Title:

Ancient genomes redate the extinction of *Sussemionus*, a subgenus of *Equus*, to late Holocene

Dawei Cai[1,†,*], Siqi Zhu[1,†], Mian Gong[2,†], Naifan Zhang[1], Jia Wen[2], Qiyao Liang[1], Weilu Sun[1], Xinyue Shao[1], Yaqi Guo[1], Yudong Cai[2], Zhuqing Zheng[2], Wei Zhang[3], Songmei Hu[4], Xiaoyang Wang[5], He Tian[3], Youqian Li[3], Wei Liu[3], Miaomiao Yang[4], Jian Yang[5], Duo Wu[6], Ludovic Orlando[7,*], and Yu Jiang[b,*]

[1]Research Center for Chinese Frontier Archaeology, Jilin University, Changchun, 130012, China.

[2]Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling, 712100, China.

[3]Heilongjiang Provincial Institute of Cultural Relics and Archaeology, Harbin, 150008, China.

[4]Shaanxi Provincial Institute of Archaeology, Xi'an, 710054, China.

[5]Ningxia Institute of Cultural Relics and Archaeology, Yinchuan, 750001, China.

[6]College of Earth and Environmental Sciences; MOE Key Laboratory of Western China's Environmental Systems, Lanzhou University, Lanzhou 730000, China.

[7]Centre d'Anthropobiologie et de Génomique de Toulouse (CAGT), CNRS UMR 5288, Université de Toulouse, Université Paul Sabatier, Toulouse, France.

†D.C., S.Z., and M.G. contributed equally to this work.

*For correspondence: caidw@jlu.edu.cn; ludovic.orlando@univ-tlse3.fr; yu.jiang@nwafu.edu.cn.

Abstract

The exceptionally-rich fossil record available for the equid family has provided textbook examples of macroevolutionary changes. Horses, asses and zebras represent three extant subgenus of *Equus* lineage, while the *Sussemionus* subgenus is another remarkable *Equus* lineage ranging from North America to Ethiopia in Pleistocene. We sequenced 26 archaeological specimens from northern China in Holocene showing morphological features reminiscent of *Equus ovodovi*, a species representative of *Sussemionus*, and further confirmed them as this species by genetic analyses. Thus, we present the first high-quality complete genome of the *Sussemionus* that we sequenced to 12.0× depth-of-coverage and demonstrate that it survived until ~3,500 years ago, despite the continued demographic collapse during the Last Glacial Maximum and the great human expansion in East Asia. We also confirmed the *Equus* phylogenetic tree, and found *Sussemionus* diverged from the ancestor of non-caballine equids ~2.3-2.7 Million years ago and admixture events could have taken place between them. Our works suggest the small genetic diversity but not the enhanced

inbreeding mainly limited the chances of survival of the species, and illustrates how ancient DNA can inform on extinction dynamics and the long-term resilience of species surviving in cryptic population pockets.

Introduction

Today, all of the seven extant species forming the horse family belong to one single genus, *Equus*. It emerged in North America some 4.0-4.5 million years ago (L. Orlando et al., 2013), and first spread to Eurasia ~2.6 million years ago, via the Beringia land bridge (Lindsay, Opdyke, & Johnson, 1980). This first vicariance and expansion out-of-America led to the emergence of the ancestors of zebras, hemiones and donkeys, a group collectively known as non-caballine equids. Another expansion through Beringia occurred later, bringing caballine equids (i.e. those most closely related to the horse) into the Old-World, where they survived until their domestication some ~5,500 years ago (Gaunitz et al., 2018; Outram et al., 2009).

In the recent years, ancient DNA (aDNA) data have revealed that the genetic diversity of non-caballine *Equus* was considerably larger diversity in the past than it is today (Librado & Orlando, 2020; Ludovic Orlando, 2020), especially as the first mitochondrial DNA (mtDNA) data of *Equus* (*Sussemionus*) specimens were collected (hereafter referred to as Sussemiones) (Eisenmann, 2010). This lineage radiated across North America, Africa, and Siberia, and developed multiple adaptations to a whole range of arid and humid environments (Eisenmann, 2010). Sussemiones was first believed to have become extinct during the Middle Pleistocene as the last known

specimen showing typical morpho-anatomy dated back to approximately 500,000 years ago (Vasiliev, 2013). However, DNA results obtained on multiple osseous remains within the radiocarbon range and showing morphological traits reminiscent of the Eurasian Sussemiones species indicated that the lineage in fact survived until the Late Pleistocene (Druzhkova et al., 2017; L. Orlando et al., 2009; Vilstrup et al., 2013; Yuan et al., 2019). Pioneering publications indicated survival dates 40-50 kya in southeastern Siberia, Russia (Proskuryakova cave) (L. Orlando et al., 2009; Vilstrup et al., 2013), ~32 kya at the Denisova cave (Druzhkova et al., 2017), and ~12.6 kya at northeastern China (Yuan et al., 2019).

Despite an abundant fossil material, only a limited number of Sussemiones specimens have been investigated for Ancient mitochondrial DNA (aDNA), which showed that Sussemiones formed a non-caballine equine lineage. However, the exact placement of Sussemiones could not be resolved (Heintzman et al., 2017; L. Orlando et al., 2009; Vilstrup et al., 2013). In this study, we have carried out archaeological excavations in three Holocene sites in China, and uncovered equine samples showing distinct morphological features when compared to horses and donkeys (Figure 1—figure supplement 1). The whole mitochondrial and nuclear genome data allowed us to unveil the phylogenetic placement of Sussemiones within the *Equus* evolutionary tree, the timing of its divergence to other non-caballine equids, signatures of demographic collapse and adaptations specific to this lineage, and the extinction dynamics of this lineage.

Results

Archaeological samples and sequencing data

All the equine specimens investigated in this study were excavated from three archaeological sites in China (Figure 1 and Table S1) (Honghe, Heilongjiang Province (Figure 1—figure supplement 2); Muzhuzhuliang, Shaanxi Province; Shatangbeiyuan, Ningxia Province). They showed morphological and genetic signatures distinct from those of extant horses and donkeys. The morphological differences were especially marked in the second and third molars, which appeared to be smaller than in modern horses, and were reminiscent of the third molars paracones and metacones observed in Sussemiones specimens (Figure 1B). Combined, these samples were radiocarbon dated to 3,477-4,481 calibrated years before the present (cal BP), including the latest sample HH13H with 3,477-3,637 cal BP (Table S2). They could, thus, represent some of the latest surviving Sussemiones individuals prior to their extinction.

We next aimed at genetically characterizing and identifying the taxonomic status of these specimens using high-throughput DNA sequencing technologies. We extracted ancient DNA from a total of 26 specimens and sequenced the whole nuclear genome at ~0.002 to 12.0 times coverage, including three samples from Honghe provided 12.0×, 3.5× and 1.0× nuclear genome (Table S1). Comparison of the X chromosome and autosomal coverage revealed the presence of 15 males and 11 females.

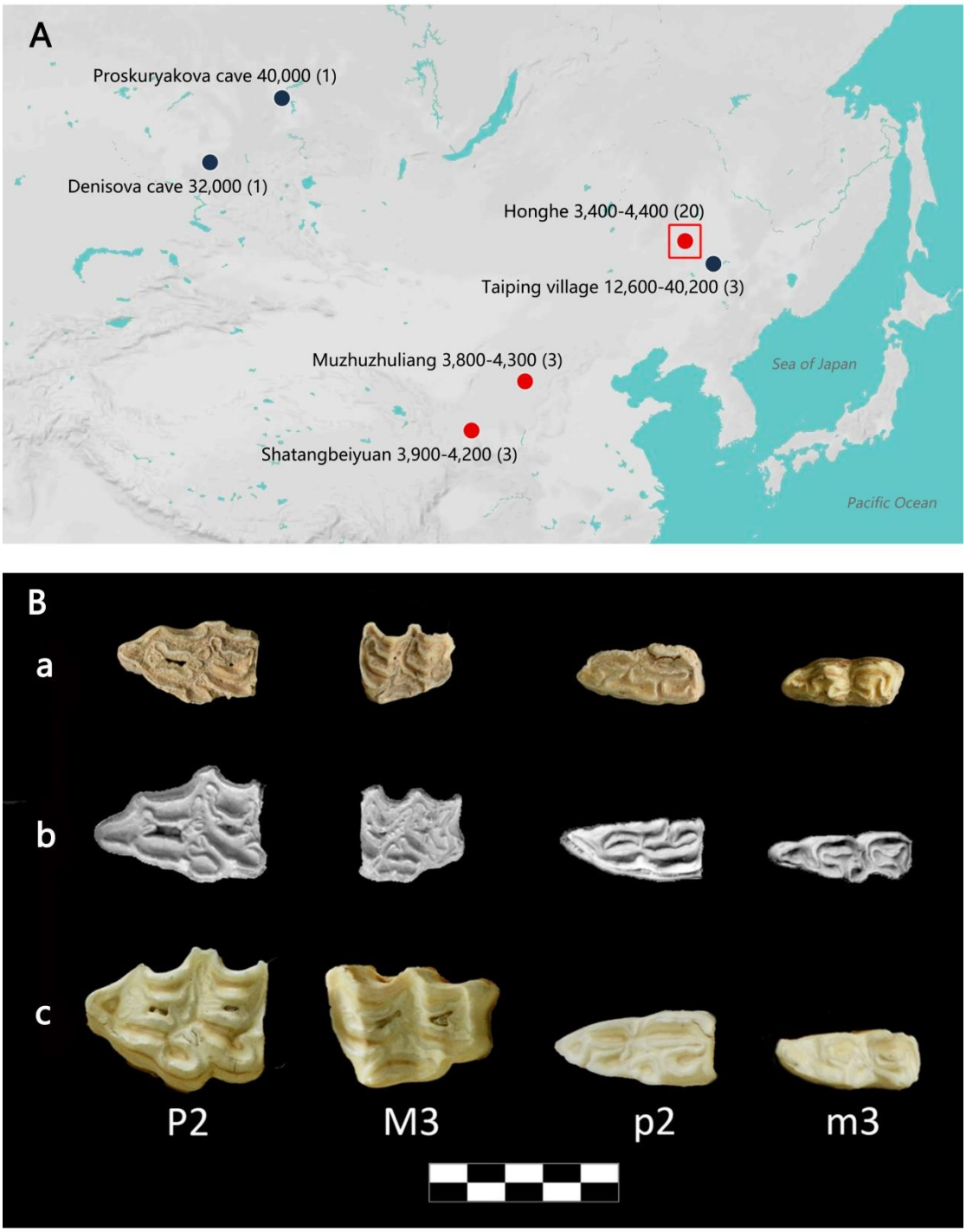


Figure 1. Sampling distribution and archaeological research of *E.* (*Sussemionus*) *ovodovi*. (**A**) *E.* (*Sussemionus*) *ovodovi* geographic range. The three red circles indicate the archaeological sites analyzed in this study. The site (Honghe) that delivered the complete genome sequence at 12.0-fold average depth-of-coverage

(HH06D) is highlighted with a square. The black circles indicate sites that provided complete mitochondrial genome sequences in previous studies (Druzhkova et al., 2017; L. Orlando et al., 2009; Vilstrup et al., 2013; Yuan et al., 2019). The temporal range covered by the different samples analyzed is given in years before present (YBP) and follows the name of each site. Numbers between parentheses indicate the number of samples for which DNA sequence data could be generated. (**B**) *Facies masticatoria dentis* of P2, M3, p2 and m3 for the *E. (Susseminous) ovodovi* samples of the Honghe site analyzed here (a), *E. Sussemionus* (b), and *E. caballus* (c). Teeth from the right side are shown, except for *E. Sussemionus*. The erupted teeth of the samples of the Honghe site appear to be smaller than those of the *E. Sussemionus* specimen.

The following figure supplements are available for figure 1:

Figure supplement 1. Archaeological material investigated in this study.

Figure supplement 2. Aerial view of the Honghe site.

Taxonomic status

To assess whether the sequenced specimens belonged to the same taxonomic group or comprised different species, we carried out a Principle Component Analysis (PCA) including all the equine species sequenced at the genome level (depth of coverage ≥ 1×) (Figure 2A and Figure 2—figure supplements 3 and 4). For this, we downloaded 11 previously-published equine genomes representing all extant species of equids and the extinct quagga zebra (Huang et al., 2015; Jonsson et al., 2014; Kalbfleisch et al., 2018; L. Orlando et al., 2013; Renaud et al., 2018) (Table S3). All the Chinese

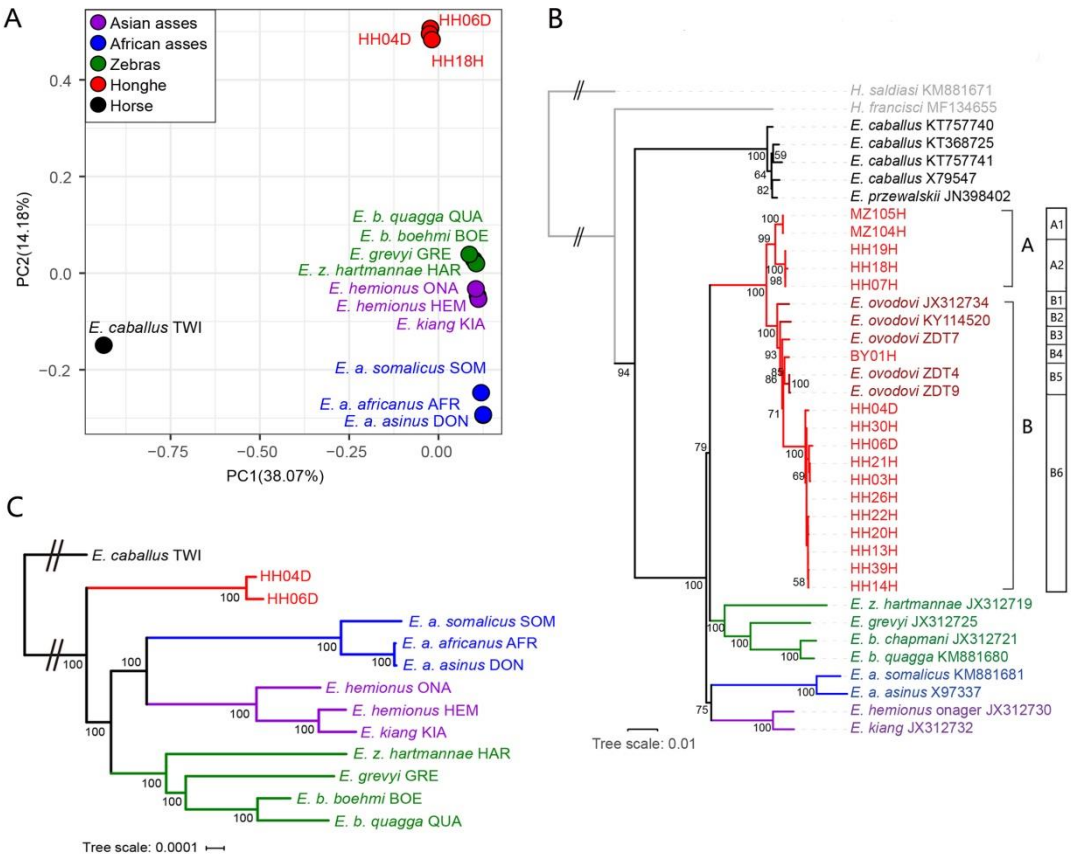
specimens analyzed in this study were found to cluster together along the first two PCA components, in a group that was distinct from all other equine species (Figure 2A and Figure 2—figure supplement 3) but closer to non-caballine equine species than to the horse (Figure 2A). This suggested that they were all members of a unique taxonomic group, most related to non-caballine equids.

Maximum likelihood (ML) phylogenetic analyses including the nearly-complete 17 mitochondrial genomes reported in this study (Table S1, depth of coverage no less than 1×) confirmed their clustering with non-caballine equids, within a single monophyletic group that also included five previously-characterized Sussemiones specimens (Figure 2B and Figure 2—figure supplements 5 and 6). This grouping was supported with maximal (100%) bootstrap support. This, and the PCA clustering, indicated that the different excavation sites investigated in this study in fact all provided specimens that belonged to the *E.* (*Sussemionus*) *ovodovi* species.

We also inferred maternal lineages of *E. ovodovi* using complete mitogenomes. Phylogenetic analyses showed that the species were mainly divided into two major haplogroups (A and B), of which most of the specimens were characterised by the haplogroup B, including all previously reported sequences, whereas 2 Muzhuzhuliang and 3 Honghe individuals belonged to the new haplogroup A (Figure 2B). Further distinction could also be carried out in the two haplogroups (Figure 2B).

To further assess phylogenetic affinities, we used the two genomes characterized to at least 3× average depth-of-coverage (HH04D and HH06D) to replace Sussemiones within the equine phylogenetic tree. To achieve this, we used ML phylogenetic

reconstruction and an alignment of the coding sequences of the protein coding genes (Figure 2C and Figure 2—figure supplement 7A). This showed that the Chinese ancient specimens branched off before the radiation leading to modern asses and zebras (Figure 2C). Similar tree topologies were recovered using whole-genome SNPs by TreeMix (Pickrell & Pritchard, 2012) (Figure 2—figure supplements 8 and 9). Combined with the analysis of the occlusal surface of the molars, in particular the absence of the caballine notch, the shape of metacones and protocones, and the reduced tooth size (Figure 1B), our analyses allowed us to conclude that the material analyzed represented small specimens of the extinct *Equus* (*Sussemionus*) *ovodovi*. This lineage, thus, survived in China during the Holocene, and until cal 3,477-4,481 cal BP, which is approximately ~8,300-9,300 years after the latest known specimen to date (Druzhkova et al., 2017; L. Orlando et al., 2009; Vilstrup et al., 2013; Yuan et al., 2019).



