Genetic variability of Akhal-Teke horses bred in Italy

Maria C. Cozzi[*], Maria G. Strillacci[*], Paolo Valiati, Elisa Rogliano, Alessandro Bagnato and Maria Longeri

Department of Veterinary Medicine, Università degli Studi di Milano, Milan, Italy
[*] These authors contributed equally to this work.

ABSTRACT

Background. The Akhal-Teke horse (AKH) is native of the modern Turkmenistan area. It was introduced in Italy from 1991 to 2000 mainly as an endurance horse. This paper characterizes the genetic variability of the whole Italian AKH horse population and evaluates their inbreeding level by analyzing microsatellite markers and mitochondrial D-Loop sequences.

Methods. Seventeen microsatellite marker loci were genotyped on 95 DNA samples from almost all the AKH horses bred in Italy in the last 20 years. Standard genetic variability measures (H_o, H_e, F_{IS}) were compared against the same variables published on other eight AKH populations. In addition, 397 bp of mtDNA D-loop region were sequenced on a sub-group of 22 unrelated AKH out of the 95 sampled ones, and on 11 unrelated Arab horses. The haplotypes identified in the Italian population were aligned to sequences of AKH (56), Arab (five), Caspian Pony (13), Przewalskii (two) and Barb (15) horses available in GenBank. The Median Joining Network (MJN), Principal Component Analysis (PCA) and Neighbor-joining (NJ) tree were calculated on the total 126 sequences.

Results. Nucleic markers showed a high degree of polymorphism ($H_o = 0.642$; $H_e = 0.649$) and a low inbreeding level ($F_{IS} = 0.016$) in Italian horses, compared to other AKH populations (ranged from -0.103 AKH from Estonia to 0.114 AKH from Czech Republic). High variability was also recorded in the D-Loop region. 11 haplotypes were identified with haplotype diversity (hd), nucleotide diversity (π) and average number of nucleotide differences (k) of 0.938, 0.021 and 6.448, respectively. When all the 126 D-Loop sequences were compared, 51 haplotypes were found, and four were here found only in the Italian AKH horses. The 51 haplotypes were conformed to eight recognized mtDNA haplogroups (A, C, F, G, L, M, P and Q) and confirmed by MJN analysis, Italian horses being assigned to five haplogroups (A, C, G, L and M). Using a PCA approach to the same data, the total haplotypes were grouped into two clusters including A+C+M+P and G+F haplogroups, while L and Q haplogroups remained ungrouped. Finally, the NJ algorithm effectively discretizes only the L haplogroup. All the above data univocally indicate good genetic variability and accurate management of the Akhal-Teke population in Italy.

Subjects Biodiversity, Genetics, Veterinary Medicine, Zoology
Keywords Horse, Akhal-Teke, mtDNA D-loop, Microsatellite markers, Genetic variability, Biodiversity, Sequencing, Median joining network, Neighbor-joining tree, PCA

INTRODUCTION

In the recent past, several molecular markers and techniques have been used to investigate the genetic origin and variability within and across local endangered populations to disclose biodiversity and to support conservation plans (*Cothran, van Dyk & van der Merwe, 2001*; *Bömcke, Gengler & Cothran, 2011*; *Kusza et al., 2013*).

Mitochondrial DNA (mtDNA) has also been widely used to estimate genetic diversity, phylogenetic relationships and maternal origins among closely related modern populations. Particularly for such purposes, the mtDNA D-loop hypervariable region 1 has been sequenced in equids (*Beja-Pereira et al., 2004*; *McGahern et al., 2006*; *Royo et al., 2005*; *Prystupa et al., 2012*; *Zhang et al., 2012*; *Hristov et al., 2016*; *Sziszkosz et al., 2016*; *Cieslak et al., 2010*; *Cieslak et al., 2017*; *Cozzi et al., 2017*). The distribution of haplogroups of the hypervariable region 1 and the complete sequence of the mtDNA D-loop from samples of horse remains from archaeological sites have been recently analyzed by several study teams to date horse domestication events (*Jansen et al., 2002*; *Lippold et al., 2011*; *Achilli et al., 2012*; *Librado et al., 2017*). These authors find a large number of maternal lineages in the domestic horse gene pool (*Vilà et al., 2001*; *Cieslak et al., 2010*). The geographic structuring of mtDNA is low today, but there is also evidence that some maternal lineages ware regionally confined (*Cieslak et al., 2010*). Recently, Elsner and collaborators (*2017*) tracked the dynamics of horse population migrations due to climate change in the Pleistocene-Holocene transition from nine archaeological sites adjacent to the Swiss Jura Mountains, providing a novel approach to the evolutionary study of the impact of human encroachment and habitat change on the development of horse populations.

Akhal-Teke (AKH) is a very ancient horse breed that is native of central Asia, precisely the southern region of modern Turkmenistan, and has adapted well for endurance capacity to cover long distances in harsh climate conditions of Central Asian steppes. The breed during the last 100 years has been selected strongly for speed (*Leisson et al., 2011*). As reported by several authors using protein markers, microsatellites, Single Nucleotide Polymorphisms (SNP) and mitochondrial DNA, the AKH belongs to the clade of Middle Eastern Breeds that includes Arabian, Caspian pony and Turkoman horses (*Cothran, van Dyk & van der Merwe, 2001*; *Luis et al., 2007*; *Bömcke, Gengler & Cothran, 2011*; *Priskin et al., 2010*; *Achilli et al., 2012*; *Petersen et al., 2013*).

The AKH breed is renowned for the distinguishing metallic shine of its hair. Moreover, natural athleticism makes a well-rounded sports horse of this breed, which is good at dressage, show jumping, eventing, racing, and endurance riding competitions. This ancient breed represents a unique genetic resource that differs from all the other breeds worldwide in terms of type, conformation and use. The pedigree recording started in 1885, and the Studbook was closed in 1932. The AKH was introduced in Italy from 1991 to 2000 mainly as an endurance horse.

This paper conducts a genetic variability analysis based on genomic DNA loci (microsatellite markers) and mtDNA D-loop of AKH horses reared in Italy. Furthermore, a meta-analysis was performed using horse mtDNA sequences available in GenBank, besides

a comparison with literature data on the genetic variability of AKH horses bred in other countries.

MATERIAL AND METHODS

DNA extraction, genotyping and sequencing

The Italian AKH Association's reference laboratory for genetic testing for parentage control at the Department of Veterinary Medicine (University of Milan) maintains the bio-bank of the biological samples of all AKH horses bred in Italy since they were first imported (107).

All applicable international, national and/or institutional guidelines for the care and use of animals were followed. The bio-banked samples were approved by the Ethics Committee of Università degli Studi di Milano Prot. OPBA_56_2016.

