

Revising the recent evolutionary history of equids using ancient DNA

Ludovic Orlando^{a,1}, Jessica L. Metcalf^b, Maria T. Alberdi^c, Miguel Telles-Antunes^d, Dominique Bonjean^e, Marcel Otte^f, Fabiana Martin^g, Véra Eisenmann^h, Marjan Mashkourⁱ, Flavia Morello^j, Jose L. Prado^k, Rodolfo Salas-Gismondil^l, Bruce J. Shockey^{m,n}, Patrick J. Wrinn^o, Sergei K. Vasil'ev^p, Nikolai D. Ovodov^q, Michael I. Cherry^r, Blair Hopwood^s, Dean Male^b, Jeremy J. Austin^b, Catherine Hänni^a, and Alan Cooper^{b,2}

^aInstitut de Génomique 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; ^bAustralian Centre for Ancient DNA, School of Earth and Environmental Sciences, University of Adelaide, Adelaide, SA 5005, Australia; ^cDepartamento de Paleobiología, Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas, José Gutiérrez Abascal 2, 28006 Madrid, Spain; ^dAcademia das Ciências de Lisboa, 1249-122 Lisboa, Portugal; ^eCentre de Recherches de la Grotte Scladina and Archéologie Andennaise, 5300 Sclayn, Belgium; ^fService de Préhistoire, Université de Liège, 4000 Liège, Belgium; ^gArqueóloga Fundación and Centro de Estudios del Cuaternario de Fuego-Patagonia y Antártica, Casilla 737, Punta Arenas, Chile; ^hUnité Mixte de Recherche 5143, Centre National de la Recherche Scientifique, and Paléobiodiversité et Paléoenvironnements, Muséum National d'Histoire Naturelle, CP 38, 75005 Paris, France; ⁱUnité Mixte de Recherche 5197, 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; ^jUniversidad de Magallanes and Centro de Estudios del Cuaternario de Fuego-Patagonia y Antártica, Punta Arenas, Chile; ^kNúcleo Consolidado sobre Investigaciones Arqueológicas y Paleontológicas del Cuaternario Pampeano, and Departamento de Arqueología, Universidad Nacional del Centro, Del Valle 5737, B7400JWI Olavarría, Argentina; ^lDepartamento de Paleontología de Vertebrados, Museo de Historia Natural, and Universidad Nacional Mayor de San Marcos, Lima 14, Peru; ^mManhattan College, New York, NY 10471; ⁿDivision of Paleontology, American Museum of Natural History, New York, NY 10021; ^oDepartment of Anthropology, University of Arizona, Tucson, AZ 85721-0030; ^pInstitute of Archaeology and Ethnography, Siberian Branch, Russian Academy of Sciences, Novosibirsk 630090, Russian Federation; ^qLaboratory of Archaeology and Paleogeography of Central Siberia, Institute of Archaeology and Ethnography, Siberian Branch, Russian Academy of Sciences, Academgorodok, Krasnoyarsk 660036, Russian Federation; and ^rDepartment 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 extinct 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 revise the currently recognized extinction times for two hemione-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 Othaniel 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 widely known examples of macroevolutionary change (2). The original linear model of gradual modification of fox-sized animals (Hyracotheres 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*, *Nannipus*, *Neohipparion*, *Onohippidium*, *Plesippus*, *Pliohippus*, *Protohippus*, *Pseudhipparion*, *Sinohippus*, Old World *Hipparion*). 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 noncaballines: 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 quaggas (*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),

Author contributions: L.O., C.H., and A.C. designed research; L.O., J.L.M., M.T.A., M.T.-A., D.B., M.O., F. Martin, V.E., M.M., F. Morello, J.L.P., R.S.-G., B.J.S., P.J.W., S.K.V., N.D.O., M.I.C., B.H., D.M., and J.J.A. performed research; C.H., A.C., M.T.A., M.T.-A., D.B., M.O., F. Martin, V.E., M.M., F. Morello, J.L.P., R.S.-G., B.J.S., P.J.W., S.K.V., N.D.O., and M.I.C. contributed new reagents/analytic tools; L.O. and V.E. analyzed data; and L.O., J.L.M., and A.C. wrote the paper.

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: Paleogenetics 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/0903672106/DCSupplemental.

