools that had no gene flow between them. Giara horses were clearly separated from the other breeds examined and showed traces of ancient separation from horses of other breeds that share the same mitochondrial lineage. On the other hand, the data from the Sarcidano horses fit well with variation among breeds from the Iberian Peninsula and North-West Europe: genetic relationships among Sarcidano and the other breeds are consistent with the documented history of this breed.

Key words: Mitochondrial DNA; Giara horse; Sarcidano horse; Haplogroup attribution; Domestication

INTRODUCTION

Island environments enable the persistence of relic varieties of species or breeds. Among the 15 horse breeds officially recognized by the Italian Registry of Autochthonous Equine Breeds, 2 breeds, which are inbred and phenotypically distinctive, are from the island of Sardinia (Figure 1).

The Giara horses take their name by the basalt plateau in the central/south region of Sardinia where they live in the wild. The Giara plateau extends over an area of 45 km^2 at an altitude of about 500-600 m above sea level. The steep mountain slopes limit connections with the surrounding valleys and prevent migration of the horses. The average height of Giara horse approaches that of a pony, but it is considered a miniature horse. The bioecological features of this population make it a rich livestock heritage to safeguard and a guarantee of protection for the natural environment where it lives (Gratani, 1980).

The Sarcidano horses are concentrated on a single farm in Laconi (in central-western Sardinia). The history of how the breed was introduced onto the island is controversial and mostly unknown; nonofficial records suggest a descent from the ancient Spanish horse, an ancestor of the Andalusian breed.

At the end of the 18th century, Cetti (1774) described the presence on Sardinia of 3 different types of horses, the "*selvaticus*" (wild), the "*vulgar*" (common or ordinary), and the "*di razza*" (thoroughbred) horse. Cetti's description of the height and temperament of "*selvaticus*" is consistent with the descriptions of the present-day Giara horse. On the other hand, the "*vulgar*" Cetti's horse description fits well with that of the present population of the Sarcidano breed in terms of character, phenotype, and work attitude. At present, there are 481 registered Giara and 108 Sarcidano horses (http://www.anagrafeequidi.it/index.php?id=217).

The aim of this study was to shed light on the genetic structure of the Giara and Sarcidano horses in a global context to depict their genetic relationships with other present-day

breeds. To achieve this goal, we compared a sample of the 2 Sardinian breeds with a large number of horses genotyped to date. For this comparison, we selected a 247-bp long internal portion of the hyper-variable region segment I (HVS-I) of the control displacement loop (d-loop) of mitochondrial DNA (mtDNA). This choice was made for three reasons: 1) sequence variation in these mtDNA regions is solely generated by the sequential accumulation of new mutations along radiating maternal lineages; 2) maternal lineages are presumably more stable because of the practice to move stallions for reproduction and breed improvement, so mtDNA sequences should present breed-specific motifs that relate our samples to a geographical origin; and 3) a large number of d-loop sequences from many horse breeds is available in GenBank.

Figure 1. Range of distribution in Giara and Sarcidano breeds.

To date, only limited data (Cozzi et al., 2004; Achilli et al., 2012) on mtDNA variation in Sardinian horse breeds are available. A comparison of a limited number of Sardinian horses with other Italian breeds suggests a reduced relationship with the other Italian populations (Cozzi et al., 2004).

MATERIAL AND METHODS

Materials

Using standard procedures, total DNA was extracted from peripheral blood samples of 24 horses of the Giara breed and 16 horses of the Sarcidano breed. The Giara and Sarcidano

horses had been bred in semiferal conditions, and animals were randomly selected by capture.

The sequences produced in this study were pooled with 5 sequences from Giara (GRH1-5; GenBank accessions AY462426-AY462430) and 5 from Sarcidano (SRH1-5; AY462451-AY462455) breeds previously reported by Cozzi et al. (2004), and with 2 additional HVS-I sequences identified in the complete mtDNA genome sequences (Gia01, JN398411 and Gia02, JN398407) reported by Achilli et al. (2012). The final sample set used in the analyses was obtained from 31 Giara and 21 Sarcidano horses.

In addition, all the 150 complete mtDNA genome sequences available from literature [NC_001640 (Xu and Arnason, 1994); EF597512-14 (Xu et al., 2007); AP012267-70 (Goto et al., 2011); EU939445 (Jiang et al., 2011); HQ439441-500 (Lippold et al., 2011); and JN398377-457 (Achilli et al., 2012)] were used to obtain more reliable and informative HVS-I patterns of sites defining haplogroups.

Finally, we produced a dataset of typical regional breeds by selecting 1192 HVS-I horse sequences from the literature for which frequency population data were available and which were reported for at least 15 individuals. Taking into account the haplotype frequencies, we obtained 1232 HVS-I samples belonging to 45 breeds (including Giara and Sarcidano) representing 6 geographic Old World macroareas: the Iberian Peninsula (N = 220), Central Europe (N = 400), Northwest Africa (N = 40), the Arabian Peninsula (N = 70), Central Eurasia (N = 113), and the Far East (N = 389).

Methods

The HVS-I of the d-loop region was amplified by the polymerase chain reaction (PCR) using two primers from a published horse sequence (GenBank accession No. X79547): forward 5'-AACGTTTCCTCCCAAGGACT-3' and reverse 5'-GTAGTTGGGAGGGTTGCTGA-3' (Ishida et al., 1994; Xu and Arnason, 1994). The amplicon obtained was a 397-bp fragment included between the $tRNA^{Pro}$ gene and the large central conserved sequence block from nucleotide position 15382-15778. The PCR products were purified by using ExoSAP-IT (USB Corporation) and sequenced using the BigDye Terminator Kit (Applied Biosystems) on an ABI PRISM 377 DNA Sequencer equipped with the Sequencing Analysis and Sequence Navigator programs (Applied Biosystems). Sequence alignments were performed with the BioEdit 7.0.5.2 software (Hall, 1999).

Intra- and interpopulation level methods were conducted with the Arlequin 3.5. software (Excoffier et al., 2005) (http://cmpg.unibe.ch/software/arlequin3): intrapopulation level variation was estimated with both standard (gene diversity; Nei, 1987) and molecular indices as pairwise differences (Tajima, 1993) and nucleotide diversity (Tajima, 1983; Nei, 1987). Population genetic structure was obtained by hierarchical analysis of the total variance subdivided in percentage of variance within the breeds, among breeds within groups and among groups by using molecular analysis of variance (AMOVA) (Excoffier et al., 1992) taking into account the number of mutations between molecular haplotypes. In both cases, the F-statistic was set at a significance level of 0.05, obtained by 10,000 permutation tests. The matrix of interpopulation pairwise distances (Tajima, 1993, Arlequin software) was summarized in two dimensions by using multidimensional scaling (MDS) analysis as implemented by the STATISTICA ’99 software (StatSoft, Tulsa, OK, USA) and plotted on an MS Excel graphic.