Ninety-five of the 107 blood samples (males and females) were genotyped using 17 microsatellite markers of StockMarks for the Horses® Genotyping Kit (utilized for parentage testing), including 12 from the official ISAG panel (http://www.isag.us/Docs/ EquineGenParentage2016.pdf). All PCR products were run on ABI PRISM® 310 Genetic Analyzer, and fragments were analyzed with the Genescan® and Genotyper® software (http://tools.thermofisher.com/content/sfs/brochures/cms_040300.pdf).

According to the available pedigrees (first generation/parents), the horses with no common mare and stallion ($n = 22$) were sequenced, and so were 11 additional Arabian horses reared in Italy and bio-banked at the University of Milan. The 379 bp fragment of the mtDNA D-loop region (nucleotide position: 15,382–15,778) was amplified and sequenced by classic Sanger methods, using previously published primer pairs (*Cozzi et al., 2004*). The PCR products were purified utilizing SEPHADEX® G-50 (SIGMA) and sequenced for both strands by using the Big Dye® Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's guidelines. The PCR products were processed on an ABI PRISM®310 Genetic Analyzer, and the nucleotide sequence was determined using the ABI PRISM® DNA Sequencing Analysis Software.

Statistical analysis on microsatellite data

Allelic frequencies, genetic equilibrium according to Hardy–Weinberg (HW) (*P*-val), linkage disequilibrium (LD) and inbreeding coefficient (F_{IS}) were estimated using the GENEPOP software (*Raymond & Rousset, 1995*). LD between a pair of loci was also tested for genotypic data using the likelihood-ratio test of Arlequin v.3.5.2.2 software (*Excoffier & Lischer, 2010*). The number of alleles (NA), the effective number of alleles (NE) and the observed (H_o) and expected (H_e) heterozygosity per locus and on the entire population were calculated with POPGENE v.1.32 (*Yeh, Yang & Boyle, 1997*). Allelic richness (AR) per locus and in the population was also used to estimate genetic population diversity. AR was standardized for variations in sample size, and was calculated using FSTAT 2.9.3 (*Goudet, 2002*), based on 79 diploid individuals with no missing genotypes.

Statistical analysis on sequence data

Sequences were checked for the presence of ambiguous bases by using the software Chromas v.2.5.1 (http://www.technelysium.com.au/).

In order to define the haplotypes (H-) all obtained sequences were aligned with the complete horse mtDNA sequences X79547 (*Xu & Arnason, 1994*) and JN398377 (*Achilli et al., 2012*), widely used as reference sequences, using the ClustalW algorithm implemented in BioEdit (*Hall, 1999*).

The number of polymorphic sites, the parsimony informative and the singleton site, the number of haplotypes, the haplotype diversity (hd), the nucleotide diversity (π) and the average number of nucleotide differences (k) were estimated using the DnaSP5 v.5.10.01 software (*Librado & Rozas, 2009*). Haplotype pairwise genetic differentiation was calculated using Arlequin v.3.5.2.2 (*Excoffier & Lischer, 2010*).

In order to define the genetic relationships among our mtDNA sequences of AKH and those available in Genbank for other AKH lines and for other populations originated from the same Middle Eastern clade, a total of 104 sequences (two reference mtDNA, two Przewalski, five Arabian horse, 15 Barb horse and 13 Caspian pony, 56 Akhal-Teke and the 11 Arabian horses bred in Italy) were considered. All the sequences were trimmed at 254 bp to maximize the number of sequences included. Haplogroup definition was performed according to *Achilli et al. (2012)*. The mtDNA sequences from GenBank database used for haplogroup definition and relationship investigation were reported in Table S1.

Three different approaches were used to investigate the haplotype and haplogroup relationship:

 i) the Median-Joining Networks (MJN) analysis performed by Network v.5.0 software (http://www.fluxus-engineering.com); the default setting of Epsilon (0) was applied and the mutational hot spots present in the 254bp fragment were down-weighted as suggested by *Jansen et al. (2002)* and *Cieslak et al. (2010)*;

 ii) the Neighbor-joining (NJ) tree (*Saitou & Nei, 1987*) based on the Kimura-2-parameter model distances (*Kimura, 1980*) was calculated among the haplotypes with MEGA v.7.0.14 software (*Kumar, Stecher & Tamura, 2016*); the bootstrap analysis, running 1,000 bootstrap replicates, was applied to check the robustness of the resulting dendrogram and a cut-off bootstrap at a significant value of 50 was applied;

 iii) the Principal Component Analysis (PCA) with a minimum spanning tree based on haplotype sequences was calculated using the PAST v.2.7 software (*Hammer, Harper & Ryan, 2001*).

The pairwise population genetic differentiation (F_{ST}) was calculated by Arlequin v.3.5.2.2 (*Excoffier & Lischer, 2010*) and used to evaluate genetic differentiation among Italian AKH, Italian Arabian, and the 93 horse GenBank sequences.

All the horse mtDNA sequences obtained in this study were deposited (GenBank ID: MF580226–MF580244).

RESULTS

Microsatellite markers

The results of microsatellite analyses are reported in Table 1. A total of 112 allelic variants were found on 17 microsatellites, ranging from four (HTG6 and HTG7) to 10 alleles (ASB2). The microsatellite LEX03 (nine alleles), mapping on X chromosome, was excluded from the statistical analyses.

Table 1   Measures of genetic variation in AKH bred in Italy by 16 microsatellites.

Locus	NA	NE	AR	P-val $\pm$ SE	H_o	H_e	F_{IS}
AHT4	6	2.6	5.9	0.019 $\pm$ 0.01	0.653	0.617	-0.058
HMS7	6	2.6	5.9	0.676 $\pm$ 0.01	0.589	0.623	0.055
HTG4	6	1.7	5.8	0.026 $\pm$ 0.01	0.379	0.410	0.075
VHL20	9	3.8	8.8	0.442 $\pm$ 0.02	0.758	0.740	-0.024
AHT5	5	3.8	5.0	0.093 $\pm$ 0.01	0.755	0.740	-0.021
ASB2	10	5.3	9.8	0.511 $\pm$ 0.02	0.810	0.816	0.007
HMS6	6	2.8	5.8	0.139 $\pm$ 0.01	0.558	0.643	0.133
HTG6	4	2.2	3.9	0.071 $\pm$ 0.01	0.453	0.548	0.174
HMS2	8	3.6	7.8	0.429 $\pm$ 0.02	0.716	0.728	0.017
HMS3	7	2.9	6.9	0.154 $\pm$ 0.01	0.653	0.665	0.019
HTG10	7	3.5	6.8	0.346 $\pm$ 0.01	0.800	0.720	-0.113
HTG7	4	2.0	3.8	0.685 $\pm$ 0.01	0.505	0.493	-0.026
ASB17	9	5.2	8.9	0.280 $\pm$ 0.02	0.625	0.813	-0.061
ASB23	6	3.3	6.0	0.167 $\pm$ 0.01	0.600	0.702	0.147
CA425	6	2.1	6.0	0.015 $\pm$ 0.01	0.532	0.522	-0.018
HMS1	4	2.5	4.0	0.226 $\pm$ 0.01	0.642	0.609	-0.055
Mean $\pm$ SD	6.4 $\pm$ 1.8	3.1 $\pm$ 1.1	6.4	0.050 $\pm$ 0.01	0.642 $\pm$ 0.1	0.649 $\pm$ 0.1	0.016

Notes.