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 descendants 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 \approx 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 *Equus*-like (22), was shown by aDNA sequences to actually be conspecific with caballines and did not form a separate subgenus within living *Equus* (18). In North America, aDNA revealed that a group of stilt-legged equid species morphologically similar to the hemiones (Asian asses) were in fact New World endemics, and that many putative species of this New World stilt-legged horse (NWSL) group should potentially be synonymized (16). Last, within Europe, aDNA has shown that the extinct European ass (*Equus hydruntinus*) was not related to zebras or primitive Pliocene equids as previously suggested, but was sufficiently closely related to the extant hemiones (Asian wild ass) that separate taxonomic status was questionable (17).

In this paper, we report 29 mitochondrial DNA (mtDNA) sequences (22 control region, HVR-1; and 7 cytochrome *b*, *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 examined. 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. Phylogenetic 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 caballids (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 Scladina 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 (67 differences; average number of differences per base per fragment = 0.884%). As expected, a large majority of the variants (74.6%) consisted of GC \rightarrow 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 Grevy's zebra hairs 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 Scladina (Belgium) and Zagheh (Iran) showed high similarity with the two *E. hydruntinus* haplotypes reported in ref. 17 (Fig. 1 and Fig. S2 B and C), as did the new Iranian *E. hemionus* sample analyzed here (CH28). These are consistent with the ancient sequences being authentic.

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 MY; 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 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*) clade \approx 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*),

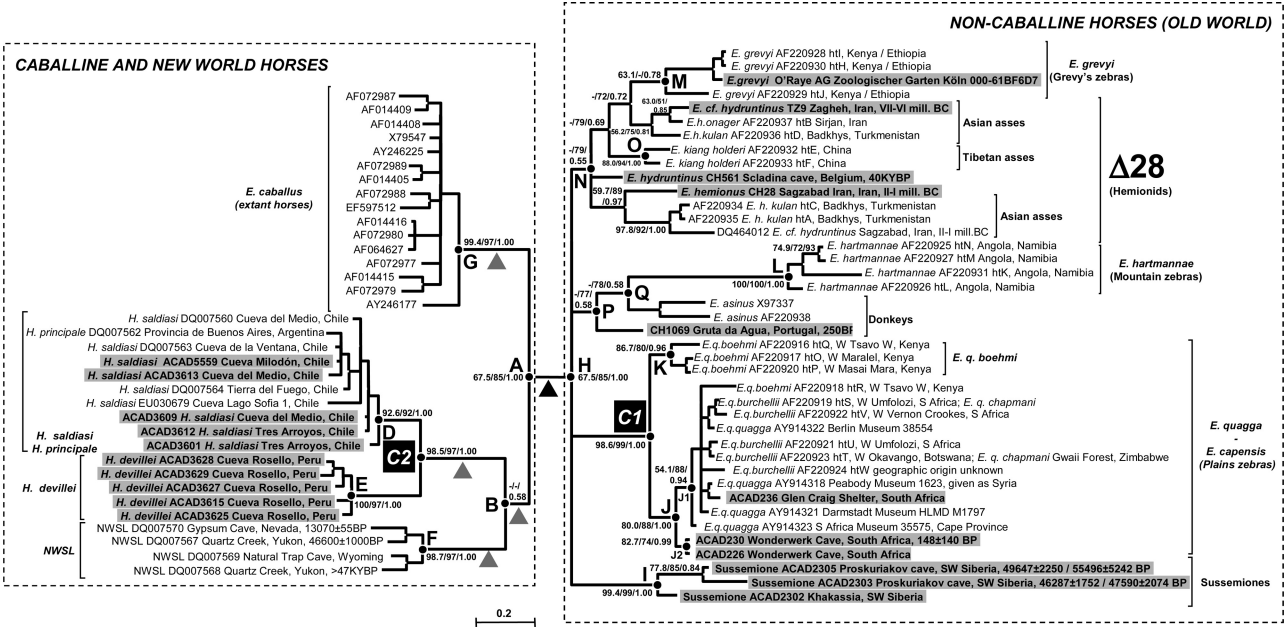


Fig. 1. Phylogenetic relationships among equids. Unrooted phylogenetic tree based on maximum likelihood analysis of the HVR-1 dataset. Sequences generated in this study are highlighted in gray. Numbers above/below branches correspond to node bootstrap values, approximate likelihood ratio test (aLRT SH-like), 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 were retrieved at the University of Colorado Laboratory for AMS Radiocarbon Preparation and Research (CURL) and the University of California-Irvine Accelerator Mass Spectrometry (UCIAMS) Facility for sussemione samples ACAD2303 and ACAD2305 (see *SI Text*).

