

Mitochondrial D-loop sequence variation among Italian horse breeds

Maria Cristina COZZI^{a*}, Maria Giuseppina STRILLACCI^a,
Paolo VALIATI^a, Barbara BIGHIGNOLI^a, Mario CANCEDDA^b,
Marta ZANOTTI^a

^a Istituto di Zootechnica, Facoltà di Medicina Veterinaria, Università degli Studi di Milano,
Via Celoria 10, 20133 Milano, Italy

^b Dipartimento di Biologia Animale, Facoltà di Medicina Veterinaria, Università degli Studi
di Sassari, Via Vienna 2, 07100 Sassari, Italy

(Received 3 December 2003; accepted 17 June 2004)

Abstract – The genetic variability of the mitochondrial D-loop DNA sequence in seven horse breeds bred in Italy (Giara, Haflinger, Italian trotter, Lipizzan, Maremmano, Thoroughbred and Sarcidano) was analysed. Five unrelated horses were chosen in each breed and twenty-two haplotypes were identified. The sequences obtained were aligned and compared with a reference sequence and with 27 mtDNA D-loop sequences selected in the GenBank database, representing Spanish, Portuguese, North African, wild horses and an *Equus asinus* sequence as the outgroup. Kimura two-parameter distances were calculated and a cluster analysis using the Neighbour-joining method was performed to obtain phylogenetic trees among breeds bred in Italy and among Italian and foreign breeds. The cluster analysis indicates that all the breeds but Giara are divided in the two trees, and no clear relationships were revealed between Italian populations and the other breeds. These results could be interpreted as showing the mixed origin of breeds bred in Italy and probably indicate the presence of many ancient maternal lineages with high diversity in mtDNA sequences.

mitochondrial DNA / D-loop / biodiversity / phylogeny / Italian horse

1. INTRODUCTION

Mitochondrial DNA (mtDNA) analysis has often been used in evolutionary studies. Feral and domestic equine cells contain a large number of maternally inherited mitochondria (from 100 to 1000) [1, 23]. The D-loop region of mtDNA is particularly interesting due to the high variability level [1, 23], the moderate mutation rate estimated at one site every 6000 years in humans [18], the matrilineal transmission and the lack of recombination [21].

* Corresponding author: cristina.cozzi@unimi.it

Many mtDNA equine studies based on the D-loop control region analysis address questions of phylogeny and evolution [2, 4–10, 12–15, 22, 24].

The aim of this study was to investigate genetic diversity of the mtDNA D-loop hypervariable region in seven Italian horse populations, in order to evaluate their matrilineal relationships. The Italian autochthonous breeds considered were the following: Giara, Haflinger, Lipizzan, Maremmano, and a small feral Sardinian population called Sarcidano. Italian Trotter and Thoroughbred horses were also included. We considered the relationships within these horse breeds bred in Italy and Spanish, Portuguese, North African and Wild horses (the Mongolian wild horse) in order to provide information about the origin of the Italian populations. An *Equus asinus* sequence was used as the outgroup.

2. MATERIALS AND METHODS

Total DNA was extracted, following standard procedures, from peripheral blood samples of five horses for each of the following breeds: Giara (GRH), Haflinger (HFL), Italian trotter (ITR), Lipizzan (LPZ), Maremmano (MAH), Sarcidano (SRH) and Thoroughbred (TBH). The Giara and Sarcidano were bred in feral conditions and samples were selected at random. The horses in other breeds were selected by pedigree analysis, in order to obtain information about maternal lineages.

The D-loop region was amplified by the polymerase chain reaction (PCR) using two primers specifically designed from a published horse sequence (GeneBank X79547): forward 5'-AACGTTTCCTCCCAAGGACT-3' and reverse 5'-TCAGCAACCCTCCCAACTAC-3' [5, 24]. The amplicon obtained was a 397-bp fragment between the tRNA^{Pro} and the large central conserved sequence block (sites 15382 and 15778), which is considered as the most polymorphic mtDNA region [24].

