SUPPORTING INFORMATION APPENDIX

Supporting Materials and Methods

Sequencing Protocol

The entire horse mtDNA was amplified in 11 overlapping PCR fragments, using a set of oligonucleotides with matching annealing temperatures (Table S10). Oligonucleotides were checked (through GenBank BLAST) in order to avoid amplification of nuclear insertions of mitochondrial sequences (numts) (1). After PCR, the fragments were purified using the ExoSAP-IT® enzymatic system (Exonuclease I and Shrimp Alkaline Phosphatase, GE Healthcare) and standard dideoxysequencing was performed by using a set of 33 nested primers (Table S11) specifically designed for this protocol. An ABI 3730 sequencer with 96 capillaries was employed for separation of the sequencing ladders. Complete sequences were aligned, assembled, and compared using the program Sequencher 4.9 (Gene Codes). Traces were generally of excellent quality and there was extensive overlap between reads with most observed mutations determined by at least two independent sequencing reactions. At least two independent operators read each sequence and any potentially ambiguous base call was tested by additional and independent PCR and sequencing reactions.

Figure S1A

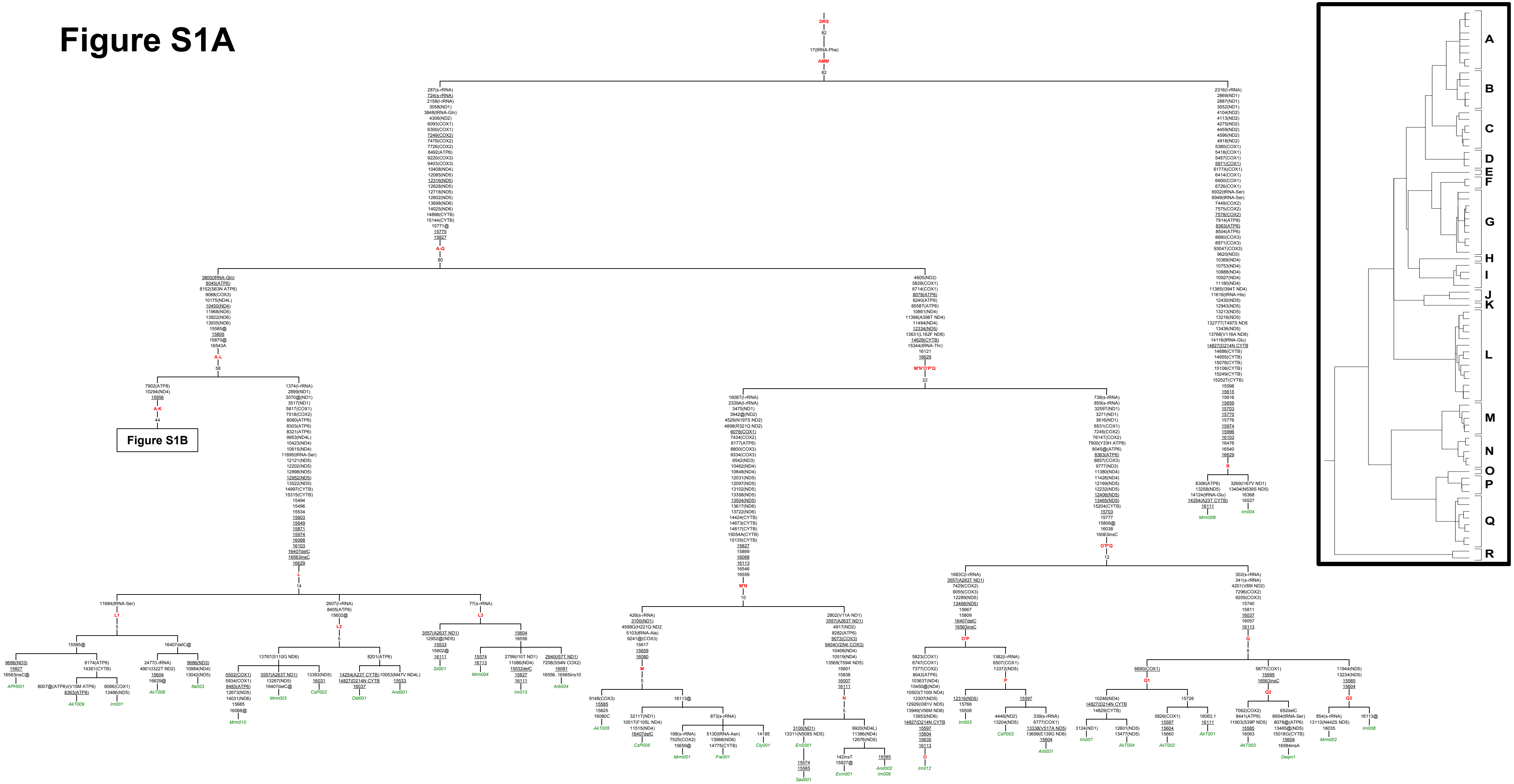
DRS
AMM
A-Q
A-L
A-K
Figure S1B
L
L1
L2
L3
M'N'O'P'Q
M'N
M
N
O'P'Q
O'P
O
P
Q
Q1
Q2
Q3
R
APH001
AkT009
Irn001
AkT008
Ita003
Mrm010
Mrm003
CsP002
Old001
And001
Sil001
Mrm004
Irn013
Arb004
AkT005
CsP005
Mrm001
Fre001
Cly001
EnS001
Sad001
Exm001
And002
Irn006
Irn012
Irn003
CsP003
Arb001
Irn007
AkT004
AkT002
AkT001
AkT003
Deqin1
Mrm002
Irn008
Mrm009
Irn004
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R