Haplogroups attribution was performed following the nomenclature rules of Achilli et al. (2012). To increase the power of imputation of the HVS-I sequences, we pooled together all the available horse complete mtDNA genome sequences. Polymorphic sites occurring among the total of 150 sequences were exported as an Excel spreadsheet and the haplotypes were organized in haplogroups following a hierarchical and parsimonious order, and the haplogroup name was assigned to the unclassified data. Three of them were eliminated because their polymorphisms were not consistent with phylogeny as previously observed by Lippold et al. (2011) and Achilli et al. (2012). The mutational pattern was dissected in order to define the following variables:

- variation associated univocally with the haplogroup; this variation is due to the most informative mutations since they are present in all of the haplotypes of the same haplogroup and absent in other haplogroups;
- variation associated univocally with the super-haplogroup; this variation is due to mutations that allow allocation to a unique clade represented by individuals that share the same mutation because it is ancestral for all the haplogroups that compose it;
- variation due to mutations occurring in a unique haplogroup but not in all of the haplotypes of this haplogroup;
- variation due to informative mutations in the allelic association;
- variation due to poorly informative mutations that are the result of recurrence or reversion.

The haplotypic pattern of the HVS-I region was extracted and a probabilistic algorithm defining haplogroup attribution was generated when only the HVS-I sequence was available. Haplotype variation and haplotype sharing into the haplogroups were evaluated by clustering in the Network program 5.0.0, and default parameters were used for obtaining the median-joining network trees (Bandelt et al., 1999). A weight of 0 were assigned to mutations classified as recurrent. In addition, mismatch distribution of the number of pairwise differences between haplotypes among haplogroups and associated demographic parameters including Harpending's raggedness index (r), Tajima's D (D), and Fu's (F_S) statistics were calculated using the Arlequin 3.5 software.

RESULTS

Diversity indices

The 52 sequences obtained from the Sardinian horses consisted of 29 different haplotypes on the segment ranging from nucleotide position 15494 to 15740 (Table 1).

We calculated the diversity indices from 43 native breeds scattered in the Old World (see Table 2, for details).

We compared diversity indices from Old World breeds to those obtained from the Sardinian horses: both the haplotype diversity values of the Giara (0.847 ± 0.053) and Sarcidano (0.905 ± 0.047) breeds were higher than the average estimates for the other 43 breeds, but only the Sarcidano breed was above the median of the distribution (Table 3). Other molecular diversity indices that were also considered (see Table 3) showed that the heterogeneity of the Sarcidano sample (mean number of pairwise differences, 6.738 ± 3.309 and nucleotide diversity, 0.027 ± 0.015) was comparable with the highest values reported for the other breeds, whereas Giara horses showed a lower molecular diversity (mean number of pairwise differences, 3.933 ± 2.025 and nucleotide diversity, 0.016 ± 0.009).

Table 1. Mitochondrial control region sequence variation in Sardinian horses.

Samples	Giara	Sarcidano	Haplogroup	15494	15495	15496	15505	15507	15526	15534	15538	15540	15542	15569	15575	15576	15585	15589	15597	15599	15601	15602	15603	15604	15617	15618	15619	15623	15627	15633	15634	15635	15639	15644	15649	15650	15659	15665	15666	15667	15672	15678	15680	15692	15694	15695A	15703	15704	15708	15709	15712	15716	15720	15732	15740
Reference X79547	#	#		T	T	A	T	C	T	C	A	A	C	T	G	C	G	T	A	T	T	C	T	G	T	A	A	T	C	A	T	C	T	T	A	A	T	T	G	A	G	C	C	A	C	-	T	C	A	C	A	C	G	C	A
JAR38	1	0	B	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	.	.	.	-	-	.	.	.	.	.	.	A	.	.
JAR69	1	0	B	.	C	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	A	.	.	.	G	.	.	A	.	.	.	.	T	.	-	-	.	.	.	.	.	.	A	.	.
JAR90	1	0	E	.	C	.	.	.	.	.	.	.	T	.	.	.	A	.	G	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	A	.	.	.	.	.	.	-	-	.	.	.	.	.	.	A	.	.
JAR40, JAR64, JAR70, JAR72, JAR87, JAR93, JAR95, GRH1, GRH3, GRH5	10	0	G	.	C	.	.	.	.	.	.	.	T	.	.	.	A	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	G	.	.	A	.	.	.	.	.	.	-	C	.	.	.	.	.	A	.	.
JAR55	1	0	G	.	C	.	.	.	.	.	.	.	T	.	.	.	A	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	G	.	.	A	.	.	.	.	.	.	-	C	.	.	A	.	.	A	.	.
JAR58	1	0	G	.	C	.	.	.	.	.	.	.	T	.	.	.	A	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	G	.	.	A	.	.	.	.	.	.	-	C	A	G	.	.	.	A	.	.
JAR59	1	0	G	.	C	.	.	.	.	.	.	.	T	.	.	.	A	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	G	.	.	A	.	.	.	.	.	.	-	C	.	.	G	.	.	A	.	.
JAR67	1	0	G	.	C	.	G	.	.	.	.	.	T	.	.	.	A	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	G	.	.	A	.	.	.	.	.	.	-	C	.	.	.	G	.	A	G	.
JAR71	1	0	G	.	C	.	.	.	.	.	.	.	T	.	.	.	A	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	G	.	.	A	.	.	.	.	.	.	-	C	.	.	.	.	.	A	.	C
JAR77	1	0	G	.	C	.	.	T	.	.	.	.	T	G	.	.	A	.	G	.	.	T	.	.	.	.	.	G	.	.	.	T	A	.	.	G	.	.	A	T	.	.	A	.	.	-	C	.	.	.	.	.	A	.	T
JAR92	1	0	G	.	C	.	.	.	.	.	.	.	T	.	.	.	A	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	G	.	.	A	C	.	.	.	.	.	-	C	.	.	.	.	.	A	.	.
JAR81	1	0	G	.	C	.	.	.	.	.	.	.	T	.	.	.	A	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T	.	A	.	G	.	.	A	.	.	.	.	.	.	-	C	.	.	.	.	.	A	.	.
GIA02	1	0	G	.	C	.	.	.	.	.	.	.	T	.	.	.	A	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	G	.	.	A	.	.	.	.	.	.	-	C	.	.	C	.	.	A	.	.
JAR34	1	0	G	.	C	.	.	.	.	.	.	.	T	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	.	.	.	-	C	.	.	.	.	.	A	.	.
JAR39, JAR41, JAR45,GRH2, GRH4	5	0	G	.	C	.	.	.	.	.	.	.	T	.	.	.	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	G	.	.	A	.	.	.	.	.	.	A	C	.	.	.	.	.	A	.	.
GIA01	1	0	G	.	C	.	.	.	.	.	.	.	T	.	.	.	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	G	.	.	A	.	.	.	.	.	.	-	C	.	.	C	.	.	A	.	.
SAR22	0	1	G	.	C	.	.	.	.	.	.	.	T	.	.	.	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	.	.	.	-	C	.	.	.	.	.	A	.	.
SAR10	0	1	G	.	C	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	.	.	.	-	.	.	.	.	.	.	A	.	.
SAR9	0	1	I	.	C	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	-	.	.	.	T	.	.	A	.	.
SAR1, SAR6, SAR26, SAR27, SRH3	0	5	I	.	C	.	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	-	.	.	.	T	.	.	A	.	.
SRH1	0	1	I	.	C	.	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	-	.	.	.	T	.	.	A	.	.
SAR24	0	1	I	.	C	.	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	-	.	.	.	G	.	A	A	.	.
SAR28	0	1	I	.	C	.	G	.	.	.	G	.	.	.	.	.	A	G	.	G	.	T	G	.	.	.	.	.	.	.	.	.	.	A	.	G	.	.	.	.	.	.	.	.	T	-	.	.	.	.	.	.	A	.	.
JAR35, JAR57, SAR13, SAR15, SAR20, SRH4, SRH5	2	5	L	C	C	G	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	A	.	.
SAR18	0	1	L	C	C	G	.	.	T	.	.	.	.	.	.	.	.	-	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	A	.	.
SAR12	0	1	L	C	C	G	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	T	-	.	.	.	.	.	.	A	.	.
SRH2	0	1	L	C	C	G	.	.	T	.	G	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	A	.	.
SAR17	0	1	M	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	T	.	.	C	C	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	A	.	.
SAR4	0	1	M	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	A	.	.