NA, number of alleles; NE, effective number of alleles; AR, allelic richness; HW, Hardy-Weinberg; P-val, P values; FIS, observed (Ho) and expected (He) heterozygosity values and inbreeding coefficient values; SE, standard error; SD, standard deviation.

The mean number of alleles (NA) calculated on 16 microsatellites was 6.4, whereas the effective mean number of alleles (NE) was 3.1. The analyses of the Hardy–Weinberg genetic equilibrium (P-val), calculated using HW formulae, showed three microsatellites (AHT4, HTG4 and CA425) that were not in genetic equilibrium ($P < 0.05$). This drift is probably caused by strongly unbalanced allelic frequencies in the small population originally imported to Italy from 1991 to 2000 (24 horses: 17 females and seven males).

According to the position (cM) of the microsatellites as reported by *Swinburne et al. (2006)*, LD analyses were performed only for the markers located on the same chromosome: HMS6 and HTG7 (chr 4), HMS3 and HTG4 (chr 9) and ASB2, HTG6 and HMS1 (chr 15) (Table S2). LD between a pair of loci was tested in all samples together and by groups of individuals: those imported in Italy from 1991 to 2000 (Founders) and the ones born in Italy after 2000 (New population). Considering all samples together, all pairs of loci were in LD, except for HTG7 vs HMS6 (Table S2). The group of Founders was in equilibrium for all pair of loci; instead, for the New population the LD was observed for ASB2 vs HMS1 (GENEPOP and Arlequin) and ASB2 vs HTG6 (Arlequin) pairs of markers (Table S2). Hence, we calculated the NE based on the observed allelic frequencies that ranged from 1.7 (HTG4) to 5.3 (ASB2), thus revealing that some loci had a very unbalanced representation of allele variants. These included AHT4 and HTG4, two of the three loci that were not balanced, which present an NE value of 2.6 and 1.7, respectively.

Allelic richness (AR) is a parameter that is used to estimate diversity in the population. It ranges from 4 (HMS1) to 9.8 (ASB2), with a mean value of 6.4, comparable to the AR of 6.09 reported by *Warmuth et al. (2011)*.

H_o and the H_e values ranged from 0.379 (HTG4) to 0.810 (ASB2), with a mean value of 0.642 and, from 0.410 (HTG4) to 0.816 (ASB2), with a mean value of 0.649, respectively The F_{IS} values per locus that ranged from -0.113 (HTG10) (outbreeding) to 0.174 (HTG6) (inbreeding); the overall F_{IS} value for the population was 0.016.

We estimated the diversity parameters on the AKH horses imported in Italy from 1991 to 2000. The values obtained were compared with those estimated in the AKH population born in Italy (after 2000). The two groups of horses are in HW equilibrium and have similar genetic values, except for the F_{IS} value that denotes, in the first nucleus, an apparent excess of heterozygosis Table S3.

Mitochondrial DNA sequences

Table 2 shows the eleven haplotypes found in AKH reared in Italy (H1 to H11). The 22 polymorphic sites identified were 19 parsimony informative and three singleton sites (15495, 15538 and 15709). All the polymorphic sites reported in Table 2 were transition sites. The maximum composite likelihood estimate of the nucleotide substitution pattern, identified for the 22 sequences, were 32.5% (A), 26.6% (T), 29.1% (C), and 11.8% (G), depending on the order of nucleotide composition in the vertebrate mitochondrial genome (A>C>T>G). The A+T content at the mtDNA D-loop region was 59.1% and exceeded G+C content (40.9%). *Zhang et al. (2012)* for Chinese Native horse breeds and *Sziszkosz et al. (2016)* for Hungarian Gidran horses report similar rich A+T content values in horse mtDNA, precisely 55.8% and 64.1% respectively.

Nucleotide differences compared to reference sequences ranged from 1 (H1) to 11 (H11). The most frequent haplotype found in our samples was H4 (18.2%), while the less frequent haplotypes were H1, H6, H8, H10 and H11 recorded in only one individual each (4.5%).

The values of variability parameter haplotype diversity (hd), nucleotide diversity (π) and average number of nucleotide differences (k) were 0.938, 0.021 and 6.448, respectively.

The pairwise distance matrix (Fig. 1) represents the number of nucleotide substitutions in each haplotype, where Hap_R represent the haplotype of the reference sequences. Inter-haplotypic differences ranged from 1 (Hap_R vs Hap_1 and Hap_10 vs Hap_11) to 13 (Hap_4 vs Hap_10 and Hap_4 vs Hap_11).

A total of 51 haplotypes with 36 variable sites (seven singleton and 29 parsimony informative) were identified by the alignment of the 126 mtDNA D-loop sequences, showing an hd value of 0.971. The detailed sequence variations of haplotypes and the identified haplogroups were reported in Table S4.

Four haplotypes (H16, H18, H21 and H22) were found exclusively in Italian AKH and were not reported in GenBank for the same breed. Additionally, our AKH sequences were grouped in five haplogroups A, C, G, L and M, as defined by *Achilli et al. (2012)*.

The MJN (Fig. 2A) showed the relationships among the 51 haplotypes occurring in the 126 mtDNA sequences, and discretized them in the eight haplogroups. Relationships

Table 2 Polymorphic sites in the horse Italian AKH mtDNA D-loop sequence nucleotide ranging from 15385 to 15862 positions, according to the reference sequences X79547 and JN398377.

Haplotypes	Position of nucleotide substitutions																						N	Freq.%
	15494 T	15495 T	15496 A	15521 G	15534 C	15538 A	15542 C	15585 G	15597 A	15602 C	15603 T	15604 G	15617 T	15635 C	15649 A	15650 A	15659 T	15666 G	15703 T	15709 C	15720 G	15737 T		
X79547	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Reference sequences	
JN398377	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		
H1	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	4.5
H2	.	C	.	.	.	.	.	.	.	T	.	.	C	.	.	.	C	.	.	.	A	.	3	13.6
H3	.	C	.	A	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	A	C	3	13.6
H4	.	C	.	.	.	.	T	A	G	T	.	.	.	T	.	G	.	A	C	.	A	.	4	18.2
H5	.	C	.	.	.	.	T	A	G	T	.	.	.	.	.	G	.	A	.	.	A	.	2	9.1
H6	.	C	.	.	.	.	.	.	.	T	.	.	.	.	.	G	.	.	.	.	A	.	1	4.5
H7	.	C	.	.	.	.	.	A	.	T	.	.	.	.	.	.	.	.	.	.	A	.	2	9.1
H8	.	C	.	.	G	.	.	A	.	T	.	.	.	.	.	.	.	.	.	T	A	.	1	4.5
H9	C	C	G	.	T	.	.	.	.	.	C	.	.	.	G	.	.	.	.	.	A	.	1	4.5
H10	C	C	G	.	T	.	.	A	.	T	C	A	.	.	G	.	.	.	.	.	A	.	3	13.6
H11	C	C	G	A	T	.	.	A	.	T	C	A	.	.	G	.	.	.	.	.	A	.	1	4.5

Notes.