Figure S1B

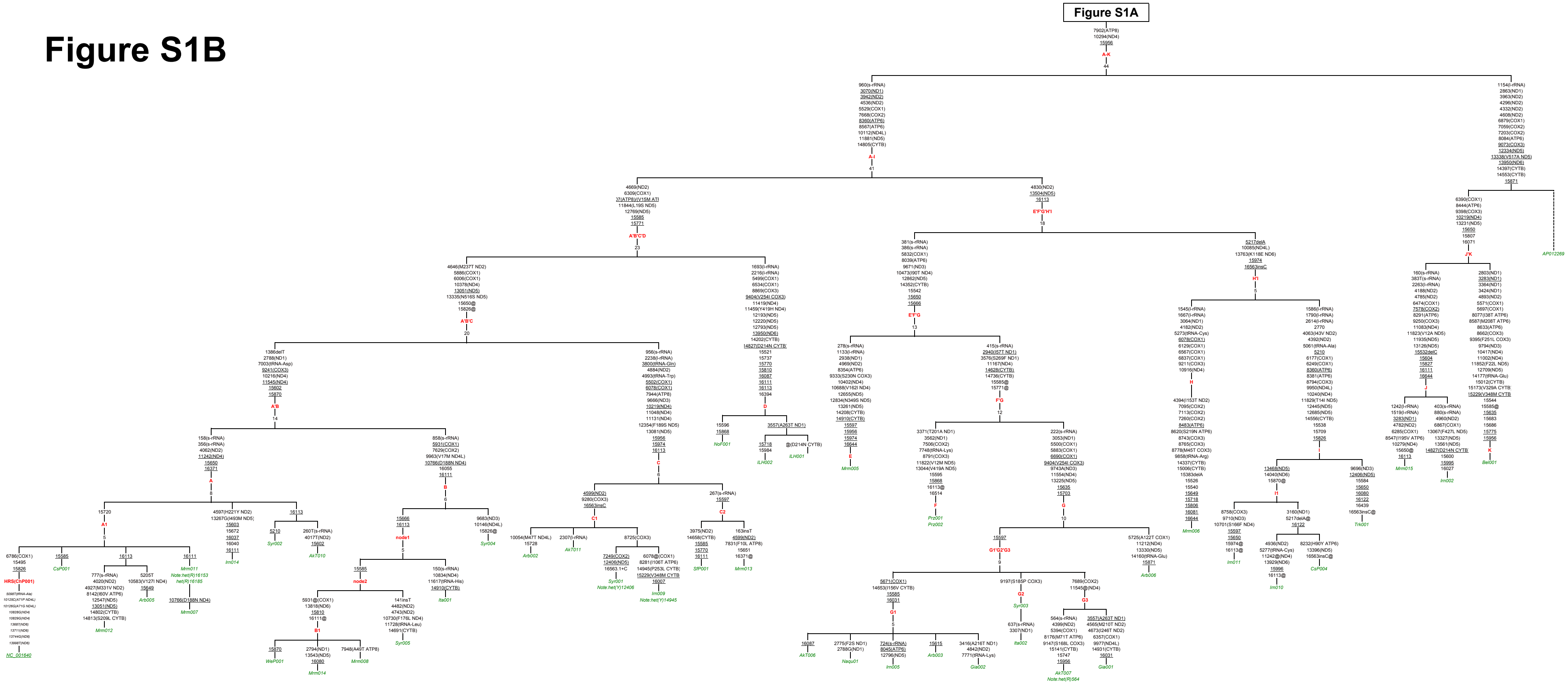
Figure S1A
7902(ATP8)
10294(ND4)
15956
A-K
44
960(s-rRNA)
3070(ND1)
3942(ND2)
4536(ND2)
5529(COX1)
7668(COX2)
8360(ATP6)
8567(ATP6)
10112(ND4L)
11881(ND5)
14805(CYTB)
A-I
41
4669(ND2)
6309(COX1)
7(ATP8)/(V15M AT
11844(L19S ND5)
12769(ND5)
15585
15771
A'B'C'D
23
4646(M237T ND2)
5886(COX1)
6006(COX1)
10378(ND4)
13051(ND5)
13335(N516S ND5)
15650@
15826@
A'B'C
20
1386delT
2788(ND1)
7003(tRNA-Asp)
9241(COX3)
10216(ND4)
11545(ND4)
15602
15870
A'B
14
158(s-rRNA)
356(s-rRNA)
4062(ND2)
11242(ND4)
15650
16371
A
8
15720
A1
5
6786(COX1)
15495
15826
HRS(ChP001)
5098T(tRNA-Ala)
10125C(A71P ND4L)
10126G(A71G ND4L)
10828G(ND4)
10829G(ND4)
13687(ND6)
13711(ND6)
13744G(ND6)
13998T(ND6)
NC_001640
15585
CsP001
16113
777(s-rRNA)
4020(ND2)
4927(M331V ND2)
8142(I60V ATP6)
12547(ND5)
13051(ND5)
14802(CYTB)
14813(S209L CYTB)
Mrm012
5205T
10583(V127I ND4)
15649
Arb005
16111
Mrm011
Note:het(R)16153
het(R)16185
10766(D188N ND4)
Mrm007
4597(H221Y ND2)
13267G(I493M ND5)
15603
15672
16037
16040
16111
Irn014
16113
5210
Syr002
260T(s-rRNA)
4017T(ND2)
15602
AkT010
858(s-rRNA)
5931(COX1)
7629(COX2)
9963(V17M ND4L)
10766(D188N ND4)
16055
16111
B
6
15666
16113
node1
5
15585
node2
5931@(COX1)
13818(ND6)
15810
16111@
B1
15870
WeP001
2794(ND1)
13543(ND5)
16080
Mrm014
7948(A49T ATP8)
Mrm008
141insT
4482(ND2)
4743(ND2)
10730(F176L ND4)
11728(tRNA-Leu)
14691(CYTB)
Syr005
150(s-rRNA)
10834(ND4)
11617(tRNA-His)
14910(CYTB)
Ita001
9683(ND3)
10146(ND4L)
15826@
Syr004
956(s-rRNA)
2238(l-rRNA)
3800(tRNA-Gln)
4884(ND2)
4993(tRNA-Trp)
5502(COX1)
6078(COX1)
7944(ATP8)
9666(ND3)
10219(ND4)
11048(ND4)
11131(ND4)
12354(F189S ND5)
13081(ND5)
15956
15974
16113
C
6
4599(ND2)
9280(COX3)
16563insC
C1
10054(M47T ND4L)
15728
Arb002
2307(l-rRNA)
AkT011
8725(COX3)
7249(COX2)
12406(ND5)
16563.1+C
Syr001
Note:het(Y)12406
6078@(COX1)
8281(I106T ATP6)
14945(P253L CYTB)
15229(V348M CYTB
16007
Irn009
Note:het(Y)14945
267(s-rRNA)
15597
C2
3975(ND2)
14658(CYTB)
15585
15770
16111
SfP001
163insT
4599(ND2)
7831(F10L ATP8)
15651
16371@
Mrm013
1693(l-rRNA)
2216(l-rRNA)
5499(COX1)
6534(COX1)
8869(COX3)
9404(V254I COX3)
11419(ND4)
11459(Y419H ND4)
12193(ND5)
12220(ND5)
12793(ND5)
13950(ND6)
14202(CYTB)
14827(D214N CYTB
15521
15737
15770
15810
16087
16111
16113
16394
D
15596
15868
NoF001
3557(A263T ND1)
15718
15984
ILH002
@(D214N CYTB)
ILH001
4830(ND2)
13504(ND5)
16113
E'F'G'H'I
18
381(s-rRNA)
386(s-rRNA)
5832(COX1)
8039(ATP6)
9671(ND3)
10473(I90T ND4)
12862(ND5)
14352(CYTB)
15542
15650
15666
E'F'G
13
278(s-rRNA)
1133(l-rRNA)
2938(ND1)
4969(ND2)
8354(ATP6)
9333(S230N COX3)
10402(ND4)
10688(V162I ND4)
12655(ND5)
12834(N349S ND5)
13261(ND5)
14208(CYTB)
14910(CYTB)
15597
15956
15974
16644
E
Mrm005
415(s-rRNA)
2940(I57T ND1)
3576(S269F ND1)
11167(ND4)
14628(CYTB)
14736(CYTB)
15585@
15771@
F'G
12
3371(T201A ND1)
3562(ND1)
7506(COX2)
7748(tRNA-Lys)
8791(COX3)
11822(V12M ND5)
13044(V419A ND5)
15595
15868
16113@
16514
F
Prz001
Prz002
222(s-rRNA)
3053(ND1)
5500(COX1)
5883(COX1)
6690(COX1)
9404(V254I COX3)
9743A(ND3)
11554(ND4)
13225(ND5)
15635
15703
G
10
15597
G1'G2'G3
9
5671(COX1)
14653(I156V CYTB)
15585
16031
G1
5
16087
AkT006
2775(F2S ND1)
2788G(ND1)
Naqu01
724(s-rRNA)
8045(ATP6)
12796(ND5)
Irn005
15615
Arb003
3416(A216T ND1)
4842(ND2)
7771(tRNA-Lys)
Gia002
9197(S185P COX3)
G2
Syr003
637(s-rRNA)
3307(ND1)
Ita002
7689(COX2)
11545@(ND4)
G3
564(s-rRNA)
4399(ND2)
5394(COX1)
8176(M71T ATP6)
9147(S168L COX3)
15141(CYTB)
15747
15956
AkT007
Note:het(R)564
3557(A263T ND1)
4565(M210T ND2)
4673(I246T ND2)
6357(COX1)
9977(ND4L)
14931(CYTB)
16031
Gia001
5725(A122T COX1)
11212(ND4)
13330(ND5)
14160(tRNA-Glu)
15871
Arb006
5217delA
10085(ND4L)
13763(K118E ND6)
15974
16563insC
H'I
5
1545(l-rRNA)
1667(l-rRNA)
3064(ND1)
4182(ND2)
5273(tRNA-Cys)
6078(COX1)
6129(COX1)
6567(COX1)
6837(COX1)
9211(COX3)
10916(ND4)
H
4394(I153T ND2)
7095(COX2)
7113(COX2)
7260(COX2)
8483(ATP6)
8620(S219N ATP6)
8743(COX3)
8765(COX3)
8778(M45T COX3)
9858(tRNA-Arg)
14337(CYTB)
15006(CYTB)
15383delA
15526
15540
15649
15718
15806
16081
16644
Mrm006
1586(l-rRNA)
1790(l-rRNA)
2614(l-rRNA)
2770
4063(I43V ND2)
4392(ND2)
5061(tRNA-Ala)
5210
6177(COX1)
6249(COX1)
8360(ATP6)
8381(ATP6)
8794(COX3)
9950(ND4L)
10240(ND4)
11829(T14I ND5)
12445(ND5)
12685(ND5)
14556(CYTB)
15538
15709
15826
I
13468(ND5)
14040(ND6)
15870@
I1
8758(COX3)
9710(ND3)
10701(S166F ND4)
15597
15650
15974@
16113@
Irn011
3160(ND1)
5217delA@
16122
4936(ND2)
5277(tRNA-Cys)
11242@(ND4)
13929(ND6)
15996
16113@
Irn010
8232(H90Y ATP6)
13396(ND5)
16563insC@
CsP004
9696(ND3)
12406(ND5)
15584
15650
16080
16122
16439
16563insC@
Trk001
1154(l-rRNA)
2863(ND1)
3963(ND2)
4296(ND2)
4332(ND2)
4608(ND2)
6879(COX1)
7059(COX2)
7203(COX2)
8084(ATP6)
9073(COX3)
12334(ND5)
13338(V517A ND5)
13950(ND6)
14397(CYTB)
14553(CYTB)
15871
6390(COX1)
8444(ATP6)
9398(COX3)
10219(ND4)
13231(ND5)
15650
15807
16071
J'K
160(s-rRNA)
383T(s-rRNA)
2263(l-rRNA)
4188(ND2)
4785(ND2)
6474(COX1)
7578(COX2)
8291(ATP6)
9250(COX3)
11083(ND4)
11823(V12A ND5)
11935(ND5)
13126(ND5)
15532delC
15604
15827
16111
16644
J
1242(l-rRNA)
1519(l-rRNA)
3283(ND1)
4782(ND2)
6285(COX1)
8547(I195V ATP6)
10279(ND4)
15650@
16113
Mrm015
403(s-rRNA)
880(s-rRNA)
4960(ND2)
6867(COX1)
13067(F427L ND5)
13327(ND5)
13561(ND5)
14827(D214N CYTB
15600
15995
16027
Irn002
2803(ND1)
3283(ND1)
3364(ND1)
3424(ND1)
4893(ND2)
5571(COX1)
5697(COX1)
8077(I38T ATP6)
8587(M208T ATP6)
8633(ATP6)
8662(COX3)
9395(F251L COX3)
9794(ND3)
10417(ND4)
11002(ND4)
11852(F22L ND5)
12709(ND5)
14177(tRNA-Glu)
15012(CYTB)
15173(V329A CYTB
15229(V348M CYTB
15544
15585@
15635
15683
15686
15775
15956
K
Bel001
AP012269

Figure S1. A putative most parsimonious tree of complete mtDNA sequences from horses. This tree encompasses 83 sequences and was rooted by using a published donkey (*E. asinus*) sequence. Mutations are shown on the branches and are numbered according to a new proposed horse reference sequence (HRS, see Table S1). This mtDNA differs from the GenBank RefSeq NC001640 (reported in italics in the tree for the sake of clarity, but disregarded in all calculations) at the following sites: 357+C, 2227delT, 5098A(tRNA-Ala), 5240delA, 5279delA, 10125G(R71G ND4L), 10126C(R71P ND4L), 10828C, 10829C(V209L ND4), 13687(I34N ND6), 13711(G42H ND6), 13744T(G53I ND6), 13998G(E138D ND6), 15388delT. Taking into account the widespread use of the RefSeq NC001640, the original nucleotide numbering was retained. In view of this, we have removed one of the two Cs at nps 357-358 and artificially introduced an N at nps 2227, 5240, 5279 and 15388 (GenBank accession n. JN398377). Mutations are transitions unless a base is explicitly indicated. Suffixes indicate transversions (to A, G, C, or T), indels (+, d), heteroplasmy (h) or reversions (@). Recurrent mutations are underlined. A basal motif for haplogroup H was extrapolated by a comparison with the GenBank sequence EU939445. However, this and another three GenBank records (AY584828, EF597512 and NC001640) were excluded from phylogenetic analyses due to a suspicious and apparently unjustified surplus of mutations in the reported sequences. Further information on the 81 novel sequences is reported in Table S1. The branching of a recently published Przewalski's horse sequence (AP012269) (2), predating the node of haplogroup J'K in the phylogeny, is indicated by a dotted line. The tree shown is only one out of many nearly most parsimonious trees, which mainly differ in the different reconstruction of mutations at highly recurrent sites.