171

172 **Figure 2.** Genetic relationships within the *Equus* genus. Honghe (HH),

173 Muzhuzhuliang (MZ) and Shatangbeiyuan (BY) specimens are shown in red, while

174 Asian asses, African asses, zebras and horses are shown in whereas purple, blue, green

175 and black, respectively. (**A**) PCA based on genotype likelihoods, including horses and

176 all other extant non-caballine lineages (16,293,825 bp, excluding transitions). Only

177 specimens whose genomes were sequenced at least to 1.0× average depth-of-coverage

178 are included. (**B**) Maximum likelihood tree based on 5 mitochondrial partitions

179 (representing a total of 15,399 bp). Previously published *E. ovodovi* sequences are

180 shown in deep red. The tree was rooted using *Hippidion saldiasi* and *Haringtonhippus*

181 *francisci* as outgroups (not shown). Node supports were estimated from 1,000

182 bootstrap pseudo-replicates and are displayed only if greater than 50%. The black line

183 indicates the mitochondrial haplogroup A and B. (**C**) Maximum likelihood tree based
184 on sequences of 19,650 protein-coding genes with specimens sequenced at least 3.0×
185 average coverage (representing 32,756,854 bp).
186 The following figure supplements are available for figure 2:
187 **Figure supplement 1.** DNA Damage patterns for HH06D.
188 **Figure supplement 2.** Error profiles of the 26 ancient genomes characterized in this
189 study.
190 **Figure supplement 3.** Principal Component Analysis (PCA) based on genotype
191 likelihoods using the horse reference genome.
192 **Figure supplement 4.** Principal Component Analysis (PCA) based on genotype
193 likelihoods using the donkey reference genome.
194 **Figure supplement 5.** RAxML-NG (GTR+GAMMA model) Maximum Likelihood
195 phylogeny of complete mitochondrial sequence data.
196 **Figure supplement 6.** Bayesian mitochondrial phylogeny based on 6 partitions and
197 using *Hippidion Saldiasi* as outgroup.
198 **Figure supplement 7.** Exome-based Maximum likelihood phylogeny rooted by the
199 horse lineage.
200 **Figure supplement 8.** Treemix analysis of based on genome-wide SNP data
201 conditioned on transversions using the horse reference genome.
202 **Figure supplement 9.** Treemix analysis of based on genome-wide SNP data
203 conditioned on transversions using the donkey reference genome.
204

Interspecies admixture and demographic modeling

Bifurcating trees fail to capture possible admixture events between lineages. Yet, previous research has unveiled pervasive admixture within equids, even amongst extant equids showing different chromosomal numbers (Jonsson et al., 2014). We thus next assessed whether the genomic data showed evidence for gene flow between Sussemiones and other non-caballine equids. To achieve this, we first applied D-statistics (Soraggi, Wiuf, & Albrechtsen, 2018) to the genome sequence underlying 26 individual genomes and detected that *E. ovodovi* shared an excess of derived polymorphisms with asses than relative to zebras (Figure 3—figure supplements 1 and 2). This suggested that at least one admixture event could have taken place between Sussemiones and the ancestor of asses after their divergence from zebras.

We next leveraged the ancient genome characterized to high depth-of-coverage (HH06D) to reconstruct the equine demographic history using G-PhoCS (Gronau, Hubisz, Gulko, Danko, & Siepel, 2011). More specifically, we first selected members of each equine lineage representing a total number of 10 genomes, and assumed that the genus *Equus* emerged some 4.0-4.5 Mya, following previous estimates (L. Orlando et al., 2013). G-PhoCS analysis confirmed previous analyses indicating that the zebras and asses linages diverged ~2.0 Mya and that the deepest divergence within zebras and asses took place prior to ~1.5 Mya (Jonsson et al., 2014) (Figure 3). It revealed that the Sussemiones lineage diverged from the ancestor of extant non-caballine equids ~2.3-2.7 Mya, in line with the fossil record (Eisenmann, 2010). Allowing for migrations provided support for gene flow between Sussemiones and the

ancestor of asses and zebras (Figure 3). However, weak to no migrations were detected between Sussemiones and extant equids (Table S6). Importantly, the admixture between Sussemiones and the ancestor of asses seems to have been stronger than that between Sussemiones and the ancestor of zebras, in line with the results of D-statistics. G-PhoCS also supported the presence of significant unidirectional gene-flow prior to ~2.3-2.7 Mya, from the horse branch into the ancestral branch to all non-caballine equids, including Sussemiones (total migration rate 2.2-9.2%, Table S7). This is consistent with previous HMMCoal analyses applied to whole genome sequences of all extant equine species, which indicated significant gene-flow between the deepest branches of the *Equus* phylogenetic tree until 3.4 Mya, mostly from a caballine lineage into the ancestor of all non-caballine equids (Jonsson et al., 2014).

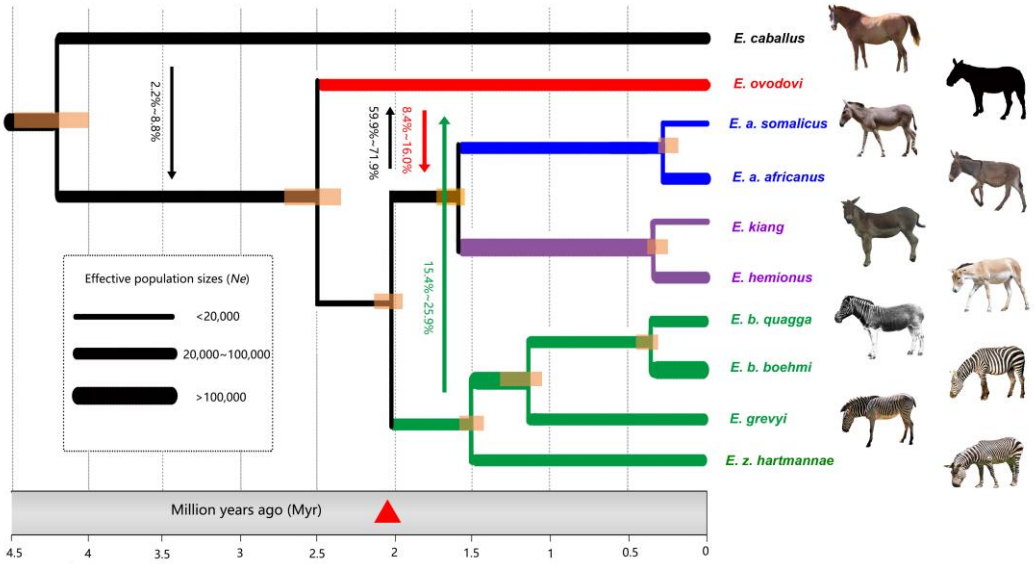


Figure 3. Demographic model for extinct and extant equine lineages as inferred by G-PhoCS (Gronau et al., 2011)**.** Node bars represent 95% confidence intervals. The

width of each branch is scaled with respect to effective population sizes (N_e). Independent N_e values were estimated for each individual branch of the tree, assuming constant effective sizes through time. Migration bands and probabilities of migration (transformed from total migration rates) are indicated with solid. The red triangle indicates the earliest *Sussemionus* evidence found in the fossil record. (Images: *E. caballus* by Infomastern, *E. a. somalicus* by cuatrok77, *E. kiang* by Dunnock_D, *E. a. africanus* by jay galvin, *E. hemionus* by Cloudtail the Snow Leopard, *E. z. hartmannae* by calestyo, *E. b. quagga* by Internet Archive Book Images, *E. b. boehmi* by GRIDArendal, and *E. grevyi* by 5of7.)

The following figure supplements are available for figure 3:

Figure supplement 1. *D*-statistics in the form of (zebra, ass; *E. ovodovi*, outgroup), using sequence alignments against the horse reference genome.

Figure supplement 2. *D*-statistics in the form of (zebra, ass; *E. ovodovi*, outgroup), using sequence alignments against the donkey reference genome.

Figure supplement 3. NJ tree of selected samples based on 15,324 candidate 'neutral' loci identified using sequence alignments against the horse reference genome.

Dynamic demographic profiles, heterozygosity and inbreeding levels

We next leveraged the high-coverage Sussemiones genome characterized here to explore further the demographic dynamics of this lineage until its extinction. When modeled as constant through time, population sizes in G-PhoCS indicated that most lineages, including Sussemiones, consisted of small populations, excepting the

Burchell's zebra (Table S8). Pairwise Sequential Markovian Coalescent (PSMC) analyses, however, provided us evidence for population size variation through time. First, the Sussemiones demographic trajectory was found to diverge from that of other non-caballine equids (specifically, *E. hemionus*) after ~2.0 Mya, confirming the divergence date estimate retrieved by G-PhoCS (Table S8). Second, we found that the Sussemiones demographic trajectory constantly increased during the last million year but stay at a level which was lower than that of other lineages for a long time, until it reached a peak between 74-84 kya. It was, then, followed by an approximately 45-fold collapse until 13 kya (Figure 4). The Sussemiones population size experienced a 1.86-fold collapse between 35-42 kya, which is almost coincident with the timing of the great human expansion to Eurasia (ie. 35-45 kya, (Henn, Cavalli-Sforza, & Feldman, 2012)). The lineage maintained extremely reduced population sizes through the Last Glacial Maximal (LGM, 19-26 kya) (Clark et al., 2009) and the Holocene, until it finally became extinct.

Importantly, the sample sequenced to sufficient coverage (HH06D) showed minimal heterozygosity and moderate inbreeding levels identified by the fraction of the segments within ROH (Figure 5). Strikingly, this is true in spite of the increased sequencing error rates of this genome, which likely inflate our estimates. The limited population sizes and genetic diversity but not the enhanced inbreeding may have limited the chances of survival of the species, ultimately leading to extinction.

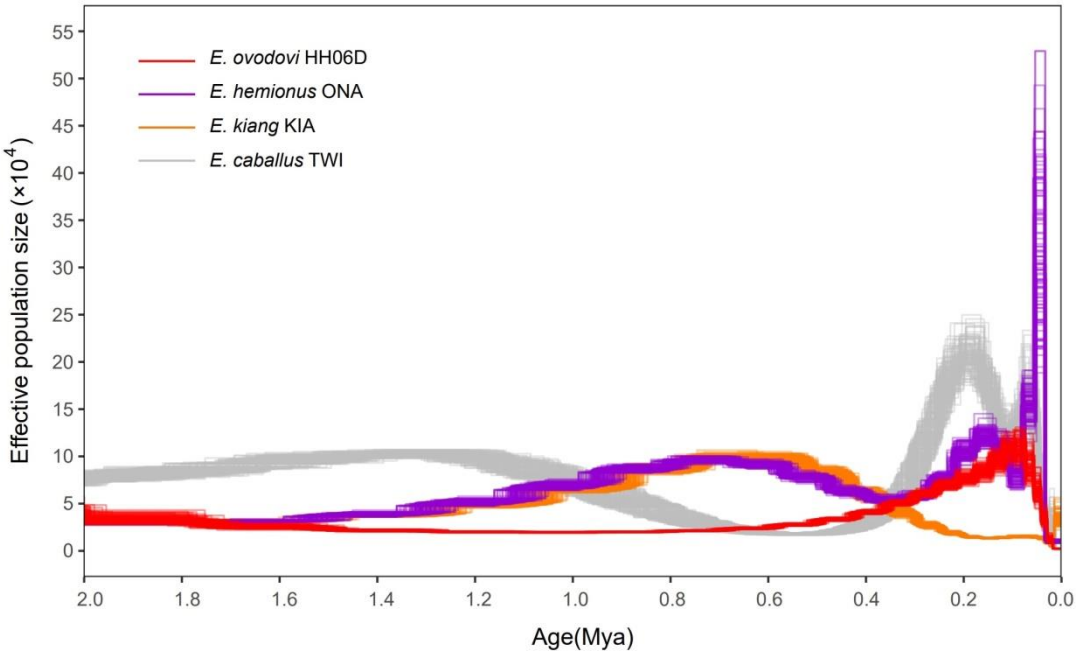


Figure 4. PSMC profiles (100 bootstrap pseudo-replicates) of four Eurasian equine species (*E. ovodovi* HH06D, *E. caballus* TWI (Kalbfleisch et al., 2018), *E. hemionus* ONA and *E. kiang* KIA) (Jonsson et al., 2014)**.** The y axis represents effective population size (×10,000), and the x axis represents millions of years before present. Faded lines show bootstrap values.

The following figure supplements are available for figure 4:

Figure supplement 1. PSMC bootstrap pseudo-replicates for samples with and without transitions.

Figure supplement 2. Determining the uniform false-negative rate (uFNR) that was necessary for PSMC scaling.

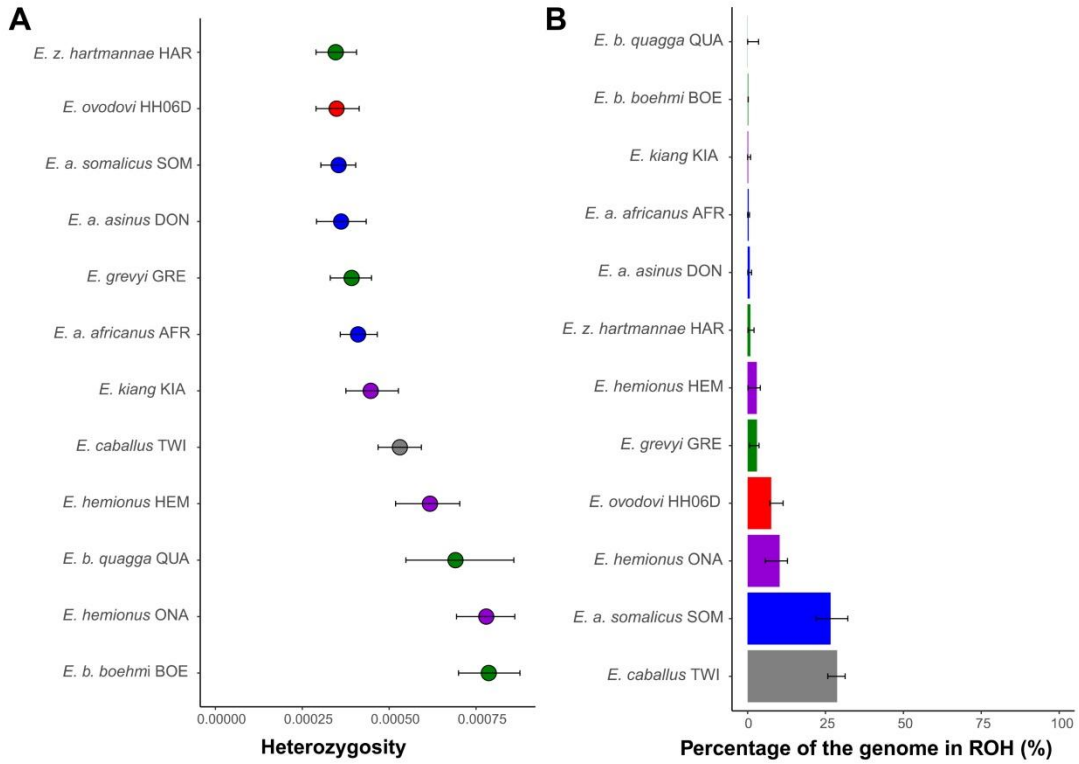


Figure 5. The Heterozygosity and Inbreeding levels of extinct and extant equine lineages. (**A**) Individual heterozygosity outside Runs-of-Homozygosity (ROH). (**B**) Fraction of the genome in ROH. Estimates were obtained excluding transitions and are shown together with their 95% confidence intervals. The colors mirror those from Figure 2.

The following figure supplements are available for figure 5:

Figure supplement 1. Heterozygosity rates outside Runs-of-Homozygosity (ROH) together with 95% confidence intervals.

Figure supplement 2. The fraction of the genome segments consisting of ROHs together with 95% confidence intervals.

Discussion

Phylogenetic placement of *Equus* (*Sussemionus*) *ovodovi*

In this study, we have characterized the first nuclear genomes of the now-extinct equine lineage, *E.* (*Sussemionus*) *ovodovi*, the last surviving member of the subgenus *Equus Sussemionus*. We demonstrated that this lineage survived in China well into the Holocene with the most recent specimens analyzed dating to ~3,477-3,637 cal BP. This is almost 9,000 years after the latest specimens previously documented in the fossil record (Druzhkova et al., 2017; Vilstrup et al., 2013; Yuan et al., 2019). Our work, thus, shows that *Sussemionus* represents the last currently known *Equus* subgenus to become extinct. Our work also adds to the list of recently identified members of the horse family that were still alive at the time horses and donkeys were first domesticated, some ~5,500 years ago (Fages et al., 2019; Gaunitz et al., 2018; Rossel et al., 2008). In contrast to those divergent members that were identified in Siberia and Iberia and both belonged to the horse species (Fages et al., 2019; Schubert, Jonsson, et al., 2014), Sussemiones members were most closely related to non-caballine equids. This is in agreement with previous studies (Der Sarkissian et al., 2015; Druzhkova et al., 2017; Heintzman et al., 2017; L. Orlando et al., 2009; Vilstrup et al., 2013; Yuan et al., 2019), which could, however, not fully resolve the exact phylogenetic placement of this species within non-caballine clade lineages as topological tests based on mitochondrial genomes received low confidence support (Der Sarkissian et al., 2015; Druzhkova et al., 2017; Heintzman et al., 2017; L. Orlando et al., 2009; Vilstrup et al., 2013; Yuan et al., 2019). Our study solved this question by reporting the first whole genome phylogeny of Sussemiones, which

confirmed with maximal bootstrap support this species as a unique basal lineage of non-caballine equids.

