Revising the recent evolutionary history of equids using ancient DNA

Ludovic Orlando1,a, Jessica L. Metcalf1,b, Maria T. Alberdi1,c, Miguel Telles-Antunes1,d, Dominique Bonjean1,e, Marcel Otte1,f, Fabiana Martin1,g, Véra Eisenmann1,h, Marjan Mashkour1,i, Flavia Morello1,j, Jose L. Prado1,k, Rodolfo Salas-Gismondi1,l, Bruce J. Shockey1,m,n, Patrick J. Wrinn1,o, Sergei K. Vasil’ev1,p, Nikolai D. Ovodov1,q, Michael I. Cherry1,q, Blair Hopwood1,q, Dean Male1,q, Jeremy J. Austin1,q, Catherine Hänni1,q, and Alan Cooper1,q,e

1Institut de Géonomique Fonctionnelle de Lyon, Université de Lyon 1, Ecole Normale Supérieure de Lyon, and Institut National de la Recherche Agronomique, Centre National de la Recherche Scientifique, 69364 Lyon Cédex 07, France; 2Australian Centre for Ancient DNA, School of Earth and Environmental Sciences, University of Adelaide, SA 5005, Australia; 3Departamento de Paleobiología, Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas, José Gutiérrez Abascal 2, 28006 Madrid, Spain; 4Academia das Ciências de Lisboa, 1249-122 Lisboa, Portugal; 5Centro de Recherches de la Grotte Sldaina and Arqueología Andenneise, 5300 Sclaym, Belgium; 6Service de Préhistoire, Université de Liège, 4000 Liège, Belgium; 7Arqueológico Fundación and Centro de Estudios del Cuaternario de Fuego-Patagonia y Antártica, Casilla 737, Punta Arenas, Chile; 8Unité Mixte de Recherche S143, Centre National de la Recherche Scientifique, and Paléobiodiversité et Paléoenvironnements, Muséum National d’Histoire Naturelle, CP 38, 75005 Paris, France; 9Unité Mixte de Recherche S197, Centre National de la Recherche Scientifique, and Département d’Ecologie et Gestion de la Biodiversité, Bâtiment d’Anatomie Comparée, Muséum National d’Histoire Naturelle, F-75005 Paris, France; 10Universidad de Magallanes and Centro de Estudios del Cuaternario de Fuego-Patagonia y Antártica, Punta Arenas, Chile; 11Núcleo Consolidado sobre Investigaciones Arqueológicas y Paleontológicas del Cuaternario Pampeano, and Departamento de Arqueología, Universidad Nacional del Centro, Del Valle 5737, B7400JWI Olavarría, Argentina; 12Departamento de Paleontología de Vertebrados, Museo de Historia Natural, and Universidad Nacional Mayor de San Marcos, Lima 14, Peru; 13Manhattan College, New York, NY 10471; 14Division of Paleontology, American Museum of Natural History, New York, NY 10021; 15Department of Anthropology, University of Arizona, Tucson, AZ 85721-0030; 16Institute of Archaeology and Ethnography, Siberian Branch, Russian Academy of Sciences, Novosibirsk 630090, Russian Federation; 17Laboratory of Archaeology andPaleoecology of Central Siberia, Institute of Archaeology and Ethnography, Siberian Branch, Russian Academy of Sciences, Academgorodok, Krasnoyarsk 660036, Russian Federation; 18Department of Botany and Zoology, University of Stellenbosch, Matieland 7602, South Africa

Edited by Raymond L. Bernor, National Science Foundation, Arlington, VA, and accepted by the Editorial Board October 20, 2009 (received for review April 15, 2009)

The rich fossil record of the family Equidae (Mammalia: Perissodactyla) over the past 55 MY has made it an icon for the patterns and processes of macroevolution. Despite this, many aspects of equid phylogenetic relationships and taxonomy remain unresolved. Recent genetic analyses of extant equids have revealed unexpected evolutionary patterns and a need for major revisions at the generic, subgeneric, and species levels. To investigate this issue we examine 35 ancient equid specimens from four geographic regions (South America, Europe, Southwest Asia, and South Africa), of which 22 delivered 87–688 bp of reproducible aDNA mitochondrial sequence. Phylogenetic analyses support a major revision of the recent evolutionary history of equids and reveal two new species, a South American hippidion and a descendant of a basal lineage potentially related to Middle Pleistocene equids. Sequences from specimens assigned to the giant extinct Cape zebra, Equus capensis, formed a separate clade within the modern plain zebra species, a phenotypically plastic group that also included the extinct quagga. In addition, we recognized extinction times for two equine-related equid groups. However, it is apparent that the current dataset cannot solve all of the taxonomic and phylogenetic questions relevant to the evolution of Equus. In light of these findings, we propose a rapid DNA barcoding approach to evaluate the taxonomic status of the many Late Pleistocene fossil Equidae species that have been described from purely morphological analyses.