The inset in Figure S1A shows the tree based on a Bayesian approach performed with Bayes 3.1.2 (3) in order to double-check the topology of the horse phylogeny. We set the HKY85 evolutionary model with the gamma-distributed rate variation. The analysis was run using four incrementally heated chains with the default temperature; over 500,000 generations long until the average standard deviation of split frequencies reached the threshold value of 0.01. When calculating summary statistics, we skipped 25% of the initial trees to obtain a potential scale reduction factor (PSRF) value reasonably close to 1.0 for all calculated parameters. The resulting tree topology was very similar to the results obtained with the MP approach.

Figure S2

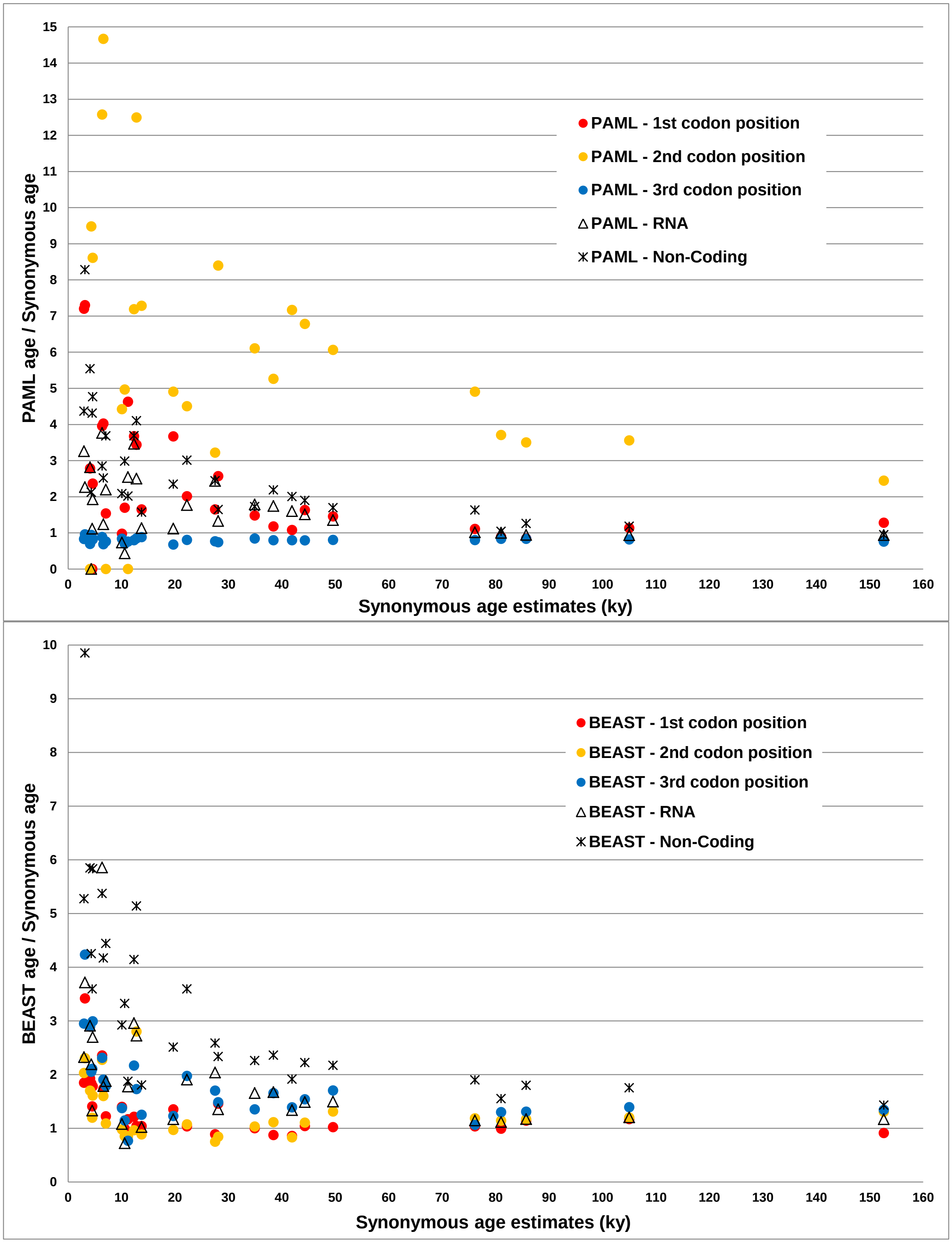

Figure S2. Comparison of age estimates obtained with the five partitions. Ages calculated both with PAML (top) and BEAST (bottom) using the same five datasets (1^{st}, 2^{nd} and 3^{rd} codons, RNA genes, and non-coding region) were weighted on the synonymous estimates. With both methods, the ratio values are inversely correlated with age. See Table S3 for further details.

Figure S3

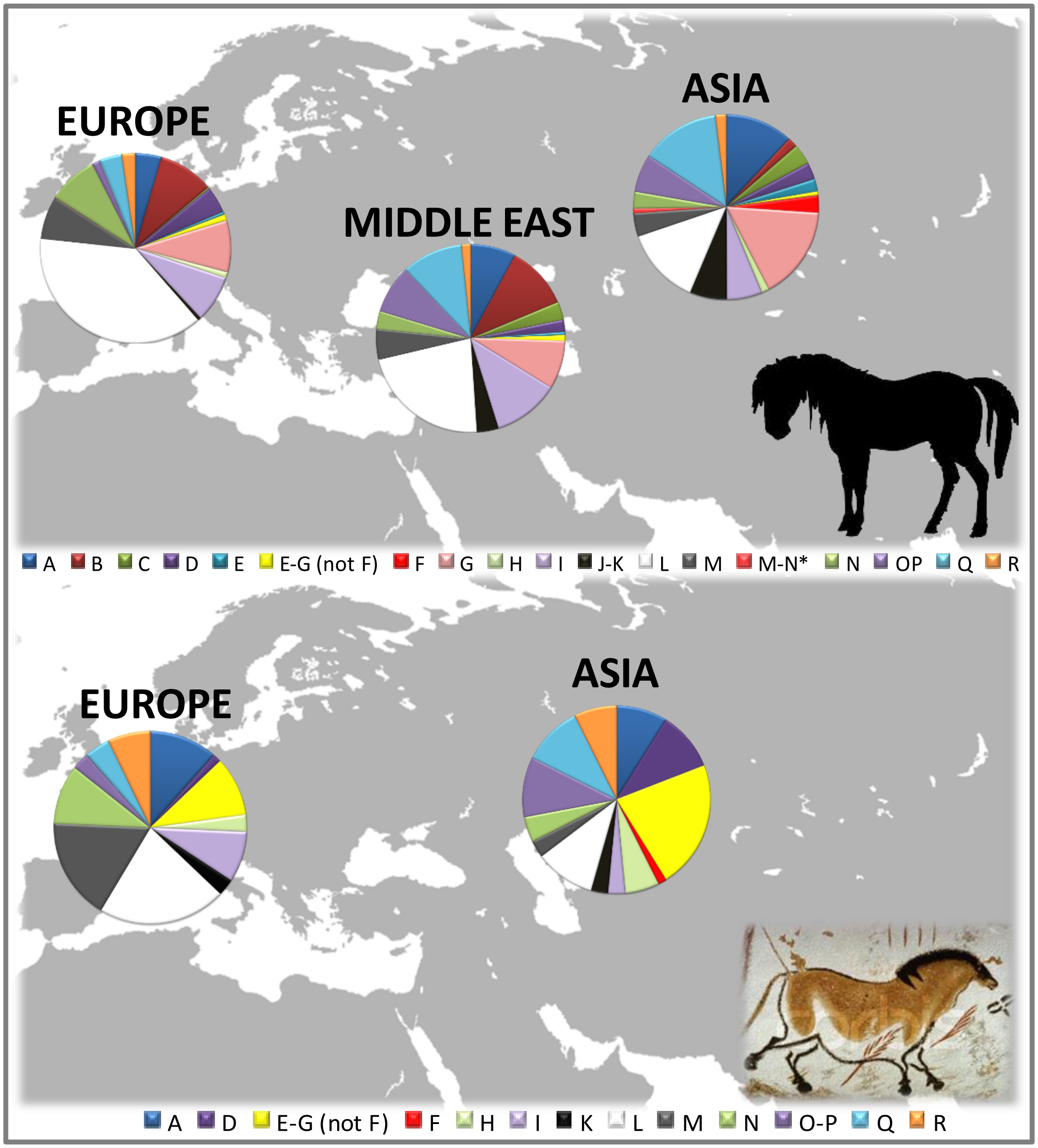

Figure S3. Spatial frequency distributions of horse haplogroups in Eurasia in modern (above) and ancient (below) DNA samples. A total of 2028 modern and 138 ancient control-region sequences are included. Population samples and corresponding frequency values are listed in Tables S6-S8.

