

19. B. J. Finlayson-Pitts, J. N. Pitts, *Chemistry of the Upper and Lower Atmosphere—Theory, Experiments, and Applications* (Academic Press, San Diego, CA, 2000).
20. J. F. Hopper, W. Hart, *J. Geophys. Res.* **99**, 25315 (1994).
21. U. K. Klänning, T. Wolff, *Ber. Bunsenges. Phys. Chem.* **89**, 243 (1985).
22. Some indirect light (twilight) is present at midday before 2 March, and even earlier further to the south, so the Br₂ and BrCl increases on day 60 to 61 could be photoinitiated.
23. K. W. Oum, M. J. Lakin, B. J. Finlayson-Pitts, *Geophys. Res. Lett.* **25**, 3923 (1998).
24. J. P. D. Abbatt, G. C. G. Waschewsky, *J. Phys. Chem. A* **102**, 3719 (1998).
25. R. Sander, P. J. Crutzen, *J. Geophys. Res.* **101**, 9121 (1996).
26. B. A. Michalowski et al., *J. Geophys. Res.* **105**, 15131 (2000).
27. T. X. Wang, D. W. Margerum, *Inorg. Chem.* **33**, 1050 (1994).
28. T. Koop, A. Kapilashrami, L. T. Molina, M. J. Molina, *J. Geophys. Res.* **105**, 26393 (2000).
29. S. Ghosal, A. Shbeeb, J. C. Hemminger, *Geophys. Res. Lett.* **27**, 1879 (2000).
30. E. M. Knipping et al., *Science* **288**, 301 (2000).
31. D. Toom-Sauntry, L. A. Barrie, *J. Geophys. Res.*, in press.
32. P. B. Shepson and D. Toom-Sauntry, unpublished data.
33. W. P. Bartlett, D. W. Margerum, *Environ. Sci. Technol.* **33**, 3410 (1999).
34. J. J. Orlando, B. Ramacher, G. Tyndall, *Geophys. Res. Lett.* **27**, 2633 (2000).
35. C. W. Spicer et al., *Nature* **394**, 353 (1998).
36. R. R. Dickerson et al., *J. Geophys. Res.* **104**, 21385 (1999).
37. C. W. Spicer, D. V. Kenny, W. J. Shaw, K. M. Busness, E. G. Chapman, *Environ. Sci. Technol.* **28**, 412A (1994).
38. We are grateful to the Department of Energy's Atmospheric Chemistry Program, and particularly P. Lunn, for support of this work; the Meteorological Service of Canada for providing transport to Alert and housing during the study; and the personnel of Canadian Forces Station Alert for logistical support. We are especially grateful for the on-site support provided by A. Gallant and P. Brickell, organizational support and comments on the manuscript by L. Barrie, use of a snow probe developed by A. L. Sumner, and ozone calibrations provided by K. Anlauf. We also thank J. N. Pitts Jr., J. C. Hemminger, P. Jungwirth, D. Tobias, and R. B. Gerber for helpful discussions and M. J. Molina for providing preprints before publication.

29 August 2000; accepted 6 December 2000

Widespread Origins of Domestic Horse Lineages

Carles Vilà,^{1*} Jennifer A. Leonard,² Anders Götherström,³ Stefan Marklund,⁴ Kaj Sandberg,⁴ Kerstin Lidén,³ Robert K. Wayne,² Hans Ellegren¹

Domestication entails control of wild species and is generally regarded as a complex process confined to a restricted area and culture. Previous DNA sequence analyses of several domestic species have suggested only a limited number of origination events. We analyzed mitochondrial DNA (mtDNA) control region sequences of 191 domestic horses and found a high diversity of matriline. Sequence analysis of equids from archaeological sites and late Pleistocene deposits showed that this diversity was not due to an accelerated mutation rate or an ancient domestication event. Consequently, high mtDNA sequence diversity of horses implies an unprecedented and widespread integration of matriline and an extensive utilization and taming of wild horses. However, genetic variation at nuclear markers is partitioned among horse breeds and may reflect sex-biased dispersal and breeding.