Polymorphic sites based on the same reference sequence are given above each column. Variable nucleotides are indicated. All samples have been assigned to one of the main haplogroups and grouped by the occurrence of haplogroup informative nucleotide. JAR = Giara horses from the present study; GRH = Giara horses by Cozzi et al., 2004; GIA = Giara horses by Achilli et al., 2012; SAR = Sarcidano horses from the present study; SRH = Sarcidano horses by Cozzi et al., 2004. 247-bp segment of the control region from 31 Giara and 21 Sarcidano horses aligned to the mtDNA reference X79547 (Xu and Arnason, 1994).

Table 2. Forty-five Old World native breeds.

Macroarea	Breed	Code	References and GenBank accessions
North-western Africa	Barb	BAR	Jansen et al., 2002; EF686021-45
Iberian Peninsula	Andalusian	AND	Mirol et al., 2002; Jansen et al., 2002; Royo et al., 2005; Luís et al., 2006a
	Asturcòn	AST	Mirol et al., 2002; Royo et al., 2005; HQ827083-90
	Caballo de Corro	CCO	Royo et al., 2005; HQ827099-103
	Garrano	GAR	Royo et al., 2005; Luís et al., 2006a; AY246231-4
	Jaca Navarra	JAN	HQ827104-HQ827118
	Losino	LOS	Mirol et al., 2002; Royo et al., 2005; HQ827119-29
	Lusitano	LUS	Jansen et al., 2002; Luís et al., 2006a; AY246242-7
	Marismeno	MAR	Royo et al., 2005; HQ827136-45
	Pottoka	POT	Mirol et al., 2002; Royo et al., 2005; HQ827156-61
	Sorraia	SOR	Jansen et al., 2002; Luís et al., 2006a,b; HQ827162-3; AY246259-65
Central Europe	Exmoor	EXM	Jansen et al., 2002; AY246219-24
	Fell	FEL	Bower et al., 2011
	Giara	GIA	Present study; Cozzi et al., 2004; Achilli et al., 2012
	Irish Draught	IRD	McGahern et al., 2006a
	Kerry Bog Pony	KEB	McGahern et al., 2006a
	Percheron	PER	Kakoi et al., 2007
	Rhineland Heavy Draft	RHD	Jansen et al., 2002
	Sanfratellano	SAN	Zuccaro et al., 2009; Guastella et al., 2011
	Sarcidano	SAR	Present study; Cozzi et al., 2004
	Scottish Highland	SCH	Jansen et al., 2002; Bower et al., 2010
	Senner	SEN	Jansen et al., 2002
	Shetland	SHE	Hill et al., 2002; Jansen et al., 2002; Bower et al., 2011; AY246253-8
	Sicilian Indigenous	SII	Zuccaro et al., 2009; Guastella et al., 2011
	Sicilian Oriental Purebred	SOP	Zuccaro et al., 2009; Guastella et al., 2011
Arabian Peninsula	Arabian	ARA	Bowling et al., 2000; Mirol et al., 2002; Jansen et al., 2002; AY246180-5
Central Eurasia	Akhal-Teke	AKT	McGahern et al., 2006b; AY246174-9
	Anatolian	ANA	Hill et al., 2002
	Kazahk	KAZ	Lei et al., 2009
	Mesenskay	MES	McGahern et al., 2006b
	Vyayskaya	VYA	McGahern et al., 2006b
	Yakut	YAK	McGahern et al., 2006b
Far East	Baise	BAI	GQ203128-GQ203143; GQ222059-60
	Cheju	CHE	Yang et al., 2002; AY246201-8
	Debao	DEB	EU826536; FJ392562-80; GQ203125-7
	Guanzhong	GUA	Lei et al., 2009
	Guizhou	GUI	Lei et al., 2009
	Mongolian	MON	Jansen et al., 2002; McGahern et al., 2006b; Kakoi et al., 2007
	NingQiang	NIN	Lei et al., 2009
	Tibetan	TIB	DQ986464-79
	Hokkaido	HOK	Kakoi et al., 2007
	Misaki	MIS	Kakoi et al., 2007
	Taishu	TAI	Kakoi et al., 2007
	Tokara	TOK	Kakoi et al., 2007
	Yonaguni	YON	Kakoi et al., 2007

Genetic structure

We used AMOVA on the basis of the pairwise difference distance method (Excoffier et al., 1992; Weir, 1996) to determine the genetic structure of the group composed by the Giara and the Sarcidano populations. A relevant and significant percentage of interpopulation variation (45.1%) was detected when compared with the intrapopulation variation (54.9%; $P < 10^{-5}$). Therefore, taking into account that in all of the 45 worldwide diffused breeds analyzed, 25% ($P < 10^{-5}$) of the total variation is allocated to the among-breeds source, we can confidently predict the presence of a genetic barrier between the two Sardinian breeds. The same value

was obtained after grouping into macrogeographical areas as reported in column 1 of Table 2, and no variance was attributable to the differences between groups. Just a small, but significant, amount of variation was related to the east (Arabian Peninsula, Central Eurasia, Far East) and west (Northwest Africa, Iberian Peninsula, Central Europe) groupings (2.89%; P = 0.01).