such as hemiones and donkeys (i.e., two noncaballine Old World equids), and less than that between hippidions and different *Equus* species (such as caballines horses or donkeys). We therefore support the Peruvian specimens as a distinct species within the same genus as hippidiforms from Patagonian/Chile. The extinct Peruvian taxa is referable to *Onohippidium devillei* according to the mean length and width of metatarsals, which are similar to the equids from the type locality of Tarija, Bolivia (33), and should be renamed *Hippidium devillei*. Our analysis therefore shows that *Onohippidium* is a junior synonym genus of *Hippidium* in agreement with Alberdi and Prado (34, 35) and in contrast to a reanalysis of facial characters, dental patterns, and metatarsal proportions (19). The molecular dating estimates (Table S4) suggest that the hippidiforms (from South America) separated from the North American taxa (caballines and NWSL) after 3.2–3.6 MYA, consistent with the final formation of the Panamanian isthmus ≈ 3.0 –3.7 MYA (36). Though we have only limited sample coverage, especially in Peru, we speculate that after this first vicariant event, the Peruvian *Hippidium* species may have specialized in higher altitude Andean habitats (almost all specimens from *H. devillei* come from high altitude environments) (37), and was genetically isolated from a more generalist species (*H. saldiasi/principale*) that extended from Brazil to Patagonia and showed large size plasticity (the adult weight of *H. principale* was twice that of *H. saldiasi*) (see refs. 35–37).

Africa. Four specimens identified as giant Cape zebra (*E. capensis*) were found to have good-quality DNA preservation. Three specimens were from Wonderwerk Cave, south of Kuruman in central southern Africa, and one from Glen Craig Shelter near Port Elizabeth on the southeast coast. Fragments up to 477-bp long could be amplified, and the six target HVR-1 and *cyt b* gene fragments were characterized (687 bp in total) apart from one Wonderwerk specimen (ACAD227), which yielded only the barcode fragment (see below and Table S1A). Maximum likelihood and Bayesian analyses placed the sequences amongst the plains

zebra group (Fig. 1, nodes C1 and J; Fig. S2B and C), which consists of six subspecies that lack population genetic structure (9) and includes the extinct quagga (8) (*E. q. quagga*). The Cape zebra sequences clustered amongst the two southern subspecies (*E. q. quagga*, *E. q. burchellii*), with the Wonderwerk specimens ACAD226 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 AF220918). Although only limited sequence information was retrieved (Table S1A), sample ACAD227 showed minimum genetic distance with node K *E. q. boehmi* specimens (i.e., over 87 nucleotides, one substitution is observed versus four or more for J1/J2 haplotypes). In addition, the Glen Craig specimen (ACAD236) grouped within the southern clade, suggesting limited reproductive isolation (if any) amongst Cape quaggas and plains zebras, or potential issues with the morphological identification of the fossil material. In support of the latter possibility, radiocarbon dating of one Wonderwerk specimen (ACAD230) yielded a recent date (cal. 148 ± 120 BP) in contrast to an expected Pleistocene age. Though further specimens and dates are obviously required, all specimens morphologically identified as Cape zebras yielded sequences that grouped phylogenetically within plains zebras. The genetic distance between the giant Cape zebra and extant plains zebras is within the range of the intraspecific diversity found within other zebra species (11) (Fig. 2 and Table S3), and so we do not consider separate species (or subspecies) status is warranted (until independent information from nuclear loci is available). In contrast, Churcher (24) used the large size of giant Cape zebra fossils (estimated 400 kg and 150 cm height at the withers) to support conspecific status with Grevy's zebra, the largest extant wild equid. However, skulls are considered to be the best taxonomic indicators for equids at the morphological level (38), and the skull proportions of a giant Cape zebra from Elandsfontein (South Africa) showed a close affinity with true Cape quaggas (23), in accordance with our genetic findings. Together, these findings suggest that the giant Cape zebra

