Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/6/38/eabb0030/DC1

Supplementary Materials for

Ancient DNA shows domestic horses were introduced in the southern Caucasus and Anatolia during the Bronze Age

Silvia Guimaraes, Benjamin S. Arbuckle, Joris Peters, Sarah E. Adcock, Hijlke Buitenhuis, Hannah Chazin, Ninna Manaseryan, Hans-Peter Uerpmann, Thierry Grange, Eva-Maria Geigl*

*Corresponding author. Email: eva-maria.geigl@ijm.fr

Published 16 September 2020, *Sci. Adv.* **6**, eabb0030 (2020)
DOI: 10.1126/sciadv.abb0030

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/6/38/eabb0030/DC1)

Supplementary Materials

1. Description of the archaeological sites

The locations of the archaeological sites are shown in Fig. S1.

1.1. Anatolia

Aşıklı Höyük

Hijlke Buitenhuis, Joris Peters

Aşıklı Höyük is an aceramic Neolithic settlement mound located next to the Melendiz River. It covers an area of 3.5-4 ha, rising 15.35 m above the Melendiz Ülain at the north and 13.16 m at the south. Under

the direction of Prof. Dr. U. Esin (University of Istanbul), excavations were carried out from 1989 until 2003. Renewed excavations directed by Prof. Dr. M. Özbaşaran (University of Istanbul) started in 2009 and continued since then. Five levels, designated as 1-5 from top to bottom respectively, could be documented so far. The earliest settlement phase excavated thus far is represented by Level 5. Radiocarbon dates allow concluding that early site occupation dates to ~8,400 BCE (*1, 2*). The faunal remains of the site are currently analyzed by a team of zooarchaeologists including H. Buitenhuis, N. Munro, J. Peters, N. Pöllath, and M.C. Stiner. A most important finding is the fact that caprine management was already practiced by the site inhabitants (*3-6*) and developed further on during the occupation in the later levels to a full caprine husbandry. So far, two species of wild equids have been evidenced in the archaeofauna. Remains of wild horses (*E. ferus*) form a small part of this assemblage, whereas the majority belongs to hemione-like animals, likely *E. hydruntinus* (*4, 7*). The faunal assemblage is curated on site at the Aşıklı Höyük dighouse, Kızılkaya village, Aksaray, Turkey. All specimens were exported with the permission of the Aksaray Museum and the Turkish Ministry of Culture and Tourism.

Acemhöyük

Benjamin Arbuckle

Acemhöyük is a large mound located at an elevation of 950 meters on the alluvial fan of the Melendiz river near the Tüz Gölü (Salt Lake) in Central Turkey. The mound is approximately 800 x 600 meters in dimension rising 20 meters above the surrounding Yesilova plain and represents the remains of a large, fortified Bronze Age urban center. The site has been excavated since 1962 by Dr. N. Özgüç, Ankara University in association with the Turkish Historical Society and the General Directorate of Antiquities and Museums and by Prof. Dr. A. Öztan of the same institution since 1989 (*8, 9*). Acemhöyük consists of 12 levels with deposits representing Chalcolithic to Medieval occupations.

The faunal assemblage at Acemhöyük is dominated by domestic sheep, goats, cattle and pigs with equids representing a tertiary resource (*10, 11*). The Acemhöyük faunal assemblage is curated on site at the Acemhöyük dighouse, Yeşilova, Aksaray, Turkey. All specimens were exported with the permission of the Konya Museum and the Turkish Ministry of Culture and Tourism.

24 equid specimens were submitted for ancient DNA analysis. Twenty of these samples derive from deposits stratigraphically assigned to levels XI, X, VII, V and IV representing the late Early Bronze Age (EBA) and earliest Middle Bronze Age (MBA) (associated radiocarbon dates c. 2100-1900 BCE). Many of these specimens were identified only to the genus level (*Equus* sp.) and could potentially represent wild horse (*Equus ferus*), domestic horse (*Equus caballus*), wild hemione/hydruntine (*Equus hemionus hydruntinus*),

domestic donkey (*Equus asinus*), or one of several hybrids (horse x donkey, donkey x horse, donkey x hemione). Four of these specimens (AC14371, AC14379, AC14380, AC14427) derive from equid burials and are identified based on dental morphology and body size as donkeys or hemiones. An additional 7 of these late EBA/early MBA specimens were also identified as donkeys or hemiones rather than horses. One specimen, identified as *Equus sp.*, derives from level III, the Assyrian Colony period (c. 1700s BC). Two specimens (*Equus caballus*) are from Hellenistic contexts (level I), while the last specimen (*Equus caballus*) is thought to come from a pit perhaps dating to the Ottoman period.