DNA taxonomy | equid evolution | macroevolution | phylogeny | ancient DNA

The original sequence of horse fossils found in the 1870s by paleontologist Othniel Charles Marsh, and popularized by Thomas Huxley (1), has been enriched by a large fossil record over the years and has now become one of the most well-known examples of macroevolutionary change (2). The original linear model of gradual modification of fox-sized animals (Hyracothere horses) to the modern forms has been replaced by a more complex tree, showing periods of explosive diversification and branch extinctions over 55 MY (3). The end of the Early Miocene (15–20 MYA) marks a particularly important transition, separating an initial phase of small leafy browsers from a second phase of more diverse animals, exhibiting tremendous body-size plasticity and modifications in tooth morphology (4). This explosive diversification has been accompanied by several stages of geographic extension from North America to the rest of the New and Old Worlds, so that by the end of the Miocene (5 MYA) more than a dozen distinct genera are represented in the fossil record (4) (Astrohippus, Calippus, Cormohipparion, Dinohippus, Hippidion, Nanohippus, Neohipparion, Onohippidium, Pleissippus, Plohippus, Protohippus, Pseudhipparion, Sinohippus, Old World Hippipparion). Nearly all of this diversity is now extinct, with all living members of the Equidae assigned to a single genus Equus (4–5), which includes caballines (true horses, E. caballus and E. przewalskii) and non-caballines: hemionids (E. hemionus and E. kiang for Asian and Tibetan wild asses, respectively), African wild asses (E. africanus, i.e., wild ancestors of the domestic donkeys E. asinus), and zebras (E. quagga, E. hartmannae/E. zebra, and E. grevyi for plains, mountain, and Grevy’s zebras, respectively) (6–9). Plains zebras (E. burchelli) have also been considered to be a species distinct from the extinct quagga (E. quagga), both belonging to the plains zebra group (10–11).

Importantly, many of the past equid lineages became extinct as recently as the Late Quaternary (Late Pleistocene and Holocene)


The authors declare no conflict of interest.

This article is a PNAS Direct Submission. R.L.B. is a guest editor invited by the Editorial Board.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. GQ324584–GQ324612 and GU062887).

1To whom correspondence may be addressed at: Paleogenetix and Molecular Evolution, Institut de Génomique Fonctionnelle de Lyon, Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364 Lyon cédex 07, France. E-mail: ludovic.orlando@ens-lyon.fr.

2To whom correspondence may be addressed at: Australian Centre for Ancient DNA, Darling Building, University of Adelaide, North Terrace Campus, South Australia 5005, Australia. E-mail: alan.cooper@adelaide.edu.au.

This article contains supporting information online at www.pnas.org/cgi/content/full/0903672106DCSupplemental.
raising the potential for ancient DNA (aDNA) analyses (12). As a consequence, several aspects of the classical paleontological model of equid evolution can now be tested with molecular data. Indeed, the field of aDNA started with the publication of mitochondrial DNA sequences from a museum specimen of the quagga (13), an equid related to Plains zebras that became extinct in the wild in the late 1870s. Over recent decades, aDNA sequences have been reported from other extinct members of the Equidae family (11, 14–18) and have revealed unexpected patterns in equid evolution with varying degrees of conflict to morphological and paleontological phylogenies. For instance, the morphologically distinct South American hippidions were formerly considered to be descendents of the Pliohippines, a basal North American group that diverged from the lineage that gave rise to the genus Equus before 10 MYA (19). Hippidions exhibited a similar distinctive nasal morphology to Pliohippines, and appeared in the South American fossil record ~2.5 MYA (20). However, mitochondrial sequences of hippidions from Patagonia and the Buenos Aires Province form a tight clade within a larger paraphyletic group of Equus, suggesting either that hippidions and living equids belong to the same genus or that living equids should be split into several genera (15, 16, 18, 21). Similarly, the South American subgenus (Amerhippus), which was contemporaneous with hippidion but quite different taxonomically, was shown to diverge from the lineage that gave rise to the genus Equus after 20 MYA (22). In this paper, we report 29 mitochondrial DNA (mtDNA) sequences (22 control region, HVR-1; and 7 cyt b) from 22 fossils of extinct equids from four different geographic regions (South America, Europe, Southwest Asia, and southern Africa). The evolutionary history and phylogeographic structure of hippidiforms was investigated using new specimens from Patagonia and Peru. In Eurasia, the taxonomic relationship of hemionids and E. hydruntinus was further analyzed using Siberian hemionid fossils and additional specimens of E. hydruntinus, which also permitted the extinction date to be dated. Last, we address the phylogenetic relationship of the Late Pleistocene giant Cape zebra (Equus capensis) (23–25) from South Africa to the other zebra species. Phylogeographic analyses and comparison with the genetic diversity found within and between extant equid groups reveal that though two new extinct species (a hippidion and a sussemione) are supported, the temporal and geographic variation in body size of others may have been underestimated, as suggested previously for cabal-lids (16), leading to taxonomic oversplitting.