The reaction profiles included the following: one cycle of denaturation at 94 °C for 9 min followed by 30 cycles of denaturation at 94 °C for 60 s, annealing at 48 °C for 45 s and extension at 74 °C for 1 min; a final extension at 74 °C for 30 min. PCR products were purified and sequenced using the BigDye Terminator Kit (Applied Biosystems) on an ABI PRISM 377 DNA Sequencer equipped with Sequencing Analysis[®] and Sequence Navigator[®] (Applied Biosystems).

Mitochondrial DNA sequences were compared with a reference sequence from a "Swedish horse" (GeneBank X79547) by the BLAST2 SEQUENCES programme [19].

Our sequences were compared with the following mtDNA D-loop sequences selected in GenBank database: Spanish horses AF466006-16; Portuguese horses AY246231-5, AY246243, AY246247; North African horses AJ246181, AJ246185, AJ413660, AJ413668; wild horses (the Mongolian wild horse) AJ413830-2. *Equus asinus* sequence accession number X97337 was chosen as the outgroup.

Multiple alignments between our sequences and the literature ones were performed using CLUSTAL W software [20]. Kimura two-parameter distances, calculated on the basis of an equal substitution rate per site [11], were estimated by PHYLIP software package version 3.5c [3]. Cluster analysis using the Neighbour-joining method [17] was performed by the same programme to obtain a phylogenetic tree viewed in TreeView software [16]. A bootstrap analysis on 1000 replicates was applied in order to evaluate the robustness of the dendrogram.

3. RESULTS

We identified 22 haplotypes among the 35 horses (Tab. I). For each breed we identified from 2 to 5 haplotypes (Tab. II).

The identified haplotypes differed from the reference sequence (GeneBank X79547) by 5-12 sites and from each other by 1-15 sites, within the 397 bp amplicons (Tab. I).

We found 29 base substitution sites in comparison with the reference sequences. The detected mutations corresponded to transitions and we did not find inversions (Tab. I). One deletion, also reported in three sequences (AF481311, AF481320, AF481322) by Hill *et al.* [4], was identified in Thoroughbred samples. Two substitutions (positions 15945 and 15720), already mentioned in other breeds [2, 8, 12], were identified (Tab. II).

The dendrogram reported in Figure 1 was performed with the Neighbour-joining algorithm using the Kimura two-parameter distances estimated on the D-loop sequences.

The clustering of haplotypes shows seven clades A to G (Fig. 1). Clade A joins horses belonging to haplotypes 9, 10, 13, 15, 18, 19 and 22. They differ from each other by 1-7 nucleotide substitutions (Tab. I). Clade B is represented by haplotype 1 that includes six horses, whereas clade C joins the haplotypes 6 and 21, differing from each other by one mutation (Tab. I). Haplotypes 2, 4 and 8 are clustered in clade D and they have 3-4 nucleotide substitution differences. Haplotype 2 shows a characteristic substitution site at position 15601, whereas haplotypes 4 and 8 presented characteristic nucleotide substitutions at

Table I. The haplotypes and nucleotide substitutions identified relative to the reference sequence GenBank X79547. In the column “sample” each breed corresponds to the following GeneBank accession numbers: GRH=AY462426-30; HFL=AY462431-35; ITR=AY462421-25; LPZ=AY462436-40; MAH=AY462446-50; SRH=AY462451-55; TBH=AY462441-45.