Suitable habitat and geographic distribution

Previous zooarchaeological and environmental research indicated an ecological range for Sussemiones overlapping with the grasslands located east of the Altay Mountains and west of the Yenisei River during the Late Pleistocene (Khenzykhenova et al., 2016; Malikov, 2016; Plasteeva, 2015; Shchetnikov, Klementiev, Filinov, & Semeney, 2015; Shunkov, 2018). Recent research also reported this species in northeastern China (~12,600-40,200 YBP), where similar climatic and ecological conditions were found at the time (Yuan et al., 2019). It could, thus, be speculated that Sussemiones was adapted to an environment with moderately dry climatic conditions and steppe landscapes (Yuan et al., 2019). However, our study identified Sussemiones in three late Holocene sites from China that have mild and humid environmental conditions. In addition, two distinct mitochondrial haplogroups from 22 individuals have been defined from the six known sites, suggesting that Sussemiones had adapted to different environmental regions. It also suggests that the species could adapt to a wider variety of habitats than previously hypothesized, and rejects the contention that the species became extinct as it could not survive in warmer climatic conditions (Yuan et al., 2019).

Interestingly, the Sussemiones specimens identified in this study were excavated from sites in northeastern China located at almost the same latitude as those

355 Sussemiones localities known so far from Russia, but also at lower latitudes (Figure
356 1A). This implies that the geographic range of *E. ovodovi* was larger than previously
357 expected and included at least Northern China and Southern Siberia. In the absence of
358 identified fossils from Mongolia, whether those two regions were in contact or
359 separated remains unknown. Further work is necessary to establish whether or not the
360 species survived in other pockets both within and outside China.
361

362 ## Demographic history with ancestral interspecific admixture

363 Our analyses reveal that the divergence between Sussemiones and the most recent
364 common ancestor of all extant non-caballine equids took place some ~2.3-2.7 Mya,
365 right before the divergence of zebras and asses. Post-divergence admixture events
366 with the lineage ancestral to asses and zebras on the one hand, and the lineage
367 ancestral to all extant zebras, were also identified (Figure 3 and Table S7). Our results,
368 thus, reveal non-caballine ancestral lineages occupying partly sympatric distributions
369 that were, consequently, different than those of their descendants, in which zebras are
370 restricted in Africa and Asian asses in Asia. Whether the admixture events identified
371 here directly involved the Sussemiones lineage or one (or more) ghost lineage(s)
372 closely related to Sussemiones requires further research.
373

374 ## Limited genetic diversity before extinction

375 The demographic profile of Sussemiones shows that after the peak of population size
376 culminating some ~74 kya, Sussemiones went through a slow and continuous decline

377 until 13 kya (Figure 4). This time period encompasses several major climate changes
378 (especially the LGM) and the great human expansion to Eurasia (~35-45 kyr BP)
379 (Henn et al., 2012). The effective size of Sussemiones populations that survived in
380 Northern China until at least ~3,500 years ago, remained extremely small, as indicated
381 by their extremely reduced heterozygosity levels compared to other extant and extinct
382 equine species, although extensive inbreeding was not detected (Figure 5). So
383 combined with a degree of inbreeding, the reduced genetic diversity available
384 ultimately resulted in the extinction of the lineage, in a process reminiscent of what
385 was previously described for the woolly mammoth (Palkopoulou et al., 2015).
386 In conclusion, our study clarifies the phylogenetic placement, speciation timing and
387 evolutionary history of the now-extinct *Equus Sussemionus* equine subgenus. This
388 group did not remain in reproductive isolation from other equine lineages, but
389 contributed to the genetic makeup of the ancestors of present-day Asiatic asses, while
390 receiving genetic material from the ancestors of African zebras. This supports
391 geographic distributions at least partly overlapping at the time, thus, not identical to
392 those observed today. The species demographic trajectory experienced a steady
393 decline from ~74 kya and during a period witnessing both important climatic changes
394 and the Great human expansion across Asia (Henn et al., 2012). It survived with
395 minimal genetic diversity the Pleistocene-Holocene transition, and for at least eight
396 millennia before it became extinct, which providing insights into the extinction of
397 large animals since Holocene.
398

Additional information

Acknowledgments

We thank High-Performance Computing (HPC) of Northwest A&F University (NWAFU) for providing computing resources.

Funding

Funder	Grant reference number	Author
Major Projects of the National Social Science Foundation of China	17ZDA221	Dawei Cai
European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme	681605	Ludovic Orlando
National Natural Science Foundation of China	31822052	Yu Jiang

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Additional files

Supplementary files

• Supplementary file 1. Supplementary tables that support the analysis and results above.

• Transparent reporting form

Materials and Methods

Genome Sequencing

Prior to DNA extraction, the outer surface of the sample was cleaned with a brush. The cleaned sample was subsequently cut into smaller pieces and soaked in 10% bleach for 20 min, rinsed with ethanol and distilled water, and then UV-irradiated for 30 min on each side. Finally, powder was obtained by drilling with a dental drill (Traus 204，Korea). Ancient DNA was extracted from the sample powder by using a modified silica spin column method (Yang, Eng, Waye, Dudar, & Saunders, 1998) in dedicated ancient DNA facilities at Jilin University (JLU). For each specimen, a total of 200mg powder was added with 3.9ml EDTA (0.465mol/L) and placed in the refrigerator at 4 °C for 12 hours for decalcification, and then 0.1 ml Proteinase K (0.4mg/mL) were added and incubated overnight in a rotating hybridization oven at 50℃ (220rpm/min). After centrifugation, the supernatant was transferred into an Amicon® Ultra-4 centrifugal filter device (Merck Millipore Ltd, 10000 Nominal Molecular Weight Limit), reduced to less than 100ul, and purified with QIAquick® PCR Purification Kit (QIAGEN), according to the manual instructions.

Before preparation of DNA libraries, we first PCR targeted short fragments of the mitochondrial hypervariable region to select those samples positive for the presence

of equine DNA (which was further confirmed through Sanger sequencing). For this, we used the oligonucleotide primers L15473 5' –CTTCCCCTAAACGACAACAA–3' and reverse primer H15692 5' –TTTGACTTGGATGGGGTATG–3' ; and forward primer L15571 5' –AATGGCCTATGTACGTCGTG–3' and reverse primer H15772 5' –GGGAGGGTTGCTGATTTC–3'from (Dawei et al., 2007), and the amplification conditions therein.

Double-stranded single-indexed libraries were prepared using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (NEB #E7645S) and NEBNext® Multiplex Oligos for Illumina® Index Primers Set 1 and 2 (NEB #E7335S, #E7500S), following the manufacturer's instructions with minor modifications. Specifically, the extracted DNA (50ul) were end repaired and A-tailed by adding 7µl of NEBNext Ultra II End Prep Reaction Buffer and 3 µl of NEBNext Ultra II End Prep Enzyme Mix, and incubated for 40 min at 20 ℃ and then 30 minutes at 65 ℃. The adaptor was ligated to the dA-tailed DNA fragments by adding 30ul of NEBNext Ultra II Ligation Master Mix, 1ul of NEBNext Ligation Enhancer and 2.5ul of NEBNext Adaptor for Illumina (Dilution 1:10), and incubated for 20 min at 20 ℃. The adaptor was then linearized by adding 3ul of USER™ Enzyme and performing an incubation for 15 min at 37°C. The adaptor-ligated DNA were cleaned without size selection using the MinElute® PCR Purification Kit (QIAGEN, Germany), following the instructions provided by the manufacturer. PCR enrichment was performed by using 30ul of NEBNext Ultra II Q5 Master Mix, 1ul of Index Primer, 1ul of Universal PCR Primer and 18ul of adaptor-ligated DNA. PCR cycling conditions comprised an initial denaturation at

98 ℃ for 30s, 14-16 cycles of 98 ℃ for 10s, 65 ℃ for 75s, and a final extension at 65 ℃ for 5min. PCR amplified DNA libraries were purified using Agencourt AMPure XP Beads, following the manufacturer's instructions, and Illumina sequencing was performed on HiSeq X Ten platform using 150bp paired-end reads. Overall, we sequenced a total of 28 DNA libraries and generated 2,727,843,803 read pairs.

All pre-PCR procedures were conducted in a dedicated ancient DNA laboratory at JLU that is physically separated from the post-PCR laboratory. To remove potential contaminant DNA, working areas and benches were frequently cleaned with bleach and UV exposure. Lab experiments were carried out wearing full body suits, facemasks, and gloves. To detect contamination, mock blank controls were included in each experimental step, including DNA extraction, DNA library preparation and PCR setup.

Data Processing

Sequencing reads were processed and aligned against the horse (EquCab3.0 (Kalbfleisch et al., 2018)) and donkey (Renaud et al., 2018) reference genomes using the PALEOMIX pipeline (Schubert, Ermini, et al., 2014) with default parameters, except that we followed the recommendations from (Schubert et al., 2012) and disabled seeding. Briefly, paired-end (PE) reads longer than 25 nucleotides were trimmed with AdapterRemoval v2.2 (Schubert, Lindgreen, & Orlando, 2016) and aligned against the reference genomes using BWA (Li & Durbin, 2009), retaining alignments with mapping qualities superior to 25. PCR duplicates were then removed

474 using Picard MarkDuplicates (http://broadinstitute.github.io/picard/). Finally, all
475 ancient and modern reads were locally realigned around indels using GATK
476 (McKenna et al., 2010).

477 Postmortem DNA damage and average sequencing error rates were determined
478 with mapDamage2.0 (Jonsson, Ginolhac, Schubert, Johnson, & Orlando, 2013)
479 (Figure 2—figure supplement 1) and ANGSD (Korneliussen, Albrechtsen, & Nielsen,
480 2014) (Figure 2—figure supplement 2), respectively. Further rescaling and trimming
481 procedures were implemented following (Gaunitz et al., 2018) to limit the impact of
482 remnant nucleotide mis-incorporations in subsequent analyses. For each of the DNA
483 libraries examined, the base composition of the position preceding read starts on the
484 horse reference genome showed an excess of Guanine and, to a lesser extent, of
485 Adenine residues (Figure 2—figure supplement 1). This is in line with depurination
486 driving post-mortem DNA fragmentation (Briggs et al., 2007). Additionally, error rate
487 estimates for each nucleotide substitution class indicated the predominance of C→T
488 and G→A mis-incorporations (Figure 2—figure supplement 2). Such
489 mis-incorporation rates were particularly inflated towards read ends, but not read
490 starts (Figure 2—figure supplement 1). This is in line with the DNA nucleotide
491 mis-incorporation profiles expected for the type of DNA library constructed
492 (Seguin-Orlando et al., 2015), and indicates Cytosine deamination at 5'-overhanging
493 ends as the most prominent post-mortem DNA degradation reactions (Jonsson et al.,
494 2013).

GATK HaplotypeCaller was used to obtain individual gvcf files with "--minPruning 1 --minDanglingBranchLength 1" to increase sensitivity. Then we employed GATK GenotypeGVCFs for genotyping with the option "--includeNonVariantSites" in order to retain non-variant loci. The vcf files were further filtered in TreeMix and G-PhoCS analysis.

Principal component analysis (PCA)

The genotype likelihood framework implemented in ANGSD helped mitigate various error rates in ancient and modern genomes. Using EquCab3 (Kalbfleisch et al., 2018) as the reference genome, ANGSD was run using the following options: "-only_proper_pairs 1 -uniqueOnly 1 -remove_bads 1 -minQ 20 -minMapQ 25 -C 50 -baq 1 -skipTriallelic 1 -GL 2 -SNP_pval 1e-6 -rmTrans 1". This provided a dataset consisting of a total of 16,293,825 transversions when the horse was included, and 10,094,431 transversions when the horse was excluded (i.e. when analyses were restricted to non-caballine genomes only). In these analyses, only specimens sequenced to an average depth of coverage ≥ 1× were retained. Principal Component Analyses were carried out using the PCAngsd package (Meisner & Albrechtsen, 2018) (Figure 2A). To assess the impact of potential reference bias, all analyses were repeated after mapping the sequence data against the donkey reference (Figure 2—figure supplement 4).

Phylogenetic inference

Mitochondrial phylogeny

Cleaned reads were mapped against the mitochondrial genome (Genbank accession no. NC_001640), following the same procedure as when mapping against the nuclear genome. Samples showing an average depth-of-coverage <1× were disregarded, leaving a total of 17 individuals for further analyses. After removing duplicates, consensus mitochondrial sequences were generated using ANGSD (-doFasta 2 -doCounts 1 -setMinDepth 3 -uniqueOnly 1 -remove_bads 1 -minQ 25 -minMapQ 25). Multiple alignment was performed together with the comparative mtDNA sequences downloaded from GenBank (Table S4) using MUSCLE v3.8.31 (Edgar, 2004), with default parameters. The alignments were then split into six partitions (1^{st}, 2^{nd} and 3^{rd} codon positions, rRNA, tRNA, and control region).

Two Maximum Likelihood (ML) trees based on all 6 partitions and excluding the control region (positions 15,469-16,660 of the horse reference mitochondrial genome) were both reconstructed using RAxML-NG v.0.9.0 (A. M. Kozlov, Darriba, Flouri, Morel, & Stamatakis, 2019) with GTR+GAMMA substitution model. A total of 1,000 bootstrap pseudo-replicates were carried out to assess node robustness (Figure 2—figure supplement 5). BEAST 2.5.1.0 (Bouckaert et al., 2019) was used to perform Bayesian phylogenetic reconstruction and to estimate split times. The six partitions described above were used, for which the best substitution model was determined using modelgenerator (version 0.85, (Keane, Creevey, Pentony, Naughton, & McLnerney, 2006)) and a Bayesian Information Criterion. We applied together with the Coalescent Constant Population model and a strict LogNormal correlated

539 molecular clock for 50 million generations (sampling frequency = 1 every 1,000).
540 Convergence was assessed visually using Tracer v1.6 and posterior date estimates
541 were retrieved using 10% as burn-in. The final consensus tree was produced using
542 TreeAnnotator 2.5.1.0 (Drummond & Rambaut, 2007) and plotted using ITOL
543 (Letunic & Bork, 2016) (Figure 2—figure supplement 6).
544
545 **Autosomal phylogeny**
546 As for autosomes, we reconstructed a (ML) phylogenetic tree as implemented
547 in the PALEOMIX phylo pipeline for phylogenomic reconstructions (Schubert,
548 Ermini, et al., 2014). This analysis was based on the coding sequence (CDS) of
549 protein-coding genes annotated in EquCab3.0, partitioning data according to 1st, 2nd
550 and 3rd codon positions. Maximum Likelihood phylogenetic inference was performed
551 using ExaML v3.0.21 (Alexey M. Kozlov, Aberer, & Stamatakis, 2015) and RAxML
552 v8.2.12 (Stamatakis, 2014), under the GAMMA substitution model with 100 bootstrap
553 pseudo-replicates (Figure 2C and Figure 2—figure supplement 7A). We also repeated
554 the same procedure after mapping against the donkey reference genome and got the
555 same topology (Figure 2—figure supplement 7B).
556 Additionally, we extracted biallelic single nucleotide polymorphisms (SNPs) from
557 the dataset generated in Section 5 using bcftools v1.9 (Li et al., 2009). Both variant
558 datasets obtained following mapping against the horse and donkey reference genomes
559 were used in this analysis to rule out reference bias. We applied filters composed of
560 minimum phred-scaled quality score quality (QUAL) = 20, sites for all individuals

below 2 or twice the mean coverage, and allowed up to three individuals with missing data per site. After disregarding transitions, a total of 18,803,101 (mapping against horse genome) and 19,459,070 (mapping against donkey genome) transversions were finally used as input for TreeMix (Pickrell & Pritchard, 2012) with parameters "-k 500 -root TWI", and considering an increasing number of migrations edges ($0 \leq m \leq 3$; Figure 2—figure supplements 8 and 9, Table S5).

Admixture analyses with *D*-statistics

D-statistics were calculated to investigate potential introgression between *E. ovodovi* and other non-caballines (Figure 3—figure supplement 1) using the doAbbababa2 programme in ANGSD (Soraggi et al., 2018). Individuals were grouped by their respective species. D-statistics were computed in the form (((H1, H2), H3), Outgroup) considering only the autosomal sites from bam files mapping against the horse reference with the following options: "-minQ 20 -minMapQ 25 -remove_bads 1 -only_proper_pairs 0 -uniqueOnly 1 -baq 1 -C 50". The horse reference genome was used as the Outgroup. H1 and H2 denoted any non-caballine genomes except *E. ovodovi* while H3 denoted the *E. ovodovi*. Confidence intervals were estimated applying a jackknife procedure and 5-Mb windows. Z-scores with absolute values higher than 3 were considered to be statistically significant. To eliminate the bias of the reference genome, we also rerun the same analysis using sequence alignments against the donkey reference genome (Figure 3—figure supplement 2).