Table S1. Sources and haplogroup affiliation for the horse complete mtDNA sequences.

# [a]	ID	Haplogroup	Breed	Geographic Origin	GenBank #
01	HRS [b] (ChP01)	A	Chincoteague Pony	America North	JN398377
02	CsP01	A	Caspian Pony	Middle East	JN398378
03	Mrm12	A	Maremmano	Europe South	JN398379
04	Arb05	A	Arabian	Middle East	JN398380
05	Mrm11	A	Maremmano	Europe South	JN398381
06	Mrm07	A	Maremmano	Europe South	JN398382
07	Irn14	A	Unspecified Iranian Breed	Middle East	JN398383
08	Syr02	A	Unspecified Syrian Breed	Middle East	JN398384
09	AkT10	A	Akhal-Teke	Asia Centre	JN398385
10	WeP01	B	Westphalian	Europe Centre	JN398386
11	Mrm14	B	Maremmano	Europe South	JN398387
12	Mrm08	B	Maremmano	Europe South	JN398388
13	Syr05	B	Unspecified Syrian Breed	Middle East	JN398389
14	Ita01	B	Unspecified Italian Breed	Europe South	JN398390
15	Syr04	B	Unspecified Syrian Breed	Middle East	JN398391
16	Arb02	C	Arabian	Middle East	JN398392
17	AkT11	C	Akhal-Teke	Asia Centre	JN398393
18	Syr01	C	Unspecified Syrian Breed	Middle East	JN398394
19	Irn09	C	Unspecified Iranian Breed	Middle East	JN398395
20	SfP01	C	Suffolk Punch	Europe North	JN398396
21	Mrm13	C	Maremmano	Europe South	JN398397
22	NoF01	D	Norwegian Fjord	Europe North	JN398398
23	ILH02	D	Icelandic Horse	Europe North	JN398399
24	ILH01	D	Icelandic Horse	Europe North	JN398400
25	Mrm05	E	Maremmano	Europe South	JN398401
26	Prz01	F	Przewalskii	Asia Centre	JN398402
27	Prz02	F	Przewalskii	Asia Centre	JN398403
28	AkT06	G	Akhal-Teke	Asia Centre	JN398404
29	Naqu01	G	Naqu	Asia Centre	EF597513
30	Irn05	G	Unspecified Iranian Breed	Middle East	JN398405
31	Arb03	G	Arabian	Middle East	JN398406
32	Gia02	G	Giara Horse (Sardinia)	Europe South	JN398407
33	Syr03	G	Unspecified Syrian Breed	Middle East	JN398408
34	Ita02	G	Unspecified Italian Breed	Europe South	JN398409
35	AkT07	G	Akhal-Teke	Asia Centre	JN398410
36	Gia01	G	Giara Horse (Sardinia)	Europe South	JN398411
37	Arb06	G	Arabian	Middle East	JN398412
38	Mrm06	H	Maremmano	Europe South	JN398413
39	Irn11	I	Unspecified Iranian Breed	Middle East	JN398414
40	Irn10	I	Unspecified Iranian Breed	Middle East	JN398415
41	CsP04	I	Caspian Pony	Middle East	JN398416
42	Trk01	I	Trakhener	Europe North	JN398417
43	Mrm15	J	Maremmano	Europe South	JN398418
44	Irn02	J	Unspecified Iranian Breed	Middle East	JN398419
45	Bel01	K	Belgian Draft	Europe Centre	JN398420
46	AkT09	L	Akhal-Teke	Asia Centre	JN398422
47	Irn01	L	Unspecified Iranian Breed	Middle East	JN398423
48	APH01	L	American Paint Horse	America North	JN398421

49	AkT08	L	Akhal-Teke	Asia Centre	JN398424
50	Ita03	L	Unspecified Italian Breed	Europe South	JN398425
51	Mrm03	L	Maremmano	Europe South	JN398427
52	Mrm10	L	Maremmano	Europe South	JN398426
53	CsP02	L	Caspian Pony	Middle East	JN398428
54	Old01	L	Oldenburg	Europe North	JN398429
55	And01	L	Andalusian	Europe South	JN398430
56	Sil01	L	Silesian	Europe Centre	JN398431
57	Mrm04	L	Maremmano	Europe South	JN398432
58	Irn13	L	Unspecified Iranian Breed	Middle East	JN398433
59	Arb04	L	Arabian	Middle East	JN398434
60	AkT05	M	Akhal-Teke	Asia Centre	JN398435
61	CsP05	M	Caspian Pony	Middle East	JN398436
62	Cly01	M	Clydesdale	Europe North	JN398439
63	Mrm01	M	Maremmano	Europe South	JN398437
64	Fre01	M	Friesian	Europe Centre	JN398438
65	EnS01	N	English Shire	Europe North	JN398440
66	Sad01	N	Saddlebred	America North	JN398441
67	Exm01	N	Exmoor Pony	Europe North	JN398442
68	And02	N	Andalusian	Europe South	JN398443
69	Irn06	N	Unspecified Iranian Breed	Middle East	JN398444
70	Irn12	O	Unspecified Iranian Breed	Middle East	JN398445
71	Irn03	P	Unspecified Iranian Breed	Middle East	JN398446
72	CsP03	P	Caspian Pony	Middle East	JN398447
73	Arb01	P	Arabian	Middle East	JN398448
74	AkT02	Q	Akhal-Teke	Asia Centre	JN398449
75	AkT01	Q	Akhal-Teke	Asia Centre	JN398450
76	Irn07	Q	Unspecified Iranian Breed	Middle East	JN398451
77	AkT04	Q	Akhal-Teke	Asia Centre	JN398452
78	AkT03	Q	Akhal-Teke	Asia Centre	JN398453
79	Deqin1	Q	Deqin	Asia Centre	EF597514
80	Mrm02	Q	Maremmano	Europe South	JN398454
81	Irn08	Q	Unspecified Iranian Breed	Middle East	JN398455
82	Mrm09	R	Maremmano	Europe South	JN398456
83	Irn04	R	Unspecified Iranian Breed	Middle East	JN398457
	DRS	DRS	*Equus asinus*	Europe North?	NC_001788

[a] Numbered as in the tree of Fig. 2.

[b] This is the newly proposed horse reference sequence (HRS).

Table S2. Age estimates[a] calculated on the protein coding genes (by considering only synonymous mutations) and on the entire genome partitioned into five datasets (1st, 2nd and 3rd positions of the codons, RNAs and non-coding region).

	ML (synonymous mutations)		**ML (all substitutions)**		**Bayesian (all substitutions)**	
Node	**Age**	**95% CI**	**Age**	**95% CI**	**Median**	**95% HPD**
A-R (AMM)	152.64	{122.64; 182.64}	133.46	{96.52; 170.40}	157.23	{109.46-214.05}
A-Q	105.01	{84.94; 125.09}	98.41	{72.28; 124.53}	120.67	{89.74-165.22}
A-L	85.71	{68.28; 103.13}	79.85	{58.68; 101.03}	97.61	{69.20-130.82}
A-K	81.02	{64.56; 97.48}	76.21	{55.96; 96.47}	91.04	{66.88-124.00}
A-I	49.56	{37.73; 61.40}	55.55	{40.47; 70.62}	66.73	{47.65-89.88}
A-D	38.44	{25.66; 51.21}	44.51	{31.39; 57.63}	51.58	{35.00-70.67}
A-C	27.51	{16.39; 38.62}	33.41	{22.71; 44.11}	37.92	{25.41-53.07}
A-B	12.33	{1.23; 23.42}	20.55	{12.84; 28.26}	23.03	{14.66-33.40}
A	3.15	{0.00; 6.34}	8.98	{5.30; 12.67}	10.69	{6.54-16.02}
B	7.07	{0.00; 16.19}	9.63	{5.03; 14.23}	11.37	{6.82-17.27}
C	4.59	{0.00; 9.80}	7.63	{3.49; 11.78}	9.71	{5.73-14.73}
D	0.00	-	3.93	{0.60; 7.27}	6.44	{3.07-10.82}
E-I	44.30	{33.34; 55.26}	52.20	{37.93; 66.47}	61.28	{43.86-82.78}
E-G	28.09	{17.16; 39.03}	33.17	{22.24; 44.11}	40.38	{27.27-56.91}
F-G	19.68	{8.84; 30.53}	23.55	{14.91; 32.18}	27.25	{16.94-39.07}
G	6.59	{0.00; 14.92}	9.16	{5.36; 12.96}	11.6	{7.26-17.07}
H-I	41.91	{30.93; 52.89}	48.97	{34.93; 63.01}	54.52	{37.37-74.36}
I	10.59	{3.96; 17.21}	12.21	{6.34; 18.08}	13.63	{7.59-21.78}
J-K	34.92	{22.70; 47.14}	41.22	{27.87; 54.57}	47.9	{31.44-67.79}
J	11.19	{3.76; 18.63}	15.50	{8.14; 22.87}	16.18	{7.51-25.81}
L	6.37	{0.00; 13.96}	9.57	{5.66; 13.48}	13.19	{8.32-19.6}
M-Q	76.13	{56.98; 95.28}	74.22	{53.04; 95.39}	87.99	{61.39-120.52}
M-N	12.79	{0.00; 30.28}	21.61	{12.86; 30.35}	27.44	{17.22-41.07}
M	4.11	{0.00; 12.08}	6.63	{2.38; 10.89}	8.87	{5-13.96}
N	4.33	{0.00; 23.42}	4.93	{1.48; 8.38}	7.86	{4.13-12.67}
O-Q	22.23	{12.39; 32.06}	29.18	{18.71; 39.65}	34.38	{21.53-49.83}
O-P	13.75	{6.14; 21.36}	15.97	{8.95; 22.99}	17.41	{9.6-27.53}
P	4.53	{0.00; 12.32}	6.63	{2.61; 10.66}	7.54	{4.06-12.24}
Q	10.07	{3.90; 16.25}	9.86	{5.49; 14.24}	11.93	{7.28-17.95}
R	2.98	{0.00; 7.13}	6.99	{2.15; 11.82}	7.66	{3.52-13.32}

[a] Age estimates are indicated in ky.