N, number of samples per haplotype; Freq. %, haplotypes frequency in percentage; X79547 and JN398377, ID Reference sequences.
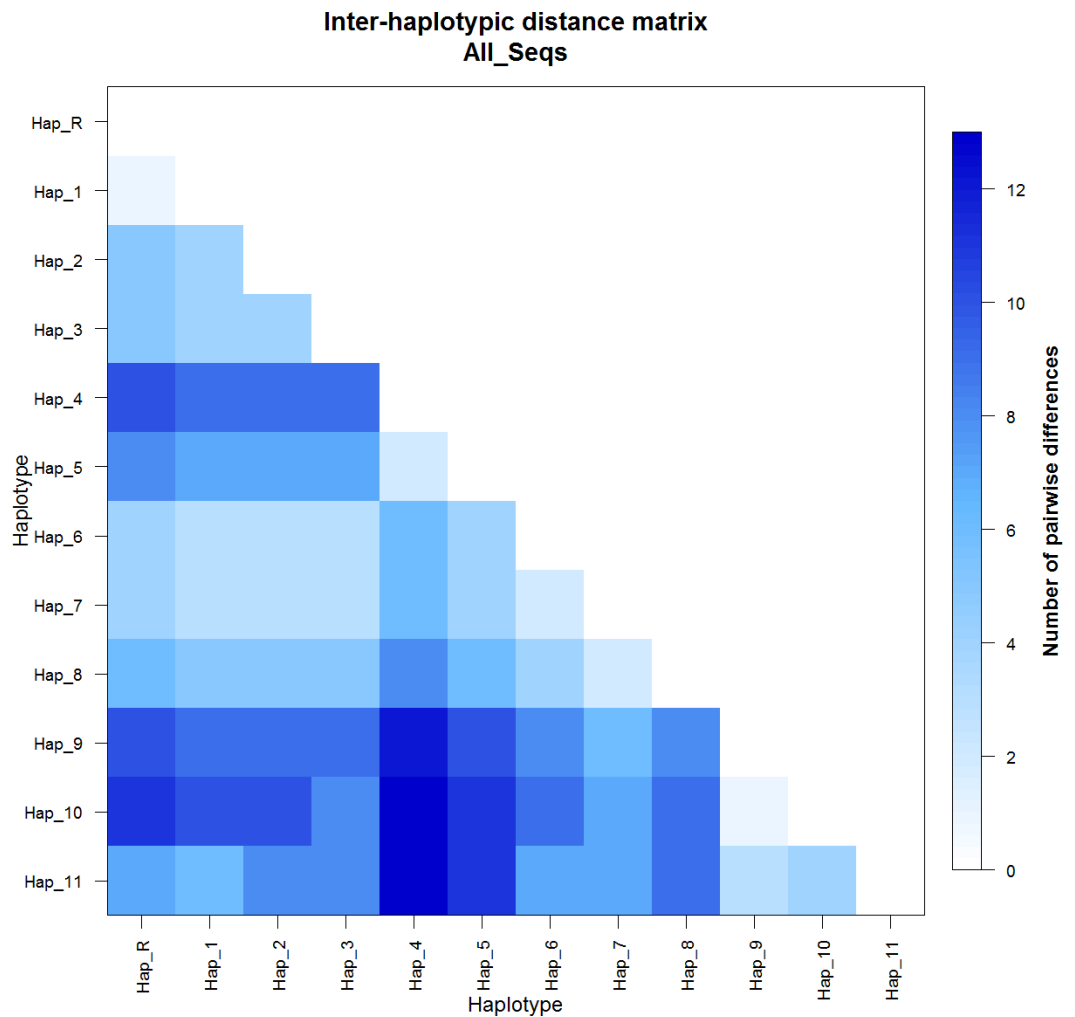
Figure 1 **Inter-haplotypic pairwise distance matrix based on haplotypes pairwise genetic differentiation, with the legend of the color code.** Hap_R = reference sequences X79547 and JN398377.

Full-size ⊡ DOI: 10.7717/peerj.4889/fig-1

among sequences were also defined in the NJ tree (Fig. 2B) and in the PCA analyses (Fig. 2C) that, nevertheless, are unable to dissect the variation among haplogroups as well as the MJN. Finally, the matrix of pairwise population genetic differentiation (F_{ST}) showed a remarkable similarity between our AKH and other AKH populations, as expected (Table S4).