G-PhoCS demographic model

Data preparation and filtering

In order to model the equine evolutionary history, we selected a total of 10 individuals representing each individual lineage and used their high-coverage genomes as input for G-PhoCS (Gronau et al., 2011). Genotypes were called by GATK and candidate 'neutral' loci were identified by applying the following filters:

(1) The simple repeats track available for the reference genome was obtained from Ensembl v99 release; corresponding regions were masked.

(2) All exons of protein-coding genes were discarded together with their 10 kb flanking regions; this was done based on the GTF format annotation file of the reference genome available from Ensembl v99 Genome Browser.

(3) We identified conserved noncoding elements (CNEs) using phastCons scores (based on the 20-way Conservation track provided for the mammal clade according to the genomic coordinates of the human reference) downloaded from the Table Browser of UCSC. All CNEs and their 100 bp flanking regions were masked, using liftOver to convert human genome coordinates into EquCab3.0 horse genome coordinates.

(4) Exons of noncoding RNA genes together with their one kilobase flanking regions were removed, based on the annotations available for the reference genome.

(5) Gaps in the reference genome were disregarded.

Besides the various hard filters described above, regions/sites likely to (1) be enriched for misaligned bases, and to (2) have high false negative rates during read alignment or variant detection were masked as missing data. Different individuals

may be treated differently depending on the result of genotyping in Section 5 depending on the presence of (1) indels, (2) triallelic sites, (3) positions with depth of coverage twice the mean depth recorded for each individual, and; (4) transition sites.

We selected 1 kb loci located with minimum inter-locus distance of 30 kb from the intervals that pass all the criteria described above. Then consensus sequences were generated for each individual from the vcf file generated in Section 5 using bcftools 'consensus' command, with IUPAC codes indicating heterozygous genotypes (--iupac-codes) and "N" representing masked sites (--mask and --missing 'N').

Finally, we excluded contiguous intervals if the total amount of missing bases was greater than 50% of the region length, resulting in a final collection of 15,324 loci using the horse reference genome (autosomes only). Neighbor-joining trees were constructed to confirm the topology before the inferring the population divergence (Figure 3—figure supplement 3).

MCMC setup

We used default global settings (Gronau et al., 2011), including a Gamma prior distribution (α= 1, β= 10,000) for all mutation-scaled population sizes (θ) and a Gamma prior distribution (α= 0.002, β= 0.00001) for all mutation-scaled migration rate (m). The initial parameter value of mutation-scaled divergence times (τ) was first set individually for each population. Then we ran ~100,000-200,000 iteration tests and manually evaluated the convergence by checking the achieve acceptance ratios (*ie*. accept if around 30-70%) or using Tracer v1.6

(http://tree.bio.ed.ac.uk/software/tracer/). For each test, we updated the input of the initial τ and all fine-tuned parameters based on previous results to get the appropriate value. The final results in Figure 3 are based on 500,000 MCMC iterations.

Parameter calibration

We assumed an average generation time (g) of 8 years, considering the mutation rate μ (per year) could be variable when using different sequences. The coalescent time of the *Equus* (4.0-4.5 Mya) (L. Orlando et al., 2013) was used to bound the parameter μ. Effective population sizes (Ne) and divergence times (T) were estimated by calibrating θ and τ parameter using g and μ (Table S6), given by: $Ne = \theta/(4\mu g)$ and $T = \tau/\mu$ (Gronau et al., 2011).

Inferring gene flow

Total migration rates (M) were estimated by a mutation-scaled version (m) given by: $M = m\tau_m$, where τ_m is the mutation-scaled time span of the migration band. We then converted such rates, M , into a probability of migration using the formula: $p = 1\text{-}e^{-M}$ (where p is the probability of gene flow), according to the method presented in (vonHoldt et al., 2016).

The migration model implemented in G-PhoCS makes it possible to detect gene flow between any two lineages by introducing migration bands manually to the demographic model. However, it remains difficult to detect weak migration events. Additionally, scenarios including a large number of migration bands can lead to

649 spurious results. To address this, we first inferred a demographic model with no
650 migration bands, and then introduced several migration bands corresponding to five
651 independent scenarios (Table S6). A significant migration band was considered
652 supported if both the 95% Bayesian credible interval of total migration rate (*M*) did
653 not include 0 and the mean value of *M* was estimated to be greater than 0.03.
654 Settings for the migration bands between extant caballines are based on previous
655 research (Jonsson et al., 2014). The significant migration band from horse to the
656 non-caballine ancestor were identified (Table S6), in line with previous work (Jonsson
657 et al., 2014). However, no other non-negligible ($M > 3\%$) migration bands was found
658 in our analyses (Table S6).
659 We then tried to estimate the migration events between *E. ovodovi* and other
660 branches. We added all possible migration bands between *E. ovodovi* and
661 non-caballine branches into the demographic model except the migration bands
662 between *E. ovodovi* and the ancestor of non-caballines, as the model is often
663 underpowered to infer migration between sister populations. All of the migration
664 bands were separated into four demographic models. Only three migration bands were
665 shown significant (Table S6).
666 Finally, the total four migration bands were combined into one demographic model
667 (Table S7) and compared the estimates to the one including no migration (Table S8).
668

669 # Demographic trajectories with PSMC

670 ## PSMC analyses

In order to reconstruct the past demographic dynamics of the *E. ovodovi* lineage, we applied the PSMC algorithm (version 0.6.5-r67) (Li & Durbin, 2011) to the sample HH06D (12.0×) , as well as three other Eurasian equine species (*E. caballus* TWI, *E. hemionus* ONA and *E. kiang* KIA).

We first obtained the diploid consensus sequences after mapping against the horse genome for the autosomes of each specimens using bcftools 'mpileup' command and the 'vcf2fq' command from vcfutils.pl with the following filters: mapping quality ≥ 25; adjust mapping quality =50; minimum depth-of-coverage = 8; maximum depth-of-coverage ≤ 99.5% quantile of the coverage distribution; minimum RMS mapping quality = 10; filtering window size of indels = 5.

After filtering the bases with Phred quality scores strictly lower than 35, we ran PSMC with following command: 'psmc -N25 -t15 -r5 -p "4+25*2+4+6" '. Calibration was carried out using a generation of 8 years and mutation rate of 7.242×10^{-9} per generation per site, following previous work (Jonsson et al., 2014). However, as for the mis-incorporation pattern and high error rate of HH06D (Figure 2—figure supplements 1 and 2), we also performed analyses without transitions using mutation rates of 2.3728×10^{-9} that was obtained assuming that the most recent common ancestor of living equine species emerged 4 Mya (L. Orlando et al., 2013).

We found a great expansion of HH06D in the past 50,000 years when retaining transitions but not when conditioning on transversions (Figure 4—figure supplement 1). The former is thus likely spurious and at least partly driven by severe post-mortem

DNA damage signatures in the sequence data. We therefore only used the latter when considering the ancient HH06D specimen.

False negative rate correction

The HH06D genome (12.0×) was corrected assuming a uniform false-negative rate (uFNR) following (L. Orlando et al., 2013), as the average depth-of-coverage is lower than the recommended 20×. To identify the correction value of uFNR for HH06D, we randomly down-sampled reads of SOM genome (21.0×) using DownsampleSam function of Picard Tools to down-scale sequence data to the same average depth-of-coverage as that obtained for HH06D. This indicated that a value of 0.22 was the most suitable uFNR value for rescaling the HH06D PSMC profile (Figure 4—figure supplement 2A). The KIA and the ONA genomes, which also showed limited coverage, were also rescaled following the same procedure (Figure 4—figure supplement 2B-C). Finally, PSMC confidence intervals were assessed from 100 bootstrap pseudo-replicates (Figure 4).

Heterozygosity Inference and Inbreeding

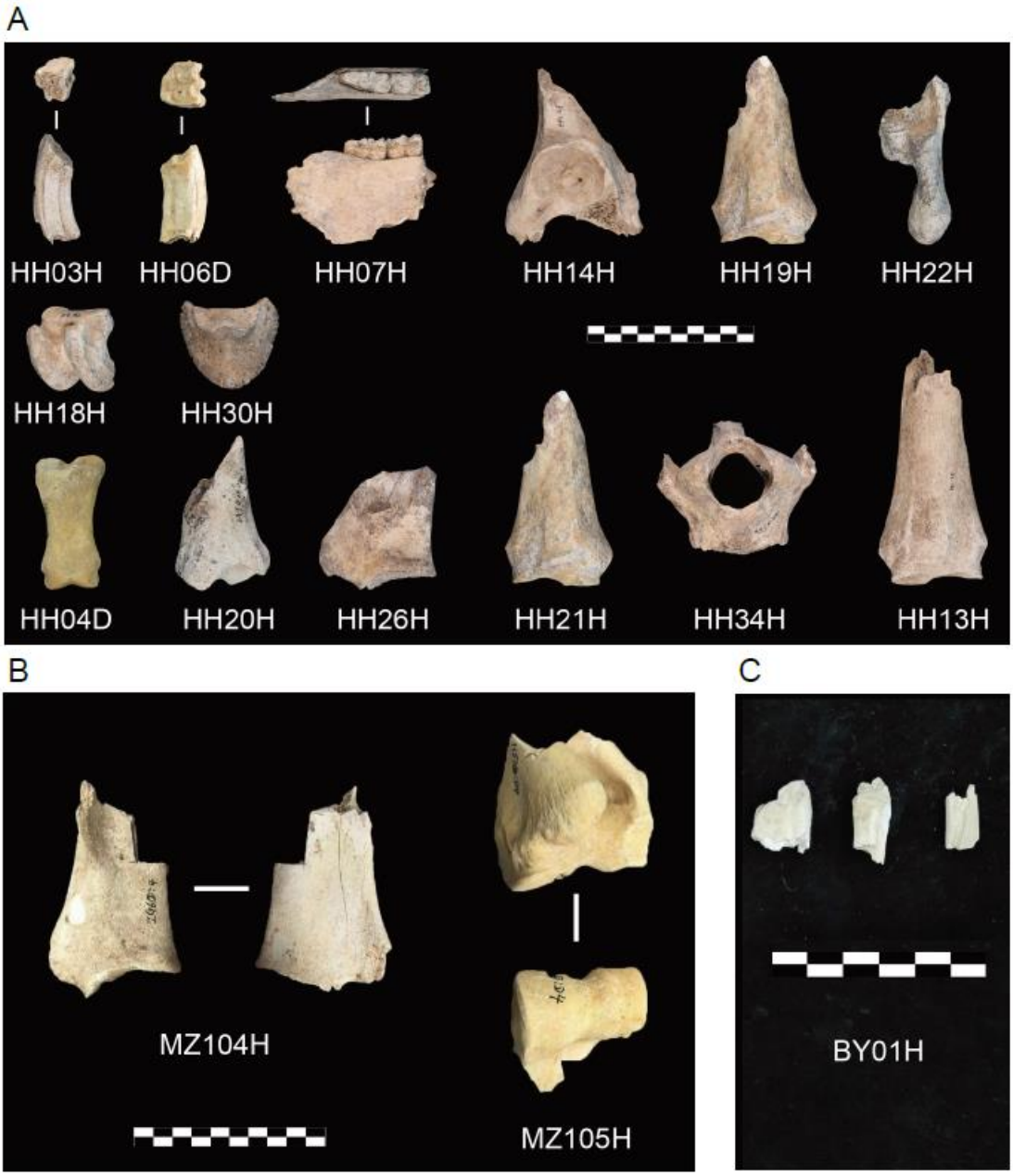
Global heterozygosity rates and inbreeding levels were inferred for high coverage individuals (>10×) using ROHan (Renaud, Hanghoj, Korneliussen, Willerslev, & Orlando, 2019) with default parameters, except that transitions were excluded (--tvonly) (Figure 5—figure supplement 1). To limit the impact of remnant mis-incorporations, we used the attached estimateDamage.pl script to estimate

714 damage for all ancient samples prior to heterozygosity computation. Inbreeding was

715 co-estimated together with genome-wide heterozygosity levels from the total ROH

716 length (Figure 5—figure supplement 2).

717



718

719 **Figure 1—figure supplement 1.** Archaeological material investigated in this study.

720 (A) Honghe (HH), (B) Muzhuzhuliang (MZ) and (C) Shatangbeiyuan (BY).

721

722

723 **Figure 1—figure supplement 2.** Aerial view of the Honghe site.

724

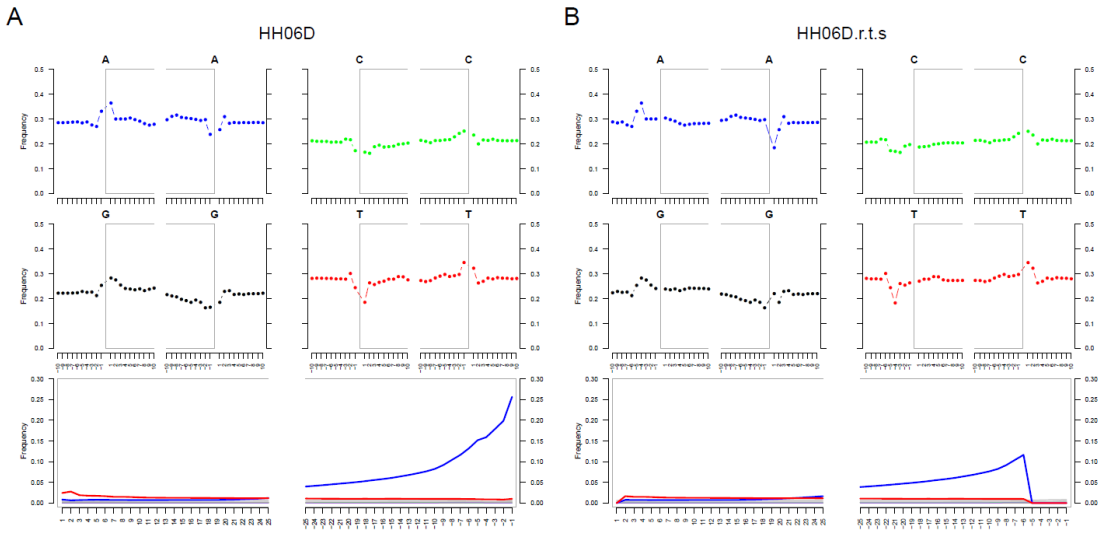


Figure 2—figure supplement 1. DNA Damage patterns for HH06D. (**A**) Before rescaling and trimming and (**B**) after rescaling and trimming the region comprising the five first and last nucleotides sequenced.

730

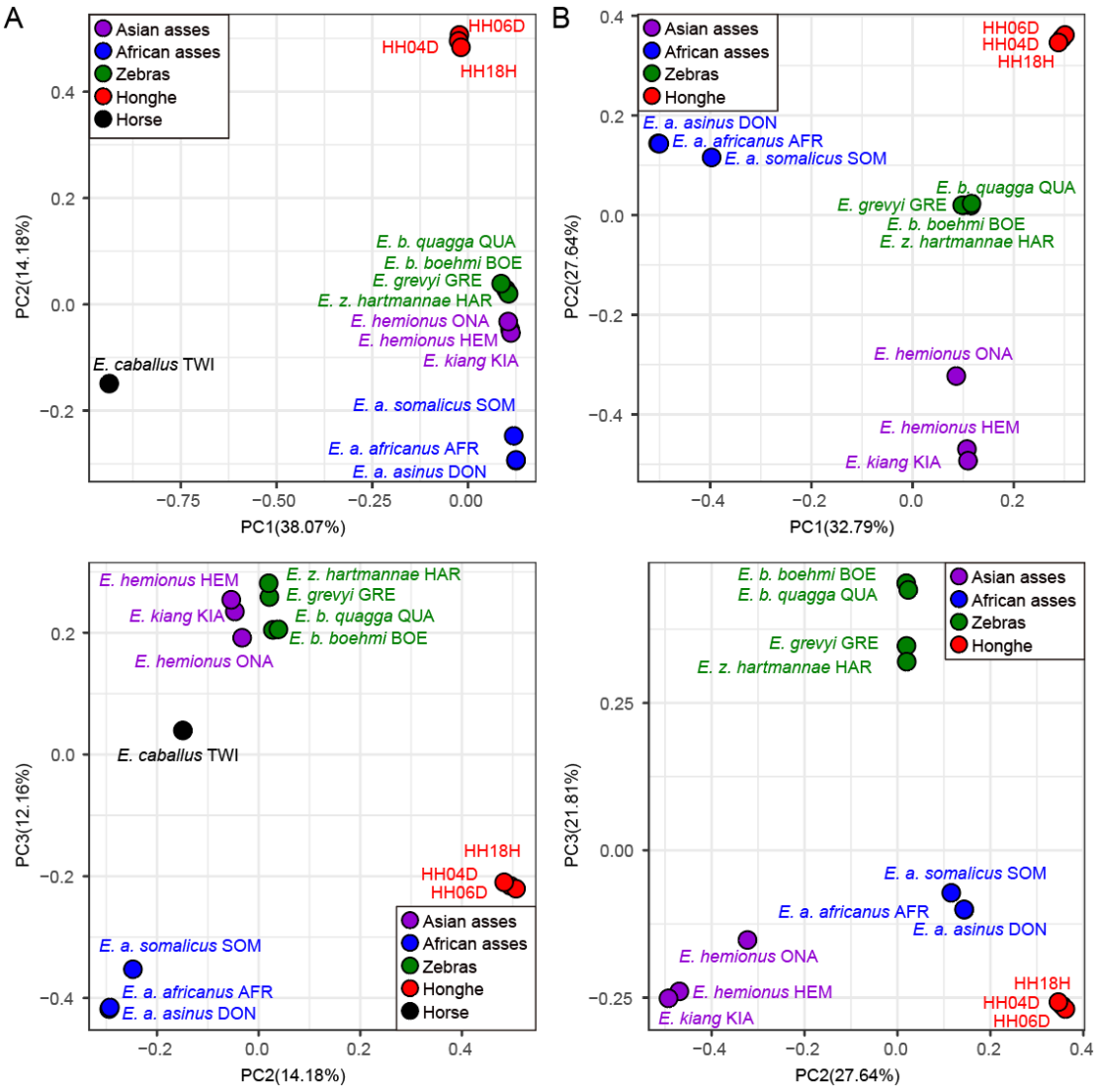
731 **Figure 2—figure supplement 2.** Error profiles of the 26 ancient genomes

732 characterized in this study. After trimming and rescaling, reads showing mapping

733 quality scores inferior to 25 and bases showing quality scores inferior to 20 were

734 disregarded.