Table S3. PAML and Beast age estimates [a] of relevant nodes in the horse phylogeny for each of the five partitions (1st, 2nd and 3rd positions of the codons, RNAs and non-coding regions).

NODE	**1st CODON POSITION**		**2nd CODON POSITION**		**3rd CODON POSITION**		**RNA**		**NON-CODING**	
PAML										
	Age estimate	**95% CI**	**Age estimate**	**95% CI**	**Age estimate**	**95% CI**	**Age estimate**	**95% CI**	**Age estimate**	**95% CI**
A-R (AMM)	195.89	{107.9; 283.87}	373.56	{211.62; 535.5}	115.81	{94.7; 136.92}	143.5	{80.8; 206.21}	144.72	{84.74; 204.71}
A-Q	119.11	{62.29; 175.92}	373.56	{211.62; 535.5}	86.14	{71.53; 100.75}	98.31	{60.87; 135.75}	124.39	{82.91; 165.87}
A-L	72.25	{45.6; 98.9}	300.45	{177.77; 423.13}	71.54	{59.05; 84.03}	80.91	{52.76; 109.07}	107.95	{71.41; 144.48}
A-K	72.25	{45.77; 98.73}	300.45	{178.75; 422.15}	67.7	{55.65; 79.76}	80.91	{52.91; 108.92}	84.47	{50.89; 118.04}
A-I	72.25	{45.94; 98.56}	300.45	{178.75; 422.15}	39.88	{31.21; 48.54}	67.09	{43.8; 90.38}	84.07	{55.07; 113.07}
A-D	45.32	{21.35; 69.28}	202.3	{61.95; 342.65}	30.61	{21.76; 39.46}	67.09	{43.65; 90.53}	84.07	{55.07; 113.07}
A-C	45.32	{21.35; 69.28}	88.63	{23.86; 153.41}	21.14	{13.44; 28.83}	67.09	{43.65; 90.53}	67.12	{36.17; 98.07}
A-B	45.32	{21.35; 69.28}	88.63	{23.86; 153.41}	9.82	{4.5; 15.15}	42.71	{11.66; 73.76}	45.45	{21.59; 69.32}
A	23	{5.57; 40.43}	88.63	{24.84; 152.43}	3.02	{0.93; 5.1}	7.14	{0; 15.36}	26.09	{9.94; 42.23}
B	10.86	{0; 23.26}	0	{0; 12.76}	5.41	{2.12; 8.71}	15.53	{0; 32.73}	26.04	{6.13; 45.96}
C	10.86	{0; 23.26}	39.56	{0; 92.56}	3.76	{0.58; 6.95}	8.85	{0; 21.03}	21.9	{4.1; 39.69}
D	16.5	{0; 37.28}	0	{0; 12.76}	0	{0; 0.25}	0	{0; 1.83}	11.77	{0.1; 23.44}
E-I	72.25	{45.94; 98.56}	300.45	{178.75; 422.15}	35.19	{27.07; 43.3}	67.09	{43.8; 90.38}	84.07	{55.09; 113.06}
E-G	72.25	{45.77; 98.73}	235.85	{103.36; 368.35}	20.81	{12.91; 28.71}	37.27	{11.09; 63.45}	46.23	{20.65; 71.81}
F-G	72.25	{45.77; 98.73}	96.64	{13.22; 180.07}	13.35	{6.85; 19.85}	22.05	{0.75; 43.36}	46.23	{20.65; 71.81}
G	26.51	{6.9; 46.11}	96.64	{14.2; 179.09}	4.51	{1.61; 7.41}	8.15	{0; 16.37}	16.58	{2.15; 31.02}
H-I	45.32	{10.12; 80.51}	300.45	{178.75; 422.15}	33.34	{24.96; 41.72}	67.09	{43.65; 90.53}	83.96	{40.02; 127.9}
I	17.96	{0; 37.73}	52.58	{0; 154.65}	7.42	{2.49; 12.34}	4.58	{0; 14.02}	31.6	{9.63; 53.57}
J-K	51.81	{20.98; 82.65}	213.32	{60.21; 366.43}	29.47	{19.45; 39.5}	62.43	{29.25; 95.61}	60.31	{29.74; 90.89}
J	51.81	{20.98; 82.65}	0	{0; 11.78}	8.5	{2.49; 14.51}	28.5	{3.08; 53.92}	22.57	{2.51; 42.62}
L	25.22	{1.93; 48.52}	80.12	{0; 174.34}	5.66	{2.98; 8.33}	23.99	{0.1; 47.89}	18.14	{7.09; 29.19}
M-Q	84.48	{38.73; 130.23}	373.56	{211.62; 535.5}	61.18	{47.25; 75.11}	77.19	{38.68; 115.69}	124.39	{82.91; 165.87}

M-N	43.95	{0; 88.53}	159.74	{0; 322.66}	10.83	{4.56; 17.1}	31.99	{2.62; 61.37}	52.48	{21.49; 83.46}
M	11.46	{0; 27.71}	0	{0; 12.76}	2.85	{0; 6}	11.57	{0; 24.96}	22.79	{1.55; 44.04}
N	0	{0; 2.35}	41.06	{0; 121.54}	4.09	{0.26; 7.92}	0	{0; 2.13}	9.2	{0.01; 18.39}
O-Q	44.72	{12.21; 77.23}	100.15	{0; 215.96}	17.94	{10.34; 25.54}	39.37	{8.32; 70.42}	66.9	{31.01; 102.78}
O-P	22.66	{0; 46.62}	100.15	{0; 214.98}	12.17	{5.92; 18.43}	15.53	{0; 37.14}	21.59	{2.08; 41.1}
P	0	{0; 2.18}	100.15	{0; 214.98}	3.9	{0.2; 7.59}	5.05	{0; 15.09}	19.52	{0; 41.85}
Q	9.83	{0; 23.41}	44.57	{0; 133.88}	8.46	{3.89; 13.03}	7.38	{0; 17.57}	21	{8.36; 33.65}
R	21.46	{0; 51.29}	70.61	{0; 203.1}	2.48	{0; 5.93}	9.71	{0; 28.73}	13.01	{0; 28.24}

BEAST

	Median	95% HPD	Median	95% HPD	Median	95% HPD	Median	95% HPD	Median	95% HPD
A-R (AMM)	139.28	{61.56-298.18}	199.75	{97.53-346.84}	205.33	{77.94-446.53}	178.72	{91.26-322.24}	218.63	{113.22-368.37}
A-Q	123.16	{57.25-251.25}	126.14	{63.57-219.49}	146.31	{64.87-306.55}	126.44	{67.08-212.69}	184.34	{94.32-300.15}
A-L	97.94	{42.47-195.41}	100.27	{50.18-177.97}	112.48	{49.86-230.33}	100.46	{54.10-168.95}	154.26	{75.35-249.39}
A-K	80.57	{38.65-161.35}	92.73	{48.41-165.41}	105.64	{49.11-217.88}	90.49	{45.44-149.18}	125.88	{63.59-206.85}
A-I	50.76	{21.93-99.52}	65.12	{31.94-116.54}	84.58	{37.16-170.00}	74.28	{39.16-122.98}	107.65	{55.78-177.77}
A-D	33.68	{12.78-69.01}	42.91	{18.56-77.08}	63.64	{27.13-129.13}	64.4	{32.34-109.43}	90.84	{42.42-151.53}
A-C	24.52	{8.53-49.65}	20.65	{8.51-41.73}	46.82	{18.55-96.70}	56.13	{24.40-96.00}	71.15	{32.54-121.55}
A-B	14.99	{5.39-31.79}	11.89	{4.56-24.04}	26.74	{8.91-57.31}	36.5	{13.55-68.64}	51.08	{22.93-90.07}
A	10.78	{4.05-23.12}	7.27	{3.09-14.38}	13.35	{4.67-29.18}	11.71	{3.39-27.48}	31.04	{13.38-58.16}
B	8.65	{2.92-18.81}	7.71	{2.98-15.68}	12.65	{4.49-28.04}	13.29	{3.09-34.46}	31.41	{12.50-61.38}
C	8.19	{2.83-18.88}	7.4	{2.69-16.02}	13.76	{4.60-31.60}	12.4	{3.43-31.83}	26.79	{9.98-55.74}
D	5.53	{0.07-14.91}	7.39	{2.23-18.59}	7.03	{0.04-17.40}	6.63	{0.41-19.88}	16.44	{5.76-34.93}
E-I	46.26	{19.68-89.72}	49.06	{21.11-88.26}	68.29	{27.82-136.94}	66	{33.52-110.81}	98.55	{50.93-165.43}
E-G	41.01	{15.69-78.73}	23.73	{8.47-46.04}	41.75	{12.19-85.29}	38.07	{13.39-71.44}	65.65	{29.10-115.29}
F-G	26.67	{9.58-52.29}	19.09	{6.90-37.46}	24.23	{7.60-52.19}	23	{6.84-46.45}	49.48	{21.03-90.61}
G	11.45	{3.94-24.68}	10.56	{3.86-22.05}	12.56	{4.22-26.17}	11.76	{3.84-26.05}	27.49	{11.44-52.27}
H-I	36	{13.66-69.87}	34.87	{13.48-65.54}	58.3	{21.41-116.6}	56.16	{24.99-96.98}	80.28	{37.01-138.32}
I	10.51	{2.68-25.19}	9.04	{2.82-21.15}	12.13	{3.49-29.49}	7.65	{2.02-20.00}	35.21	{12.92-69.03}