The NJ tree based on Kimura-2-parameter model distances and performed by applying a cut-off value of 50 for bootstrap value significance clearly separated haplogroups L from the others (62% of bootstrap value) (Fig. 2B). Haplogroups A and Q are divided into a different branch of the dendrogram (Fig. 2B). Particularly, haplotypes H2, which belong to haplogroup A, group together with haplotypes belonging to haplogroups C in another branch of the tree. The H2 located in the same operational taxonomic unit (OTU) of H3 and H50 differs from these by one mutation 15,650 (H2 *vs* H3) and by two mutations 15,650 and 15,728 (H2 *vs* H50), respectively. The haplotype H42, formally belonging to
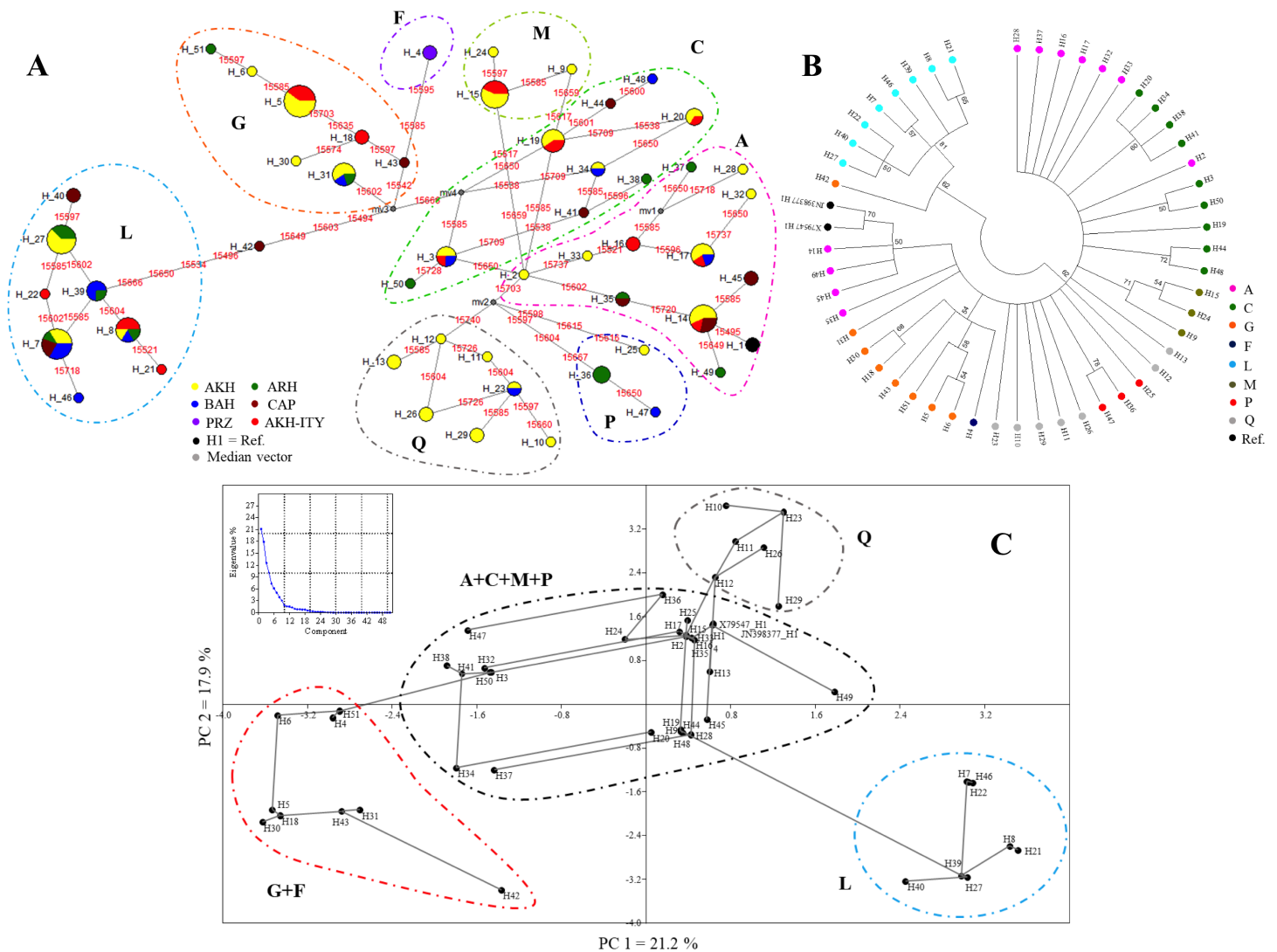
Figure 2 Clustering of haplogroups by MJN (A), NJ tree (B), PCA (C). (A) The haplotypes MJN based on the 124 horse mtDNA sequences and the two reference sequences (X79547 and JN398377). The numbers in red represent the position of nucleotide substitution; the black circle are the median vectors; circle are proportional to the numerosity of the haplotypes. AKH, Akhal-Teke; ARH, Arabian horse; BRH, Barb horse; CAP, Caspian Pony; PRZ, Przewalskii; H1, references. (B) The haplotypes NJ tree based on Kimura-2-parameter model distances. The numbers represent the robustness of the dendrogram after the cutoff bootstrap procedure was applied, with the legend of the haplogroup definition. (C) PCA analyses based on 51 haplotypes showing a minimum spanning tree. The graphic reports the percentage of variance explained by PCA components.

Full-size ⤢ DOI: 10.7717/peerj.4889/fig-2

haplogroups G in the PCA but separated with haplogroup L in NJ tree, will be considered as intermediate among haplogroups G and L as well defined in MNJ (Figs. 2A, 2B and 2C).

The PCA analysis showing a minimum spanning tree based on sequence haplotypes, is displayed in Fig. 2C. The graphic placed in Fig. 2C reports the percentage of variance explained by PCA components. The first component (PC1) that explains most of the variability (21.2%) clearly separated haplogroups G+F and L, whereas the second component (PC2) (17.9%) defines haplogroup Q, confirming the MJN and NJ tree

Table 3 Comparison with results reported by *Szontagh et al. (2005)*[a] and by *Jiskrová, Vrtková & Prausová (2016)*[b] using the same panel of 12 and 13 microsatellites.

Breed	NSTS	NS	NA	MNA	NE	H_o	H_e	F_{IS}
Akhal-Teke Italy	*12*	*95*	*78*	*6.5*	*3.07*	*0.636*	*0.645*	*0.009*
Akhal-Teke Turkmenian[a]	12	48	73	6.1	4.03	0.664	0.703	0.053
Akhal-Teke USA[a]	12	58	91	7.6	4.03	0.655	0.727	0.100
Akhal-Teke Iran[a]	12	19	70	5.8	4.20	0.728	0.725	−0.005
Akhal-Teke Italy	13	95	82	6.3	3.03	0.636	0.642	0.004
Akhal-Teke Czech Republic[b]	13	121	114	8.8	4.12	0.648	0.731	0.114
Akhal-Teke Russia[b]	13	152	74	5.7	3.43	0.669	0.668	−0.005
Akhal-Teke Estonia[b]	13	28	62	4.8	3.16	0.720	0.665	−0.103
Akhal-Teke Switzerland[b]	13	24	63	4.8	3.04	0.664	0.656	−0.029

Notes.

NSTS, number of microsatellites; NS, number of samples; NA, absolute number of alleles; MNA, Mean number of alleles; NE, effective number of alleles; H_o, observed heterozygosity; H_e, expected heterozygosity; F_{IS}, inbreeding coefficient.

evidence. Haplogroups A, C, M and P appear mixed and are not mutually well separated both on PC1 and on PC2. The PCA analysis confirms the central role of H2.

DISCUSSION

In Italy, AKH horses are reared by few breeders and are mainly trained for shows and endurance races. The small number of animals and the use of the same stallions for reproduction entail a potential risk of decreasing variability. Hence, we analyzed genetic variability by using both microsatellite DNA nuclear markers and mitochondrial DNA variations.

As mentioned above, AKH belong to the clade of Middle Eastern Breeds that involved Arabian horses, Barb horses, Turkoman horses and Caspian pony. The native area of the AKH is considered one of the two hotspots for horse diversity. In fact, the Caspian Area of Western Asia served as refugium for wild horses in the early and mid-Holocene when, due to a lack of suitable habitat as a result of forest expansion, the wild horses migrated to more comfortable areas for this steppe-adapted species (*Warmuth et al., 2011*). Later, around 3500 BC, the Caspian Area and the Kazakh steppe were horse domestication points from where horses spread across Eurasia (*Outram et al., 2009*; *Cieslak et al., 2010*; *Warmuth et al., 2012*).

The mean number of alleles (MNA) (6.5) obtained in Italian AKH was comparable to those reported by *Szontagh et al. (2005)* for the AKH population bred in Turkmenistan and in Turkoman horses, using the same panel of 12 microsatellites (Table 3).

The NE value indicates the actual impact of alleles on population variability, precisely, a small value reports an unbalanced frequency of the alleles at a certain locus. The NE is associated to the expected heterozygosity. Indeed, heterozygosity is highest when allelic frequencies are balanced, and every allele contributes to the absolute number of alleles detected per locus.