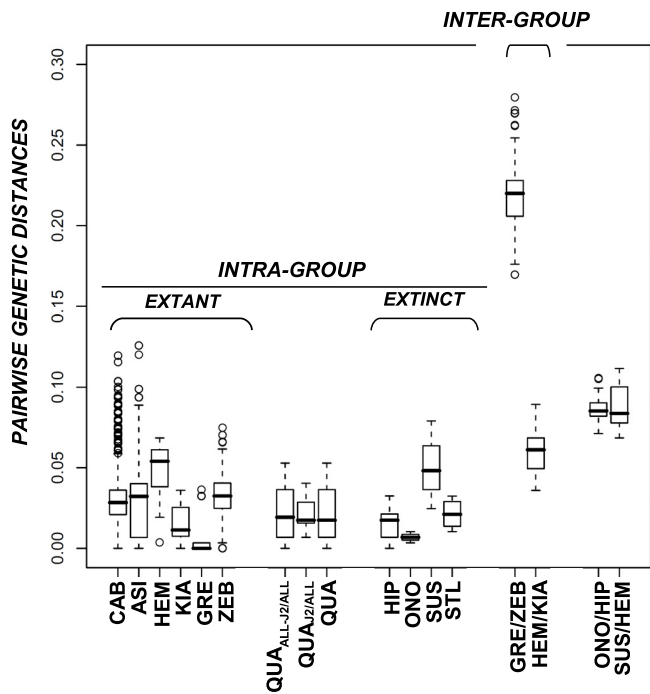


Fig. 2. Intra- and interspecific corrected (GTR+ Γ +I) pairwise distances among equids. The dataset consisted of 1,544 sequences encompassing positions 15518 and 15818 from the complete horse mitochondrial genome (GenBank accession no. X79547). Numeric values are reported on [Table S3A](#). A complete list of accession numbers used for defining each taxonomic group is provided as [SI Text](#). Parameters and distances were estimated using PhyML 3.0 and PAUP* 4.0, respectively. Bar, median; box, 50% quantiles; bars, 75% quantiles; CAB, *E. caballus*; ASI, *E. asinus*; QUA, *E. quagga* (all Plain zebras, including *E. quagga quagga*); QUA_{J2}, node J2 is defined on Fig. 1 and includes samples ACAD226 and ACAD230; QUA_{ALL-J2}, QUA but excluding samples for node J2; GRE, *E. grevyi*; ZEB, *E. hartmannae*; HEM, *E. hemionus*; KIA, *E. kiang*; HIP, *Hippidion saldasil/principale*; ONO, *Hippidion devillei* (Peruvian hippidions; ACAD3615, ACAD3625, ACAD3627, ACAD3628, and ACAD3629); STL, New World Stilt-Legged horses; SUS, Sussemiones (Khakassia, SW Siberia; ACAD2302, ACAD2303, and ACAD2305). Only two distributions (GRE/ZEB and HEM/KIA) are shown for the intergroup genetic distances, given that all others are included within these two.

formed part of the same diverse taxonomic group as plains 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 noncaballine 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 hemione; Fig. 2 and [Table S3](#)), 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 mor-