J-K	35.02	{8.90-73.65}	36.13	{12.11-71.90}	47.3	{14.04-101.45}	57.82	{18.17-106.31}	78.96	{30.38-138.97}
J	13.09	{2.04-36.17}	9.57	{2.45-25.59}	8.62	{0.88-21.91}	19.91	{3.47-50.49}	20.99	{6.02-48.06}
L	15.02	{4.12-35.77}	14.5	{4.58-32.09}	14.74	{4.47-35.38}	37.32	{9.59-76.59}	34.24	{14.06-65.44}
M-Q	79.05	{32.90-158.51}	90.28	{45.53-160.75}	81.14	{27.96-175.01}	87.2	{39.38-155.16}	144.71	{69.67-243.12}
M-N	13.35	{4.06-32-08}	35.84	{13.64-72.25}	22.14	{6.76-54.76}	34.84	{10.02-74.72}	65.72	{25.96-121.66}
M	7.94	{2.52-18.32}	7	{2.56-15.99}	11.91	{3.73-27.82}	11.97	{2.80-31.08}	24.06	{7.25-52.38}
N	7.96	{2.19-19.14}	8.92	{2.79-21.41}	8.9	{1.78-20.32}	9.5	{1.82-28.98}	18.42	{6.66-37.15}
O-Q	23.04	{7.31-51.88}	23.88	{9.07-49.85}	43.96	{12.11-99.46}	42.37	{12.78-82.70}	79.91	{34.75-147.79}
O-P	14.33	{3.95-33.33}	12.2	{3.17-27.17}	17.25	{4.28-43.43}	14.06	{2.91-38.16}	24.85	{7.80-53.68}
P	6.38	{1.23-15.36}	5.43	{0.57-12.04}	9.55	{2.58-21.41}	6.03	{1.17-14.63}	16.28	{5.60-34.18}
Q	14.1	{4.42-32.47}	9.88	{3.34-21.20}	13.87	{4.11-33.83}	10.88	{3.22-26.78}	29.46	{11.55-56.55}
R	5.51	{0.19-16.57}	6.05	{1.31-16.13}	8.79	{1.27-23.56}	6.93	{0.16-27.55}	15.72	{3.74-37.97}

[a] Age estimates are indicated in ky.

Table S4. Diagnostic mutational motifs of the newly defined horse mtDNA haplogroups. A comparison between the new and the old nomenclatures based on partial sequences is shown.

Revised Nomenclature of Nodes and Haplogroups	**Diagnostic Mutational Motif [a]**	**Nomenclature by Cai and Jansen[b]**	**Nomenclature by Cieslak[c]**
A1	6786, 15495, 15826	A4'A5	D2’
A	15720		A’D’D1’H’H1’
A'B	158, 356, 4062, 11242, 15650, 16371		
B	858, 5931, 7629, 9963, 10766, 16055, 16111	A3	D3
B1	5931back, 13818, 15810, 16111back		
A'B'C	1386del, 2788, 7003, 9241, 10216, 11545, 15602, 15870		
C	956, 2238, 3800, 4884, 4993, 5502, 6078, 7944, 9666, 10219, 11048, 11131, 12354, 13081, 15956, 15974, 16130, 16131, 16113		
C1	4599, 9280, 16563insC		
C2	267, 15597		
A'B'C'D	4646, 5886, 6006, 10378, 13051, 13335, 15650back, 15826back	A6 [e]	
D	1693, 2216, 5499, 6534, 8869, 9404, 11419, 11459, 12193, 12220, 12793, 13950, 14202, 14827, 15521, 15737, 15770, 15810, 16087, 16111, 16394, 16113	E	G’G1’G2’G3’G4’Gx’
A-I	4669, 6309, 8007, 11844, 12769, 15585, 15771		
E'F'G'H'I	4830, 13504, 16113, 16130, 16131		
E'F'G	381, 386, 5832, 8039, 9671, 10473, 12862, 14352, 15542, 15650, 15666		
***E*[d]**	*278, 1133, 2938, 4969, 8354, 9333, 10402, 10688, 12655, 12834, 13261, 14208, 14910, 15597, 15956, 15974, 16644*		
F'G	415, 2940, 3576, 11167, 14628, 14736, 15585back, 15771back	A7 [f]	X3’X3a’X3b’X3c2’X3d’
***F*[d]**	*3371, 3562, 7506, 7748, 8791, 11822, 13044, 15595, 15868, 16113back, 16514*	*A2*	
G	222, 3053, 5500, 5883, 6690, 9404, 9743A, 11554, 13225, 15635, 15703	A1	X3c’X3c1’
G1'G2'G3	15597		
G1	5671, 14653, 15585, 16031		
G2	9197		
G3	7689, 11545back		
H'I	5217del, 10085, 13763, 15974, 16563insC		
***H*[d]**	*1545, 1667, 3064, 4182, 5273, 6078, 6129, 6567, 6837, 9211, 10916*		*X4’*
I	1586, 1790, 2614, 2770, 4063, 4392, 5061, 5210, 6177, 6249, 8360, 8381, 8794, 9950, 10240, 11829, 12445, 12685, 14556, 15538, 15709, 15826	B1'B2	I’I1’I2’
I1	13468, 14040, 15870back		
A-K	960, 3070, 3942, 4536, 5529, 7668, 8360, 8567, 10112, 11881, 14805		
J'K	1154, 2863, 3963, 4296, 4332, 4608, 6390, 6879, 7059, 7203, 8084, 8444, 9073, 9398, 10219, 12334, 13231, 13338, 13950, 14397, 14553, 15650, 15807, 15871, 16071		
J	160, 383T, 2263, 4188, 4785, 6474, 7578, 8291, 9250, 11083, 11823, 11935, 13126, 15532del, 15604, 15827, 16111, 16644		
***K*[d]**	*2803, 3283, 3364, 3424, 4893, 5571, 5697, 8077, 8587, 8633, 8662, 9395, 9794, 10417, 11002, 11852, 12709, 14177, 15012, 15173, 15229, 15544, 15585back, 15635, 15683, 15686, 15775, 15956, 16130, 16131, 16331, 16339, 16347*		*X5’X6b’*
A-L	7902, 10294, 15956		
L	1374, 2899, 3070back, 3517, 5817, 7518, 8060, 8303, 8321, 9953, 10423, 10615, 11695, 12121, 12202, 12898, 12952, 13522, 14997, 15315, 15494, 15496, 15534, 15603, 15649,	D'D1'D2'D3 [g]	X1’X2’

	15871, 15974, 16068, 16103, 16407del, 16563insC, 16629		
L1	11684		
L2	2607, 8405, 15602back		
L3	77		
A-Q	3800, 8045, 8152, 9088, 10175, 10450, 11968, 13922, 13935, 15585back, 15806, 15870back, 16130, 16131, 16543A		
M'N'O'P'Q	4605, 5829, 6714, 8078, 8240, 8558T, 10861, 11396, 11494, 12334, 13631, 14628, 15344, 16121, 16629		
M'N	1608T, 2339A, 3475, 3942back, 4526, 4898, 6078, 7434, 8177, 8800, 9334, 9542, 10462, 10648, 12031, 12097, 13102, 13358, 13504, 13617, 13722, 14424, 14673, 14817, 15054A, 15135, 15827, 15869, 16068, 16113, 16546, 16559		
M	426, 3100, 4599G, 5103, 9241back, 15617, 15659, 16080	C1	B’B1’B2’B3’
N	2802, 3557, 4917, 8282, 9073, 9404, 10406, 10519, 13569, 15601, 15838, 16007, 16111	C2	F
O'P'Q	738, 859, 3259T, 3271, 3616, 6531, 7245, 7614T, 7900, 8045back, 8363, 8857, 9777, 11380, 11426, 12169, 12232, 12406, 13465, 15204, 15703, 15777, 15806back, 16038, 16063insC		
O'P	1683C, 3557, 7429, 9055, 12289, 13468, 15667, 15809, 16563insC, 16407del	F1	K3’
O[d]	*5823, 6747, 7377, 8042, 10363T, 10450back, 10503, 12307, 12929, 13949, 13953, 14827, 15597, 15604, 15635, 16113, 16360*		*K3a1’*
P	1382, 6507, 13372		
Q	302, 341, 4201, 7296, 9205, 15740, 15811, 16037, 16057, 16113	F2'F3	K2’
Q1	6690		
Q2	5877, 15995, 16563insC		
Q3	11944, 13234, 15585, 15604		
A-R	287, 724, 2158, 3058, 3848, 4306, 6093, 6300, 7249, 7479, 7726, 8492, 9220, 9403, 10408, 12085, 12316, 12628, 12718, 12802, 13698, 14025, 14898, 15144, 15771back, 15775, 15827		
R	2316, 2869, 2887, 3052, 4104, 4113, 4275, 4459, 4596, 4918, 5380, 5418, 5457, 5671, 6177A, 6414, 6600, 6726, 6932, 6949, 7449, 7575, 7578, 7914, 8363, 8504, 8680, 8971, 9304T, 9620, 10369, 10753, 10888, 10927, 11180, 11385, 11616, 12430, 12943, 13213, 13216, 13277T, 13436, 13768, 14116, 14827, 14886, 14955, 15078, 15108, 15249, 15252T, 15598, 15615, 15616, 15659, 15703, 15770, 15776, 15974, 15996, 16103, 16476, 16540, 16629	G	X7

[a] Mutational motifs are relative to the new proposed *E. caballus* reference sequence (HRS, see Fig. S1 and Table S1 for additional details), which is a member of an A sub-haplogroup.