In the Italian AKH horses we identified a lower effective number of alleles (NE) compared to the other AKH populations. Indeed, in the Italian population the founder effect is still

having a strong impact on the population's genetic composition due to the low number of stallions and dams imported from abroad to constitute the first Italian nucleus in 1991–2000 (**Founders**). Particularly, the low number of stallions used in the last twenty years by breeders appears to have strongly influenced allele frequencies, as can be derived by the deviation from HW equilibrium for some microsatellite markers as shown in Table 1.

Several authors reported the influence of inbreeding, founder effect and population subdivision on LD. Nevertheless as pointed out by *Farnir et al. (2000)*, *McRae et al. (2002)* and *Miller et al. (2015)* after one generation, it can be observed that the major part of the alleles in most of the offspring could be in the same phase as in the parents. The LD here found suggests a possible bottleneck effect occurred when the Founder group has been imported in Italy (Table S2). Despite that, the Italian Akhal-Teke population maintained a large genetic variability and low degree of inbreeding over the years.

The MNA value for Italian AKH, comparing with values obtained by *Jiskrová, Vrtková & Prausová (2016)* using a panel of 13 microsatellites, is lower respect to the Czech Republic AKH (Table 3). Additionally, the effective number of alleles is similar to those found by *Jiskrová, Vrtková & Prausová (2016)* in the small AKH population bred in Estonia and in Switzerland (Table 3).

The overall F_{IS} values obtained using the 17 microsatellites panel (0.016) and both 12 and 13 microsatellites panels (0.009 and 0.004), revealed a low level of inbreeding, i.e., the Turkmenian population, indicating the fact that Italian breeders are prone to avoid mating between related individuals as much as possible. This trend can be highlighted by comparing the F_{IS} values in Italian horse to other populations (Table 3). US and Czech AKH horses appear to have a higher level of inbreeding, compared to Italian AKH, which reflect good breeding practice of Italian breeders.

Finally, H_o and H_e in the Italian population are similar to those observed in the other AKH populations (Table 3), showing that, overall, selection pressure for endurance and athleticism is not intense and genetic variability appear to be conserved in the worldwide AKH population.

The results of Italian AKH mtDNA analyses reveal the presence of 11 haplotypes with a high haplotype diversity (hd = 0.938) and a high nucleotide diversity ($\pi = 0.021$), indicating a high level of variability in the Italian population. Many authors report similar results for several horse breeds, specifically, in modern AKH (hd = 0.94; $\pi = 0.023$) (*Priskin et al., 2010*), Hungarian Gidran horse (hd = 0.913; $\pi = 0.021$ (*Sziszkosz et al., 2016*), Anatolian horse (hd = 0.98; $\pi = 0.026$) (*Koban et al., 2012*), Caspian horse (hd = 0.96; $\pi = 0.019$) (*Prystupa et al., 2012*) and in Chinese horse breeds (hd range 0.97–0.98; π range 0.019–0.024) (*Zhang et al., 2012*).

The high genetic variability of mtDNA is proof of the broad genetic base of founder mares in the AKH population, though only the sire lines were historically documented (http://www.akhal-teke.org/at-lines.html), whereas no information was recorded for the female lines.

As expected, we found a great similarity between our AKH sequences and all the AKH sequences from GenBank, in contrast with *Priskin et al. (2010)*, who identified significant differences in F_{ST} values ($F_{ST} = 0.075$, $P = 0.027$) among modern AKH, and the 18 Kazakh

AKH deposited by *McGahern et al. (2006)* in GenBank (Table S5). It is interesting to note the low F_{ST} values found with one of the two Caspian Pony populations (AKH_ITY *vs.* CAP_a $F_{ST} = 0$) and with the two Arab horse populations (AKH_ITY *vs.* ARH_a $F_{ST} = 0$; AKH_ITY *vs.* ARH_ITY $F_{ST} = 0.027$) (Table S5). These small F_{ST} values confirmed that AKH belong to the same phylogenetic group of the Caspian Pony and Arab horse as reported by several authors (*Cothran, Juras & Macijauskiene, 2005*; *McGahern et al., 2006*; *Cieslak et al., 2010*; *Achilli et al., 2012*).

Haplotype H2, which belongs to haplogroup A, both on MJN and on PCA, appears to be the central point, and it is the link among all the haplogroups (Fig. 2A and 2C). The central role of H2 is due to the two parsimony sites 15,602 and 15,720 and the singleton site 15,495 that are diagnostic mutational motives for more than one haplogroup (*McGahern et al., 2006*; *Achilli et al., 2012*).

Haplogroups A, C, M, P and Q are clearly separated on the MJN, but the relationships among these haplogroups are not well defined in the PCA and in the NJ tree. In fact, the PCA is able to separate only the haplogroup Q form others, whereas in the NJ tree they are mixed with the P haplogroup (Fig. 2B and 2C). The 22 mtDNA sequences of Italian AKH horses were grouped in the five haplogroups A, C, G, L and M.

Consistently with literature, (*McGahern et al., 2006*; *Cieslak et al., 2010*; *Achilli et al., 2012*), haplogroup A **is** the most frequent (23.5%) in modern Middle Eastern horse populations as well as in the Western central Eurasia, one of the primary areas of horse domestication event, that comprises the AKH origin area, (*Warmuth et al. (2012)*. In agreement with *Achilli et al. (2012)*), haplogroup F was represented only by the two Przewalskii samples sequenced (1.9% of analyzed sequences). The separation of haplogroup F from others appears clear in the MJN, while it is grouped with G in NJ and PCA analyses.

The four median vectors identified by haplogroups C-G and G-L should be interpreted as incomplete sequences that limited the connection under the maximum parsimony hypothesis and it could be explained by the presence of ancestors not included in the analyses and with extinct maternal sequences. In fact, without the median vector the shortest connection among analyzed sequences would be lost (http://www.fluxus-engineering.com/Network5000_user_guide.pdf). Hence, the haplotype H42, which formally belonged to haplogroups G, has to be considered as intermediate among haplogroups G and L (Fig. 2A). The three different methods used to evaluate the genetic relationships among haplogroups clearly indicate that haplogroup L is separated from the other by 4 parsimony informative mutations. In the overall mtDNA sequences considered herein, the G and the L ones have a frequency of 15.6%. *Achilli et al. (2012)* reported a frequency of 8.3% for haplogroup G and 22.4% for haplogroup L in the Middle East. This haplogroup (L) is considered ancient, as it has been reported in horses in the Iberian Peninsula since the Neolithic Bronze Age (*McGahern et al., 2006*; *Cieslak et al., 2010*; *Achilli et al., 2012*).

CONCLUSION

This study evaluated the genetic variability of the AKH horse bred in Italy since 2000. Despite the low number of horses and the use of few stallions, during these 17 years, a

large genetic variability with a low degree of inbreeding has maintained the breed over the years, confirming a well-balanced breeding practice. In fact, microsatellite markers showed a high level of heterozygosity that was comparable to findings in AKH populations reared in other countries.