The domestication of the horse has profoundly affected the course of civilization. Horses provided meat, milk, and enhanced transportation and warfare capabilities that led to the spread of Indo-European languages and culture and the collapse of ancient societies (1, 2). Horse remains become increasingly common in archaeological sites of the Eurasian grassland steppe dating from about 6000 years ago, suggesting the time and place of their first domestication (3–5). Two alternative hypotheses for the origin of the domestic horse from wild populations can be formulated. A restricted origin hypothesis postulates that the domestic horse was developed through selective breeding of a limited

wild stock from a few foci of domestication. Thereafter, domestic horses would have been distributed to other regions. Under this hypothesis, domestication is a complex and improbable process requiring multigeneration selection on traits that permit stable coexistence with humans. Another alternative could be that domestication involved a large number of founders recruited over an extended time period from throughout the extensive Eurasian range of the horse. In this multiple origins scenario, horses may have been independently captured from diverse wild populations and then increasingly bred in captivity as wild numbers dwindled. Consequently, early domestic horses may not represent a stock highly modified by selective breeding.

These two hypotheses for the origin of the domestic horse make distinct predictions with regard to genetic variation in maternally inherited mtDNA. The restricted origin hypothesis predicts that mitochondrial diversity of the horse should be limited to a few founding lineages and those added subsequently by mutation. In contrast, a multiple origins hypothesis predicts diversity greater than that

typically found in a single wild population and divergence among lineages that well precedes the first evidence of domestication.

Phylogenetic analysis of 37 different mtDNA control region sequences from domestic horses deposited in GenBank, 616 base pairs (bp) in length (6), revealed at least six divergent sequence clades (clades A to F, Fig. 1A). After correcting for multiple hits and ignoring indels, the mean divergence observed between sequences was 2.6% (range: 0.2 to 5%). The average divergence between donkey (*Equus asinus*) and horses was 16.1% (range: 14.3 to 19.1%). Assuming that horses diverged from the lineage leading to extant stenoid equids (zebras and asses) at least 2 million years ago (Ma), as the fossil record suggests (7), or about 3.9 Ma, according to molecular data (8), we can estimate an average rate of equid mtDNA sequence divergence of 4.1% or 8.1% per million years. Therefore, modern horse lineages coalesce at about 0.32 or 0.63 Ma, long before the first domestic horses appear in the archaeological record (4). Even clade D, having a more recent coalescence time, has a mean sequence divergence of 0.8% (range: 0.2 to 2.0%), which predicts an origin at least 0.1 Ma. These results show that domestic horse lineages have an ancient origin. Thus, given the 6000-year origin suggested by the archaeological record, numerous matriline must have been incorporated into the gene pool of the domestic horse.

To expand the representation of modern and ancient breeds, we sequenced 355 bp of the left domain of the mtDNA control region in 191 horses from 10 distinct breeds (9), including some that are very old such as the Icelandic pony, Swedish Gotland Russ, and British Exmoor pony. A Przewalski's horse was also sequenced. We found 32 different sequences, and a search of GenBank provided 38 additional haplotypes for the same region. We compared all these sequences with those obtained from DNA isolated from long bone remains of eight horses preserved frozen in Alaskan permafrost deposits from a locality near Fairbanks, Alaska, dated 12,000 to 28,000 years ago (10, 11). Additionally, we sequenced DNA of eight

¹Department of Evolutionary Biology, Uppsala University, Norbyvägen 18D, S-75236 Uppsala, Sweden.

²Department of Organismic Biology, Ecology and Evolution, University of California, Los Angeles, CA 90095–1606, USA. ³Archaeological Research Laboratory, Stockholm University, S-10691 Stockholm, Sweden. ⁴Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, S-75007 Uppsala, Sweden.

*To whom correspondence should be addressed. E-mail: carles.vila@ebc.uu.se

REPORTS

ceeds that between late Pleistocene and modern sequences suggest a massive and unprecedented retention of ancestral matriline. The observation that the sequences obtained from the late Pleistocene Alaskan horses cluster in two discrete groups indicates that the diversity of mtDNA lineages in single natural populations might have been limited. Consequently, the high diversity of matriline observed among modern horses suggests the utilization of wild horses from a large number of populations as founders of the domestic horse. A single geographically restricted population would not suffice as founding stock.

Although the initial founding of the domestic horse involved incorporation of multiple matriline, the development of phenotypically distinct breeds may be a different process characterized by a limited founding stock and restricted breeding (18). Consistent with limited genetic exchange among breeds, we found that haplotype frequencies differed significantly in all but one of 45 pairwise breed comparisons

(exact test, $P < 0.05$) (19). However, the sequences found in the 10 sampled modern and ancient breeds never defined monophyletic groups as might be expected if they were derived from a limited founding stock (Fig. 2A). Additionally, genetic diversity within breeds is high; on average, 7.4 haplotypes occurred per breed (range: 3 to 9), and the nucleotide diversity per breed averaged 0.022 (range: 0.012 to 0.027), comparable to that found in large wild ungulate populations (20). Moreover, genetic diversity within breeds is not a recent phenomenon, because Viking Age horse bones 1000 to 2000 years old and from a restricted area in northern Europe also show a diversity of control region lineages (Fig. 1B). The high haplotypic diversity of ancient horse breeds and of a Viking Age population suggests that domestic horse populations were founded by a diversity of matriline that, as suggested by the archaeological record, was augmented by trade (21, 22).