[b] The old nomenclature (4, 5) was haplotype-oriented and cannot be used for phylogenetic naming of branches.

[c] The "new" nomenclature introduced by Cieslak et al. (6) is even less preferable than the old one.

[d] An accurate diagnostic motif cannot be determined yet due to the availability of only one sequence.

[e] A6 would correspond to a paragroup.

[f] A7 is probably a sister group to F'G.

[g] D’D1, D2, and D3 might be further sub-classified as L1 (11684), L2 (15602@), and L3 (15604).

Table S5. List of mtDNAs from Przewalski's horses.

GenBank Acc. Numb.	Control-Region Haplotype (*vs* HRS)	Hg	Sequence Range	Reference	PubMed Identifier
AF014409	15386d 15495 15542 15595 15602 15650 15666 15720 15868 15870 16130 16131 16213-16276d 16371 16391A 16407A	**F**	15369-16660	(7)	10376300
AF055876	15542 15595 15602 15650 15666 15720 15868 15870	**F**	15521-16120	(8)	9883508
AF055877	16514	**F**	16377-16540	(8)	9883508
AF055879	HRS	?	16377-16540	(8)	9883508
AF072994	15495 15542 15595 15602 15650 15666 15676N 15678N 15720 15868 15870 16063N	**F**	15468-16076	(9)	
AF072995	15495 15542 15595 15602 15650 15666 15720 15868 15870 15974 15975 16063N	**F**	15468-16074	(9)	
AF326635	15495 15542 15595 15602 15650 15666 15720	**F**	15479-15828	(10)	11161199
AJ413830	15495 15508A 15542 15595 15602 15650 15666 15720	**F**	15427-15740	(5)	12130666
AJ413831	15495 15508A 15542 15595 15602 15650 15666 15720	**F**	15427-15740	(5)	12130666
AJ413832	15495 15542 15595 15602 15650 15666 15720	**F**	15427-15740	(5)	12130666
AP012267 [a]	15495 15542 15595 15602 15650 15666 15720 15868 15870 16130 16131N 16211-16226N 16228-16327N 16371 16514 16591-16593N 16595N	**F**	15469-16660	(2)	21803766
AP012268 [a]	15495 15542 15595 15602 15650 15666 15720 15868 15870 16130 16131N 16245-16248N 16251N 16264N 16270-16326N 16328N 16371 16421-16424N 16435N 16453N 16514 16593-16594N	**F**	15469-16660	(2)	21803766
AP012270 [a]	15495 15542 15595 15602 15650 15666 15720 15868 15870 16130-16131N 16208-16254N 16291-16339N 16371 16392-16422N 16424N 16457-16460N 16509N 16514 16554-16555N 16557N 16574-16601N 16603-16604N	**F**	15469-16660	(2)	21803766
Prz001 & Prz002	15495 15542 15595 15602 15650 15666 15720 15868 15870 16130 16131 16371 16514	**F**	15469-16660	This study	
AF055878	15569 15585 15602 15720 15771 15775 15870 15871 16037 16113	**(pre)J K**	15521-16120	(8)	9883508
AP012269 [a]	15495 15569 15585 15602 15720 15771 15775 15870 15871 16037 16113 16130 16131N 16284N 16288-16327N 16371 16557N	**(pre)J K**	15469-16660	(2)	21803766
DQ900930 [b]	*15495 15595 15597 15602 15604 15703 15720 15740*	**F**	15493-15756	(11)	

[a] For these samples, complete mtDNA sequences are available.

[b] This is the only ancient specimen bearing an mtDNA haplotype likely belonging to haplogroup F, similarly to modern Przewalski's horses.

Table S6. References and PubMed identifiers for the analyzed control-region sequences from modern and ancient horse samples.

Reference	PubMed IDentifier	Number of Subjects
MODERN DNA DATA		
(12)	20926431	104
(13)	19789983	29
(14)	19744143	116
(15)	18680492	43
(16)	17897600	12
(17)	17707216	5
(18)	17265183	289
(19)	17132900	3
(20)	16978181	213
(21)	16489143	88
(22)	16251517	96
(23)	15932397	172
(24)	15496286	34
(25)	12427390	25
(26)	12354144	31
(27)	12139508	101
(28)	12137332	65
(5)	12130666	307
(10)	11161199	32
(29)	10690354	24
(30)	10612231	9
(7)	10376300	13
(8)	9883508	2
(31)	7985837	3
(9)	-	261
This study	-	186
Total		**2263**

ANCIENT DNA DATA (from Pleistocene to Middle Ages)		
(6)	21187961	52
(32)	19943892	4
(13)	19789983	26
(4)	-	22
(11)	-	8
(20)	16978181	4
(33)	15932398	7
(34)	15040002	3
(10)	11161199	4
(9)	-	8
Total		**138**

Table S7. Haplogroup affiliation of modern horse control-region sequences from different geographic areas [a].

Haplogroup [b]	**Geographic Area** [c]							
	Africa	**America North**	**America South**	**Asia**	**Europe**	**Middle East**	**Unspecified**	**Total**[b]
	N = 27	**N = 104**	**N = 71**	**N = 587**	**N = 1249**	**N = 192**	**N = 33**	**N = 2263**
A	…	3 (2.88)	3 (4.23)	70 (11.93)	56 (4.49)	15 (7.81)	3 (9.09)	150 (6.63)
B	2 (7.41)	24 (23.08)	9 (12.68)	10 (1.70)	117 (9.38)	21 (10.94)	3 (9.09)	186 (8.22)
C	…	…	…	21 (3.58)	5 (0.32)	6 (3.13)	…	32 (1.37)
D	1 (3.70)	…	…	17 (2.90)	57 (4.57)	4 (2.08)	2 (6.06)	81 (3.58)
E	0	1 (0.96)	…	13 (2.21)	6 (0.48)	1 (0.52)	…	21 (0.93)
E-G general (non-F)	…	…	…	4 (0.68)	13 (1.04)	2 (1.04)	…	19 (0.84)
F	…	…	…	18 (3.07)	…	…	…	18 (0.80)
G	…	7 (6.73)	2 (2.82)	96 (16.35)	109 (8.73)	16 (8.33)	3 (9.09)	233 (10.30)
H	…	…	…	8 (1.36)	13 (1.04)	…	1 (3.03)	22 (0.97)
I	1 (3.70)	…	4 (5.63)	36 (6.13)	100 (8.01)	22 (11.46)	5 (15.15)	168 (7.43)
J-K	…	1 (0.96)	1 (1.41)	38 (6.47)	7 (0.56)	7 (3.65)	…	54 (2.39)
L	21 (77.78)	61 (58.65)	34 (47.89)	79 (13.46)	475 (38.06)	43 (22.40)	8 (24.24)	721 (31.87)
M	…	3 (2.88)	4 (5.63)	24 (4.09)	91 (7.29)	10 (5.21)	1 (3.03)	133 (5.88)
M'N	…	…	…	5 (0.85)	1 (0.08)	…	…	6 (0.27)
N	1 (3.70)	3 (2.88)	9 (12.68)	17 (2.90)	106 (8.49)	6 (3.13)	1 (3.03)	143 (6.32)
O'P	1 (3.70)	…	3 (4.23)	39 (6.64)	17 (1.36)	16 (8.33)	3 (9.09)	79 (3.49)
Q	…	1 (0.96)	2 (2.82)	81 (13.80)	48 (3.85)	20 (10.42)	3 (9.09)	155 (6.85)
R	…	…	…	11 (1.87)	28 (2.24)	3 (1.56)	…	42 (1.86)

[a] MtDNA control-region sequences are from GenBank. Reference sources are listed in Table S6.

[b] Haplogroup classification is based on control-region motifs, as reported in Table S4.

[c] Haplogroup frequencies (%) are in parentheses. Sequences with an ambiguous haplogroup affiliation were not included.

Table S8. Haplogroup affiliation of ancient horse control-region sequences from different continents [a].

Haplogroup	**Geographic Area**		
	Asia	**Europe**	**Total**
	N=68	**N=70**	**N=138**
A	6 (8.82)	8 (11.43)	14 (10.14)
D	7 (10.29)	1 (1.43)	8 (5.80)
E'G	15 (22.06)	7 (10.00)	22 (15.94)
F	1 (1.47)	…	1 (0.72)
H	4 (5.88)	2 (2.86)	6 (4.35)
I	2 (2.94)	6 (8.57)	8 (5.80)
K	2 (2.94)	2 (2.86)	4 (2.90)
L	7 (10.29)	15 (21.43)	22 (15.94)
M	2 (2.94)	12 (17.14)	14 (10.14)
N	3 (4.41)	7 (10.00)	10 (7.25)
O'P	7 (10.29)	2 (2.86)	9 (6.52)
Q	7 (10.29)	3 (4.29)	10 (7.25)
R	5 (7.35)	5 (7.14)	10 (7.25)

[a] MtDNA control-region sequences are from GenBank. Reference sources are listed in Table S6.
[b] Haplogroup classification is based on control-region motifs, as reported in Table S4.
[c] Haplogroup frequencies (%) are in parentheses. Sequences with an ambiguous haplogroup affiliation were not included.

Table S9. Geographic origin, dating and haplogroup affiliation of horse control-region sequences from ancient horse samples predating domestication times [a].