735



736

737 **Figure 2—figure supplement 3.** Principal Component Analysis (PCA) based on

738 genotype likelihoods using the horse reference genome. (**A**) Including and (**B**)

739 excluding the outgroup individual underlying the horse reference genome (TWI)

740 (Kalbfleisch et al., 2018). Sequence data were aligned against the horse reference

741 genome (Kalbfleisch et al., 2018).

742

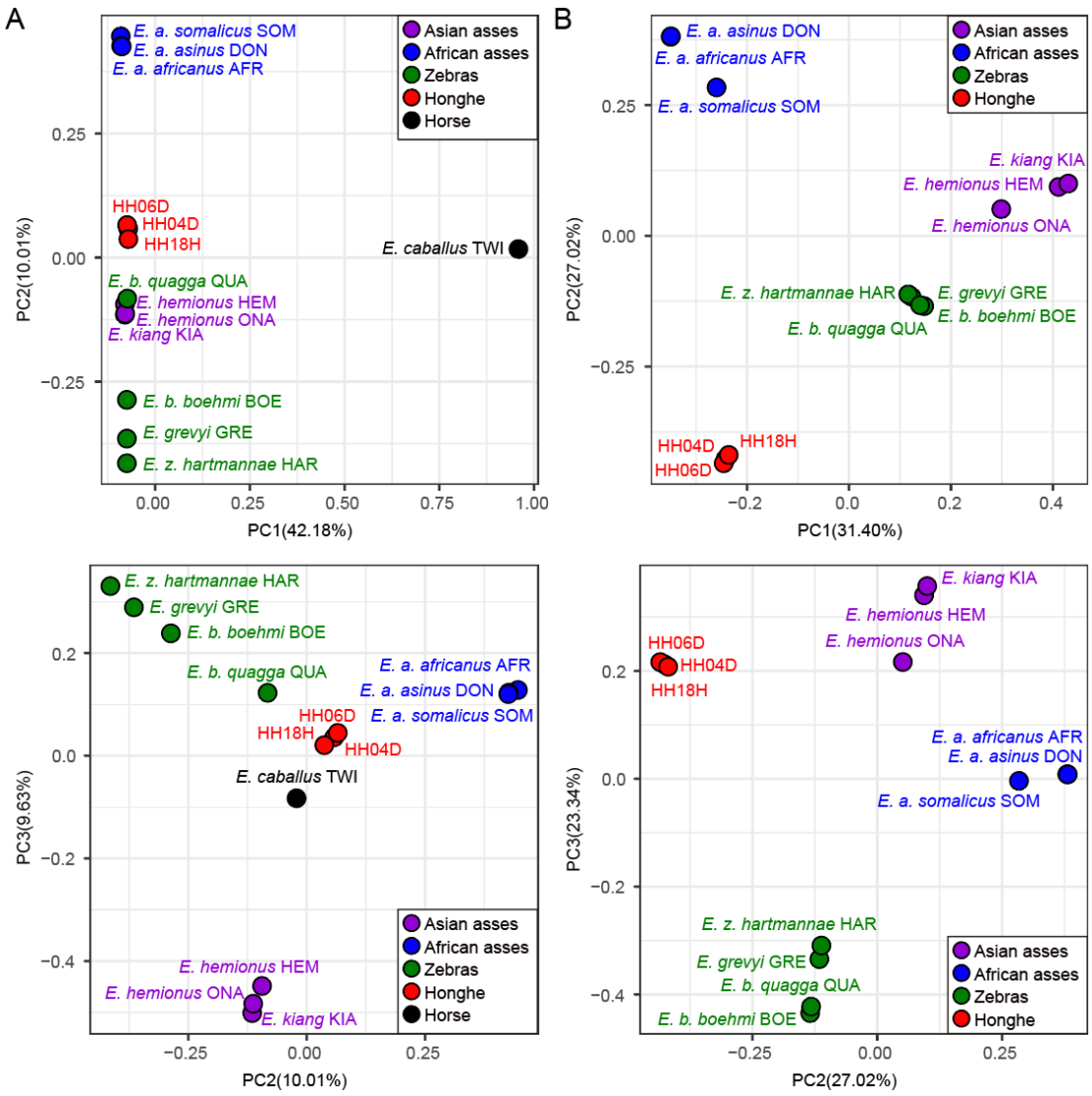


Figure 2—figure supplement 4. Principal Component Analysis (PCA) based on genotype likelihoods using the donkey reference genome. (**A**) Including and (**B**) excluding the outgroup individual underlying the horse reference genome (TWI) (Kalbfleisch et al., 2018). Sequence data were aligned against the donkey reference genome (Renaud et al., 2018).

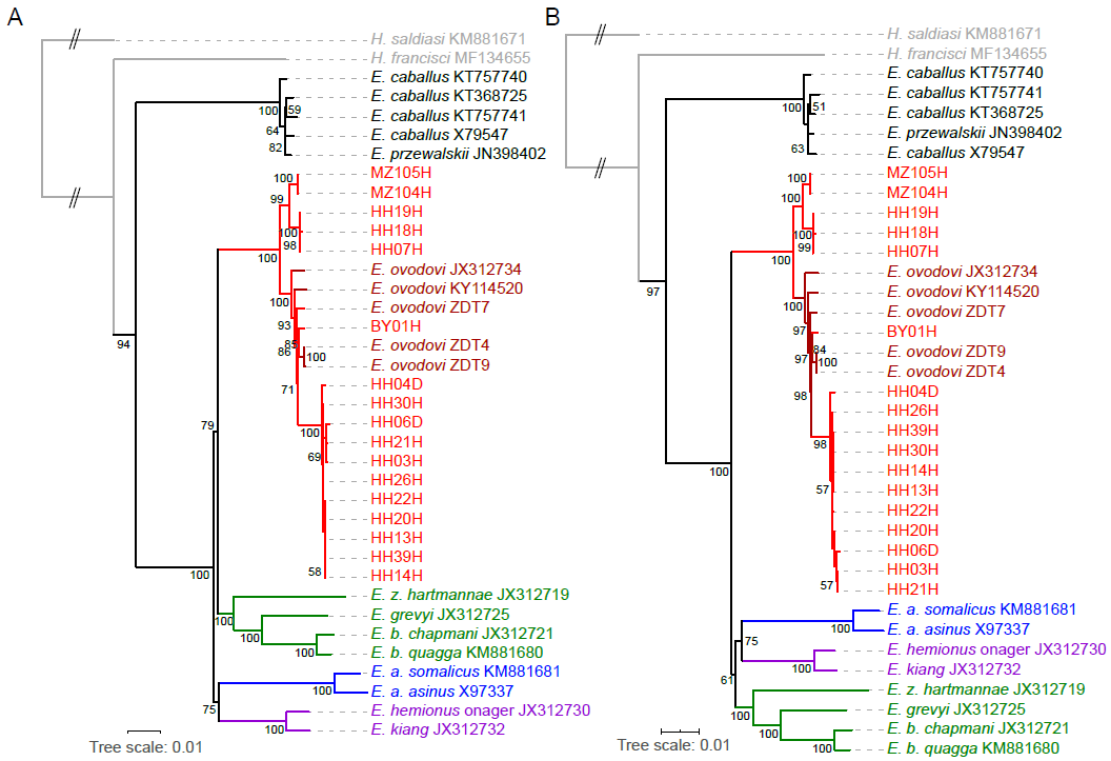


Figure 2—figure supplement 5. RAxML-NG (GTR+GAMMA model) Maximum Likelihood phylogeny of complete mitochondrial sequence data. (**A**) Including the control region. (**B**) Excluding the control region. Node support was estimated from 1,000 bootstrap pseudo-replicates and the tree was manually rooted using *Hippidion Saldiasi*.

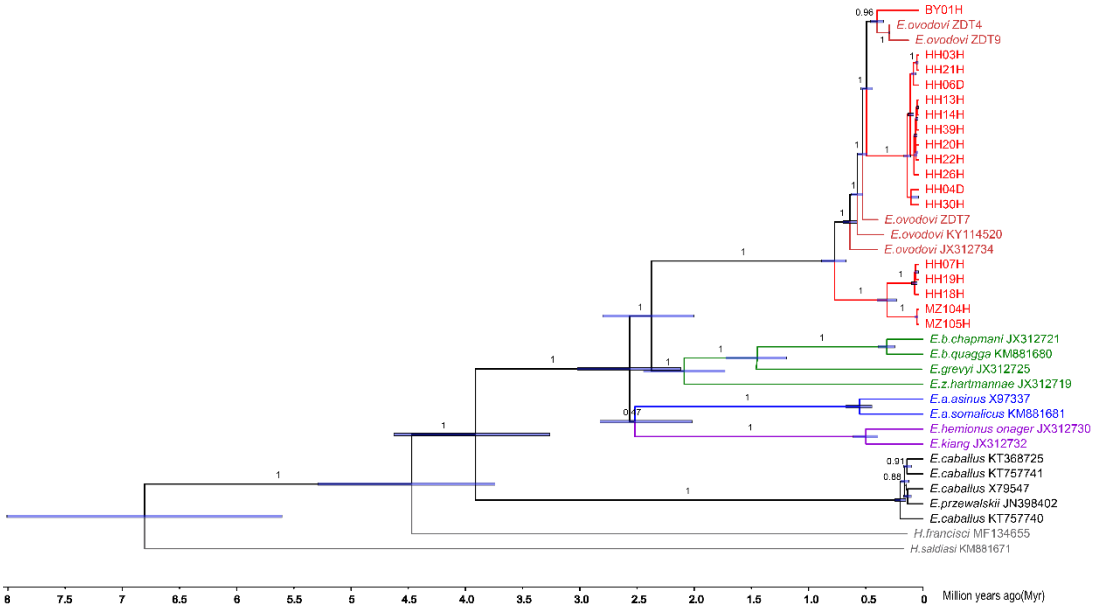


Figure 2—figure supplement 6. Bayesian mitochondrial phylogeny based on 6 partitions and using *Hippidion Saldiasi* as outgroup. The tree was reconstructed using a total number of 50 million MCMC states in BEAST (sampling frequency = 1 every 10,000, burn-in = 10%). The substitution models applied to the six sequence partitions were the TrN+I+G model (1st codon position = 3,802 sites), the TrN+I model (2nd codon position = 3,799 sites), the GTR+I+G model (3rd codon position = 3,799 sites), the HKY+I model (transfer RNAs = 1,517 sites), the TrN+I+G model (ribosomal RNAs = 2,556 sites) and the HKY+I+G model (control region = 1,192 sites).

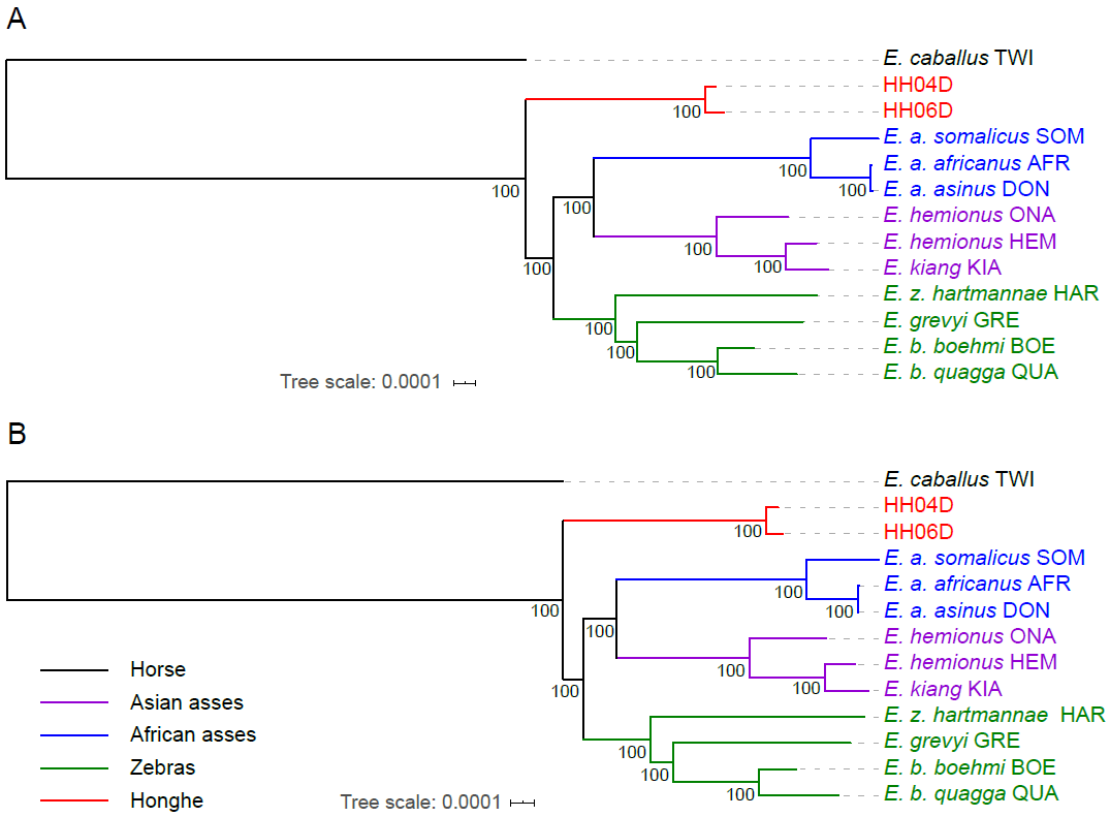


Figure 2—figure supplement 7. Exome-based Maximum likelihood phylogeny rooted by the horse lineage. (**A**) Using sequence alignments against the horse reference genome (Kalbfleisch et al., 2018). (**B**) Using sequence alignments against the donkey reference genome (Renaud et al., 2018). Node supports were estimated from 100 bootstrap pseudo-replicates.

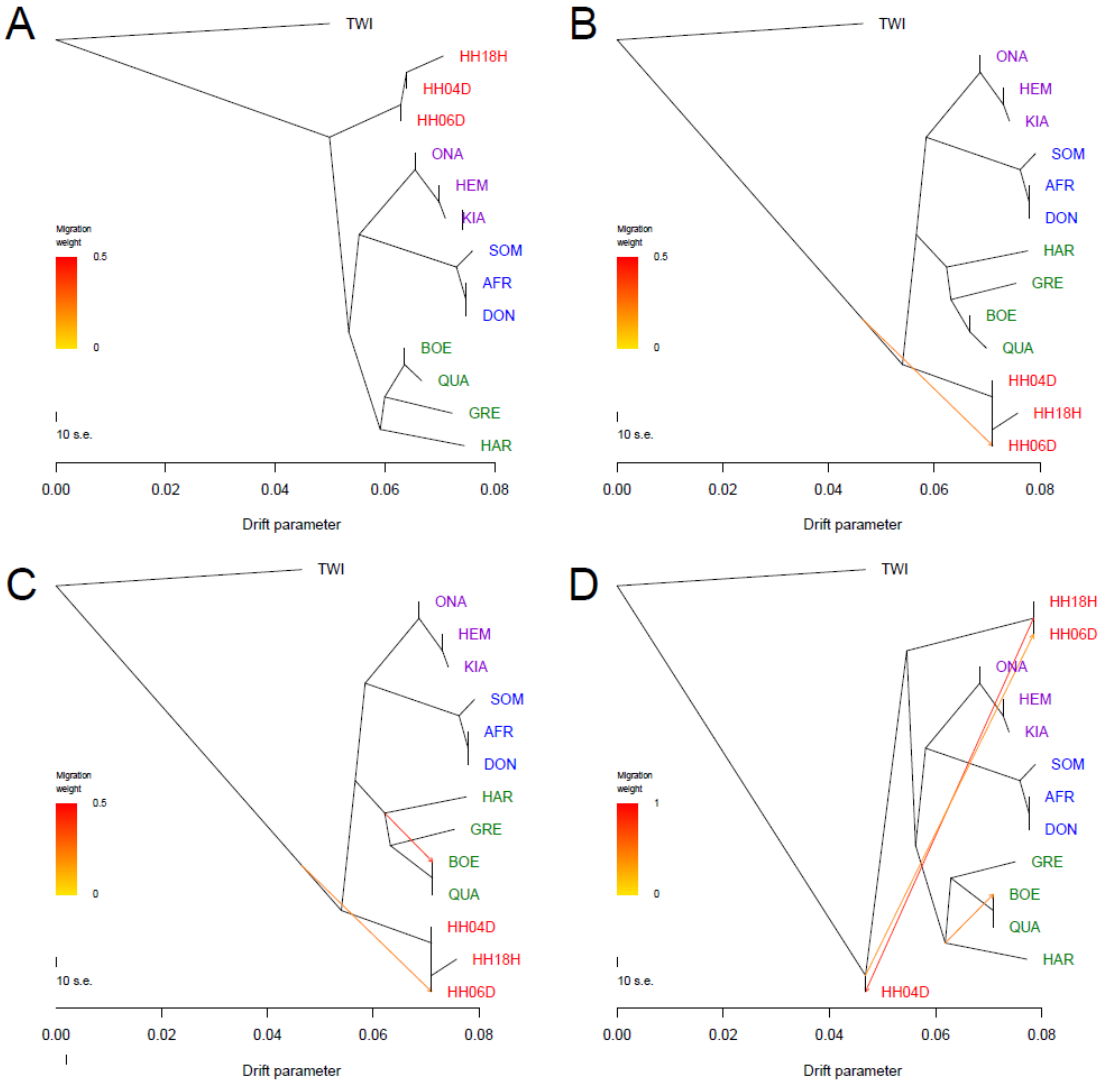


Figure 2—figure supplement 8. Treemix analysis of based on genome-wide SNP data conditioned on transversions using the horse reference genome. Sequence data were mapped against the horse reference genome (Kalbfleisch et al., 2018). A total of 0 to 3 migration edges were considered. The result of each analysis is shown in panels (**A**) to (**D**), respectively. Considering additional migration edges did not improve the variance explained by the TreeMix model (Table S5).