phologically similar to the three Khakassian specimens have been recovered from cave sites in the Altai Mountains (Denisova, Kaminnaya, Logovo Gieny, Okladnikov, and Strashnaya) 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. S4 G](#)) 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. S2 A](#)). 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 S3 C](#)) 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 elephantids (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, Grevy's zebras nest within paraphyletic hemionids, 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 brown bears (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 Adelaide ($n = 24$) (Table S1 A and B) according to appropriate techniques and controls (29). In Lyon, a maximum of three different samples were coextracted at any time to limit the potential for cross-contamination. Mock extractions and the three different amplification controls were included in each experiment to monitor contamination. DNA was extracted using either a phenol/chloroform method (59, 60) or a silica-based method adapted from Rohland et al. (61). The latter method is designed to maximize recovery of PCR-amplifiable DNA from ancient bone and tooth specimens while minimizing coextraction 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 15518–16063 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 S1 A). In addition, 121 PCR amplicons (16 samples) were independently analyzed by direct sequencing at ACAD (two additional amplicons were cloned and sequenced for ACAD227; Table S1 A). 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 S1 A), 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 BYPASS-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) neogeus* sequences (EU030680 and EU030681). 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 (SI 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 (SI Text). A RY-recoded rooted dataset was used for further phylogenetic analyses to limit possible substitution 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. S2 D–F, Table S2, and SI Text), and the monophyly of the principal nodes was assessed using likelihood-based statistical tests implemented in PAUP*4.0 [Kishino-Hasegawa (KH) and Shimodeira-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+ Γ +I) pairwise distance matrices were recovered from PAUP* 4.0b10 (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. S2 B and C). Bayesian analyses were applied to the unrooted HVR-1 dataset to assess divergence times, using BEAST (67) and assuming a HKY+ Γ +I model of molecular evolution. The age of specimens was recovered from either radiocarbon dating or the literature (SI Text, dating; Table S1 A). 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 Marilynne Duffraisie and Marta Kasper for technical help; Benjamin Gillet for his help in the Palgene platform (Centre National de la Recherche Scientifique/Ecole Normale Supérieure de Lyon); Francis Thackeray (Transvaal Museum, Pretoria, South Africa) for confirming the morphological identity of the *E. capensis* specimens; Bastien Llamas for giving helpful feedback on the map; Jane Wheeler [(Coordinadora de Investigación y Desarrollo de Camélidos Sudamericanos (CONOPA)] for assistance with sampling the Peruvian material; Oxford Radiocarbon Accelerator Unit, CU INSTAAR, and UCI 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 EGIDE), 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 grevyi* hairs used in this study, and Muséum National d'Histoire Naturelle, Museo de Historia Natural (permit no. RDN 1740/INC), Museum of the Polytechnical University of Ecuador (Solymar López and José Luis Román Carrión), and Dr. A. Derevianko for morphological information about Portuguese and Russian equid fossils.