Çadır Höyük

Benjamin Arbuckle

Çadır Höyük is a 32-meter-high mound located in the Kanak Su basin situated in the Yozgat Province of north central Turkey. The modestly sized 240 x 185-meter mound boasts an occupational history spanning 6000 years (ca. 5200 BCE to the thirteenth century CE). This sequence has been explored since 1993 by the Çadır Höyük Archaeological Project led by Ron Gorny and more recently Greg McMahon and Sharon Steadman (*12, 13*). The Çadır Höyük faunal assemblage is curated on site at the Çadır Höyük dighouse, Peyniryemez, Sorgun, Turkey. Specimens were exported with the permission of the Yozgat Archaeology Museum and the Turkish Ministry of Culture and Tourism.

21 specimens derived from throughout the Çadır Höyük occupational sequence were submitted for ancient DNA analysis. These include two of specimens (CD2017, CD2015) from the Deep Sounding which were recovered from a Middle Chalcolithic context with an associated radiocarbon date of 4500 cal BCE (Beta-418460). Due to this early date, these specimens are thought to represent wild Anatolian horse (*Equus ferus*). Another (wild?) horse specimen (CD1875) was recovered from a Late Chalcolithic context with an associated radiocarbon date placing it in the late fourth millennium BCE (Beta418461: 3350 cal BCE). Six specimens derive from Early and Late Bronze Age contexts; ten specimens come from Iron Age deposits and include horse, donkey and *Equus* sp identifications. Finally, two horse specimens derive from Byzantine deposits.

Boğazköy-Hattusa

Benjamin Arbuckle

Boğazköy-Hattusa is an UNESCO World Heritage site located in north central Turkey (Çorum province) that represents the remains of a major urban center dating to the Middle and Late Bronze Age (2000-1200 BCE). Boğazköy is perhaps best known as the capital city of the Hittite empire and has been excavated by

the Deutsches Archäologisches Institut under the supervision of Dr. Andreas Schachner. One specimen was provided from Hattusa for ancient DNA analysis. This specimen, an incisor tooth likely representing a domestic donkey, derives from Late Bronze Age (Hittite) deposits (excavation year 2013, context 407, bag#820).

Köşk Höyük

Benjamin Arbuckle

The site of Köşk Höyük (originally Köşk-Pınar (*14*)) represents the remains of a small agro-pastoral village located on the eastern edge of the Bor Plain in central Turkey at an elevation of 1400 meters (*15*). The site consists of a shallow mound, about half a hectare in area, situated on the edge of a limestone ridge overlooking a permanent spring. The location of the site provides ready access to the resources of the Konya-Ereğli Plain to the west, the foothills and uplands of the Taurus Mountains to the east and south, and the most direct route from the west and south to the obsidian sources of the nearby Melendiz Mountains. Excavations at Köşk began with a series of soundings done by Dr. Nimet Özgüç (Ankara University) in 1981, followed by a more extensive excavation campaign led by Dr. U. Silistreli, in association with the Nigde Museum from 1983 to 1992, and finally by Prof Dr. Aliye Öztan from 1995 to 2012 (*15, 16*). The stratigraphy of Köşk has been divided into prehistoric levels V-I. Stratigraphic levels V-II have been radiocarbon dated to between 6200-5500 cal BCE and represent the Early Chalcolithic period in central Anatolia. In this period the faunal assemblage indicates that although domestic sheep and goats are the most abundant mammalian taxa, wild taxa are abundant as well. Wild cattle, boar, as well as hare and foxes were hunted and trapped at the site. However, wild equids including both wild horse (*Equus ferus*) and European wild ass (*Equus hemionus hydruntinus*) are the most abundant wild taxa representing a surprising 15-38% of the fauna from EC contexts (*17*).

24 specimens from Köşk Höyük were sent for ancient DNA analysis. These specimens derive from the Early Chalcolithic levels IV and III representing the Early Chalcolithic period. 14 specimens were identified based on dental and skeletal morphology as Anatolian wild horse (*Equus ferus*); five specimens were identified as European wild ass (*Equus hemionus hydruntinus*); while the remaining five specimens were identified to genus as *Equus* sp (Arbuckle). All specimens were exported with the permission of the Niğde Archaeology Museum and the Turkish Ministry of Culture and Tourism.

Troy

Hans-Peter Uerpmann

The excavation of the site of Troy, located close to the Dardanelles under the hill of Issarlik in Western Turkey, has identified a succession of 9 towns and 46 occupational levels. The complex stratigraphy extends over 20 m. The bones used for the present study originate from three different levels, Troy I, Troy IV and Troy VI. Troy I is a small fortified village and corresponds to the first human occupation of the site during the Early Bronze Age. The bone samples analysed in the present study originate from Troy IV and were assigned to *E. caballus*. Troy IV is dated to the Middle Bronze Age and corresponds to the phase of decline of the town whose subsistence mode seems to have been based on agriculture and handcraft rather than trading. Troy VI corresponds to the golden age of the town at the end of the Bronze Age as witnessed by the construction of palaces.