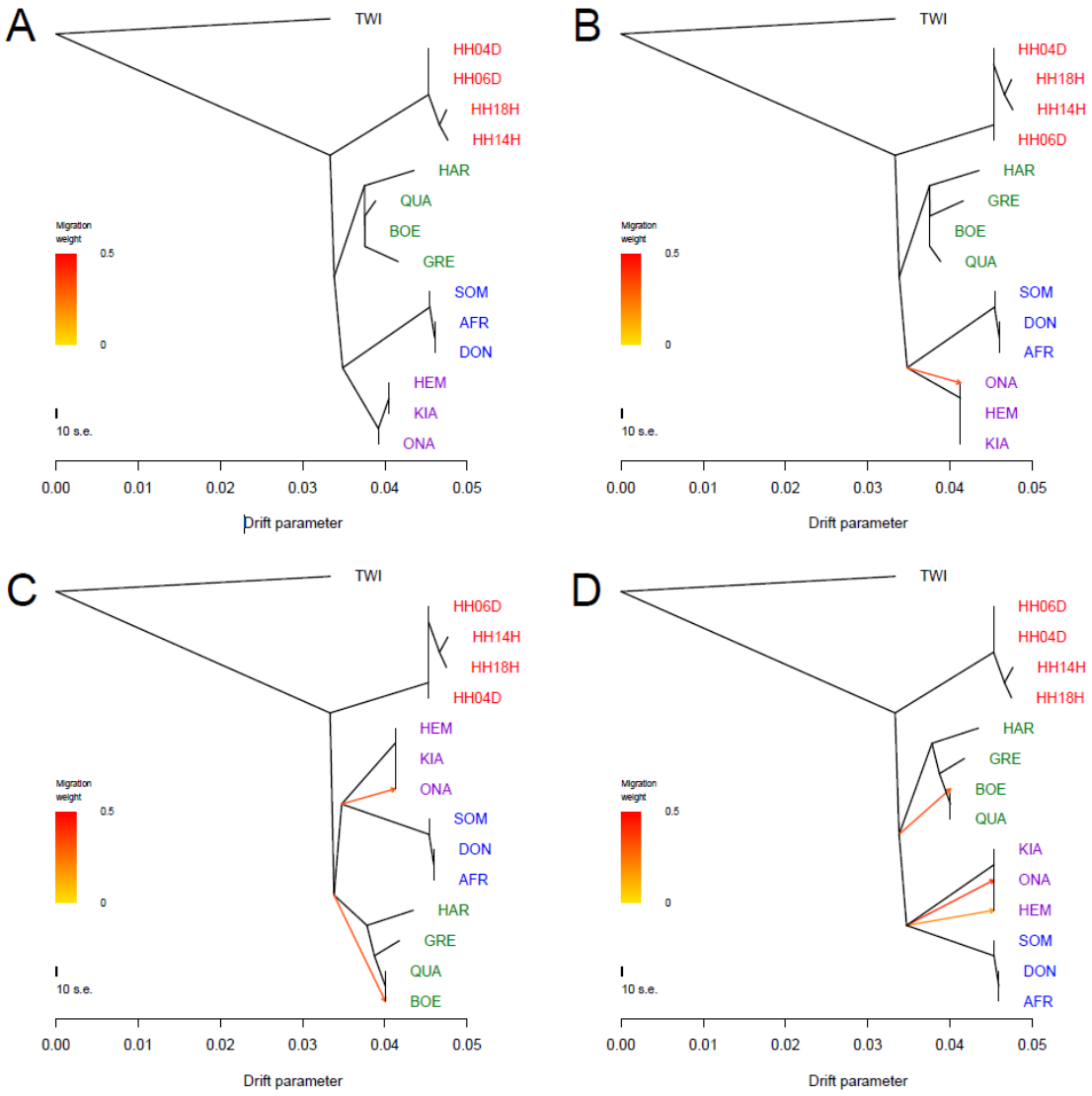


Figure 2—figure supplement 9. Treemix analysis of based on genome-wide SNP data conditioned on transversions using the donkey reference genome. Sequence data were mapped against the donkey reference genome (Renaud et al., 2018). A total of 0 to 3 migration edges were considered. The result of each analysis is shown in panels (**A**) to (**D**), respectively. Considering additional migration edges did not improve the variance explained by the TreeMix model (Table S5).

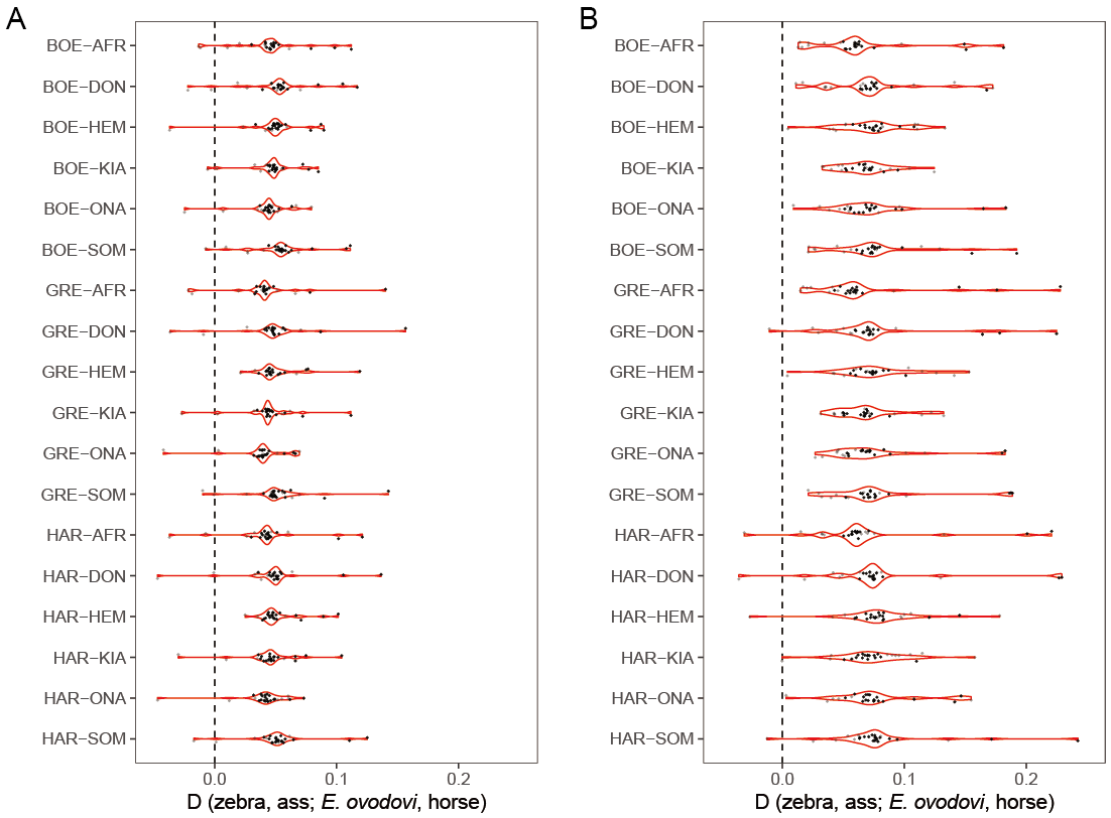


Figure 3—figure supplement 1. *D*-statistics in the form of (zebra, ass; *E. ovodovi*, outgroup), using sequence alignments against the horse reference genome. Significantly positive D-statistics are indicative of an excess of shared derived polymorphisms between *E.* (*Sussemionus*) *ovodovi* and extant assess, which is compatible with admixture between both lineages. (**A**) Including transitions. (**B**) Excluding transitions.

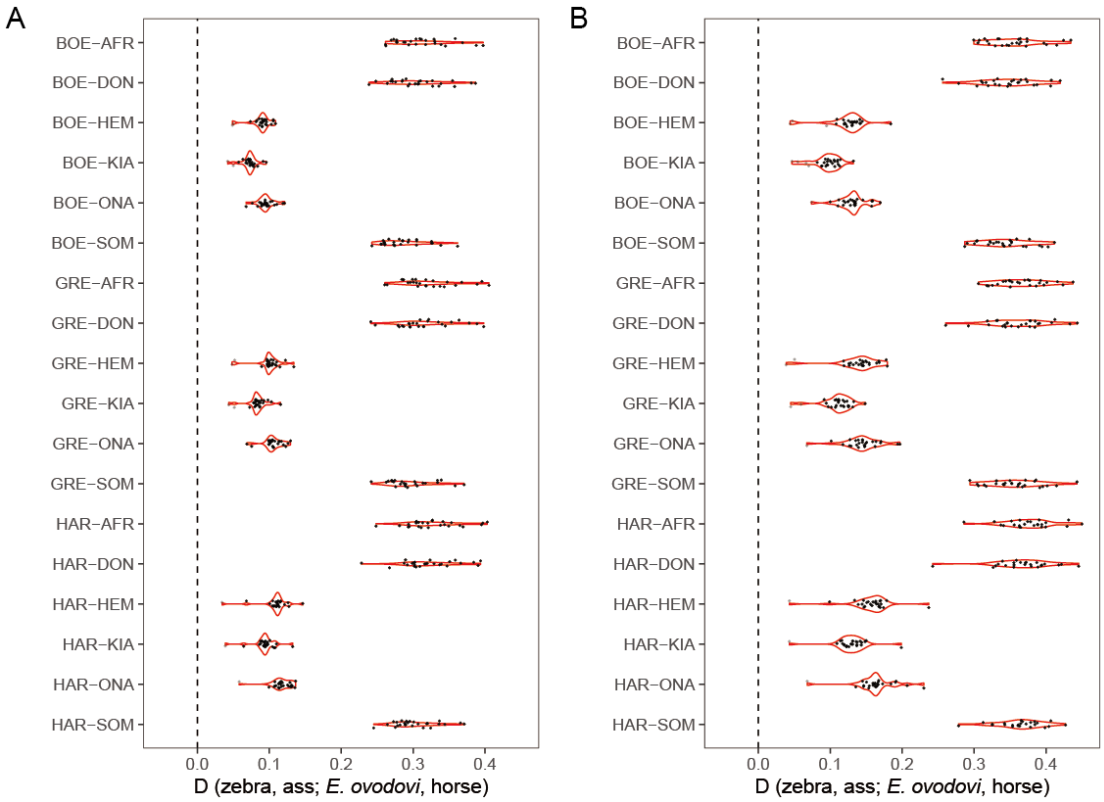
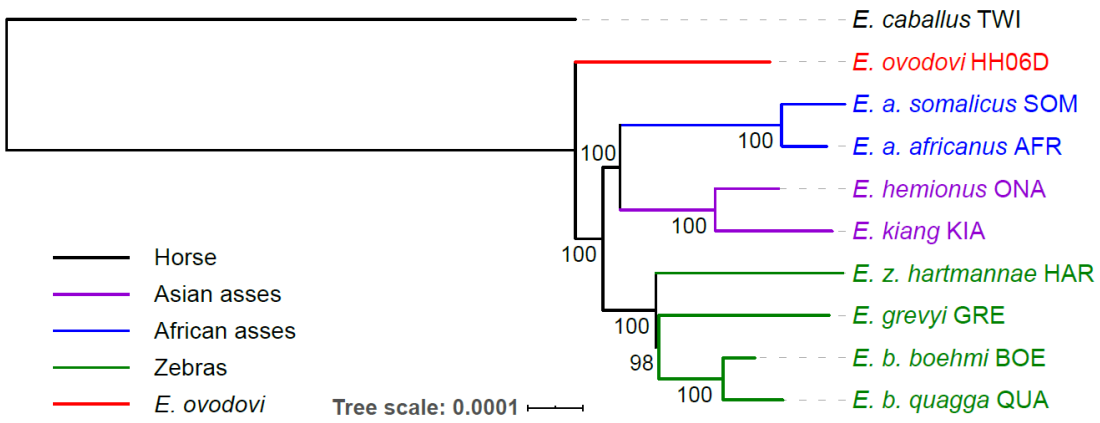


Figure 3—figure supplement 2. *D*-statistics in the form of (zebra, ass; *E. ovodovi*, outgroup), using sequence alignments against the donkey reference genome. Significantly positive D-statistics are indicative of patterns of shared derived polymorphisms between *E.* (*Sussemionus*) *ovodovi* and extant assess, which is compatible with admixture between both lineages. (**A**) Including transitions. (**B**) Excluding transitions.



806

807 **Figure 3—figure supplement 3.** NJ tree of selected samples based on 15,324

808 candidate 'neutral' loci identified using sequence alignments against the horse

809 reference genome (detailed in **Data preparation and filtering**). Node supports were

810 assessed from 1,000 bootstrap pseudo-replicates.