Results and Discussion

Authenticity of aDNA Sequences. We successfully recovered aDNA sequences from 22 of the 35 specimens (Fig. S1 and Table S1A): one horse (E. caballus) from the Schladina cave, Belgium; 10 Hippidion specimens (five each from Patagonia and Peru); six E. (cf.) hydruntinus (two from Europe, three from Russia, and one from the Middle East); one E. hemionus (from the Middle East); and four E. capensis (from South Africa). Thirteen fossils did not yield any aDNA information despite extensive extraction and PCR attempts (Table S1B), and by default, also acted as cross-contamination controls in different extraction experiments. The poor DNA yields from these fossils are in agreement with their preservation state [high thermal age (26); see SI Text]. Thirteen of the samples were successfully amplified and sequenced in two independent laboratories [Lyon and Australian Centre for Ancient DNA (ACAD)], using different extraction methods, PCR conditions, and sequencing strategies (cloning/sequencing of different amplicons in Lyon versus direct sequencing at ACAD). It is notable that of the 6,569 bp duplicated at ACAD (from 64 amplicons), comparison with the final consensus sequences revealed only minor discrepancies (average number of differences per base per fragment = 0.88%). As expected, a large majority of the variants (74.6%) consisted of GC→AT substitutions and are likely to derive from postmortem cytosine deamination characteristic of ancient templates (27, 28). The base recovered from the majority of independent amplicons was accepted as authentic, in agreement with standard recommendations (27, 29), and importantly, the extent of damage-induced errors was well below the intraspecific levels of genetic diversity among equids (Table S3). It is noteworthy that no modern equid DNA was stored in the aDNA laboratories, and though modern Grey's zebra hair were analyzed in the Lyon post-PCR laboratory, this haplotype was not found in any ancient sample. Of the new sequences, the mtDNA HVR-I sequences of the Patagonian Hippidion saldiasi specimens were highly similar to previously reported Hippidion sequences (15–17) (Fig. 1). The mtDNA HVR-I and/or cyt b sequences of the E. (cf.) hydruntinus from Schladina (Belgium) and Zagheh (Iran) showed high similarity with the two E. hydruntinus haplotypes reported in ref. 17 (Fig. 1 and Fig. S2 A and B). Similarly, although the strongest support was obtained for the placement of hippidiforms within a larger paraphyletic group of Equus (with merged datasets, Table S2, and with marginal support from SH tests: P values = 0.065–0.083; Table S5, topologies R–V), likelihood-based topological tests cannot significantly reject alternative topologies and the exact position of the root, previously reported between nodes A and H (Fig. 1) (15, 16) should be considered preliminary. The lack of resolution is complicated by the short divergence time among caballines and New World horses (circa 0.5 MYA; nodes A and B/B1/B2; Table S4) and the lack of a close outgroup, as has been noted with mammoths (30, 31). When the rhino was used as an outgroup, the data were RY coded to reduce possible mutation saturation artifacts resulting from this the deep divergence (55 MYA), but this removed support for most nodes (Table S2).