Only a fraction of the animal remains found during the ongoing excavations have been studied. Nevertheless, the archaeozoological data of more than 120,000 bone finds are stored in a data-base (*18*).

Lidar Höyük

Joris Peters

Southwest of the village Lidar (SE-Anatolia, Turkey) lies the impressive settlement mound of Lidar Höyük, measuring 200 by 240 m at the base and rising 23 m above the Euphrates River terrace. Situated near a major passage across the Euphrates River, the site of Lidar Höyük was already inhabited in the Early Bronze Age. Although occupation lasted until Medieval Times, the site was not occupied permanently, as the archaeological record shows.

Excavations carried out between 1979 and 1986 under the direction of Prof. Dr. H. Hauptmann† produced a rich archaeofauna totalling 62,453 vertebrate remains (*19*). 3.8% of this assemblage pertains to equids, remains of donkeys outnumbering those of horses and hemiones. Difficulties arose to assign some larger specimens either to (small) horse, mule or Asiatic wild ass (*19*). The 1987 excavation produced another sample of equid bone, from which specimens were selected for ancient DNA analysis, including a Bronze Age horse radius collected in Area S47.

1.2. Caucasus

Georgia

Didi Gora

Hans-Peter Uerpmann

This site in the province of Kachetia is located on the plain of the Alazani in Eastern Georgia between the Black and the Caspian Seas. In the past it was probably located on a river bank. It comprises a history of three millennia, from the Early Bronze Age (c. 2800 BCE) to the Early Iron Age (c. 1,000 BCE) and is one of the largest settlement hills during the Neolithic and Bronze Age in this region. Archaeological evidence points to a semi-nomadic, seasonal lifestyle of the inhabitants, to the exploitation of ore deposits and to extensive trading contacts with other regions in the Near East. The Bronze Age layers are poor in archaeological artifacts except for obsidian flints, but very rich in faunal remains.

These faunal remains, which were analyzed in the present study, are very well preserved suggesting a rapid burial. The genetically analyzed equid remains had been attributed to *E. caballus* based on the relative size of the bones and on their weight since the bones were heavily fragmented (*20*).

Tqisbolu Gora

The site of Tqisbolu Gora was occupied between 3,800 and 2,800 years ago and was thus partly contemporaneous to Didi Gora. The youngest layers of Didi Gora correspond to Iron Age layers in Tqisbolu Gora. Three equid species were identified on the two sites. The remains of domestic horses (*E. caballus*) are bigger than the remains of the small equids from the site. The latter could belong to *E. hemionus*, *E. hydruntinus* or to the domestic donkey *E. asinus* (*20*).

In the present study, four remains were analyzed genetically. Three of them had been attributed to *E. caballus* and one to *E. hydruntinus* based on the relative size of the bones.

Armenia

Lusakert

Nina Manaseryan

The analyzed horse remains originate from the Middle Palaeolithic Cave Lusakert 1 in the Hrazdan valley of Armenia.

Mets Sepasar

Nina Manaseryan

Mets Sepasar is a Bronze Age habitat and necropolis in Shirak. The mountain Mets Sepasar is located on the Shirak plateau in the northwest of Armenia. The archaeological studies, led by L.Yeganyan, started out from the foothills of the mountain. As a result of these excavations an Early Bronze Age settlement has been discovered. In the middle of the monument, in a volcanic cone, a sanctuary was found showing traces of festivities accompanied by sacrifices.

Tsaghkahovit and Gegharot

Hannah Chazin

Gegharot and Tsaghkahovit are multi-phase sites located in the Tsaghkahovit Plain, Armenia and have been the site of ongoing archaeological research by the joint Armenian-American Project ArAGATS (*21-23*). Gegharot was occupied during both the Early and Late Bronze Age. The Late Bronze Age occupation consisted of a walled fortress and nearby cemetery. Tsaghkahovit was occupied during the Late Bronze Age, as well as during the Achaemenid period. The Late Bronze Age occupation at Tsaghkahovit consisted of a walled fortress, extra-mural room blocks, and an associated cemetery. Two of the equid specimens were from Late Bronze Age tombs (Gegharot Kurgan 1 and Tsaghkahovit BC12), and the remainder were from Late Bronze middens excavated in the Tsaghkahovit Residential Complex. Equid specimens from these sites were identified by Hannah Chazin and Belinda Monahan.