Table 3. Intrapopulation level variation of 45 Old World native breeds.

Breed	Individuals	Haplotypes	Polymorphic sites	Sum of square frequencies	Haplotype diversity		Pairwise differences		Nucleotide diversity	
					h	SD	MNPD	SD	π	SD
BAR	40	14	23	0.165	0.856	0.040	4.659	2.332	0.019	0.010
AND	30	15	22	0.104	0.926	0.026	5.400	2.677	0.022	0.012
AST	21	9	21	0.134	0.910	0.035	7.381	3.596	0.030	0.016
CCO	19	4	10	0.357	0.678	0.088	3.719	1.965	0.015	0.009
GAR	18	14	20	0.080	0.974	0.025	6.327	3.147	0.026	0.014
JAN	15	14	22	0.076	0.991	0.028	6.381	3.203	0.026	0.015
LOS	23	15	24	0.081	0.961	0.022	6.719	3.288	0.027	0.015
LUS	21	10	15	0.147	0.895	0.039	5.943	2.953	0.024	0.013
MAR	22	9	15	0.169	0.870	0.044	5.931	2.942	0.024	0.013
POT	21	18	26	0.066	0.981	0.023	6.600	3.247	0.027	0.015
SOR	30	5	12	0.422	0.598	0.059	2.232	1.265	0.009	0.006
EXM	18	6	20	0.247	0.797	0.066	5.209	2.644	0.021	0.012
FEL	17	8	19	0.177	0.875	0.053	5.500	2.783	0.022	0.013
GIA	31	15	31	0.180	0.847	0.053	3.933	2.025	0.016	0.009
IRD	59	28	31	0.070	0.946	0.017	6.373	3.063	0.026	0.014
KEB	39	17	26	0.090	0.934	0.020	5.614	2.753	0.023	0.012
PER	15	3	10	0.662	0.362	0.145	2.800	1.566	0.011	0.007
RHD	24	15	24	0.118	0.920	0.040	6.754	3.298	0.027	0.015
SAN	20	11	20	0.130	0.916	0.041	6.295	3.117	0.025	0.014
SAR	21	13	26	0.138	0.905	0.047	6.738	3.309	0.027	0.015
SCH	31	16	26	0.086	0.944	0.021	6.185	3.021	0.025	0.014
SEN	19	2	1	0.900	0.105	0.092	0.105	0.183	0.000	0.001
SHE	66	15	26	0.134	0.880	0.023	6.441	3.088	0.026	0.014
SII	20	13	30	0.120	0.926	0.043	7.132	3.492	0.029	0.016
SOP	20	1	0	1.000	0.000	0.000	0.000	0.000	0.000	0.000
ARA	70	37	37	0.042	0.972	0.007	6.231	2.995	0.025	0.013
AKT	24	14	27	0.108	0.931	0.033	6.859	3.345	0.028	0.015
ANA	15	11	17	0.102	0.962	0.034	5.333	2.727	0.022	0.012
KAZ	18	16	31	0.068	0.987	0.023	6.131	3.059	0.025	0.014
MES	18	11	26	0.136	0.915	0.050	5.549	2.797	0.022	0.013
VYA	18	10	17	0.124	0.928	0.037	5.216	2.646	0.021	0.012
YAK	20	12	20	0.105	0.942	0.030	6.021	2.995	0.024	0.014
BAI	18	16	37	0.068	0.987	0.023	7.974	3.888	0.032	0.018
CHE	73	15	24	0.114	0.899	0.017	4.974	2.447	0.020	0.011
DEB	23	15	31	0.093	0.949	0.028	6.727	3.292	0.027	0.015
GUA	27	10	20	0.180	0.852	0.039	5.556	2.754	0.022	0.012
GUI	62	27	33	0.082	0.933	0.019	5.799	2.812	0.023	0.013
MON	35	17	27	0.084	0.943	0.019	6.424	3.116	0.026	0.014
NIN	27	16	26	0.092	0.943	0.027	5.880	2.898	0.024	0.013
TIB	16	14	22	0.078	0.983	0.028	6.742	3.355	0.027	0.015
HOK	28	3	13	0.865	0.140	0.087	1.325	0.850	0.005	0.004
MIS	26	3	10	0.731	0.280	0.107	1.563	0.963	0.006	0.004
TAI	16	3	14	0.594	0.433	0.138	3.600	1.927	0.015	0.009
TOK	19	1	0	1.000	0.000	0.000	0.000	0.000	0.000	0.000
YON	19	2	1	0.501	0.526	0.040	0.526	0.460	0.002	0.002

MNPD = mean number of pairwise differences.

The relationship between the different breeds was inferred by estimating the pairwise differences between breeds and interpolating the data into an MDS graphic (Figure 2).

This graph shows a large presence of outliers representing low variability (Figure 2A and see also Table 3) and clustering of the majority of breeds having higher variation (Figure

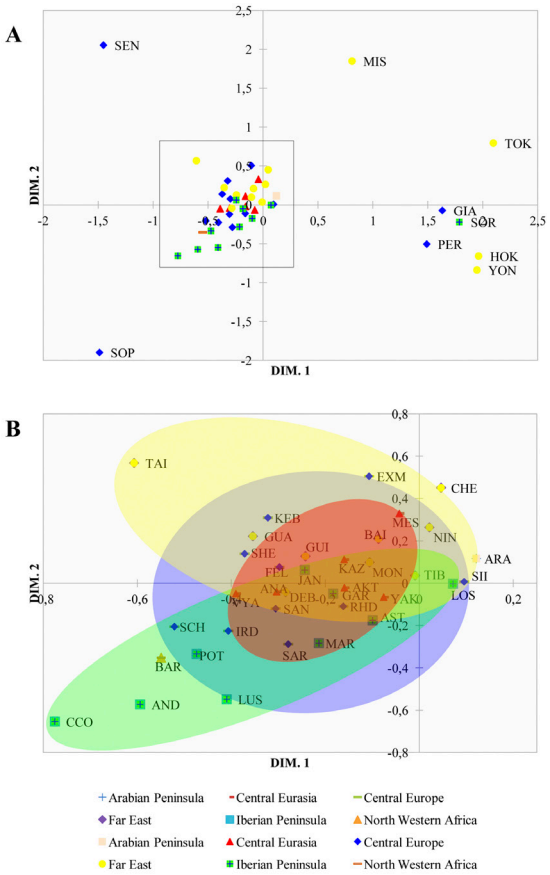

Figure 2. Multidimensional scaling (MDS) plot computed from the matrix of the pairwise differences of the mitochondrial HVS-I sequences. Each symbol represents the breeds from the 6 geographic Old World macroareas. Breed codes are as in Table 2. In the section A of the graph only breeds outside the main cluster are indicated. The breeds enclosed in the square are highlighted in the section B. Colors of the spots: green - Iberian Peninsula, blue - Central Europe, red - Central Eurasia, yellow - Far East. D-star: Raw stress = 30.54; Alienation = 0.12; D-hat: Raw stress = 24.05; Stress = 0.11.