High levels of mtDNA variation within and between horse breeds may reflect a bias

toward females in breeding and trade. Consequently, we assessed variation in 15 hypervariable, biparentally inherited, microsatellite loci (23–27). The observed microsatellite heterozygosity was moderately high for all breeds, varying from $49.4\% \pm 5.7\%$ to $62.6\% \pm 5.8\%$, and the mean number of alleles per locus varied from 3.6 ± 0.3 to 4.5 ± 0.4 . Allele frequencies differed significantly between breeds (exact test, $P < 0.01$). The microsatellite divergence between breeds was more marked than observed with mtDNA sequences. Individuals from the same breed generally clustered together in a neighbor-joining tree based on allele sharing distance (Fig. 2B), and 95% of individuals could be classified correctly to breed with an assignment test (28). These results show that maternal gene flow has dominated the genetic exchange between breeds and/or that female effective population size within breeds is larger than that of males. A sex bias in ancient breeding is consistent with modern

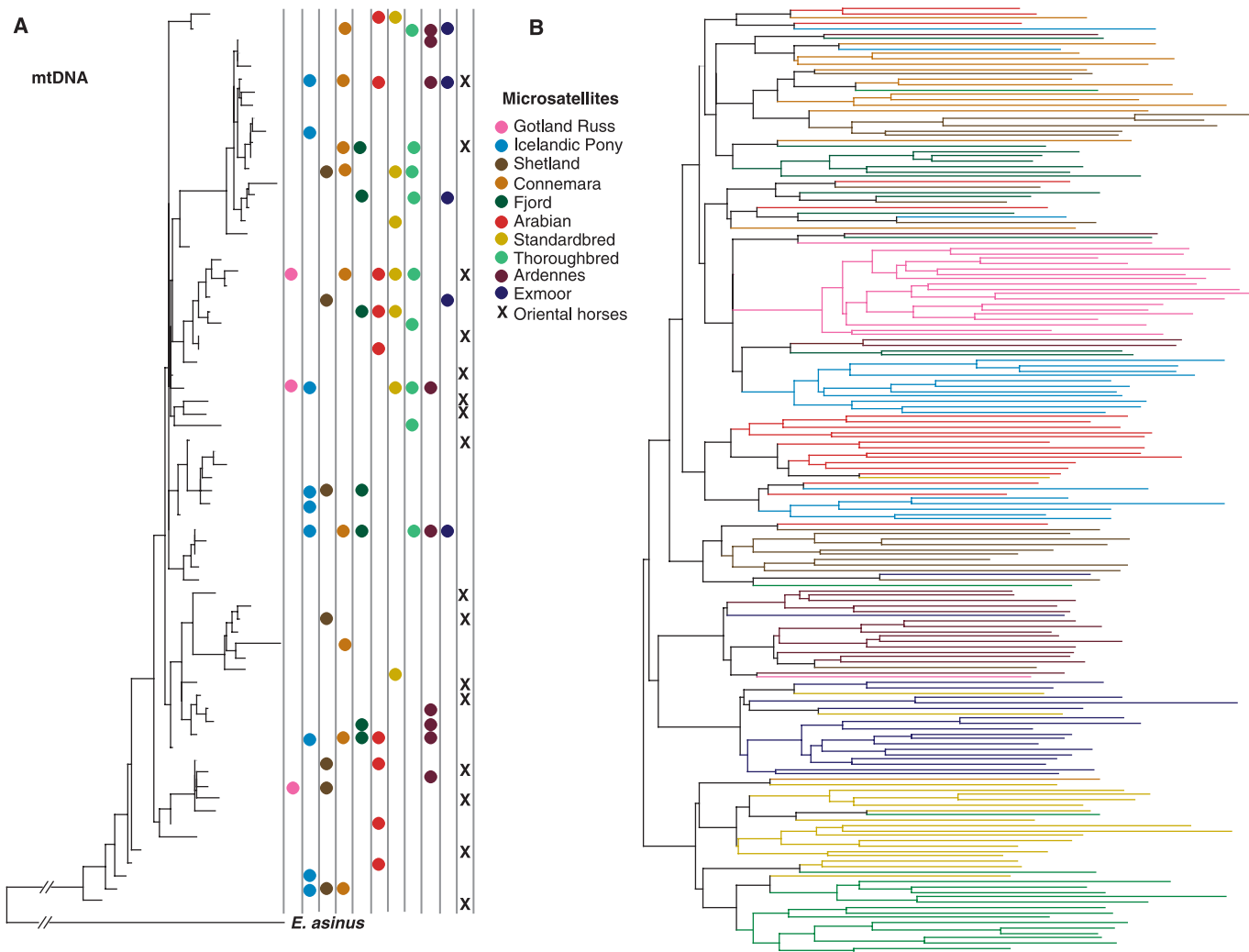


Fig. 2. Genetic diversity in horse breeds. **(A)** Distribution of haplotypes found in 10 horse breeds and in horses of the Orient from GenBank. This distribution is superimposed on a neighbor-joining tree of modern horse haplotypes based on the same 355 bp of the control region sequence, as

in Fig. 1B. Each one of the observed haplotypes is indicated by a colored symbol corresponding to its position in the relationship tree. **(B)** Unrooted neighbor-joining tree of horses from 10 breeds as in **(A)** based on the proportion of shared alleles between individuals.

REPORTS

breeding practices in which select breeding males are used as stud for 15 to 20 or more females (29).