2. Paleogenetic Analysis

Rationale of the experimental approach

Due to the age of the archaeological samples and the climatic conditions of the area and to the fact that most of the bones that were appropriate to answer the scientific questions were not petrous bones, DNA was poorly preserved. Thus, we used a method that we have developed for poorly preserved archaeological specimens (*24*). This powerful multiplex PCR assay is highly sensitive when optimized primers are used and allows high-throughput genotyping of a large sample number since it is coupled to next-generation sequencing (*24*). We used it previously successfully to genotype various difficult ancient samples (*24-27*). This approach is more efficient and cost-effective than sequence capture, in particular for samples with highly degraded DNA, and is informative enough for the specific questions addressed in this study. Here, we optimized the primers *in silico* and *in vitro*, i.e., tested and optimized the PCR conditions prior to the analysis of the ancient specimens. We amplified the ancient DNA extracts in the

presence of dUTP and degraded products of previous PCR amplifications using UDG (28). This UDG treatment eliminates in addition the products of cytosine deamination, thus increasing the accuracy of retrieved sequences (*28*). Moreover, prior to each PCR amplification, we eliminated contaminating DNA molecules in reagents using the UVD decontamination procedure (*29*). This multi-level decontamination procedure enabled us to produce reliable, authentic amplicons. Finally, we performed independently each PCR amplification three times and sequenced the bulk PCR products using NGS to determine the authentic sequences.

All pre-PCR procedures were carried out in the high containment facility of the Jacques Monod Institute in Paris physically separated from areas where modern samples are analyzed and dedicated exclusively to ancient DNA analysis using the strict procedures for contamination prevention previously described (*29, 30*).

2.1. DNA Extraction

The external surface of the specimens was removed with a sterile blade to minimize the environmental contamination. For each bone sample, roughly 0.2 g was ground to a fine powder in a freezer mill (Freezer Mill 6750, Spex Certiprep, Metuchen, NJ), which was then suspended in 2mL of extraction buffer containing 0.5M EDTA, 0.25M Di-sodium hydrogen phosphate (Na_2HPO_4), pH 8.0, 0.14M 2-mercaptoethanol and 0.25 mg/mL of proteinase K and incubated under agitation at 37°C for 48 hours. Blank extractions were carried out for each extraction series. Samples were then centrifuged and the supernatant was purified with a modified protocol of the Qiaquick PCR Purification Kit (Qiagen, Hilden, Germany) using the manifold device, as previously described (*30-32*).

2.2. Multiplex PCR Amplification

2.2.1. Mitochondrial DNA

Primers were designed in order to cover the complete mitochondrial genome with minimal primer dimer propensity from a multiple alignment of all the equids mitogenomes sequences present in Genbank in 2014 using the software Oligo 7 as previously described (*24*). The primers were then tested for efficiency and dimer formation using quantitative real-time PCR (qPCR) (*24*). We selected a total of 20 primer pairs (see paragraph SI 3.1.). To assess the phylogenetic consistency of our molecular assay with the selected fragments, we built a maximum likelihood tree with PHYML (*33*) for both the complete mitogenomes (GTR + G + I) and the concatenation of the 20 amplified fragments without primers (TN93 + G + I) (*34*). This

confirmed the preservation of the salient features of the tree topology when using only the selected fragments (Fig. 1). We optimized the multiplex PCR as described before (*24*) and the primer pairs were combined in three different multiplex reactions (see paragraph SI 3.4.1.). To protect against cross-contamination between samples, we used the UNG-coupled PCR system [40]. Each reaction was carried out in a 30 µl reaction volume with a final composition of 50 mM Tris/HCl, 10 mM KCl, 5 mM $(NH_4)_2SO_4$, 4 mM $MgCl_2$, pH 8.3 (@25°C), 1 mg/ml BSA (bovine serum albumin), 0.15 µM of each primer, 0.25 mM dA/G/CTPs and 0.50 mM dUTP, 2 U of Fast Start DNA polymerase, 0.5 U of uracil-N-Glycosylase (UNG) and 4 µl of DNA. The $MgCl_2$ solution, reaction buffer and BSA were decontaminated by exposing the solutions in UV-pervious tubes (Qubit®, Life Technologies) to UV light for 10 minutes at short distance as described (*29*). A negative control was performed for each multiplex reaction that was processed in the same way as the samples throughout the whole experimental procedure. The cycling program consisted of 15 minutes at 37°C (carry-over contamination prevention and aDNA damage products through digestion by UNG of dUTP-labeled amplicons and of cytosine deamination products), 95°C for 10 minutes (inactivation of UNG and activation of the Fast Start DNA polymerase), followed by 40 cycles at 95°C for 15 seconds and 60°C for 1 minute and a final extension step at 72°C for 4 minutes. All extracts were tested for inhibition as previously described (*35*).