Haplotype	Sample	Position of nucleotide substitutions																													
		15489	15491	15494	15495	15496	15512	15526	15532	15534	15538	15540	15542	15585	15596	15597	15601	15602	15603	15604	15617	15635	15649	15650	15659	15666	15667	15703	15709	15718	15720
	X79547	A	A	T	T	A	T	T	C	C	A	A	C	G	A	A	T	C	T	G	T	C	A	A	T	G	A	T	C	C	G
1	2 HFL, 1, 5 ITR, 3 LPZ, 4, 5 SRH	.	.	C	C	G	.	.	.	T	T	C	.	.	.	G	A
2	2, 4 ITR	.	.	C	A	.	.	C	T	A
3	3 ITR, 3 SRH	.	.	C	G	.	.	A	.	.	T	G	T	.	A	
4	1 MAH	.	.	C	A	.	.	T	.	.	C	.	.	.	C	A	
5	2 MAH	.	.	C	G	.	T	.	A	G	C	.	.	A	
6	5 LPZ, 3 MAH	.	.	C	C	G	.	.	.	T	.	.	A	.	.	C	.	.	C	.	.	G	A	
7	4 MAH	.	.	C	A	.	.	T	.	A	C	.	.	A	
8	1 LPZ, 5 MAH	.	.	C	T	.	.	C	.	.	.	C	A	
9	1 TBH	.	C	C	C	G	.	.	T	.	.	A	.	.	.	T	C	A	.	.	G	A	
10	2 TBH	.	.	C	C	G	.	.	T	.	G	.	A	.	.	C	.	.	C	.	.	G	A	
11	3 TBH	.	.	C	.	.	C	.	.	.	G	.	A	.	.	T	G	T	.	A	
12	4 TBH	.	.	C	G	.	.	.	G	.	T	G	T	.	A	
13	5 TBH	.	.	C	C	G	.	.	T	.	G	T	C	A	.	.	G	A	
14	1 SRH	.	.	C	G	.	.	A	.	.	.	T	C	G	T	.	A	
15	2 SRH	.	.	C	C	G	.	.	T	.	G	T	C	G	A	
16	1, 3, 5 GRH	.	.	C	T	A	.	G	.	T	.	.	.	T	.	G	.	A	.	C	.	.	A	
17	2, 4 GRH	.	.	C	T	.	G	.	T	T	.	G	.	A	.	C	.	.	A	
18	2 LPZ	.	.	C	C	G	C	.	.	T	.	.	T	A	.	.	T	C	.	.	.	G	A	
19	1 HFL	G	.	C	C	G	.	.	T	.	G	C	.	.	.	G	A	
20	3 HFL	.	.	C	G	.	T	.	.	G	.	T	T	.	G	.	A	.	C	.	.	A	
21	4 HFL, 4 LPZ	.	.	C	C	G	.	.	T	.	.	A	.	.	.	T	C	.	.	.	G	A	
22	5 HFL	.	.	C	C	G	.	.	T	T	C	A	.	.	G	A	

Table II. Mitochondrial DNA haplotypes in each breed.

Breed	N samples	N haplotype	Haplotype
GRH	5	2	16, 17
HFL	5	5	1, 19, 20, 21, 22
ITR	5	3	1, 2, 3
LPZ	5	5	1, 6, 8, 18, 21
MAH	5	5	4, 5, 6, 7, 8
SRH	5	4	1, 3, 14, 15
TBH	5	5	9, 10, 11, 12, 13

positions 15617 and 15659 (Tab. I). Clade E joins the haplotypes 3, 12 and 14 differing from each other by 1-3 site substitutions. They present characteristic substitutions at positions 15538 and 15709, whereas haplotype 12 is also characterised by site 15596 (Tab. I). Clade F joins the haplotypes 16, 17 and 20, differing from each other by 1-2 nucleotide substitutions and have characteristic mutations in positions 15635 and 15666. The haplotypes 5 and 7 are present in clade G. Haplotype 5 has a characteristic substitution in position 15667. Haplotypes 11 and 18 show larger distances than the other haplotypes and they are separated in the dendrogram. They showed identifying mutations in positions 15526, 15718 and 15512 respectively (Tab. I).

We also analysed our data in a wider context. Only 7 haplotypes (1, 3, 7, 11, 17, 21 and 22) showed a 100% alignment upon comparison of our sequences with those with the same 397 bp length or longer found in the GeneBank database.

Our data from Thoroughbreds compared with those present in the literature on the same breed showed a similarity between haplotype 11 and the *founder*-haplotype K identified by Hill *et al.* [4]. The other haplotypes differed from Hill's haplotypes by 1 to 12 nucleotide substitutions.