- Huxley TH (1870) Address delivered at the anniversary meeting of the Geological Society of London, February 18. *Proceedings of the Geological Society of London* (Taylor and Francis, London), pp 16–38. [Reported in Gould SJ (1987) Life's little joke. *Nat Hist* 96:16–25.]
- Gould SJ (1994) *Hen's Teeth and Horse's Toe: Further Reflections in Natural History* (Norton, New York).
- Simpson GG (1951). *Horses* (Oxford Univ Press, Oxford, U.K.).
- MacFadden BJ (2005) Fossil horses—Evidence for evolution. *Science* 307:1728–1730.
- Azzaroli A (2002) Phylogeny of the genus *Equus* L. *Palaeontogr Ital* 89:11–16.
- Groves CP, Willoughby DP (1981) Studies on the taxonomy and phylogeny of the genus *Equus*. *Mammalia* 45:321–354.
- Oakenfull EA, Lim HN, Ryder OA (2000) A survey of equid mitochondrial DNA: Implications for the evolution, genetic diversity and conservation of *Equus*. *Cons Genet* 1:341–353.
- Groves CP, Bell CH (2004) New investigations on the taxonomy of the zebras genus *Equus*, subgenus. *Hippotigris Mammal Biol* 69:182–196.
- Lorenzen ED, Arcander P, Siegmund HR (2008) High variation and very low differentiation in wide ranging plains zebra (*Equus quagga*): Insights from mtDNA and microsatellites. *Mol Ecol* 17:2812–2824.
- Eisenmann V (1979) Evolutionary characters and phylogeny of the genus *Equus* (Mammalia, Perissodactyla) (Translated from French). *CR Acad Sci Paris D* 288:497–500.
- Leonard JA, et al. (2005) A rapid loss of stripes: The evolutionary history of the extinct quagga. *Biol Lett* 1:291–295.
- Pääbo S, et al. (2004) Genetic analyses from ancient DNA. *Annu Rev Genet* 38:645–679.
- Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC (1984) DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312:282–284.
- Vilà C, et al. (2001) Widespread origins of domestic horse lineages. *Science* 291:474–477.
- Orlando L, Eisenmann V, Reynier F, Sondaar P, Hänni C (2003) Morphological convergence in *Hippidion* and *Equus* (*Amerhippus*) South American equids elucidated by ancient DNA analysis. *J Mol Evol* 57(Suppl 1):S29–S40.
- Weinstock J, et al. (2005) Evolution, systematics, and phylogeography of pleistocene horses in the new world: A molecular perspective. *PLoS Biol* 3:e241.
- Orlando L, et al. (2006) Geographic distribution of an extinct equid (*Equus hydruntinus*: Mammalia, Equidae) revealed by morphological and genetical analyses of fossils. *Mol Ecol* 15:2083–2093.
- Orlando L, et al. (2008) Ancient DNA clarifies the evolutionary history of american late pleistocene equids. *J Mol Evol* 66:533–538.
- MacFadden BJ (1997) Pleistocene horses from Tarija, Bolivia, and validity of the genus *Onhippidium* (Mammalia: Equidae). *J Vertebr Paleontol* 17:199–218.
- Reguera MA, Candelà AM, Alonso RN (2007) Biochronology and biostratigraphy of the Uquia Formation (Pliocene–early Pleistocene, NW Argentina) and its significance in the Great American Biotic Interchange. *J S Am Earth Sci* 23:1–16.
- Alberdi MT, Prado JL, Prieto A (2005) Considerations on the paper “Morphological Convergence in *Hippidion* and *Equus* (*Amerhippus*) South American equids elucidated by ancient DNA analysis” by Ludovic Orlando, Véra Eisenmann, Frédéric Reynier, Paul Sondaar, and Catherine Hänni. *J Mol Evol* 61:145–147.
- Azzaroli A (1998) The genus *Equus* in North America—The Pleistocene species. *Palaeont Ital* 85:1–60.
- Eisenmann V (2000) *Equus capensis* (Mammalia, Perissodactyla) from Elandsfontein. *Palaeont Afr* 36:91–96.
- Churcher CS (1986) The extinct Cape zebra. *Sagittarius* 1:4–5.
- Thackeray JF (1988) Zebras from Wonderwerk cave, northern Cape Province, South Africa: Attempts to distinguish *Equus burchelli* and *E quagga*. *South African J Sci* 84:99–101.
- Smith CI, Chamberlain AT, Riley MS, Stringer C, Collins MJ (2003) The thermal history of human fossils and the likelihood of successful DNA amplification. *J Hum Evol* 45:203–217.
- Hofreiter M, et al. (2001) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res* 29:4793–4799.
- Brotherton P, et al. (2007) Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Res* 35:5717–5728.
- Cooper A, Poinar H (2000) Ancient DNA: Do it right or not at all. *Nature* 289:1139.
- Orlando L, Hänni C, Douady CJ (2007) Mammoth and elephant phylogenetic relationships: *Mammot Americanum*, the missing outgroup. *Evol Bioinform Online* 3:45–51.
- Rohland N, et al. (2007) Proboscidean mitogenomics: Chronology and mode of elephant evolution using mastodon as outgroup. *PLoS Biol* 5:e207.
- Prado JL, Alberdi MT (1996) A cladistic analysis of the horses of the tribe equine. *J Paleontol* 39:663–680.
- Shockey BJ, et al. (2009) New Pleistocene cave faunas of the Andes of central Perú: Radiocarbon ages and the survival of low latitude, Pleistocene DNA. *Palaeontologia Electronica* 12(3):15A.
- Alberdi MT, Prado JL (1993) Review of the genus *Hippidion* Owen, 1869 (Mammalia: Perissodactyla) from the Pleistocene of South America. *Zool J Linn Soc* 108:1–22.