2A and B enlarged). The Giara sample lies in an outlier position because of its low variability. In contrast, the Sarcidano sample is located in an area of generally high variability, included in the Western European group of breeds.

Haplogroup assignment

The phylogenetic analysis involving 147 complete mtDNA genome sequences reported from the literature allowed us to infer the HVS-I haplotype patterns and assign them to the 18 (A-R) haplogroups (Achilli et al., 2012) with better confidence. HVS-I was affected by variation that differed in the quality and degree of information (see Methods). The haplogroups D, F, H, I, L, M, N, Q, and R were defined by highly informative mutations linked univocally to their haplogroup (see Table 4).

Table 4. Relative frequency of HVS-I polymorphisms linked to the haplogroups defined on the whole mitochondrial genome.

Haplogroup	A	B	C	D	E	F	G	H	I	JK	J	K	L	M	N	O	P	Q	R	# Associated haplogroups
# Sequences	13	12	9	5	1	5	16	2	10	1	2	1	30	10	6	2	8	12	2	
Variant																				
15720G	0.769	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1
15737C	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1
15521A	0.000	0.000	0.000	0.800	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1
15595G	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1
15540G	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1
15526C	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1
15709T	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1
15538G	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1
15495T	0.231	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1
15534T	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	1
15496G	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	1
15494C	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	1
15685G	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	1
15601C	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1
15617C	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	1
15740G	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	1
15726A	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.417	0.000	1
15616G	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1
15598C	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1
15574A	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	1
15533G	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.067	0.000	0.000	0.000	0.000	0.000	0.000	1
15718T	0.000	0.000	0.000	0.200	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	2
15596G	0.000	0.000	0.000	0.400	0.000	0.000	0.000	0.000	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	2
15603C	0.077	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	2
15659C	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.900	0.000	0.000	0.000	0.000	1.000	2
15667G	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	0.000	0.000	2
15615G	0.000	0.000	0.000	0.000	0.000	0.000	0.063	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	2
15544C	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.167	0.000	0.000	0.000	0.000	2
15651A	0.000	0.000	0.111	0.000	0.000	0.000	0.000	0.000	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	2
15602C	0.923	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.433	0.000	0.000	0.000	0.000	0.000	0.000	3
15542T	0.000	0.000	0.000	0.000	1.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	3
15635T	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	3
15649G	0.077	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	3
15666A	0.000	0.917	0.000	0.000	1.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	4
15703C	0.000	0.000	0.000	0.000	0.000	0.000	0.938	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	5
15604A	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.300	0.000	0.000	0.500	0.625	0.583	0.000	5
15597G	0.000	0.000	0.333	0.000	0.000	0.200	0.875	0.000	0.200	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.625	0.083	0.000	7
15650G	0.000	1.000	1.000	0.000	1.000	1.000	1.000	0.000	0.800	0.000	0.500	1.000	0.000	0.000	0.000	0.000	0.000	0.083	0.000	9
15585A	0.077	0.583	0.222	0.200	1.000	0.000	0.625	1.000	0.900	1.000	1.000	0.000	0.667	0.200	0.500	0.000	0.000	0.250	0.000	14

Mutated positions are ordered on the basis of their specificity in defining haplogroups.

The EFG clade was defined by the 15542T and 15666A mutations. Haplogroup F was distinguished in its clade by the haplogroup F-specific mutation 15595G, whereas haplogroup G was characterized by its association with the 15635T mutation. The clade OPQR was defined by the mutation 15703T, and the inside group OP was identified by the mutation 15667G.

About 80% of the A haplotypes had the mutation 15720G, 20% of which showed an association between 15720G and 15495G. Overall, the 20% of the A haplotypes had no diagnostic sites and the accuracy of their haplogroup attribution could only be confirmed when HVS-I-specific sites were available.

The B haplotypes were attributed on the basis of an allelic association/exclusion criterion in 90% of cases on the basis of the 15666A mutation if this mutation was not associated with the 15542T mutation typical of the EFG clade. However, 10% of the B sequences did not contain any diagnostic nucleotide site.

The haplogroup C and the JK clade, well defined by specific variants in the coding region, were associated with hypervariable mutations in HVS-I. For this reason, when these mutations were available, inference of haplogroup attribution was conducted by either a comparison or an exclusion criterion.

Following the algorithm shown in Table 4, we classified the HVS-I mtDNA of the 45 typical breeds. In total, 237 haplotypes were identified from the breed dataset and 229 were attributed to haplogroups. Eight haplotypes (3.4%) were unambiguously attributable (**Table S1**). The polymorphism 15602C appeared to be phylogenetically recursive in the L lineage and was also observed in all A and B lineages; analogously, the 15585A, 15597G, 15604A, and 15650G mutations were not specifically associated with the haplogroups, and for this reason we attributed a null phylogenetic weight to these markers.

All of the HVS-I sequences of the Sardinian horses were assigned to corresponding haplogroups on the basis of univocal mutations and other sufficiently informative polymorphisms, so further sequencing or genotyping of specific diagnostic sites in the coding region was not required (see the haplogroup attribution in Table 1).

Twenty-six of 31 HVS-I sequences from the Giara breed belonged to the G haplogroup. In fact, this haplogroup was associated with the pattern 15542T, 15666A, 15650G, and 15635G in all of the 16 complete mtDNA sequences belonging to this haplogroup, and was more variable in the 15597G and 15703C variants. Overall, the G haplogroup represented 84% of the Giara maternal lineages.

The 21 Sarcidano sequences belonged mainly to haplogroups I (43%) and L (38%). Haplogroup L was defined by the HVS-I pattern 15494C, 15496G, 15534T, 15602C, 15603C, and 15649G present in 30 haplogroup L whole mtDNA genome sequences. On the other hand, the I haplogroup in the HVS-I of the 10 complete mtDNA genome sequences was defined by 15709T and 15538G.