Phylogenetic Analyses. South America. Both the maximum likelihood and Bayesian phylogenetic analyses (Fig. 1) of the mtDNA HVR sequences identified three monophyletic New World groups: caballine horses (node G); NWSL horses (node F); and hippidiforms (node C2), in agreement with earlier studies (15, 16, 18). The exact branching order of the three groups (relative to the Old World noncaballine horses) could not be resolved with either the HVR-1 dataset alone (node B shows weak supports in Fig. 1; Table S5, topologies L–Q) or various combinations of HVR-1 plus cyt b/HVR-2 data (Table S5, topologies R–V). Similarly, although the strongest support was obtained for the placement of hippidiforms within a larger paraphyletic group of Equus (with merged datasets, Table S2, and with marginal support from SH tests: P values = 0.065–0.083; Table S5, topologies R–V), likelihood-based topological tests cannot significantly reject alternative topologies and the exact position of the root, previously reported between nodes A and H (Fig. 1) (15, 16) should be considered preliminary. The lack of resolution is complicated by the short divergence time among caballines and New World horses (circa 0.5 MYA; nodes A and B/B1/B2; Table S4) and the lack of a close outgroup, as has been noted with mammoths (30, 31). When the rhino was used as an outgroup, the data were RY coded to reduce possible mutation saturation artifacts resulting from this the deep divergence (55 MYA), but this removed support for most nodes (Table S2).

According to our molecular dating estimates, the different equid lineages (hippidiforms, NWSL, caballines, and noncaballines) originated 3.7–4.3 MYA (95% confidence range: 2.8–6.2 MYA; Table S4). This directly contrasts with classical palaeontological models of hippidiform origins as descendants of the Pliohippines (divergence time with the Equus lineage >10 MYA) (19) or as a lineage diverging from a (Dinohippus, Astrohippus, and Equus) clad ~7–8 MYA (32), and considerably reduces the time gap between the supposed divergence of the hippidiform lineages and their first appearance in the fossil record 2.5 MYA (20). A deep split within hippidiforms separates one cluster (Fig. 1, node D) from Patagonia and Argentina (Hippidion saldiasi and Hippidion principale, respectively) from a second (Fig. 1, node E) consisting of specimens from a high-altitude cave in Peru. The genetic differentiation between these clusters is greater than intraspecific levels of genetic diversity in any current equid species (Fig. 2 and Table S3) but comparable to that between indisputable species from the same genus (Equus),
Fig. 1. Phylogenetic relationships among equids. Unrooted phylogenetic tree based on maximum likelihood analysis of the HVR-1 dataset. Sequences generated in this study are highlighted in gray. Numbers above/below branches correspond to node bootstrap values, approximate likelihood ratio test (aLRT SH-like), and Bayesian posterior probabilities. \( \Delta 28 \): 28-bp deletion in the mitochondrial HVR-1 (positions 15533–15560 according to the complete horse mitochondrial genome: GenBank accession no. X79547) (Fig. 2A). The potential locations of the root are reported with gray (alternative)/black (most likely) triangles according to likelihood-based topological tests (Fig. S4 and Table S5) and previous reports (15, 16). C1 and C2 nodes have been used as calibration points for molecular dating. Two radiocarbon dates retrieved (Table S1), with the Wonderwerk specimens ACAD223 and ACAD230 exhibiting haplotypes outside the previously known variation (node J2), but inside that of the northern subspecies (E. q. boehmi, node K, with the exception of specimen ACAD226 and ACAD230 exhibiting haplotypes outside the pre-
formed part of the same diverse taxonomic group as plain zebras and quaggas (E. q. quagga), and that this group possessed a marked plasticity in both morphological size (reminiscent of Hippidion) and coat color (8).