The Lipizzan bred in Italy showed haplotypes that are also frequent in other breeds. The haplotypes 1 and 21 were similar respectively to the Allegra and Montaura haplotypes identified by Kavar *et al.* [9] and were considered more frequent in the Lipizzan maternal lines.

In Figure 2 a Neighbour-joining dendrogram is shown using the mtDNA D-loop sequences selected from the GenBank database.

4. DISCUSSION

The mitochondrial DNA D-loop region is very polymorphic as reported by many authors (2, 4, 5, 6, 7, 8, 9, 10, 12, 13, 22, 24). Twenty-two haplotypes were identified in our samples. Haflinger, Lipizzan, Maremmano and Thoroughbred showed the highest variability (5), while Giara showed the lowest variability (2) (Tab. II). By contrast the Sarcidano horse, also bred in feral conditions, showed a high level of variability (Tab. II).

However, the five Giara samples were selected at random, while selection for the other breeds used pedigree information to maximise maternal diversity.

The presence in Thoroughbreds and in Lipizzans bred in Italy of haplotypes more frequent in maternal lines and considered in Lipizzan "ancestral" [9], is in accordance with the wide genetic base of the maternal lines of these breeds [4, 8, 9].

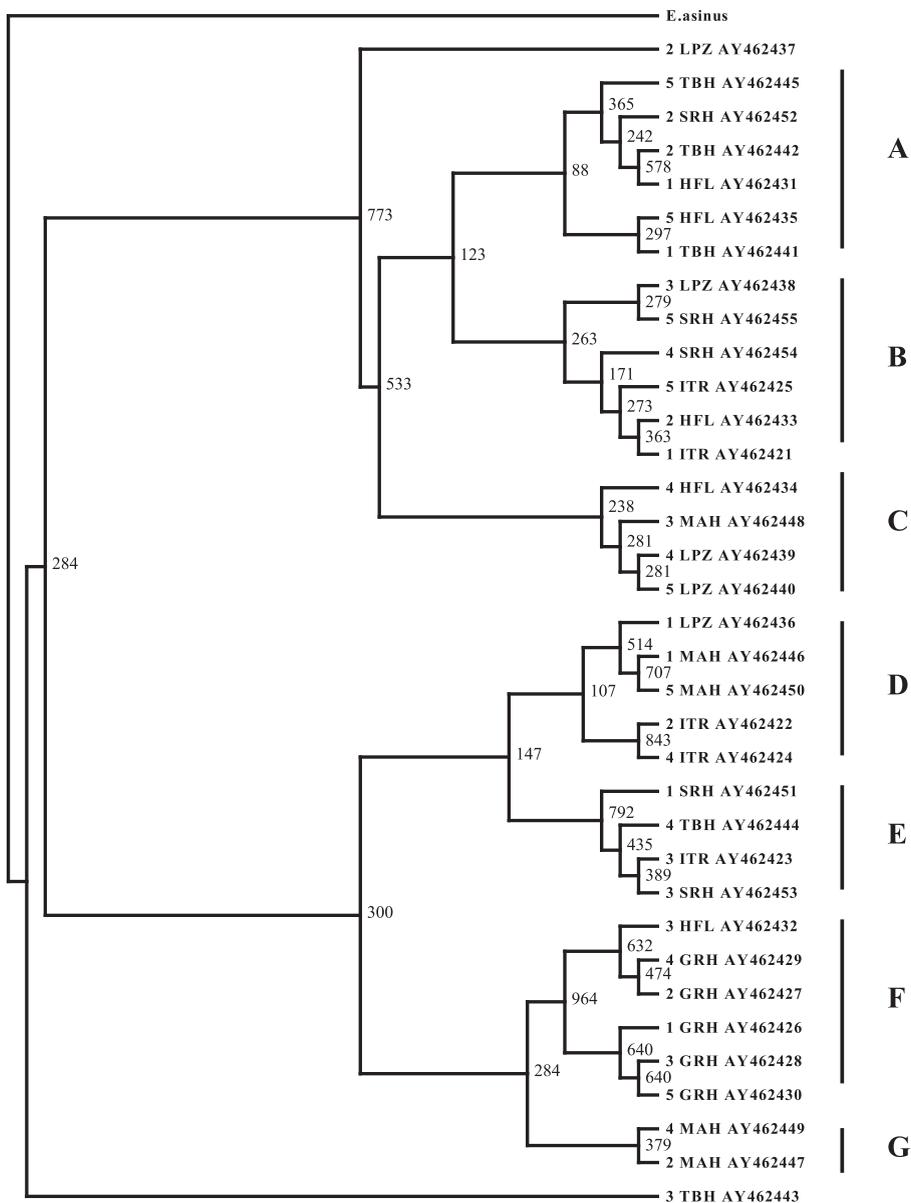


Figure 1. Neighbour-joining tree relating mtDNA haplotypes in horses bred in Italy. The capital letter on the right indicates cluster groups referring to different haplotypes.

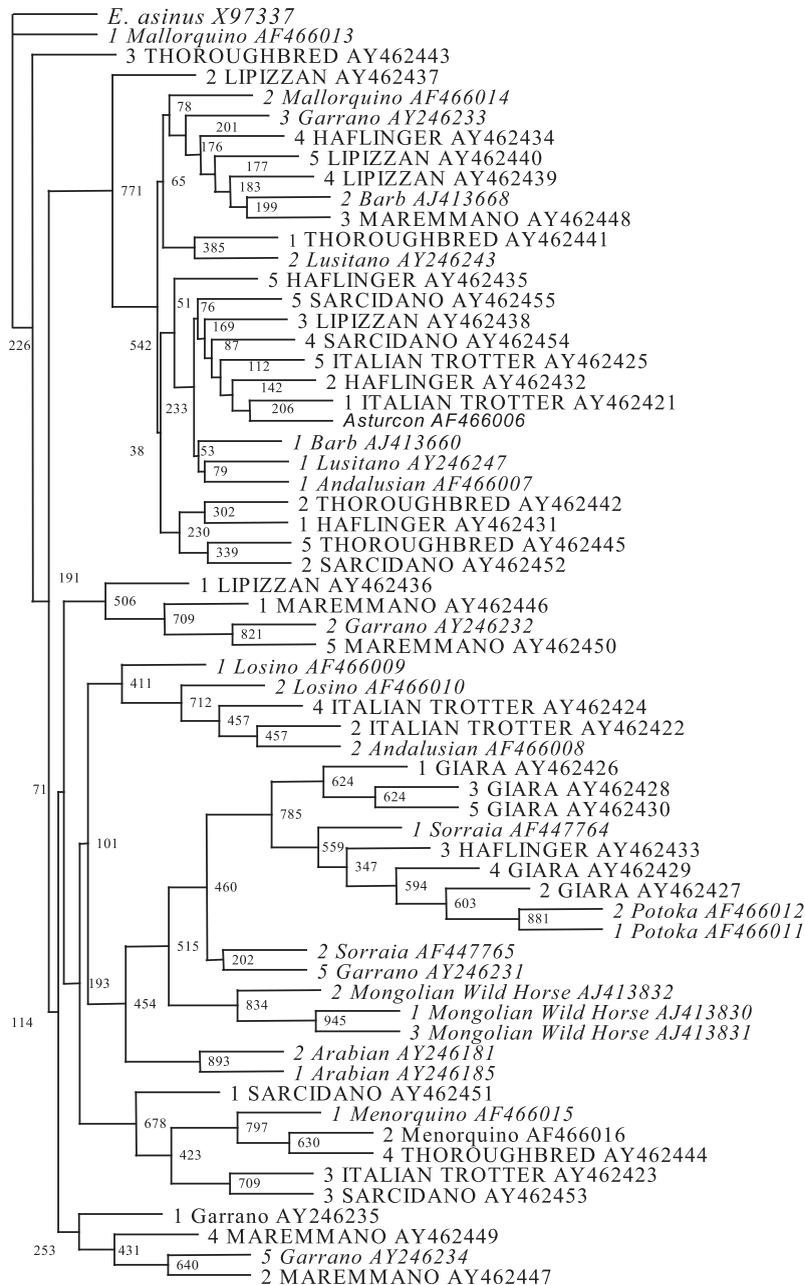


Figure 2. Neighbour-joining tree of breeds bred in Italy (in capital letters) including Spanish, Portuguese, North African and Wild horses (Mongolian wild horses) with their sequence accession numbers (in italics).