GenBank #	Continent /Geographic Area	Age of the Specimen	Hg	Reference	PubMed Identifier
FJ204318	Asia (Siberia)	Late Pleistocene	**E-G**	(6)	21187961
FJ204324	Asia (North Siberia)	Late Pleistocene	**E-G**	(6)	21187961
FJ204351	Europe (Germany)	Late Glaciation-Mesolithic (15-14000 BCE)	**M**	(6)	21187961
FJ204352	Europe (Germany)	Late Glaciation-Mesolithic (14-11000 BCE)	**M**	(6)	21187961
FJ204382	Europe (Spain)	Mesolithic/Neolithic (5200-4900 BCE)	**M**	(6)	21187961
DQ683528	Europe (Iberia)	Neolithic/Bronze Age	**M**	(32)	19943892
DQ683526	Europe (Iberia)	Neolithic/Bronze Age	**A**	(32)	19943892
DQ683532	Europe (Iberia)	Neolithic/Bronze Age	**H**	(32)	19943892
DQ683531	Europe (Iberia)	Neolithic/Bronze Age	**L**	(32)	19943892

[a] These are a subset of the ancient samples listed in Table S6.

Table S10. Oligonucleotide pairs used to amplify the horse mitochondrial genome in 11 overlapping PCR fragments[a].

PCR ID Number	Fragment Length (bps)	Name [b]	Sequence (5' → 3')	Tm (°C)
1	1989	175For	AGGAGCAGGTATCAAGCACAC	59.25
		2163Rev	TTAGTCATTCCCGCCTCTTC	59.27
2	1944	1679For	TCCCAACAATCAACCCAAAC	59.35
		3622Rev	TTGGTCATATCGGAATCGTG	59.37
3	2028	3124For	ATCAGGATGGGCCTCAAAC	59.46
		5151Rev	AATTTCGTGGGATGGTAGCC	59.57
4	2085	4842For	GACCATATTCCCATCCACAAAC	59.61
		6926Rev	GGGTTCGATTCCTTCCTTTC	59.75
5	1894	6544For	CACTGATTCCCTCTATTCTCAGG	59.79
		8437Rev	ACGGCTAGGGCTACAGGTTG	59.87
6	1992	7978For	TATTCGCCTCTTTCGCTACC	59.88
		9969Rev	GGCCTACGAGGGATACTGTG	59.88
7	1878	9538For	CCTCGCTACTCGTACTCATCG	59.98
		11415Rev	GCGGTGATGGTGATATTGG	60.05
8	1799	11096For	ATCGTAGCCGTCCTCATCC	60.05
		12894Rev	GTGGTGGTGAAGGGTATTGC	60.24
9	1852	12546For	CAGGCGTCTTCCTGCTAATC	60.28
		14397Rev	GCAGATGTGAGTGACGGATG	60.31
10	1867	14001For	CCCTCCACTTACAATCAGCAC	60.44
		15867Rev	ATGGCCCTGAAGAAAGAACC	60.59
11	1860	15364For	AAACCAGAAAAGGGGGAAAA	61.07
		00563Rev	TGGCGAATAGCTTTGTTGTG	61.56

[a] The annealing temperature for all PCR reactions is 55°C.
[b] Primer names correspond to the (5') nucleotide position in HRS.

Table S11. Oligonucleotides used for sequencing the horse mitochondrial genome.

PCR ID Number	**Name [a]**	**Sequence (5’ → 3’)**	**Tm (°C)**
1	204For	AGCTCATAACACCTTGCTCA	56.09
1	715For	CAAGGTACCGAAGTAAGCAC	55.08
1	1234For	TAAGGGAACGATGAAAGATG	55.31
2	1729For	TCAACACATAGAAGCAATAATGT	54.26
2	2224For	CCTTAACCTTCAGGGACAAC	56.23
2	2834For	TTGAACGAAAAGTCTTAGGC	54.85
3	3231For	TCCTAATAAGCGGATCATTC	54.50
3	3800For	TACAGGAATTGAACCTGCTC	55.37
3	4341For	ACTAGCCCCAATATCAATCC	55.61
4	4913For	CCCCGTTAATTGTTATATCC	53.74
4	5437For	GCTGGAATAGTAGGAACTGC	54.21
4	6042For	TCCAATCCTTTATCAACACC	54.97
5	6636For	CTTCCCACAACATTTCCTT	55.02
5	7119For	CGACCACACACTAATAATCG	54.14
5	7661For	GCTTTATACCAATTGTCCTTG	54.16
6	8081For	CAATCGCCTAATCTCAATTC	54.98
6	8591For	CCAAGCCTACGTATTCACTC	54.99
6	9128For	GGCCTATTCATCACAATTTC	54.63
7	9604 For	CATATGAATGCGGATTTGAC	56.47
7	10171For	GAGTAGACCACGTACAAAACC	53.90
7	10691For	GCACTAATCTCTATCCAAAACC	54.79
8	11161For	CGGCCTTACATCATCAATAC	55.63
8	11650For	CCGAGAAAGTATGCAAGAAC	55.12
8	12098For	TGATACATGCACTCAGATCC	54.35
9	12600For	TCCAGTCACTTACCCTATGC	55.24
9	13141For	CATCAAACGCCTCTTAATTG	56.00
9	13686For	AGCCCCTATAATTTCCTCAC	54.94
10	14056For	AACCCCACAAAACTAACAAC	54.15
10	14496For	CTACGGCTCTTACACATTCC	54.99
10	15003For	GTACTTCCTGTTTGCCTACG	55.08
11	15433For	CACCCAAAGCTGAAATTCTA	55.55
11	16656Rev	GAAGAAGGGTTGACAGATTTAG	54.83
11	418Rev	TACGGCTTAACTGGGTTTTA	55.20

[a] Primer names correspond to the (5’) nucleotide position in HRS.

Supporting References

1. Nergadze SG, et al. (2010) Mitochondrial DNA insertions in the nuclear horse genome. *Anim Genet* 41 Suppl 2:176-185.
2. Goto H, et al. (2011) A massively parallel sequencing approach uncovers ancient origins and high genetic variability of endangered Przewalski's horses. *Genome biology and evolution*.
3. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574.
4. Cai D, et al. (2009) Ancient DNA provides new insights into the origin of the Chinese domestic horse. *J Archaeol Sci* 36:835-842.
5. Jansen T, et al. (2002) Mitochondrial DNA and the origins of the domestic horse. *Proc Natl Acad Sci U S A* 99:10905-10910.
6. Cieslak M, et al. (2010) Origin and history of mitochondrial DNA lineages in domestic horses. *PLoS One* 5:e15311.
7. Kim KI, et al. (1999) Phylogenetic relationships of Cheju horses to other horse breeds as determined by mtDNA D-loop sequence polymorphism. *Anim Genet* 30:102-108.
8. Oakenfull EA, Ryder OA (1998) Mitochondrial control region and 12S rRNA variation in Przewalski's horse (*Equus przewalskii*). *Anim Genet* 29:456-459.
9. GenBank (2011) in *http://www.ncbi.nlm.nih.gov*.
10. Vilà C, et al. (2001) Widespread origins of domestic horse lineages. *Science* 291:474-477.
11. Cai D, et al. (2007) Mitochondrial DNA analysis of Bronze Age horses recovered from Chifeng region, Inner Mongolia, China. *Prog Nat Sci* 17:544-550.
12. Bower MA, et al. (2010) The cosmopolitan maternal heritage of the Thoroughbred racehorse breed shows a significant contribution from British and Irish native mares. *Biol Lett* 7:316-320.
13. Priskin K, et al. (2010) Mitochondrial sequence variation in ancient horses from the Carpathian Basin and possible modern relatives. *Genetica* 138:211-218.

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

15. Pérez-Gutiérrez LM, De la Pena A, Arana P (2008) Genetic analysis of the Hispano-Breton heavy horse. *Anim Genet* 39:506-514.

16. Glazewska I, Wysocka A, Gralak B, Sell J (2007) A new view on dam lines in Polish Arabian horses based on mtDNA analysis. *Genet Sel Evol* 39:609-619.

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

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

19. Iwanczyk E, Juras R, Cholewinski G, Cothran EG (2006) Genetic structure and phylogenetic relationships of the Polish Heavy horse. *J Appl Genet* 47:353-359.

20. McGahern AM, et al. (2006) Mitochondrial DNA sequence diversity in extant Irish horse populations and in ancient horses. *Anim Genet* 37:498-502.

21. Luis C, Bastos-Silveira C, Cothran EG, Oom Mdo M (2006) Iberian origins of New World horse breeds. *J Hered* 97:107-113.

22. Royo LJ, et al. (2005) The origins of Iberian horses assessed via mitochondrial DNA. *J Hered* 96:663-669.

23. Lopes MS, et al. (2005) The Lusitano horse maternal lineage based on mitochondrial D-loop sequence variation. *Anim Genet* 36:196-202.

24. Cozzi MC, et al. (2004) Mitochondrial D-loop sequence variation among Italian horse breeds. *Genet Sel Evol* 36:663-672.

25. Kavar T, Brem G, Habe F, Solkner J, Dovc P (2002) History of Lipizzan horse maternal lines as revealed by mtDNA analysis. *Genet Sel Evol* 34:635-648.

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

27. Hill EW, et al. (2002) History and integrity of thoroughbred dam lines revealed in equine mtDNA variation. *Anim Genet* 33:287-294.

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

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

30. Kavar T, Habe F, Brem G, Dovc P (1999) Mitochondrial D-loop sequence variation among the 16 maternal lines of the Lipizzan horse breed. *Anim Genet* 30:423-430.

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

32. Lira J, et al. (2010) Ancient DNA reveals traces of Iberian Neolithic and Bronze Age lineages in modern Iberian horses. *Mol Ecol* 19:64-78.

33. Keyser-Tracqui C, et al. (2005) Mitochondrial DNA analysis of horses recovered from a frozen tomb (Berel site, Kazakhstan, 3rd Century BC). *Anim Genet* 36:203-209.

34. Di Bernardo G, et al. (2004) Genetic characterization of Pompeii and Herculaneum Equidae buried by Vesuvius in 79 AD. *J Cell Physiol* 199:200-205.