Eurasia. We analyzed two Middle Eastern hemiones from Iran (CH28, CH30) and six Equus (cf.) hydruntinus samples (one Middle Eastern, TZ9; two European, CH561 and CH1069; and three from Khakassia, Russia, southwest Siberia) to clarify the taxonomic relationships and geographic range of these groups. Surprisingly, the Siberian specimens did not cluster with E. hydruntinus nor hemiones but formed a new monophyletic group with strong support (Fig. 1, node I) and no specific affinity with other non-caballine horses (Table S5, topologies J–K4). Cyt b sequences further support this arrangement (Fig. S2 B and C). Interestingly, this new clade does not possess a 28-bp deletion in the first HVR-1 PCR fragment that unite E. hydruntinus specimens from Belgium (CH561) and Crimea (17), and the other Asian and Middle Eastern hemiones (Figs. 1 [CH28] and 2A) relative to other equids. The genetic distances separating the three Russian specimens from hemiones are of a similar magnitude to other species (e.g., those separating donkey and hemion) and together this evidence suggests that clade I should be considered a distinct taxonomic group, representing a new species, distinct from E. hydruntinus, with no extant relative. Isolated bones that are morphologically similar to the three Khakassian specimens have been recovered from cave sites in the Altai Mountains (Denisova, Kaminnaya, Logovo Gieny, Okladnikov, and Strachnaya) as well as from alluvial sites in the Altai piedmont plain (39). In contrast with our molecular findings, previous morphological analyses suggested similar metapodial proportions to classical E. hydruntinus from Western Europe, the Crimea, the Middle East, and Caucasus, although the former was larger and more robust. Our analysis of tooth morphological data (Fig. S4) shows more consistency with the molecular phylogeny in that there are poor affinities with hemiones and E. hydruntinus, and instead there are strong similarities with the archaic Middle Pleistocene Sussemiones from Süssenborn, Germany (40). The age to the most recent common ancestor of the three samples analyzed here provided a conservative estimate of 1.1–1.3 MYA (95% confidence range: 0.4–2.3 MYA; Table S4) for the emergence of this new species. Sussemiones possibly differentiated much earlier (circa 2 MYA) in Beringia, as suggested by the lower cheek series found at Lost Chicken, Alaska, and were extremely successful and dispersed across all Eurasia and Africa (41). Interestingly, Sussemiones were supposedly extinct well before the Late Pleistocene (40), but radiocarbon dating of two Khakassian specimens revealed that the Siberian species survived at least until circa 45–50 KYBP (SI Text).

To further examine the extinction date of E. hydruntinus, we examined a specimen from 17th-century Portugal (CH1069), purported to have been the last representative of this species (42). However, HVR-1 (median spanning network; Fig. S2 B and C), cyt b (Fig. S2 D and E), and cranial morphological data (Fig. S4G) indicate that this specimen showed a strong relationship to donkeys. This implies that E. hydruntinus, which first appeared in the fossil record 350 KYA in Lunel-Viel (43) (southern France), did not survive until the Middle Ages but most probably became extinct in the Iron Age (44). Interestingly, the genetic distance between E. hydruntinus and E. hemionus is within the diversity observed within hemiones, in agreement with previous studies of short (144–288 bp) HVR-1 sequences (17). Although the combination of the relative average dimensions of the metapodials and upper-cheek teeth parameters of E. hydruntinus differed from any other hemione (45), the genetic data suggests that E. hydruntinus was a subspecies of E. hemionus. However, it is worth noting that the kiang and onager/kulan are currently classified as separate species based on differences in coat color, morphology, geographic distribution, and the number of chromosomes, and yet show poor mitochondrial differentiation (7, 8, 46). Therefore, the data are still compatible with E. hydruntinus being a separate species, which perhaps experienced multiple introgression events from different stocks (from onagers in Zagheh, Iran, sample TZ9; and from kulan in Sagzabad, Iran, sequence DQ464012 reported in ref. 17). Consequently, we suggest that the taxonomic status of Equus hydruntinus as a separate species or a subspecies of E. hemionus should be clarified with genetic information from nuclear loci.