811

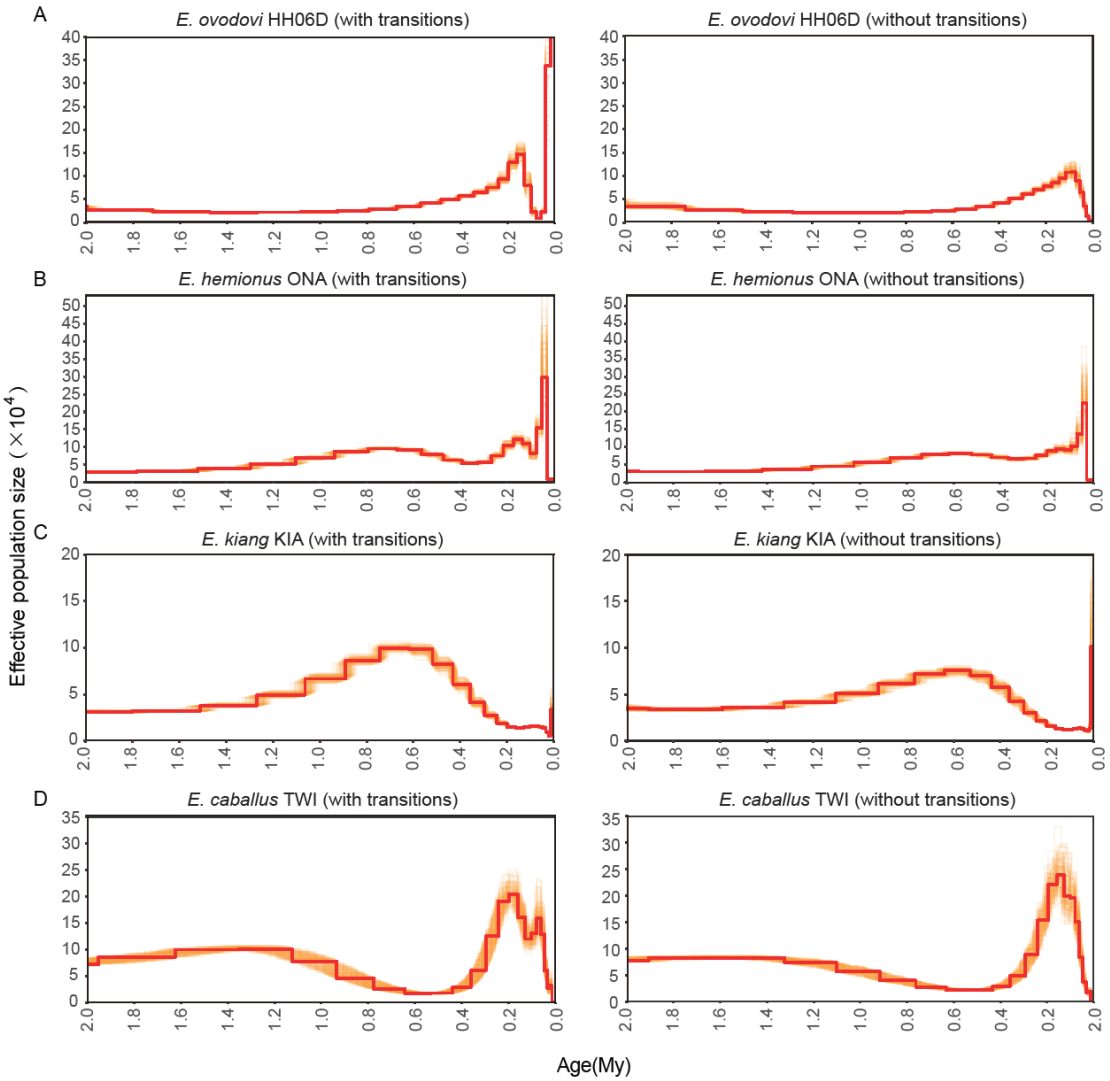
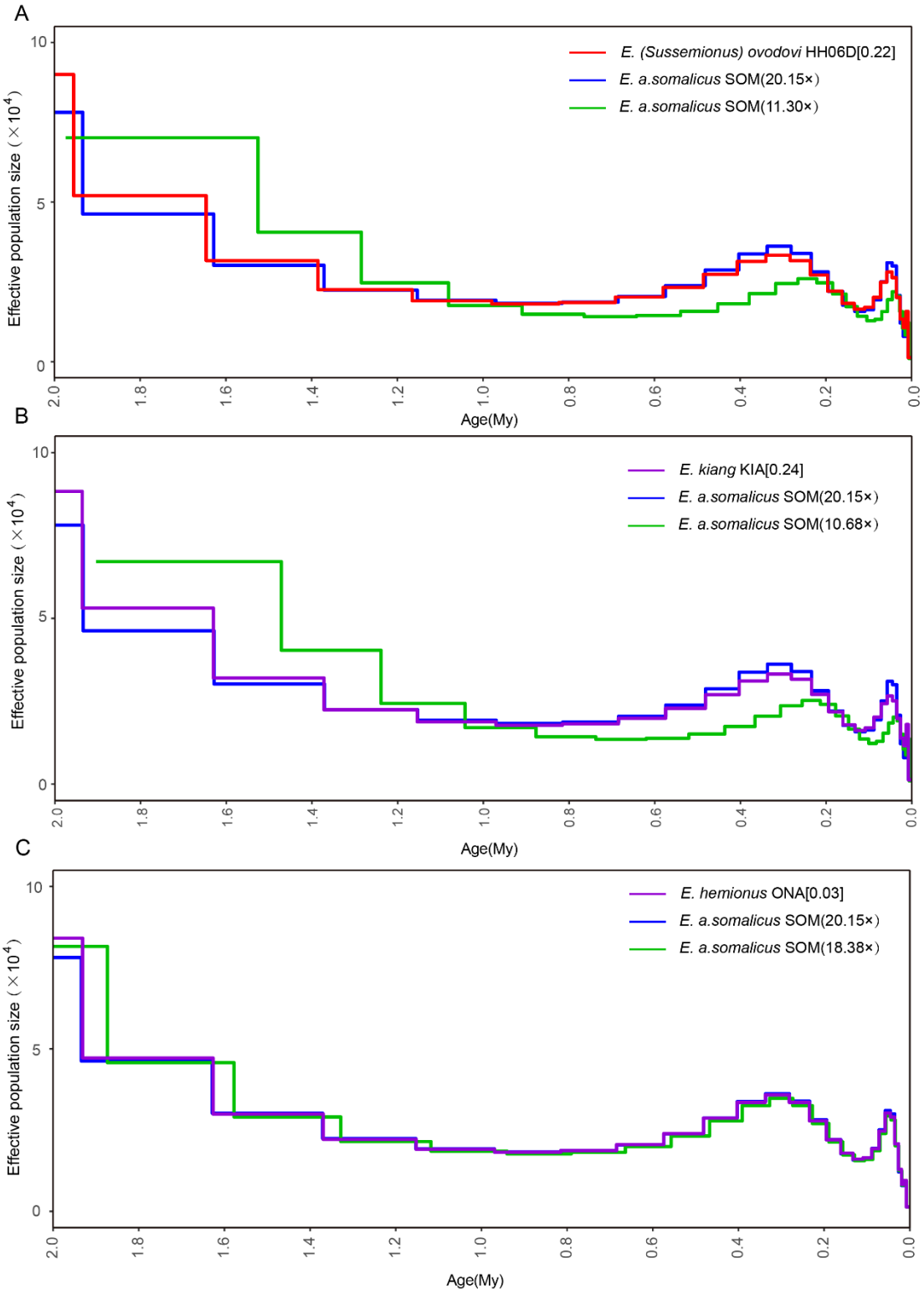


Figure 4—figure supplement 1. PSMC bootstrap pseudo-replicates for samples with (left) and without (right) transitions. (**A**) HH06D, (**B**) ONA, (**C**) KIA and (**D**) TWI. The *E. ovodovi* genome still included, even after rescaling, a significant proportion of nucleotide mis-incorporations pertaining to post-mortem DNA damage. This resulted in the presence of an excessive fraction of singleton mutations along this lineage, and the artefactual expansion observed in the most recent time range.



820

821 **Figure 4—figure supplement 2.** Determining the uniform false-negative rate (uFNR)

822 that was necessary for PSMC scaling. (**A**) HH06D (11.30×), (**B**) KIA (10.68×) and

823 (**C**) ONA (18.38×).

824

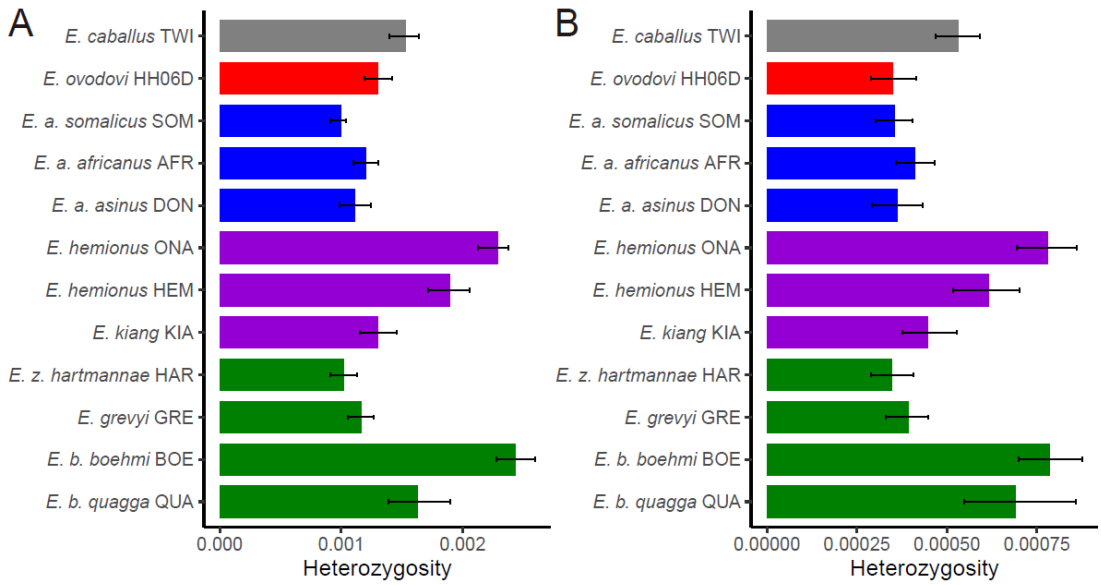


Figure 5—figure supplement 1. Heterozygosity rates outside Runs-of-Homozygosity (ROH) together with 95% confidence intervals. (**A**) Including transitions and (**B**) excluding transitions.

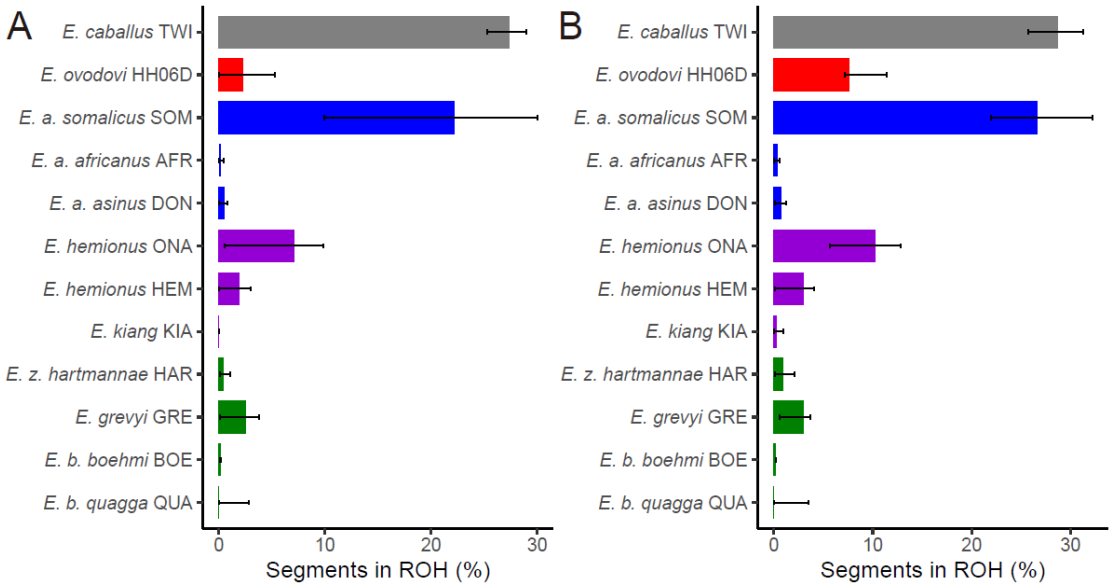


Figure 5—figure supplement 2. The fraction of the genome segments consisting of ROHs together with 95% confidence intervals. (**A**) Including transitions and (**B**) excluding transitions.

References

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchene, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kuhnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Muller, N. F., Ogilvie, H. A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M. A., Wu, C. H., Xie, D., Zhang, C., Stadler, T., & Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol, 15*(4), e1006650. doi:10.1371/journal.pcbi.1006650

Briggs, Adrian W., Stenzel, Udo, Johnson, Philip L. F., Green, Richard E., Kelso, Janet, Prüfer, Kay, Meyer, Matthias, Krause, Johannes, Ronan, Michael T., Lachmann, Michael, & Pääbo, Svante. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences, 104*(37), 14616-14621. doi:10.1073/pnas.0704665104

Clark, Peter U., Dyke, Arthur S., Shakun, Jeremy D., Carlson, Anders E., Clark, Jorie, Wohlfarth, Barbara, Mitrovica, Jerry X., Hostetler, Steven W., & McCabe, A. Marshall. (2009). The Last Glacial Maximum. *Science, 325*(5941), 710-714. doi:10.1126/science.1172873

Dawei, Cai, Lu, Han, Chengzhi, Xie, Shengnan, Li, Hui, Zhou, & Hong, Zhu. (2007). Mitochondrial DNA analysis of Bronze Age horses recovered from Chifeng region, Inner Mongolia, China. *Progress in Natural Science*.

Der Sarkissian, C., Vilstrup, J. T., Schubert, M., Seguin-Orlando, A., Eme, D., Weinstock, J., Alberdi, M. T., Martin, F., Lopez, P. M., Prado, J. L., Prieto, A., Douady, C. J., Stafford, T. W., Willerslev, E., & Orlando, L. (2015). Mitochondrial genomes reveal the extinct Hippidion as an outgroup to all living equids. *Biol Lett, 11*(3). doi:10.1098/rsbl.2014.1058

856 Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC*
857 *Evol Biol, 7*, 214. doi:10.1186/1471-2148-7-214
858 Druzhkova, Anna S., Makunin, Alexey I., Vorobieva, Nadezhda V., Vasiliev, Sergey K., Ovodov, Nikolai D.,
859 Shunkov, Mikhail V., Trifonov, Vladimir A., & Graphodatsky, Alexander S. (2017). Complete
860 mitochondrial genome of an extinct Equus (Sussemionus) ovodovi specimen from Denisova
861 cave (Altai, Russia). *Mitochondrial DNA Part B, 2*(1), 79-81.
862 doi:10.1080/23802359.2017.1285209
863 Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput.
864 *Nucleic Acids Res, 32*(5), 1792-1797. doi:10.1093/nar/gkh340
865 Eisenmann, V. (2010). Sussemionus, a new subgenus of Equus (Perissodactyla, Mammalia). *C R Biol,*
866 *333*(3), 235-240. doi:10.1016/j.crvi.2009.12.013
867 Fages, Antoine, Hanghøj, Kristian, Khan, Naveed, Gaunitz, Charleen, Seguin-Orlando, Andaine,
868 Leonardi, Michela, McCrory Constantz, Christian, Gamba, Cristina, Al-Rasheid, Khaled A. S.,
869 Albizuri, Silvia, Alfarhan, Ahmed H., Allentoft, Morten, Alquraishi, Saleh, Anthony, David,
870 Baimukhanov, Nurbol, Barrett, James H., Bayarsaikhan, Jamsranjav, Benecke, Norbert,
871 Bernáldez-Sánchez, Eloísa, Berrocal-Rangel, Luis, Biglari, Fereidoun, Boessenkool, Sanne,
872 Boldgiv, Bazartseren, Brem, Gottfried, Brown, Dorcas, Burger, Joachim, Crubézy, Eric,
873 Daugnora, Linas, Davoudi, Hossein, de Barros Damgaard, Peter, de los Ángeles de Chorro y de
874 Villa-Ceballos, María, Deschler-Erb, Sabine, Detry, Cleia, Dill, Nadine, do Mar Oom, Maria,
875 Dohr, Anna, Ellingvåg, Sturla, Erdenebaatar, Diimaajav, Fathi, Homa, Felkel, Sabine,
876 Fernández-Rodríguez, Carlos, García-Viñas, Esteban, Germonpré, Mietje, Granado, José D.,
877 Hallsson, Jón H., Hemmer, Helmut, Hofreiter, Michael, Kasparov, Aleksei, Khasanov, Mutalib,

878 Khazaeli, Roya, Kosintsev, Pavel, Kristiansen, Kristian, Kubatbek, Tabaldiev, Kuderna, Lukas,
879 Kuznetsov, Pavel, Laleh, Haeedeh, Leonard, Jennifer A., Lhuillier, Johanna, Liesau von
880 Lettow-Vorbeck, Corina, Logvin, Andrey, Lõugas, Lembi, Ludwig, Arne, Luis, Cristina, Arruda,
881 Ana Margarida, Marques-Bonet, Tomas, Matoso Silva, Raquel, Merz, Victor, Mijiddorj,
882 Enkhbayar, Miller, Bryan K., Mochlov, Oleg, Mohaseb, Fatemeh A., Morales, Arturo,
883 Nieto-Espinet, Ariadna, Nistelberger, Heidi, Onar, Vedat, Pálsdóttir, Albína H., Pitulko, Vladimir,
884 Pitskhelauri, Konstantin, Pruvost, Mélanie, Rajic Sikanjic, Petra, Rapan Papeša, Anita,
885 Roslyakova, Natalia, Sardari, Alireza, Sauer, Eberhard, Schafberg, Renate, Scheu, Amelie,
886 Schibler, Jörg, Schlumbaum, Angela, Serrand, Nathalie, Serres-Armero, Aitor, Shapiro, Beth,
887 Sheikhi Seno, Shiva, Shevnina, Irina, Shidrang, Sonia, Southon, John, Star, Bastiaan, Sykes,
888 Naomi, Taheri, Kamal, Taylor, William, Teegen, Wolf-Rüdiger, Trbojević Vukičević, Tajana, Trixl,
889 Simon, Tumen, Dashzeveg, Undrakhbold, Sainbileg, Usmanova, Emma, Vahdati, Ali,
890 Valenzuela-Lamas, Silvia, Viegas, Catarina, Wallner, Barbara, Weinstock, Jaco, Zaibert, Victor,
891 Clavel, Benoit, Lepetz, Sébastien, Mashkour, Marjan, Helgason, Agnar, Stefánsson, Kári, Barrey,
892 Eric, Willerslev, Eske, Outram, Alan K., Librado, Pablo, & Orlando, Ludovic. (2019). Tracking
893 Five Millennia of Horse Management with Extensive Ancient Genome Time Series. *Cell*.
894 doi:https://doi.org/10.1016/j.cell.2019.03.049

895 Gaunitz, C., Fages, A., Hanghoj, K., Albrechtsen, A., Khan, N., Schubert, M., Seguin-Orlando, A., Owens,
896 I. J., Felkel, S., Bignon-Lau, O., de Barros Damgaard, P., Mittnik, A., Mohaseb, A. F., Davoudi, H.,
897 Alquraishi, S., Alfarhan, A. H., Al-Rasheid, K. A. S., Crubezy, E., Benecke, N., Olsen, S., Brown,
898 D., Anthony, D., Massy, K., Pitulko, V., Kasparov, A., Brem, G., Hofreiter, M., Mukhtarova, G.,
899 Baimukhanov, N., Lougas, L., Onar, V., Stockhammer, P. W., Krause, J., Boldgiv, B., Undrakhbold,