Intrahaplogroup variation of the 3 main lineages found in the Giara and Sarcidano horses (haplogroups I, G, and L) was examined in all the available breed data by using the clustering method of the median-joining network. All of these maternal lineages produced a nascent star-like structure of the networks suggesting recent growth.

Haplogroup G (Figure 3) was poorly differentiated in eastern breeds, while a relevant number of subtypes were present in the breeds from Central Europe and the Iberian Peninsula, indicating differences in demographic growth between the eastern and western macroareas. In particular, the longest branches were often shared between samples from different breeds. Haplogroup G was the main haplogroup in the Giara horses. In this breed, the ancestral form

of the haplogroup G evolved into several derivate haplotypes that are scarcely shared with other breeds. By contrast, haplogroup G was rare in the Sarcidano horses and its derivative lineages were shared with samples from the Iberian Peninsula.

The network of haplogroup I (Figure 3) showed a strong signal of population size expansion in breeds from north-western Europe and less differentiation in other macroareas. It was common in the Sarcidano maternal lineages where it was represented by derived and private lineages.

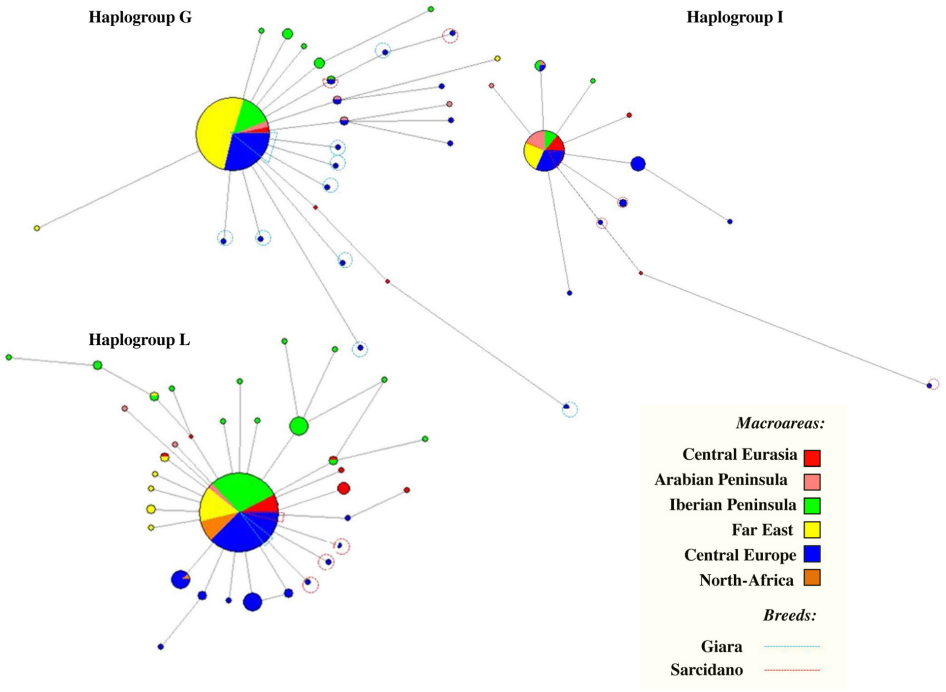

Figure 3. G, I and L intrahaplogroup variations analyzed by neighbor-joining networks.

The network of the haplotypes belonging to the haplogroup L (Figure 3) also showed a recent worldwide growth, but, unlike for haplogroups G and I, the expansion had been similar in eastern and western breeds, even if there were relatively fewer eastern L lineages than western lineages. The location of major evolution was identified as the Iberian Peninsula, but the haplotype sharing of new lineages among macroareas and breeds was very low. The Sarcidano breed displayed emergence of 3 new haplotypes in the evolution of the haplogroup, whereas this was rarely observed in the Giara breed.

Mismatch distribution of the G, I, and L haplogroups in breeds from western and eastern Eurasia and from Sardinia indicated a recent expansion that seems to have occurred earlier in haplogroup L than in the other 2 haplogroups, as indicated by a greater number of pairwise differences (Figure 4).

For the same reason, it is postulated that an earlier western Eurasian expansion took place in the haplogroups I and G. However, correlated r, D, and F_S statistics gave negative values not significantly different from 0, indicating that rare alleles were not more frequent

than expected from a null-neutral hypothesis in an equilibrium population. Moreover, pairwise differences did not fit well into an unimodal mismatch distribution model. This could be explained by a stationary population size or by very slow growth of the population.

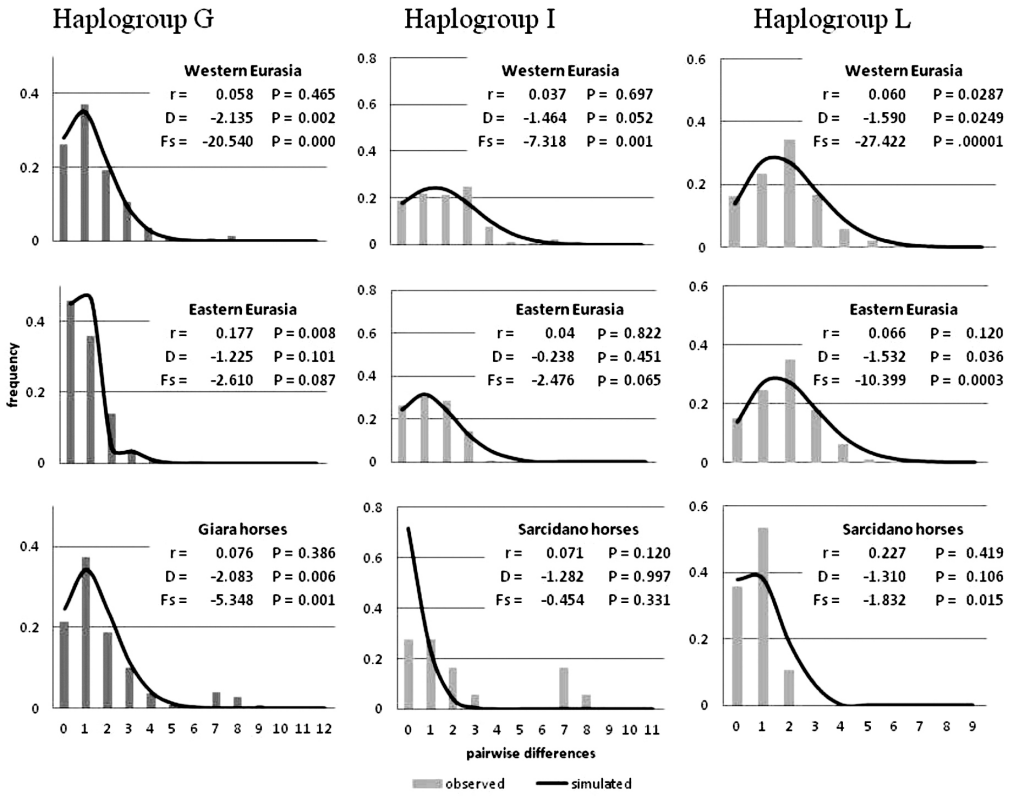

Figure 4. Intrahaplogroup variation analyzed by mismatch distribution analysis in Western and Eastern Eurasian breeds and in Sardinian breeds. Mismatch distributions were established for the haplogroups G, I, and L. Black lines represent the expected mismatch distribution of a stationary population. Histograms represent the observed mismatch distribution from segregating sites of the aligned sequences of HVS-I sequences in horse mtDNA.