2.2.2. Coat color

We chose a set of eight SNPs in seven genes for detecting basic colors (bay, black, chestnut and grey), diluted phenotypes (silver and cream), spotted or painted phenotypes (overo, tobiano and sabino) and leopard spotting, as described previously (*36*) (see paragraph SI 3.2.) We optimized the multiplex PCR as described before (*24*) and all the primers were selected and combined in a single multiplex reaction (see paragraph SI 3.4.2.). Each reaction was carried out in a 30 µl reaction volume with a final composition of 50 mM Tris/HCl, 10 mM KCl, 5 mM $(NH_4)_2SO_4$, 4 mM $MgCl_2$, pH 8.3 (@25°C), 1 mg/ml BSA (bovine serum albumin), 0.15 µM of each primer, 0.25 mM dA/G/CTPs and 0.50 mM dUTP, 2 U of Fast Start DNA polymerase, 0.5 U of uracil-N-Glycosylase (UNG) and 4 µl of DNA. A negative control was performed for each multiplex reaction that was processed in the same way as the samples throughout the whole experimental procedure. The cycling program consisted of 15 minutes at 37°C (carry-over contamination prevention through digestion by UNG of dUTP-labeled amplicons), 95°C for 10 minutes (inactivation of UNG and activation of the Fast Start DNA polymerase), followed by 50 cycles at 95°C for 20 seconds, 55°C for 30 seconds, 72°C for 30 seconds and a final extension step at 72°C for 4 minutes.

2.2.3. Y chromosome

We analyzed four anonymous Y-linked fragments, Y2B17, Y3B1, Y3B12, Y3B19 and two fragments of the amelogenin gene, AME2 and AME3, as previously described (*37*) (see paragraph SI 3.3). We optimized the multiplex PCR as described before (*24*) and all the primers were selected and combined in two multiplex reactions (see paragraph SI 3.4.3.). Each reaction was carried out in a 30 µl reaction volume with a final composition of 50 mM Tris/HCl, 10 mM KCl, 5 mM $(NH_4)_2SO_4$, 4 mM $MgCl_2$, pH 8.3 (@25°C), 1 mg/ml BSA (bovine serum albumin), 0.15 µM of each primer, 0.25 mM dA/G/CTPs and 0.50 mM dUTP, 2 U of Fast Start DNA polymerase, 0.5 U of uracil-N-Glycosylase (UNG) and 4 µl of DNA. A negative control was performed for each multiplex reaction that was processed in the same way as the samples throughout the whole experimental procedure. The cycling program consisted of 15 minutes at 37°C (carry-over contamination prevention through digestion by UNG of dUTP-labeled amplicons), 95°C for 10 minutes (inactivation of UNG and activation of the Fast Start DNA polymerase), followed by 50 cycles at 95°C for 20 seconds, 55°C for 30 seconds, 72°C for 30 seconds and a final extension step at 72°C for 4 minutes.

For each sample, we then performed the six multiplex PCRs with different primer combinations (mtDNA, coat color and Y chromosome) that were pooled at the end in a 96-well plate. Each extract was amplified in triplicate so that, at the end, there were three pools for each extract.

2.3. ION Torrent sequencing

The following steps were performed using an automate Tecan Freedom Evo 100 equipped with a 4 channel liquid handling arm suitable for disposable tips, a gripper to move objects, a double thermoblock and an automated solution for vacuum solid phase extraction. We first performed End Repair with the NEBNext end repair module (New England Biolabs) using 20 µL of multiplex PCR (containing no more than 5 picomoles of amplicon products) in a 50 µL reaction volume with only 0.1 µL of End Repair Enzymes. The reaction was incubated for 30 minutes at 25°C and then purified using the NucleoSpin 96 PCR clean-up kit as recommended by the manufacturer (Macherey-Nagel ref. 740658). The DNA was eluted in 50 µL and 20 µL were used for the ligation of the sample-specific Ion Torrent barcoded adaptors (1 µL of annealed A+P1, 20 µM) using the NEB Next ligation module (New England Biolabs) in a 30 µL reaction volume with 1 µL of Quick Ligase. The reaction was incubated for 30 minutes at 16°C. After the ligation, 60 µL of binding buffer (i.e., NT buffer, Macherey-Nagel) was added and all samples were pooled in a tube before purification on a silica column. The pooled barcoded PCR products were size selected using either an E-Gel SizeSelect (Life technology) or the Caliper Labchip XT (Perkin Elmer) to purify the PCR products ligated

to the adapters. The selected product was subjected to nick repair and amplification with the NEB OneTaq Hot Start in a 40 µL reaction volume containing Ion Torrent primers A and P1 (0.5 µM). The reaction was incubated for 20 min at 68°C (nick repair step). The DNA library was then amplified using the following program: initial denaturation at 94°C for 5 min, (94°C for 15 sec, 60°C for 15 sec, 68°C for 40 sec) for 6 cycles, final elongation at 68°C for 5 min. Products were finally purified with a Qiaquick PCR purification kit (Qiagen). The size distribution and concentration of the library was assessed on the Agilent 2100 Bioanalyzer. Emulsion PCR and Ion Sphere Particle enrichment were conducted with the Ion OneTouch System (Life Technologies) using the Ion OneTouch 200 Template kit v2 DL according to the manufacturer's protocol. The DNA library was sequenced on the Ion Torrent Personal Genome Machine (PM) Sequencer using the Ion PGM 200 Sequencing Kit and Ion 314 semiconductor sequencing chips (Life Technologies).