Genetic Taxonomy of Late Pleistocene and Holocene Equids. The short (<89 bp) first fragment of HVR-1 appears to be a potentially useful barcode for the higher-level taxonomic identification of fossil equids (Fig. S2A). This fragment can be amplified in caballine horses, zebras, asses, and hemiones using our PCR conditions, and, importantly, contains the informative 28-bp deletion characteristic of the (E. hydruntinus/E. hemionus/E. kiang) lineage. It shows higher genetic distances between, rather than within, extant taxonomic groups (Table S3C) and could potentially be used to assess the taxonomic status of the many paleontological equid species. For instance, it could be used to test the suggestion that the many species of Late Pleistocene New World hemione-like equids (e.g., E. francisci, E. tau, E. quinni, E. cf. hemionus, E. cf. kiang) are conspecific within a single species of NWSL horse (16). This rapid assessment method could be a useful tool to disentangle the temporal and regional body-size variation and marked anatomical
convergence observed within Late Pleistocene and Holocene equids. Of course, an mtDNA HVR genetic survey would be only a first step in identifying major taxonomic groups, which would need to be examined with morphological and other molecular studies. Indeed, the relatively short sequences we have generated cannot clearly resolve the exact branching order of major taxonomic groups and the bootstrap values, posterior probabilities, and approximate likelihood ratio tests (aLRT, SH-like) for these nodes were generally not strong. Recently, the use of whole mitochondrial genome sequences from Pleistocene fossils has helped resolve the rapid radiation events within elephants (31) and ursids (47). Such a mitogenomic approach is likely to uncover the phylogenetic relationships within caballines and New World horses (Fig. 1, nodes A–G), and potentially within extant noncaballine horses as well, which are still poorly understood at the molecular level (Fig. 1, node H). For example, in Fig. 1, Greyeyi’s zebras nest within paraphyletic hemonids, and mountain zebras do not cluster with plains zebras but with donkeys, in striking contrast with traditionally accepted taxonomy based on coat patterns, behavior, morphology, geographic separation (6, 7, 10, 48), and the 28-bp deletion (Fig. 1) (17). However, current datasets provide only moderate (if any) bootstrap and aLRT supports and posterior probabilities (Table S2) for these branches, and they should be regarded with caution. Besides, likelihood-based topological tests show significant support for the standard topology (in accordance with the current taxonomy) (KH and SH tests P values ≤0.003 for model E7, and KH test P value = 0.008 for model E7 vs. E6; Table S5). It seems likely that a (mito)genomic approach will be needed to solve major phylogenetic splits within equids and rapid radiation events (Table S4).

In this study, two species of extinct equids have been identified by mtDNA. The first is related to hippidiforms and corresponds to the paleontological genus Onohippidium, which should be reclassified within the Hippidion genus as Hippidion devillei. The second is a unique basal lineage of Old World equid, which appears morphologically related to the Middle Pleistocene Sussemiones, yet survived in southwestern Siberia at least until circa 45–50 KYBP. In addition, we propose to synonymize a variety of other species [Cape zebras, quaggas, and plains zebras (11, 13) and potentially E. hydruntinus and E. hemionus]. This reinforces similar suggestions for hippidiforms, NWSL, and caballids (15, 16, 18) and E. hydruntinus (17). This pattern of taxonomic oversplitting does not appear to be restricted to equids but is widespread amongst other Quaternary megafauna [e.g., Late Pleistocene bison (49); Holarctic cave lions (50); New World bison (51), and ratite moas (52, 53)]. Together, these findings suggest that the morphological plasticity of large terrestrial vertebrates across space and time has generally been underestimated, opening the way to detailed studies of the environmental, ecological, and epigenetic factors involved. Interestingly, in this regard the human lineage shows a rich fossil record over the last 6 MY, spreading over seven possible genera and 22 species (54). The exact number of taxonomic groups that should be recognized is still debated, even within our own genus (55), and in this context it is pertinent to consider the degree of taxonomic oversplitting, from species to generic levels, that aDNA has revealed amongst Late Pleistocene equids and other megafauna. A further important implication of this finding is that the number of megafaunal extinctions and loss of taxonomic diversity from the Pleistocene to modern day may not have been nearly as large as previously thought, at least at the species or subspecies level. Conversely, at the molecular level, aDNA studies on a wide range of large mammal taxa (49, 50, 56, 57) have revealed that the loss of genetic diversity over this time period has been much larger than previously recognized with major implications for the conservation biology of surviving populations (58).

Materials and Methods

**aDNA Extraction, Amplification, and Sequencing.** The samples were extracted, amplified, and sequenced in specialist aDNA laboratories in Lyon (n = 27) and Oxford (n = 24) (Table S1). Primers (Table S3) were designed in accordance to appropriate tech of PCR inhibitors. A 183-bp fragment of the cyt b gene and six short overlapping fragments encompassing 546 bp of the horse mtDNA HVR-I (positions 15158–16063) from the complete horse mitochondrial genome; GenBank accession no. X79547 were targeted by PCR, as described in Orlando et al. (15, 18). All PCR products were cloned and sequenced, and a total of 330 PCR products and 2,230 clones were analyzed at Lyon (Table S1A). In addition, 121 PCR amplicons (16 samples) were independently analyzed by direct sequencing at ACAAD (two additional amplicons were cloned and sequenced for ACAD227; Table S1A). For most samples, each sequence position was determined from at least two independent amplifications to avoid possible artefactual substitutions induced by DNA damage (27). However, for four samples (reported in italics in Table S1A), it was only possible to amplify a certain fragment once. No individual site with a significant posterior probability of DNA damage could be detected in four independent BAYPASS-degr analyses (62).