Relationships among the Italian horse breeds obtained using Kimura two-parameter distances are reported in Figure 1.

According to the large number of haplotypes identified in the analysed sample, our breeds except Giara spread out in the dendrogram clades.

In the dendrogram performed using mtDNA data, Giara appears to be homogeneous and clustered in a unique clade (F) (Fig. 1).

The wide variability of the D-loop sequences among our populations may be caused by the multiple origins of the breeds bred in Italy, in accordance with the results of other authors studying different horse populations [8, 10, 13]. In reference to these breeds, the high variability of the mtDNA haplotypes within Italian populations is probably due to the important role played by other horse populations influencing the evolution of Italian horse breeds.

The analysis of our sequences in a wider context is reported in Figure 2. All the horse breeds bred in Italy but Giara appear to be spread out in the different clades. Giara is grouped in a unique clade joint with Sorraia, Garrano and Potoka, considered as very ancient breeds (Fig. 2). The two Sorraia haplotypes are closely related to the two Giara haplotypes 16 and 17. In fact haplotype 17 aligns 100% with Sorraia AF447764, whereas it differs from Sorraia AF447765 by 3 nucleotide substitutions. Haplotype 16 differs from Sorraia AF447764 by just one substitution and from Sorraia AF447765 by two substitutions. The close relationships between Giara and the other breeds in the cluster could be interpreted by the presence of common ancient maternal lineages.

However, the poor number of Giara horses sampled prevents us from making any conclusion on this hypothesis and the question needs to be investigated more.

The distribution of our haplotypes in the different clades suggests that, as reported by some authors [7, 13, 22], the modern horse mtDNA sequences do not define monophyletic groups. In particular compared to wild progenitors, modern horse populations are not derived from a single stock of wild horses. The horse domestication probably involved several distinct populations. The initially domesticated horses spread out and incorporated wild mares forming different mtDNA clusters [7]. In this case, the phylogenetic differences detected in our breeds could be explained by the presence of a very ancient mitochondrial diversity.

In conclusion, we provide a preliminary sequence characterisation and phylogenetic study by mitochondrial D-loop DNA polymorphism of seven Italian horse breeds. Horse populations bred in Italy are the result of multiple origins since they retain very ancient mitochondrial diversity.

Accession numbers

GenBank accession numbers for the sequences presented in our study are: GRH AY462426-30, HFL AY462431-35, ITR AY462421-25, LPZ AY462436-40, MAH AY462446-50, SRH AY462451-55, TBH AY462441-45.

ACKNOWLEDGEMENTS

This paper was supported by Italian MURST (research program FIRST). We would like to thank all breeder associations for providing sample collection.