3. Information to PCR primers used

3.1. PCR primers for mitochondrial DNA

Primers	Primer seq (5'-3')	Primer coordinates	Amplicon length (bp)	Sequence length (bp)	Gene region
SNP2F	GGTACAGCTTTTTAGATACAGGTTA	1554-1578	81	34	16S
SNP2R	GGCTGCTTTTAAGCCAACTATG	1613-1634			
SNP4F	TTCAACTCCTCTCCCTAACAA	2749-2769	78	37	ND1
SNP4R	AATGCTACGGCGAGCAAGAT	2807-2826			
SNP8F	AATACGCAAACATCATCATGATAAA	3450-3474	91	44	ND1
SNP8R	ATTGAGTAGAGTTCTGGCAGGT	3519-3540			
SNP10F	GAATCGGATTTGAAATAAATCTACT	4025-4049	73	25	ND2
SNP10R	TCTATGGTTCGGGGATTGTACTT	4075-4097			
SNP11F	TCGCGCATATAGGATGAATAAC	4487-4508	120	73	ND2
SNP11R	AAGAGCTGTGGATAAATAGTATGAA	4582-4606			
SNP13F	AATAAAATGACAATTCGAAACCAAA	4872-4896	73	25	ND2
SNP13R	GAGGAGTAGGGAGGATATAACAA	4922-4944			
SNP15F	GAATTAGGCCAACCTGGGAC	5479-5498	75	35	COX1
SNP15R	TACGAATGCATGGGCGGTTA	5534-5553			
SNP17F	GTATGGGCTCACCACATGTTT	6220-6240	119	76	COX1
SNP17R	GTGGCTAGTCAGCTGAATACTT	6317-6338			
SNP21F	TAGGCCTCCCTATTGTAATTCTG	8008-8030	95	49	ATP6
SNP21R	TTGAATTGAGATTAGGCGATTGT	8080-8102			
SNP25F	CTCAACAGCCCTTATTACGTTTA	8522-8544	76	30	ATP6
SNP25R	GGCTTGGATTATAGCTACTGCGA	8575-8597			
SNP26F	TCCCCTTACTCAACACCTCA	9030-9049	83	40	COX3
SNP26R	ACGGTTTCCTTCTATTAGGCTAT	9090-9112			
SNP29F	TCCTAAACACCCACTTCACACT	10057-10078	91	47	ND4L
SNP29R	AGTAGGGATAATCCTAGAGCTG	10126-10147			
SNP31F	ATCCCCTATCAGCCCCACTT	10380-10399	71	31	ND4L
SNP31R	GTGTTGGCTGGCTATGAGTA	10431-10450			
SNP32F	CTGCCACTAATACTCATAGCCAG	10421-10443	79	32	ND4L
SNP32R	GCATGGTGATGTAGAGTTTTTTTC	10476-10499			
SNP36F	GAAAGTATGCAAGAACTGCTAAT	11654-11676	71	25	Ser (AGY)
SNP36R	TCCTTTAAAAGTTTGAGAGAGCC	11702-11724			
SNP37F	TCCTCCCTCATACTAGTTTCAC	11795-11816	116	70	ND5
SNP37R	GCATATGAGATAGTGTTTTTTACA	11887-11910			
SNP41sh1F	ATCAGCAATTCCCTACATCGG	14637-14657	77	35	CytB
SNP41sh1R	TAAGGGTGGCTTTGTCTACTG	14693-14713			

SNP41sh2F	GGTGGATTCTCAGTAGACAAA	14683-14703	79	37	CytB
SNP41sh2R	GGGCTGTGATGATGAAGGGTA	14741-14761			
HVR1F	ATTTCTTCCCCTAAACGACAAC	15469-15490	96	52	D-loop
HVR1R	ATATTGCATGTCAGGTGGGTAT	15543-15564			
HVR5F	CCGCGGGAAATCAGCAAC	15749-15775	121	73	D-loop
HVR5R	GAATGGCCCTGAAGAAAGAAC	15849-15869			

3.2. Primers for alleles involved in coat color

Colour	Primer name	Gene region	Primer seq (5'-3')	Amplicon length (bp)	Sequence length (bp)
Chestnut	MC1R-F	MC1R-E1	GCACTCACCCATGTACTACTTCA	71	29
	MC1R-R		GCACGTTGCTCATGCTCAC		
Black	ASIP-F	ASIP	AAGAAATCCAAAAAGATCAGC	77	38
	ASIP-R		ATGAGAAGTCCAAGGCCTACCT		
Cream	MATP-F	MATP (SLC45A2)	GCCATAACCATCACCATGATAG	65	24
	MATP-R		GGCCCTACAATGAAGTCAG		
Tobiano	KIT13F	KIT (intron 13)	CGTCATGACTCATTCGTGAGAA	63	18
	KIT13R		GCTCTGAAGGTAACAAGCAACTA		
	KIT16F	KIT (intron 16)	TTTAAATGGCTTTCTTTTCTCC	59	15
	KIT16R		TGCCAAGTCCCTATGAATACAC		
Silver	SILV11-F	SILV	TCCTTCTTCTTCTCCCAAATCA	52	9
	SILV11-R		GAGCTGAGCCCTGCTTCATAA		
Overo	ENDRB-F	EDNRB	CAGTAGTGTCCTGCCTAGTGTTC	63	17
	ENDRB-R		TGATTCTCAGCAGTGTGGAGTTT		
Leopard	LP-F	LP	AGATCGAGCTGCTGAACTGG	62	22
	LP-R		TCTCCATGATCCCAAGCAAT		