DISCUSSION

The mtDNA genetic structure of the lineages observed in the domestic breeds can be used to infer their demographic and domestication history (Kavar and Dovc, 2008; Cieslak et al., 2010; Georgescu et al., 2011). To infer information about the demographic history and origin of the 2 Sardinian Giara and Sarcidano breeds from the genetic data, we compared these data with those from 43 typical breeds of the Old World.

The examined breeds displayed large molecular variation. These differences were not solely due to the total number of different haplotypes because a large amount of this molecular variation was due to the presence of different haplogroups in the same breed. The A-R haplogroups represent extant maternal lineages transmitted from a wild ancestral mare to the

present-day mares: these haplogroups were defined on the complete mtDNA phylogeny as the cutoff of lineages that lived 10,000 years ago and whose haplogroups were transmitted during the process of domestication (Achilli et al., 2012). Therefore, the distinction between molecular variation that is produced before and after the formation of the present-day breeds is particularly relevant. Breeds that have more than 1 haplogroup show an ancestral variability that arose long before breed formation. In addition, as observed in the main lineage networks, the phylogeny of the single haplogroups showed a low degree of evolution: the molecular pattern of the HVS-I belonging to the same haplogroup was barely differentiated because only a limited number of new mutations had arisen in the ancestral haplotype. For this reason, breeds with a single haplogroup had lower molecular variation and were placed as outliers in the MDS graphics.

The emerging "star-like" phylogeny of the haplogroups indicated a population bottleneck followed by a small expansion in population size. In fact, the population size of breeds had remained relatively unchanged for a long time as indicated by the non-significant values of the D and F_S statistics (Figure 4). The presence of a single (and usually frequent) ancestral haplotype shared by the majority of populations suggests founding of recent breeds from the same genetic pool.

As a consequence, all of the breeds that have maintained high haplogroup variability shared the same ancient variation and tended to cluster in the MDS graphic. Furthermore, the partition of the variability determined in AMOVA confirmed that only a small amount of variation was due to differences between breeds. For this reason, the different geographical areas of the breeds' origins tended to overlap (Figure 2A).

Nevertheless, we observed that the variability of breeds from the western European steppes, where, according to archeological records, the domestication originated (Outram et al., 2009), was central and entirely enclosed in the overall variability. Therefore, this area also represents the point of lineage radiation, and diversification in other geographical macroareas appears to be incipient and has not yet been completed.

In addition to suggesting a recent origin of the current breeds, our data also suggest that the homogeneity of the genetic pool may have been stable until recently when warrior peoples repeatedly migrated in several waves from the Central Eurasian steppes into Europe during the Middle Age. An east-west distribution of the variation was also apparent, probably generated from isolation by distance and weakly detected as significant by AMOVA when the breeds were merged into the two east and west Eurasia groups.

A genetic structuring occurred between the 2 Sardinian breeds (Figure 2) analyzed in this study, which were clearly separated and located in 2 different groups as determined by the mitochondrial lineage variation: the Sarcidano horse breed predominantly consisted of the haplogroups I and L, and other less frequent groups. This haplogroup variability, as mentioned above, is the result of molecular and haplotype diversity, which account for the presence of the Sarcidano among the most genetically variable breeds. The MDS analysis, which is based on genetic distances between breeds, effectively positioned the Sarcidano in the group with high variation, showing a greater affinity with the Iberian breeds as well as with those from north-western Europe, where the haplogroups L and I are highly represented. Haplogroup L is the most representative of the Iberian Peninsula, and Spanish influence in the Sardinia Island was strong until 1700. Moreover, none of the derivative Sarcidano horse L or I haplotypes is shared with the other continental breeds, suggesting that there was no recent gene flow from outside into the island.

The Giara breed consisted almost exclusively of the haplogroup G, a very common worldwide haplogroup in horses and typical of many other outlier breeds. Unlike other breeds with little haplogroup variation, the Giara sample showed a significant molecular variation. Mismatch distribution in the Giara G haplogroup was comparable to those identified in the western eurasian population group. We therefore conclude that ancestors of the Giara horse in the past may have occupied an area that was larger than the one where they are found now, albeit always within the boundaries of the island. This interpretation is supported by historical records reporting the widespread presence in Sardinia of a horse described as "wild" that was phenotypically similar to the Giara horse (Cetti, 1774).

In conclusion, in this study we have first described the distribution of the current genetic diversity in typical breeds of horses and compared genetic differences among the various breeds and among groups of breeds from different geographical areas. We also inferred that the genetic diversity in the Sardinian Giara and Sarcidano breeds is the result of recent evolution. We further demonstrated that the genetic system used is powerful of discerning past and present evolutionary patterns.

As to the question of horse domestication, we agree with other authors that horses are a notable exception to the theory that holds that domestication is the result of a very small number of independent, often geographically separated taming events, as observed for all domesticated species (Bruford et al., 2003). The abundance of very differentiated mtDNA lineages indicates that horse domestication probably involved a large number of wild captures. Our analysis support the hypothesis that the area of horse domestication was originally restricted and then gave rise to the high number of present-day horse mitochondrial lineages (Jansen et al., 2002; Forster et al., 2012); however, the subsequent recruitment of local mares from wild horse populations into domesticated herds is less apparent. According to our results, multiple radiation events of lineages from the original place of wild capture better explain the genetic structure of today's domestic horses.

ACKNOWLEDGMENTS

We thank Barbara Wilkens, Marco Apollonio, and Ercole Contu for useful discussions and helpful comments and breed owners for providing sample collections. Research supported by MURST ex60% and the "Fondazione Banco di Sardegna" to L. Morelli and P. Francalacci. A. Useli was supported by the Autonomous Region of Sardinia and by the Programma Operativo Fondo Sociale Europeo 2007-2013.