- Alberdi MT, Prado JL (1998) Comments on Pleistocene horses from Tarija, Bolivia, and the validity of the genus *Onhippidium* (Mammalia: Equidae), by B. J. MacFadden. *J Vert Paleontol* 18:669–672.
- Ibaraki M (1997) Closing of the Central American Seaway and Neogene coastal upwelling along the Pacific coast of South America. *Tectonophysics* 281:99–104.
- Alberdi MT, Prado JL, Ortiz Jaureguizar E (1995) Patterns of body size changes in fossil and living Equini (Perissodactyla). *Biol J Linn Soc* 54:349–370.
- Eisenmann V, Brink JS (2000) Koffiefontein quaggas and true Cape quaggas: The importance of basic skull morphology. *South Afric J Sci*, 96:529–533.
- Vasiliev SK, Derevianko AP, Markin SV (2006) Large mammal fauna of the Sartan period from the northwestern Altai. *Arch Ethn Anthropol Eurasia* 2:2–22.
- Eisenmann V (2006) Pliocene and pleistocene equids: Paleontology versus molecular biology. *Late Neogene and Quaternary Biodiversity and Evolution: Regional Developments and Interregional Correlations*, eds Kahlke RD, Maul LC, Mazza P. Proceedings of the 18th International Senckenberg Conference (Sixth International Palaeontological Colloquium in Weimar), Vol 1. *Courier Forschungsinstitut Senckenberg* 256:71–89.
- Eisenmann V, Howe J, Pichardo M. Old World hemionines and New World slender species (Mammalia, Equidae). *Palaeovertebrata*, in press.
- Antunes MT (2006) The Zebra (Equidae) and its extinction in Portugal, with an appendix on the noun zebro and the modern zebra. *Equids in Time and Space*, ed Mashkour M (Oxbow Books, Oxford, UK), pp 210–235.
- Bonifay MF (1991) *Equus hydruntinus* Regalia minor n.ssp. from the caves of Lunel-Viel (Hérault, France). *Beihfte Tübinger Atlas Vorderen Orients Reihe A* 19:178–216.
- Wilms C (1989) The extinction of the European wild ass (Translated from German). *Germania* 67:143–148.
- Burke A, Eisenmann V, Amblar G (2003) The systematic position of *Equus hydruntinus*, an extinct species of Pleistocene equid. *Quat Res* 59:459–469.
- Ryder OA, Chemnick LG (1990) Chromosomal and molecular evolution in Asiatic wild asses. *Genetica* 83:67–72.
- Krause J, et al. (2008) Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evol Biol* 8:220.
- Oakenfull EA, Clegg JB (1998) Phylogenetic relationships within the genus *Equus* and the evolution of the α and θ globin genes. *J Mol Evol* 47:772–783.
- Shapiro B, et al. (2004) Rise and fall of the Beringian steppe bison. *Science* 306:1561–1565.
- Barnett R, et al. (2009) Phylogeography of lions (*Panthera leo* spp.) reveals three distinct taxa and a late Pleistocene reduction in genetic diversity. *Mol Ecol* 18:1668–1677.
- Barnes I, Mathews P, Shapiro B, Jensen D, Cooper A (2002) Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science* 295:2267–2270.
- Bunce M, et al. (2003) Extreme reversed sexual size dimorphism in the extinct New Zealand moa *Dinornis*. *Nature* 425:172–175.
- Huynen L, Millar CD, Scofield RP, Lambert DM (2003) Nuclear DNA sequences detect species limits in ancient moa. *Nature* 425:175–178.
- Wood B, Lonergan N (2008) The hominin fossil record: Taxa, grades and clades. *J Anat* 212:354–376.
- Wood B, Collard M (1999) The human genus. *Science* 284:65–71.
- Knapp M, et al. (2009) First DNA sequences from Asian cave bear fossils reveal deep divergences and complex phylogeographic patterns. *Mol Ecol* 18:1225–1238.
- Hofreiter M (2008) Long DNA sequences and large data sets: Investigating the Quaternary via ancient DNA. *Quat Sci Rev* 27:2586–2592.
- Hofreiter M, Stewart J (2009) Ecological change, range fluctuations and population dynamics during the Pleistocene. *Curr Biol* 19:R584–R594.
- Loreille, et al. (2001) Ancient DNA analysis reveals divergence of the cave bear, *Ursus spelaeus*, and brown bear, *Ursus arctos*, lineages. *Curr Biol* 11:200–203.
- Orlando L, et al. (2006) Revisiting neanderthal diversity with a 100,000 year old mtDNA sequence. *Curr Biol* 16:R400–R402.
- Rohland N, Hofreiter M (2007) Ancient DNA extraction from bones and teeth. *Nat Protocol* 2:1756–1762.
- Mateiu LM, Rannala BH (2008) Bayesian inference of errors in ancient DNA caused by postmortem degradation. *Mol Biol Evol* 25:1503–1511.
- Galtier N, Gouy M, Gautier C (1996) SeaView and PHYLO.WIN, two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Sci* 12:543–548.
- Swofford DL (2000) PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4 (Sinauer Associates, Sunderland, MA).
- R Development Core Team (2007) R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna).
- Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48.
- Drummond AJ, Rambaut (2007) A BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.
- Eisenmann V (1980) Extinct and extant horses (*Equus sensu lato*): Skulls and superior cheek teeth (Translated from French). *Cah Paleont*, CNRS eds (Paris).
- Geraads D, et al. (1986) The Pleistocene hominid site of Ternifine, Algeria: New results on the environment, age, and human industries. *Quaternary Res* 25:380–386.