The mtDNA analyses on the Italian AKH confirmed the breed's high matrilineal variability, and added four new haplotypes that were never described before in the breed. The results also confirmed the close genetic relationship with the other AKH populations and the Middle Eastern horses represented by Arabian horse, Caspian Pony and Barb horse breeds, estimated as one of the most ancient and primitive clades of *Equus caballus*.

ACKNOWLEDGEMENTS

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Maria C. Cozzi and Maria G. Strillacci conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Paolo Valiati and Elisa Rogliano performed the experiments, analyzed the data, authored or reviewed drafts of the paper, approved the final draft.
- Alessandro Bagnato prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Maria Longeri conceived and designed the experiments, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.

Animal Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

The bio-banked samples were approved by the Ethics Committee of Università degli Studi di Milano Prot. OPBA_56_2016. The permits state that no additional permit is needed to use DNA samples, and blood samples are part of the historical samples and collections, and maintained for years in the University of Milan bio-bank.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:
  GenBank: MF580226–MF580244.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.4889#supplemental-information.

REFERENCES

Achilli A, Olivieri A, Soares P, Lancioni H, Hooshiar Kashani B, Perego UA, Nergadze SG, Carossa V, Santagostino M, Capomaccio S, Felicetti M, Al-Achkar W, Penedo MC, Verini-Supplizi A, Houshmand M, Woodward SR, Semino O, Silvestrelli M, Giulotto E, Pereira L, Bandelt HJ, Torroni A. 2012. Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proceeding of the National Academy Science of the United States of America* **109**:2449–2454 DOI 10.1073/pnas.1111637109.

Beja-Pereira Al, England PR, Ferrand N, Jordan S, Bakhiet AO, Abdalla MA, Mashkour M, Jordana J, Taberlet P, Luikart G. 2004. African origin of domestic donkey. *Science* **304**:1781 DOI 10.1126/science.1096008.

Bömcke E, Gengler N, Cothran EG. 2011. Genetic variability in the Skyros pony and its relationship with other Greek and foreign horse breeds. *Genetics and Molecular Biology* **34**:68–76 DOI 10.1590/S1415-47572010005000113.

Cieslak J, Wodas L, Borowska A, Cothran EG, Khanshour AM, Mackowski M. 2017. Characterization of the Polish Primitive Horse (Konik) maternal lines using mitochondrial D-loop sequence variation. *PeerJ* **5**:e3714 DOI 10.7717/peerj.3714.

Cieslak M, Pruvost M, Benecke N, Hofreiter M, Morales A, Reissmann M, Ludwig A. 2010. Origin and history of mitochondrial DNA lineages in domestic horses. *PLOS ONE* **5**:e15311 DOI 10.1371/journal.pone.0015311.

Cothran EG, Van Dyk E, Van der Merwe FJ. 2001. Genetic variation in the feral horses of the Namib Desert. Namibia. *Journal of the South African Veterinary Association* **72(1)**:18–22.

Cothran EG, Juras R, Macijauskiene V. 2005. Mitochondrial DNA D.loop sequence variation among 5 maternal lines of the Zemaitukai horse breed. *Genetics and Molecular Biology* **28(4)**:677–681 DOI 10.1590/S1415-47572005000500006.

Cozzi MC, Strillacci MG, Valiati P, Bighignoli B, Cancedda M, Zanotti M. 2004. Mitochondrial D-loop sequence variation among Italian horse breeds. *Genetics Selection Evolution* **36**:663–672 DOI 10.1186/1297-9686-36-6-663.

Cozzi MC, Valiati P, Cherchi R, Gorla E, Prinsen RT, Longeri M, Bagnato A, Strillacci MG. 2017. Mitochondrial DNA genetic diversity in six Italian donkey breeds (*Equus asinus*). *Mitochondrial DNA Part A* **28**:1–10 DOI 10.1080/24701394.2017.1292505.

Elsner J, Hofreiter M, Schibler J, Schlumbaum A. 2017. Ancient mtDNA diversity reveals specific population development of wild horses in Switzerland after the Last Glacial Maximum. *PLOS ONE* **12**:e0177458 DOI 10.1371/journal.pone.0177458.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**:564–567 DOI 10.1111/j.1755-0998.2010.02847.x.

Farnir F, Coppieters W, Arranz JJ, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mni M, Nezer C, Simon P, Vanmanshoven P, Wagenaar D, Georges M. 2000. Extensive genome-wide linkage disequilibrium in cattle. *Genome Research* **10**:220–227 DOI 10.1101/gr.10.2.220.

Goudet J. 2002. FSTAT, a program to estimate and test gene diversities and fixation indices. *Available at http://www.unil.ch/izea/softwares/fstat.html* (accessed on 09 July 2017).

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**:95–98.

Hammer Ø, Harper DAT, Ryan PD. 2001. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica* **4**:1–9.

Hristov P, Yordanov G, Ivanova A, Mitkov I, Sirakova D, Mehandzyiski I, Radoslavov G. 2016. Mitochondrial diversity in mountain horse population from the South-Eastern Europe. *Mitochondrial DNA a DNA Mapping Sequencing and Analaysis* **1**:1–6 DOI 10.1080/24701394.2016.1186667.

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. *Proceeding of the National Academy of Sciences of the United States of America* **99**:10905–10910 DOI 10.1073/pnas.152330099.

Jiskrová I, Vrtková I, Prausová M. 2016. Genetic diversity of populations of Akhal-Teke horses from the CzechRepublic, Russia, Estonia and Switzerland. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* **64**:461–466 DOI 10.11118/actaun201664020461.

Kimura M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**:111–120 DOI 10.1007/BF01731581.

Koban E, Denizci M, Aslan O, Aktoprakligil D, Aksu S, Bower M, Balcioglu BK, Ozdemir Bahadir A, Bilgin R, Erdag B, Bagis H, Arat S. 2012. High microsatellite and mitochondrial diversity in Anatolian native horse breeds shows Anatolia as a genetic conduit between Europe and Asia. *Animal Genetics* **43**:401–409 DOI 10.1111/j.1365-2052.2011.02285.x.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**:1870–1874 DOI 10.1093/molbev/msw054.

Kusza S, Priskin K, Ivankovic A, Jedrzejewska B, Podgorski T, Jávor A, Mihók S. 2013. Genetic characterization and population bottleneck in the Hucul horse based on microsatellite and mitochondrial data. *Biological Journal of the Linnean Society* **109**:54–65 DOI 10.1111/bij.12023.

Leisson K, Alev K, Kaasik P, Jaakma Ü, Seene T. 2011. Myosin heavy chain pattern in the Akhal-Teke horses. *Animal* **5**:658–662 DOI 10.1017/S1751731110002375.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**:1451–1452 DOI 10.1093/bioinformatics/btp187.