3.3. Sequences of Y-chromosome primers

Primer name	Gene region	Primer seq (5'-3')	Amplicon length (bp)	Sequence Length (bp)
Y2B17-1F	Y2B17	ATGTCAGGATGCCAACTGGTC	74	34
Y2B17-1R		GTGCCCACAAGACAATGTC		
Y2B17-4F	Y2B17	AGTGGCYCCCTGAAGACATT	76	36
Y2B17-4R		CGACRGGAAAAGGGACAGTG		
Y2B17-5F	Y2B17	GGCCACCACTGTCCCTTTT	67	28
Y2B17-5R		CAGGACAACAACCCCACTCA		
Y2B17-6F	Y2B17	GGCGACATGGCAGCACTA	64	27
Y2B17-6R		ACCTGAGCCAGTGCAGTGA		
Y3B1-9F	Y3B1	CTCTCTCTCTAWTGGGGCAGCAG	58	16
Y3B1-9R		GTCACCKGCCAGTCTCAGC		
Y3B1-10F	Y3B1	GCTTCAGCCTGGCTGAGACT	56	16
Y3B1-10R		GCACTCCCACTGGTGAAGGA		
Y3B1-12F	Y3B1	AATGTGGATTCCAGGCAGCTTCT	73	27
Y3B1-12R		TTTTGTCACTGAGGACTCTTGCA		
Y3B12-16F	Y3B12	GCCCCTGGTGACTCCCAGT	68	30
Y3B12-16R		GGCCTTGGCGAGGAAGAAC		

Y3B19-18F Y3B19-18R	Y3B19	GGCTTGATCCAAACTGCT TCTTTTTTATTGGATTTATTGC	70	30
YAme2-21F YAme2-21R	Amelogenin	TCTAAGGGAGATATTCTGCACGATT TGCAATTTGTATTGGTGCTGGAGTG	71	21
YAme2-23F YAme2-23R	Amelogenin	CATAGGATTTATAATGTTTCTACTTTACC ACAATTATTTTGGGCATTATTGA	71	19
YAme2-24F YAme2-24R	Amelogenin	GGAAAATGGTGGGTAGGATAGG AAGATCCCCTTCACCCTACAC	66	23
YAme3-26F YAme3-26R	Amelogenin	TTGTCCCATCACACTCCTACTCG CCATGTAAGGAGATGAAACTGAG	56	10
YAme3-27F YAme3-27R	Amelogenin	GGAATATAAGAATAGTTGCTCAAA ATGGTAGTACAGCATTTGAAATCAAT	72	22
YAme3-28F YAme3-28R	Amelogenin	GAGTCTCCAGATTTGCGATTAACT TGTGAAAGAGGAATAGTTTCC	69	24

3.4. Multiplex PCR conditions

3.4.1. Mitochondrial DNA

Multiplex:	**1**	**2**	**3**
	SNP2	SNP29	SNP8
	SNP31	SNP15	SNP21
	SNP11	SNP4	SNP26
	SNP17	SNP32	SNP36
	SNP41sh2	SNP25	SNP10
	HVR1	SNP37	SNP13
	HVR5	SNP41sh1	

PCR program: 1 cycle 37°C 15 min, 95°C 10 min; 40 cycles 95°C 15 sec, 60°C 60 sec ; 1 cycle 72°C, 4 min

3.4.2. Multiplex Coat color

Multiplex:	**1**
	ASIP
	KIT13
	KIT16
	MATP
	MC1R
	SILV11
	LP
	EDNRB

PCR program : 1 cycle 37°C, 15 min, 95°C, 10 min; 50 cycles 95°C 20 sec, 55°C 30 sec, 72°C 30 sec, ; 1 cycle 72°C 4 min

3.4.3. Multiplex Y chromosome

Multiplex:	**1**	**2**
	P-Y2B17-1	P-Y3B1-10
	P-Y2B17-5	P-Y2B17-4
	P-Y2B17-6	P-YAme2-20
	P-Y3B1-9	P-YAme2-22