Supplementary material

REFERENCES

Achilli A, Olivieri A, Soares P, Lancioni H, et al. (2012). Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proc. Natl. Acad. Sci. U. S. A.* 109: 2449-2454.

Bandelt HJ, Forster P and Rohl A (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16: 37-48.

Bower MA, Campana MG, Whitten M, Edwards CJ, et al. (2011). The cosmopolitan maternal heritage of the Thoroughbred racehorse breed shows a significant contribution from British and Irish native mares. *Biol. Lett*. 7: 316-320.

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

Bruford MW, Bradley DG and Luikart G (2003). DNA markers reveal the complexity of livestock domestication. *Nat. Rev. Genet.* 4: 900-910.

Cetti F (1774). I Quadrupedi di Sardegna. In: Storia Naturale di Sardegna (2000). Ilisso Biblioteca Sarda, Nuoro.

Cieslak M, Pruvost M, Benecke N, Hofreiter M, et al. (2010). Origin and history of mitochondrial DNA lineages in domestic horses. *PLoS One* 5: e15311.

Cozzi MC, Strillacci MG, Valiati P, Bighignoli B, et al. (2004). Mitochondrial D-loop sequence variation among Italian horse breeds. *Genet. Sel. Evol.* 36: 663-672.

Excoffier L, Smouse PE and Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.

Excoffier L, Laval G and Schneider S (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* 1: 47-50.

Forster P, Hurles ME, Jansen T, Levine M, et al. (2012). Origins of the domestic horse. *Proc. Natl. Acad. Sci. U. S. A.* 109: E3148.

Georgescu SE, Manea MA, Dudu A and Costache M (2011). Phylogenetic relationships of the Hucul horse from Romania inferred from mitochondrial D-loop variation. *Genet. Mol. Res.* 10: 4104-4113.

Goto H, Ryder OA, Fisher AR, Schultz B, et al. (2011). A massively parallel sequencing approach uncovers ancient origins and high genetic variability of endangered Przewalski's horses. *Genome Biol. Evol.* 3: 1096-1106.

Guastella AM, Zuccaro A, Criscione A, Marletta D, et al. (2011). Genetic analysis of Sicilian autochthonous horse breeds using nuclear and mitochondrial DNA markers. *J. Hered.* 102: 753-758.

Gratani L (1980). Cavallo della Giara. Istituto di Incremento Ippico della Sardegna. RAS, Cagliari.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41: 95-98.

Hill EW, Bradley DG, Al-Barody M, Ertugrul O, et al. (2002). History and integrity of thoroughbred dam lines revealed in equine mtDNA variation. *Anim. Genet*. 33: 287-294.

Ishida N, Hasegawa T, Takeda K, Sakagami M, et al. (1994). Polymorphic sequence in the D-loop region of equine mitochondrial DNA. *Anim. Genet.* 25: 215-221.

Jansen T, Forster P, Levine MA, Oelke H, et al. (2002). Mitochondrial DNA and the origins of the domestic horse. *Proc. Natl. Acad. Sci. U. S. A.* 99: 10905-10910.

Jiang Q, Wei Y, Huang Y, Jiang H, et al. (2011). The complete mitochondrial genome and phylogenetic analysis of the Debao pony (*Equus caballus*). *Mol. Biol. Rep.* 38: 593-599.

Kakoi H, Tozaki T and Gawahara H (2007). Molecular analysis using mitochondrial DNA and microsatellites to infer the formation process of Japanese native horse populations. *Biochem. Genet*. 45: 375-395.

Kavar T and Dovc P (2008). Domestication of the horse: Genetic relationships between domestic and wild horses. *Livest. Sci.* 116: 1-14.

Lei CZ, Su R, Bower MA, Edwards CJ, et al. (2009). Multiple maternal origins of native modern and ancient horse populations in China. *Anim. Genet*. 40: 933-944.

Lippold S, Matzke NJ, Reissmann M and Hofreiter M (2011). Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. *BMC Evol. Biol.* 11: 328.

Luís C, Bastos-Silveira C, Cothran EG and Oom MM (2006a). Iberian origins of New World horse breeds. *J. Hered.* 97: 107-113.

Luís C, Bastos-Silveira C, Costa-Ferreira J, Cothran EG, et al. (2006b). A lost Sorraia maternal lineage found in the Lusitano horse breed. *J. Anim. Breed. Genet.* 123: 399-402.

McGahern AM, Edwards CJ, Bower MA, Heffernan A, et al. (2006a). Mitochondrial DNA sequence diversity in extant Irish horse populations and in ancient horses. *Anim. Genet*. 37: 498-502.

McGahern A, Bower MA, Edwards CJ, Brophy PO, et al. (2006b). Evidence for biogeographic patterning of mitochondrial DNA sequences in Eastern horse populations. *Anim. Genet.* 37: 494-497.

Mirol PM, Peral GP, Vega-Pla JL and Dulout FN (2002). Phylogenetic relationships of Argentinean Creole horses and other South American and Spanish breeds inferred from mitochondrial DNA sequences. *Anim. Genet.* 33: 356-363.

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

Outram AK, Stear NA, Bendrey R, Olsen S, et al. (2009). The earliest horse harnessing and milking. *Science* 323: 1332-1335.

Royo LJ, Alvarez I, Beja-Pereira A, Molina A, et al. (2005). The origins of Iberian horses assessed via mitochondrial DNA. *J. Hered.* 96: 663-669.

Tajima F (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437-460.

Tajima F (1993). Measurement of DNA Polymorphism. In: Mechanisms of Molecular Evolution. Introduction to Molecular Paleopopulation Biology (Takahata N and Clark AG, eds.). Japan Scientific Societies Press, Sinauer Associates, Inc., Tokyo, 37-59.

Weir BS (1996). The second National Research Council report on forensic DNA evidence. *Am. J. Hum. Genet.* 59: 497-500.

Xu S, Luosang J, Hua S, He J, et al. (2007). High altitude adaptation and phylogenetic analysis of Tibetan horse based on the mitochondrial genome. *J. Genet. Genomics* 34: 720-729.

Xu X and Arnason U (1994). The complete mitochondrial DNA sequence of the horse, *Equus caballus*: extensive heteroplasmy of the control region. *Gene* 148: 357-362.

Yang YH, Kim KI, Cothran EG and Flannery AR (2002). Genetic diversity of Cheju horses (*Equus caballus*) determined by using mitochondrial DNA D-loop polymorphism. *Biochem. Genet*. 40: 175-186.

Zuccaro A, Bordonaro S, Longeri M, Cozzi MC, et al. (2009). Mitochondrial DNA control region variation in Sanfratellano horse and two other Sicilian autochthonous breeds. *Ital. J. Anim. Sci*. 8: 180-182.