Wild horses were widely distributed throughout the Eurasian steppe during the Upper Paleolithic [35,000 to 10,000 years before present (B.P.)], but in many regions, they disappeared from the fossil record about 10,000 years ago (3, 4). Horse remains became increasingly frequent in archaeological sites of southern Ukraine and Kazakhstan starting about 6000 years ago, where limited evidence from bit wear on teeth suggests that some horses could have been ridden (29, 30). By the beginning of the Iron Age, wild horse populations had declined, and today, only one putative wild population, the Przewalski's horse, remains (4). Therefore, a scenario consistent with the archaeological record and genetic results posits that, initially, wild horses were captured over a large geographic area and used for nutrition and transport. As wild populations dwindled because of exploitation or environmental changes (31), increased emphasis was placed on captive breeding, allowing for multiple matrilineal lines of a single ancestral species to be integrated into the gene pool of domestic horses. This contrasts with previous notions of domestication as a complex multigeneration process that begins with relatively few individuals selected for behavioral characteristics, such as docility and sedentary habits, as a prerequisite to coexistence with humans.

Domestic species such as dogs, cattle, sheep, and goats were established several thousand years before the horse was domesticated (4). The geographic spread of these species was a likely result of their expansion from a limited number of domestication centers. In the horse, the extensive and unparalleled retention of ancestral matrilineal lines suggests that widespread utilization occurred primarily through the transfer of technology for capturing, taming, and rearing wild caught animals (29, 30). In contrast, the export of domesticated horses from a geographically restricted center of origin would have resulted in a more limited diversity of matrilineal lines. Consequently, in the history of the domestic horse, transfer of technology rather than selective breeding may have been the critical innovation leading to their widespread utilization. Moreover, the high value of horses in primitive societies (1, 2) placed a premium on the rapid acquisition of this technology from neighboring communities.