S., Erdenebaatar, D., Lepetz, S., Mashkour, M., Ludwig, A., Wallner, B., Merz, V., Merz, I., Zaibert, V., Willerslev, E., Librado, P., Outram, A. K., & Orlando, L. (2018). Ancient genomes revisit the ancestry of domestic and Przewalski's horses. *Science, 360*(6384), 111-114. doi:10.1126/science.aao3297

Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G., & Siepel, A. (2011). Bayesian inference of ancient human demography from individual genome sequences. *Nat Genet, 43*(10), 1031-1034. doi:10.1038/ng.937

Heintzman, P. D., Zazula, G. D., MacPhee, R., Scott, E., Cahill, J. A., McHorse, B. K., Kapp, J. D., Stiller, M., Wooller, M. J., Orlando, L., Southon, J., Froese, D. G., & Shapiro, B. (2017). A new genus of horse from Pleistocene North America. *Elife, 6*. doi:10.7554/eLife.29944

Henn, Brenna M., Cavalli-Sforza, L. L., & Feldman, Marcus W. (2012). The great human expansion. *Proceedings of the National Academy of Sciences, 109*(44), 17758-17764. doi:10.1073/pnas.1212380109

Huang, J., Zhao, Y., Bai, D., Shiraigol, W., Li, B., Yang, L., Wu, J., Bao, W., Ren, X., Jin, B., Zhao, Q., Li, A., Bao, S., Bao, W., Xing, Z., An, A., Gao, Y., Wei, R., Bao, Y., Bao, T., Han, H., Bai, H., Bao, Y., Zhang, Y., Daidiikhuu, D., Zhao, W., Liu, S., Ding, J., Ye, W., Ding, F., Sun, Z., Shi, Y., Zhang, Y., Meng, H., & Dugarjaviin, M. (2015). Donkey genome and insight into the imprinting of fast karyotype evolution. *Sci Rep, 5*, 14106. doi:10.1038/srep14106

Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P. L., & Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics, 29*(13), 1682-1684. doi:10.1093/bioinformatics/btt193

921 Jonsson, H., Schubert, M., Seguin-Orlando, A., Ginolhac, A., Petersen, L., Fumagalli, M., Albrechtsen, A.,
922 Petersen, B., Korneliussen, T. S., Vilstrup, J. T., Lear, T., Myka, J. L., Lundquist, J., Miller, D. C.,
923 Alfarhan, A. H., Alquraishi, S. A., Al-Rasheid, K. A., Stagegaard, J., Strauss, G., Bertelsen, M. F.,
924 Sicheritz-Ponten, T., Antczak, D. F., Bailey, E., Nielsen, R., Willerslev, E., & Orlando, L. (2014).
925 Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proc Natl Acad*
926 *Sci U S A, 111*(52), 18655-18660. doi:10.1073/pnas.1412627111
927 Kalbfleisch, Theodore S., Rice, Edward S., DePriest, Michael S., Walenz, Brian P., Hestand, Matthew S.,
928 Vermeesch, Joris R., O′ Connell, Brendan L., Fiddes, Ian T., Vershinina, Alisa O., Saremi,
929 Nedda F., Petersen, Jessica L., Finno, Carrie J., Bellone, Rebecca R., McCue, Molly E., Brooks,
930 Samantha A., Bailey, Ernest, Orlando, Ludovic, Green, Richard E., Miller, Donald C., Antczak,
931 Douglas F., & MacLeod, James N. (2018). Improved reference genome for the domestic horse
932 increases assembly contiguity and composition. *Communications Biology, 1*(1), 197.
933 doi:10.1038/s42003-018-0199-z
934 Keane, Thomas M., Creevey, Christopher J., Pentony, Melissa M., Naughton, Thomas J., & McLnerney,
935 James O. (2006). Assessment of methods for amino acid matrix selection and their use on
936 empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC*
937 *Evolutionary Biology, 6*(1), 29. doi:10.1186/1471-2148-6-29
938 Khenzykhenova, Fedora I, Shchetnikov, Alexander A, Sato, Takao, Erbajeva, Margarita A, Semenei,
939 Elena Y, Lipnina, Ekaterina A, Yoshida, Kunio, Kato, Hirofumi, Filinov, Ivan I, & Tumurov, Erdem
940 G. (2016). Ecosystem analysis of Baikal Siberia using Palaeolithic faunal assemblages to
941 reconstruct MIS 3-MIS 2 environments and climate. *Quaternary international, 425*, 16-27.

942 Korneliussen, Thorfinn Sand, Albrechtsen, Anders, & Nielsen, Rasmus. (2014). ANGSD: Analysis of Next
943 Generation Sequencing Data. *BMC Bioinformatics, 15*(1), 356.
944 doi:10.1186/s12859-014-0356-4
945 Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: A fast, scalable, and
946 user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*.
947 doi:10.1093/bioinformatics/btz305
948 Kozlov, Alexey M., Aberer, Andre J., & Stamatakis, Alexandros. (2015). ExaML version 3: a tool for
949 phylogenomic analyses on supercomputers. *Bioinformatics, 31*(15), 2577-2579.
950 doi:10.1093/bioinformatics/btv184 %J Bioinformatics
951 Letunic, I, & Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and
952 annotation of phylogenetic and other trees. *Nucleic Acids Research, 44*(Web Server issue),
953 W242-W245.
954 Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform.
955 *Bioinformatics, 25*(14), 1754-1760. doi:10.1093/bioinformatics/btp324
956 Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome
957 sequences. *Nature, 475*(7357), 493-496. doi:10.1038/nature10231
958 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R.
959 (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16),
960 2078-2079. doi:10.1093/bioinformatics/btp352
961 Librado, Pablo, & Orlando, Ludovic. (2020). Genomics and the Evolutionary History of Equids. *Annual*
962 *Review of Animal Biosciences*. doi:10.1146/annurev-animal-061220-023118

963 Lindsay, Everett H., Opdyke, Neil D., & Johnson, Noye M. (1980). Pliocene dispersal of the horse Equus
964 and late Cenozoic mammalian dispersal events. *Nature, 287*(5778), 135-138.
965 Malikov, Dmitriy G. (2016). The large mammals of North-Minusinsk basin in the Last Glacial period.
966 *Quaternary international, 420*, 208-220. doi:10.1016/j.quaint.2015.10.055
967 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler,
968 D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce
969 framework for analyzing next-generation DNA sequencing data. *Genome Res, 20*(9),
970 1297-1303. doi:10.1101/gr.107524.110
971 Meisner, Jonas, & Albrechtsen, Anders. (2018). Inferring Population Structure and Admixture
972 Proportions in Low-Depth NGS Data. *Genetics, 210*(2), 719-731.
973 doi:10.1534/genetics.118.301336
974 Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E.,
975 Petersen, B., Moltke, I., Johnson, P. L., Fumagalli, M., Vilstrup, J. T., Raghavan, M.,
976 Korneliussen, T., Malaspinas, A. S., Vogt, J., Szklarczyk, D., Kelstrup, C. D., Vinther, J., Dolocan,
977 A., Stenderup, J., Velazquez, A. M., Cahill, J., Rasmussen, M., Wang, X., Min, J., Zazula, G. D.,
978 Seguin-Orlando, A., Mortensen, C., Magnussen, K., Thompson, J. F., Weinstock, J., Gregersen,
979 K., Roed, K. H., Eisenmann, V., Rubin, C. J., Miller, D. C., Antczak, D. F., Bertelsen, M. F., Brunak,
980 S., Al-Rasheid, K. A., Ryder, O., Andersson, L., Mundy, J., Krogh, A., Gilbert, M. T., Kjaer, K.,
981 Sicheritz-Ponten, T., Jensen, L. J., Olsen, J. V., Hofreiter, M., Nielsen, R., Shapiro, B., Wang, J., &
982 Willerslev, E. (2013). Recalibrating Equus evolution using the genome sequence of an early
983 Middle Pleistocene horse. *Nature, 499*(7456), 74-78. doi:10.1038/nature12323

984 Orlando, L., Metcalf, J. L., Alberdi, M. T., Telles-Antunes, M., Bonjean, D., Otte, M., Martin, F.,
985 Eisenmann, V., Mashkour, M., Morello, F., Prado, J. L., Salas-Gismondi, R., Shockey, B. J., Wrinn,
986 P. J., Vasil'ev, S. K., Ovodov, N. D., Cherry, M. I., Hopwood, B., Male, D., Austin, J. J., Hanni, C.,
987 & Cooper, A. (2009). Revising the recent evolutionary history of equids using ancient DNA.
988 *Proc Natl Acad Sci U S A, 106*(51), 21754-21759. doi:10.1073/pnas.0903672106
989 Orlando, Ludovic. (2020). The Evolutionary and Historical Foundation of the Modern Horse: Lessons
990 from Ancient Genomics. *Annual Review of Genetics, 54*(1), 563-581.
991 doi:10.1146/annurev-genet-021920-011805
992 Outram, A. K., Stear, N. A., Bendrey, R., Olsen, S., Kasparov, A., Zaibert, V., Thorpe, N., & Evershed, R. P.
993 (2009). The earliest horse harnessing and milking. *Science, 323*(5919), 1332-1335.
994 doi:10.1126/science.1168594
995 Palkopoulou, Eleftheria, Mallick, Swapan, Skoglund, Pontus, Enk, Jacob, Rohland, Nadin, Li, Heng,
996 Omrak, Ayça, Vartanyan, Sergey, Poinar, Hendrik, Götherström, Anders, Reich, David, & Dalén,
997 Love. (2015). Complete Genomes Reveal Signatures of Demographic and Genetic Declines in
998 the Woolly Mammoth. *Current Biology, 25*(10), 1395-1400.
999 doi:https://doi.org/10.1016/j.cub.2015.04.007
1000 Pickrell, Joseph K., & Pritchard, Jonathan K. (2012). Inference of Population Splits and Mixtures from
1001 Genome-Wide Allele Frequency Data. *PLOS Genetics, 8*(11), e1002967.
1002 doi:10.1371/journal.pgen.1002967
1003 Plasteeva, Natalya A. (2015). Equus (Sussemionus) ovodovi Eisenmann et Vasiliev, 2011 from the Late
1004 Pleistocene of Western Siberia. *Russian Journal of Theriology, 14*(2), 187-200.

1005 Renaud, G., Hanghoj, K., Korneliussen, T. S., Willerslev, E., & Orlando, L. (2019). Joint Estimates of
1006 Heterozygosity and Runs of Homozygosity for Modern and Ancient Samples. *Genetics*.
1007 doi:10.1534/genetics.119.302057
1008 Renaud, G., Petersen, B., Seguin-Orlando, A., Bertelsen, M. F., Waller, A., Newton, R., Paillot, R., Bryant,
1009 N., Vaudin, M., Librado, P., & Orlando, L. (2018). Improved de novo genomic assembly for the
1010 domestic donkey. *Sci Adv, 4*(4), eaaq0392. doi:10.1126/sciadv.aaq0392
1011 Rossel, Stine, Marshall, Fiona, Peters, Joris, Pilgram, Tom, Adams, Matthew D., & O'Connor, David.
1012 (2008). Domestication of the donkey: Timing, processes, and indicators. *Proc Natl Acad Sci U*
1013 *S A, 105*(10), 3715-3720. doi:10.1073/pnas.0709692105 %J Proceedings of the National
1014 Academy of Sciences
1015 Schubert, M., Ermini, L., Der Sarkissian, C., Jonsson, H., Ginolhac, A., Schaefer, R., Martin, M. D.,
1016 Fernandez, R., Kircher, M., McCue, M., Willerslev, E., & Orlando, L. (2014). Characterization of
1017 ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis
1018 using PALEOMIX. *Nat Protoc, 9*(5), 1056-1082. doi:10.1038/nprot.2014.063
1019 Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J. F., Al-Rasheid, K. A., Willerslev, E., Krogh, A., &
1020 Orlando, L. (2012). Improving ancient DNA read mapping against modern reference genomes.
1021 *BMC Genomics, 13*, 178. doi:10.1186/1471-2164-13-178
1022 Schubert, M., Jonsson, H., Chang, D., Der Sarkissian, C., Ermini, L., Ginolhac, A., Albrechtsen, A.,
1023 Dupanloup, I., Foucal, A., Petersen, B., Fumagalli, M., Raghavan, M., Seguin-Orlando, A.,
1024 Korneliussen, T. S., Velazquez, A. M., Stenderup, J., Hoover, C. A., Rubin, C. J., Alfarhan, A. H.,
1025 Alquraishi, S. A., Al-Rasheid, K. A., MacHugh, D. E., Kalbfleisch, T., MacLeod, J. N., Rubin, E. M.,
1026 Sicheritz-Ponten, T., Andersson, L., Hofreiter, M., Marques-Bonet, T., Gilbert, M. T., Nielsen, R.,

1027 Excoffier, L., Willerslev, E., Shapiro, B., & Orlando, L. (2014). Prehistoric genomes reveal the
1028 genetic foundation and cost of horse domestication. *Proc Natl Acad Sci U S A, 111*(52),
1029 E5661-5669. doi:10.1073/pnas.1416991111
1030 Schubert, M., Lindgreen, S., & Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming,
1031 identification, and read merging. *BMC Res Notes, 9*, 88. doi:10.1186/s13104-016-1900-2
1032 Seguin-Orlando, A., Gamba, C., Sarkissian, C., Ermini, L., Louvel, G., Boulygina, E., Sokolov, A.,
1033 Nedoluzhko, A., Lorenzen, E. D., Lopez, P., McDonald, H. G., Scott, E., Tikhonov, A., Stafford, T.
1034 W., Jr., Alfarhan, A. H., Alquraishi, S. A., Al-Rasheid, K. A. S., Shapiro, B., Willerslev, E.,
1035 Prokhortchouk, E., & Orlando, L. (2015). Pros and cons of methylation-based enrichment
1036 methods for ancient DNA. *Sci Rep, 5*, 11826. doi:10.1038/srep11826
1037 Shchetnikov, A. A., Klementiev, A. M., Filinov, I. A., & Semeney, E. Yu. (2015). Large mammals from the
1038 Upper Neopleistocene reference sections in the Tunka rift valley, southwestern Baikal Region.
1039 *Stratigraphy and Geological Correlation, 23*(2), 214-236. doi:10.1134/s0869593815020057
1040 Shunkov, M.V. (2018). The Denisova cave – everything changes, but nothing disappears. *Science First*
1041 *Hand*.
1042 Soraggi, Samuele, Wiuf, Carsten, & Albrechtsen, Anders. (2018). Powerful Inference with the
1043 D-Statistic on Low-Coverage Whole-Genome Data. *G3: Genes|Genomes|Genetics, 8*(2),
1044 551-566. doi:10.1534/g3.117.300192
1045 Stamatakis, Alexandros. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of
1046 large phylogenies. *Bioinformatics, 30*(9), 1312.

1047 Vasiliev, S. K. (2013). Large Mammal Fauna from the Pleistocene Deposits of Chagyrskaya Cave
1048 Northwestern Altai (based on 2007–2011 Excavations). *Archaeology, Ethnology and*
1049 *Anthropology of Eurasia, 41*(1), 28-44. doi:https://doi.org/10.1016/j.aeae.2013.07.003
1050 Vilstrup, J. T., Seguin-Orlando, A., Stiller, M., Ginolhac, A., Raghavan, M., Nielsen, S. C., Weinstock, J.,
1051 Froese, D., Vasiliev, S. K., Ovodov, N. D., Clary, J., Helgen, K. M., Fleischer, R. C., Cooper, A.,
1052 Shapiro, B., & Orlando, L. (2013). Mitochondrial phylogenomics of modern and ancient
1053 equids. *PLoS One, 8*(2), e55950. doi:10.1371/journal.pone.0055950
1054 vonHoldt, Bridgett M., Cahill, James A., Fan, Zhenxin, Gronau, Ilan, Robinson, Jacqueline, Pollinger,
1055 John P., Shapiro, Beth, Wall, Jeff, & Wayne, Robert K. (2016). Whole-genome sequence
1056 analysis shows that two endemic species of North American wolf are admixtures of the
1057 coyote and gray wolf. *Sci Adv, 2*(7), e1501714. doi:10.1126/sciadv.1501714
1058 Yang, Dongya Y., Eng, Barry, Waye, John S., Dudar, J. Christopher, & Saunders, Shelley R. (1998).
1059 Improved DNA extraction from ancient bones using silica-based spin columns. *American*
1060 *Journal of Physical Anthropology, 105*(4), 539-543.
1061 doi:10.1002/(sici)1096-8644(199804)105:4<539::Aid-ajpa10>3.0.Co;2-1
1062 Yuan, Jun-Xia, Hou, Xin-Dong, Barlow, Axel, Preick, Michaela, Taron, Ulrike H., Alberti, Federica, Basler,
1063 Nikolas, Deng, Tao, Lai, Xu-Long, Hofreiter, Michael, & Sheng, Gui-Lian. (2019). Molecular
1064 identification of late and terminal Pleistocene Equus ovodovi from northeastern China. *PLoS*
1065 *One, 14*(5), e0216883. doi:10.1371/journal.pone.0216883
1066