REFERENCES

- [1] Aquadro C.F., Greenberg B.D., Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals, *Genetics* 103 (1983) 287–312.
- [2] Bowling A.T., Del Valle A., Bowling M., A pedigree-based study of mitochondrial D-loop DNA sequence variation among Arabian horses, *Anim. Genet.* 31 (2000) 1–7.
- [3] Felsenstein J., PHYLIP (Phylogeny Inference Package) version 3.55, Department of Genetics, University of Washington, Seattle, 1995.
- [4] Hill E.W., Bradley D.G., Al-Barody M., Ertugrul O., Splan R.K., Zakharov I., Cunningham E.P., History and integrity of Thoroughbred dam lines revealed in equine mtDNA variation, *Anim. Genet.* 33 (2002) 287–294.
- [5] Ishida N., Hasegawa T., Takeda K., Sakagami M., Onishi A., Inumaru S., Polymorphic sequence in the D-loop region of equine mitochondrial DNA, *Anim. Genet.* 25 (1994) 215–221.
- [6] Ishida N., Oyunsuren T., Mashima S., Mukoyama H., Saitou N., Mitochondrial DNA sequences of various species of genus *Equus* with special reference to the phylogenetic relationship between Przewalskii's wild horse and domestic horse, *J. Mol. Evol.* 41 (1995) 180–188.
- [7] Jansen T., Forster P., Levine M.A., Oelke H., Hurler M., Renfrew C., Weber J., Olek K., Mitochondrial DNA and the origins of the domestic horse, *PNAS* 99 (2002) 10905–10910.
- [8] Kavar T., Habe F., Brem G., Dovè P., Mitochondrial D-loop sequence variation among the 16 maternal lines of the Lipizzan horse breed, *Anim. Genet.* 30 (1999) 423–430.
- [9] Kavar T., Brem G., Habe F., Sölkner J., Dovè P., History of Lipizzan horse maternal lines as revealed by mtDNA analysis, *Genet. Sel. Evol.* 34 (2002) 635–648.
- [10] Kim K.-I., Yang Y.-H., Lee S.-S., Park C., Ma R., Bouzat J.L., Lewin H.A., Phylogenetic relationships of Cheju horses to other horse breeds as determined by myDNA D-loop sequence polymorphism, *Anim. Genet.* 30 (1999) 102–108.

- [11] Kimura M., A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences, *J. Mol. Evol.* 16 (1980) 111–120.
- [12] Luís C., Bastos-Silveira C., Cothran E.G., Oom M.M., Variation in the mitochondrial control region sequence between the two maternal lines of the Sorraia horse breed, *Genet. Mol. Biol.* 25 (2002) 309–311.
- [13] Mirol P.M., Peral García P., Vega-Pla J.L., Dulout F.N., Phylogenetic relationships of Argentinean Creole horses and other South American and Spanish breeds inferred from mitochondrial DNA sequences, *Anim. Genet.* 33 (2002) 356–363.
- [14] Oakenfull E.A., Ryder O.A., Mitochondrial control region and 12S rRNA variation in Przewalski's horse (*Equus przewalskii*), *Anim. Genet.* 29 (1998) 456–459.
- [15] Oakenfull E.A., Lim H.N., Ryder O.A., A survey of equid mitochondrial DNA: Implications for the evolution, genetic diversity and conservation of *Equus*, *Conserv. Genet.* 1 (2000) 341–355.
- [16] Page R.D.M., TreeView (Win32), University of Glasgow, Division of Environmental and Evolutionary biology, Institute of Biomedical and Life Sciences, Glasgow, 1998.
- [17] Saitou N., Nei M., The Neighbor-joining method: a new method for reconstructing phylogenetic trees, *Mol. Biol. Evol.* 4 (1987) 406–425.
- [18] Stoneking M., Sherry S.T., Redd A.J., Vigilant L., New approaches to dating suggest a recent age for the human mtDNA ancestor, *Philosophical transaction of the Royal Society of London, B* 337 (1992) 167–175.
- [19] Tatusova T.A., Madden L.T., BLAST2 SEQUENCES, a new tool for comparing protein and nucleotide sequences, *FEMS Microbiol. Lett.* 174 (1999) 247–250.
- [20] Thompson J.D., Higgins D.G., Gibson T.J., CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.* 22 (1994) 4673–4680.
- [21] Vigilant L., Pennington R., Harpending H., Kocher T.D., Wilson A.C., Mitochondrial DNA sequences in single hairs from a southern African population, *Proc. Natl. Acad. Sci. USA* 86 (1989) 9350–9354.
- [22] Vilà C., Leonard J.A., Götherström A., Marklund S., Sandberg K., Lidén K., Wayne R.K., Ellegren H., Widespread origins of domestic horse lineages, *Science* 291 (2001) 474–477.
- [23] Wolstenholme D.R., Animal mitochondrial DNA: structure and evolution, *Int. Rev. Cytol.* 141 (1992) 173–215.
- [24] Xu X., Arnason U., The complete mitochondrial DNA sequences of the horse, *Equus caballus* extensive heteroplasmy of the control region, *Gene* 148 (1994) 357–362.