**Sequence Analyses.** The sequences reported here have been deposited in GenBank under accession nos. GQ324584–GQ324612 and GU062887. Available sequences from equid species were extracted from GenBank using the first 10,000 BLAST hits recovered from Equus (Amerhippus) neogene seques (EU036080 and EU036081). Both gene fragments were aligned manually using the SeaView software (63). A final HVR-1 dataset was constructed to maximize the length of included sequences and to balance taxonomic sampling (S1 Text). Unrooted and rooted phylogenies (GenBank accession nos. X97336 and Y07726 for Rhinoceros unicornis and Ceratotherium simum, respectively) were constructed using maximum likelihood and Bayesian Markov chain Monte Carlo methods (S1 Text). A Ry-recorded rooted dataset was used for further phylogenetic analyses to limit the possible adaptation of saturation due to the deep divergence of rhinos and equids. —55 MYA (Table S2). Another dataset consisting of all of the available equid sequences for the targeted 143-bp cyt b fragment was also generated and analyzed (Fig. S2D). Finally, HVR-1, HVR-2, and cyt b datasets were merged (see “merged consensus” described in Orlando et al. (17)). All datasets were used for phylogenetic inference (Fig. S2D–F, Table S2, and S1 Text), and the monophyly of the principal nodes was assessed using likelihood-based statistical tests implemented in PAUP*4.0 (Kishino-Hasegawa (KH) and Shimodaira-Hasegawa (SH) tests; Fig. S4 and Table S5) (64).

To more accurately estimate the distributions of pairwise genetic distances within and between different taxonomic groups, two other HVR-1 datasets were generated of intermediate sequence length (243 and 301 bp, respectively, according to the reference sequence X79547), but with larger numbers of haplotypes (1,866 and 1,544, respectively). Uncorrected and corrected (GTR+I+0) pairwise distance matrices were averaged from PAUP* (64), and genetic diversity estimates were calculated and plotted using the R package (65). All datasets were automatically filtered for numts, and levels of missing data greater than 5% using a Perl script. Median-joining network were reconstructed for both datasets using Network (66), weighting all sites equally (Fig. S2D). Bayesian analyses were applied to the unrooted HVR-1 dataset to assess divergence times, using BEAST (67) and assuming a HKY + I + I model of molecular evolution. The age of specimens was recovered from either radiocarbon dating or the literature (S1 Text; dating; Table S1A). A calibration point was established by assuming the emergence of the hippidiform lineage occurred within a normal distribution centered at 2.5 ± 0.1 MYA (Fig. 1, node C1), based on fossil evidence (20). A second calibration point (Fig. 1, node C2) was placed at 0.7 ± 0.1 MYA for the emergence of plains zebras, documented by the fossil E. mauritanus, which has been unambiguously placed on the plain zebra lineage (10, 68, 69).

**ACKNOWLEDGMENTS.** We thank Marilyn Duffraisse and Marta Kasper for technical help; Benjamin Gitler for his help in the Palgene platform (Centre National de la Recherche Scientifique/École Normale Superieure de Lyon); Francis Thackeray (Transvaal Museum, Pretoria, South Africa) for confirming the morphological identity of the E. capensis specimen; Bastien Llamas for giving helpful feedback on the map; Jane Wheeler (Instituto de Investigación y Desarrollo de Camelidos Sudamericanos (CONOPEA)) for assistance with the Peruvian material; Oxford Radiocarbon Accelerator Unit, CU INSTAR, and UCl for radiocarbon dating; the National Geographic Society (Research and Exploration Grant 8010–06) for support.
for the work in Peru, and the following organizations for their support of this work: Centre National de la Recherche Scientifique, Australian Research Council, Australian DIISR, France MESR and MAAE Partenariat Hubert Curien (Fast EGEIDE), and Ecole Normale Supérieure de Lyon. We also thank five anonymous reviewers, the AG Zoologischer Garten from Köln who kindly provided us with the extant Equus greyii hairs used in this study, and Museum National d'Histoire Naturelle, Musee de Historia Natural (permit no. RDN 1740/INC), Museum of the Politechnical University of Ecuador (Solymar López and José Luis Román Carrion), and Dr. A. Derviánkov for morphological information about Portuguese and Russian equid fossils.