Librado P, Gamba C, Gaunitz C, Der Sarkissian C, Pruvost M, Albrechtsen A, Fages A, Khan N, Schubert M, Jagannathan V, Serres-Armero A, Kuderna LFK7, Povolotskaya IS, Seguin-Orlando A, Lepetz S, Neuditschko M, Thèves C, Alquraishi S, Alfarhan AH, Al-Rasheid K, Rieder S, Samashev Z, Francfort HP, Benecke N, Hofreiter M, Ludwig A, Keyser C, Marques-Bonet T, Ludes B, Crubézy E, Leeb T, Willerslev E, Orlando L. 2017. Ancient genomic changes associated with domestication of the horse. *Science* **356(6336)**:442–445 DOI 10.1126/science.aam5298.

Lippold S, Matzke NJ, Reissmann M, Hofreiter M. 2011. Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. *BMC Evolutionary Biology* **11**:328 DOI 10.1186/1471-2148-11-328.

Luis C, Juras R, Oom MM, Cothran EG. 2007. Genetic diversity and relationships of Portuguese and other horse breeds based on protein and microsatellite loci variation. *Animal Genetics* **38**:20–27 DOI 10.1111/j.1365-2052.2006.01545.x.

McGahern A, Bower MA, Edwards CJ, Brophy PO, Sulimova G, Zakharov I, Vizuete-Forster M, Levine M, Li S, MacHugh DE, Hill EW. 2006. Evidence for biogeographic patterning of mitochondrial DNA sequences in Eastern horse populations. *Animal Genetics* **37**:494–497 DOI 10.1111/j.1365-2052.2006.01495.x.

McRae AF, Mcewan JC, Dodds KG, Wilson T, Crawford AM, Slate J. 2002. Linkage disequilibrium in domestic sheep. *Genetics* **160**:1113–1122.

Miller JM, Poissant J, Malenfant RM, Hogg JT, Coltman DW. 2015. Temporal dynamics of linkage disequilibrium in two populations of bighorn sheep. *Ecology and Evolution* **5(16)**:3401–3412 DOI 10.1002/ece3.1612.

Outram AK, Stear NA, Bendrey R, Olsen S, Kasparov A, Zaibert V, Thorpe N, Evershed RP. 2009. The earliest horse harnessing and milking. *Science* **323**:1332–1335 DOI 10.1126/science.1168594.

Petersen JL1, Mickelson JR, Cothran EG, Andersson LS, Axelsson J, Bailey E, Bannasch D, Binns MM, Borges AS, Brama P, Da Câmara Machado A, Distl O, Felicetti M, Fox-Clipsham L, Graves KT, Guérin G, Haase B, Hasegawa T, Hemmann K, Hill EW, Leeb T, Lindgren G, Lohi H, Lopes MS, McGivney BA, Mikko S, Orr N, Penedo MC, Piercy RJ, Raekallio M, Rieder S, Røed KH, Silvestrelli M, Swinburne J, Tozaki T, Vaudin M, Wade CM, McCue ME. 2013. Genetic diversity in the modern horse illustrated from Genome-Wide SNP data. *PLOS ONE* **8**:e54997 DOI 10.1371/journal.pone.0054997.

Priskin K, Szabó K, Tömöry G, Bogácsi-Szabó E, Csányi B, Eördögh R, Downes CS, Raskó I. 2010. Mitochondrial sequence variation in ancient horses from the Carpathian Basin and possible modern relatives. *Genetica* **138**:211–218 DOI 10.1007/s10709-009-9411-x.

Prystupa JM, Hind P, Cothran EG, Plante Y. 2012. Maternal lineages in native Canadian equine populations and their relationship to the Nordic and Mountain and Moorland pony breeds. *The Journal of Heredity* **103**:380–390 DOI 10.1093/jhered/ess003.

Raymond M, Rousset F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *The Journal of Heredity* **86**:248–249 DOI 10.1093/oxfordjournals.jhered.a111573.

Royo LJ, Alvarez I, Beja-Pereira A, Molina A, Fernández I, Jordana J, Gómez E, Gutiérrez JP, Goyache F. 2005. The origins of Iberian horses assessed via mitochondrial DNA. *The Journal of Heredity* **96**:663–669 DOI 10.1093/jhered/esi116.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**:406–425.

Swinburne JE, Boursnell M, Hill G, Pettitt L, Allen T, Chowdhary B, Hasegawa T, Kurosawa M, Leeb T, Mashima S, Mickelson JR, Raudsepp T, Tozaki T, Binns M. 2006. Single linkage group per chromosome genetic linkage map for the horse, based on two three-generation, full-sibling, crossbred horse reference families. *Genomics* **87(1)**:1–29 DOI 10.1016/j.ygeno.2005.09.001.

Sziszkosz N, Mihók S, Jávor A, Kusza S. 2016. Genetic diversity of the Hungarian Gidran horse in two mitochondrial DNA markers. *PeerJ* **2(4)**:e1894.

Szontagh B, Bán I, Bodó E, Cothran G, Hecker W, Józsa CS, Á Major. 2005. Genetic diversity of the Akhal-Teke horse breed in Turkmenistan based on microsatellite analysis. In: Bodó I, Alderson L, Langlois B, eds. *Conservation genetics of endangered horse breeds. EAAP publication N 116.* Wageningen: Wageningen Academic Publishers, 123–129.

Vilà C, Leonard JA, Gotherstrom A, Marklund S, Sandberg K, Liden K, Wayne RK, Ellegren H. 2001. Widespread origins of domestic horse lineages. *Science* **291**:474–477 DOI 10.1126/science.291.5503.474.

Warmuth V, Eriksson A, Bower MA, Barker G, Barrett E, Hanks BK, Li S, Lomitashvili D, Ochir-Goryaeva M, Sizonov GV, Soyonov V, Manica A. 2012. Reconstructing the origin and spread of horse domestication in the Eurasian steppe. *Proceedings of the National Academy of Sciences of the United States of America* **22**:8202–8206.

Warmuth V, Eriksson A, Bower MA, Cañon J, Cothran EG, Distl O, Glowatzki-Mullis ML, Hunt H, Luís C, Do Mar Oom M, Yupanqui IT, Ząbek T, Manica A. 2011. European domestic horses originated in two holocene refugia. *PLOS ONE* **6**:e18194 DOI 10.1371/journal.pone.0018194.

Xu X, Arnason U. 1994. The complete mitochondrial DNA sequence of the horse Equus caballus: extensive heteroplasmy of the control region. *Gene* **148**:657–662.

Yeh FC, Yang R, Boyle TJB. 1997. POPGENE: Microsoft Window-based freeware for population genetic analysis. Version 1.31. *Available at https://sites.ualberta.ca/~fyeh/popgene.html* (accessed on 09 July 2017).

Zhang T, Lu H, Chen C, Jiang H, Wu S. 2012. Genetic Diversity of mtDNA D-loop and maternal origin of three chinese native horse breeds. *Asian-Australasian Journal of Animal Sciences* **25(7)**:921–926 DOI 10.5713/ajas.2011.11483.