	P-Y3B1-12	P-YAme2-24
	P-Y3B12-16	P-YAme3-26
	P-Y3B19-17	P-YAme3-27
		P-YAme3-28

PCR program : 1 cycle 37°C, 15 min, 95°C, 10 min; 50 cycles 95°C 20 sec, 55°C, 30 sec, 72°C, 30 sec ; 1 cycle 72°C 4 min

Supplementary Tables:

Table S1: samples analyzed with dates and results

TabS2: List of SNPs used for determination of the mitochondrial DNA haplogroups

TabS3: List of diagnostic SNPs for the determination of horse mitochondrial DNA haplogroups in both the complete mitogenomes and in the reduced assay used in the present study

TabS4: List of diagnostic SNPs assayed to determine horse mitochondrial DNA haplogroups

TabS5: List of SNPs obtained from the Y-chromosome assay

TabS6: List of genotypes obtained from the coat color assay

Supplementary Figures:

Fig. S1: Map showing the archaeological sites from where the analyzed samples originated

Fig. S2: coat colors found in the present study

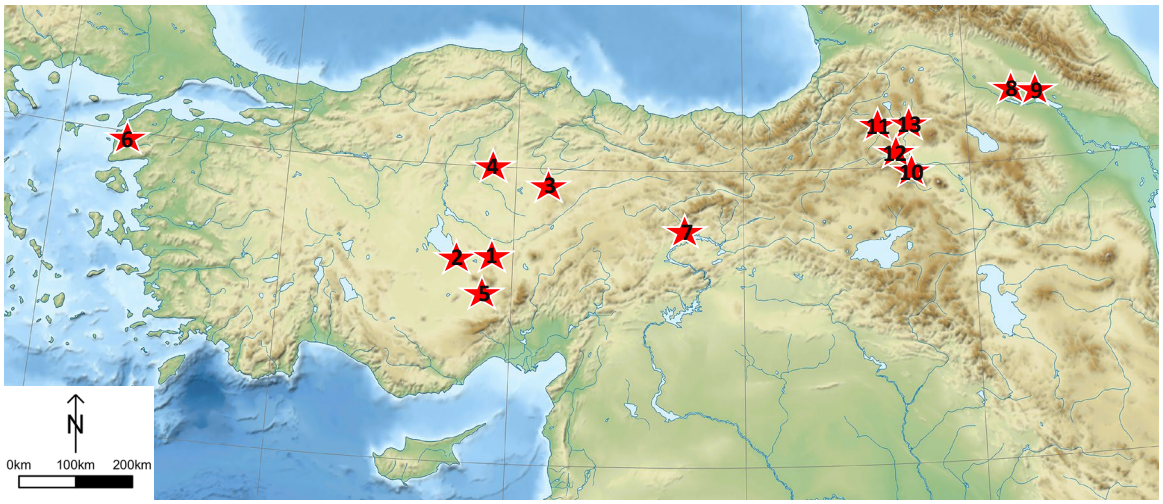


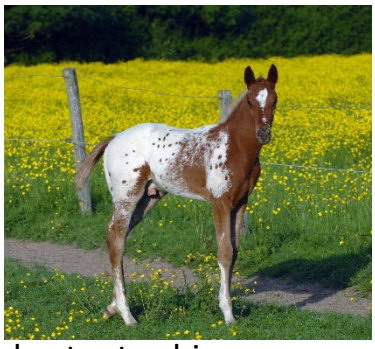
Fig. S1: Map showing the locations of the archaeological sites from which the analyzed specimens originate. Aşıklı Höyük (1), Acemhöyük (2), Çadır Höyük (3), Boğazköy-Hattusa (4), Köşk Höyük (5), Troy (6), Lidar Höyük (7), Didi Gora (8), Tqisbolu Gora (9), Lusakert (10), Mets Sepasar (11), Tsaghkahovit (12), Gegharot (13)

Bay
©IFCE-ALaurioux

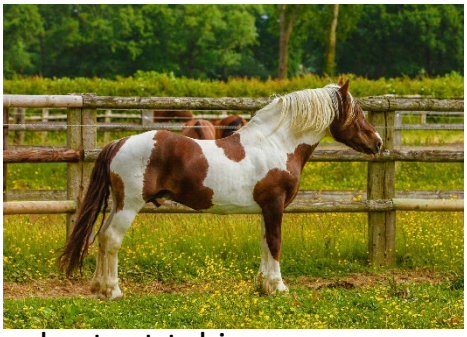
bay-tobiano
©IFCE-Y.RIVALAIN

Chestnut
©IFCE

Chestnut-silver (mane)
©IFCE-PSV



chestnut-sabino-blanket-spotted
©IFCE-A.LAURIOUX



chestnut-tobiano
©IFCE-A.BASSALER

Leopard
©IFCE-A.LAURIOUX

Black
©IFCE-L.GERARD

Fig. S2: Horses showing the coat colors evidenced in the present study. Copyright: Institut du cheval et de l'équitation ifce, France.