References and Notes

1. D. W. Anthony, *Curr. Anthropol.* **27**, 291 (1986).
2. J. M. Diamond, *Nature* **350**, 275 (1991).
3. D. W. Anthony, in *Horses Through Time*, S. L. Olsen, Ed. (Roberts Rinehart for Carnegie Museum of Natural History, Boulder, CO, 1996), pp. 57–82.
4. J. Clutton-Brock, *A Natural History of Domesticated Mammals* (Cambridge Univ. Press, Cambridge, ed. 2, 1999).
5. D. Bennett, R. S. Hoffmann, *Mamm. Species* **628**, 1 (1999).
6. Maximum likelihood, neighbor-joining, and maximum parsimony phylogenetic trees were constructed with the program PAUP* (32). To select the model of DNA substitution that best fit the data for neighbor-joining analysis, we used a hierarchical likelihood ratio test approach implemented in the program MODELTEST 3.0 (33). A Tamura and Nei (34) model was selected, assuming a proportion of invariable sites of 0.52 and a gamma rate heterogeneity of 0.76. Gene and nucleotide diversity measures for each breed, as well as overall and interbreed pairwise differentiation tests, were conducted with the program Arlequin 1.1 (35).
7. A. Forstén, *Ann. Zool. Fenn.* **28**, 301 (1992).
8. M. George, O. A. Ryder, *Mol. Biol. Evol.* **3**, 535 (1986).
9. Blood samples were extracted from 191 horses from 10 breeds: Arabian (sample size, $n = 20$), Connemara (16), Exmoor (19), Fjord (16), Gotland Russ (20), Icelandic pony (20), Shetland (20), Standardbred (20), Swedish Ardennes (20), and Thoroughbred (20). Genomic DNA was extracted from whole blood samples after digestion with proteinase K and separation with organic solvents. Amplification of the mtDNA control region (from the tRNA-proline region to the central domain) by polymerase chain reaction (PCR) followed by analysis of single-strand conformation polymorphism (SSCP) was performed according to Marklund *et al.* (36). Direct sequencing of PCR products representing different SSCP variants was carried out with dye terminator chemistry on an ABI 377 instrument (Perkin-Elmer).
10. J. A. Leonard, R. K. Wayne, A. Cooper, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1651 (2000).
11. DNA was amplified successfully from eight permafrost-preserved bones from the American Museum of Natural History (FAM numbers are collection numbers): Pleist1, FAM 4695; Pleist2, FAM 4305; Pleist3, FAM 6735; Pleist4, FAM 6206; Pleist5, FAM 133514; Pleist6, FAM 6300; Pleist7, FAM 6171; and Pleist8, FAM 6282. All these samples were obtained from metapodials, originated from the Fairbanks area, Alaska, and were carbonated to the late Pleistocene (12,000 to 28,000 years B.P.). DNA also was extracted from eight horse bones from archaeological sites in southern Sweden (from the Historical Museum in Stockholm) and Estonia (provided by L. Luõgas) dating to 1000 to 2000 years ago: Anc1, SK1, tooth from sacrifice in Skedemosse (bog) in Öland, Sweden, 200 to 500 A.D.; Anc2, SK2, tooth from sacrifice in Skedemosse in Öland, Sweden, 200 to 500 A.D.; Anc3, VenA9, tooth, from Vendel grave 9 in Uppland, Sweden, Viking Age; Anc4, A17, tooth, from Alsike grave 9 in Uppland, Sweden, Viking Age; Anc5, Bi842, tooth, from Birka grave 842 in Uppland, Sweden, Viking Age; Anc6, HALLA, tooth, from Broa i Hall in Gotland, Sweden, Late Viking Age; Anc7, EstP61, tooth, from Pada, Estonia, Late Iron Age; and Anc8, EstP66, tooth, from Pada, Estonia, Late Iron Age. Extractions were done in laboratories exclusively devoted to the analysis of ancient DNA to avoid cross-contamination of samples. Permafrost samples were extracted in Los Angeles [see (10)] and archaeological samples in Stockholm. Three sets of primers amplifying fragments of about 160 bp were designed to amplify the same region as above: pair 1, 5'-GCCATCACTCCCAAGCT and 5'-ACATGCTTAT-TATTCATGG; pair 2, 5'-CCACCTGACATGCAATAT and 5'TGTGACTGGAAATGATTTG; and pair 3, 5'TCGT-GCATACCCCATCCAA and 5'CCTGAAGTAGGAACCA-GATG. No recent horse samples were amplified in laboratories that extracted the ancient samples. We sequenced DNA from multiple extractions and independent PCR amplifications to confirm the authenticity of DNA sequences. Three overlapping fragments were sequenced for each specimen, and overlapping portions were always identical. The poor preservation of DNA in ancient samples and the high copy number of mtDNA in the cell imply that nuclear inserts, if they exist, are unlikely to be amplified (37). None of the ancient or recent samples showed evidence of superimposed sequences, and replications always gave identical results. Sequences have been deposited in GenBank (accession numbers AF326635 to AF326686).
12. M. C. Winans, in *The Evolution of Perissodactyls*, D. R. Prothero, R. M. Schoch, Eds. (Oxford Univ. Press, New York, 1989), pp. 262–297.
13. K. Benirschke, N. Malouf, R. J. Low, H. Heck, *Science* **148**, 382 (1965).
14. E. A. Oakenfull, O. A. Ryder, *Anim. Genet.* **29**, 456 (1998).
15. C. Vilà *et al.*, *Science* **276**, 1687 (1997).
16. H. Mannen, S. Tsuji, R. T. Loftus, D. G. Bradley, *Genetics* **150**, 1169 (1998).
17. E. Giuffra *et al.*, *Genetics* **154**, 1785 (2000).
18. F. E. Zeuner, *A History of Domesticated Animals* (Harper & Row, New York, 1963).
19. M. Raymond, F. Rousset, *Evolution* **49**, 1280 (1995).
20. P. Arcander, P. W. Kat, B. T. Simonsen, H. R. Siegmund, in *Molecular Genetic Approaches in Conservation*, T. B. Smith, R. K. Wayne, Eds. (Oxford Univ. Press, New York, 1996), pp. 399–412.
21. U. E. Hagberg, *The Archeology of Skedemosse II* (KVAA 46, Almqvist & Wiksell Intl, Stockholm, 1967).
22. H. Braathen, *Ryttergraver: Politiska Strukturer i Eldre Rikssamlingstid* (VARIA 19, Universitetets Oldsaksamling, Oslo, 1989).
23. H. Ellegren, M. Johansson, K. Sandberg, L. Andersson, *Anim. Genet.* **23**, 133 (1992).
24. G. Guerin, M. Bertaud, Y. Amigues, *Anim. Genet.* **25**, 62 (1994).
25. S. Marklund, H. Ellegren, S. Eriksson, K. Sandberg, L. Andersson, *Anim. Genet.* **25**, 19 (1994).
26. L. Coogle, R. Reid, E. Bailey, *Anim. Genet.* **28**, 309 (1997).
27. Fifteen microsatellite loci were genotyped from the blood samples: HMS1, HMS3, HMS5, HMS6, HMS7, HTG6, HTG8, HTG10, HTG14, HTG15, LE034, LE037, LE038, LE039, and LE040 (23–26). PCR products were radiolabeled with ^{32}P and were separated in 6% polyacrylamide gels and visualized by autoradiography. Known size standards were included at regular intervals on the gels. PCR conditions and profiles are available upon request. All breeds were polymorphic for all loci except four that were monomorphic for locus LE039, for which only two alleles were observed (one with an overall frequency of 92%). Measures of genetic diversity and exact tests of genic differentiation were done with the program GENEPOP 3.1b (38). An assignment test that uses likelihood ratios to assign individuals to breeds according to their allelic composition was used to assess differentiation (28). We used the distance $d_{ij} = 1 - P_{ij}$, where P_{ij} is the proportion of shared alleles between individuals i and j , to construct a neighbor-joining tree of individuals.
28. D. Paetkau, W. Calbert, I. Stirling, C. Strobeck, *Mol. Ecol.* **4**, 347 (1995).
29. M. A. Levine, *J. Anthropol. Archaeol.* **18**, 29 (1999).
30. D. W. Anthony, D. R. Brown, *Antiquity* **65**, 22 (1991).
31. S. L. Olsen, in *Horses Through Time*, S. L. Olsen, Ed. (Roberts Rinehart for Carnegie Museum of Natural History, Boulder, CO, 1996), pp. 35–56.
32. D. L. Swofford, *PAUP* Phylogenetic Analysis Using Parsimony and Other Methods* (Sinauer, Sunderland, MA, 1998).
33. D. Posada, K. A. Crandall, *Bioinformatics* **14**, 817 (1998).
34. K. Tamura, M. Nei, *Mol. Biol. Evol.* **10**, 512 (1993).
35. S. Schneider, J.-M. Kueffer, D. Roessli, L. Excoffier, *Arlequin ver. 1.1: A Software for Population Genetic Data Analysis* (Genetics and Biometry Laboratory, University of Geneva, Switzerland, 1997).
36. S. Marklund, R. Chaudhary, L. Marklund, K. Sandberg, L. Andersson, *Anim. Genet.* **26**, 193 (1995).
37. R. K. Wayne, J. A. Leonard, A. Cooper, *Annu. Rev. Ecol. Syst.* **30**, 457 (1999).
38. M. Raymond, F. Rousset, *J. Hered.* **83**, 248 (1995).
39. Supported by the Swedish Research Council for Agriculture and Forestry, the Swedish Royal Academy of Sciences, the NSF, and the Bank of Sweden Tercentenary Foundation. We thank the American Museum of Natural History in New York and the Historical Museum in Stockholm for permission to sample their collections; L. Luõgas for providing archaeological horse samples from Estonia; H. Persson for assisting in the microsatellite typing; and J. Diamond, B. Van Valkenburgh, M. Kohn, and K.-P. Koepfli for commenting on the manuscript.

12 October 2000; accepted 13 December 2000